GEROS Evaluation Quality Assurance Tool

Version: September 2021

Generation of Innovation Leaders (GIL) Programme Evaluation

REPORT RATING SUMMARY			
Overall Rating		50%	Fair
● ● ● ● ●	Exceptional (96% - 100%)	5	
● ● ● ●	Highly Satisfactory (87.5% - 95.99%)	4	
● ● ● -	Satisfactory (62.5% - 87.49%)	3	
● ● - -	Fair (35% - 62.49%)	2	**Meets UNICEF/UNEG standards for evaluation reports in some regards, but not all. Decision makers may continue to use the evaluation with caution, but substantive improvements are possible.**
● - - -	Unsatisfactory (0% - 34.99%)	1	

REPORT DETAILS	
Title of the evaluation report	Generation of Innovation Leaders (GIL) Programme Evaluation
Report sequence number	Lebanon/59/2024/20433
Region	MENA
Year of report	2021
Office	Lebanon
Coverage (countries)	Lebanon
ToRs present	Yes
Date of review (dd/mmm/yyyy)	December 2, 2024
Name of review firm	IOD PARC

CLASSIFICATION OF EVALUATION REPORT	
Management of evaluation *(Managerial control and oversight of evaluation*	UNICEF managed
Unicef goal areas *(Alignment with strategic plan priorities)*	
Every child survives and thrives	No
Every child learns	Yes
Every child is protected from violence and exploitation	No
Every child lives in a safe and clean environment	No
Every child has an equitable chance in life	Yes
Gender equality (cross-cutting)	Yes
Humanitarian action (cross-cutting)	No
Evaluation object	Programme
Evaluation type	Summative and formative
Evaluation strategy	Mixed methods
Evaluation design *(primary method used)*	Non-experimental
Evaluation level	Impact
Geographic scope	National
Primary SDG(s) covered *(number)*	4, 8, 9, 10

EQA Summary: *The rater will provide top line issues for this evaluation relevant for feedback to senior management (positive and negative), summarizing how the evaluation report meets or fails to meet all criteria. As relevant, the rater will highlight best practice/added value elements and the level of complexity of the evaluation.*

This is a broadly adequate evaluation that provides a reasonable level of analysis and utility for stakeholders. The evaluation provides a decent level of analysis of the evaluation subject and scope. On the positive level,
> The methodology is broadly appropriate and supports an overall sense that the analysis is relevant
> The subject, findings, conclusions and lessons learnt are for the most part clear in their presentation.
Nevertheless, several areas for improvement can be identified that would have given much greater clarity to the focus and analysis of the evaluation. Critical elements include,
> Providing a theory of change or intervention logic - such as in a figure - to clarify the original expectations and ambition of the GIL programme, review the associated assumptions, give greater focus to the evaluation questions, and better locate the recommendations in the programme framework and requirements
> Developing the recommendations after the conclusions and lessons learnt rather than locating them against each of the current long-list of questions
Further details are provided below.

Recommendations for Improvement: *The rater will identify topline recommendations to improve the evaluation, and be specific to the sections of the report where shortcomings were found. As relevant, resources will be cited to assist evaluation managers in overseeing future evaluations.*

> The evaluation process and report would have greatly benefited from the development, revision and testing of a theory of change (or intervention logic). At a minimum it could have used and presented a summary of the GIL results framework. This is not evident from the evaluation matrix and would have allowed the evaluation to make clear what had and had not been achieved against the programme design and plan, and the ways in which the GIL programme was adjusted over time.
> Use of a theory of change would have supported the GIL programme by helping locate the activities and recommendations against a set of clear pathways that take the beneficiaries from training into employment/self-employment.
> Use of stakeholder mapping and analysis would have built better understanding of the programme approach and helped located the recommendations against duty bearers
> Given the evaluation found significant data gaps, more attention could have been given to exploring the causes of these gaps by examining the GIL programme's monitoring systems in order to provide clear learning as to what improvements could be made in the conclusions and recommendations.
> The evaluation includes 32 evaluation questions against the five criteria of relevance, effectiveness, efficiency, impact and sustainability. The evaluation questions as they stand are far too detail focused and there are too many of them. Each of the criteria could have been developed into a high level evaluation question (EQ) with the 32 existing EQs clustered around appropriate sub-questions and re-termed as ‘lines of enquiry’ or similar. This would have provided a much clearer focus and sequencing of the findings, conclusions and recommendations
> The findings would have been more clearly presented by the use of findings statements that sit against each of the evaluation lines of enquiry (as termed EQs in the report) before the findings narrative then when on to introduce and analyse the relevant evidence. Without this, many of the findings are discursive and it is hard for the reader to understand exactly what the finding is or how the evidence presented is being used to set a clear argument. This could have helped the report flow and clarity.
> The presentation of recommendations under each of the findings sections is confusing and leads to a long-list of recommendations that is not appropriate to the needs of users. The recommendations would have been better developed after the conclusions and summary of lessons learnt (Sections 4.1 and 4.2). Instead, the current long-list of recommendations under the findings sub-sections could have been reworked as “lessons learnt” specific to each of the evaluation questions (i.e. lines of enquiry) in the findings sub-sections and later summarised in Section 4.2.
> The resulting long-list of more than twenty recommendations is under-developed and it is unclear who should be doing what. Many could have been clustered across multiple findings – for example the current recommendations 5, 9, 10, 11, 13, 16, 18 and 19 all refer in different ways to how the GIL programme might strengthen linkages from training into employment and self-employment. A single recommendation on strengthening linkages to employment would likely have improved their utility
> Limiting the evaluation to 6-7 recommendations would have been more appropriate - for example, these could have responded directly to the 7 existing conclusions
> It is important for the recommendations to be more fully developed to identify duty bearers, timelines and their implications. As has been done with the conclusions, the recommendations could have been clustered around different aspects of 'what next' from across the different findings.

