# GEROS Evaluation Quality Assurance Tool

Version 2016.4

|  |  | *Notes for QA* |
|---|---|---|
| **Title of the Evaluation Report** | **Evaluation of the training programme for continuous professional development of social protection staff** |  |
| **Report sequence number** | **2017/001** |  |
| **Region** | **CEE/CIS** |  |
| **Year of Report** | **2017** |  |
| **Office** | **Macedonia** |  |
| **Coverage (countries)** | **Macedonia** |  |
| **ToRs present** | **Yes** |  |
| **Date of Review** | **21/04/2017** |  |
| **Name of reviewer** | **ImpactReady** |  |
| **Classification of Evaluation Report** |  |  |
| **Management of Evaluation** *(Managerial control and oversight of evaluation decisions)* |  | **UNICEF managed** |
| **SPOA Correspondence** *(Alignment with SPOA focus area priorities)* | **Health** | **No** |
|  | **HIV/AIDS** | **No** |
|  | **WASH** | **No** |
|  | **Nutrition** | **No** |
|  | **Education** | **No** |
|  | **Child protection** | **Yes** |
|  | **Social inclusion** | **No** |
|  | **Gender equality (cross-cutting)** | **Yes** |
|  | **Humanitarian action (cross-cutting)** | **No** |
| **Evaluation object** |  | **Programme** |
| **Evaluation type** |  | **Formative** |
| **Evaluation strategy** |  | **Mixed methods** |
| **Evaluation design** |  | **Case study** |
| **Evaluation level** |  | **Impact** |
| **Geographic Scope** |  | **National** |
|  |  |  |
| **SECTION A: BACKGROUND (weight 5%)** |  | Rating |

| Question 1. | Is the object of the evaluation clearly described? | 78% |
|---|---|---|
| | Clear and relevant description of the intervention, including: location(s), timelines, cost/budget, and implementation status | **Mostly** |
| | Clear and relevant description of intended beneficiaries by type (i.e., institutions/organizations; communities; individuals...), by geographic location(s) (i.e., urban, rural, particular neighbourhoods, town/cites, sub-regions...) and in terms of numbers reached (as appropriate to the purpose of the evaluation) | **Yes** |
| | Description of the relative importance of the object to UNICEF (e.g. in terms of size, influence, or positioning | **Mostly** |
| Question 2. | Is the context of the intervention clearly described? | 78% |
| | Clear and relevant description of the context of the intervention (policy, socio-economic, political, institutional, international factors relevant to the implementation of the intervention) | **Mostly** |
| | Clear and relevant description (where appropriate) of the status and needs of the target groups for the intervention | **Mostly** |
| | Explanation of how the context relates to the implementation of the intervention | **Yes** |
| Question 3. | Is the results chain or logic well articulated? | 100% |
| | Clear and complete description of the intervention's intended results | **Yes** |
| | Intervention logic presented as a coherent theory of change, logic chain or logic framework | **Yes** |
| Question 4. | Are key stakeholders and their contributions clearly identified? | 83% |
| | Identification of implementing agency(ies), development partners, primary duty bearers, secondary duty bearers, and rights holders | **Yes** |
| | Identification of the specific contributions and roles of key stakeholders (financial or otherwise), including UNICEF | **Mostly** |

| • • • - | **Executive Feedback on Section A** | Overall rating for Section |
|---|---|---|
| The rater will briefly (3-5 sentences) assess top line issues for this section relevant for feedback to senior management (positive and negative), summarizing here how the evaluation report meets or fails to meet the criteria above (judgement). As relevant, the | | **Satisfactory** |
| **Overall, the report includes a strong background section and description of the object of evaluation. The child protection reform process in Macedonia is well outlined and the positioning of the programme within this process is well explained. However, it would be useful to provide more information on the current contextual situation regarding child protection in Macedonia and to identify the specific needs of the children and the child protection staff that the reform process and the training programme are attempting to address. The report explains that the programme used no Theory of Change model but an elaborate model was retroactively produced by the evaluation team. While stakeholders are identified, the cost of the programme and UNICEF's specific contributions to the programme are not clearly specified.** | | **3** |

