UNICEF GEROS Meta-Analysis 2018 Regional Profiles
Headquarters (HQ)

Overview

Introduction
This brief provides an overview of the quality of the evaluation reports submitted to UNICEF’s Global Evaluation Reports Oversight System (GEROS) by the HQ during 2018, and provides a limited annual comparison and trend analysis. The brief presents: 1) the HQ 2018 evaluation portfolio; 2) trends in the overall quality of evaluation reports for the region; and 3) trends in the quality of evaluation reports per assessment criteria for the HQ. This brief is intended to accompany the global 2018 GEROS Meta-Analysis report.

Regional Performance Compared to Global Trends
A trend analysis from 2011 until today indicates that the quality of HQ evaluation reports was well above the global average from 2011 – 2014 (with 100 percentage of evaluations rated as satisfactory or highly satisfactory in 2012, 2013, and 2014) until the overall quality was halved in 2015, reaching a low of only 50 percent of evaluations rated satisfactory or higher. Report quality then fluctuated considerably from 2016 to 2018 by rising to 80 percent in 2016, dropping to 60 percent in 2017, and then again rising to 91 percent in 2018. The percentage of evaluations rated satisfactory or higher commissioned by HQ was 4 percent above the global average in 2018.
**Exhibit 1. Percentage of Reports Rated Satisfactory or Highly Satisfactory from 2011 - 2018**

<table>
<thead>
<tr>
<th>Year</th>
<th>No. of Reports Rated Satisfactory or Above</th>
<th>Satisfactory (Region)</th>
<th>Satisfactory (Global)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2011</td>
<td>7</td>
<td>42%</td>
<td>100%</td>
</tr>
<tr>
<td>2012</td>
<td>7</td>
<td>62%</td>
<td>100%</td>
</tr>
<tr>
<td>2013</td>
<td>9</td>
<td>69%</td>
<td>100%</td>
</tr>
<tr>
<td>2014</td>
<td>4</td>
<td>74%</td>
<td>100%</td>
</tr>
<tr>
<td>2015</td>
<td>10</td>
<td>50%</td>
<td>100%</td>
</tr>
<tr>
<td>2016</td>
<td>5</td>
<td>80%</td>
<td>100%</td>
</tr>
<tr>
<td>2017</td>
<td>5</td>
<td>72%</td>
<td>100%</td>
</tr>
<tr>
<td>2018</td>
<td>11</td>
<td>91%</td>
<td>87%</td>
</tr>
</tbody>
</table>

**HQ 2018 Evaluation Portfolio**

In 2018, HQ produced a total of 11 evaluation reports, placing it as the third region to produce the least number of evaluations that year, along with LACR.

In 2018, HQ evaluations assessed objects of evaluation that touched upon a wide range of thematic areas, with humanitarian action and education comprising the greatest percentage of the region’s portfolio (at 14 percent each) followed by health (13 percent); gender equality (12 percent); child protection (12 percent); WASH (12 percent); social inclusion (10 percent); nutrition (8 percent); and HIV/AIDS (5 percent). The thematic breakdown of evaluations commissioned by HQ in 2018 was overall quite similar to the global average. However, HQ commissioned nearly twice as many humanitarian evaluations. The 2018 HQ portfolio was comprised of a generally even breakdown per 2018 – 2021 Strategic Plan goal area, with the largest percentage (20 percent) of evaluations assessing cross-cutting priorities.

In 2018, 46 percent of HQ evaluations were programme evaluations, which was in line with the global average (49 percent). The office commissioned a much higher percentage of thematic evaluations compared to the global average, with 36 percent (4/11 evaluations) of its evaluation profile dedicated to this area (compared to 5 percent at the global level). The office also commissioned one strategy and one system evaluation.

The office also focused more on learning with 91 percent of its portfolio consisting of either formative and formative and summative evaluations compared to the global average of 72 percent.
In 2018, HQ drew on a variety of evaluation designs. The most common were theory-based at 28 percent of its portfolio (3/11 evaluations), case study (18 percent), and quasi-experimental (18 percent). The office used a larger percentage of case study designs than the global average and a smaller percentage of quasi-experimental designs, which is to be expected since case studies are particularly well suited to cover several countries through thematic global evaluations. The rest of its portfolio was comprised of four evaluations that drew on other designs (longitudinal, retrospective, systematic review, and other).

HQ drew primarily on mixed methods (91 percent) and some exclusively quantitative strategies (9 percent). The office did not use any exclusively qualitative evaluation strategies. All evaluations from HQ in 2018 were at the outcome (82 percent) or impact levels (18 percent) with no evaluations exclusively at the output level. The majority (64 percent) of evaluations were at the global level while 36 percent (4/11 evaluations) were at the national level. All evaluations at HQ were managed exclusively by UNICEF.

**Trends in the Overall Quality of Evaluation Reports for HQ**

In 2018, HQ had the 4th highest overall score for evaluation report quality among regions tied with WCAR at 3.01/4.0. The region scored 8.8 percent below the highest-scoring region (EAPR at 3.36) and 7 percent above the lowest-scoring region (ROSA at 2.73). As shown in Exhibit 2, the average quality of evaluations at HQ increased by 12.3 percent between 2017 and 2018.

