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 WCAR during 2018, and provides a limited annual comparison and trend analysis. The brief presents: 1) the WCAR 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 WCAR. 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 WCAR was in line with the global average trend of steady improvement in evaluation report quality, with 69 percent of its evaluations rated as satisfactory or higher in 2013 before dropping substantially to 33 percent in 2014 (41 percent below the global 2014 average). The percentage of reports rated as satisfactory or highly satisfactory then steadily increased reaching 74 percent in 2016 before falling to 59 percent in 2017. In 2018, the region dramatically increased its percentage of satisfactory or highly satisfactory evaluations, ultimately exceeding the global average for the first time (by 8 percent)
to hit an all-time high of 95 percent. This is indeed a very high level of quality and will require dedicated attention and investment to sustain these gains moving forward.

Exhibit 1. Percentage of Reports Rated Satisfactory or Highly Satisfactory from 2011 - 2018

WCAR 2018 Evaluation Portfolio

In 2018, WCAR produced a total of 20 evaluation reports, which is the largest evaluation portfolio the region has experienced since the beginning of GEROS, placing it as the region (along with ESAR) to produce the most evaluations in 2018. As seen in Exhibit 2, fifteen (15) countries produced evaluations in 2018 with Niger producing the largest number of reports (3 reports) followed by WCARO, Mali, and Benin (2 reports each).
In 2018, the largest percentage of WCAR’s evaluation portfolio was aligned with the health sector (17 percent) and education sector (17 percent), followed by gender equality (15 percent), child protection (13 percent), nutrition (12 percent), WASH (9 percent), humanitarian action (7 percent), HIV/AIDS (6 percent), and social inclusion (4 percent). The prevalence of these sectors in evaluations in 2018 is generally aligned with the overall Strategic Plan Objective Areas (SPOAs) across regions. Furthermore, WCARO made some efforts to increase the number of cross-sectoral evaluations, such as the one of the convergent programming in Mali and the other three covering, among others, CSD interventions.

In 2018, the WCAR evaluation portfolio was most aligned with the 2018 – 2021 UNICEF Strategic Plan Goals “cross-cutting priorities”, every child survives and thrives”, and “every child learns” at 22 percent each. The remainder of the 2018 portfolio was aligned with “every child is protected from violence and exploitation” (17 percent), “every child lives in a safe and clean environment” (11 percent), and “every child has an equitable chance in life” (6 percent).

While 60 percent of the 2018 WCAR evaluation portfolio consisted of programmes (12 evaluations), the region also evaluated six (6) projects (30 percent of its evaluation portfolio), one (1) joint programme and one (1) country programme. The region had a slightly stronger focus than the global average on summative evaluations with 70 percent of evaluations being either summative or summative and formative in nature. Additionally, 18/20 evaluations (90 percent) used mixed methods, with two (2) evaluations drawing on purely qualitative methods. In 2018, all evaluations were at the outcome (70 percent) or impact (30 percent) levels.

In 2018, evaluations in WCAR used primarily four types of evaluation designs: quasi-experimental (35 percent), theory-based (25 percent), case studies (15 percent), and participatory (10 percent). Three evaluation reports (15 percent of the portfolio) did not specify evaluation design used. The proportion of the use of case study designs in the region was three times higher than the global average.
While over half (65 percent) of evaluations commissioned by the region were managed exclusively by UNICEF, the region also jointly managed three (3) evaluations with a national government counter-part (15 percent); one (1) was managed jointly with at least another UN agency, and one (1) with an organisation external to the UN. Two (2) evaluation reports did not specify their management arrangement.

In 2018, all of the region’s evaluations but one were focused at the national level. The remaining evaluation was led by the WCARO and covered Cameroon, Chad, Niger and Nigeria (WCARO 2018-001).

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

In 2018, WCAR had the 4th highest overall score (along with HQ) for evaluation report quality among regions at 3.01/4.0. The region scored 8.8 percent below the highest scoring region (EAPR) and 7 percent above the region with the lowest score (ROSA). As shown in Exhibit 3, the average quality of evaluations in WCAR increased substantially by 10.3 percent between 2017 and 2018.

**Exhibit 3: Overall Score of WCAR Compared with Other Regions from 2016 - 2018**

As shown in Exhibit 4, fifteen countries produced evaluation reports in 2018 in WCAR with all obtaining an average score of satisfactory or above with the exception of Guinea that scored 1.95. Liberia and Nigeria ranked in the highly satisfactory ranking (3.5 or above) with ratings of 3.65 and 3.70 respectively.

---

1 (Guinea submitted a WASH Sustainability Check that did not qualify as a full-fledged evaluation because of the more limited number of evaluation criteria guiding the overall line of inquiry).
By looking at a breakdown of evaluations per assessment rating, we can see from Exhibit 5 that in 2018, 95 percent of evaluation reports in WCAR were rated as satisfactory or higher. This is more than a 30 percent improvement since 2017. Additionally, no reports were rated as highly satisfactory in 2017 but 10 percent of reports received this rating in 2018. Also, of interest is that the percentage of reports in the upper satisfactory category increased from 24 percent to 45 percent from 2017 to 2018. If the region continues to improve the quality of its evaluation reports, a percentage of reports in the upper satisfactory category may transition to highly satisfactory.
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 WCAR per assessment criterion in 2018.

