Background
Background

Context

• UNICEF’s GEROS aims to ensure that evaluations managed or commissioned by UNICEF meet high quality standards.

• Based on the Global Evaluation Compact, which commits to collaborate in strengthening the evaluation function within UNICEF.

Objectives

• Provide UNICEF with an independent assessment on the quality of evaluation reports.

• Strengthen their internal evaluation capacity.

• Report on the quality of evaluation reports.

• Contribute to the EO’s corporate knowledge.
Methodology- Review Process

- Review tool based on UNICEF Adapted UNEG Evaluation Report Standards.
- Classification of reports according to region, geographic scope, management type, purpose, results-level, MTSP correspondence, level of independence and stage of the evaluation.
- Total of 79 reports reviewed
- Reports rated according to a four-point performance scale
### Methodology- Performance Scorecard

<table>
<thead>
<tr>
<th>Dark Green</th>
<th>Light Green</th>
<th>Amber</th>
<th>Red</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outstanding, Best Practice</td>
<td>Highly Satisfactory</td>
<td>Mostly Satisfactory</td>
<td>Unsatisfactory</td>
</tr>
</tbody>
</table>

The report / individual sections of the report meet UNICEF’s Report Quality Standards. It is a report of good quality.

The report / individual sections of the report do not meet UNICEF’s Report Quality Standards.
Methodology- Meta Evaluation

• Analysis on quantitative and qualitative data from the 79 reports reviewed

• Overall ratings, rating by report type, ratings by report section.

• Emerging themes from the qualitative data extracted to support and explain the results from the quantitative analysis.
Limitations

- Rating of reports based on only the final evaluation reports.

- Key questions in the evaluation review tool can be subjective, leading to possible inconsistencies (mitigated through peer review process).

- Limited data for 2009, comparison across the four years not always possible.
Findings

Overall Ratings
• Very few reports reviewed for EAPRO and MENARO. Overall ratings for these two regions must take into consideration the number of reports.
Overall ratings for 2012

- The quality of reports submitted to UNICEF increased sharply between 2011 and 2012, rising by 20%.
- The percentage of good quality reports (Highly Satisfactory and Outstanding/Best Practice) rose by 20% between 2011 to 2012.
Improvement 2009-2012
Overall Regional Trends

- With few exceptions, the quality of UNICEF sponsored evaluations has improved across most of the Regions when compared with previous years.
Trends by Type and Scope of Evaluation

Findings
Geography

% of reports by geographic scope:
- National: 53%
- Sub-national: 33%
- Multi-region/global: 6%
- Multi-country: 4%
- Regional: 4%

% of good quality reports:
- National: 100%
- Sub-national: 33%
- Multi-region/global: 100%
- Multi-country: 33%
- Regional: 100%
Management of the Evaluation

- UNICEF: 47% reports by management, 70% good quality reports
- Externally Managed: 14% reports by management, 64% good quality reports
- Joint with country: 14% reports by management, 55% good quality reports
- Joint with UN: 9% reports by management, 57% good quality reports
- Joint with other: 9% reports by management, 71% good quality reports
- Not Clear: 5% reports by management, 25% good quality reports
- Country-led UNDAF: 3% reports by management, 0% good quality reports
- UNDAF: 0% reports by management, 0% good quality reports
Purpose of the Evaluation

% of reports by purpose

Programme: 41%
Pilot: 16%
Project: 15%
At Scale: 11%
RTE: 8%
Humanitarian: 4%
Policy: 3%
Country Programme: 1%
Regional/Multi-country programme: 1%

% of good quality reports

Programme: 59%
Pilot: 54%
Project: 58%
At Scale: 67%
RTE: 83%
Humanitarian: 33%
Policy: 100%
Country Programme: 100%
Regional/Multi-country programme: 100%
Results

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Impact</th>
</tr>
</thead>
<tbody>
<tr>
<td>71%</td>
<td>42%</td>
</tr>
<tr>
<td>30%</td>
<td>52%</td>
</tr>
<tr>
<td>18%</td>
<td></td>
</tr>
</tbody>
</table>

% of reports by results level

% of good quality reports
Stage

<table>
<thead>
<tr>
<th>Stage</th>
<th>% of reports by stage</th>
<th>% of good quality reports</th>
</tr>
</thead>
<tbody>
<tr>
<td>Formative</td>
<td>24%</td>
<td>68%</td>
</tr>
<tr>
<td>Summative</td>
<td>39%</td>
<td>58%</td>
</tr>
<tr>
<td>Formative and Summative</td>
<td>37%</td>
<td>62%</td>
</tr>
</tbody>
</table>
MTSP Correspondence

