UNICEF GEROS Meta-Analysis 2018 Regional Profiles
Europe and Central Asia Region (ECAR)
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 ECAR during 2018, and provides a limited annual comparison and trend analysis. The brief presents: 1) the ECAR 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 ECAR. 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 ECAR has consistently remained above the global average in terms of the production of satisfactory evaluation reports, hitting a peak in 2016 at 100 percent of its evaluation reports being rated as satisfactory or higher. The trend throughout the years has been upwards with two exceptions, i.e. one major drop to 64 percent in 2015 and a second drop to 81 percent in 2017. However, the region went on to improve its performance again to 92 percent in 2018, where the region remains 5 percent above the global average.

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

ECAR 2018 Evaluation Portfolio
In 2018, ECAR produced a total of 12 evaluation reports, placing it as the third region (along with EAPR) to produce the most evaluations that year. As seen in Exhibit 2, Romania and Kazakhstan produced the largest number of reports per country at 3 reports each.
In 2018, nearly one fourth (23 percent) of the ECAR evaluation portfolio covered child protection issues while nearly another fourth (21 percent) looked at social inclusion, placing the region above the global average for both SPOA priorities. The remaining portfolio consisted of gender equality and education (both at 14 percent) followed by health (12 percent); humanitarian action (7 percent); HIVAIDS (5 percent); and WASH and nutrition (both at 2 percent).

The two 2018 – 2021 Strategic Plan goals that represented the largest portion of the 2018 ECAR portfolio were: every child is protected from violence and exploitation (at 26 percent) and every child has an equitable chance in life (at 24 percent). The other goals areas were: cross-cutting priorities (18 percent); every child learns (16 percent); every child survives and thrives (13 percent); and every child lives in a safe and clean environment (3 percent).

While nearly half of all evaluations commissioned within the global evaluation portfolio were programme evaluations in 2018, the ECAR portfolio reflects greater diversity, with projects accounting for 33 percent; system evaluations comprising 17 percent; pilot/innovation evaluations comprising another 17 percent, and programmes at 17 percent. Thematic and strategy evaluations comprised 8 percent each.

In 2018, 67 percent of ECAR evaluations were either formative or formative and summative with 33 percent exclusively summative. The region’s choice of evaluation methods is very similar to the global average, with 92 percent of its evaluations drawing on mixed methods (with 8 percent drawing on purely qualitative methods).

While the global portfolio drew on a variety of evaluation designs in 2018, ECAR focused on four designs, which were quasi-experimental (42 percent), participatory (33 percent), theory-based
(17 percent), and experimental (8 percent). ECAR also focused more of its evaluations on the impact level than the global average, with 67 percent of evaluations assessing impact in 2018 (compared to 46 percent), with the remaining 33 percent assessing outcome-level results in 2018.

In 2018, 83 percent of the region’s evaluations were focused at the national level while 8 percent looked at the regional level (the remaining 9 percent looked at other levels). All evaluations were managed exclusively by UNICEF in 2018.

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

In 2018, ECAR had the 2nd highest overall score (after EAPR) for evaluation report quality among regions at 3.13/4.0. The region scored 1 percent above the next highest-scoring region (MENA at 3.09) and 10 percent above the lowest-scoring region (ROSA at 2.73). As shown in Exhibit 3, the average quality of evaluations in ECAR increased by 4.5 percent between 2016 and 2017 but then decreased by a slight 1.5 percent in 2018.

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

As shown in Exhibit 4, eight countries produced evaluation reports in 2018 in ECAR with all obtaining an average score of satisfactory or above with the exception of Armenia that scored 2.45. One country (Republic of Uzbekistan) obtained an average score of 3.65 in the highly satisfactory ranking (3.5 or above).

**Exhibit 4: Overall score per country in ECAR for 2018**
By looking at a breakdown of evaluations per assessment rating, we can see from Exhibit 5 that in 2016, all of the evaluation reports from ECAR were rated as satisfactory or higher, but the region struggled to maintain all of its evaluation reports above the satisfactory threshold in 2017 where the rate of reports rated as satisfactory or highly satisfactory dropped to 80 percent. The region was able to bring the percentage of evaluations rated as satisfactory or better up in 2018 to a rate of 92 percent, which was an increase of 12 percent.

