



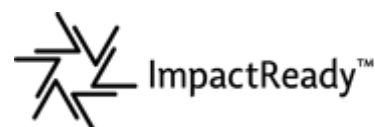
UNICEF GEROS Meta-Analysis 2015

*An independent review of UNICEF evaluation report quality
and trends, 2009-2015*

10 August 2016 – v2.1

Principal Authors: Joseph Barnes, Susie Turrall, Sara Vaca

Evaluation Manager: Ada Ocampo



Acronyms

CEE/CIS	Central and Eastern Europe/Commonwealth of Independent States (Regional Office)
COs	Country Offices
EAPRO	East Asia and Pacific Regional Office
EMOPs	Emergency Operations
EO	Evaluation Office
ESARO	East and Southern Africa Regional Office
EQA	Evaluation Quality Assessment
GEROS	Global Evaluation Report Oversight System
HQ	Headquarters
HRBAP	Human Rights Based Approach to Programming
LACRO	Latin America and Caribbean Regional Office
M&E	Monitoring and Evaluation
MENARO	Middle East and North Africa Regional Office
MTSP	Medium Term Strategic Plan
N/A	Not Applicable
OECD/DAC	Organisation for Economic Co-operation and Development/Development Assistance Committee
RBM	Results-based Management
ROs	Regional Offices
ROSA	Regional Office of South Asia
RTE	Real-time evaluation
SPOA	Strategic Plan Objective Area
SWAP	System-wide Action Plan
ToC	Theory of Change
TORs	Terms of Reference
UN	United Nations
UNDAF	United Nations Development Assistance Framework
UNEG	United Nations Evaluation Group
UNICEF	United Nations Children's Fund
WASH	Water, Sanitation and Hygiene
WCARO	West and Central Africa Regional Office

Cover photo: Ulrik Pedersen // ImpactReady



Executive Summary

Introduction

This review is a meta-analysis of the quality of the evaluation reports submitted to UNICEF's [Global Evaluation Reports Oversight System \(GEROS\)](#)¹ during 2015. It synthesizes results of 90 evaluation reports, reviewed by two independent consultancy teams. It shares findings on a global level, as well as highlighting trends across regions, sectors trends and quality assessment criteria. This report contributes to a wider body of knowledge, of similar GEROS meta-analysis reports produced each year since GEROS began in 2010.

The purpose of the meta-analysis is to contribute to achieving the three revised (2016) objectives of GEROS (particularly objective 1):

Objective 1: Enabling environment for senior managers and executive board to make informed decisions based on a clear understanding of the quality of evaluation evidence and usefulness of evaluation reports;

Objective 2: Feedback leads to stronger evaluation capacity of UNICEF and partners;

Objective 3: UNICEF and partners are more knowledgeable about what works, where and for who.

GEROS is underpinned by United National Evaluation Group (UNEG) norms and standards, UN System Wide Action Plan on gender equality (UN SWAP) and other UNICEF-adapted standards, including equity and human-rights based approaches. The system consists of rating evaluation reports commissioned by UNICEF Country Offices, Regional Offices and HQ divisions. All reports and the results of their quality assessment are made available in the UNICEF Global Evaluation and Research Database (ERDB), as well as made publicly available on the UNICEF external website. GEROS is an organization-wide system.

Methodology

Quality reviews of evaluations completed in 2015 were carried out over a 14-month period, from January 2015 to February 2016. Two separate review teams undertook reviews of 2015 evaluations, Universal Management Group (for reports submitted to onto the ERDB from January to December 2015) and ImpactReady LLP (for reports submitted to onto the ERDB from January to February 2016). Both teams included evaluation experts with a broad range of relevant sectoral knowledge and linguistic capabilities (English, French and Spanish). Reviews were delegated within teams according to 'fit' with thematic and language expertise.

Evaluation quality assessment was carried out for each report using 64 questions and a scale of four ratings: Outstanding, Highly Satisfactory, Mostly Satisfactory, and Unsatisfactory. The meta-analysis was conducted once all of the evaluation reports had been assessed, submitted to UNICEF EO and accepted. Quantitative data was compiled regarding scores for different aspects of the reports using Excel. Qualitative analysis of reviewer comments was used to explore any causal links. In addition, the reviews were searched to explore good practice from the reports. Particular attention was given to SWAP reporting. Quantitative and qualitative data

¹ <https://icon.unicef.org/apps02/cop/edb/SitePages/Home.aspx>



were triangulated, and compared with longitudinal data on findings from four previous years to map key trends and patterns.

For the purposes of efficiency, the meta-analysis unit of assessment for quality is an evaluation report – a proxy of overall evaluation quality used by many UNEG members and SWAP. Whilst this serves the purpose of assessing quality, it is inevitably not a complete picture of evaluation quality and this limitation must be considered in interpreting the implications of findings.

Findings

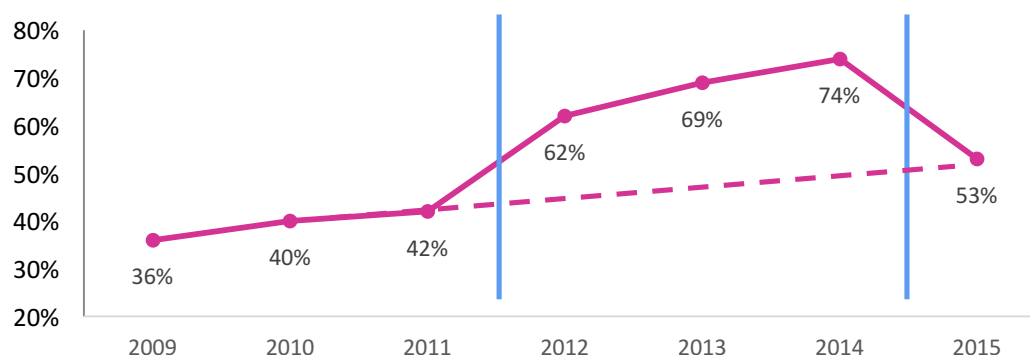
There has been a sharp increase in evaluation reports submitted, which marks a return to pre-2014 levels. Regional offices in Africa submitted the most reports. More reports were submitted from education and child protection sectors than other sectors.

The number of reports reviewed in 2015 was 90, a significant increase from the 69 reports reviewed the previous year. This marks a return to the levels of evaluations seen during the 2011-13 period (85-96 evaluation reports submitted per year).

For the purpose of the meta-analysis, 'Good quality' reports are classified as those that are 'outstanding' or 'highly satisfactory', which constitutes just over half (54%) of the reports.

When analysed against previous years' data, it can be seen that there has been a steady increase in quality from 2009 to 2011, a sharp increase in 2012 and then a steady increase until 2014. 2015 represents a sharp year-on-year decrease from 74% to 53%. Both the sharp increase (2012) and decrease (2015) correlate with changes in service providers (as illustrated on the graph below). Seen as part of a long-term trend, however, there is a steady improvement in evaluation report quality since 2009 and other evidence suggests that changes in performance are more feasibly explained by the proportion of outcome-level evaluations.

Figure: Long term trends in evaluation report quality 2009-2015 showing phases in different GEROS service providers



In terms of the quality of the reports across regions, ROSA has retained the same percentage of good quality reports, whereas some regions have gone down (HQ², CEE/CIS, LACRO and EAPRO) and some have gone up (ESARO, MENARO and WCARO).

Overall the strongest sections of reports in 2015 were 'Purpose, objectives and scope' (Section B) and 'Report is well structured, clear and logical' (Section F): both of which are

² HQ reports included 2 Programme Division evaluations for the first time (previously only evaluations from the Evaluation Office were rated).



directly influenced by evaluation Terms of Reference. ‘Findings and Conclusions’ (Section D) was also rated relatively strongly. The recommendations and lessons learned section (Section E) receive the lowest rating.

A notable pattern for 2015 is a sharp drop year-on-year in the quality of the methodology section from 71% in 2014 to 56% in 2015 (but this is an improvement on 2013 – 55%). Section D findings and conclusions also reduced in quality, returning to its 2013 level. Other areas show a decline, albeit less notable, from 2014 levels.

The review identified the following regional evaluation strengths and areas for improvement in 2016 based on quantitative analysis of sub-criteria data.

Table: Relative strengths and weaknesses of 2015 evaluation sub-sections

Strengths	Weaknesses
Purpose, objectives and scope	Theory of Change
Data collection	Stakeholder Participation
Assessing strengths & weaknesses of the evaluation object	Lessons Learned

Quantitative analysis of alignment with Strategic Plan Objective Areas (including cross-cutting issues) found that evaluations of humanitarian action (as a cross cutting issue) and WASH rated most strongly (although with small sample sizes). Other areas in which a majority of reports are rated as meeting UNICEF standards are child protection, HIV and AIDS, social inclusion, and multi-sector evaluations. Health and education (both with large sample sizes) rate lower. Nutrition and gender equality areas both had very small samples (1).

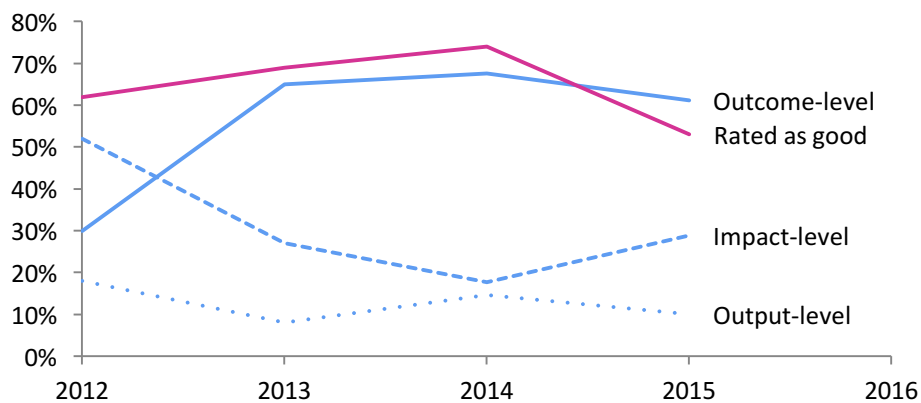
UN SWAP evaluation performance indicator criteria were assessed according to the standards and scoring system established by UNEG. The aggregated average score for 2015 was 6.3, which is classified as **Approaching Requirements**. This represents a year-on-year improvement, with UNICEF reporting a rating of 6 (Approaching Requirements) in 2014. According to UNEG reporting cycle for 2015 this rating is consistent with similar entities that are subject to independent external review or peer review, including UNDP, UN Women and UNESCO.

A review of the historical data on the classifications of evaluations does reveal a potential explanatory factor for the overall performance trend. The percentage of reports rated as meeting UNICEF/UNEG standards appears to be proportional to the percentage of reports that are outcome-level evaluations³. The implication is that a more diverse portfolio of evaluation types makes it more difficult to consistently deliver reports to UNEG standards. The portfolio of reports in terms of levels of evaluation was more diverse in 2015 than in 2014, helping to explain the performance pattern identified in GEROS 2015.

³ **Output:** Causal effects deriving directly from programme activities, and assumed to be completely under programme control; **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; **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.



Figure: Percentage of evaluation reports rated as good vs the result-level of evaluations, 2012-2015



Conclusions

Conclusion 1: The strengths and weaknesses of evaluation reports remain similar to previous years. Elements of the evaluation that are influenced by the ToR (purpose and objectives) are an organisational strength, whilst areas for improvement include improving theories of change, stakeholder participation, and lessons learned.

The findings of the meta-analysis continue the pattern of the preceding meta-evaluations. Comparative strengths are seen in the purpose and objectives of evaluations (both of which are established by the terms of reference). Evaluation reports are also stronger in terms of describing the object and context, and the specification of relevant methodologies (although this saw some decline in 2015).

In contrast to these strengths are persistent areas of weakness in evaluation reports over the years that GEROS has been generating evidence. These include presenting and using theories of change; the mapping and participation of stakeholders; including cost analyses; applying ethics standards; integrating human rights (including gender and equity) especially in terms of analysis; ensuring usefulness of recommendations; and identifying lessons learned that are clear and generalizable.

It is possible to conclude from these observations that the current process of commissioning and managing evaluations is insufficient in terms of addressing persistent areas of weakness. It is proposed that three issues represent particularly acute priorities for action: 1) establishing and critiquing theories of change, 2) enhancing stakeholder participation in evaluations, and 3) enhancing the quality of evaluation lessons to generate new knowledge.

Conclusion 2: The variations in the proportion of reports that are rated as meeting UNEG standards is best explained in terms how diverse the overall evaluation portfolio is. Greater diversity in terms of the types of evaluation being undertaken seems to slow – or even regress – the rate at which the quality of reports improves over time. The long-term trend, however, remains one of improvement in quality.

The headline result from the meta-analysis appears to show that, after a six-year run of constant improvement, the proportion of evaluation reports meeting UNEG standards dropped for the first time, with a year-on-year decline of 21 percentage points (although this includes relatively large variations across the individual regions). However, this substantive drop in the

overall rating may hide a number of explanatory factors that require a more nuanced reading of the data.

The findings identified three possible causal stories that might explain the performance of evaluation reports in 2015. The most plausible of these explanations emerges from an analysis of trends in the diversity of evaluation reports using the proxy of the 'result-level' (output, outcome, impact) that evaluations seek to assess. This finds that the proportion of reports that are aimed at evaluating outcome-level changes is proportional to the overall quality rating.

The implication of this conclusion for UNICEF is that additional support may be required to ensure that evaluations of higher levels of complexity (such as joint evaluations, impact evaluations and country-led evaluations) reach the same level of report quality as the more-frequent programme/outcome evaluations do. It also highlights the importance of interpreting the 'immediate' feedback of GEROS through a more nuanced and long-term lens that takes into account changes in the characteristics of the evaluation portfolio.

Conclusion 3: UNICEF is approaching UN SWAP commitments with regard to integration of gender equality, but significant scope remains for enhancing the use of gender analysis in developing findings, conclusions and recommendations. This places UNICEF in a similar position to its comparable sister agencies.

Whilst 62 per cent of reports were found to have specifically addressed the issue of equity (as required by the UNICEF standards), the integration of gender into evaluation designs and methods remains inconsistent. The entire portfolio of reports was found to meet SWAP requirements in terms of evaluation scope, indicators, criteria, questions, and methods. However, a gap in the integration of gender – and the cause of an overall rating of 'approaching requirements' – is insufficient evidence of gender analysis being used to form findings, conclusions and recommendations.

Shortfalls in performance with regard to this criterion are closely related to the earlier conclusion that stakeholder mapping and participation is a critical gap in UNICEF's current evaluation practice. As a first step, promoting stakeholder mapping through the application of human rights role analysis, power analysis, and similar tools available within UNICEF's guidance is thus considered to be a significant priority if SWAP standards are to be conclusively met by 2018.

Conclusion 4: Whilst there are recurrent shortfalls in overall report quality, a wide range of evaluative capacities are also evident from examples of high quality reports. However, these capacities seem to be available only in specific regions and strategic plan objective areas – suggesting a strong need for better internal learning and knowledge exchange in UNICEF.

As previously concluded, a number of dimensions are persistently weak in UNICEF evaluations, and the systems intended to strengthen these aspects are therefore insufficient. However, it was also found that these overall gaps are not weak in all cases. For example, cost analysis (which is only considered adequate in 51% of reports) is addressed strongly in evaluations of health – with 81% rated as good. Similarly, LACRO is notable for developing strong recommendations and ROSA for developing strong evaluation frameworks, neither of which are universal strengths of the evaluation function.

This combination of findings suggests that a wide range of evaluative capabilities is available within UNICEF, but it can be concluded from the uneven distribution of quality that these



capabilities exist in relative isolation. The implication of this conclusion is that additional scope exists to enhance cross-fertilisation of evaluation experience, capacity and knowledge within UNICEF as a means to target organisational capability gaps.

Recommendations

Recommendation 1: Review the strategy for evaluation systems-strengthening to prioritise enhancements to the quality of less-frequent and complex types of decentralised evaluations.

EVALUATION OFFICE

This meta-analysis has concluded, based on longitudinal analysis of data from the GEROS system, that a decrease in the proportion of outcome-level evaluations and their replacement with less-frequent types of evaluations leads to a drop in the number of reports that reach UNEG standards. It is hypothesised that this is a result of organisational capacity and experience being concentrated around programme and outcome evaluations.

Given the benefit to UNICEF's knowledge-base of increasing the number and quality of diverse evaluations (such as country-led evaluations, joint evaluations, and impact evaluations), in addition to the commitment to ensure that all evaluations reach UNEG standards, it is recommended that priority is given in systems-strengthening efforts to enhance the capacity of regional and country offices to support more diverse types of evaluations. This recommendation is also considered to be relevant in view of developing the decentralised capabilities needed to support the UNEG commitment to enhancing national evaluation capacity in order to meet the Global Goals.

Recommendation 2: Ensure evaluators clearly elaborate comprehensive stakeholder mapping, analysis of human rights roles (e.g. duty bearers and rights holders), and examination of an Object's theory of change within the evaluation inception report.

EVALUATION MANAGERS

The evidence from 2015 and previous years suggests that the aspects of evaluation that are primarily determined at the TOR stage (including purpose and objectives) are consistently among the stronger-performing aspects of evaluation reports. After the TORs, the next evaluation stage and product is the inception phase and report. The purpose of this phase is to outline the evaluation approach in full detail and the production (and approval) of an inception report acts as a quality control point. It is recommended that all evaluation managers take full advantage of the inception phase and inception report to ensure that persistent gaps in evaluation quality are sufficiently addressed before proceeding to data collection.

Based on the findings and conclusions of this meta-analysis and previous meta evaluations, it is recommended that priority is given to ensuring the quality of following three elements of an inception report:

- ❖ Detailed elaboration of the Object's theories of change – at a minimum a presentation and explanation of the logframe, but ideally a theory-based analysis of the underlying assumptions used to design an intervention;
- ❖ Comprehensive stakeholder mapping that identifies social role groups and not just institutions – ideally by applying human rights role analysis, gender role analysis, power mapping, and similar techniques;



- ❖ Elaboration of how inclusion and participation will be ensured during the evaluation process for each of the identified stakeholder groups, including how ethical standards will be implemented in concrete terms.

Implementing this recommendation will help to ensure that data is generated which is of sufficient quality to undertake robust evaluative analysis, including gender analysis to the required SWAP standards.

Recommendation 3: Review the strategy for internal learning and knowledge sharing – especially for evaluation focal persons – to focus on addressing the persistent performance gaps identified through the GEROS process.

EVALUATION OFFICE AND REGIONAL OFFICES

So as to take advantage of the evaluation capabilities within the organisation, it is recommended to revisit relevant global and regional learning and knowledge management strategies for evaluation so as to give greater emphasis to the persistent performance gaps identified by GEROS. For example, global and regional internal evaluation learning events (such as the webinar on developing recommendations that has previously been held) could be themed based on issues that represent persistent gaps in UNICEF's evaluation capability. These should seek to draw on colleagues' experiences and knowledge where there is evidence of high-performing examples in isolated cases – such as cost analysis in the health sector. Priority themes include: undertaking cost analysis in the context of social programmes, developing relevant lessons learned, integrating human rights, gender and equity into evaluation, applying ethical standards, and reconstructing and critiquing theories of change.

Recommendation 4: In order to meet SWAP standards by 2018, UNICEF needs to prioritise the inclusion of gender responsive evaluation frameworks, methods and analysis in all evaluations through: 1) increasing awareness of UNICEF staff around the SWAP evaluation performance indicators, 2) specifying gender requirements in all evaluation ToRs, and 3) include assessment of SWAP in all regional evaluation helpdesks.

REGIONAL OFFICES AND EVALUATION MANAGERS

UNICEF is currently rated as 'Approaching Requirements' for the System Wide Action Plan for gender equality (SWAP) evaluation performance indicator, and has committed to fully meet the required standard by 2018. To assist in meeting the required standard – as independently assessed by GEROS – a detailed list is provided on page 41 of the full report of concrete features that will ensure future reports meet or exceed UNEG standards for SWAP. It is recommended that this list is used by evaluation managers for quality assurance and shared with evaluators to improve the performance of UNICEF evaluations in regards to SWAP. Doing so will also contribute to enhancing the performance of evaluation reports with regard to equity.



Contents

Acronyms.....	2
Executive Summary	3
Introduction.....	3
Methodology.....	3
Findings.....	4
Conclusions.....	6
Recommendations.....	8
Introduction.....	12
Background to GEROS	13
Purpose, Scope and Objective of GEROS	13
Methodology.....	14
Overview of evaluation reports included in the meta analysis.....	14
Limitations	17
Findings.....	18
Overall findings on evaluation report quality	18
Areas of progress and priorities for improvement.....	21
Overall regional trends	30
Overall thematic performance	31
UN SWAP performance and trends.....	32
Other observable trends.....	34
Conclusions.....	36
Recommendations	39
Good practices	42
Regional performance summaries	45
EAPRO.....	45
ESARO.....	46
CEE/CIS.....	47
LACRO.....	48
Other (HQ).....	49
MENARO.....	50
ROSA.....	51
WCARO.....	52
Appendices.....	53
Annex 1. Terms of Reference.....	53
Annex 2. GEROS Methodology.....	57
Annex 3. GEROS evaluation quality assessment indicators.....	61
Annex 4. List of reports assessed	71
Annex 5. Quality of evaluation sub-sections	78
Annex 6: UN SWAP Calculations	79

Tables

Table: Relative strengths and weaknesses of 2015 evaluation sub-sections	5
Table 1: Number of reports included in GEROS, 2011-2015	15
Table 2: Classification of reports relating to humanitarian action in GEROS 2015	16
Table 3: Priority areas for action to strengthen evaluation identified in meta evaluations, 2012-2014	22
Table 4: Percentage of evaluation reports meeting UNICEF standards for each objective and cross-cutting area.....	32
Table 5: Performance according to UN SWAP evaluation criteria, 2015	33



Table 6: Notable evaluation practices identified in 2015 evaluations42
 Table A1: Number of reports included in GEROS, 2011-2015.....57
 Table A2: Qualitative ratings categories used in GEROS 201558
 Table A3: UN SWAP ratings specified by UNEG58
 Table A4: Number of GEROS reviews undertaken by independent GEROS providers.....60

Figures

Figure: Long term trends in evaluation report quality 2009-2015 showing phases in different GEROS service providers4
 Figure: Percentage of evaluation reports rated as good vs the result-level of evaluations, 2012-2015.6
 Figure 1: Number of reports submitted to GEROS per zone..... 15
 Figure 2: Trend in assessment of evaluation reports 2011-2015 16
 Figure 3: Distribution of the main unit of analysis in reports by Strategic Plan Objective Areas and cross-cutting issues..... 17
 Figure 4: Overall distribution of quality ratings for 2015 evaluation reports 18
 Figure 5: Long term trends in evaluation report quality 2009-2015 showing phases in different GEROS service providers 19
 Figure 6: Trends 2011-2015 in the percentage of good evaluation reports from different zones.....20
 Figure 7: Performance of quality assessment sections 2011-2015.....21
 Figure 8: Quality ratings for Section A (Object and context) 2011-201523
 Figure 9: Quality ratings for Section B (purpose, objectives and scope) 2011-201524
 Figure 10: Quality ratings for Section C (methods, gender, human rights and equity) 2011-201525
 Figure 11: Long term patterns in the inclusion of human rights, gender equality and equity (reports rated good)26
 Figure 12: Quality ratings for Section D (findings and conclusions) 2011-2015.....27
 Figure 13: Quality ratings for Section E (recommendations and lessons learned) 2011-201529
 Figure 14: Quality ratings for Section F (structure, logic and clarity) 2011-201530
 Figure 15a: Cumulative percentage of reports rated good in each region, 2011-201531
 Figure 15b: Cumulative number of reports rated good in each region, 2011-2015.....31
 Figure 16: Performance of UNICEF reports across the SWAP criteria33
 Figure 17: Percentage of evaluations rated good for different geographic scopes (blue) and management arrangements (pink)34
 Figure 18: Percentage of different types of evaluation rated good.....34
 Figure 19: Percentage of evaluations rated good for different result levels (blue) and purposes (pink)35
 Figure 20: Percentage of evaluation reports rated as good vs the result-level of evaluations, 2012-201535



Introduction

This review is a meta-analysis of the quality of the evaluation reports submitted to UNICEF's Global Evaluation Reports Oversight System (GEROS) during 2015. It synthesizes results of 90 evaluation reports, reviewed by two independent consultancy teams. It shares findings on a global level, as well as highlighting trends across regions, sectors trends and quality assessment criteria.

