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 MENA during 2018, and provides a limited annual comparison and trend analysis. The brief presents: 1) the MENA 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 MENA. 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 MENA performance has experienced a sustained improvement in the overall quality of its evaluations over the last four years. Initially, MENA produced satisfactory evaluation reports at a slightly lower rate than the global average, but then the percentage of MENA reports rated satisfactory or higher plunged by 41 percent below the global average to 33 percent 2014. It is, however, worth noting that in 2014 the region produced only one report. Since then, the region has increased the number of evaluations commissioned and the average report quality has increased steadily until today. As indicated in
Exhibit 1 below, a total of 88 percent of the MENA evaluation reports were rated as satisfactory or highly satisfactory in 2018, which was 1 percent above the global average for that year.

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

![Chart showing percentage of reports rated satisfactory or highly satisfactory from 2011 to 2018.](chart)

**MENA 2018 Evaluation Portfolio**

In 2018, the MENA produced the highest number of evaluations for the region since 2011, with a total of 8 evaluation reports. This number still places it as the region to produce the fewest evaluations in 2018 following ROSA, which produced 9 reports. As seen in Exhibit 2, Palestine and Lebanon were the only countries to produce more than one report, with 3 and 2 reports respectively.

**Exhibit 2. Number of Evaluation Reports per Country in MENA in 2018**
In 2018, the largest percentage of MENA’s evaluation portfolio was aligned with humanitarian action and child protection (19 percent each). Its percentage of humanitarian evaluations was slightly over twice as high as the global average (8 percent). The remaining portfolio was aligned with social inclusion (16 percent); gender equality (13 percent); health (12 percent); education (12 percent); nutrition (6 percent); and HIV/AIDS (3 percent).

In 2018, the MENA portfolio was most aligned with the 2018 – 2021 Strategic Plan Goals “every child is protected from violence and exploitation” and “cross-cutting priorities” (at 24 percent each), followed by “every child has an equitable chance in life” (20 percent); “every child survives and thrives” (16 percent); and “every child learns” (16 percent). In 2018, the region did not commission any evaluations that were aligned with the goal “every child lives in a safe and clean environment”.

While three quarters of the region’s 2018 portfolio consisted of programme evaluations, MENA also commissioned one pilot/innovation evaluation and one system evaluation. Evaluations in the region were mainly focused on learning with 7/8 reports (87 percent) being either formative or formative and summative, which is higher than the 2018 global average (72 percent).

In 2018, 5/8 of MENA evaluations drew on quasi-experimental designs (63 percent), while one evaluation used a longitudinal design (Lebanon 2018-001) and another one was theory-based (State of Palestine (2018-002)) \(^1\). Also, the region privileged the use of mixed methods in 7/8 of its evaluations, with only one evaluation drawing on purely qualitative methods. Over half (62 percent) of evaluations assessed results at the impact level while 38 percent examined results at the outcome level.

All evaluations in 2018 were commissioned at the country level and were national in scope. Similarly, most evaluations (7/8) were managed exclusively by UNICEF with the management of one evaluation unclear from the report.

\(^1\) There was one evaluation whose design was not clear from the report.
Trends in the Overall Quality of Evaluation Reports for MENA

In 2018, MENA had the 3rd highest overall score for evaluation report quality among regions at 3.09/4.0. The region scored 6.8 percent below the highest-scoring region (EAPR at 3.36) and 2.5 percent above the global overall score (2.99). As shown in Exhibit 3, MENA had in 2018 its highest overall score over the past three years, with the quality of its reports improving considerably compared to 2017 (5.8 percent increase).

Exhibit 3: Overall score of MENA compared with other regions from 2016 - 2018

As shown in Exhibit 4, five countries produced evaluation reports in MENA in 2018 with all but one obtaining an average score of satisfactory or above (Tunisia scored 2.00). No countries in MENA obtained an average score in the highly satisfactory ranking (3.50 or above).

Exhibit 4: Overall score per country in MENA for 2018
By looking at a breakdown of evaluations per assessment rating, we can see from Exhibit 5 that 88 percent of MENA evaluations in 2018 were rated as satisfactory or higher, with an increase of 7 percent of reports rated in the upper satisfactory and highly satisfactory categories compared to 2017.

**Exhibit 5: Percentage of Reports per Assessment Rating for MENA from 2016 – 2018**
2018 MENA 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 MENA per assessment criterion in 2018.

As can be seen in Exhibit 6, MENA was strongest at clearly explaining the purpose, objectives and scope of the evaluation (3.63) followed by a strong background section (3.25), evaluation methods (3.25), and executive summary (3.25).

No countries in MENA scored below 75 percent for the assessment criterion on purpose, objectives and scope, with two countries (Lebanon – with two reports – and Jordan) achieving perfect scores (4/4) for this criterion.

In 2018, the region was weakest at integrating evaluation principles (2.75), followed by conclusions and lessons learned, and recommendations (both scoring 3.00). It is worth noting that the section on the integration of evaluation principles was the weakest across regions and MENA evaluations were still, on average, 0.2 percent above the global average for this assessment criterion. On the other hand, while the global average for the integration of evaluation principles has been improving overall since 2016, MENA evaluation reports experienced a slight decrease in quality of 2.8 percent between 2017 and 2018.

