



Global Evaluation Report Oversight System- Meta-Evaluation 2014

Prepared for: UNICEF Evaluation Office



Background

Background

Context

- **GEROS aims to ensure that the evaluations managed or commissioned by UNICEF uphold high quality standards.**
- **Through the Geros process, UNICEF monitors its progress, identify its strengths and areas for improvement with regards to evaluation reports.**

Objectives

- 1. Provide senior managers with a clear assessment of evaluation reports;**
- 2. Strengthen internal evaluation capacity;**
- 3. Report on the quality of evaluation reports;**
- 4. Contribute to the EO's corporate knowledge management and organisational learning.**

Methodology

- The **GEROS** assessment tool contains **58** questions derived from the **UNICEF-adapted UNEG** evaluation report standards.
- Reports are rated according to a **four-point performance scale**, and **qualitative comments** are provided to support each rating.

Dark Green	Light Green	Amber	Red
Outstanding, Best Practice	Highly Satisfactory	Mostly Satisfactory	Unsatisfactory



The report meets UNICEF's Report Quality Standards.
The report is of good quality.



The report does not meet UNICEF's Report Quality Standards.

Methodology

- **Quality assurance is conducted on each review to ensure consistency and harmonization of reviews. The following methods were used:**
 - **Use of a small team of reviewers with complementary skills**
 - **Team briefings and discussions at the start of a new cycle**
 - **Senior and peer reviews on at least 20% of reviews**
 - **All reviews were read and checked for quality control**
 - **UNICEF's feedback was integrated**

Meta-Evaluation

- **Quantitative and qualitative data from the 69 reports reviewed were aggregated and analysed.**
- **Overall ratings and ratings by section were considered.**
- **When possible, year-to-year comparisons were made.**
- **Qualitative data was analysed in light of the quantitative data: emerging themes were drawn out in order to support the quantitative results obtained.**

Limitations

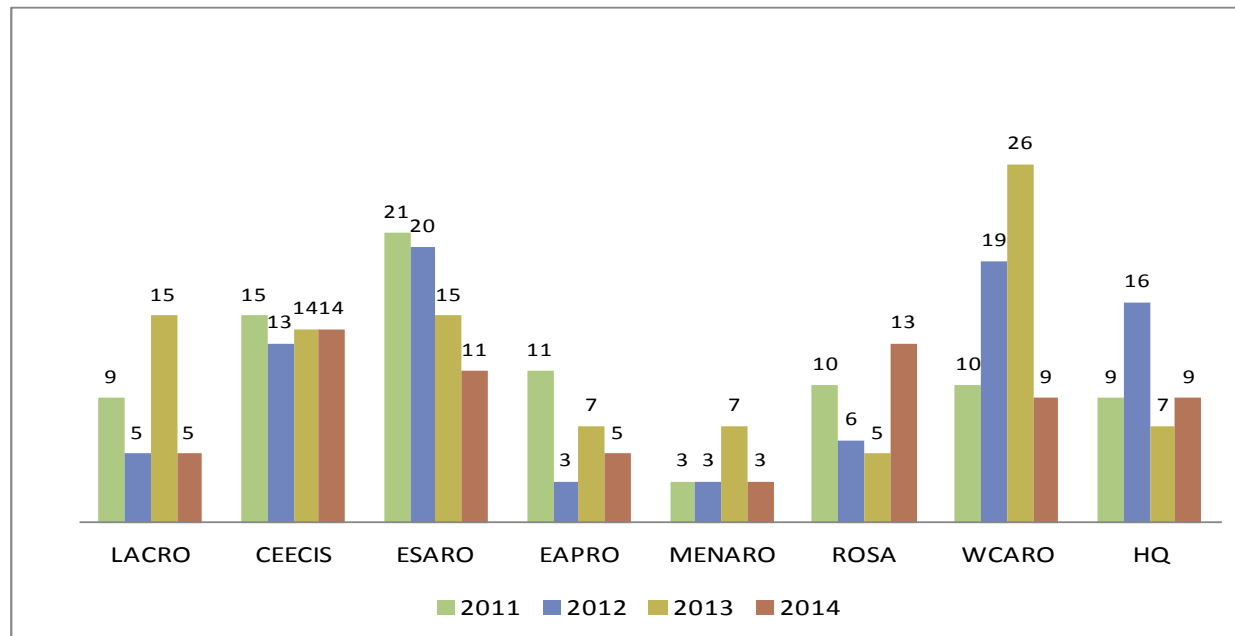
- **Use of best judgment to apply criteria to each report. Inclusion of TORs helped justify certain choices, approaches or foci. Efforts were made to train reviewers and check their work to ensure consistency.**
- **Some unique or special cases required deeper analysis in order to ensure consistency 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 or case studies).**