This report contributes to a wider body of knowledge, of similar GEROS meta-analysis reports produced each year since GEROS began in 2010.

The report was commissioned by UNICEF Evaluation Office. The key audiences are UNICEF's evaluation office, regional offices and evaluation managers. It will be used to monitor progress, analyse strengths and identify current challenges to improve evaluations.

The report is structured in the following way:

- ❖ Background to GEROS
- ❖ Methodology
- ❖ Summary of key findings from previous years
- ❖ Findings
- ❖ Conclusions
- ❖ Recommendations
- ❖ Good practices



Background to GEROS

UNICEF's evaluation system is decentralised, which reflects the decentralised nature of the organisation. Evaluation, according to the 2013 Evaluation Policy, is a shared function, with key roles distributed across senior leaders and oversight bodies, heads of offices, sector programme staff and technical evaluation staff. UNICEF's evaluation office and regional offices collaborate in order to strengthen the organisation's evaluation function.

While the decentralized nature of the evaluation function ensures that evidence generated is relevant to the local context, it poses the challenge of setting up a consistent system to ensure good quality and credibility. UNICEF's Evaluation Office (EO) set up the Evaluation Quality Oversight System (GEROS) in 2010 to assess quality of evaluations and further inform developments of the organisation's evaluation function.

The GEROS process includes review of the quality of the evaluation reports and a meta-analysis each year by an external independent team. UNICEF adapted UNEG standards are used as the criteria of quality assessment (Annex 2). Annual Meta Evaluation Reports have been produced each year since 2009. In 2015 an independent review of GEROS was undertaken with the findings and recommendations of this review being incorporated into the (forthcoming) 2016 cycle.

Purpose, Scope and Objective of GEROS

The purpose of GEROS is to support strengthening of the evaluation function to meet UNEG standards, ensure accountability, and promote use of robust evaluative evidence.

As noted above, GEROS was subject to an independent review in 2015 that made recommendations for refining the system. This included an adjustment of the existing objectives to be more results-based, which was implemented in March 2016.

The purpose of this meta-analysis is to support the three revised objectives of GEROS (particularly objective 1):

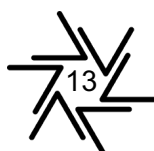
Objective 1: Enabling environment for senior managers and executive board to make informed decisions based on a clear understanding of the quality of evaluation evidence and usefulness of evaluation reports

Objective 2: Feedback leads to stronger evaluation capacity of UNICEF and partners

Objective 3: UNICEF and partners are more knowledgeable about what works, where and for who.

GEROS is underpinned by United National Evaluation Group (UNEG) standards, UN System Wide Action Plan on gender equality (UN SWAP) and other UNICEF-adapted standards, including equity and human-rights based approaches. The system consists of rating evaluation reports commissioned by UNICEF Country Offices, Regional Offices and HQ divisions. All reports and the results of their quality assessment are made available in the UNICEF Global Evaluation and Research Database (ERDB), as well as made publicly available on the UNICEF external website. GEROS is an organization-wide system.

The 2015 assessment analysed the extent to which the GEROS has achieved its expected objectives; determined the conceptual clarity of the GEROS approach and the adequacy of its methodology and tools; and identify constraining and enabling factors for an effective system



implementation. The assessment proposed recommendations and lessons to inform the forthcoming cycles of GEROS implementation that will take effect from the 2016 cycle of evaluation quality assurance.

Methodology

This meta-analysis was conducted in May 2016 once all of the evaluation reports had been assessed, submitted to UNICEF EO and accepted. Quantitative data was compiled regarding scores for different aspects of the reports using Excel. Analysis was carried out across multiple axes:

- ❖ Regional trends (regional and country levels)
- ❖ Trends by quality assessment criteria (*including across time*)
 - Object of the evaluation
 - Evaluation purpose, objectives and scope
 - Evaluation Methodology, Gender, Human Rights And Equity
 - Findings and Conclusions
 - Recommendations and Lessons Learned
 - Report Structure, Logic and Clarity
- ❖ Type of management
- ❖ Purpose
- ❖ Scope
- ❖ Results level
- ❖ Strategic Plan Objective Area correspondence
- ❖ UN SWAP performance and trends since 2013 (when it was introduced)

Qualitative analysis of reviewer comments was used to explore any causal links. In addition the reviews were trawled to explore good practice from the reports. Quantitative and qualitative data were triangulated, and compared with longitudinal data on findings from four previous years to map key trends and patterns.

A full explanation of the methodology is included in Annex 2.

Overview of evaluation reports included in the meta analysis

Evaluation quality reviews of evaluations commissioned in 2015 were carried out over a 14-month period, from January 2015 to February 2016. Two separate review teams undertook reviews of 2015 evaluations, Universal Management Group (January-December 2015) and ImpactReady LLP (January-February 2016). Both teams included evaluation experts with a broad range of relevant sectoral knowledge and linguistic capabilities (English, French and Spanish). Reviews were delegated within teams according to 'fit' with thematic and language expertise.

The list of evaluation reports to be reviewed was finalised by the UNICEF Evaluation Office, which involved monitoring the Evaluation and Research Database and undertaking an initial filtering to ensure that all reports were correctly classified as evaluations. In total, 90 evaluation reports were reviewed for 2015, which was a return to similar levels as 2011-13 (there was drop in 2014 to 69 reports).



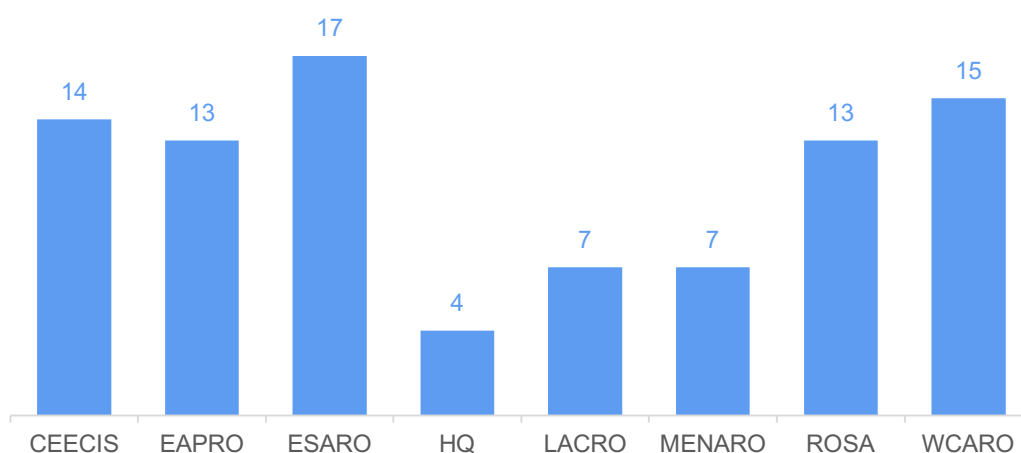
There has been a sharp increase in evaluation reports rated, which marks a return to pre-2014 levels. In 2014, the number of reports reviewed had dropped considerably, decreasing from 96 to 69. The previous year (2013) there had been 79 rated reports, and in the range of 80-100 reports each year prior to that. For 2015, 100% of evaluation reports submitted to the Evaluation and Research Database have been reviewed. Regional offices in Africa submitted the most reports. More reports were submitted from education and child protection sectors than other sectors.

Table 1: Number of reports included in GEROS, 2011-2015

Year	Number of Reports Reviewed
2011	88
2012	85
2013	96
2014	69
2015	90

The number of reports submitted by each region varies, as shown in the graph below. In 2015 more reports were submitted from the East and Southern Africa Regional Office (17 reports), and the West and Central Africa Regional Office (15): collectively there were more from Africa than other continents. The Central and Eastern Europe and Commonwealth of Independent States (CEE/CIS) submitted a few less (14 reports), as did the East Asia and Pacific Regional Office (EAPRO) and Regional Office of South Asia (who each submitted 13 reports). There were fewer reports from the Latin America and Caribbean Regional Office (LACRO) and Middle East and North Africa Regional Office, who each submitted 7 reports. The least reports were submitted from the headquarters, that submitted 4 reports overall – 2 from the Evaluation Office (corporate evaluations) and 2 from HQ units.

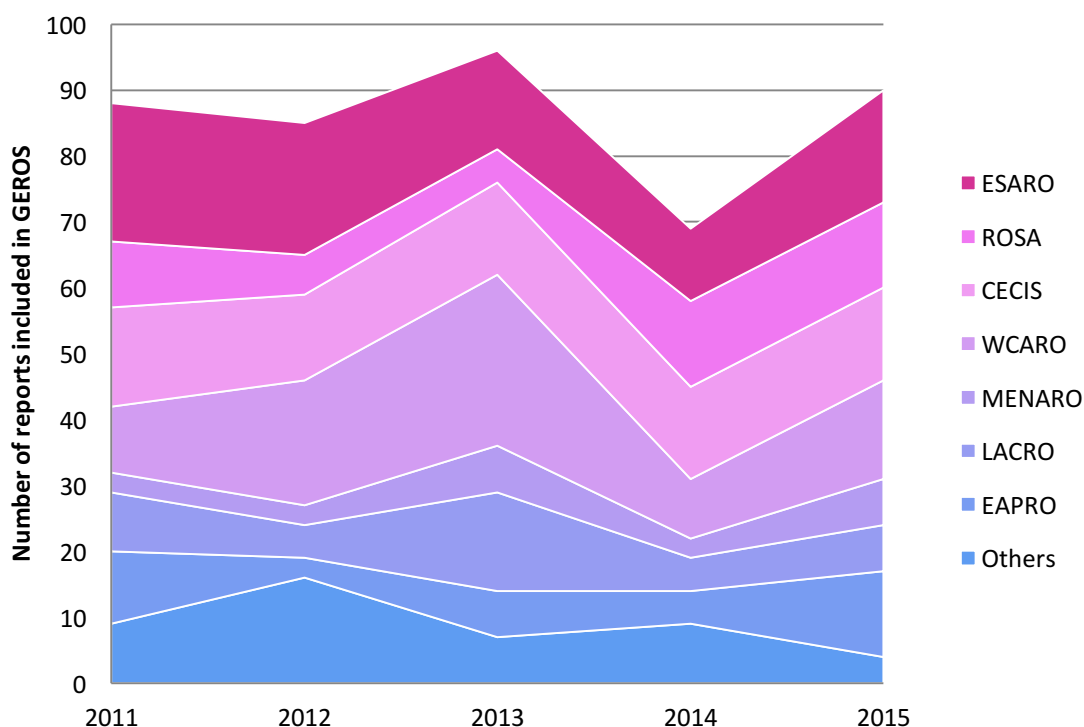
Figure 1: Number of reports submitted to GEROS per zone



Analysis of the data from previous years shows a few trends. Submissions from EAPRO have increased from 2014 and returned to 2011 levels. LACRO submitted a similar amount to 2014, which is marginally less than previous years. WCARO has submitted more this year than last year, but still not as many as during 2012 and 2013. The number of evaluation reports

submitted from CEE/CIS is fairly constant. ROSA has submitted more reports this year than previous years, and ESARO has increased again, back to 2011 and 2012 levels.

Figure 2: Trend in assessment of evaluation reports 2011-2015



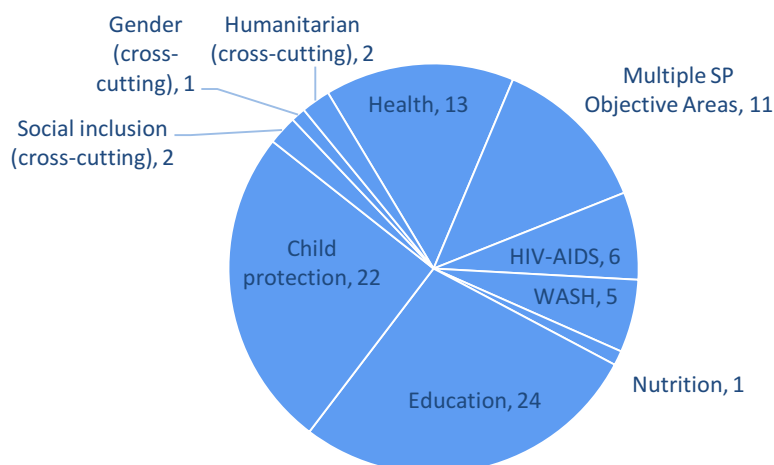
In terms of the sector focus of the object of the evaluations - or correspondence to UNICEF's Strategic Plan Objective Area (SPOA) - the chart below illustrates that in 2015 the sector that submitted more reports was education with 28% of reports, which was marginally higher than child protection (25%). Health submitted 15% of reports. Whilst multiple evaluations included coverage of humanitarian action and social inclusion issues, for each of these categories only 2% were classified as being exclusively about humanitarian action as a cross-cutting issue⁴ or inclusion-related for the purposes of GEROS (see Table 2 regarding humanitarian evaluations).

Table 2: Classification of reports relating to humanitarian action in GEROS 2015

Label in meta-analysis	Description
Evaluations of humanitarian action	Evaluations with humanitarian action as the main thematic unit of analysis (cross-cutting)
Humanitarian evaluations	Evaluations with humanitarian-specific designs or methods (type)

⁴ Additional evaluations may have been undertaken in a humanitarian context or mainstreamed humanitarian issues.

Figure 3: Distribution of the main unit of analysis in reports by Strategic Plan Objective Areas and cross-cutting issues



Limitations

The GEROS team had to address some challenges during the process of reviewing the reports. Although there are objective standards within GEROS reporting, it is assessed by individuals who use their best professional judgement. Furthermore, the change in service provider may have added to this, although four of the reviewers had been on the review teams from previous providers.

The independent review of GEROS also found weaknesses in the current tool, including the process by which indicators and sections were weighted to provide an overall rating. This tool is being revised for 2016.

Various quality assurance mechanisms were put in place to reduce variation as discussed above, including the role of the service provider's coordinator as 'Master Reviewer', peer review systems and UNICEF Evaluation Office feedback to ensure consistent standards.

Adjustments to the template in previous years – such as the wording of the colour coding in 2012 – reduces the ability to compare results over multiple years (as noted within the findings where relevant).

GEROS is grounded in a theory of change which presupposes that increased awareness of and accountability for evaluation quality will lead to better quality evaluations (in terms of both process and evidence), which in turn will lead to greater use of evaluations. For the purposes of efficiency, the unit of assessment of quality is the evaluation reports – a proxy of overall evaluation quality used by many UNEG members and SWAP. Whilst this serves the purpose of assessing quality, it is inevitably not a complete picture of evaluation quality and this limitation must be considered in interpreting the implications of findings.

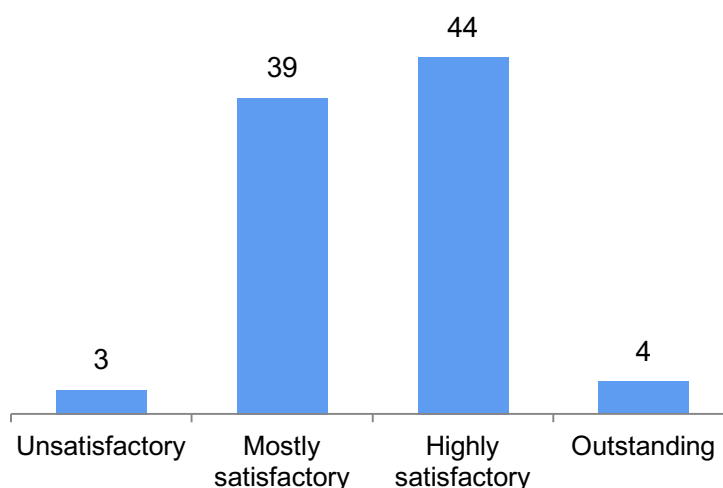
Findings

The following findings emerged from the quantitative and qualitative analysis (including triangulation) of data from the GEROS Evaluation Quality Assurance and UN SWAP reviews of 90 UNICEF evaluations delivered in 2015.

Overall findings on evaluation report quality

In 2015, the majority of reports (87%) were equally divided between the ratings of 'highly satisfactory' or 'mostly satisfactory'. As with previous years, a minor percentage of reports were rated as 'outstanding' – a rating normally reserved for exemplary examples of good practice – or entirely 'unsatisfactory'. For the purpose of the meta-analysis, 'Good quality' reports are classified as those that are 'outstanding' or 'highly satisfactory', which constitutes just over half (54%) of the reports (Figure 4).

Figure 4: Overall distribution of quality ratings for 2015 evaluation reports



The previous meta-evaluation reports recorded a continual increase in the quality of evaluations 2009-2014. From the start of GEROS and during the 2009-11 period, there was a steady increase in report quality from 36% in 2009, to 40% in 2010 and 42% in 2011. In 2012 the quality jumped up to 62%, then to 69% in 2013 and 74% in 2014. This jump coincided with a change in the GEROS service provider.

By comparison, the percentage of UNICEF evaluation reports that have been rated as good quality has declined in 2015, and is lower than for 2012-14, but higher than 2009-11. However, this still fits a long-term trend of gradual improvement (see Figure 5) and the absolute number of evaluation reports rated Good has remained consistent year-on-year (33 in 2015, 32 in 2014).

When analysed against previous years' data (Figure 5), it can be seen that there has been a steady increase in quality from 2009 to 2011, a sharp increase in 2012 and then a steady increase until 2014. 2015 represents a sharp year-on-year decrease from 74% to 53%. In 2014, for example 6% were scored 'outstanding', and 68% 'highly satisfactory' whereas in 2015 this stands at 8% as 'outstanding' (higher) but 44% as 'highly satisfactory' (lower). Both the sharp increase (2012) and decrease (2015) correlate with changes in service provider. This is illustrated on the graph below. Seen as part of a long-term trend, however, there is a

steady improvement in evaluation report quality since 2009. The apparent adjustment in quality ratings could also be due to a variety of factors that are discussed throughout the report.

Figure 5a: Long term trends in evaluation report quality 2009-2015 showing phases in different GEROS service providers

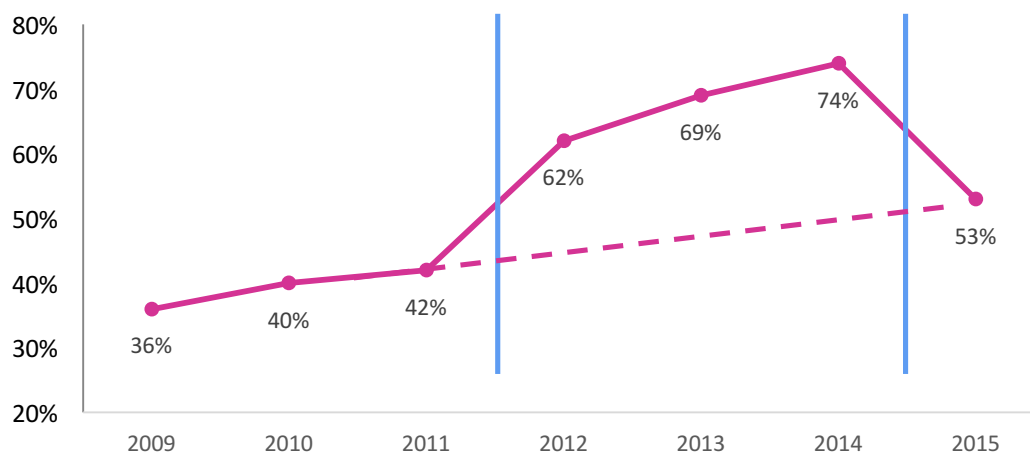
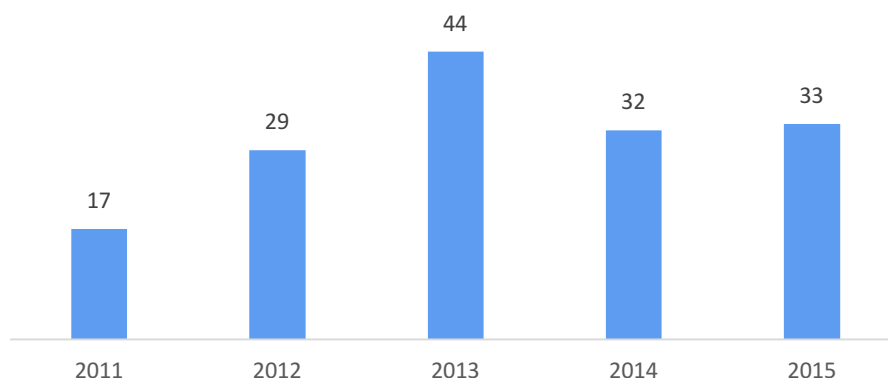
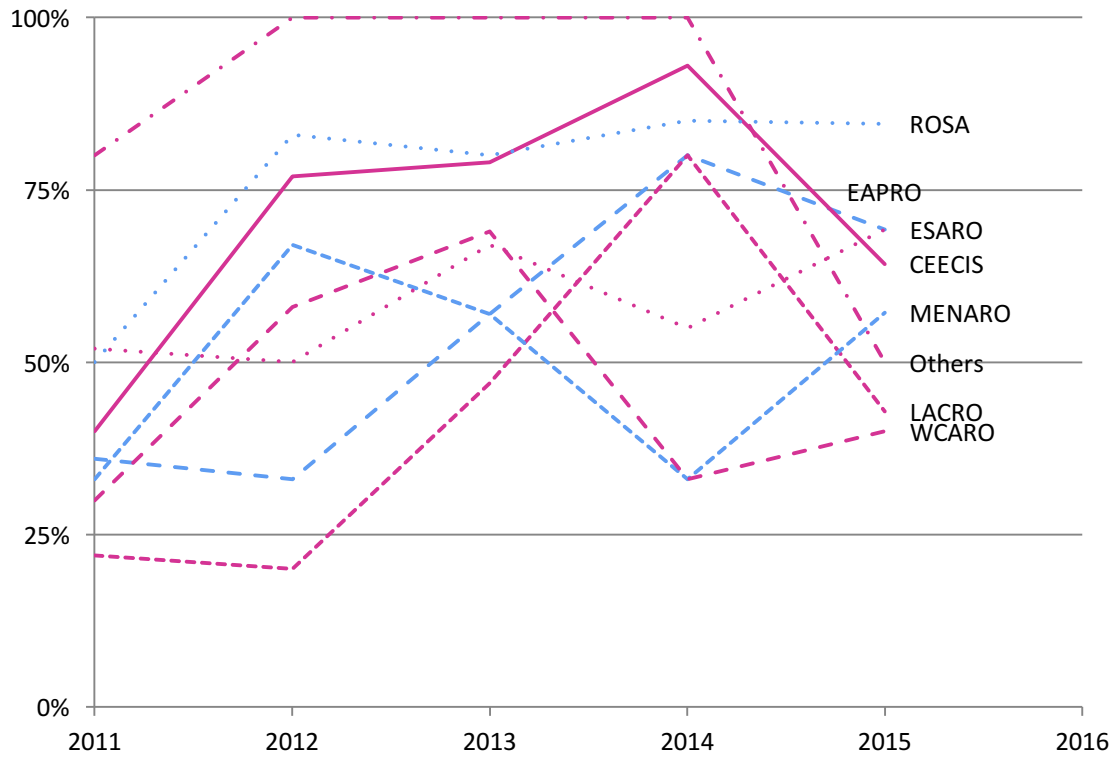


Figure 5b: Long term trends in numbers of reports rated Good 2011-2015



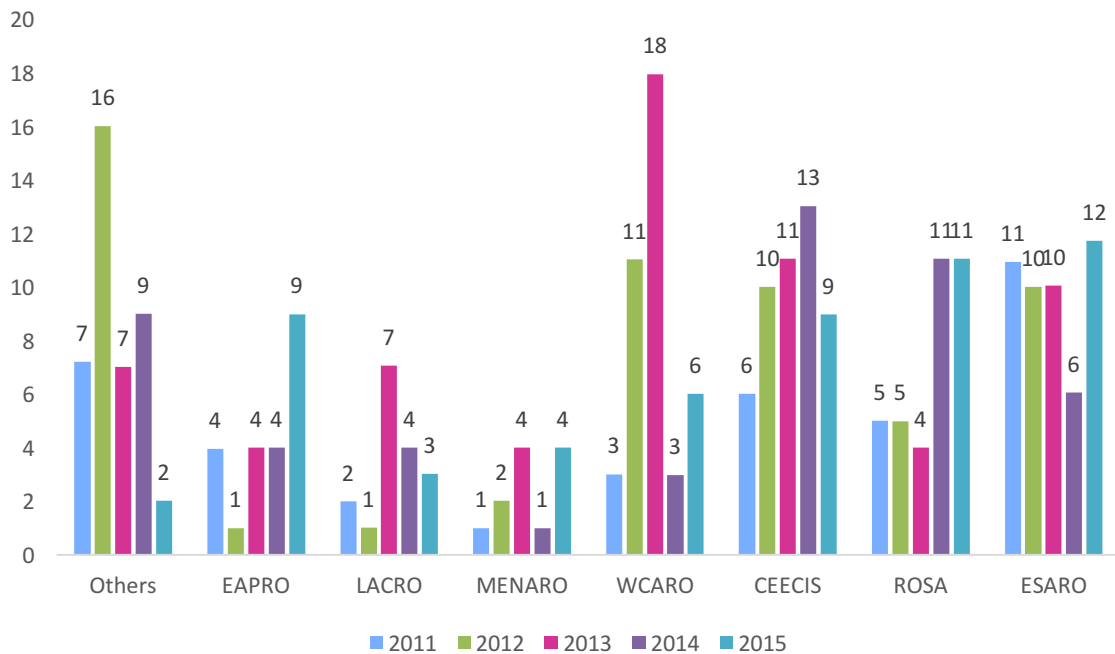
In terms of the quality of the reports across regions, it can be seen that ROSA has retained the same percentage of good quality reports, whereas some regions have gone down (CEE/CIS, LACRO and EAPRO) and some have gone up (ESARO, MENARO and WCARO – see figure 6a). Four regions improved the absolute numbers of reports rated as good in 2015: EAPRO, MENARO, WCARO and ESARO (see figure 6b)

Figure 6a: Trends 2011-2015 in the percentage of good evaluation reports from different zones



(Note: colours are used for differentiating regions and do not relate to a value).