Exhibit 5: Percentage of Reports per Assessment Rating for ECAR from 2016 – 2018
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 ECAR 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 (3.42) and the evaluation background (3.33).

In 2018, half 4/8 of ECAR offices (Montenegro, Uzbekistan, ECARO regional office, and Bulgaria) received perfect scores (4/4) for evaluation purpose, objectives and scope. A total of 2/8 countries (Montenegro and Bulgaria) in EAPR received perfect scores (4/4) for their evaluation report background sections.

In 2018, the region was weakest at integrating evaluation principles (2.83), followed by presenting a logical and accessible report structure (3.0). The presentation of robust findings (3.08) and the use of grounded and actionable recommendations (3.17) could also benefit from some improvement.

With the exception of Uzbekistan (who received a perfect score) and Armenia (who received a score of 2.0), most countries in ECAR had an average score for evaluation principles close to the regional average of 2.83. While Montenegro received a perfect score for evaluation report structure and presentation, three offices (ECARO, Bulgaria and Belarus) scored low at 2.0. Half of the offices scored near the regional average of 3.08 for evaluation findings with ECARO and Armenia scoring both at 2.0. A similar trend can be seen for recommendations where most offices scored near the regional average of 3.17 but with Uzbekistan and Bulgaria receiving a perfect score and Armenia receiving a score of 2.0.

### 2018 ECAR Quality of Evaluation Reports by Assessment Criteria
The placement of ECAR compared to other regions in 2018 varies depending on the assessment criterion, with the region ranking within the top three for the background, purpose, methods, conclusions and lessons learned, and evaluation principles. Also, the region ranked fourth for the recommendations after EAPR, WCAR and HQ. On the other hand, ECAR score for the executive summary criterion was comparatively low (3.08 percent) which was only superior to those of ESAR and ROSA.

**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, although ECAR had the second highest overall rating per region, it scored second last in terms of its integration of gender equality and the empowerment of women (GEEW) principles into its evaluations. In 2018, it had an average score of 59 percent, which is approaching SWAP requirements. This highlights the need for targeted improvement around GEEW moving forward.
A closer look at the three different SWAP criteria in Exhibit 8 shows that evaluations in the ECAR struggled the most in 2018 with SWAP criterion 2, which partially integrated SWAP principles in using a gender responsive methodology, methods and tools, and data analysis techniques. 2018 evaluations in ECAR also struggled with criterion 3 to produce a meaningful gender analysis throughout the findings, conclusions and recommendations and only partially integrated SWAP for criterion 3. However, ECAR satisfactorily integrated GEEW in the evaluation scope of analysis (SWAP criterion 1), which was the strongest gender criterion for the region. That being said, ECAR still ranked third lowest under this criterion behind ROSA and ESAR.
When broken down by office, as we can see in Exhibit 9, the quality of GEEW integration into evaluation design and assessment varies greatly. In 2018, both Montenegro and Uzbekistan received a perfect SWAP score while ECARO and Armenia scored poorly at only 33 percent (missing SWAP requirements). This suggests an opportunity for the exchange of GEEW knowledge and skills within the region.
2018 Overall Regional and Country Strengths

- **Clear Evaluation Purpose, Objectives and Scope**: ECAR evaluations in 2018 presented clear evaluation 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 Background Section**: ECAR evaluations 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, reports generally include good quality theories of change (TOCs), which is important especially considering that 17 percent of ECAR evaluations in 2018 were theory-based.

2018 Overall Regional and Country Weaknesses

- **Evaluation Principles**: While many evaluations adopted human rights language throughout reports, 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, but did not always provide a meaningful gender analysis throughout the report. Many such analyses did not sufficiently explore key aspects of gender transformation. The gender analysis was 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 good practices 8.A and 8.C-G, in Annex V).

- **Report Structure and Presentation**: In 2018, some evaluation reports from ECAR were excessively long to be accessible to evaluation users. Additionally, some reports were missing important information from the annexes.

- **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 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 good practice 4.G in Annex V).

- **Recommendations**: Many recommendations from ECAR evaluations in 2018 were too general to be highly useful to evaluation users. They often lacked detail around how they can be implemented and which specific stakeholder(s) should be responsible for their
implementation. Additionally, many recommendations continued to lack information around the extent to which stakeholders were involved in their development and/or validation. (To learn more about how to improve in this area, see good practice 6.A 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 ECAR.