Overall Ratings

Findings

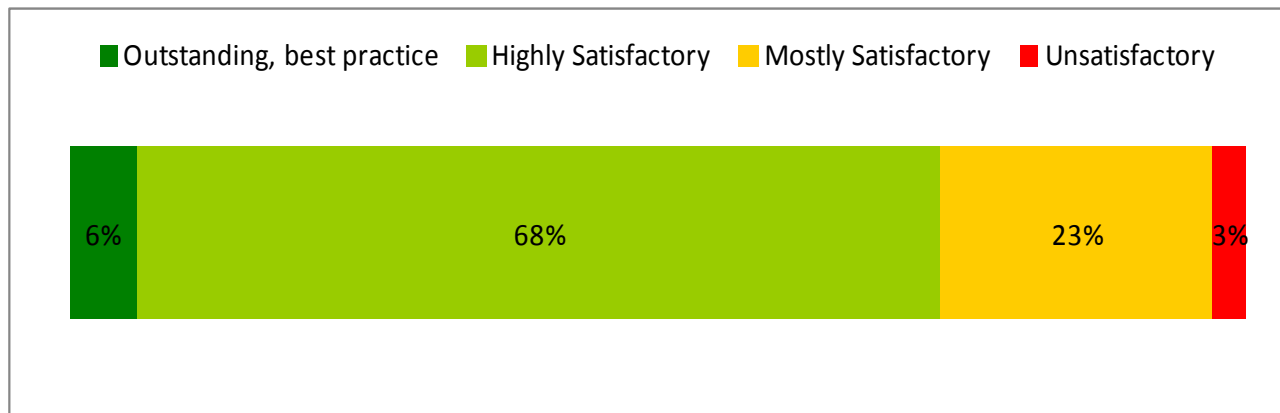
Reports Reviewed per Region, per Year

- **Between 2013 and 2014, reports reviewed decreased from 96 to 69. Many reasons could explain the decrease, but no firm explanation.**
- **The number of reports submitted by CEE/CIS and HQ was stable, while reports from ROSA increased.**