Figure 6b: Trends 2011-2015 in the number of good evaluation reports from different zones



Areas of progress and priorities for improvement

Analysis of quality assessments relates to six subsections of the GEROS assessment tool, which are derived from to UNICEF reporting standards. Each sub-section comprises a varying number of relevant questions, with a short checklist of issues that should be covered.

- ❖ Section A: Object of the evaluation
- ❖ Section B: Evaluation purpose, objectives and scope
- ❖ Section C: Evaluation methodology, gender and human rights
- ❖ Section D: Findings and conclusions
- ❖ Section E: Recommendations and lessons learned
- ❖ Section F: Structure, logic and clarity of the report

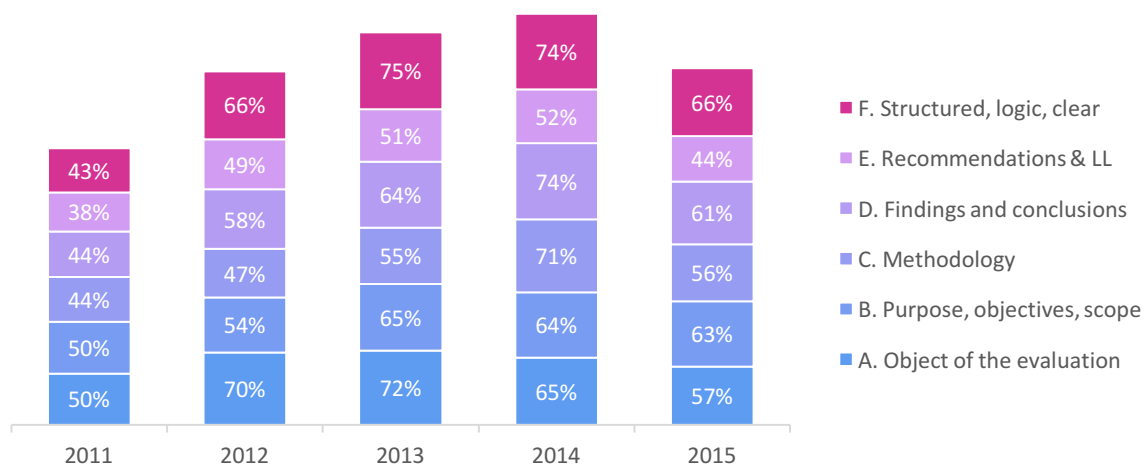
Overall it can be seen from Figure 7 that the strongest sections in 2015 were ‘Purpose, objectives and scope’ (Section B) and ‘Report is well structured, clear and logical’ (Section F): both of which are directly influenced by evaluation Terms of Reference. ‘Findings and Conclusions’ (Section D) was also rated relatively strongly. The recommendations and lessons learned section (Section E) receive the lowest rating (44%).

In terms of progress over the years, there are key areas that have been recorded for each year. In 2012, reports were analysed to be particularly strong in articulating the underlying ‘Object and context’ of the intervention being evaluated (Section A). The 2013 meta-analysis found that, since 2012, reports improved most in terms of their description of the evaluation framework, their structure and overall methodology (Sections B and C).

The 2014 meta evaluation did not document progress as specifically, but found that the highest rated sections of evaluation reports were “Methodology” (Section C – 71% good quality ratings) and “Findings and Conclusions” (Section D – 74% good quality ratings). The lowest rated section of evaluation reports was “Recommendations and lessons learned” (Section E), with only 52% good quality ratings.

A notable pattern for 2015 is a sharp drop year-on-year in the quality of the methodology section from 71% in 2014 to 56% in 2015 (but this is an improvement on 2013 – 55%). Section D findings and conclusions also reduced in quality, returning to its 2013 level. Other areas show a decline, albeit less notable, from 2014 levels.

Figure 7: Performance of quality assessment sections 2011-2015



The table below summarises the shortcomings that were highlighted in the meta-analysis reports over the last four years. In many cases specific progress was mentioned from one year to another but they continue to be identified as areas for improvement in 2015.

Table 3: Priority areas for action to strengthen evaluation identified in meta evaluations, 2012-2014

	2012	2013	2014	2015
Conclusions that did not add value to the report findings	⚠			
Insufficient justification for the selection of evaluation criteria	⚠	⚠	⚠	
Absent or weak description of the evaluated object's theory of change	⚠	⚠	⚠	⚠
Limited mention of ethical considerations and safeguards	⚠	⚠	⚠	⚠
Weak cost analysis	⚠		⚠	⚠
Poor incorporation of gender, human rights from start to finish of an evaluation	⚠			⚠
Insufficient mapping or analysis of stakeholder contributions		⚠	⚠	⚠
Lack of clear identification of target stakeholders for recommendations		⚠	⚠	⚠
Lessons learned unclear and not generalizable			⚠	⚠
Insufficiently robust analysis of contribution and causality				⚠

Object of the evaluation [Section A]

The description of the object of the evaluation scored slightly lower than the last three years but is higher than 2011 levels. Reports did well in terms of describing the object of the evaluation and implementation status; but scope remains for improving the identification and role of stakeholders, and articulation and use of the theory of change.

Section A of the Assessment tool covers the quality of the description of the object of the evaluation: description of specific object – e.g. project/ programme/ policy being evaluated; context; theory of change; identification of stakeholders; and implementation status. The section scored 'outstanding' in 3% of reports, 'highly satisfactory' in 53% of reports, mostly satisfactory in 41% of reports and 2% unsatisfactory. It can be seen that overall 'good' ratings are 56% (outstanding and highly satisfactory) have been higher in previous years (for example 66% in 2014, and 72% in 2013), but 2015 is higher than 2011 (50%).

Figure 8: Quality ratings for Section A (Object and context) 2011-2015



Further exploration of this section reveals that the majority (71%) of reports did well at describing the object of the evaluation and explaining the context and how it related to the object being evaluated. Of those that were deemed 'mostly satisfactory' this was largely due to partial discussion of the context, for example limited discussion of the social, political and economic context. They did particularly well at describing the implementation status (79% as good) which is often a relatively straightforward section unless significant changes had happened which required further embellishment.

The identification of stakeholders, and their contributions received a lower scoring (58% deemed to be good). These shortcomings are noteworthy, particularly given that this is an area that has been highlighted as an area for improvement over previous years. There were very few reports in which there was an explicit list of all stakeholders (or mention of a stakeholder analysis/ mapping exercise) and only primary stakeholders were often included. There was limited comprehensive discussion of their roles, and rarely specific mention of financial contributions (notably this included UNICEF).

The theory of change is another area which has been highlighted as a key shortcoming of reports over previous years, and continues to be so this year. This was the lowest score of 'good quality' in section A, at 48% and it is noteworthy that 10% of reports were deemed unsatisfactory. In a couple of 'outstanding' cases the evaluators were lauded for reconstructing a theory of change given a lack of one. In others there was positive feedback for the inclusion of a clear project causal chain/ results logic / theory of change. However in many reports that did not do as well, it was observed there was inclusion of parts of the results chain but there was no discussion or diagram of the causal chain bringing the elements (outputs, outcomes and impacts) together.

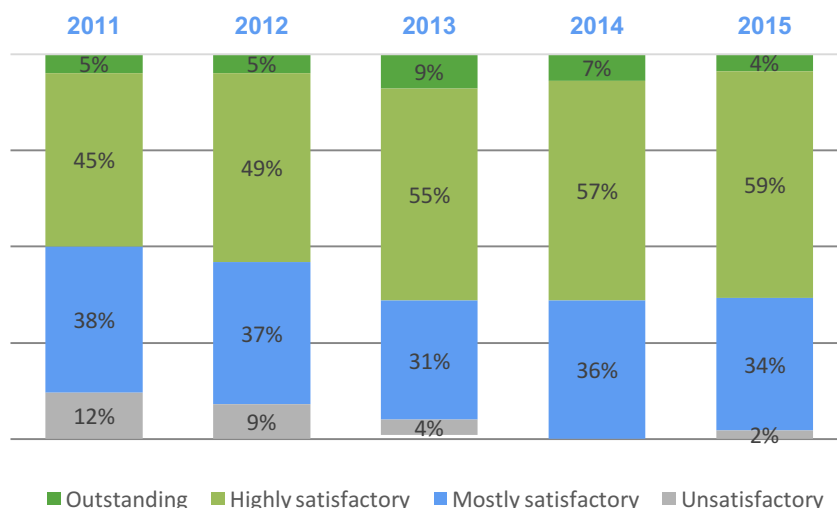
"The Object is well described and a theory of change reconstructed for the purpose of evaluation. Key stakeholders and potential users of evaluation have been clearly identified. However, the report could have included a brief outline of the overall social scenario in the country and the social protection status to enable better understanding of the rationale behind the Object." Example of review of a Section A, rated 'highly satisfactory' from *Evaluation Of Family Support Services And Stakeholders Contribution To Related Services/Systems*

“The object and context are well described in the report though in such a short way that it is hardly found complete. The logic model is presented in a simplistic way and the theory of change is not reconstructed. Stakeholders are mentioned but their role not further developed.” Example of review of a Section A, marked ‘mostly satisfactory’ from *Évaluation Du Projet De Soutien Au Redemarrage Et Au Suivi Des Activites Educatives En Republique Centrafricaine*

Evaluation purpose, objectives and scope [Section B]

Section B focuses upon the evaluation’s purpose, objectives and scope as well as the evaluation framework, which sets out the criteria used within the evaluation. It is the second highest scoring area of all the sections. In 2015, 4% were scored as outstanding, 59% as highly satisfactory, 34% as mostly satisfactory and 2% as unsatisfactory. The ‘good quality’ ratings are similar to the last two previous years (63%, compared to 64% in both 2014 and 2013) and well above 2011 levels (50%). According to evaluation theory, the evaluation framework has important implications for the quality of the later stages of the evaluation, so maintaining strength in this area is an important result.

Figure 9: Quality ratings for Section B (purpose, objectives and scope) 2011-2015



Analysis of sub-section ratings shows that evaluation reports have generally done well at setting out the purpose, objectives and scope of the evaluations, with 81% being rated as ‘good’ (of which 7% were outstanding). The reviews found that many reports were clear and explicit in terms of the purpose, why the evaluation was needed, the objectives of the evaluation, and what it intended to achieve. Where reports fell short was a lack of description about what information was needed from the evaluation, how it would be used, and the scope of the evaluation in terms of what it would and would not cover.

The evaluation framework was rated ‘good quality’ for 58% of reports, 5% of which were outstanding. This was also a notable increase from last year (53% rated as ‘good’ in 2014). In terms of those reports that did well, they were commended for clear, explicit and well justified criteria. The majority used the standard OECD DAC criteria of relevance, effectiveness, efficiency, sustainability, and impact when appropriate. In those evaluation reports that did not do as well (including 10% that were rated unsatisfactory), it was often due to a lack of explanation or insufficient justification for the criteria used.

“Evaluation purpose is well described, the objectives are clearly stated and the scope is well delimited. The evaluation criteria are also present and explained, even including the evaluation questions and indicators.” Example of a review of a highly satisfactory Section B: Evaluation report of the Technogirl programme in South Africa

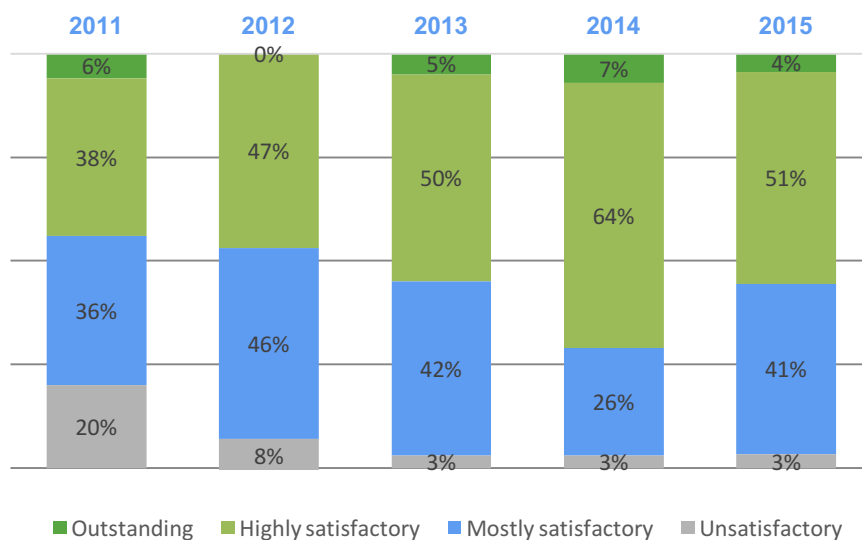
“This section sets out the objectives of the evaluation and how they were altered according to time constraints and feasibility adequately. It does not discuss how information from the evaluation will be used. The evaluation criteria were justified based upon the ToRs only rather than any wider / deeper discussion and justification.” Example of a review of a mostly satisfactory Section B: Review of the Life Skills Education Programme: Maldives

Evaluation methodology, gender, human rights and equity [Section C]

Section C assesses a broad range of criteria about the evaluation methodology, including data analysis and collection; ethics; results based management; human rights, gender and equity; stakeholder participation and methodological robustness.

The graph below illustrates that 55% of the evaluation reports were perceived to be ‘good’ with 4% of those being outstanding. This is lower than 2013-14 reports but higher than 2011’s report.

Figure 10: Quality ratings for Section C (methods, gender, human rights and equity) 2011-2015



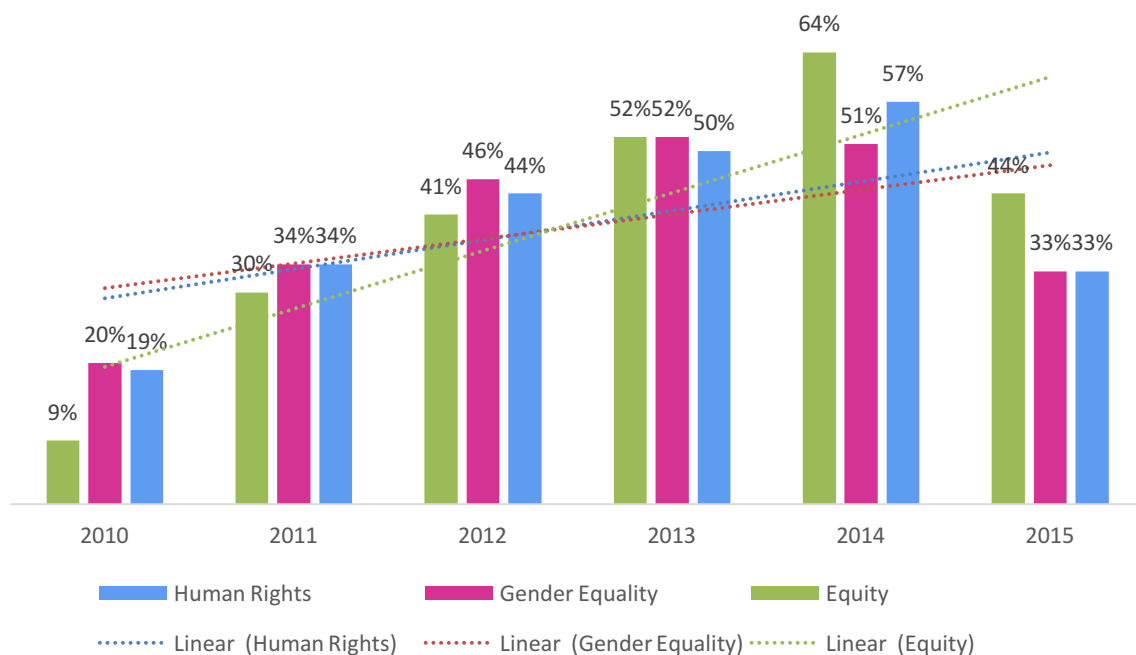
In 2015, the greatest area of strength in the methodology section continues to be data collection (81% good quality ratings, with 13% as outstanding). These reports included not only the data collection, data analysis and sampling methods (where appropriate) but also were complimented for giving the rationale for selection of methods and any limitations found.

Methodological robustness was also a strength this year with 70% of reports receiving good ratings, and it was notable that good reports were commended for having provided strong contribution analysis, and in many cases a counterfactual was considered inappropriate. Results-based management also continues to be a strong part of reports with 65% rated good, with sufficient discussion and analysis of the object’s own monitoring system and then appropriate use of the evaluation framework.

In terms of the less strong areas, ethics scored the lowest with 56% rated as good and, notably, 22% as unsatisfactory (as did not include any mention of ethical issues). This is, however, significantly higher than 2014 (43% good rating). Those reports that were considered good were commended for their consideration of relevant ethical issues including informed consent, protection of confidentiality, participation, gender responsive data collection techniques, and independence. Of the less satisfactory reports, there was lower attention to ethical issues, or they referred to generic ethical standards rather than explain how those standards relevant to the evaluation were implemented. For example, ethical issues that were identified often did not go further to discuss how safeguards could be put in place to satisfy the standards.

Human rights, gender and equity is a high priority for UNICEF, and progress in this area may reflect the efforts that UNICEF has invested. Despite gains in this area, however, scope remains for further progress. In 2012, reports were considered to be weak in all three areas (46% of good quality ratings), there was improvement in 2013 and 2014 in particular in human rights and equity. This year overall ratings for human rights, gender equality and equity are 57% good, maintaining the same level of quality as 2014. A disaggregated analysis (see figure 11) reveals a similar pattern as the overall ratings (discussed elsewhere in the report). Whilst all long term trends are positive, it is notable that inclusion of equity has improved faster than human rights or gender equality, which is in line with UNICEF strategic focus.

Figure 11: Long term patterns in the inclusion of human rights, gender equality and equity (reports rated good)



Positive reviews mention that the description of the object of the evaluation includes relevant issues (for example, children’s rights, gendered disaggregation of evidence, and discussion of gender implications). More critical reviews found that evaluation reports may mention the relevant terms within the appropriate section, but these are not carried through to findings, conclusions and recommendations.

“The report’s methodology discusses evaluation design including realistic rationale for selection of data, detail about the methods and why they are appropriate. The criteria

used to select case study locations were appropriate, as was the use of a more reflective and in-depth case study approach. The ethical safeguards were sufficient but could have gone further. Equity issues for HIV vulnerable populations, and gender issues are considered, more than human rights. Stakeholder participation is discussed in a clear, transparent and specific way. Overall a robust approach.” Example of a ‘highly satisfactory’ Section C: IATT on YKP - Joint Evaluation of HIV-Related Capacity Development Initiatives on Young Key Populations in Asia and the Pacific in the period 2010-2014

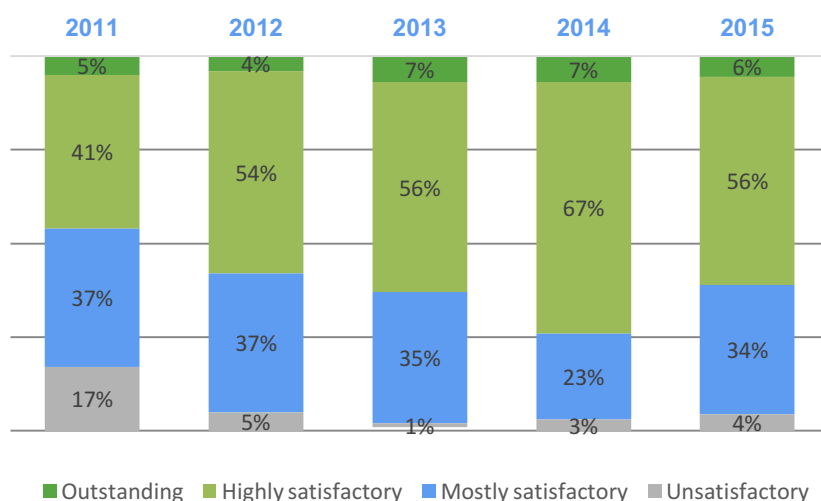
“The ToRs themselves lay down the principal elements of the methodology to be adopted for the evaluation. This and the limited time frame available for field work seem to have constrained the evaluators. While the report describes in detail the methodology adopted, it does not discuss the appropriateness of the methodology used and what could have been done in addition. In addition to the largely qualitative approach, field surveys should have been organized to gather relevant data on gender differences, inclusion and equity.” Example of a Mostly Satisfactory Section C: Evaluation Of The EU-Funded Project Widening Horizons And Creating Opportunities For Sustainable Livelihoods In North And East Sri Lanka In Partnership With EU-Unicef (2012-14)

Findings and Conclusions [Section D]

Section D focuses on findings and conclusions and covers the completeness and logic of findings, cost analysis, contribution and causality, completeness and insights of conclusions. Of the 90 reports, there were 6% outstanding reports, 56% ‘highly satisfactory’ (62% good ratings overall), 34% ‘mostly satisfactory’ and 4% ‘unsatisfactory’. In comparison to previous years, there were less ‘good quality’ ratings (outstanding or highly satisfactory), being 74% in 2014 and 63% in 2013, but this was significantly higher than in 2011 (46% good rating).

This year’s ‘good quality’ reports were commended for marshalling strong evidence, thorough analysis, clearly presented findings and relevant, actionable conclusions. In many reports there was scope for further analysis and insights from conclusions so that they were not simply summaries of findings. Those reports which had weaker findings and conclusions included critiques of short conclusions that offered no additional insights and findings that did not present sufficient evidence to support their propositions.

Figure 12: Quality ratings for Section D (findings and conclusions) 2011-2015



The sub-section ratings show that the highest scoring section of ‘findings and conclusions’ was ‘strengths and weaknesses’, or providing a balanced view of each of the strengths and weaknesses of the object of the evaluation of which a very high 83% received a ‘good’ rating. This is higher than in previous years (74% in 2014, 63% in 2013, and significantly higher than 46% in 2011).

Reports also showed that completeness of findings retained its high score with 67% good rating (67% in 2014, 62% in 2013). Cost analysis, whilst the lowest scoring of all of the sub-sections at 56% good rating, has progressed significantly since last year (34%) although 18% of reports are still considered unsatisfactory with no cost analysis at all. As of 2014, there is some reviewer discretion as to whether a cost analysis is feasible where it is not requested within the ToRs. Those reports that were outstanding or highly satisfactory were commended for various aspects of cost analysis, including thorough analytical examination of cost structures, and including projected vs actual expenditure.