SECTION RATINGS			
SECTION A:	**EXECUTIVE SUMMARY (weight 5%)**	67%	Comments on Rating
Question 1.	Can the executive summary inform decision-making?		
i	Is clearly presented, serves as a standalone document useful for informing decision making, and is of relevant conciseness and depth for key users (Maximum of 5 pages unless otherwise specified in ToR).	Partially	The executive summary works as a stand alone document. However it is lengthy at over 6 pages. Greater precision could have been provided in the presentation of findings which are provided as a discussion narrative that lacks precision. In the absence of any conclusions it is hard to determine if there is any differential priority to the recommendations. The recommendations appear as a long-list that is quite difficult to relate to the general discussion of findings as presented.
ii	Includes all necessary elements (overview of the intervention, evaluation purpose, objectives and intended audience, evaluation methodology, key conclusions on findings, lessons learned if requested, key recommendations) as per the ToR.	Partially	The executive summary includes most of the necessary elements, namely an overview of the GIL programme, the evaluation purpose and objectives, intendied audience and users, methodology and recommendations. However, no review of the substantial differences in the evaluation questions between the inception and final report is provided. The evaluation report lacks a set of findings statements, and does not provide clear conclusions. This makes the recommendations appear to lack grounding in the findings when presented in the executive summary.
iii	Includes all significant information needed to understand the intervention and the evaluation AND does not introduce new information from what is presented in the rest of the report.	Yes	The summary provides sufficient information on the GIL programme and results, and does not introduce new information from the rest of the report.
SECTION B:	**BACKGROUND (weight 5%)**	36%	Comments on Rating
Question 2.	Is the object of the evaluation clearly described?		
i	Clear and relevant description of the intervention, including: location(s), timelines, cost/budget, and implementation status.	Partially	The report includes a good description of the GIL intervention, locations, timeline, implementation and associated national processes; it is unfortunate that no examination of funding or budgeting (including future prospects) is provided.
ii	Clear and relevant description of intended rightsholders (beneficiaries) and duty bearers (state and non-state actors with responsibilities regarding the intervention) 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, with disaggregation by gender, age, disability . . . (as appropriate to the purpose of the evaluation).	Partially	Although different stakeholders are identified and overall numbers of beneficiaries reached, this is distributed in the report and there is little further analysis of their specific interests. No use of rights based language is present.
Question 3.	Is the context of the intervention clearly described?		
i	Clear and relevant description of the context of the intervention (i.e. relevant policy, socio-economic, political, cultural, power/privilege, institutional, international factors) and how context relates to the implementation of the intervention.	Partially	The context analysis provides an adequate and relevant description of the Lebanon context and how changes in economic, COVID and political factors have affected Lebanon's youth. However, it lacks a summary of youth opportunities and gaps across the different Governorates that would have given some of the background to the geographical variances later identified in the findings. The context analysis would have also benefited from a stronger examination of the evolving socioeconomic context (beyond employment) has affected Lebanon's youth, and the extent to which this has been addressed in the country's youth policies and associated institutional arrangements.
ii	Linkages drawn to the SDGs and relevant targets and indicators for the area being evaluated.	No	Indirect reference to the SDGs is provided in the literature / footnotes only, while no SDG linkages are described in the main narrative.
iii	Clear and relevant description (where appropriate) of the status and needs of the rightsholders/beneficiaries of the intervention.	Partially	The status and needs of youth as the key rightsholders is partially described. While the context analysis provides a description and review of youth unemployment, and identifies gaps in the representation of youth in political processes, it does not relate this summary to a clear analysis of how Lebanon's policy, political, socioeconomic, educational, institutional environment has affected Lebanon's youth as rightsholders. The analysis instead jumps forward to a summary of the GIL programme rationale and the contextual factors that have influenced the programme, rather than the youth targeted by the programme.
Question 4.	Are key stakeholders, their relationships and contributions clearly identified?		
i	Identification of implementing agency(is), development partners, right holders, and additional duty bearers and other stakeholders; and of linkages between them (e.g., stakeholder map) (if relevant).	No	The evaluation report lacks any substantive stakeholder analysis. An annex mapping the main stakeholders and their interests would have been helpful and adequate withough adding significantly to the narrative. The Inception Report (IR) to some extent provides a level of stakeholder analysis in the determination of 7 categories of stakeholders for interviews. However this analysis is not referenced in the evaluation report or developed further.
ii	Identification of the specific contributions and roles of key stakeholders (financial or otherwise), including UNICEF.	Partially	The contributions of UNICEF and its national partners including Governorates is to some extent covered but is distributed in the text. Further development a single paragraph on users of the evaluation (currently repeated on pages 9, 10 and 28) with an annex providing the evaluation stakeholder analysis would have been sufficient.
SECTION C:	**EVALUATION PURPOSE, OBJECTIVES AND SCOPE (weight 5%)**	60%	Comments on Rating
Question 5.	Is the purpose of the evaluation clearly described?		

