Prepared for: UNICEF Evaluation Office
Background

Context

• UNICEF’s GEROS aims to ensure that the evaluations managed or commissioned by UNICEF uphold the high quality standards set for them.
• Through the GEROS process and the information obtained from annual meta-evaluations, UNICEF can monitor its progress, identify its strengths and be aware of its areas for improvement with regards to evaluation.

Objectives

• Provide senior managers with a clear and succinct assessment of evaluation reports;
• Strengthen internal evaluation capacity;
• Report on the quality of evaluation reports; and
• Contribute to the EO’s corporate knowledge management and organisational learning.
Methodology – Review Process

• The GEROS assessment tool contains 58 questions derived from the UNICEF-adapted UNEG evaluation report standards.

• Reports are classified based on region, geographic scope, management type, purpose, results-level, MTSP correspondence, level of independence and approach.

• A total of 96 reports were reviewed in this exercise.

• Reports are rated according to a four-point performance scale, and qualitative comments are provided to support each rating given.
## 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 section of the report meets UNICEF’s Report Quality Standards. The report/section is of good quality.

The report/individual section of the report does not meet UNICEF’s Report Quality Standards.
Methodology – Meta-evaluation

- Quantitative and qualitative data from the 96 reports reviewed were aggregated and analysed.

- Overall ratings, ratings by report type and report section were considered.

- Qualitative data was analysed in light of the quantitative data: emerging themes were drawn out in order to support the quantitative results obtained.
Limitations

- Use of best judgment to apply criteria to each report. Inclusion of ToRs and Inception Report often helped justify certain choices, approaches or foci.

- Some unique or special cases required deeper analysis in order to ensure consistency of ratings.

- Different evaluation types and designs put the flexibility of the template to the test, and may have influenced rating consistency (e.g. impact evaluations, case studies, and evaluation research).

- Limited data from 2009 sometimes made it difficult to compare GEROS results over the full five-year implementation period.
Overall Ratings

Findings
Reports Reviewed Per Region Per Year

• Between 2012 and 2013, reports reviewed increased from 79 to 96. The most important portion of this increase was generated by WCARO and TACRO. CEE/CIS and ESARO continue to be important contributors.
Overall Ratings for 2013

• In 2013, the quality of reports submitted continued to increase, though less sharply than in 2012.
The number of good quality reports (i.e. Highly Satisfactory and Outstanding/Best Practice) increased by 7 percentage points.
Overall Regional Trends

- In most cases, the quality of UNICEF evaluation reports stayed the same or improved since last year’s review process. In some instances, these improvements were considerable.
Findings

Trends by Type and Scope of Evaluation
Geographic Scope

- **Sub-national**: 34%
- **National**: 47%
- **Multi-country**: 1%
- **Regional**: 6%
- **Multi-region/global**: 11%

% of reports by scope

- **Sub-national**: 58%
- **National**: 64%
- **Multi-country**: 1%
- **Regional**: 6%
- **Multi-region/global**: 11%

% of good quality report
Management of the Evaluation

% reports by management  % good quality reports

- UNICEF: 48%  - 82%
- Joint with UN: 18%  - 60%
- Joint with other: 10%  - 43%
- Joint with country: 0%  - 50%
- Country-led: 7%
- Externally Managed: 8%  - 63%
- Not Clear: 8%

© Universalia
Purpose of the Evaluation

- **3%** Pilot
- **8%** At Scale
- **2%** Policy
- **1%** RTE
- **8%** Humanitarian
- **20%** Project
- **50%** Programme
- **5%** Country Programme
- **2%** Impact Evaluation

% of reports by purpose:
- **100%**
- **100%**
- **100%**
- **100%**
- **88%**

% Good quality:
- **58%**
- **63%**
- **50%**
- **50%**
Results

% of report by results level  % Good quality evaluations

Activities and products & Outputs: 8% 6%
Outcome: 65% 45%
Impact: 27% 18%
Approach

- **Formative**: 29%
- **Summative**: 33%
- **Formative and Summative**: 38%

- **% of reports by approach**: 64%, 66%, 75%
MTSP Correspondence

% of reports by MTSP correspondence

- 31%
- 11%
- 3%
- 11%
- 1%
- 28%
- 10%
- 3%
- 57%
- 55%
- 82%
- 100%
- 100%
- 100%
- 70%

% of good quality reports

- 100%
- 100%
- 100%

Young child survival & development
Basic education & gender equality
HIV/AIDS & Children
Child Protection
Policy advocacy & partnerships
Multi-Sectoral
Cross-cutting
Organizational performance
Level of Independence

- Independent external
- Independent Internal

Percentage changes from 2009 to 2013:
- Independent external: 46% to 82%
- Independent Internal: 45% to 10%
Report Language

- English: 76% Total, 70% % Good quality reports
- French: 15% Total
- Spanish: 9% Total
Key Findings

Geographic Scope
• As in previous years, most reports focused on initiatives at the national and sub-national levels.
• While the quality of sub-national reports decreased compared to 2012, the number of good quality national-level reports increased.

Management of the Evaluation
• The quality of UNICEF-managed or joint-UN reports has continued to improve, with a notable increase in the quality of joint-UN evaluations especially.

Purpose of the Evaluation
• Programme and project evaluations continue to represent the most important proportion of evaluations reviewed, while most other types of evaluations have decreased in number. The quality of most types of reports has either stayed the same or increased since 2012.