Contribution and causality was rated as 66% good, which indicates that that most reports do well in assigning contribution for results to stakeholders, and identifying causal reasons for accomplishments and failures.

“The section is rated outstanding on the basis of the marshalling and use of evidence in the findings section to systematically address the evaluation criteria. The synthesis of strengths and weaknesses is particularly useful.” Example of an outstanding Section D: *Water, Sanitation & Hygiene In Schools Programme Mid-Term Evaluation Report*

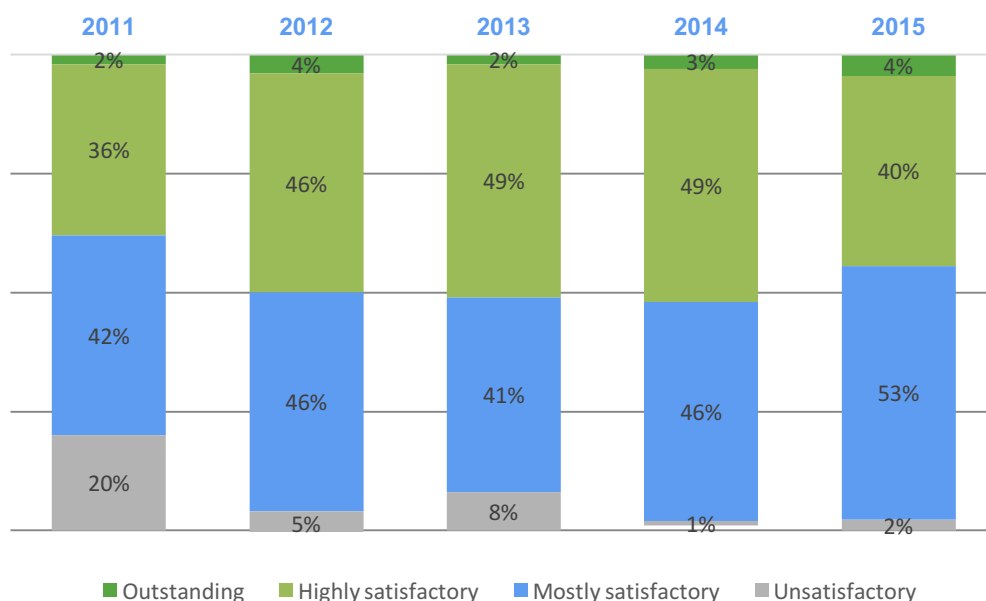
“The recommendations are relevant and actionable and they are based on the evidence from the findings and conclusions. They reflect a great understanding of the complementing organizations and demonstrate an effort to be relevant to specific stakeholders by including them in the formulation process. However, lessons learned do not indicate how they could apply in other contexts.” Example of a review with a Highly-satisfactory Section D: *Evaluation of Activity-Based Learning as a Means of Child-Friendly Education*

“Findings and conclusions were present in the report in a structured and clear manner. Important elements such as the analysis of causation and cost-value analysis were not covered.” Example of a mostly satisfactory Section D: *Évaluation du Projet de Soutien Au Redemarrage et au Suivi des Activités Educatives en République Centrafricaine.*

Recommendations and lessons learned [Section E]

Good quality ratings are the lowest within the ‘recommendations and lessons learned’ section, as they were in 2014 and 2013. The scores were slightly lower than the previous three years at 44%, (52% in 2014, 41% in 2013, 50% in 2012) and higher than 2011 (38%). Those reports that were unsatisfactory remained very low at 2% whilst the mostly satisfactory increased to 53%.

Figure 13: Quality ratings for Section E (recommendations and lessons learned) 2011-2015



Analysis of the sub-section ratings shows that relevance and clarity of recommendations continues to be the strongest sub-section, with 69% good quality ratings (lower than the 75% from last year). Qualitative comments from reviewers showed that recommendations were strongly linked to findings and conclusions, grounded in evidence and clearly stated. Those that had lower ratings often fell down on the lack of prioritisation of recommendations.

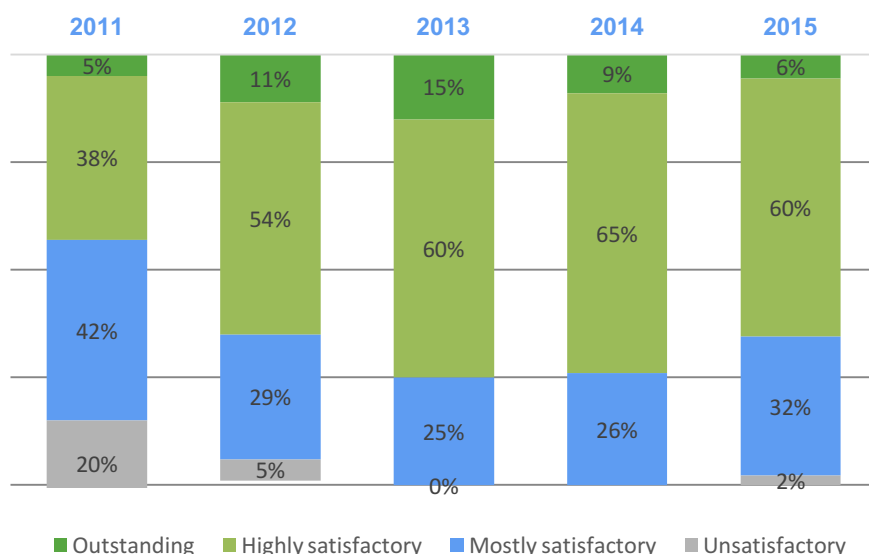
The usefulness of recommendations retained a relatively low rating (54%) and showed that there is still scope for improvement in terms of identifying the target group for action and ensuring that recommendations were realistic. A notably high 18% were considered unsatisfactory. Qualitative data showed that most reports omit to explain the process of reaching the recommendations.

Lessons learned sections remained the weakest area for Section E, with 27% of reports considered unsatisfactory, 38% as mostly satisfactory, and only 34% as good. In some reports, no lessons were included at all. In other reports, lessons were listed but were already common practice or knowledge. Very few reports demonstrated the good practice of identifying lessons that contribute to a wider body of knowledge (and are not purely repetition of commonly understood knowledge) and how they are applicable in other contexts.

Report structure, logic and clarity [Section F]

Section F focuses on the structure, logic, and clarity of the report, including style and presentation, and executive summary. It was the section with the highest ratings overall. The ratings show that it is lower than in previous years, with 66% classified as ‘good quality’ (74% in 2014, 80% in 2013 and 66% in 2012), but significantly higher than 2011 (43%).

Figure 14: Quality ratings for Section F (structure, logic and clarity) 2011-2015



Further analysis reveals that the executive summaries were strong (74% good rating with only), with many commended as containing all the necessary elements and able to stand-alone. A few reports did not include executive summaries and were therefore considered unsatisfactory. Style and presentation also scored well, with 71% deemed as good quality, and only 2% unsatisfactory. Qualitative data from reviewers supports this with reviews including mention of the logical, clear structure of reports.

“The report is logically structured, well presented and well written. The executive summary can be used by itself to inform decision making.” Example of a highly satisfactory rating of Section F: *Evaluation of the Government of Albania and United Nations Programme of Co-operation, 2012-16*

Overall regional trends

Analysis of regional trends of good quality reports provides the following insights. In absolute terms (numbers of reports), in 2015 ESARO, WCARO, MENARO and EAPRO all produced a greater number of reports rated good than in 2014 (see Figure 15b). In percentage terms, some of the regions have shown a decline in quality from 2014 (LACRO, CEE/CIS, EAPRO), which is particularly notable for HQ⁵, which declined from 100% in 2014 to 50% (Figure 15a). ROSA remained constant. MENARO, WCARO and ESARO increased slightly and returned to their 2013 levels.

⁵ It is important to note, however, that only two of the four 2015 evaluations submitted by HQ were corporate evaluations (managed by the Evaluation Office), whereas in the previous year all HQ reports were produced by the Evaluation Office.



Figure 15a: Cumulative percentage of reports rated good in each region, 2011-2015

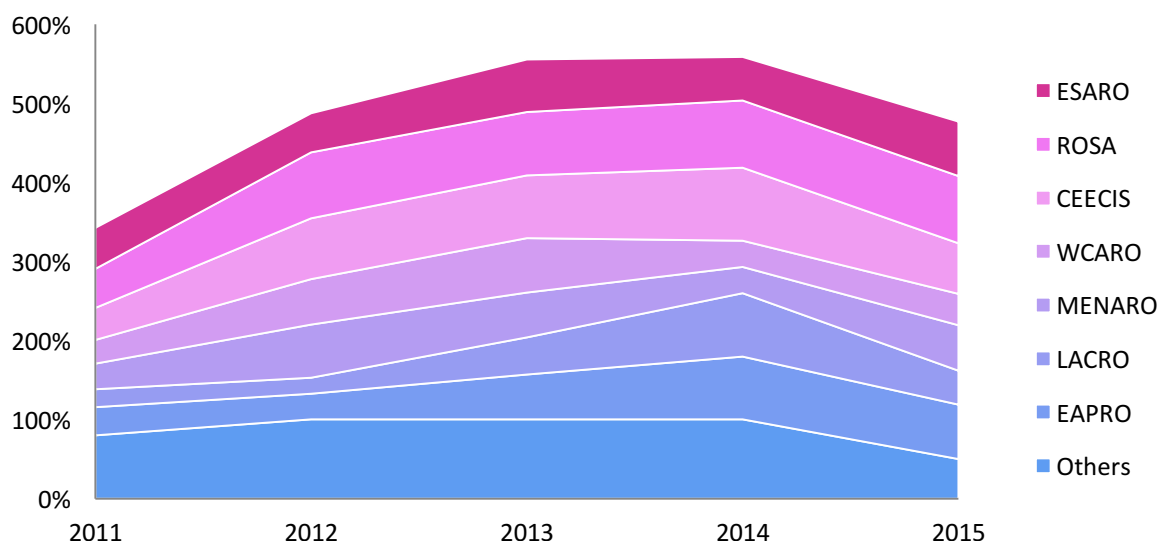
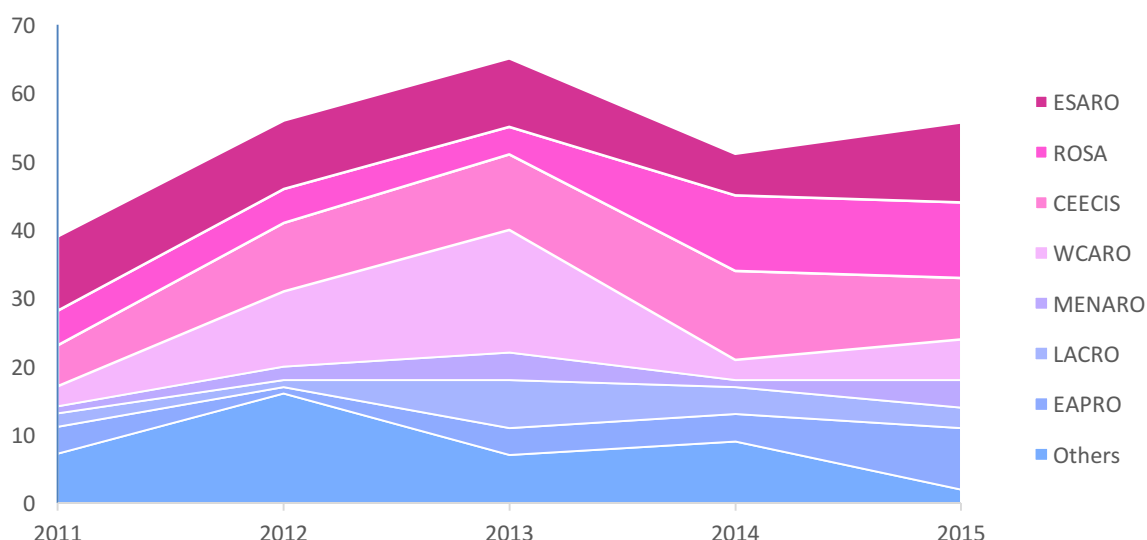


Figure 15b: Cumulative number of reports rated good in each region, 2011-2015



The review identified the regional evaluation strengths and priorities for action in 2016 based on quantitative analysis of sub-criteria data. These are presented in regional summaries at the end of the report. Globally, strengths included objectives and scope, data collection, and assessing strengths & weaknesses; whilst weaknesses included theories of change, levels of stakeholder participation, and lessons learned.

Overall thematic performance

Quantitative analysis of alignment with Strategic Plan Objective Areas (including cross-cutting issues) found that evaluations of humanitarian action and WASH rated most strongly (although with small sample sizes). Other areas in which a majority of reports are rated as meeting UNICEF standards are child protection, HIV and AIDS, social inclusion, and multi-sector evaluations. Health and education (both with large sample sizes) rate lower. Nutrition and gender equality areas both had very small samples (1).

The priority areas for action to improve health evaluations include theories of change, stakeholder participation, ethics, usefulness of recommendations and lessons learned. It should be noted, however, that health evaluations rated very highly for cost analysis, which is a weakness in the overall body of evaluations. This provides an opportunity to apply lessons from the health sector on integrating cost analysis within other strategic objective areas.

The priority areas for action to improve education evaluations include theories of change, evaluation frameworks, human rights, gender and equity, cost analysis, completeness of conclusions, and lessons learned.

Table 4: Percentage of evaluation reports meeting UNICEF standards for each objective and cross-cutting area

	Percentage rated Good	N
Humanitarian action	100%	2
WASH	80%	5
Child Protection	68%	22
Multiple	58%	12
HIV & AIDS	50%	6
Social Inclusion	50%	2
Education	42%	24
Health	31%	13
Nutrition	0%	1
Gender Equality	0%	1

UN SWAP performance and trends

In addition to GEROS reporting on the inclusion of gender equality (33% reports rated as good – see detailed findings relating to Section C, above), UN SWAP evaluation performance indicator criteria were assessed according to the standards and scoring system established by UNEG. Criteria cover the integration of GEEW within: the evaluation scope of analysis and indicators (Q1); evaluation criteria and questions (Q2); methodology; methods and tools (Q3); findings, conclusions and recommendations (Q4).

Each criterion is rated 0-3, providing an overall evaluation score of 0-12. The mean score of all evaluation reports assessed for SWAP is used to calculate the overall performance of UNICEF evaluations according to the following scale:

- ❖ 0-3,5 points = Missing requirements
- ❖ 3,51-7,5 points = Approaches requirements
- ❖ 7,51-10,5 points = Meets requirements
- ❖ 10,51-12 = Exceeds Requirements

The aggregated average score for 2015 was 6.3⁶, which is classified as **Approaching Requirements**. This represents a year-on-year improvement, with UNICEF reporting a rating of 6 (Approaching Requirements) in the 2014 cycle.

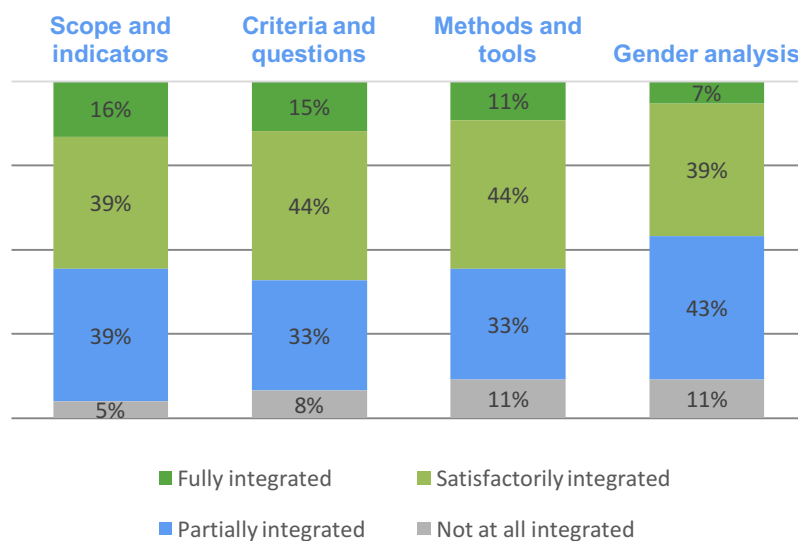
According to UNEG reporting cycle for 2015 this rating is consistent with similar entities that are subject to independent external review or peer review, including UNDP, UN Women and UNESCO. The full SWAP calculations table is included in Annex 6.

Reports were slightly stronger with regard to integrating gender in the scope, indicators, criteria and questions of evaluations. The priority for action to improve SWAP is to ensure gender analysis is used to inform evaluation findings, conclusions and recommendations.

Table 5: Performance according to UN SWAP evaluation criteria, 2015

Indicator	n	Average score	Classification
Scope and indicators	61	1.7	Satisfactorily integrated
Criteria and questions	61	1.7	Satisfactorily integrated
Methods and tools	61	1.6	Satisfactorily integrated
Gender analysis	61	1.4	Partially integrated
Overall	61	6.3 ^{See footnote}	Approaches requirements

Figure 16: Performance of UNICEF reports across the SWAP criteria



For 2014 and 2015, SWAP was implemented in parallel to the main GEROS evaluation quality assessment. It is not currently possible, therefore, to undertake detailed analysis of correlations between SWAP performance and other factors. This will be addressed through the development of a new integrated tool for the 2016 reporting cycle.

⁶ The average overall score is 6.3 based on analysis of original scores for each report. Due to rounding, the sum of the average indicator ratings displayed to one decimal place provides an inaccurate result of 6.4.

Other observable trends

This section reveals trends that emerge for ‘good quality reports’ in order to explore other causal links. A number of axes have been looked at including geographic scope, management of evaluation, purpose, level of results, level of independence, and approach. It was found that all evaluations were classified as ‘independent internal’ (independent evaluators managed by the unit responsible for the evaluation object), this cements a trend identified in the 2014 meta-evaluation report. Analysis is therefore not provided for this aspect.

With regard to management of evaluations, country-led evaluations⁷ and UNICEF-managed evaluations were strongest, although country-led was a small sample size (see Figure 16). Multi-regional and national evaluations were also strong (Figure 17), which correlates with the good performance of Country Programme Evaluations and ‘At-Scale’ evaluations (Figure 18). By comparison, less than half of joint UN evaluations, sub-national and multi-country evaluations, and pilot and project were rated as meeting UNICEF standards (Figures 17 and 18).

Figure 17: Percentage of evaluations rated good for different geographic scopes (blue) and management arrangements (pink)

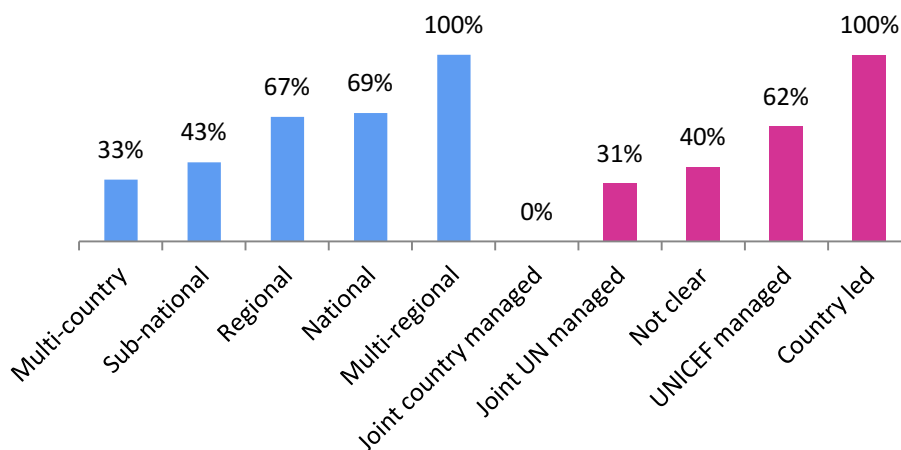
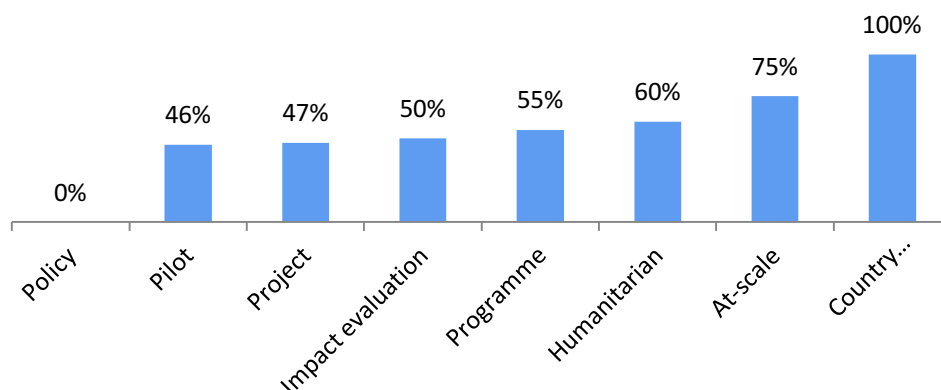


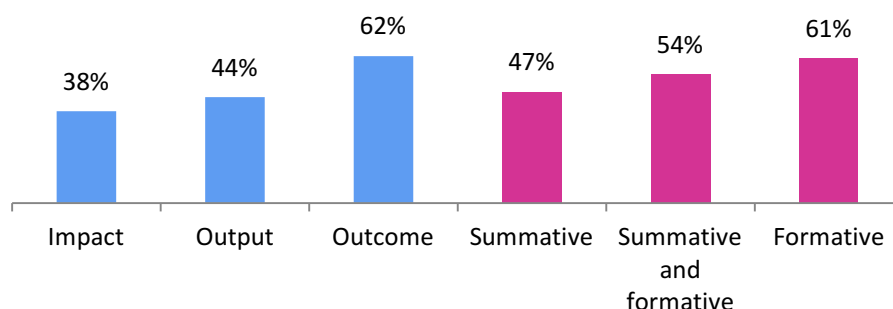
Figure 18: Percentage of different types of evaluation rated good



⁷ Country-led evaluations are commissioned and managed by national institutions (with or without technical support from UNICEF)

Quantitative analysis of meta-data reveals that outcome-level evaluations were stronger overall than output-level (generally project) or impact-level evaluations. Formative (learning and forward-looking) evaluations were notably stronger than summative (accountability and backward-looking) evaluations (Figure 19). Combined with the observations illustrated in Figures 17 and 18 this suggests that UNICEF’s decentralised evaluation capacity is most robust with regard to country-level evaluation of outcomes for the purposes of future programme development. Frequency analysis of typology data reveals that the most submitted evaluations to Geros are outcome-level (55%), country-level (89%), and programme-level (44%): suggesting that UNICEF capacity is highest in the areas that it has the most practice and experience.

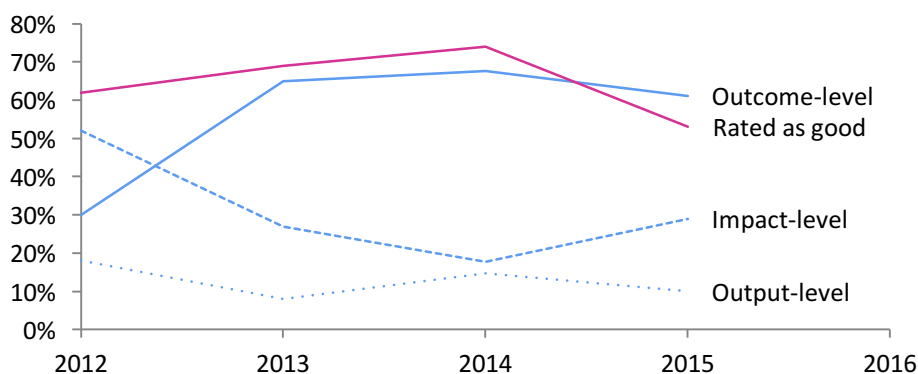
Figure 19: Percentage of evaluations rated good for different result levels (blue) and purposes (pink)



A review of the historical data on the classifications of evaluations does reveal a potential explanatory factor for the overall performance trend. The percentage of reports rated as meeting UNICEF/UNEG standards appears to be proportional to the percentage of reports that are outcome-level evaluations (as illustrated in Figure 20). The implication is that a more diverse portfolio of evaluation types makes it more difficult to consistently deliver reports to UNEG standards.