i	Purpose of evaluation is clearly defined, including why it was needed at that point in time, its intended use, and key intended users.	Partially	The purpose is "to provide a comprehensive picture of the impact of the Programme's services on its beneficiaries, the sustainability of the Innovation Labs, and the weaknesses of the Programme." While the intended users and use are identified, it has to be assumed that the evaluation was undertaken at this point in time to provide a summative analysis of the impact and sustainability of the intervention's '3D' workplan. Unfortunately, this isn't clearly stated or linked to the potential possible interests of UNICEF and its government or governorate partners.
Question 6.	Are the objectives and scope of the evaluation clear and realistic?		
i	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 (if applicable).	Yes	A clear description of the evaluation is provided. No changes to the overall objectives were made.
ii	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, if applicable, 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).	Partially	While relevant information on the evaluation's geographical scope across Governorates and timeline is provided in Section 2.2, further relevant information has not been consolidated. No analysis of known differences across Governorates is provided (e.g. Hub locations) that would have informed the sampling arrangements, or the provisional reasons for the geographical variances found in the findings. Nor is there an examination of how data availability and quality affected the evaluation scope. Data limitations are instead explored under the methodology. Most of this could have been addressed during the inception phase and in the inception report (IR).
Question 7.	Is the theory of change, results chain or logic well articulated?		
i	Clear description of the intervention's intended results, or of the parts of the results chain that are applicable to, or are being tested by, the evaluation.	Partially	The evaluation relies on a high level narrative description of the GIL programme. The reliance on broad "discover, develop, do" headings (relating to the programme process) provide a limited sense of how UNICEF entry points were expected to contribute to anticipated change pathways and impacts. There is no examination of any associated assumptions, such as Governorate support, or the integration of linkages between programme training activities and through industry partnerships or self-employment. An over-arching TOC developed retrospectively by the evaluation process would have greatly benefited the programme, the evaluation process and analysis, the presentation of learning, and development of ways forward.
ii	Causal relationship between outputs and outcomes is presented in narrative and/or graphic form (e.g., results chain, logic model, theory of change, evaluation matrix).	Partially	The over-arching "discover, develop, do" framework is activity / delivery focused and hardly responds to the impact analysis requested by the TOR. While indicators are listed in, for example, the evaluation matrices of the inception report (IR) and final evaluation report (ER), neither the evaluation matrix not the narrative provide an adequate examination of how these are meant to fit into any change pathway. A simple graphic could have greatly helped this.
iii	For theory-based evaluations, the theory of change or results framework is assessed, and if requested in the ToR, it is reformulated/improved by the evaluators.	Not Rated	Arguably, this should have been a theory based evaluation with a methodology based around retrospective development of a theory of change (ToC). However, no reference to a ToC or theory based process is provided in the ToR, IR or ER. It is not clear why this was not adopted or even explored as an option. A proper evaluability assessment and examination of data limitations during the inception could have refocused the evaluation approach but appears to have been absent until the data collection phase. Although reference to a ToC is in the abbreviations tables of the IR and ER, none is provided or referenced in either report.
SECTION D:	**EVALUATION DESIGN AND METHODOLOGY (weight 20%)**	59%	Comments on Rating
Question 8.	Does the evaluation use questions and the relevant evaluation criteria that are explicitly justified as appropriate for the purpose of the evaluation? *UNICEF evaluation standards refer to the OECD/DAC criteria - Relevance; Coherence; Effectiveness; Efficiency; Sustainability; Impact (not all are necessarily relevant for all evaluations). Evaluations should also consider equity and leaving no-one behind, gender and human rights based approach (these can be mainstreamed into other criteria). Humanitarian evaluations should also consider Coverage; Connectedness; Coordination; Protection; Security.*		