Overall Ratings for 2014

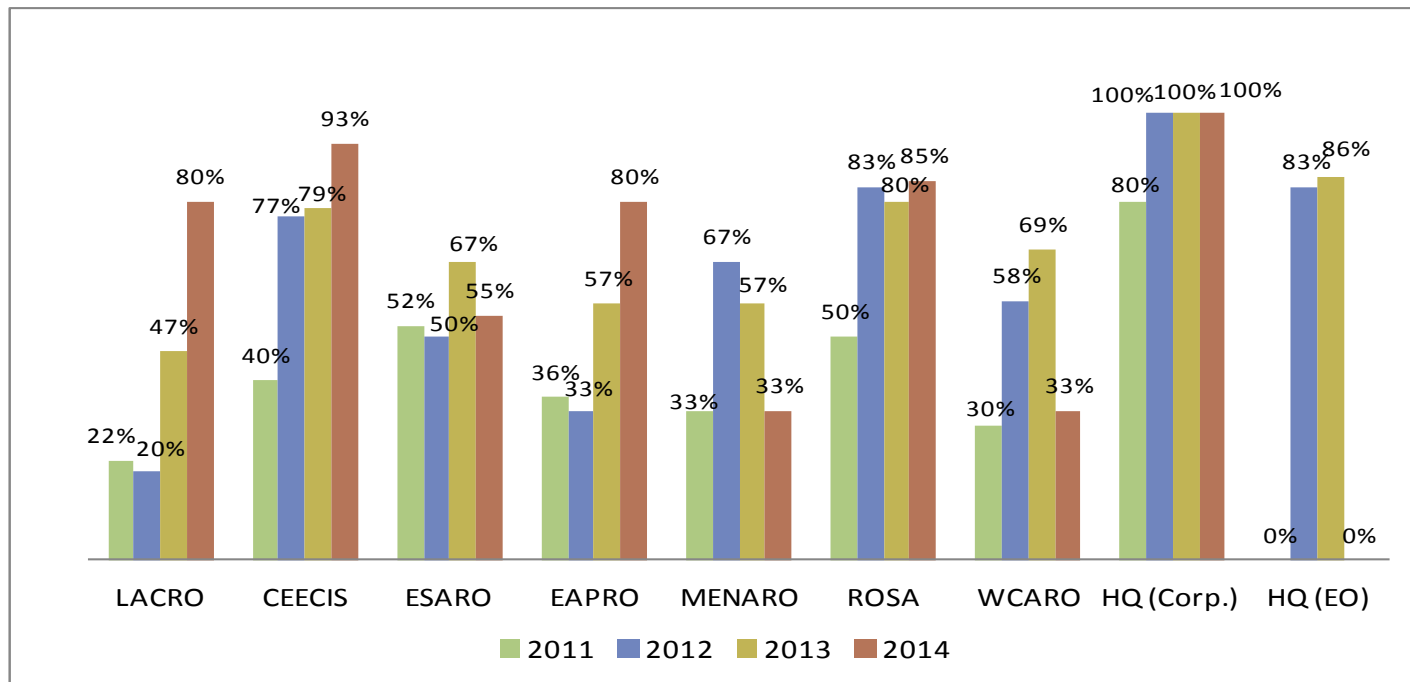
- In 2014, three-fourths of the reports were considered to be good quality reports (i.e. rated Highly Satisfactory and Outstanding/Best Practice).



- The number of good quality reports increased from 69% in 2013 to 74% in 2014.

Overall Regional Trends

- The quality of UNICEF evaluation reports by region varied since last year's review process.
- LACRO, EAPRO and MENARO submitted 5 reports or less in 2014, comparison across regions or years may be unreliable.