This observation could be considered to be consistent with the proposition that decentralised capacity in UNICEF – including the roster of evaluators that UNICEF has access to – is focused around outcome-level evaluations since UNICEF evaluation policy emphasises outcome evaluations and the majority of evaluations (and therefore experience) is with outcome-level work. Such a relationship between evaluation diversity and decentralised capacity is both viable in terms of capacity development theory, and would help to explain the performance pattern identified in Geros 2015.

Figure 20: Percentage of evaluation reports rated as good vs the result-level of evaluations, 2012-2015



Conclusions

Conclusion 1: The strengths and weaknesses of evaluation reports remain similar to previous years. Elements of the evaluation that are influenced by the ToR (purpose and objectives) are an organisational strength, whilst areas for improvement include improving theories of change, stakeholder participation, and lessons learned.

The findings of the meta-analysis continue the pattern of the preceding meta-evaluations. Persistent areas of weakness in evaluation reports include: the elaboration and use of theories of change, the mapping and participation of stakeholders, cost analysis, details on the application of ethics standards, integration of human rights (including gender and equity) especially in terms of analysis, usefulness of recommendations, and lessons learned that are insufficiently clear or generalizable.

In contrast to these persistent weaknesses are comparative strengths. The most apparent of these is in terms of the purpose and objectives of evaluations (both of which are established by the terms of reference). Evaluation reports are also stronger in terms of describing the object and context, and the specification of relevant methodologies (although this saw some decline in 2015).

It is possible to conclude from these observations that the current process of commissioning and managing evaluations is insufficient in terms of addressing persistent areas of weakness. It is proposed that three issues represent particularly acute priorities for action given both the overall level of performance and their relevance to the organisational agenda – particularly in terms of knowledge management.

Firstly, establishing and critiquing theories of change is important for three reasons: 1) UNICEF evaluations are primarily theory-based designs and thus clearly elaborating the underlying assumptions of an Object is essential to robust evaluative analysis, 2) use of theories of change provide an opportunity to aggregate knowledge and learning across UNICEF's portfolio of evaluations about the development effectiveness of particular approaches, and 3) the presentation of results chains and frameworks is consistent with UNICEF's commitment to Results Based Management.

Secondly, enhancing stakeholder participation in evaluations is considered a priority given UNICEF's leadership on human rights and equity, and investment in a considerable body of evaluation guidance on participatory approaches to evaluation (including impact evaluations). Whilst the inclusion of human rights, gender and equality analysis is also an important gap in evaluation reports, comprehensive stakeholder mapping and substantive participation is a necessary first step to generate the primary evidence and process by which this analysis can be properly undertaken.

Thirdly, lessons learned sections have persistently been found to be weak – both in the clarity of the insights generated and the applicability of these insights to wider contexts. Enhancing the quality of evaluation lessons to generate new knowledge that is both new and generalizable is considered to be a significant priority in view of UNICEF's strategic positioning as a knowledge broker.

Conclusion 2: The variations in the proportion of reports that are rated as meeting UNEG standards is best explained in terms how diverse the overall evaluation portfolio is. Greater diversity in terms of the types of evaluation being undertaken seems to slow – or even regress – the rate at which the quality of reports improves over time. The long-term trend, however, remains one of improvement in quality.

The headline result from the meta-analysis appears to show that, after a six-year run of constant improvement, the proportion of evaluation reports meeting UNEG standards dropped for the first time, with a year-on-year decline of 21 percentage points. However, this substantive drop in the overall rating may hide a number of explanatory factors that require a more nuanced reading of the data.

The findings identified three possible causal stories that might explain the performance of evaluation reports in 2015. The first of these is a return to the long-term trend after a few years of exceptional growth. Viewed over the entire period of GEROS (2009-2015), the performance this year is largely in line with the expected trend. However, whilst this may be true, it does not offer an explanation for why the change in quality is as observed.

A second possible explanation that must be explored is the correlation between large changes in performance (both positive in 2012 and negative in 2015) and the appointment of different independent service providers. Whilst this seems to be a 'neat fit' in terms of an explanation, there are several reasons that it does not represent a conclusive explanation.

Firstly, the specific risk of inconsistency between providers was explicitly addressed in the 2015 independent review of GEROS, and whilst it could not be discounted, neither was there sufficient evidence to confirm the relationship between service provider and overall quality ratings. Secondly, two service providers produced ratings for 2015 evaluations, with similar numbers of reports assessed by each firm and only a very small difference in average ratings. Furthermore, the same UNICEF staff member undertook quality assurance for both providers.

The third plausible explanation emerges from an analysis of trends in the diversity of evaluation reports using the proxy of the 'result-level' (output, outcome, impact) that evaluations seek to assess. This finds that the proportion of reports that are aimed at evaluating outcome-level changes is proportional to the overall quality rating. This third explanation has the further benefit of a plausible mechanism in the sense that report quality is theoretically associated with evaluation capacity, and UNICEF has demonstrably more experience and better average results from outcome-level programme evaluations (suggesting that capacity is focused around these types of evaluation).

It is possible to conclude, therefore, that the third explanation for UNICEF's evaluation performance – concentration of capacity around outcome-level evaluations – is the most likely, since this is the only one that fits with both the theoretical model and empirical data without contradicting evidence. The implication of this conclusion for UNICEF is that additional support may be required to ensure that less-frequent types of evaluations reach the same level of report quality as the more-frequent programme/outcome evaluations do.

The other main implication is to interpret the 'immediate' feedback of GEROS through a more nuanced and long-term lens that takes into account changes in the characteristics of the evaluation portfolio.

Conclusion 3: UNICEF is approaching requirements with regard to integration of gender equality and women and girls' empowerment, but significant scope remains for enhancing the use of gender analysis in developing findings, conclusions and recommendations. This places UNICEF in a similar position to its comparable sister agencies.

Whilst the integration of gender into evaluation designs and methods remains inconsistent, the entire portfolio of reports was scored by a small margin as satisfactorily meeting SWAP requirements in terms of evaluation scope, indicators, criteria, questions, and methods. The main gap in the integration of gender – and the cause of an overall rating of 'approaching requirements' – is insufficient evidence of gender analysis being used to form findings, conclusions and recommendations.

Gender analysis needs to go beyond the simple disaggregation of primary data according to sex, and analyse the differentiated impacts (expected and unexpected, positive and negative) on groups of people with different social roles and groups. In the case of UNICEF there is also intersectionality of different identities, such as children and adults, abled and differently abled, educated and uneducated, urban and rural, poor and middle class, and so on. Shortfalls in performance with regard to this criterion are thus closely related to the earlier conclusion that stakeholder mapping and participation is a critical gaps in UNICEF's current evaluation practice. As a first step, promoting stakeholder mapping through the application of human rights role analysis, power analysis, and similar tools available within UNICEF's guidance is thus considered to be a significant priority if SWAP standards are to be conclusively met by 2018.

Conclusion 4: Whilst there are recurrent shortfalls in overall report quality, a wide range of evaluative capacities are also evident from examples of high quality reports. However, these capacities seem to be available only in specific regions and strategic plan objective areas – suggesting a strong need for better internal learning and knowledge exchange in UNICEF.

As previously concluded, a number of dimensions are persistently weak in UNICEF evaluations, and the systems intended to strengthen these aspects are therefore insufficient. However, it was also found that these overall gaps are not weak in all cases. For example, cost analysis (which is only considered adequate in 51% of reports) is addressed strongly in evaluations of health – with 81% rated as good. Similarly, LACRO is notable for developing strong recommendations and ROSA for developing strong evaluation frameworks, neither of which are universal strengths of the evaluation function.

This combination of findings suggests that a wide range of evaluative capabilities is available within UNICEF, but it can be concluded from the uneven distribution of quality that these capabilities exist in relative isolation. The implication of this conclusion is that additional scope exists to enhance cross-fertilisation of evaluation experience, capacity and knowledge within UNICEF as a means to target organisational capability gaps.

Recommendations

Recommendation 1: Review the strategy for evaluation systems-strengthening to prioritise enhancements to the quality less-frequent and complex types of decentralised evaluations.

EVALUATION OFFICE

This meta-analysis has concluded, based on longitudinal analysis of data from the GEROS system, that a decrease in the proportion of outcome-level evaluations and their replacement with less-frequent types of evaluations leads to a drop in the number of reports that reach UNEG standards. It is hypothesised that this is a result of organisational capacity and experience being concentrated around programme and outcome evaluations.

Given the benefit to UNICEF's knowledge-base of increasing the number and quality of diverse evaluations (such as country-led evaluations of the SDGs, evaluations of complexity, and impact evaluations), in addition to the commitment to ensure that all evaluations reach UNEG standards, it is recommended that priority is given in systems-strengthening efforts to enhance the capacity of regional and country offices to support more diverse types of evaluations. This recommendation is also considered to be relevant in view of developing the decentralised capabilities needed to support the UNEG commitment to enhancing national evaluation capacity in order to meet the Global Goals.

Recommendation 2: Ensure evaluators clearly elaborate comprehensive stakeholder mapping, analysis of human rights roles (e.g. duty bearers and rights holders), and examination of an Object's theory of change within the evaluation inception report.

EVALUATION MANAGERS

The evidence from 2015 and previous years suggests that the aspects of evaluation that are primarily determined at the TOR stage (including purpose and objectives) are consistently among the stronger-performing aspects of evaluation reports. After the TORs, the next evaluation stage and product is the inception phase and report. The purpose of this phase is to outline the evaluation approach in full detail and the production (and approval) of an inception report acts as a quality control point. It is recommended that all evaluation managers take full advantage of the inception phase and inception report to ensure that persistent gaps in evaluation quality are sufficiently addressed before proceeding to data collection.

Based on the findings and conclusions of this meta-analysis and previous meta evaluations, it is recommended that priority is given to ensuring the quality of following three elements of an inception report:

- ❖ Detailed elaboration of the Object's theories of change – at a minimum a presentation and explanation of the logframe, but ideally a theory-based analysis of the underlying assumptions used to design an intervention;
- ❖ Comprehensive stakeholder mapping that identifies social role groups and not just institutions – ideally by applying human rights role analysis, gender role analysis, power mapping, and similar techniques;

- ❖ Elaboration of how inclusion and participation will be ensured during the evaluation process for each of the identified stakeholder groups, including how ethical standards will be implemented in concrete terms.

Implementing this recommendation will help to ensure that data is generated which is of sufficient quality to undertake robust evaluative analysis, including gender analysis to the required SWAP standards.

Recommendation 3: Review the strategy for internal learning and knowledge sharing – especially for evaluation focal persons – to focus on addressing the persistent performance gaps identified through the GEROS process.

EVALUATION OFFICE AND REGIONAL OFFICES

So as to take advantage of the evaluation capabilities within the organisation, it is recommended to revisit relevant global and regional learning and knowledge management strategies for evaluation so as to give greater emphasis to the persistent performance gaps identified by GEROS. For example, global and regional internal evaluation learning events (such as the webinar on developing recommendations that has previously been held) could be themed based on issues that represent persistent gaps in UNICEF's evaluation capability. These should seek to draw on colleagues' experiences and knowledge where there is evidence of high-performing examples in isolated cases – such as cost analysis in the health sector. Priority themes include: undertaking cost analysis in the context of social programmes, developing relevant lessons learned, integrating human rights, gender and equity into evaluation, applying ethical standards, and reconstructing and critiquing theories of change.

Recommendation 4: In order to meet SWAP standards by 2018, UNICEF needs to prioritise the inclusion of gender responsive evaluation frameworks, methods and analysis in all evaluations through: 1) increasing awareness of UNICEF staff around the SWAP evaluation performance indicators, 2) specifying gender requirements in all evaluation ToRs, and 3) include assessment of SWAP in all regional evaluation helpdesks.

REGIONAL OFFICES AND EVALUATION MANAGERS

UNICEF is currently rated as 'Approaching Requirements' for the System Wide Action Plan for gender equality (SWAP) evaluation performance indicator, and has committed to fully meet the required standard by 2018. To assist in meeting the required standard – as independently assessed by GEROS – the following detailed list of concrete features that will ensure future reports meet or exceed UNEG standards for SWAP. It is recommended that this list is used by evaluation managers for quality assurance and shared with evaluators to improve the performance of UNICEF evaluations in regards to SWAP. Doing so will also contribute to enhancing the performance of evaluation reports with regard to equity.

Reports submitted to GEROS will meet or exceed UN SWAP standards where they include the following features:

- ❖ A specific reference in the Objectives of the evaluation to assessing how gender was mainstreamed in the design of the object of the evaluation;
- ❖ One or more evaluation questions specifically address how gender has been integrated into the design, planning, implementation of the intervention and the results achieved;



- ❖ Inclusion of evaluation sub-questions and/or criteria that address participation and social inclusion in UNICEF interventions;
- ❖ Mainstreaming of gender into one or more indicators under each evaluation criterion – by being gender-disaggregated, gender-specific (relevant to a specific social group), or gender-focused (concerning relations between social groups);
- ❖ A standalone section highlighting gender, human rights and equity indicators attached to the evaluation framework;
- ❖ A mix of quantitative and qualitative indicators in the evaluation framework;
- ❖ A background section that includes an intersectional analysis of the specific social role groups affected by the issue that is being addressed by the evaluation object. The best reports attempt to quantify the size of these groups and differentiate the ways in which they are affected by a particular issue;
- ❖ Presentation or reconstruction of the theories of change used by the intervention and subjecting these assumptions to critical analysis from a gender perspective;
- ❖ Description of an evaluation design that includes substantial utilisation-focused and participatory elements – including the participation of a range of duty-bearers and rights-holders in scoping the evaluation and making meaning from evaluation data (i.e. not just being a source of data);
- ❖ A statement in the main report or the annexes that explains how data collection protocols ensured that women and men, girls and boys were included in ways that avoid gender biases or the reinforcement of gender discrimination and unequal power relations;
- ❖ Data analysis in all findings that explicitly and transparently triangulates the voices of different social role groups, and/or disaggregates quantitative data; and
- ❖ At least one finding, one conclusion, and one recommendation that explicitly address the extent to which the intervention contributes to transforming the structural relationships between the social role groups identified in the background section of the report.

Good practices

The following evaluation practices were identified in reports and are considered to be examples of important strengths and innovative practices.

Table 6: Notable evaluation practices identified in 2015 evaluations

Region	Example practices
CEECIS	<p>A report from Albania covered a broad scope, yet its concise style meant that it was still accessible: lessons were well distilled, and made generalizable to a wider context.</p> <p>A report in Tajikistan included a specific evaluability assessment, as well as other high quality evaluation features.</p> <p>A regional evaluation included active participation of key stakeholders at all stages of the evaluation process and the use of mixed methods.</p> <p>Evaluators in Moldova recognised the lack of monitoring data at the outcome level; discussed how they would use qualitative data to supplement; and include improved M&E in the recommendations.</p>
EAPRO	<p>An evaluation in Timor Leste did a good job of overcoming challenges in carrying out the assignment, such as the lack of a baseline.</p> <p>An evaluation in Thailand used a participatory approach to promote ownership of the evaluation findings and recommendations.</p> <p>In Fiji, an evaluation consultant suggested and added additional important and helpful evaluation criteria to the ones requested in the Terms of Reference, indicating that the consultant critically reviewed the TOR. The same evaluation report provides a detailed overview of ethical safeguards put into place to ensure the rights and well-being of stakeholders involved in the evaluation process.</p> <p>A report from Indonesia reconstructed the theory of change when one was not available.</p> <p>An evaluation from the Philippines makes clear and concrete provision for ethics. It also uses mixed methods well.</p> <p>A report from Malaysia demonstrates numerous elements of best practices including its mainstreaming of human rights, gender, and equity priorities; its detailed description of the evaluation purpose, objectives, and how the information will be used by UNICEF and other stakeholders.</p>

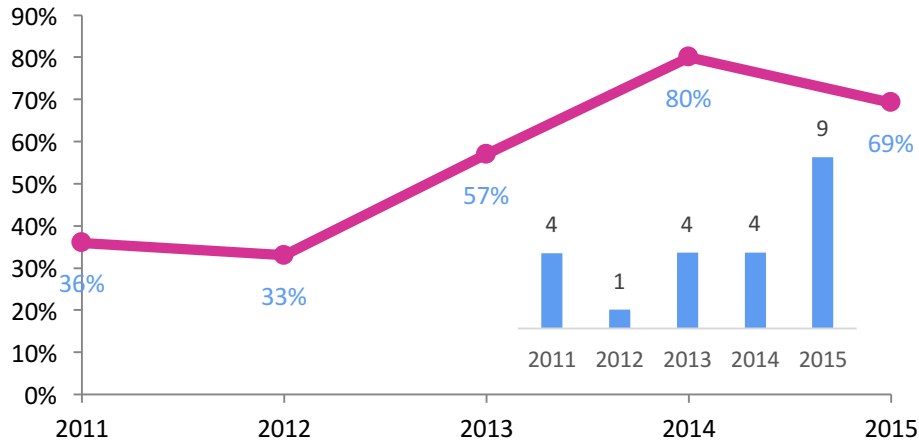
Region	Example practices
ESARO	<p>A report from South Africa includes extensive quantity of data in the form of graphs and tables that provide the findings section with great amount of field-based information.</p> <p>In Zambia, a report was found to be very good at summarising of key points in the executive summary.</p> <p>In Ethiopia, one group of evaluators did a good job at overcoming methodological limitations (problems with the counterfactual and baseline data) by including diverse stakeholders in interviews, focus groups, and observation.</p> <p>An evaluation from Tanzania was particularly strong at evaluating the programme's M&E system and goes above and beyond by providing an assessment of the challenges in the Country Programme relative to a potential Theory of Change model.</p> <p>In Malawi an evaluation team attempted to establish attribution/contribution to impact level results by using the LiST tool to create a theoretical counterfactual</p>
HQ	<p>A global programme evaluation included of a specific statement on ethics approval and was based on the use of mixed methods.</p> <p>A global evaluation included thorough and forward looking cost analysis as the basis for assessing project sustainability.</p>
LACRO	<p>In Costa Rica, a good practice in one evaluation was the attempt to construct a counterfactual through a "matching" technique, which was appropriate for assessing the effects of interventions where a stricter control group would be ethically unviable.</p> <p>In Haiti, a field visit was longer than in most of the evaluations (one month long), which was considered a good practice.</p> <p>A report from Belize included a management-endorsed action plan.</p>
MENARO	<p>A report from Djibouti includes some good practices such as including the evaluation matrix, the code of conduct and biodata of the consultant among the annexes.</p> <p>In Morocco, an evaluation incorporated an additional appropriate criteria called Coordination.</p> <p>An evaluation in Palestine included both mainstreamed and standalone gender analysis drawing on mixed QUANT/QUAL data and analysis.</p> <p>A report from Jordan was strong in terms of actionable recommendations and correctly identified lessons learned.</p>

Region	Example practices
ROSA	<p>An evaluation in the Maldives attempted to fill monitoring data gaps with alternative data collection methods.</p> <p>A good practice in an evaluation in Bangladesh was the assessment of the cost structure of the CO's work form the perspective of its impact on the strategic positioning of UNICEF as well as of the sustainability of its work.</p> <p>In Sri Lanka, one evaluation used a wide range of methods like survey, FGDs, case studies, observation and desk document review and triangulation of findings; attempts to establish a reasonable counterfactual.</p> <p>In Pakistan, one report demonstrated good practice by attempting to reconstruct a theory of change where none existed for the programme under review. It also showed good practice in its in-depth critical analysis of the programme's M&E system/results framework.</p>
WCARO	<p>An evaluation in Chad introduced additional evaluation criteria such as Coordination and HR, and Gender and Equity.</p> <p>An evaluation from Central African Republic included some extra evaluation criteria such as Coherence or Coverage, and the report acknowledged negative unexpected outcomes.</p> <p>The design of an evaluation in Mali was approved by the Institutional Review Board of the Emory University, to validate soundness and ethical questions.</p> <p>An evaluation from Ghana was particularly strong at evaluating the programme's M&E system, including its Theory of Change. It used the M&E system to create the Evaluation Framework.</p> <p>One evaluation in Liberia was based on a strong mixed-methods methodology including the use of a counterfactual.</p>

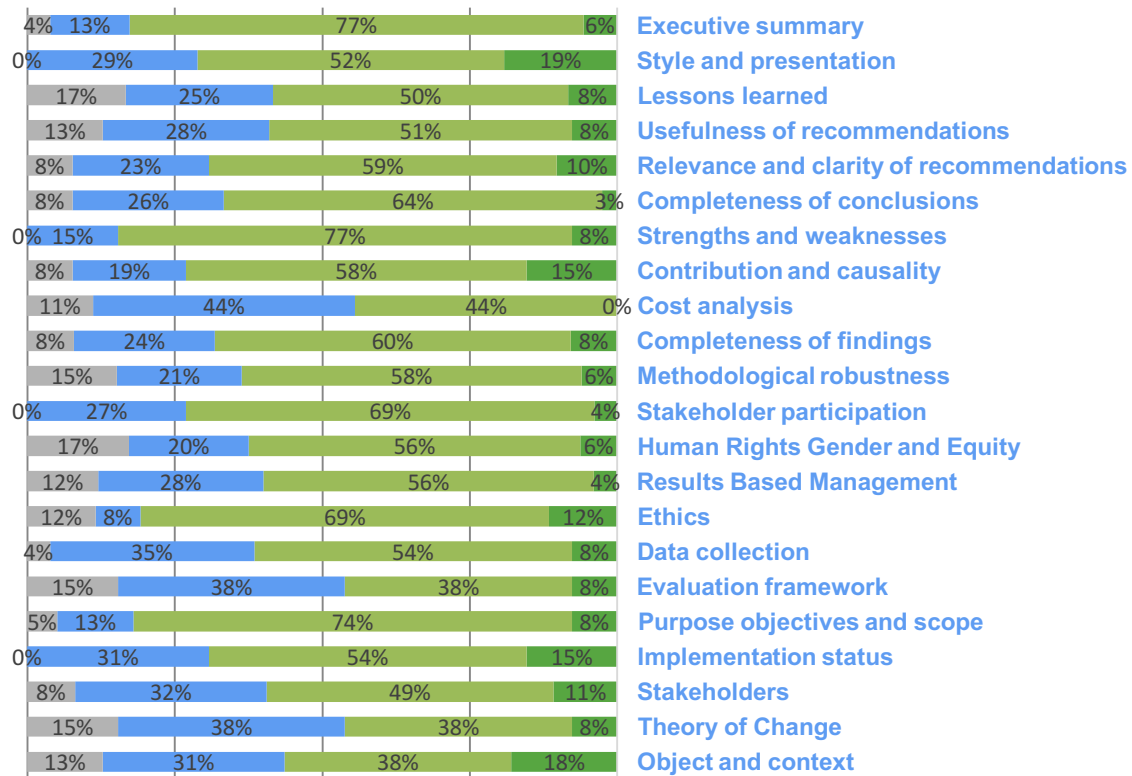
Regional performance summaries

EAPRO

Reports rated Good



Strengths	Areas for improvement
Purpose, objectives and scope	Evaluation frameworks
Assessing strengths & weaknesses	Theory of Change
	Cost analysis