| **Additional comments for Section A (recommendations for improvement)** | Weighting |
|---|---|
| **Contextual information on the actual situation of child protection in Macedonia (including the needs of both children and child protection staff) would help to better ground the evaluation within the current thematic context. Additionally, the size and importance of the programme (both overall and to UNICEF) would be clearer to the reader if the report were to include information on the overall programme costs and UNICEF's contributions.** | 0.05 |

| **SECTION B: EVALUATION PURPOSE, OBJECTIVES AND SCOPE (weight 5%)** | | Rating |
|---|---|---|
| Question 5. | Is the purpose of the evaluation clearly described? | 100% |
| | Specific identification of how the evaluation is intended to be used and to what this use is expected to achieve | **Yes** |
| | Identification of appropriate primary intended users of the evaluation | **Yes** |
| Question 6. | Are the objectives and scope of the evaluation clear and realistic? | 100% |
| | Clear and complete description of what the evaluation seeks to achieve by the end of the process with reference to any changes made to the objectives included in the ToR | **Yes** |
| | Clear and relevant description of the scope of the evaluation: what will and will not be covered (thematically, chronologically, geographically with key terms defined), as well as the reasons for this scope (e.g., specifications by the ToRs, lack of access to particular geographic areas for political or safety reasons at the time of the evaluation, lack of data/evidence on particular elements of the intervention) | **Yes** |
| • • • • | **Executive Feedback on Section B** | Overall rating for Section |

| | Highly Satisfactory |
|---|---|
| The rater will briefly (3-5 sentences) assess top line issues for this section relevant for feedback to senior management (positive and negative), summarizing here how the evaluation report meets or fails to meet the criteria above (judgement). As relevant, the rater will highlight best practice/added value elements. | |
| The report explains why the evaluation is necessary at this point in time, how it will be used, and its intended audience. The evaluation report also provides well outlined evaluation objectives that link to its purpose. Changes made to the ToRs are well outlined and justified (i.e. changes that had to be made to the evaluation process due to a missing programme results framework and corresponding indicators). The evaluation scope in terms of geographic and time coverage are also specified. | 4 |

| Additional comments for Section B (recommendations for improvement) | Weighting |
|---|---|
| This section observes good practices. No further improvement is required. | 0.05 |

| SECTION C: EVALUATION METHODOLOGY (weight 15%) | Rating |
|---|---|
| **Question 7.** Does the evaluation provide a relevant list of evaluation criteria that are explicitly justified as appropriate for the purpose of the evaluation? UNICEF evaluation standards refer to the OECD/DAC criteria. Not all OECD/DAC criteria are relevant to all evaluation objectives and scopes. Standard OECD DAC Criteria include: Relevance; Effectiveness; Efficiency; Sustainability; Impact. Evaluations should also consider equity, gender and human rights (these can be mainstreamed into other criteria). Humanitarian evaluations should consider Coverage; Connectedness; Coordination; Protection; Security | 100% |
| Clear and relevant presentation of the evaluation framework including clear evaluation questions used to guide the evaluation | Yes |
| If the framework is OTHER than UNICEF standard criteria, or if not all standard criteria of the chosen framework are included, the reasons for this are clearly explained and the chosen framework is clearly described | Not Rated |
| **Question 8.** Does the report specify methods for data collection, analysis, and | 92% |
| Clear and complete description of a relevant design and set of methods that are suitable for the evaluation's purpose, objectives and scope | Yes |
| Clear and complete description of the data sources, rationale for their selection and sampling strategy. This should include a description of how diverse perspectives are captured (or if not, provide reasons for this), how accuracy is ensured, and the extent to which data limitations are mitigated | Mostly |
| Clear and complete description of the methods of analysis, including triangulation of multiple lines and levels of evidence (if relevant)? | Yes |
| Clear and complete description of limitations and constraints faced by the evaluation, including gaps in the evidence that was generated and mitigation of bias? | Yes |
| **Question 9.** Are ethical issues and considerations described? The evaluation should be guided by the UNEG ethical standards for evaluation. As such, the evaluation report should include: | 100% |
| Explicit reference to the obligations of evaluators (independence, impartiality, credibility, conflicts of interest, accountability) | Yes |
| Description of ethical safeguards for participants appropriate for the issues described (respect for dignity and diversity, right to self-determination, fair representation, compliance with codes for vulnerable groups, confidentiality, and avoidance of harm) | Yes |
| ONLY FOR THOSE CASES WHERE THE EVALUATION INVOLVES INTERVIEWING CHILDREN: explicit reference is made to the UNICEF procedures for Ethical Research Involving Children | Not Rated |