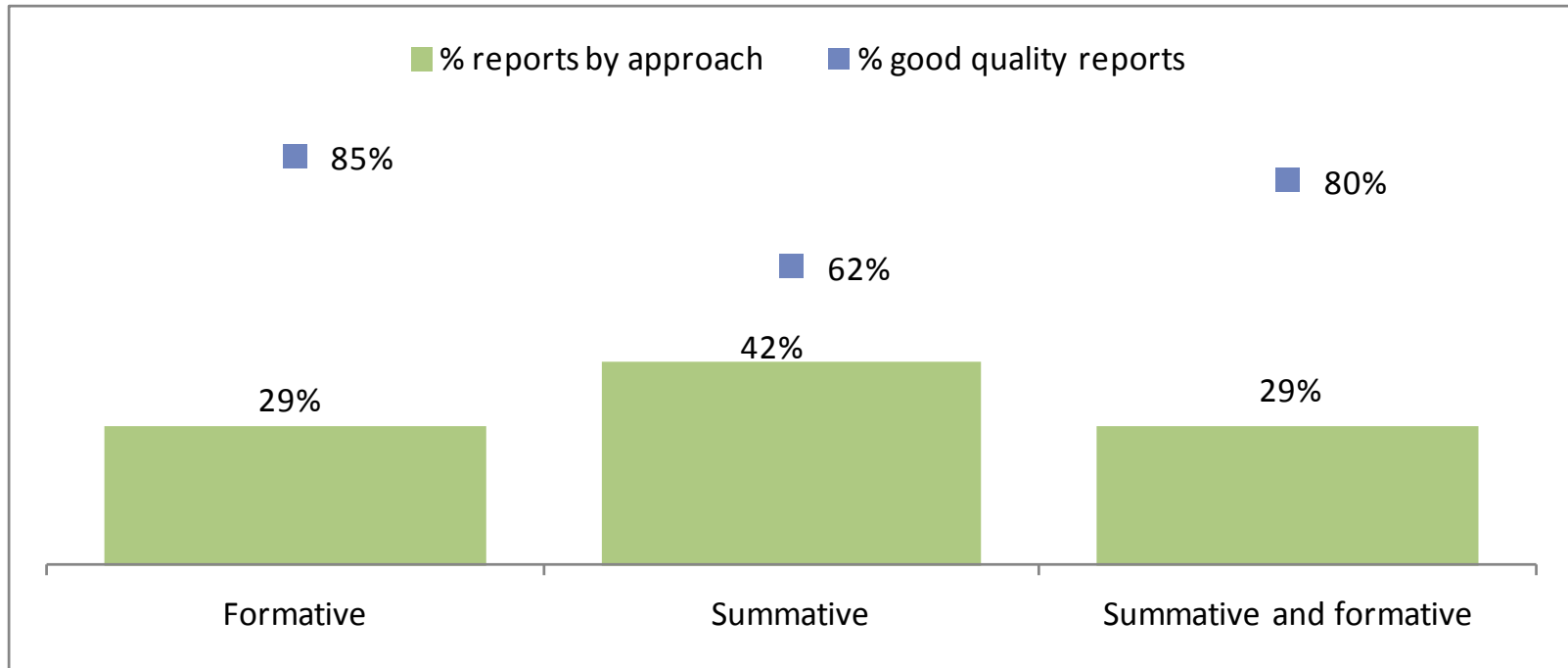




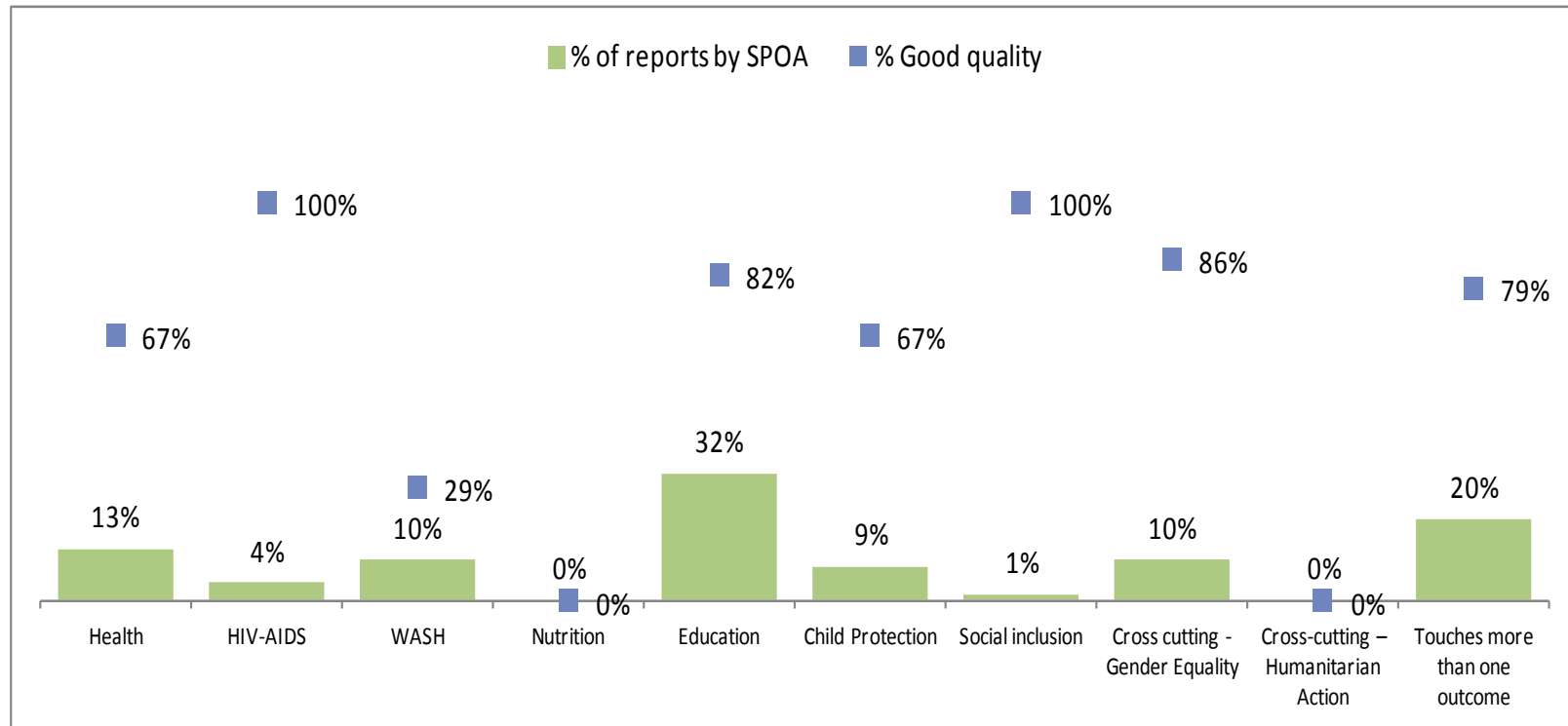
Trends by Type and Scope of Evaluation

Findings

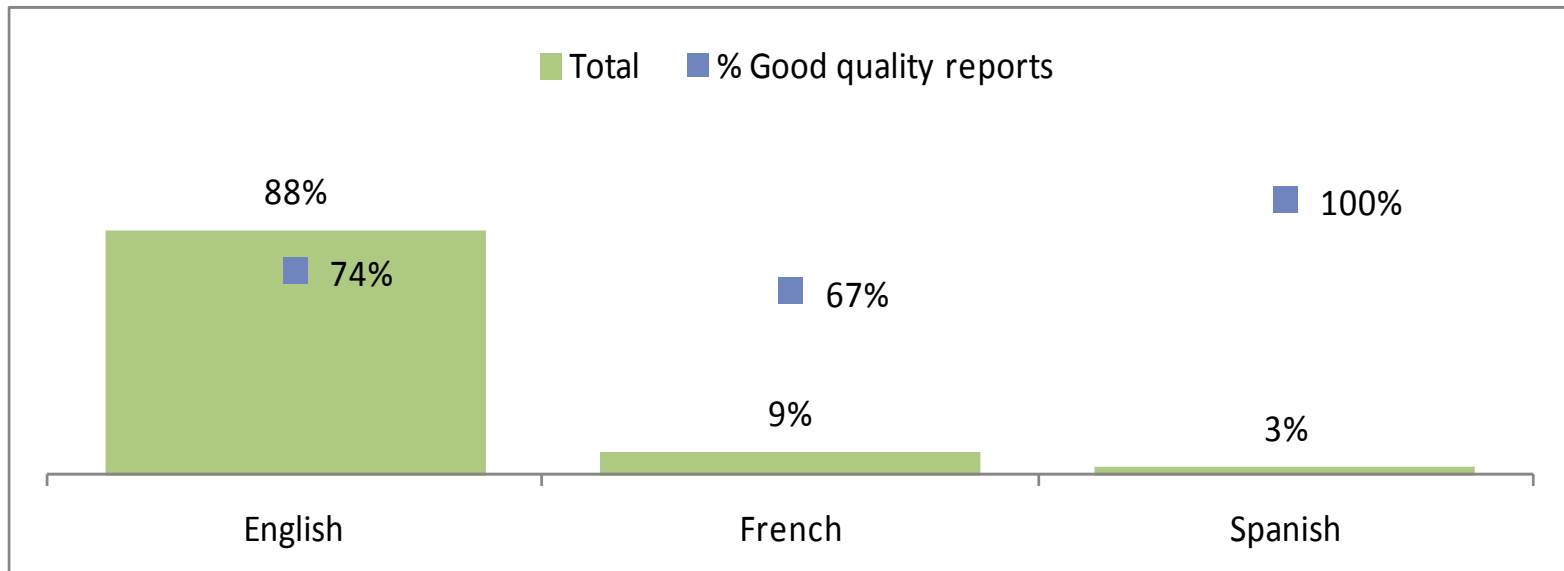
Approach



SPOA Correspondence



Report Language



Key Findings

Geographic Scope

- As in previous years, most reports focused on initiatives at the national level (88%) and they were overall good quality reports (69%).

Management of the Evaluation

- The quality of UNICEF-managed evaluations reached 80%, with good quality also evident in evaluations managed jointly with countries (63%).