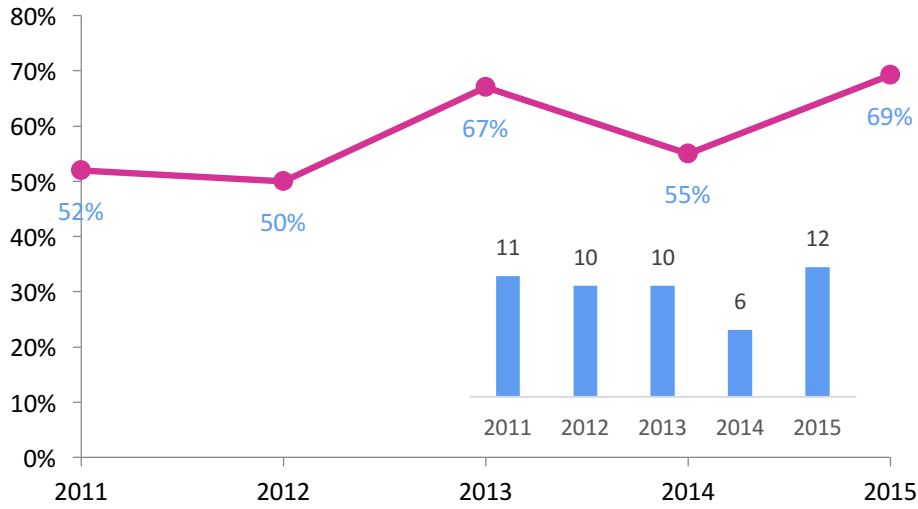


■ Outstanding ■ Yes ■ Mostly ■ No

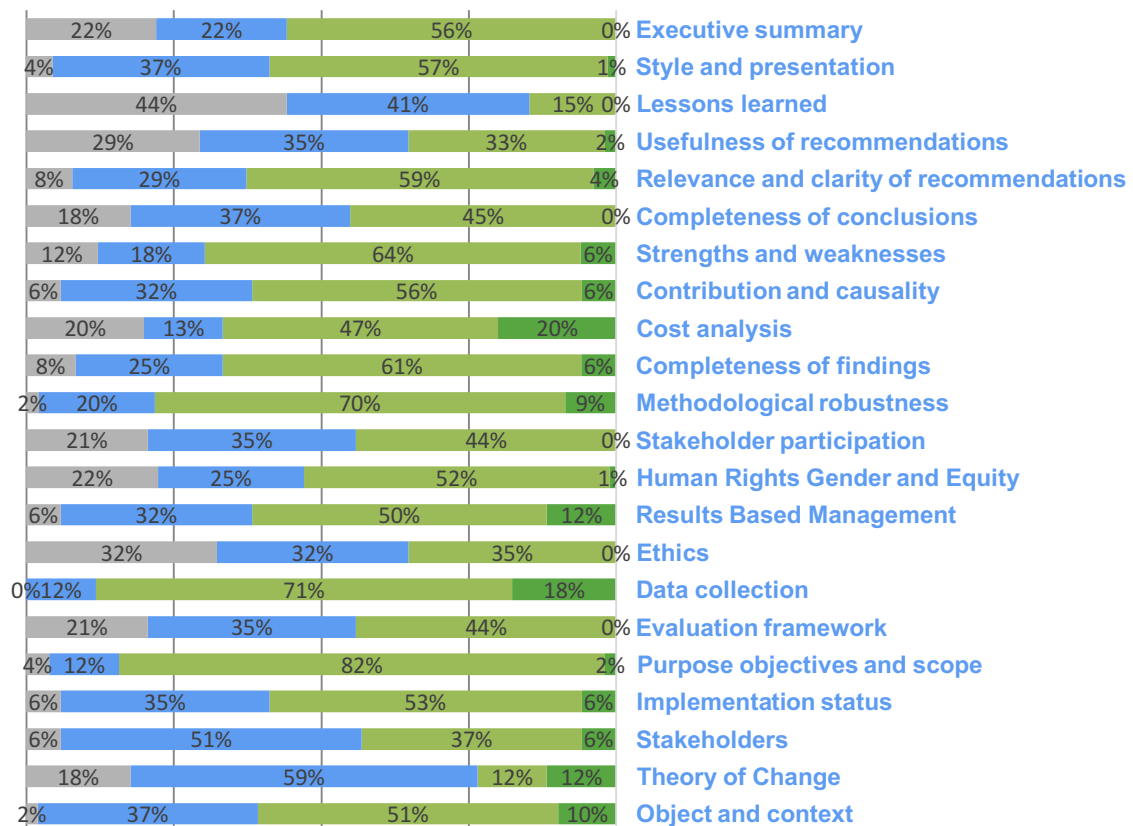


ESARO

Reports rated Good



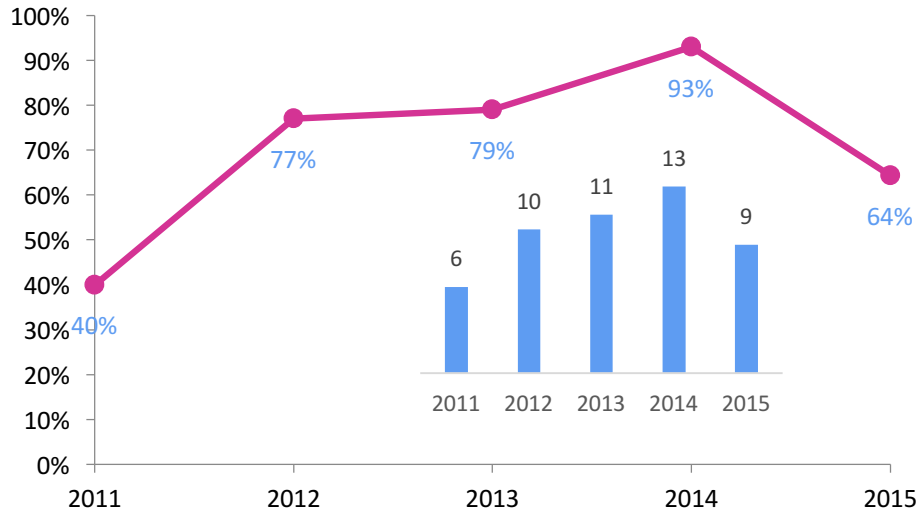
Strengths	Areas for improvement
Purpose, objectives and scope	Theory of Change
Data collection	Ethics
	Lessons learned



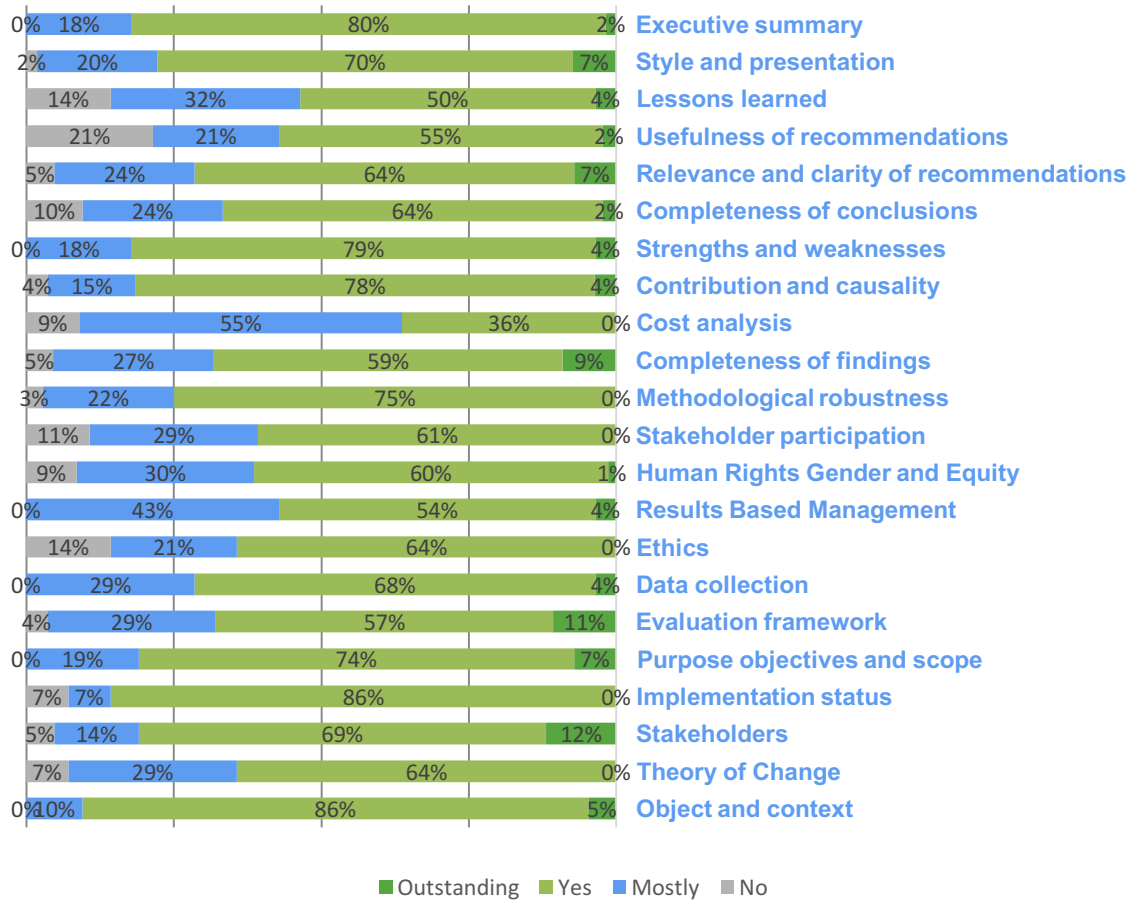
■ Outstanding ■ Yes ■ Mostly ■ No

CEE/CIS

Reports rated Good

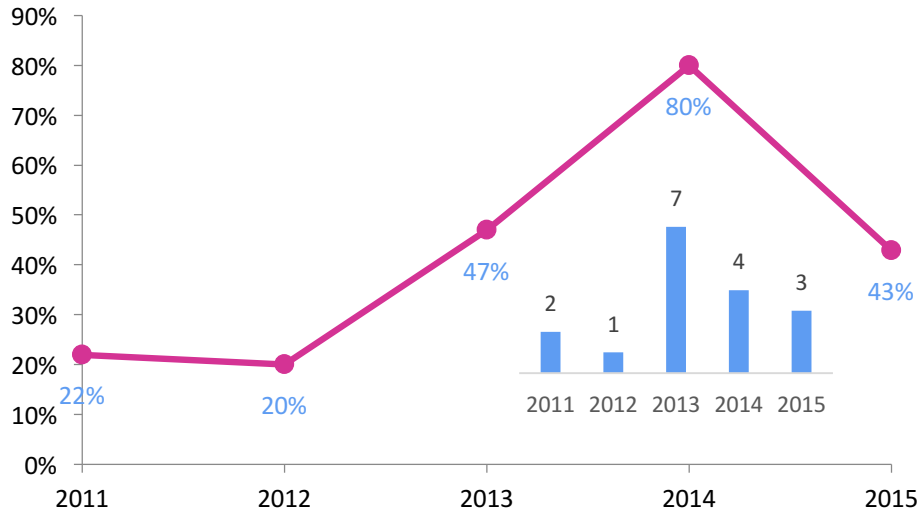


Strengths	Areas for improvement
Describing object and context	Cost analysis
Analysing contribution & causality	Usefulness of recommendations
	Assessment of RBM

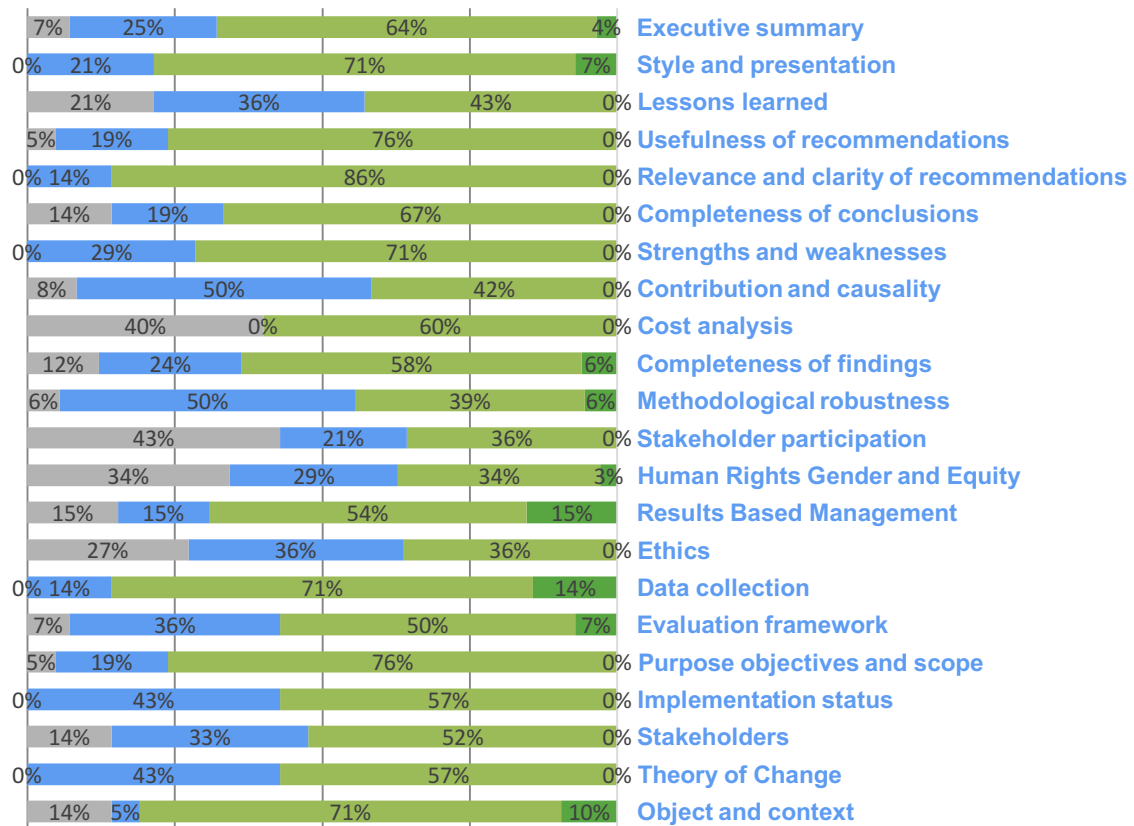


LACRO

Reports rated Good



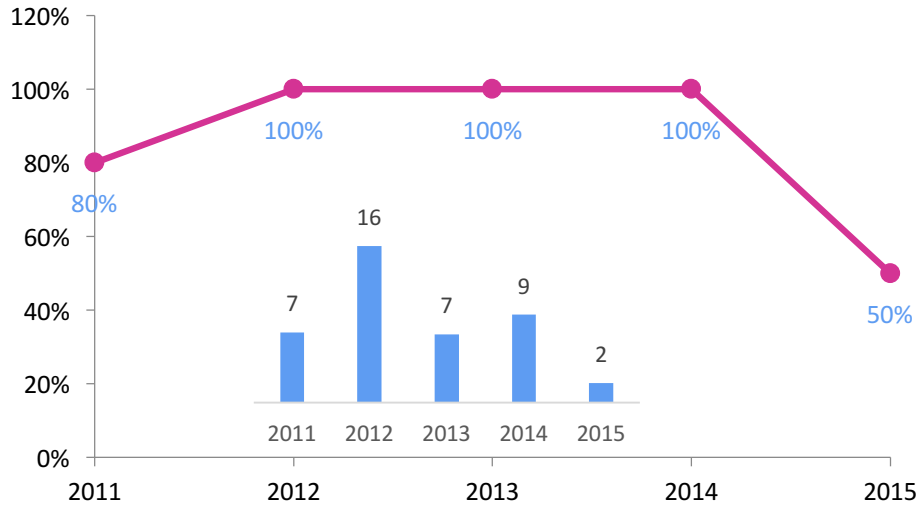
Strengths	Areas for improvement
Relevance and clarity of recommendations	Stakeholder participation
Data collection	Human rights, gender and equity
	Analysing contribution & causality



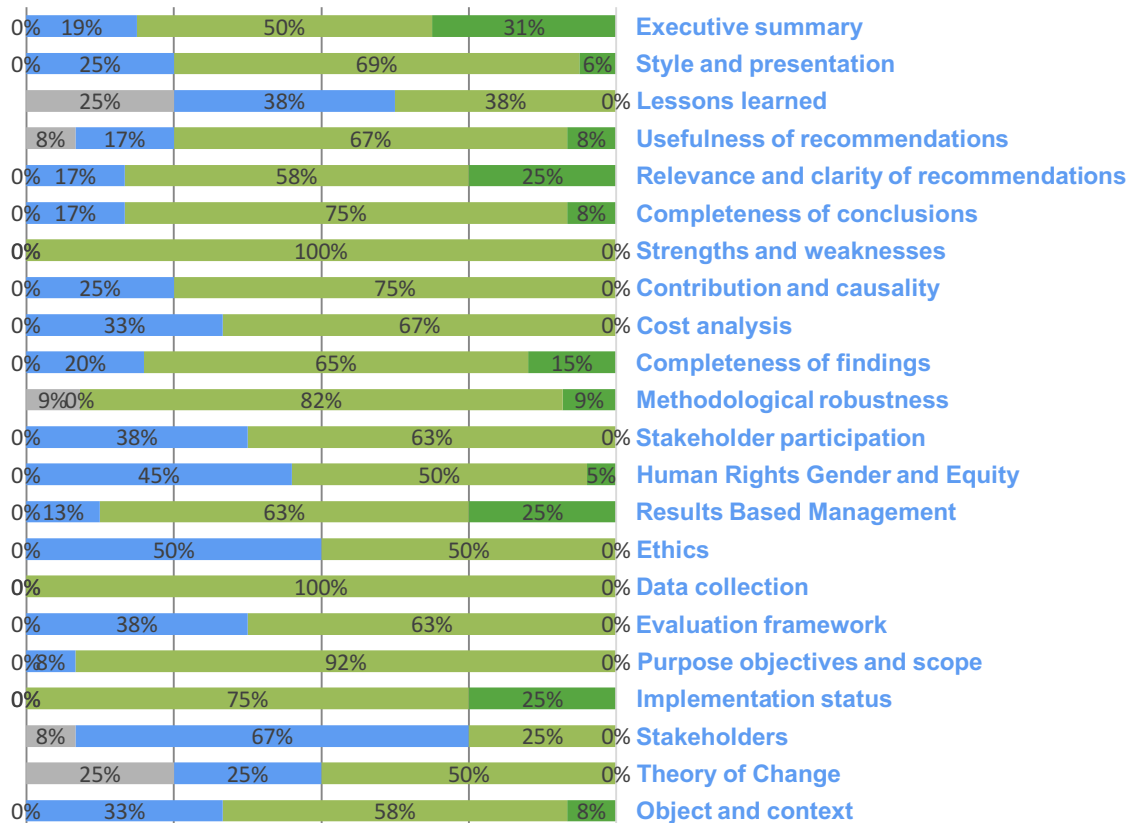
■ Outstanding ■ Yes ■ Mostly ■ No

Other (HQ)

Reports rated Good



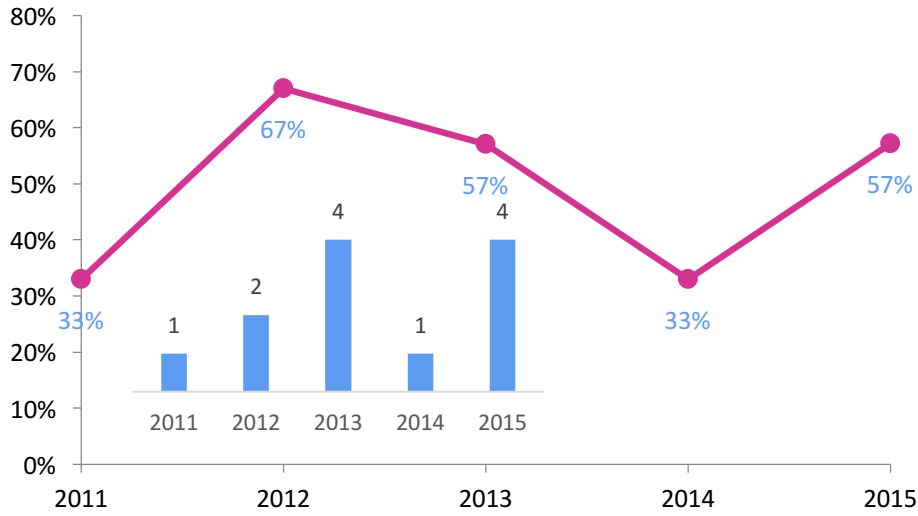
Strengths	Areas for improvement
Data collection	Stakeholder mapping
Assessing strengths & weaknesses	Theory of Change
	Ethics



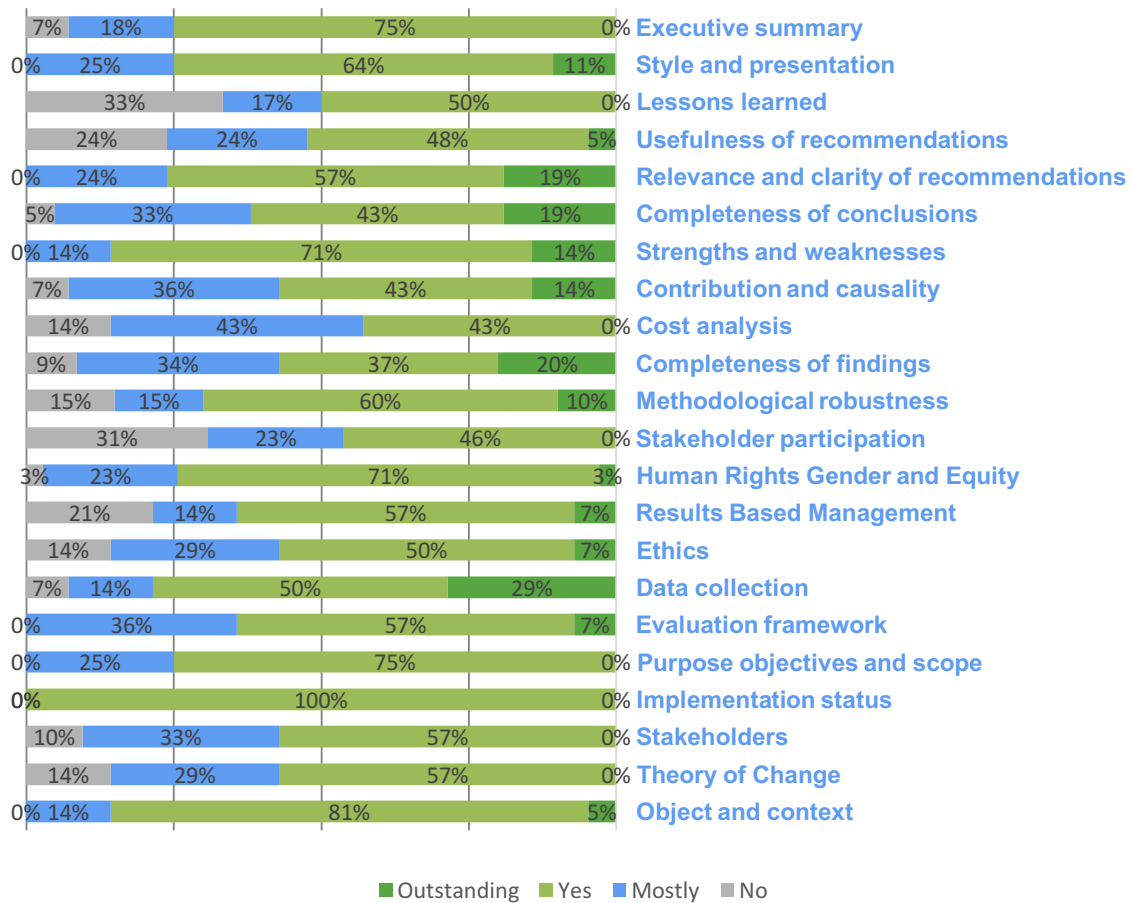
■ Outstanding ■ Yes ■ Mostly ■ No

MENARO

Reports rated Good

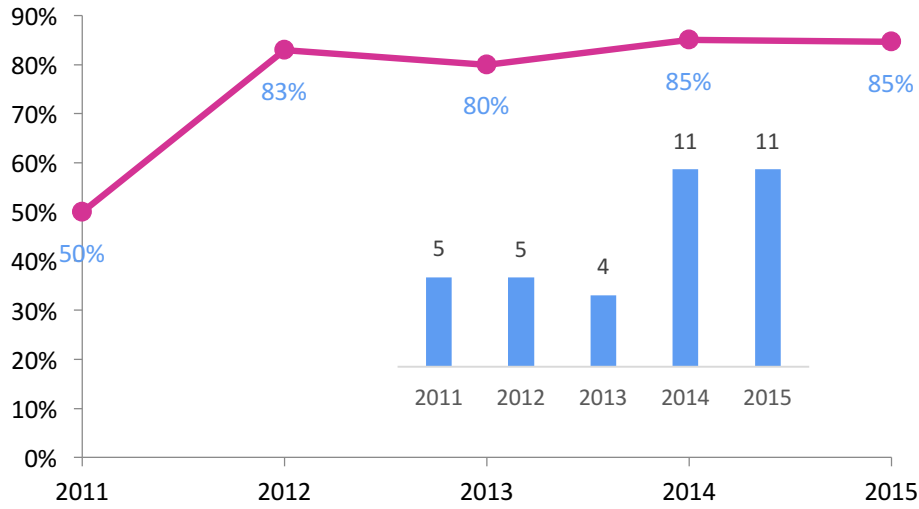


Strengths	Areas for improvement
Describing object and context	Stakeholder participation
Assessing strengths and weaknesses	Cost analysis
	Usefulness of recommendations