Multi-Sectoral: 43%
Young child survival & development: 14%
Basic education & gender equality: 13%
Cross-cutting: 10%
Child Protection: 8%
Organizational performance: 6%
HIV/AIDS & Children: 5%
Policy advocacy & partnerships: 1%

% or reports by MTSP correspondence

% of good quality reports

43% 14% 13% 10% 8% 6% 5% 1%
56% 73% 50% 100% 50% 100% 25% 0%
Level of Independence

- Independent external: 63%
- Independent internal: 28%
- Not clear: 22%

% of reports by level of independence

- 77% satisfactory reports

% of satisfactory reports
Report Language

- **English**: 84% reports, 62% good quality reports
- **French**: 13% reports, 70% good quality reports
- **Spanish**: 4% reports, 33% good quality reports

© Universalia
Key Findings

Geographic Scope
- Most evaluations continue to be at the national or sub-national level.
- The percentage of good quality reports has proportionally increased for all of the categories of geographic scope.

Management of the Evaluation
- The quality of UNICEF managed or co-managed evaluations has increased markedly since 2011

Purpose of the Evaluation
- The quality of reviewed submissions has improved for nearly all types of evaluation purposes (programme, policy, etc...).

Results
- Reports that address outcomes continue to under-perform with regards to meeting UNICEF quality standards.
Key Findings

**Stage**
- There are no significant trends across evaluations of different stages of an initiative – summative, formative, or a combination thereof.

**MTSP Correspondence**
- A larger proportion of evaluation reports were multi-sector and performance levels have increased in most of the focus areas compared to 2011.

**Level of Independence**
- There is an upward move towards externally managed independent evaluations since 2009.

**Language**
- English is the most common language of reports. No discernible differences in quality can be linked to language.
Trends by Quality of Assessment Category

Findings
Trends by Assessment Category

- Good Quality Ratings per Section—progression 2010-2012

![Bar chart showing trends in good quality ratings by assessment category from 2010 to 2012.](chart.png)
Inclusion of Human Rights, Gender, and Equity: Good Quality Ratings 2010-2012

- Human rights: 19% (2010), 34% (2011), 44% (2012)
- Gender: 20% (2010), 34% (2011), 46% (2012)
- Equity: 9% (2010), 30% (2011), 41% (2012)
### Key Findings - by Report Section

<table>
<thead>
<tr>
<th>Section</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Object of the evaluation</td>
<td>- The extent to which the object of the evaluation is clearly defined has considerably increased by 20%. In comparison to previous years, the articulation of a clear theory of change or results logic remains weak overall.</td>
</tr>
<tr>
<td>Evaluation Purpose, Objectives and Scope</td>
<td>- The proportion of reports that clearly present the purpose, objectives and scope of an evaluation increased only marginally by 4% from the previous cycle of reviews. Failure to specify the underlying questions and criteria guiding an evaluation remains problematic.</td>
</tr>
</tbody>
</table>
| Evaluation Methodology, Gender, Human Rights and Equity | - Compliance with UNICEF standards regarding gender, human rights, equity and methodological considerations continues to progress, though achievements on these issues remain weak, relative to other dimensions.  
   - Equity, in particular has markedly improved since 2010 from 9% to 41%. |
| Findings and Conclusions                          | - All areas pertaining to findings and conclusions have improved and as a result this section improved from 44% to 58%. Areas for further improvement include better analysis on cost-related factors and development of conclusions that add value to report findings. |
| Recommendations and Lessons Learned               | - Though improvements in the sections on recommendations and lessons learned are accruing at a more modest pace, evidence suggests that substantial gains in these areas could be achieved by clearly differentiating lessons learned from the conclusions and recommendations and by improving the usefulness of recommendations. |
| Report structure, logic and clarity               | - Overall, evaluation reports are clearly and logically structured, and executive summaries tend to be relatively complete, though more could be done to ensure that executive summaries can stand alone. |
Outstanding Reports

• Real-Time Independent Assessment (RTIA) of UNICEF’s Response to the Sahel Food and Nutrition Crisis, 2011–2012 (WCARO)

• Evaluation of the Roma Good Start Initiative (CEE/CIS)
Conclusions

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

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

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

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

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

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

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

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

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

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

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