| • • • - Executive Feedback on Section C | Overall rating for Section |
|---|---|
| The rater will briefly (3-5 sentences) assess top line issues for this section relevant for feedback to senior management (positive and negative), summarizing here how the evaluation report meets or fails to meet the criteria above (judgement). As relevant, the | Satisfactory |

| | | |
|---|---|---|
| **The evaluation faced several significant limitations in terms of inadequate quantitative data to draw from (i.e. outdated desk review documents), inexistent M&E data, and no programme performance management framework.  Even so, the evaluators used a mixed methods approach, (which included key informant interviews, focus group discussions, a desk review, an online survey, direct observation, and a case study) to help overcome these limitations.  The rationale for the selection of these methods is well explained and analysis methods are well outlined.  Data sources (including documents and interviewees) are identified, although it seems odd that no UNICEF staff were included in any of the informant interviews.  The ethical procedures followed by the evaluation team are exceptionally well outlined.** | 📶 | **3** |
| **Additional comments for Section C (recommendations for improvement)** | | Weighting |
| **This section observes good practices.  However, it would be helpful for the reader to understand why UNICEF staff appear to not have been included in any informant interviews.** | | 0.15 |

| SECTION D: EVALUATION FINDINGS (weight 20%) | | | Rating |
|---|---|---|---|
| Question 10. | Do the findings clearly address all evaluation objectives and scope? | | 100% |
| | | Findings marshal sufficient levels of evidence to systematically address all of the evaluation's questions and criteria | **Yes** |
| | | Reference to the intervention's results framework in the formulation of the findings | **Yes** |
| Question 11. | Are evaluation findings derived from  the conscientious, explicit and judicious use of the best available, objective, reliable and valid data and by accurate quantitative and qualitative analysis of evidence. | | 83% |
| | | The evaluation clearly presents multiple lines (including multiple time series) and levels (output, outcome, and appropriate disaggregation) of credible evidence. | **Yes** |
| | | Findings are clearly supported by and respond to the evidence presented, including both positive and negative. Findings are based on clear performance indicators, standards, benchmarks, or other means of comparison. | **Yes** |
| | | Unexpected effects (positive and negative) are identified and analysed | **Partly** |
| | | The causal factors (contextual, organizational, managerial, etc.) leading to achievement or non-achievement of results are clearly identified. For theory-based evaluations, findings analyse the logical chain (progression -or not- from implementation to results). | **Yes** |
| Question 12. | Does the evaluation assess and use the intervention's Results Based Management elements? | | 100% |
| | | Clear and comprehensive assessment of the intervention's monitoring system (including completeness and appropriateness of results/performance framework -including vertical and horizontal logic; M&E tools and their usage) | Yes |
| | | Clear and complete assessment of the use of monitoring data in decision making | Yes |
| • • • - | **Executive Feedback on Section D** | | Overall rating for Section |
| The rater will briefly (3-5 sentences) assess top line issues for this section relevant for feedback to senior management (positive and negative), summarizing here how the evaluation report meets or fails to meet the criteria above (judgement). As relevant, the | | | **Satisfactory** |
| **Findings are well written, are structured around the standard evaluation criteria, and sufficiently marshal different levels and types of evidence. The evaluation is particularly strong at developing its own assessment benchmarks in light of a missing performance management framework.  The report explains that the programme was entirely missing an M&E function.  The findings include both positive and negative information, which is used to answer the evaluation questions.  Unexpected findings are not clearly identified.** | | 📶 | **3** |
| **Additional comments for Section D (recommendations for improvement)** | | | Weighting |

| | |
|---|---|
| **This section observes good practices. However, it is useful for the reader if the report clearly identifies any unexpected findings, ideally within a sub-section, labelled as such.** | 0.2 |

