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 EAPR during 2018, and provides a limited annual comparison and trend analysis. The brief presents: 1) the EAPR 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 EAPR. 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 EAPR initially struggled to produce satisfactory evaluation reports, with the regional average for satisfactory reports dipping to well below the global average in 2012 but then increasing sharply to surpass the global average in 2014. Since then, the regional average dropped in 2015 in line with a global decrease in quality but then continued to improve until 2018. As indicated in Exhibit 1 below, a total of 92 percent of the EAPR evaluation reports were rated as satisfactory or highly satisfactory in 2018, which was 5 percent above the global average.
Exhibit 1. Percentage of reports rated satisfactory or highly satisfactory from 2011 - 2018

EAPR 2018 Evaluation Portfolio

In 2018, the EAPR produced a total of 12 evaluation reports, placing it as the third region (along with ECAR) to produce the most evaluations that year. As seen in Exhibit 2, Vietnam and Cambodia produced the largest number of reports per country at 3 reports each.

Exhibit 2. Number of evaluation reports per country in EAPR in 2018
In 2018, the EAPR evaluation portfolio was comprised primarily of five thematic areas (health, WASH, education, child protection, and social inclusion at 14 or 15 percent of the total portfolio each). The remaining elements of the portfolio were aligned with nutrition (12 percent); gender equality (10 percent); humanitarian action (4 percent); and HIV/AIDS (2 percent). In 2018, EAPR commissioned nearly twice as many WASH evaluations as the global average (15 percent compared to 8 percent).

The regional portfolio was most aligned in 2018 with the 2018-2021 Strategic Plan goal “every child survives and thrives” (19 percent). This was followed by “every child is protected from violence and exploitation” (17 percent); “every child lives in a safe and clean environment” (17 percent); “every child has an equitable chance in life” (17 percent); “every child learns” (16 percent); and “cross-cutting priorities” (14 percent).

While half of the region’s 2018 portfolio consisted of programme evaluations, EAPR commissioned more strategic evaluations (17 percent of its portfolio) compared to the global average (9 percent). The rest of the portfolio was comprised of one (1) thematic; 1 country programme; 1 pilot/innovation; and 1 project evaluation.

The region also focused more on learning with 92 percent of its portfolio consisting of either formative and formative and summative evaluations compared to the global average of 72 percent.

In 2018, EAPR drew on primarily three types of evaluation designs (quasi-experimental at 50 percent, theory-based at 33 percent, and participatory at 17 percent). Due to the significant percentage of theory-based and participatory evaluations, the region drew more heavily on qualitative evaluation methods than the global average at 17 percent compared to 7 percent. Mixed methods accounted for the remaining 83 percent.

Over half (58 percent) of all EAPR evaluations in 2018 examined outcome-level results while the remaining evaluations (42 percent) examined impact. All evaluations in 2018 were commissioned at the country level and were national in scope. Most evaluations (11/12) were managed exclusively by UNICEF while one (1) was managed jointly with another organisation outside the UN system.

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

In 2018, EAPR had the highest overall score for evaluation report quality among regions at 3.36/4.0. The region scored 5.8 percent above the next highest-scoring region (ECAR at 3.13) and 15.8 percent above the lowest-scoring region (ROSA at 2.73). As shown in Exhibit 3, the average quality of evaluations in EAPR increased by 13.3 percent between 2017 and 2018.

---

*Exhibit 3: Overall score of EAPR compared with other regions from 2016 - 2018*
As shown in Exhibit 4, eight countries produced evaluation reports in 2018 with all obtaining an average score of satisfactory or above with the exception of Lao People’s Democratic Republic that scored 2.40. A total of 3 countries (Cambodia, Democratic People’s Republic of Korea, and Philippines) obtained an average score in the highly satisfactory ranking (3.5 or higher).

**Exhibit 4: Overall score per country in EAPR for 2018**

By looking at a breakdown of evaluations per assessment rating, we can see from Exhibit 5 that 92 percent of EAPR evaluations in 2018 were rated as satisfactory or higher, with a significant
increase of reports rated in the upper satisfactory and highly satisfactory categories than in the previous year.

*Exhibit 5: Percentage of Reports per Assessment Rating for EAPR from 2016 – 2018*
2018 EAPR 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 6 provides an overview of the performance of EAPR per assessment criterion in 2018.

As can be seen in Exhibit 6, the region was strongest in 2018 at presenting the evaluation purpose, objectives and scope (3.67), a strong executive summary (3.67), and at presenting complete information in a logical and easy to use report format (3.75).