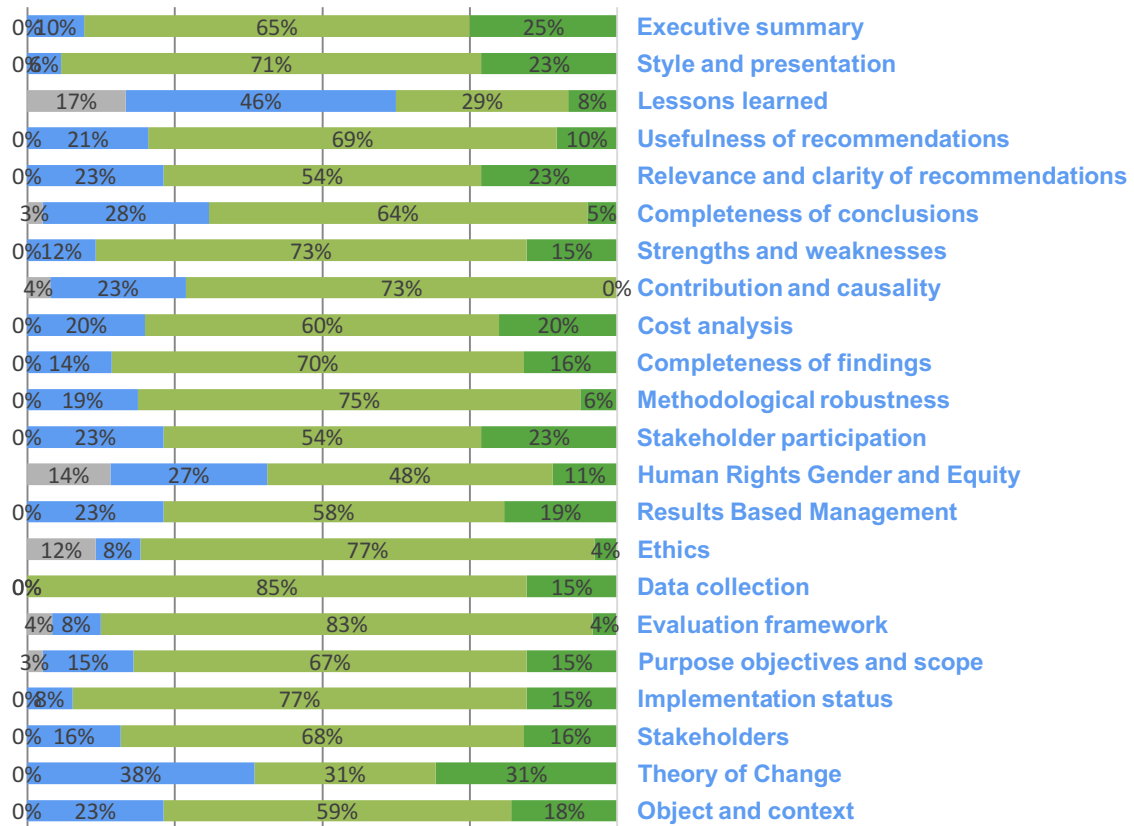


ROSA

Reports rated Good



Strengths	Areas for improvement
Data collection	Human rights, gender and equity
Evaluation Frameworks	Lessons learned
	Theory of Change

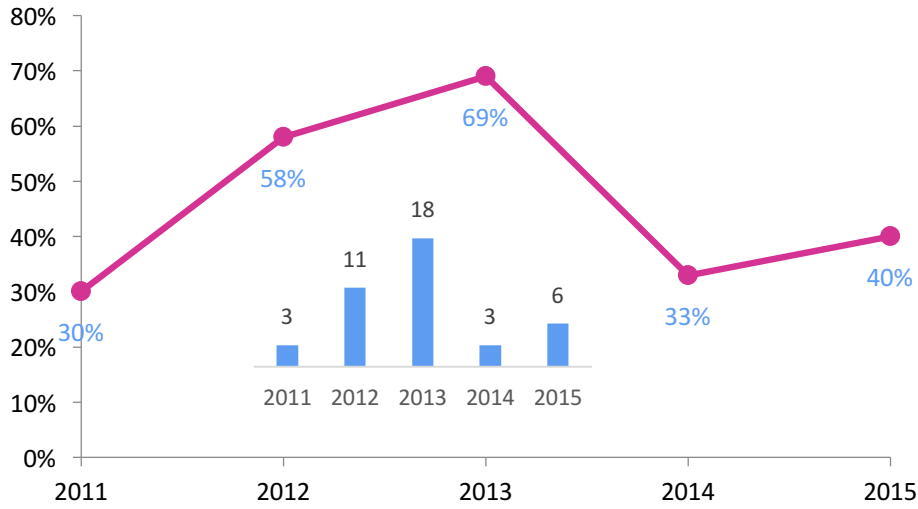


■ Outstanding ■ Yes ■ Mostly ■ No

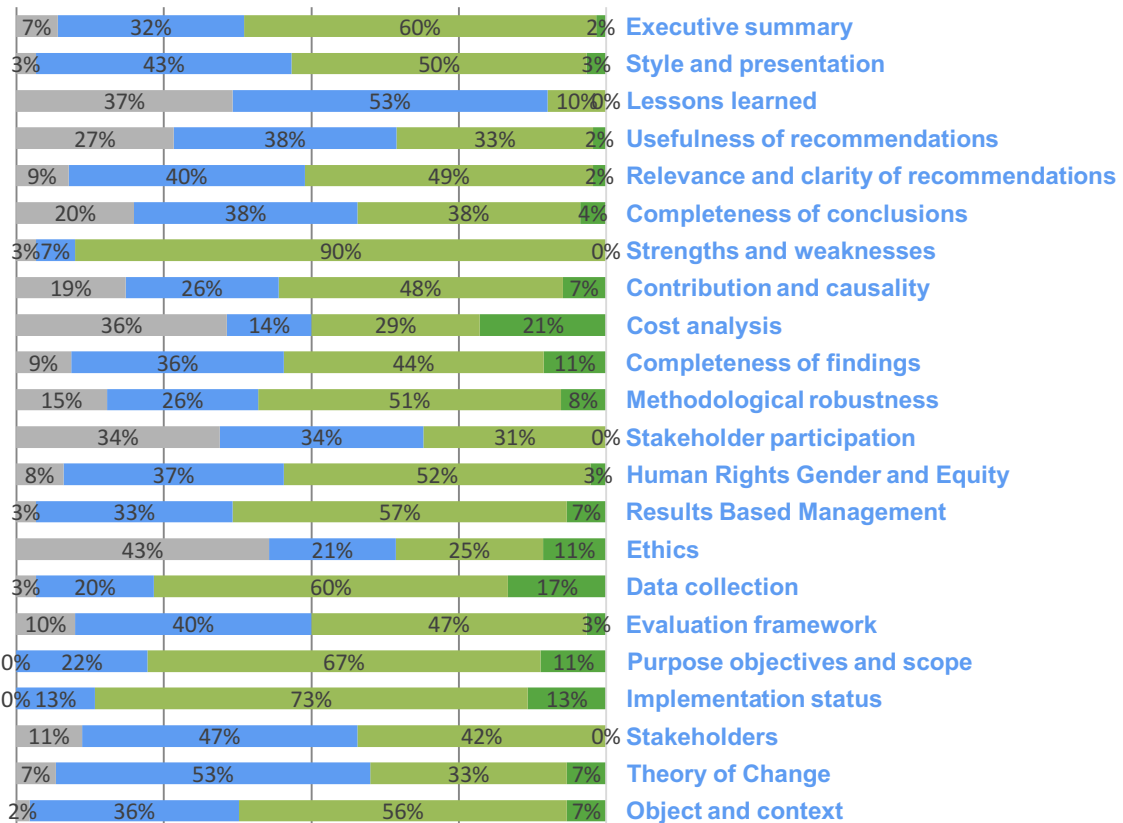


WCARO

Reports rated Good



Strengths	Areas for improvement
Assessing strengths & weaknesses	Stakeholder participation
Purpose, objectives and scope	Ethics
	Lessons learned



■ Outstanding ■ Yes ■ Mostly ■ No

Appendices

Annex 1. Terms of Reference

UNICEF Evaluation Office

Terms of Reference: Long Term Agreement (LTA) for the implementation of the Global Evaluation Reports Oversight System (GEROS). Period 2015-2018

1) Background and Introduction

The evaluation function seeks to strategically contribute to UNICEF's performance by providing good-quality evidence for learning, decision making, policy advocacy, as well as accountability purposes. Reflecting the decentralized nature of UNICEF, the majority of evaluations supported by UNICEF are managed at a decentralized level. While the decentralized nature of the evaluation function ensures that evidence generated is relevant to the local context and therefore more likely to inform national policies for children, it poses the challenge of setting up a consistent corporate system to ensure good quality and credibility.

UNICEF's Evaluation Office (EO) has in place an Evaluation Quality Oversight System (GEROS) since 2010. The GEROS is aimed at monitoring the impact of efforts to strengthen the UNICEF evaluation function globally. The system consists of rating evaluation reports commissioned by UNICEF Country Offices, Regional Offices and HQ divisions against the UNICEF/UNEG Evaluation Report Standards. All reports and the results of their quality assessment are made available in the UNICEF Global Evaluation and Research Database (EDB), as well as made publicly available on the UNICEF external website. GEROS is an organization-wide system. GEROS (annex 1)

The Global Evaluation Reports Oversight System (GEROS) has four main objectives:

1. Provide senior managers with a clear and short independent assessment of the quality and usefulness of individual evaluation reports, including those commissioned by their own offices;
2. Strengthen internal evaluation capacity by providing to commissioning offices feedback with practical recommendations on how to improve future evaluations. Commissioning offices can also use the feedback received to better assess the performance of external consultants to be hired for future evaluations.
3. Report on the quality of evaluations reports, by reviewing and assessing the quality of final evaluation reports commissioned by UNICEF Offices. Quality of evaluations reports is reported to senior management mainly through three channels: a) annual report of the EO Director to the Executive Board; b) the Global Evaluation Dashboard, and c) inclusion of this information in the Global Evaluation database;
4. Contribute to corporate knowledge management and organizational learning, by identifying evaluation reports of satisfactory quality to be used in meta-analysis to be shared within the organization, as well as facilitating internal and external sharing of satisfactory evaluations reports

The GEROS will be subject to an external assessment which will take place from June to September this year. The assessment will aim to: analyze the extent to which the GEROS has achieved its expected objectives; determine the conceptual clarity of the GEROS approach and the adequacy of its methodology and tools; and identify constraining and enabling factors

for an effective system implementation. The assessment is also expected to propose recommendations and lessons to inform the forthcoming cycles of GEROS implementation.

Against this backdrop, the Evaluation Office is looking for an institution to:

- i) Based on the findings, recommendations and lessons proceeding from the external assessment referred earlier, review and adjust the GEROS approach, methodology and tools in close collaboration with the Systemic Strengthening Unit of the Evaluation Office.
- ii) Guided by the revised GEROS approach, review and rate the quality of final evaluation reports supported by UNICEF globally, at the country, regional offices and HQ divisions; and provide feedback aimed at increasing the quality of future evaluation reports.
- iii) Further rate each evaluation report in terms of its compliance with the UN SWAP assessment tool.
- iv) Review the terms of references and inception reports produced by the regional and country offices, upon demand, as part of the regional technical assistance facilities.

2. Expected deliverables

The selected institution is expected to deliver the following:

- i) Adjusted methodological approach and tools for GEROS, based on the findings and recommendations of the external assessment.
- ii) Reviews and ratings of all evaluation reports, submitted in one year timeframe up to a maximum of 150 per year, in English, French and Spanish. Ratings of evaluation reports in Arabic and Russian will be desirable.
- iii) Ratings of all final evaluation reports using the UN SWAP assessment and meta-evaluation tool.
- iv) Executive feedbacks for each of the evaluations reviewed. The executive feedback is also expected to include the rating of each report using the UN SWAP tool.
- v) An annual global analysis of trends, key weaknesses and strengths informing whether UNICEF is meeting the requirements and criteria as set out within the UN SWAP;
- vi) A meta-analysis on the quality of evaluation reports for each GEROS one-year cycle. The meta-analysis is expected to include a Global analysis of trends, key weaknesses and strengths of the reports reviewed, including a sectoral/thematic analysis, lessons learned and good practices on Evaluation reports; and actionable recommendations to improve the GEROS system as well as the quality of Evaluation reports.
- vii) Reviews of and feedback on terms of references and inception reports upon demand by Regional Offices as part of the regional facilities.

3. Management of the system

The Senior Evaluation Specialist, Systemic Strengthening, with support from the Knowledge Management Specialist of UNICEF's Evaluation Office, will have responsibility of the overall management of the GEROS.

The selected institution is expected to appoint a project manager and to establish an explicit internal quality assurance system to ensure consistency of rating, quality and timely delivery of expected products; and overall coordination with UNICEF Evaluation Office. The project manager is expected to provide monthly updates including a tracking matrix highlighting the status of reviews, rating and executive feedback. At the end of each review cycle there will be a formal review and feedback process with both the Evaluation Office and the institution as part of the oversight and accountability role of such a global system.

4. Qualifications

- Excellent and proved knowledge of evaluation methodologies and approaches
- Proven experience on the design and implementation of quality assurance systems preferably with UN agencies
- Proven experience in designing and conducting major evaluations
- Excellent analytical and writing skills in English, French and Spanish; and Arabic and Russian are desirable
- Familiarity with UNEG/UNICEF evaluation standards is an asset
- Sectorial knowledge of UNICEF areas of intervention (Child Protection, HIV-AIDS, WASH, Education; Nutrition, Health, Social Inclusion, Gender Equality and Humanitarian Action) is an asset.

5. Duration of contract

The contract is expected to start the 1st of October 2015 and will expire on September 30th 2018.

6. Bidding documentation is as follow:

- Cover letter explaining the value added of the proposed institution
- Presentation of the institution, CVs of the project manager to be nominated for this contract, as well as of the specialists who will do the rating in English, French and Spanish are mandatory and Arabic and Russian, are desirable
- Description of the internal quality assurance system.
- Technical proposal describing how the institution will equip itself to ensure the deliverables as described in item 2 above.
- Financial proposal, with the following details:
 - o Unit cost of reviewing individual reports using the GEROS assessment tool including executive feedbacks.
 - o Unit cost of reviewing individual reports using the UN SWAP assessment tool

9. Payment

Cost of producing yearly meta-analysis on analysis of trends, key weaknesses and strengths of reports reviewed for both GEROS and the UN SWAP (two separate reports).

Unit and cost of reviewing individual terms of references for regional and country offices



Unit and cost of reviewing individual inception reports for regional and country offices.

The total fee will depend on the actual number of completed quality reviews of evaluations the unit cost per report plus the cost of the meta- analyses. This is due to the uncertainty over the total number of actual reports that will be received. All billing would be based on actual work done so as to minimize cost if expected levels of activity did not materialize.

Annex 2. GEROS Methodology

Review of evaluation reports

Evaluation quality reviews of evaluations commissioned in 2015 were carried out over a 16-month period, from January 2015 to February 2016. Two separate review teams undertook reviews of 2015 evaluations, Universal Management Group (January-December 2015) and ImpactReady LLP (January-February 2016). Both teams included evaluation experts with a broad range of relevant sectoral knowledge and linguistic capabilities (English, French and Spanish). Reviews were delegated within teams according to 'fit' with thematic and language expertise.

The list of evaluation reports to be reviewed was finalised by the UNICEF Evaluation Office, which involved monitoring the Evaluation and Research Database and undertaking an initial filtering to ensure that all reports were correctly classified as evaluations. In total, 90 evaluation reports were reviewed for 2015, which was a return to similar levels as 2011-13 (there was drop in 2014 to 69 reports).

Table A1: Number of reports included in GEROS, 2011-2015

Year	Number of Reports Reviewed
2011	88
2012	85
2013	96
2014	69
2015	90

During the handover between service providers, there was initial training for the new review team on the assessment tool. Four of the seven team members were familiar with the tool from previous cycles. For those who were new to the tool, they were individually trained and their first review was peer reviewed. The full evaluation quality assurance tool is included in Annex 3.

The GEROS assessment tool comprises 64 questions derived from the UNICEF-adapted UNEG evaluation reports, across six main sections:

- ❖ Object of the evaluation
- ❖ Evaluation purpose, objectives and scope
- ❖ Evaluation methodology, gender and human rights
- ❖ Findings and conclusions
- ❖ Recommendations and lessons learned
- ❖ Structure, logic and clarity of the report

Each of the sections is further divided into sub-sections around key areas (for example, methodological robustness, data collection, contribution and causality).

Assessment was carried out for each of the 64 questions, using a scale of four ratings: Outstanding, Highly Satisfactory, Mostly Satisfactory, and Unsatisfactory. Justifying remarks

were added per sub-section, before an assessment of ranking for each section with a narrative justification based upon evidence and examples. At the end of each section, 1-2 sentences were added for 'executive feedback' (which was used to generate to an executive feedback form for all reports). An additional 'constructive feedback' column was used to highlight issues for future reports, including examples of good practice and areas for improvement.

Table A2: Qualitative ratings categories used in GEROS 2015

Colour coding (EQA tool)	Question rating	Section rating
Dark green	Outstanding	Outstanding, best practice
Green	Yes	Highly Satisfactory
Amber	Mostly Satisfactory	Mostly Satisfactory
Red	No	Unsatisfactory

UN SWAP evaluation performance indicators were rated according to the process and standards specified by the UN Evaluation Group (UNEG):

- ❖ GEEW is integrated in the Evaluation Scope of analysis and Indicators are designed in a way that ensures GEEW-related data will be collected
- ❖ Evaluation Criteria and Evaluation Questions specifically address how GEEW has been integrated into the design, planning, implementation of the intervention and the results achieved.
- ❖ A gender-responsive Evaluation Methodology, Methods and tools, and Data Analysis Techniques are selected.
- ❖ The evaluation Findings, Conclusions and Recommendation reflect a gender analysis.

The four criteria were rated using a four-point scale (0-3) based on description of the level of integration of gender equality.

Table A3: UN SWAP ratings specified by UNEG

Score	Rating	Description
0	Not at all integrated.	Applies when none of the elements under a criterion are met.
1	Partially integrated.	Applies when some minimal elements are met but further progress is needed and remedial action to meet the standard is required.
2	Satisfactorily integrated.	Applies when a satisfactory level has been reached and many of the elements are met but still improvement could be done.
3	Fully integrated.	Applies when all of the elements under a criterion are met, used and fully integrated in the evaluation and no remedial action is required.

Once each evaluation quality assessment was completed, the review teams carried out checks to ensure completeness. The UNICEF Evaluation Office also assessed reviews to ensure consistency: providing comments and requesting further justifications where required.

Reviewers provided written response to comments received and, where appropriate and justifiable, adjusted any ratings accordingly.

Quality assurance process

The independent review co-ordinator and project manager worked closely with UNICEF in order to oversee the process, check systematic completion of reviews and executive feedback, and facilitate as consistent as possible assessments across the reviews.

Four elements were key to this:

- ❖ **Team:** the team comprised the relevant skills and linguistic skills (as mentioned above) to carry out the reviews in the time period allowed, yet was compact in an attempt to limit the degree of variation in consistency. Most reviewers were familiar with the tool having carried out the assessments in previous years, as stated above.
- ❖ **Team briefings:** those who were unfamiliar with the tool were trained in the template, and carried out an initial review which was peer reviewed.
- ❖ **Peer reviews:** the co-ordinator checked each of the reviews to ensure that it was systematically completed. In addition, there were selected peer reviews undertaken by senior evaluation experts.
- ❖ **UNICEF Feedback:** The UNICEF Evaluation Office was an integral part of the quality assurance process, reviewing each of the assessments to ensure consistency in the rating within an assessment and across assessments. In some cases reviews were highlighted to be revisited by individual reviewers, or to request a peer review.

Meta-analysis

This meta-analysis was conducted in May 2016 once all of the evaluation reports had been assessed, submitted to UNICEF EO and accepted. Quantitative data was compiled regarding scores for different aspects of the reports using Excel. Analysis was carried out across multiple axes:

- ❖ Regional trends (regional and country levels)
- ❖ Trends by quality assessment criteria (*including across time*)
 - Object of the evaluation
 - Evaluation purpose, objectives and scope
 - Evaluation Methodology, Gender, Human Rights And Equity
 - Findings and Conclusions
 - Recommendations and Lessons Learned
 - Report Structure, Logic and Clarity
- ❖ Type of management
- ❖ Purpose
- ❖ Scope
- ❖ Results level
- ❖ Strategic Plan Objective Area correspondence
- ❖ UN SWAP performance and trends

Qualitative analysis of reviewer comments was used to explore any causal links. In addition the reviews were trawled to explore good practice from the reports. Particular attention was given to SWAP reporting. Quantitative and qualitative data were triangulated, and compared with longitudinal data on findings from four previous years to map key trends and patterns.

Changes from Previous Years

Since the current GEROS process began in 2010 the process has been similar each year, with some iterative adjustments of the review assessment tool (notably in 2012 following a rapid review of the experiences and lessons of previous years). This included revising descriptions of the rating scale.

For the 2015 cycle, there were no technical changes from previous year other than integrating the (previously separate) SWAP scorecard into the main tool for some reviews. It should be noted, however, that a new Long Term Agreement for Services came into operation mid-way through the review of evaluation reports. Whilst every effort was made to ensure consistency between teams, evidence from the previous changeover suggests that performance rating consistency of the current tool can take time to settle. This was highlighted by the independent review of GEROS, and a new tool is under development for the 2016 cycle of reports in an attempt to ensure better overall consistency.

Table A4: Number of GEROS reviews undertaken by independent GEROS providers

Review team	Number of reports included in meta analysis	Percentage rated 'good'
Universalia Management Group	43	56%
ImpactReady LLP	47	51%
Total	90	53%

Limitations

The GEROS team had to address some challenges during the process of reviewing the reports. Although there are objective standards within GEROS reporting, it is assessed by individuals who use their best professional judgement. Furthermore, the change in service provider may have added to this, although four of the reviewers had been on the review teams from previous providers.

The independent review of GEROS also found weaknesses in the current tool, including the process by which indicators and sections were weighted to provide an overall rating. This tool is being revised for 2016.

Various quality assurance mechanisms were put in place to reduce variation as discussed above, including the role of the co-ordinator as 'Master Reviewer', peer review systems and UNICEF Evaluation Office feedback to ensure consistent standards.

Adjustments to the template in previous years – such as the wording of the colour coding in 2012 – reduces the ability to compare results over multiple years (as noted within the findings where relevant).

GEROS is grounded in a theory of change which presupposes that increased awareness of and accountability for evaluation quality will lead to better quality evaluations (in terms of both process and evidence), which in turn will lead to greater use of evaluations. For the purposes of efficiency, the unit of assessment of quality is the evaluation reports – a proxy of overall evaluation quality used by many UNEG members and SWAP. Whilst this serves the purpose of assessing quality, it is inevitably not a complete picture of evaluation quality and this limitation must be considered in interpreting the implications of findings.



Annex 3. GEROS evaluation quality assessment indicators

OVERALL

"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."

SECTION A

"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 change includes what the programme was meant to achieve and the pathway (chain of results) through which it was expected to achieve this.

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."

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."

"2 Is the context explained and related to the object that is to be evaluated?"

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"

"3 Does this illuminate findings?"

The context should ideally be linked to the findings so that it is clear how the wider situation may have influenced the outcomes observed."

Theory of Change

"4 Is the results chain or logic well articulated?"

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. "

Stakeholders and their contributions



"5 Are key stakeholders clearly identified?"

These include o implementing agency(ies) o development partners o rights holders o primary duty bearers o secondary duty bearers"

"6 Are key stakeholders' contributions described?"

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"

"7 Are UNICEF contributions described?"