Results
• From 2010 through 2013, the number of evaluations that aimed to provide insights on outcome- and impact-level results has generally increased. The number of reports focused on activities and products continues to decrease.
Key Findings

Approach
• The proportion of evaluations using different approaches (summative, formative, or a mix of both) witnessed minimal variation, but the quality of formative and summative reports saw greater change than the other two categories.

MTSP Correspondence
• In 2013, the largest proportion of evaluations covered young child survival and development, closely followed by multi-sectoral evaluations. In most cases, the quality of reports in each category increased between 2012 and 2013.

Level of Independence
• While a pronounced shift in favour of independent external evaluations began as of 2009, this change became even more prominent in 2013.

Language
• As in previous years, the majority of reports submitted were in English, though more reports in French and Spanish were reviewed this year than last. Further, the quality of reports in all three languages increased since 2012.
Trends by Quality of Assessment Category

Findings
Trends by Assessment Category

- Good Quality Ratings per Section — Progression 2010-2013
Inclusion of Human Rights, Gender, and Equity: Good Quality Ratings 2010-2013

<table>
<thead>
<tr>
<th></th>
<th>2010</th>
<th>2011</th>
<th>2012</th>
<th>2013</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human rights</td>
<td>19%</td>
<td>34%</td>
<td>44%</td>
<td>50%</td>
</tr>
<tr>
<td>Gender</td>
<td>20%</td>
<td>34%</td>
<td>46%</td>
<td>52%</td>
</tr>
<tr>
<td>Equity</td>
<td>9%</td>
<td>30%</td>
<td>41%</td>
<td>52%</td>
</tr>
</tbody>
</table>
### 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>• In 2013, the description of the evaluated object and its context continued to improve, but only slightly compared to 2012. Evidence suggests that the description of the theory of change and of stakeholder roles and contributions remain areas for improvement.</td>
</tr>
<tr>
<td>Evaluation Purpose, Objectives and Scope</td>
<td>• The proportion of reports that presented the evaluation framework in a coherent manner increased more considerably in 2013 than in any previous exercise. Though strengths included clear purpose, objectives and scope and a well identified list of criteria, the justification for the selection of criteria still requires greater attention.</td>
</tr>
<tr>
<td>Evaluation Methodology, Gender, Human Rights and Equity</td>
<td>• Reports have made some improvements in terms of the description of the methodology, yet this section remains among the weaker sections, overall. While methodological robustness is often satisfactory, lacunas in ethical considerations and stakeholder participation may have impacted the overall ratings of this section.</td>
</tr>
<tr>
<td>Findings and Conclusions</td>
<td>• The proportion of good quality findings and conclusions continued to increase in 2013. Cost analyses, the clear identification of findings and, to some extent, the insights provided by conclusions remain challenging.</td>
</tr>
<tr>
<td>Recommendations and Lessons Learned</td>
<td>• Despite modest gains in 2013, the number of good quality reports for Section E remained the lowest of all sections. Clearer identification of target stakeholder groups and lessons learned could help improve the ratings for this section.</td>
</tr>
<tr>
<td>Report Structure, Logic and Clarity</td>
<td>• In 2013, Section F surpassed Section A in terms of the proportion of good quality reports. The majority of evaluations were logically structured, but issues were noted regarding the annexes and executive summary of some reports.</td>
</tr>
</tbody>
</table>
Outstanding Reports

• “Expanding Early Learning Opportunities” (Serbia, CEE/CIS)

• “Determining the Results of the Koudmen Sent Lisi Pilot Programme: A Social Safety Net Programme in St. Lucia” (Barbados, TACRO)

• “Joint Evaluation of the UNFPA-UNICEF Joint Programme on Female Genital Mutilation/Cutting (FGM/C): Accelerating Change” (USA, HQ)

• “Evaluation of the UNICEF Child Protection Monitoring and Response System (CPMRS) in Thailand” (Thailand, EAPRO)

• “Evaluación Nacional de los Servicios Amigables de Salud para Adolescentes y Jóvenes en Colombia” (Colombia, TACRO)
Conclusions

• The quality of UNICEF’s evaluation reports continued to increase through 2013. Good quality ratings rose, while unsatisfactory ratings decreased 75%.

• Several trends point to an increase in the number of evaluations produced by UNICEF’s regional offices. This growth reflects additional submissions from TACRO, WCARO, EAPRO and MENARO especially, and points to increased interest among the regions in participating in the GEROS process. Through these added submissions, greater linguistic diversity is also being gradually achieved.

• Reports continue to demonstrate shortcomings in some sections of the GEROS assessment framework, especially in the areas of methodology, recommendations, lessons learned, and compliance with requirements for annexing complementary information.
Recommendations

- UNICEF should identify how GEROS helps strengthen the credibility and use of evaluation, within the broader organisational feedback system.
- Within its decentralised evaluation strategy, UNICEF should continue work to build regional and national capacities to conduct and assess evaluations.
- UNICEF should continue to update and systematically communicate its requirements for evaluation reports across its entire evaluation oversight/management system. These updates should take into account evolving standards for evaluation in the UN System.
- Special efforts should be made to strengthen certain aspects of evaluation reports that have been consistently weak in the past few years.
Lessons Learned

- Clear and systematic communication of evaluation standards and priorities favours the effective alignment of evaluations with UNICEF standards, from the outset (i.e. ToRs stage).

- While common standards help improve evaluation quality, systems should provide sufficient flexibility to account for different types of evaluations.