| SECTION E: EVALUATION CONCLUSIONS & LESSONS LEARNED (weight 15%) | | Rating |
|---|---|---|
| Question 13. | Do the conclusions present an objective overall assessment of the | 100% |
| | Clear and complete description of the strengths and weaknesses of the intervention that adds insight and analysis beyond the findings | **Yes** |
| | Description of the foreseeable implications of the findings for the future of the intervention (if formative evaluation or if the implementation is expected to continue or have additional phase) | **Yes** |
| | The conclusions are derived appropriately from findings | **Yes** |
| Question 14. | Are lessons learned correctly identified? | 100% |
| | Correctly identified lessons that stem logically from the findings, presents an analysis of how they can be applied to different contexts and/or different sectors, and takes into account evidential limitations such as generalizing from single point observations. | **Yes** |
| • • • • | **Executive Feedback on Section E** | Overall rating for Section |
| The rater will briefly (3-5 sentences) assess top line issues for this section relevant for feedback to senior management (positive and negative), summarizing here how the evaluation report meets or fails to meet the criteria above (judgement). As relevant, the rater will highlight best practice/added value elements | | **Highly Satisfactory** |
| **Both overall conclusions and conclusions that are specific to each evaluation criterion are clearly presented, are analytical, and discuss the programme's strengths and areas for improvement. Lessons learned are particularly insightful and are written in a generalizable manner that easily facilitates applying them to other similar contexts.** | | 4 |
| **Additional comments for Section E (recommendations for improvement)** | | Weighting |
| **This section observes good practices. No further improvements are required.** | | 0.15 |

| SECTION F: RECOMMENDATIONS (weight 15%) | | Rating |
|---|---|---|
| Question 15. | Are recommendations well grounded in the evaluation? | 78% |
| | Recommendations are logically derived from the findings and/or conclusions | **Yes** |
| | Recommendations are useful to primary intended users and uses (relevant to the intervention and provide realistic description of how they can be made operational in the context of the evaluation) | **Yes** |
| | Clear description of the process for developing recommendations, including a relevant explanation if the level of participation of stakeholders at this stage is not in proportion with the level of participation in the intervention and/or in the conduct of the evaluation | **Partly** |
| Question 16. | Are recommendations clearly presented? | 100% |
| | Clear identification of target group for action for each recommendation (or clearly clustered group of recommendations) | **Yes** |
| | Clear prioritization and/or classification of recommendations to support use | **Yes** |
| • • • - | **Executive Feedback on Section F** | Overall rating for Section |
| The rater will briefly (3-5 sentences) assess top line issues for this section relevant for feedback to senior management (positive and negative), summarizing here how the evaluation report meets or fails to meet the criteria above (judgement). As relevant, the rater will highlight best practice/added value elements | | **Satisfactory** |
| **Recommendations are logically derived from the findings and conclusions. They include both strategic and operational elements, which will likely be useful for the evaluation's intended users. The target group for action for each recommendation is clearly specified and the priority level of each recommendation is identified (i.e. high or medium priority). While the report explains in the methodology section that the evaluation was highly participatory and that the evaluators consulted with beneficiaries throughout the evaluation process, the extent to which beneficiaries were involved in developing the recommendations is unclear.** | | 3 |

| Additional comments for Section F (recommendations for improvement) | Weighting |
|---|---|
| **It is important for the report to clearly articulate the extent to which beneficiaries were involved in the development of the evaluation recommendations, as this provides additional validity to the recommendations. A brief sentence or paragraph explaining stakeholder involvement in the development of recommendations can be useful at the beginning of the recommendations section before the recommendations are presented.** | 0.15 |

| SECTION G: EVALUATION STRUCTURE/PRESENTATION (weight 5%) | | Rating |
|---|---|---|
| Question 17. | Does the evaluation report include all relevant information? | 67% |
| | Opening pages include: Name of evaluated object, timeframe of the evaluation, date of report, location of evaluated object, names and/or organization(s) of the evaluator(s), name of organization commissioning the evaluation, table of contents -including, as relevant, tables, graphs, figures, annexes-; list of acronyms/abbreviations, page numbers | **Yes** |
| | Annexes should include, when not present in the body of the report: Terms of Reference, Evaluation matrix, list of interviewees, list of site visits, data collection instruments (such as survey or interview questionnaires), list of documentary evidence Other appropriate annexes could include: additional details on methodology, copy of the results chain, information about the evaluator(s) | **Partly** |
| Question 18. | Is the report logically structured? | 100% |
| | The structure is easy to identify and navigate (for instance, with numbered sections, clear titles and sub-titles | **Yes** |
| | Context, purpose and methodology would normally precede findings, which would normally be followed by conclusions, lessons learned and recommendations | **Yes** |