*Exhibit 2: Overall score of HQ compared with other regions from 2016 - 2018*

By looking at a breakdown of evaluations per assessment rating, we can see from Exhibit 3 that 91 percent of HQ evaluations in 2018 were rated as satisfactory or higher, which constitutes a 41
percent increase from the previous year. Also, whereas no evaluations rated in the upper satisfactory and highly satisfactory categories in 2017, 64 percent of HQ reports did in 2018.

**Exhibit 3: Percentage of Reports per Assessment Rating for HQ from 2016 – 2018**

![Percentage of Reports per Assessment Rating for HQ from 2016 – 2018](chart.png)

**2018 HQ Quality of Evaluation Reports by Assessment Criteria**

GEROS is made up of 9 assessment criteria, as outlined in the description on GEROS assessment ratings in Annex II. Exhibit 4 provides an overview of the performance of HQ per assessment criterion in 2018. As can be seen in Exhibit 4, HQ was strongest in 2018 at presenting the evaluation purpose, objectives and scope (3.36), at providing a complete and informative background section (3.27), and at presenting actionable recommendations (3.18) as well as complete yet succinct executive summaries (3.18).

HQ had the highest average rating among all regions for evaluation principles (3.0) in 2018. On the other hand, reports from HQ ranked 2nd lowest, just above ESAR, at presenting the evaluation methodology (2.73). This was the lowest scoring criterion for HQ in 2018 followed by conclusions and lessons learned (2.82).

**Exhibit 4: 2018 Regional Evaluation Report Quality per Assessment Criterion**
UN System-Wide Action Plan for Gender Equality and Empowerment of Women (SWAP)

The “principles” assessment criterion is composed of three assessment questions, one of which is related to gender equality and SWAP principles. As Exhibit 7 indicates, HQ ranked 5th out of 8 regions in 2018 regarding its integration of gender equality and the empowerment of women (GEEW) principles into its evaluations with an average score of 64 percent, which is approaching SWAP requirements.

Exhibit 5: 2018 SWAP Scores by Region
A closer look at the three different SWAP criteria in Exhibit 6 shows that evaluations from HQ struggled most with SWAP criterion 2 by only partially integrating SWAP principles into the use of a gender responsive methodology, methods and tools, and data analysis techniques. HQ reports ranked 2nd lowest among regions for this criterion (1.55), only above ROSA (1.44), and 8.3 percent below the global average for SWAP criterion 2. On average, the region satisfactorily integrated SWAP principles into criteria 1 (evaluation scope and design) and 3 (gender analysis) at an average score of 2.09 each.

Exhibit 6: 2018 SWAP Criterion by Region

2018 Overall Strengths

- **Clear Evaluation Purpose, Objectives and Scope**: HQ evaluations in 2018 were generally strong at clearly laying out evaluation objectives and at identifying specific evaluation users and uses (i.e. UNICEF departments and government ministries as opposed to UNICEF or national governments in general). They also generally provided a good description of both what is and is not included in the evaluation scope in terms of thematic, geographic, and time-bound characteristics. (Ex: 2.B. Annex V; and EO 2018-011 in Annex IV).

- **Background Section**: Reports in HQ generally presented strong background sections that provided a good description of the object of the evaluation along with an informative explanation of the national/regional context in which the object of the evaluation operated.

- **Actionable Recommendations**: HQ evaluations in 2018 were generally strong at developing recommendations that were correctly drawn from the findings and conclusions and that were actionable, as they identified the stakeholder groups for action in each case. Recommendations were also duly prioritized. (Ex, HQ 2018-009 in Annex IV).
- **Strong Executive Summary:** HQ executive summaries were in general complete, thorough and yet succinct with a length of not more than 6 pages to effectively inform decision-makers.

### 2018 Overall Weaknesses

- **Methods:** While most HQ evaluations appear to use appropriate evaluation designs and methods, reports often do not adequately explain why these designs and methods are well suited for meeting the evaluation purpose and objectives. Additionally, reports often do not explain why evaluations diverted away from the standard OECD/DAC evaluation criteria. Evaluation limitations are usually identified but not all reports provide mitigation strategies to address the limitations. Even though reports are becoming stronger at describing how the evaluation followed ethical principles, it is common for evaluation reports to inadequately describe the ethical obligations of the evaluators. (To learn more about how to improve in this area, see good practices 3.B, 3.F, and 3.H in Annex V).

- **Conclusions:** In general, HQ reports could benefit from conclusions that present a more forward-looking analysis that goes beyond the information presented in the findings and that discusses potential future implications. (To learn more about how to improve in this area, see good practice 5.A in Annex V).

- **Lessons Learned:** HQ reports in general struggled to provide correctly identified lessons learned that contributed to the organisation’s institutional knowledge and whose relevance could make them applicable to similar initiatives taking place or projected to take place in other contexts. In some cases, evaluations did not identify any lessons learned whatsoever. In some other cases, lessons learned were confused with best practices that primarily highlight success stories as opposed to critical moments of learning from what did not work well. In total, only 36.4 percent of HQ evaluation reports presented satisfactory lessons learned. (To learn more about how to improve in this area, see good practice 5.C in Annex V).

See Annex IV for a list of strengths and weaknesses with examples and Annex V for a list of good practices emerging from 2018 evaluations, including those from HQ.