As can be seen in Exhibit 6, the region was strongest in 2018 at presenting a strong executive summary (3.50) and insightful conclusions and lessons learned (3.25).

In 2018, around half (7/15) of WCAR countries received perfect scores (4/4) for the executive summary. Only one country (Senegal at 2.0) fell below the satisfactory marker for this criterion. For conclusions and lessons learned, seven countries scored 3.0 while four countries (Nigeria, Liberia, Ghana, and Equatorial Guinea) received perfect scores. Two countries (Togo and Guinea) fell below the satisfactory marker at scores of 2.0 each.

In 2018, the region was weakest at evaluation methods (2.75) and at integrating evaluation principles (2.75), along with evaluation findings (2.90). These were also assessment criteria that scored low globally in 2018.

Across the criteria on methods, principles, and findings, around half of WCAR countries that commissioned evaluation reports in 2018 had an average rating of 3.0 for each criterion. While 9 countries had an average rating of 3.0 for methods, the top scoring country for this criterion was Nigeria with a perfect 4/4 score while the lowest scoring country was Guinea with a score of 1.0. The integration of evaluation principles shows a similar trend with 10 countries at an average score of 3.0 while Nigeria scored 4/4 and Ghana scored 1.4. A total of 9 countries scored 3.0 for evaluation findings while Liberia and Congo received perfect scores.
The placement of WCAR compared to other regions in 2018 varies depending on the assessment criterion, with the region ranking 2nd highest regarding conclusions and lessons learned as well as the quality of its executive summaries. On the other hand, the region scored comparatively low for the methods 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, WCAR had the third highest score in 2018 in terms of its integration of gender equality and the empowerment of women (GEEW) principles into its evaluations at a score of 69 percent, which meets SWAP requirements.

---

A closer look at the three different SWAP criteria in Exhibit 8 shows that evaluations in the WCAR struggled the most in 2018 with SWAP criterion 2 (with a score of 1.85), by only partially integrating SWAP principles in a gender responsive methodology, methods and tools, and data analysis techniques. Conversely, 2018 evaluations in WCAR performed better regarding SWAP criterion 1 and satisfactorily integrated GEEW in the evaluation scope of analysis with an average score of 2.15. WCAR scored particularly high for criterion 3 (at a score of 2.20), where it satisfactorily integrated SWAP principles into a gender analysis. In fact, WCAR had the second highest score across regions (after MENA) for criterion 3.

When broken down by office, as we can see in Exhibit 9, the quality of GEEW integration into evaluation varies greatly by country. In 2018, three countries (Nigeria, Mauritania, and Democratic Republic of Congo) received a perfect SWAP score, therefore meeting SWAP requirements while other countries scored poorly such as Ghana (11 percent) and Guinea (33 percent) who were missing SWAP requirements and Niger (41 percent) who was approaching...
**SWAP requirements.** This suggests an opportunity for the exchange of GEEW knowledge and skills within the region.

**Exhibit 9: 2018 SWAP Scores by WCAR Office**

2018 Overall Regional and Country Strengths

- **Complete Executive Summaries:** WCAR evaluations in 2018 in general presented very thorough yet succinct executive summaries that effectively informed decision-makers about both the interventions being assessed as well as the evaluations. (Ex: Togo 2018-001).

- **Insightful Conclusions and Lessons Learned:** WCAR evaluations were generally strong at presenting insightful and analytical conclusions that provided an added value to the information listed in the findings section. Also, conclusions tended to have a forward-looking perspective, discussing what may lay ahead for the interventions being assessed. Furthermore, lessons learned in reports from WCAR were, in general, correctly identified and presented in a way that made them relevant for wider audiences and applicable to different contexts. (Ex: Liberia 2018-001).

2018 Overall Regional and Country Weaknesses
• **Methods:** In 2018, some evaluation reports from WCAR only described the evaluation design, methods, and data sources used, but did not included a discussion around why these were appropriate in fulfilling the evaluation purpose. Some evaluations simply referenced the ToRs to list the methods to be used without providing an expert justification and/or validation of what was requested or a discussion around whether what was initially requested in the ToRs was sufficient or feasible within the evaluation context. (To learn more about how to improve in this area, see good practices 3.B and 3.C, in Annex V).

• **Evaluation Principles:** While many evaluations adopted human rights language throughout reports, only a few actually anchored the analysis within a human rights framework such as a CRC and/or CEDAW framework. While WCAR performance in terms of the integration of gender equality (SWAP criteria) was among the best across regions, reports in this region could benefit from presenting more sex-disaggregated data throughout the report. In the absence of available sex-disaggregated data, few evaluation methodologies were designed to collect qualitative sex-disaggregated data. Additionally, even if a gender analysis was often reflected in conclusions, not all evaluations provided recommendations for how to improve on gender. (To learn more about how to improve in this area, see good practices 8.A, 8.C and 8.E, in Annex V).

• **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, it was common for evaluations not to 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 this analysis varied a good deal 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 good practice example 4.G in Annex V).

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