| • • - - Executive Feedback on Section G | Overall rating for Section |
|---|---|
| The rater will briefly (3-5 sentences) assess top line issues for this section relevant for feedback to senior management (positive and negative), summarizing here how the evaluation report meets or fails to meet the criteria above (judgement). As relevant, the | **Fair** |
| **The report is logically structured in the traditional order with one section logically leading to the next. The opening pages contain all of the required information, including the names of the consultants, the commissioning agency, and the programme name and date. The annexes include important information, such as an evaluation matrix and the ToRs. However, the data collection tools are not presented (i.e. a copy of the survey, a copy of the questions used to guide the KIIs and FGDs). In terms of report clarify, all acronyms should be fully spelled out the first time they are used in the report to facilitate easy reading. Additionally, the report would benefit from some English-language editing, as there are grammatical errors throughout the report. Articles such as "a" or "the' are regularly missing.** | 2 |

| Additional comments for Section G (recommendations for improvement) | Weighting |
|---|---|
| **In terms of clarity, it would be useful if acronyms were fully spelled out the first time they are used in the report. Additionally, the report would benefit from some English-language editing, as there are grammatical errors throughout the report. Articles such as "a" or "the' are regularly missing. Additionally, it is important for a copy of the data collection tools to be included in the annexes, as this helps to validate the quality of the data collected.** | 0.05 |

| SECTION H: EVALUATION PRINCIPLES (weight 15%) | | Rating |
|---|---|---|
| Question 19. | Did the evaluation design and style consider incorporation of the UN and UNICEF's commitment to a human rights-based approach to programming, to gender equality, and to equity? | 89% |
| | Reference and use of rights-based framework, and/or CRC, and/or CCC, and/or CEDAW and/or other rights related benchmarks in the design of the evaluation | **Yes** |

| | | |
|---|---|---|
| | Clear description of the level of participation of key stakeholders in the conduct of the evaluation, and description of the rationale for the chosen level of participation (for example, a reference group is established, stakeholders are involved as informants or in data gathering) | **Mostly** |
| | Stylistic evidence of the inclusion of these considerations can include: using human-rights language; gender-sensitive and child-sensitive writing; disaggregating data by gender, age and disability groups; disaggregating data by socially excluded groups. | **Yes** |
| Question 20. | Does the evaluation assess the extent to which the implementation of the intervention addressed gender, equity & child rights? | 92% |
| | Identification and assessment of the presence or absence of equity considerations in the design and implementation of the intervention | **Yes** |
| | Identification and assessment of the presence or absence of gender in the design and implementation of the intervention | **Yes** |
| | Explicit analysis of the involvement in the object of right holders, duty bearers, and socially marginalized groups, and the differential benefits received by different groups of children | **Yes** |
| | Clear proportionality between the level of participation in the intervention and in the evaluation, or clear explanation of deviation from this principle (this may be related to specifications of the ToRs, inaccessibility of stakeholders at the time of the evaluation, budgetary constraints, etc.) | **Mostly** |
| Question 21. | Does the evaluation meet UN SWAP evaluation performance indicators? | Meets requirements |
| | Note: this question will be rated according to UN SWAP standards | 10 |
| | GEEW is integrated in the Evaluation Scope of analysis and Indicators are designed in a way that ensures GEEW-related data will be collected | **Fully integrated 3 points** |
| | integrated into the design, planning, implementation of the intervention and the results achieved. | **Fully integrated 3 points** |
| | A gender-responsive Evaluation Methodology, Methods and tools, and Data Analysis Techniques are selected. | **Satisfactorily integrated 2 points** |
| | The evaluation Findings, Conclusions and Recommendation reflect a gender analysis | **Satisfactorily integrated 2 points** |
| • • • - | **Executive Feedback on Section H** | Overall rating for Section |
| The rater will briefly (3-5 sentences) assess top line issues for this section relevant for feedback to senior management (positive and negative), summarizing here how the evaluation report meets or fails to meet the criteria above (judgement). As relevant, the rater will highlight best practice/added value elements | | **Satisfactory** |
| **The evaluation report is strong at using human rights language and child rights frameworks. Gender equality is also thoroughly mainstreamed throughout the report. The evaluation provides a particularly in-depth assessment of the extent to which the programme is improving child protection services to those children most in need, using a strong equity lens. Evaluation questions include questions around equity, human rights, and gender equality and information on these issues can be seen throughout the findings, conclusions, and recommendations. Even though the report explains that the evaluation used a highly participatory process, gaining feedback from stakeholders along the way, no reference is made in the report to an evaluation reference group and the extent to which stakeholders were involved in developing the recommendations is unclear. The use of an evaluation reference group is called for in the ToRs.** | | 3 |
| **Additional comments for Section H (recommendations for improvement)** | | Weighting |
| **The report could benefit from further clarification around how stakeholders were involved in the evaluation process. Particularly, it is important to mention if an evaluation reference group was used to guide the evaluation and it is necessary to outline the extent to which stakeholders were involved in developing the recommendations.** | | 0.15 |