In 2018, 4/8 countries (Philippines, Indonesia, Democratic People’s Republic of Korea, and Cambodia) in EAPR received perfect scores (4/4) for evaluation purpose, objectives and scope. Over half (5/8) of all countries (Timor-Leste, Philippines, Indonesia, Democratic People’s Republic of Korea, and Cambodia) in EAPR received perfect scores (4/4) for their executive summaries. This was also the case where over half (5/8) of all countries (Vietnam, Timor-Leste, Philippines, Indonesia, and Democratic People’s Republic of Korea) in EAPR received perfect scores (4/4) for their evaluation report presentation.

In 2018, the region was weakest at integrating evaluation principles (2.92), followed by findings (3.25), and evaluation methods (3.33). These areas of weakness reflect those that are most common at the global level.

Most countries scored near the regional average for evaluation principles, indicating that weaknesses around this criterion are generally shared across countries. There was greater diversity among country scores regarding findings, with some countries (Timor-Leste and Philippines) receiving a perfect score (4.0) while other countries fell below the regional average, such as Lao People’s Democratic Republic with a score of 2.0. This was also the case for evaluation methods, where the Democratic People’s Republic of Korea and Cambodia scored a perfect 4.0, while Lao People’s Democratic Republic scored only 2.0.

Exhibit 6: 2018 Regional Evaluation Report Quality per Assessment Criterion
In terms of placement compared to other regions in 2018, EAPR had the top average score for all assessment criteria with the exception of evaluation principles, where it had the second highest score at 2.92 after HQ (3.0).

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, while EAPR was a top scorer in 2018 across assessment criteria, it ranks 6th out of 8 regions regarding its integration of gender equality and the empowerment of women (GEEW) principles into its evaluations. With a score of 63 percent, EAPR was approaching SWAP requirements in 2018.

Exhibit 7: 2018 SWAP Scores by Region

A closer look at the three different SWAP criteria in Exhibit 8 shows that evaluations in the EAPR were strongest at SWAP criterion 1 and satisfactorily integrated GEEW in the evaluation scope of analysis. It was weakest at SWAP criterion 3 and 2 and only partially integrated GEEW into the evaluation methodology (criterion 2) and analysis (criterion 3). In fact, the EAPR average score was the second lowest across all regions (1.58) for criterion 3, which is 12 percent below the global score for this SWAP criterion.
When broken down by country, as seen in Exhibit 9, it appears that the quality of GEEW integration into evaluation varies greatly across countries. In 2018, Cambodia had the highest average SWAP score at 89 percent followed by Philippines at 78 percent (both meeting SWAP requirements). In contrast, Thailand had the lowest average score at 22 percent (missing requirements).
2018 Overall Regional and Country Strengths

- **Clear Evaluation Purpose, Objectives and Scope:** EAPR evaluations in 2018 presented clear objectives and were generally strong 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.

- **Strong Executive Summary:** EAPR executive summaries were generally complete and at an appropriate length to inform decision-makers. (Ex: Timor-Leste 2018-001 in Annex IV).

- **Correct Report Presentation:** Reports were generally logically structured and included most of the relevant information (including in the annexes). (Ex: Cambodia 2018-002 in Annex IV).

2018 Overall Regional and Country Weaknesses

- **Evaluation Principles:** While many evaluations adopted human rights language throughout the report, only some fully anchored the analysis within a human rights framework. In terms of gender equality (SWAP criteria), indicators that accompany non-gender specific evaluation questions in evaluation matrices were often not designed to collect sex-disaggregated data. In the absence of sex-disaggregated data, few evaluation methodologies were designed to collect qualitative sex-disaggregated data. Additionally, some reports tended to address gender equality under a standalone question, while not always providing a meaningful gender analysis throughout the report. Many such analyses did not sufficiently explore key aspects of gender transformation. The gender analysis is often reflected in conclusions, but not all evaluations provided recommendations for how to improve on gender. (To learn more about how to improve in this area, see an example from Guyana and Suriname 2018-001 in Annex IV).

- **Methods:** While most EAPR evaluations appear to have used appropriate evaluation designs and methods, reports often did not adequately explain why these designs and methods were well suited for meeting the evaluation purpose and objectives. Additionally, while evaluation limitations were usually identified, not all reports necessarily provided robust mitigation strategies to address the limitations. Finally, even though reports are becoming stronger at describing how the evaluation followed ethical principles, it is still common for evaluation reports to inadequately describe the ethical obligations of the evaluators. (To learn more about how to improve in this area, see an example from the Evaluation Office 2018-009 in Annex IV).

- **Findings:** While ample evidence was typically provided around results outputs, several evaluations struggled to present reliable and robust evidence at the outcome level. While evaluations generally discussed both the strengths and weaknesses of the object of evaluation, oftentimes evaluations did not adequately discuss positive and negative unintended effects of the evaluated object. While an increasing number of evaluations included an analysis around the results-based management system of the object of
evaluation, the quality of the analysis fluctuated greatly and the analysis often failed to assess how the system was used to inform decision-making. (To learn more about how to improve in this area, see an example from Zimbabwe 2018-003 in Annex IV).