Type of Evaluation

- Programme and project evaluations continue to represent the most important proportion of evaluations reviewed (60%). Project-level evaluations increased significantly in quality (from 53% in 2013 to 84% in 2014).

Key Findings

Approach

- A larger proportion of evaluations are summative this year (42%), but their quality has decreased slightly compared to last year. The quality of formative evaluations has improved from 64% in 2013 to 85% in 2014.

SPOA Correspondence

- The largest proportion of evaluations covered education (32%) and their quality was high (82%). The quality of WASH evaluations was lower than other categories with only 29% good quality reports.

Level of Independence

- Almost all evaluations are now conducted by independent external evaluators (99%), completing a rapid transition away from internal evaluations that began after 2012.

Language

- As in previous years, the majority of reports submitted were in English, with a declining number of reports in French and Spanish compared to 2013.

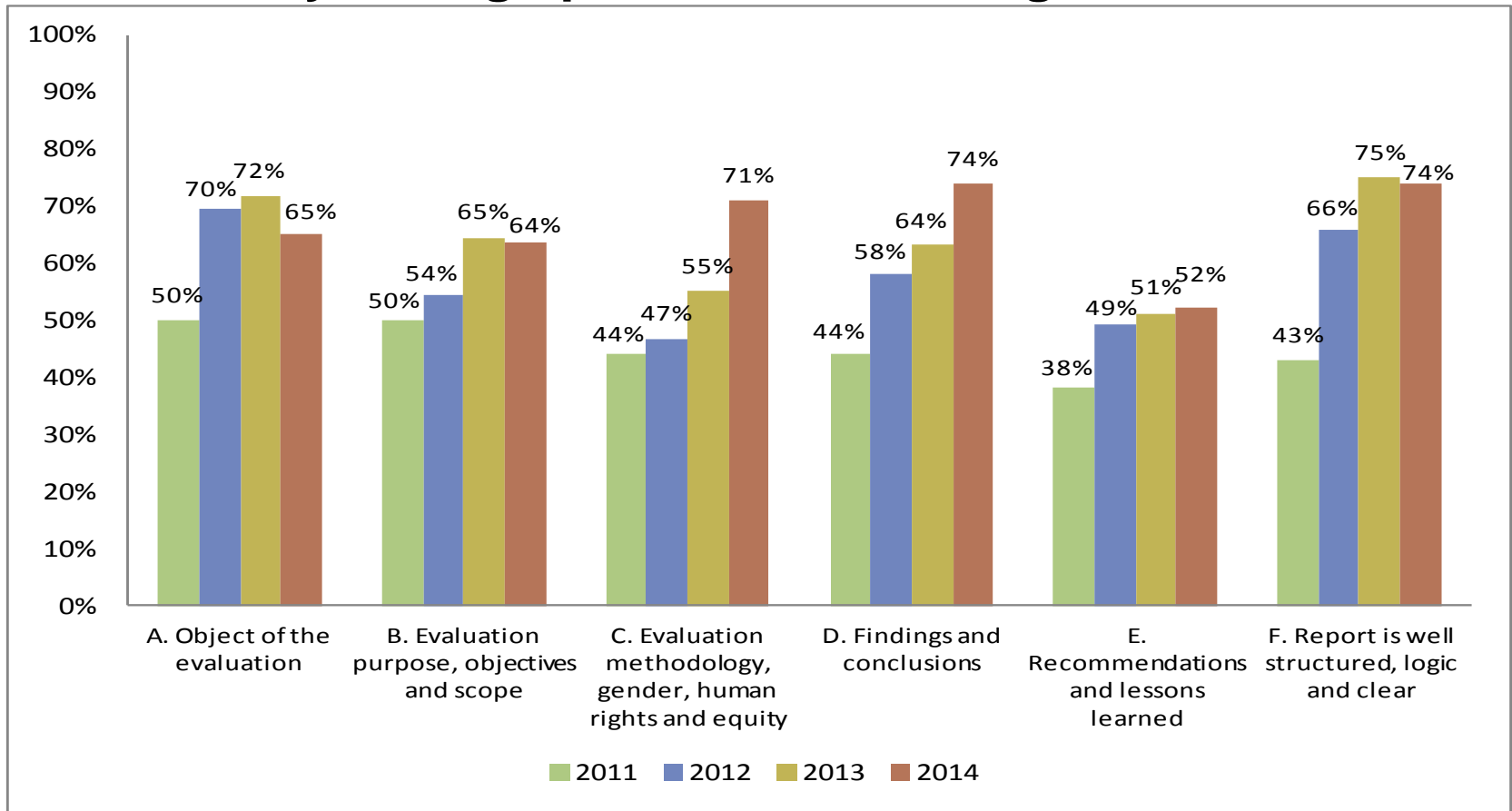
A decorative wavy line in a light blue color, starting from the left edge, dipping down, and then rising towards the right edge.