| | | |
|---|---|---|
| **SECTION I: EXECUTIVE SUMMARY (weight 5%)** | | Rating |
| Question 22. | Can the executive summary inform decision-making? | 78% |
| | An executive summary is provided that is of relevant conciseness and depth for primary intended users | **Yes** |

| | | |
|---|---|---|
| | Includes all necessary elements (overview of the intervention, evaluation purpose, objectives and intended audience, evaluation methodology, key findings, key conclusions, key recommendations) | **Mostly** |
| | Includes all the necessary information to understand the intervention and the evaluation AND does not contain information not already included in the rest of the report | **Mostly** |

| • • - - | **Executive Feedback on Section I** | Overall rating for Section |
|---|---|---|
| | The rater will briefly (3-5 sentences) assess top line issues for this section relevant for feedback to senior management (positive and negative), summarizing here how the evaluation report meets or fails to meet the criteria above (judgement). As relevant, the | **Fair** |
| | **An Executive Summary is included and is an appropriate length to inform decision-makers. It includes a summary of all of the relevant sections of the report (i.e. programme overview, evaluation purpose, findings, conclusions, lessons learned, and recommendations) but does not provide enough detailed information on the object of the evaluation to allow the reader to understand what is being evaluated without having to reference the body of the report. Information on the programme interventions and the context in which the programme operates is important information to include in the first section of the Executive Summary under the section title "Programme Background and Key Information".** | **2** |

| **Additional comments for Section I (recommendations for improvement)** | Weighting |
|---|---|
| **The Programme Background and Key Information section at the beginning of the Executive Summary does not provide enough detailed information on the object of the evaluation to allow the reader to understand what is being evaluated without having to reference the body of the report. An explanation of the programme interventions and the context in which the programme operates is important information to include in this first section.** | 0.05 |

*Weighting checksum (should equal 1)* 1

| **OVERALL SCORE (max=4, min=0)** | | |
|---|---|---|
| • • • - | **Satisfactory** | **3.1** |
| | **Meets UNICEF/UNEG standards for evaluation reports and decision makers may use the evaluation with confidence** | |

| **Lessons for improving the management and performance of future evaluations** |
|---|
| Overall, this is a satisfactory evaluation report that uses a strong methodology and presents insightful and useful findings, conclusions, lessons learned, and recommendations. This being said, there are still some areas for improvement. The quality of the contextual information on the actual situation of child protection in Macedonia (including the needs of both children and child protection staff) could be improved and would help to better ground the evaluation within the current thematic context. Additionally, the size and importance of the programme (both overall and to UNICEF) would be clearer to the reader if the report were to include information on the overall programme costs and UNICEF's contributions. While the methodology is strong, it would be helpful for the reader to understand why UNICEF staff appear to not have been included in any informant interviews. Additionally, unexpected findings should be clearly identified (preferably within a sub-section). The report could benefit from further clarification around how stakeholders were involved in the evaluation process. Particularly, it is important to mention if an evaluation reference group was used to guide the evaluation and it is necessary to outline the extent to which stakeholders were involved in developing the recommendations. The report is well structured but it needs to be edited for English-language errors. The number of errors found throughout the report (especially missing articles such as "a" or "the" reduce the report's professional tone). Additionally, it is important for a copy of the data collection tools to be included in the annexes, as this helps to validate the quality of the data collected. Finally, the Executive Summary should include a more robust description of the object of evaluation and its context so that the reader does not need to reference the body of the report to gain an overall picture of what the evaluation is assessing. |