This can involve financial or other contributions and should be specific"

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"

SECTION B**"B/ Are the evaluation's purpose, objectives and scope sufficiently clear to guide the evaluation?"**

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. 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."

Purpose, objectives and scope**"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."

"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

"12 Does the evaluation provide a relevant list of evaluation criteria that are explicitly justified as appropriate for the Purpose?"

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 TOR 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.

(This is an extremely important question to UNICEF)"

"13 Does the evaluation explain why the evaluation criteria were chosen and/or any standard DAC evaluation criteria (above) rejected?"

The rationale for using each particular non-OECD-DAC criterion (if applicable) and/or rejecting any standard OECD-DAC criteria (where they would be applicable) should be explained in the report."

SECTION C**"C/ Is the methodology appropriate and sound?"**

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.

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."

Data collection**"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."

"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"

Ethics

"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"

"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"

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. Promote gender-sensitive interventions as a core programmatic priority, To the extent possible, all relevant policies, programmes and activities will mainstream gender equality."

"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."

"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. Support for humanitarian action – achieving faster scaling up of response, early identification of priorities and strategies, rapid deployment of qualified staff and clear accountabilities and responses consistent with humanitarian principles in situations of unrest or armed conflict."

"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.

"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.

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"

SECTION D**"D/ Are the findings and conclusions, clearly presented, relevant and based on evidence & sound analysis?"**

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."

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."

"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."

"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."

"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."

Cost Analysis**"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."

Contribution and causality**"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."

"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.

(This is an extremely important question to UNICEF)"

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)."

SECTION E**"E/ Are the recommendations and lessons learned relevant and actionable?"**

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."

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."

"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."

"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"

"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."

Appropriate lessons learned**"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."

"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."

SECTION F**"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."

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"

"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"

"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"

"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."



Annex 4. List of reports assessed

Title of the Evaluation Report	Report sequence number	Country	Rating
Corporate (HQ)			
Evaluation of UNICEF's Peace Building, Education and Advocacy Programme (PBEA)	2015/008	USA	Highly satisfactory
Protecting Children from Violence: A Comprehensive Evaluation of UNICEF's Strategies and Programme Performance	2015-003	USA	Highly Satisfactory
Report on the external mid-term, formative evaluation of the optimizing HIV treatment access (OHTA) for pregnant and breastfeeding women initiative in Uganda, Malawi, Ivory Coast and the Democratic Republic of Congo (DRC)	2015/004	Programme Division	Mostly Satisfactory
Summative Evaluation of UNICEF's Integrated Health Systems Strengthening (IHSS) Programme in Ethiopia, Ghana, Malawi, Mali, Mozambique and Niger	2015/002	USA	Mostly Satisfactory
Central & Eastern Europe, Commonwealth of Independent States RO			
Evaluation of the Government of Albania and United Nations Programme of Co-operation, 2012-16	2015/001	Albania	Highly satisfactory
Independent Multi-Country Evaluation of Results Area 1: A Child's Right to a Supportive and Caring Family Environment. (Countries - Azerbaijan, Belarus, Bulgaria, Croatia, Georgia, Moldova, Montenegro, Romania, Serbia, Turkey and Ukraine)	2015/006	CEE/CIS and Baltic States	Highly satisfactory
"Progress in Reducing Health System Bottlenecks Towards Achieving the MDG 4: Evaluation of UNICEF's Contribution in Five CEE/CIS Countries"	2015/007	CEE/CIS and Baltic States	Highly satisfactory
EVALUATION OF FAMILY SUPPORT SERVICES AND STAKEHOLDERS CONTRIBUTION TO RELATED SERVICES/SYSTEMS	2015/004	Armenia	Highly satisfactory
Summative Evaluation Of Improvement Of Mother And Child Health Services In Republic Of Uzbekistan – Phase II	2015/007	Rep of Uzbekistan	Highly satisfactory
Evaluation of the project "Using textbooks in DAISY format - Primary school teacher training in using audio textbooks"	2015/001	Republic of Montenegro	Highly satisfactory
Evaluation of UNICEF's Response to the Syria Refugee Crisis in Turkey 2012-2015	2015/002	Turkey	Highly satisfactory
Enhancing Social Protection and Inclusion Systems for Children in Bosnia and Herzegovina	2015/001	Bosnia and Herzegovina	Highly satisfactory



Title of the Evaluation Report	Report sequence number	Country	Rating
MULTI-COUNTRY EVALUATION OF THE IMPACT OF JUVENILE JUSTICE SYSTEM REFORMS ON CHILDREN IN CONFLICT WITH THE LAW (2006-12)	2015/005	CEE/CIS and Baltic States	Mostly Satisfactory
End-of Project Evaluation of the Interventions Aimed to Promote an Integrated Approach to Developing Services for Most-at-Risk Adolescents and Youth Implemented during March 2012-December 2014	2015/012	Moldova	Mostly Satisfactory
Evaluation of the 'School Without Violence' Programme	2015-001	Republic of Kyrgyzstan	Mostly Satisfactory
Evaluation of Psychosocial Support Programs Implemented after Van-Erciş Earthquake	2015/001	Turkey	Mostly Satisfactory
Independent External Evaluation of the Young People's Media Network (YPMN)	2015/001	CEE/CIS and Baltic States	Mostly Satisfactory
Evaluation of UNICEF Tajikistan's work in priority districts during the 2010-2015 Country Programme	2015/001	Tajikistan	Outstanding, best practice
East Asia and the Pacific Regional Office			
Evaluation of the Thailand National Child and Youth Development Plan (NCYDP) 2012-2015	2015/003	Thailand	Highly satisfactory
Evaluation of the Water, Sanitation and Hygiene (WASH) Programme of the UNICEF Country Programme (2012-2014) implemented with the Government of Lao PDR	2015/001	Lao People's Dem Rep.	Highly satisfactory
End-of-term Evaluation of UNICEF's Solomon Islands Solar Power Pilot Project	2015/001	Fiji (Pacific Islands)	Highly satisfactory
Community Resilience and Coping with Climate Change and Natural Disasters in Vanuatu	2015/002	Fiji (Pacific Islands)	Highly satisfactory
External Evaluation of the Pacific Enable Project	2015/003	Fiji (Pacific Islands)	Highly Satisfactory
Formative Evaluation: UNICEF's Child Protection System Building Approach in Indonesia	2015/002	Indonesia	Highly satisfactory
Evaluation of the Equity Focus of Malaysia Country Programme UNICEF Malaysia Final Report	2015/001	Malaysia	Highly Satisfactory
End of Project Evaluation of Project Improving Access to Water, Sanitation, and Hygiene (WASH) in Rural Schools and Communities through Capacity Development.	2015/001	East Timor	Highly Satisfactory



Title of the Evaluation Report	Report sequence number	Country	Rating
Evaluation of Child Friendly Schools (CFS) Teacher Training Intervention Under UNICEF-Supported CFS Initiative of Ministry of Education in Timor-Leste	2015/002	East Timor	Mostly Satisfactory
WASH in Schools and Kindergartens Project	2015/001	Mongolia	Mostly Satisfactory
Evaluation of UNICEF-supported child-friendly spaces in the aftermath of Typhoon Haiyan (Yolanda) November 2013–November 2014	2015-005	Philippines	Mostly Satisfactory
Evaluation of the Intervention and Rehabilitation Program in Residential Facilities and Diversion Programs for Children in Conflict with the Law	2015/004	Philippines	Outstanding, best practice
EVALUATION OF PHASE I OF THE CAMBODIA CAPACITY DEVELOPMENT PARTNERSHIP FUND (CDPF)	2015/001	Cambodia	Unsatisfactory
Eastern and Southern Africa Regional Office			
Final report for the Evaluation of the Implementation of the Safer South Africa Programme on Violence against Women and Children	2015/002	South Africa	Highly satisfactory
Namibia Football Association's Galz and Goals Sports for Development Programme – Impact Evaluation Report	2015/004	Namibia	Highly satisfactory
Rapport final d'évaluation du Programme « Appui d'Urgence à l'Éducation Nationale Malgache (AUENM) »	2015/001	Madagascar	Highly satisfactory
Protecting Children from Violence: A Comprehensive Evaluation of UNICEF's Strategies and Programme Performance: Tanzania	2015/001	United Rep. of Tanzania	Highly Satisfactory
Evaluation report of the Technogirl programme in South Africa	2015/001	South Africa	Mostly Satisfactory
EVALUATION DE L'IMPACT DE LA MISE EN OEUVRE DU CONTRAT PROGRAMME DE REUSSITE SCOLAIRE (CPRS) SUR LA SCOLARISATION DES ENFANTS ET LA PERFORMANCE DU SYSTEME EDUCATIF DANS LES REGIONS DE DIANA, SOFIA, MELAKY ET ATSIMO ATSIANANA	2015/002	Madagascar	Mostly Satisfactory
EVALUATION OF THE INTERACTIVE RADIO INSTRUCTION (IRI) PILOT PROGRAMME IN EARLY CHILDHOOD EDUCATION IN THE EASTERN PROVINCE OF ZAMBIA	2015/007	Zambia	Mostly Satisfactory
Measuring the Impact of SMS-Based Interventions on VMMC Uptake in Lusaka Province, Zambia	2015/008	Zambia	Mostly Satisfactory

Title of the Evaluation Report	Report sequence number	Country	Rating
Evaluation of Learning Achievement in Selected Woredas in Amhara and Addis Ababa Sub-Cities Mid-Term Report	2015/013	Ethiopia	Mostly Satisfactory
EVALUATION OF THE SOCIAL CASH TRANSFER PILOT PROGRAMME, TIGRAY REGION, ETHIOPIA	2015/051	Ethiopia	Mostly Satisfactory
Final Evaluation Report YouthLead Initiative: Most Significant Change and Lessons Learnt	2015-001	South Sudan	Mostly Satisfactory
Evaluation of Tanzania UNDAF 2011-2016 Final Report	2015/007	United Rep. of Tanzania	Mostly Satisfactory
Evaluation of TUSEME Programme in Tanzania	2015/008	United Rep. of Tanzania	Mostly Satisfactory
Evaluation of Integrated Post-Primary Education (IPPE) Pilot Project in Tanzania	2015/009	United Rep. of Tanzania	Mostly Satisfactory
Report on the Summative External Evaluation of the Catalytic Initiative (CI) / Integrated Health Systems Strengthening (IHSS) Programme in Ethiopia	2015/001	Ethiopia	Mostly Satisfactory
Report on the Summative External Evaluation of the Catalytic Initiative (CI)/Integrated Health Systems Strengthening (IHSS) Programme in Malawi	2015/001	Malawi	Mostly Satisfactory
Report on the Summative External Evaluation of the Catalytic Initiative (CI) / Integrated Health Systems Strengthening (IHSS) Programme in Ethiopia	2015/001	Ethiopia	Mostly Satisfactory
Latin America and Caribbean Regional Office			
Evaluación de la Iniciativa Maternidades Seguras y Centradas en la Familia (MSCF) 2010-2014	2015/001	Argentina	Highly satisfactory
Informe Final de la Evaluación de Resultados de la Estrategia de Comunicación para el Cambio Social La Muralla soy Yo	2015/001	Colombia	Highly satisfactory
Reducing the Juvenile Population in State-supported institutions (RJPSI) in Jamaica Project	2015/002	Jamaica	Highly Satisfactory
Apoyo Técnico para la Revisión y Evaluación del Programa de Transferencia Monetaria Avancemos del Instituto Mixto de Ayuda Social (IMAS) para Contribuir a la Reducción de la Deserción y el Abandono Escolar. Informe Final de Evaluación	2015/001	Costa Rica	Mostly Satisfactory
Evaluation du Projet d'Appui au Programme National de Lutte contre le Choléra en Haïti	2015/001	Haiti	Mostly Satisfactory

Title of the Evaluation Report	Report sequence number	Country	Rating
School Community Liaison & Security Programme Process Evaluation FINAL Report	2015/001	Belize	Mostly Satisfactory
Protecting Children from Violence: A Comprehensive Evaluation of UNICEF's Strategies and Programme Performance - Mexico Country Report.	2015/001	Mexico	Mostly Satisfactory
Middle East and North Africa Regional Office			
End-Project Evaluation of the Youth Leadership, Empowerment, Advocacy, and Development in Sudan - Final Report	2015/005	Sudan	Highly Satisfactory
UNICEF'S Emergency Psychosocial Support Response for Syrian Children in Jordan: An Evaluation	2015/001	Jordan	Highly satisfactory
Evaluation of Emergency Education Response for Syrian Refugee Children and Host Communities in Jordan	2015/004	Jordan	Highly satisfactory
Rapport sur l'Evaluation de la mise en œuvre de la Stratégie Nationale pour l'Abandon Totale à Toutes Formes d'Excision	2015/001	Djibouti	Mostly Satisfactory
Programme de Développement Local et Droits des Enfants et des Jeunes	2015/001	Morocco	Mostly Satisfactory
IMPROVING COMMUNITY-BASED PSYCHOSOCIAL PROTECTION SERVICES FOR CHILDREN AND ADOLESCENTS IN EAST JERUSALEM	2015/003	Occupied Palestinian Territ.	Mostly Satisfactory
WATER, SANITATION & HYGIENE IN SCHOOLS PROGRAMME MID-TERM EVALUATION REPORT	2015/002	Occupied Palestinian Territ.	Outstanding, best practice
South Asia Regional Office			
EVALUATION OF UNICEF-SUPPORTED MOET'S INITIATIVE OF MOTHER TONGUE BASED BILINGUAL EDUCATION IN VIET NAM 2006 – 2014	2015/031	Vietnam	Highly satisfactory
IATT on YKP - Joint Evaluation of HIV-Related Capacity Development Initiatives on Young Key Populations in Asia and the Pacific in the period 2010-2014	2015/001	ROSA, Nepal	Highly satisfactory
End Project Evaluation of GAVI Supported CSOs Project to Strengthen CSO Involvement in Immunisation and Maternal-Child Health Services in Pakistan	2015/002	Pakistan	Highly satisfactory
Evaluation of the Parent to Child Transmission of HIV (PPTCT) Programme in Pakistan	2015/003	Pakistan	Highly satisfactory

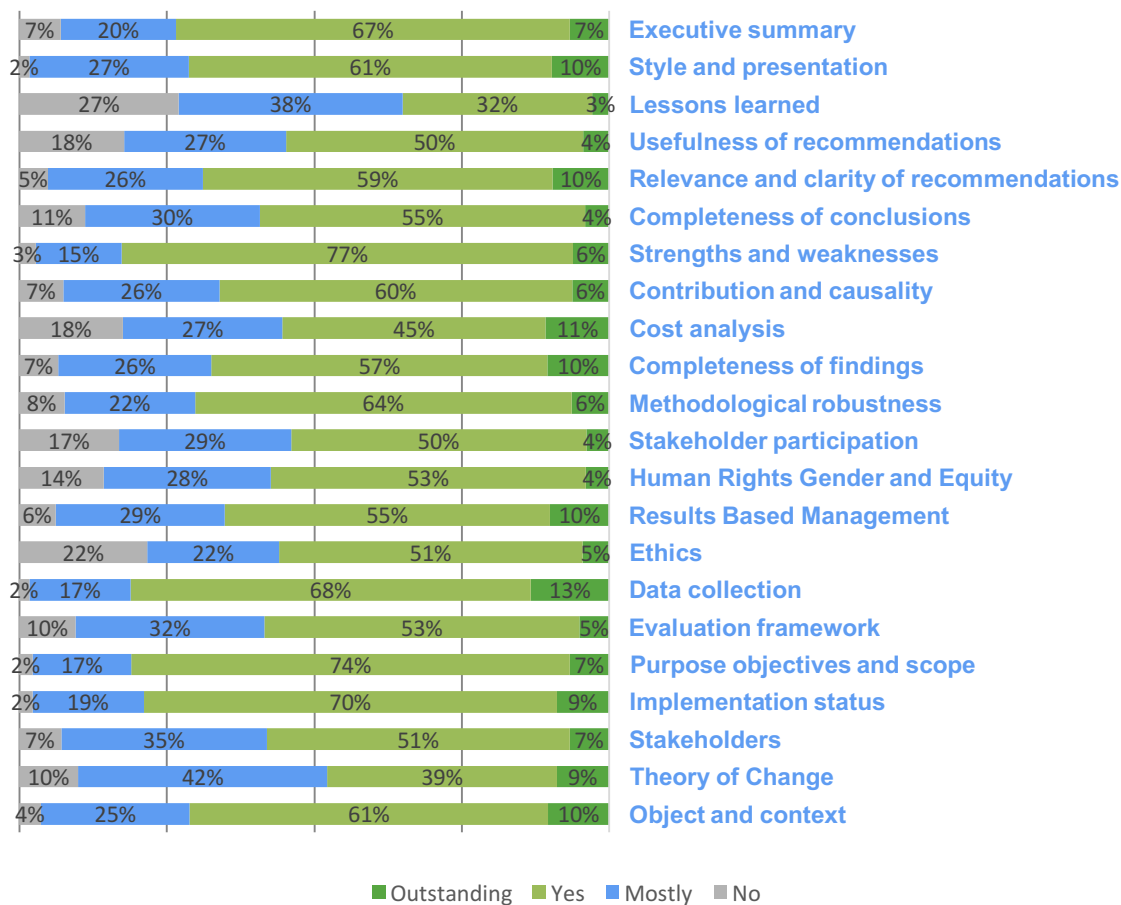


Title of the Evaluation Report	Report sequence number	Country	Rating
EVALUATION OF THE EU-FUNDED PROJECT WIDENING HORIZONS AND CREATING OPPORTUNITIES FOR SUSTAINABLE LIVELIHOODS IN NORTH AND EAST SRI LANKA IN PARTNERSHIP WITH EU-UNICEF (2012-14)	2015/005	Sri Lanka	Highly satisfactory
Sri Lanka Child-Friendly Approach Evaluation Final Report	2015/006	Sri Lanka	Highly satisfactory
Evaluation of Activity-Based Learning as a Means of Child-Friendly Education	2015/024	India	Highly satisfactory
Protecting Children from Violence: A Comprehensive Evaluation of UNICEF's Strategies and Programme Performance	2015/001	Bangladesh	Highly Satisfactory
Evaluation of Meena Radio Programme	2015/002	India	Highly satisfactory
Mid-Term Evaluation Report of the UNICEF Programme "Promoting Child Rights in Cotton Farming Areas of Pakistan"	2015/001	Pakistan	Highly satisfactory
Evaluation of Promoting the Protection and Education Rights of Children in Cotton Areas in Gujarat, Rajasthan and Maharashtra	2015/001	India	Highly Satisfactory
Review of the Life Skills Education Programme: Maldives	2015/001	Maldives	Mostly Satisfactory
Evaluation of UNICEF Strategic Positioning in Bangladesh	2015/002	Bangladesh	Mostly Satisfactory
West and Central Africa Regional Office			
DUBAI CARES WASH IN SCHOOLS INITIATIVE IN MALI	2015-002	Mali	Highly satisfactory
Evaluation à Mi - Parcours du Programme de Cours Accélérés (PCA) au Bénin	2015/001	Benin	Highly satisfactory
Protecting Children from Violence: A Comprehensive Evaluation of UNICEF's Strategies and Programme Performance	2015/002	Ghana	Highly Satisfactory
Évaluation participative à mi-parcours du programme de l'UNICEF Budgétisation Participative / Collectivités Locales Amies des Enfants (BP/CLAE)	2015-001	Senegal	Highly satisfactory
Évaluation formative du programme Cash Transfer de l'UNICEF à Kolda et Coumbacara	2015/002	Senegal	Highly satisfactory
ÉVALUATION DE LA MISE EN OEUVRE DU PROJET FRANÇAIS DE LUTTE CONTRE LES VIOLENCES DE GENRE EN MILIEU SCOLAIRE	2015/001	WCARO, Senegal	Mostly Satisfactory

Title of the Evaluation Report	Report sequence number	Country	Rating
Evaluation des interventions visant l'amélioration de la qualité de vie des populations autochtones	2015-002	Congo	Mostly Satisfactory
Formative evaluation of the project entitled "Accelerating efforts to reduce maternal, neonatal and child mortality in the Northern and Upper East regions of Ghana"	2015/010	Ghana	Mostly Satisfactory
Final Evaluation: Liberia Social Cash Transfer Programme	2015/001	Liberia	Mostly Satisfactory
Rapport de l'Évaluation externe de la Stratégie Sectorielle de l'Éducation 2015-2025 de la République du Congo	2015/001	Congo	Mostly Satisfactory
Report on the Summative External Evaluation of the Catalytic Initiative (CI)/Integrated Health Systems Strengthening (IHSS) Programme in Ghana	2015/001	Ghana	Mostly Satisfactory
Report on the Summative External Evaluation of the Catalytic Initiative (CI)/Integrated Health Systems Strengthening (IHSS) Programme in Niger	2015/001	Niger	Mostly Satisfactory
EVALUATION INDEPENDANTE DE LA COMPOSANTE SURVIE ET DEVELOPPEMENT DE L'ENFANT DU PROGRAMME DE COOPERATION TCHAD-UNICEF	2015/001	Chad	Outstanding, best practice
Evaluation externe de l'approche Ecole de Qualité Amie des Enfants (EQAmE) dans les provinces du Ganzourgou et du Namentenga	2015-009	Burkina Faso	Unsatisfactory
ÉVALUATION DU PROJET DE SOUTIEN AU REDEMARRAGE ET AU SUIVI DES ACTIVITES EDUCATIVES EN REPUBLIQUE CENTRAFRICAINE	2015/010	Central African Republic	Unsatisfactory



Annex 5. Quality of evaluation sub-sections



Annex 6: UN SWAP Calculations

	Scope and indicators	Criteria and questions	Methods and tools	Gender analysis	Overall
<i>Mean</i>	<i>1.67</i>	<i>1.66</i>	<i>1.56</i>	<i>1.41</i>	<i>6.30</i>
<i>n</i>	<i>61</i>	<i>61</i>	<i>61</i>	<i>61</i>	<i>61</i>
<i>Sum</i>	<i>102</i>	<i>101</i>	<i>95</i>	<i>86</i>	<i>384</i>
Report 1	2	2	2	2	8
Report 2	2	2	2	1	7
Report 3	1	2	2	2	7
Report 4	2	2	1	1	6
Report 5	1	1	2	0	4
Report 6	1	1	2	2	6
Report 7	1	0	1	1	3
Report 8	2	2	2	2	8
Report 9	3	3	3	1	10
Report 10	1	1	2	1	5
Report 11	3	3	3	3	12
Report 12	2	1	1	1	5
Report 13	2	2	2	2	8
Report 14	2	2	2	2	8
Report 15	2	2	3	3	10
Report 16	2	3	2	3	10
Report 17	1	1	1	1	4
Report 18	1	1	1	0	3
Report 19	1	2	2	2	7
Report 20	2	2	2	2	8
Report 21	2	2	3	3	10
Report 22	2	2	2	1	7
Report 23	2	2	2	2	8
Report 24	1	1	1	1	4
Report 25	3	2	3	2	10
Report 26	2	2	2	2	8
Report 27	3	2	2	2	9
Report 28	0	0	1	0	1
Report 29	3	3	1	1	8

UNICEF GEROS META ANALYSIS 2015

Report 30	1	2	2	2	7
Report 31	3	2	3	2	10
Report 32	2	1	2	1	6
Report 33	1	1	0	1	3
Report 34	3	3	2	2	10
Report 35	2	2	1	1	6
Report 36	1	1	1	1	4
Report 37	2	1	2	2	7
Report 38	1	1	2	2	6
Report 39	3	2	1	2	8
Report 40	1	1	0	0	2
Report 41	1	1	2	1	5
Report 42	1	2	1	2	6
Report 43	3	3	3	2	11
Report 44	2	1	2	2	7
Report 45	3	2	2	2	9
Report 46	2	2	2	1	7
Report 47	2	2	2	2	8
Report 48	2	3	1	1	7
Report 49	2	3	1	1	7
Report 50	2	2	1	1	6
Report 51	2	3	2	2	9
Report 52	1	1	1	1	4
Report 53	1	1	1	1	4
Report 54	1	0	0	0	1
Report 55	1	1	1	1	4
Report 56	1	1	0	1	3
Report 57	1	2	1	1	5
Report 58	1	1	1	1	4
Report 59	1	2	0	1	4
Report 60	0	0	0	0	0
Report 61	0	0	0	0	0