Most countries scored near the regional average for evaluation principles (2.75). However, two countries in the region appeared at the two extremes for this assessment criterion, with Jordan achieving a perfect score (4/4) while Tunisia obtained a score of 2.00. Also, regarding conclusions and lessons learned, one report from Lebanon (Lebanon 2018-001) and Tunisia struggled at presenting analytical conclusions and correctly identified lessons learned, scoring 1.00 and 2.00, respectively.

Exhibit 6: 2018 Regional Evaluation Report Quality per Assessment Criterion
In terms of placement compared to other regions in 2018, after EAPR (3.67) MENA had the 2nd highest average score for the assessment criterion on purpose, objectives and scope, with (3.63), which is 10 percent above the global average for this assessment criterion (3.23).

**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, MENA was the top scorer in 2018 regarding its integration of gender equality and the empowerment of women (GEEW) principles into evaluations, with a score of 78 percent and *meeting SWAP requirements*. MENA scored 13 percent above the global average and 7 percent above LACR, the 2nd highest scoring region in this area.

**Exhibit 7: 2018 SWAP Scores by Region**

A closer look at the three different SWAP criteria in Exhibit 8 shows that evaluations in the MENA were relatively even as to their performance in the three SWAP criteria, outscoring all other regions and *satisfactorily integrated SWAP principles* across criteria. Evaluations in MENA were strongest regarding SWAP criterion 1 (integrating GEEW in the evaluation scope of analysis) and criterion 3 (gender analysis) with scores of 2.38 each. MENA reports scored slightly lower at SWAP criterion 2, (gender responsive methodology) at 2.25.

**Exhibit 8: 2018 SWAP Criterion by Region**
When broken down by country, as seen in Exhibit 9, the quality of GEEW integration into evaluation design and assessment varies somewhat across countries. In 2018, reports from Lebanon and Jordan reached perfect SWAP scores (100 percent) followed by Djibouti at 78 percent and Palestine at 74 percent, all of which were meeting SWAP requirements. In contrast, Tunisia was missing SWAP requirements with an average score of 22 percent.

*Exhibit 9: 2018 SWAP Scores by MENA Country*
2018 Overall Regional and Country Strengths

- **Clear Purpose, Objectives and Scope:** Reports from MENA generally listed the intended use and users of the evaluation. Also, evaluations explained in much detail the purpose of the evaluations as well as the objectives sought and the scope of the analysis in terms of geographical extension, time period and aspects of the initiatives to be evaluated. A total of 6/8 evaluations (75 percent) from MENA obtained a perfect score for this assessment criterion. (Ex: Lebanon 2018-001).

- **Detailed Background Section:** MENA evaluations in 2018 were generally strong at providing good detail around the object of evaluation and its context (including information on beneficiaries and the financial value of the object of evaluation). Additionally, 87.8 percent of MENA reports obtained a perfect score for their good quality theories of change (TOCs). Finally, MENA evaluations were consistently strong at describing key stakeholders and their contributions, with all of the reports in this region scoring satisfactory or above (Ex: Palestine 2018-001).

- **Well-Justified Evaluation Methods:** MENA reports largely used appropriate evaluation methods and also tended to provide good justifications for why the methods were selected and were appropriate for serving the evaluation purpose and objectives. Additionally, mitigation strategies often accompanied identified evaluation limitations, and the ethical principles used during the conduct of the evaluation were generally well outlined (Lebanon 2018-001).

- **Strong Executive Summary:** MENA executive summaries were generally complete, included all of the necessary elements to readily inform about both the object of the evaluation and the evaluation itself, and were at an appropriate length to inform decision-
makers. All reports from MENA scored satisfactory or above in this section. (Ex: State of Palestine 2018-002). (See Annex IV for a list of strengths with examples and Annex V for a list of good practices from 2018 evaluations, including those from HQ).

2018 Overall Regional and Country Weaknesses

- **Evaluation Principles:** While MENA countries scored the highest in terms of gender equality (in all three SWAP criteria), the region could have been stronger at explicitly basing evaluation assessments on a human rights framework and at more consistently assessing the extent to which the implementation of the intervention addressed equity and child rights. Additionally, many evaluations did not sufficiently discuss the ways in which stakeholders were involved in the design and management of the evaluation or the way they played a role beyond that of informants. (To learn more about how to improve in this area, see good practice example 8.A from Uzbekistan in Annex V).

- **Conclusions and Lessons Learned:** While evaluations in MENA correctly summarized the information provided in the findings, many reports lacked a deeper and insightful analysis of the main elements presented and did not elaborate on their potential future implications. Also, some evaluations struggled at correctly identifying lessons learned that are actual contributions to general institutional knowledge and that are applicable in different contexts. Finally, some evaluations tended to provide lessons learned that were scattered throughout the findings instead of presenting them within their own section. (To learn more about how to improve in this area, see good practice example 5.A from Uzbekistan in Annex V).

- **Recommendations:** Many evaluations from MENA do not provide recommendations that are detailed enough to be easily made actionable. Also, most evaluations from MENA did not explain the process followed in developing the recommendations and did not discuss the level of stakeholder involvement in their formulation and validation. This may have compromised the degree of ownership by stakeholders and ultimately risks reducing the chances that the recommendations be implemented. (To learn more about how to improve in this area, see good practice example 6.A from Mauritania in Annex V).