Trends by Quality of Assessment Category

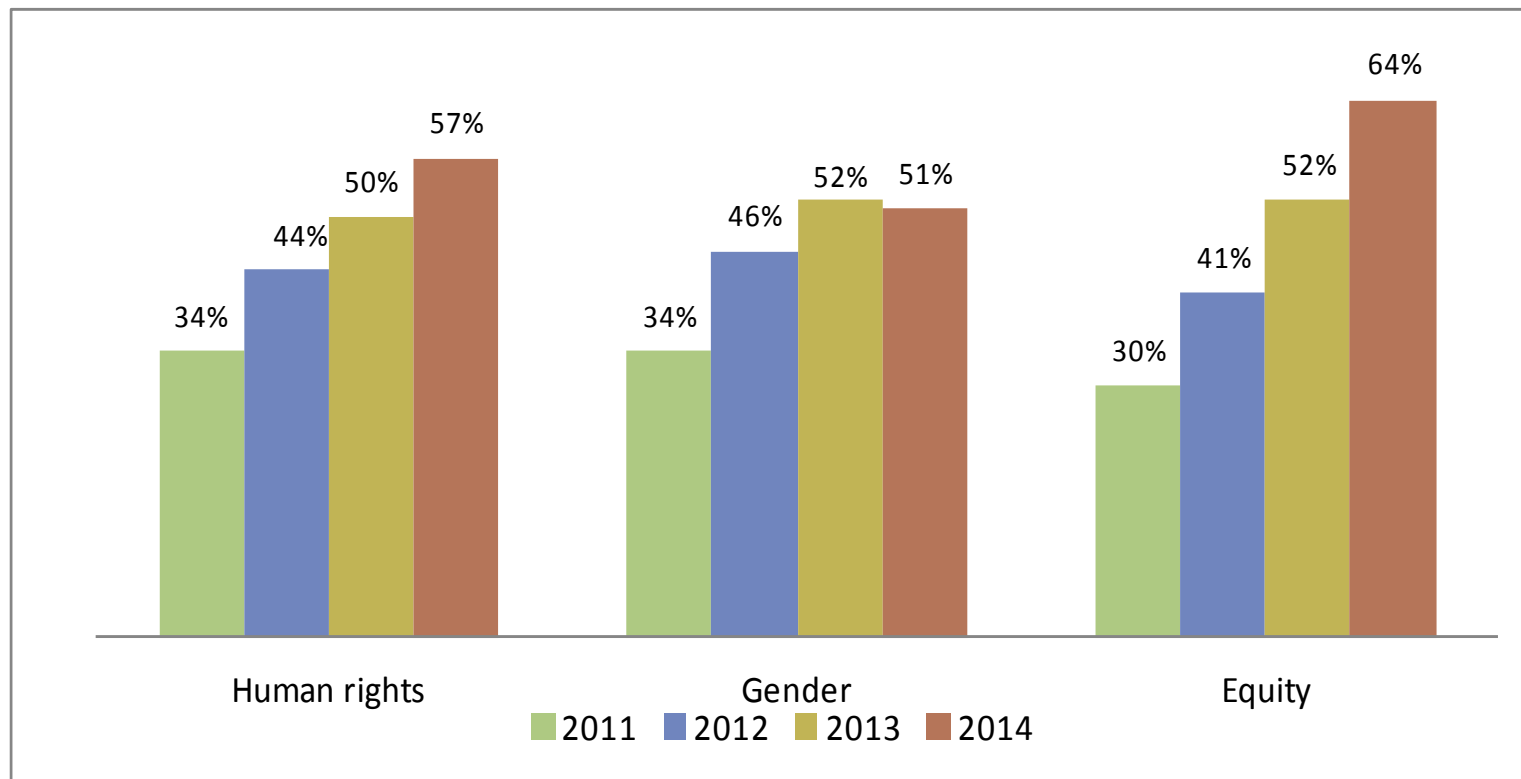
Findings

Trends by Assessment Category

- **Good Quality Ratings per Section — Progression 2011-2014**



Inclusion of Human Rights, Gender, and Equity: Good Quality Ratings 2011-2014



Key Findings by Report Section

A. Object of the Evaluation

• In 2014, the description of the evaluated object and its context declined somewhat in quality compared to 2013. Evidence suggests that the description of the theory of change and of stakeholder roles and contributions remain areas for improvement.

B. Evaluation Purpose, Objectives and Scope

• The extent to which reports met standards of evaluation purpose, objectives and scope changed little from previous years. Strengths of reports in this area included clear purpose, objectives and scope and a clear list of evaluation criteria. However, the justification for the selection of evaluation criteria still requires greater attention.

C. Evaluation Methodology, Gender, Human Rights and Equity

• Reports have made improvements in terms of the description of the methodology. While methodological robustness is often satisfactory, lacunas in ethical considerations and stakeholder participation may have impacted the overall ratings of this section.

D. Findings and Conclusions

• The presentation of findings, conclusions, contribution and causality continued to improve in 2014. Cost analysis remained particularly challenging.

E. Recommendations and Lessons Learned

• Largely unchanged since 2014, 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.

F. Report Structure, Logic and Clarity

• Good quality ratings for Section F did not significantly change compared to 2013, but this Section continues to have amongst the highest overall good quality ratings. The majority of evaluations were logically structured, but issues were noted regarding the length of executive summaries or their ability to stand alone.

Reports rated as Outstanding / Best Practice

- **“Expanding Early Learning Opportunities”
(Afghanistan, ROSA)**
- **“RKLA3 multi-country evaluation: increasing access and equity in early childhood education – final evaluation report”
(Regional, CEE/CIS)**
- **“Evaluation of Country Programme of Co-operation between Government of Uzbekistan and UNICEF 2010-2014”
(Uzbekistan, CEE/CIS)**
- **“Final Evaluation of the "Child Care System Reform"”
(Montenegro, CEE/CIS)**

Conclusions

- 1. The quality of UNICEF's evaluation reports continued to increase overall through 2014, but only moderately.**
- 2. The contribution of reports to the GEROS process by some UNICEF's regional and country offices appears to be dropping.**
- 3. Reports continue to demonstrate similar shortcomings found in 2013.**
- 4. After six years of applying the GEROS template with minor changes, it may be time to more significantly adjust the content and structure of the template to address ongoing issues.**

Recommendations

- 1. UNICEF should examine whether the increase in quality of evaluation reports, as assessed through GEROS, has resulted in senior managers having greater confidence in evaluation reports.**
- 2. Within its decentralised evaluation strategy, UNICEF should continue to build its own regional/country office evaluation capacities and national capacities to conduct relevant types of evaluations.**
- 3. Special efforts should be made to strengthen certain aspects of evaluation reports that have been consistently weak in the past few years.**

Recommendations

- 4. 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.**
- 5. As part of the periodic review of GEROS, UNICEF should consider revising the rating scale and several elements of the GEROS template in order to ensure greater precision in the messages that are provided about evaluation quality and the characteristics of evaluation reports, and to create more efficiency in applying the template.**

Lessons Learned

- **Clear and systematic communication of evaluation standards and priorities favours the effective alignment of evaluations with UNICEF standards, from the outset.**
- **Quality assurance systems such as GEROS should provide sufficient flexibility to account for different types of evaluations.**
- **Compliance with quality assurance systems such as GEROS is affected by incentives, available resources, and the perception of relevance.**
- **Quality assurance systems such as GEROS need to strike a balance between consistent application over a period of time (which allows for comparison) and making major adjustments in order to improve utility and reflect changes in the environment.**