i	Evaluation questions and sub-questions are appropriate for meeting the objectives and purpose of the evaluation. The relevant criteria are specified and are aligned with the questions.	Partially	While 5 criteria are identified in the ER (Relevance, Effectiveness, Efficiency, Impact, Sustainability), these differ from both the evaluation ToR and ER. No justification of the changes - albeit probably appropriate - is provided. A proper review of the criteria would have been helpful to the reformulation of evaluation questions (EQs), development of the methodology, and presentation of findings. It is unfortunate this did not happen. For example, there was no clear scrutiny of the Impact criterion in the IR or ER. Given the short temporal scope of the GIL programme being evaluated, and the programme's data availability and quality limitations (Section 2.3) this would have been useful. While introduction of the Effectiveness criterion was appropriate to the evaluation Scope, no justification or explanation of these changes to the criteria is provided. The evaluation lists 32 EQs in the IR and 36 in the ER. As with the criteria, changes to the EQs between the ToR, IR and ER are not explored or justified. This is also an extra-ordinarily large number of EQs and simply not realistic. A simpler approach could, for example, have provided 5 overarching EQs based on the 5 selected criteria, each with 1 to 3 sub-questions. The EQs listed would then have emerged as lines of enquiry that could have been clustered under their own sub question. This is in effect what the report has ended up doing in the presentation of findings. The EQs listed - de facto lines of enquiry - are relevant to the criteria.
ii	In addition to the questions and sub-questions, the evaluation matrix includes indicators, benchmarks, assumptions and/or other processes from which the analysis can be based and conclusions drawn.	Partially	Indicators are included in the evaluation matrix. No analysis of assumptions is provided. The use of a ToC could have included the presentation of activities and expected results along a set of inter-linked change pathways, each with their selection of associated indicators, and analysis of assumptions.
Question 9.	Does the report specify adequate methods for data collection, analysis, and sampling?		
i	Evaluation design and set of methods are relevant and adequately robust for the evaluation's purpose, objectives and scope; and are fully and clearly described.	Yes	The methodology is appropriate to the evaluation subject and scope and is well described.
ii	Data sources are appropriate - these would normally include qualitative and quantitative sources (unless otherwise specified in the ToR) - and are all clearly described.	Yes	Different data sources, including qualitative and quantitative, are adequately described.
iii	Sampling strategy is provided - it should include a description of how diverse perspectives are captured (or, if not, provide reasons for this).	Partially	While the reader is to some extent able to ascertain the sampling strategy from the narrative, this isn't clearly presented as a stand-along summary, and it is difficult to clearly ascertain how the sampling strategy was designed to ensure the different views of stakeholders were to be captured. A table and associated paragraph would have provided an adequate summary of the sampling approach and how stakeholders were included in the evaluation process.
iv	Clear and complete description of the methods of analysis.	Partially	The analysis of methods described include thematic analysis, benchmarking, case examples (successes), and mixed quantitative and qualitative methods. Each is explored in only a limited way (see also limitations: 9.vi). It is only in later sections of the report that one is able to get a clear sense of what was intended through these analytical methods. A little more depth of analysis of the matrix indicators would have improved Section 2.2.
v	Methodology allows for drawing causal connections between outputs and expected outcomes.	Partially	In the presentation of the methodology, the reader is largely left to develop their own picture of causal connections. This is inadequate to the needs of decision-makers including their ability to build a proper understanding of the GIL programme design, and how it was explored by the evaluation design. A clear description of the expected change pathways is needed, that describes how each planned activity was designed to lead to one or more specific output(s), and how these outputs were meant to contribute to the programme outcomes.
vi	Clear and complete description of limitations and constraints faced by the evaluation, including gaps in the evidence that was generated and mitigation of bias, and how these were addressed by the evaluators (as feasible).	Partially	Section 2.3 provides a good description of evaluation limitations and constraints. A key area relates to data availability and quality, where the evaluators found "*no coherent, exhaustive set of response variables (impact) and explanatory variables that would describe the situation of all beneficiaries regardless of which content partner provided support. Secondly, the scope of questions asked changed across the years*". More information on mitigating actions to address the limitations would have been useful. Again, a simple tabulation could have helped the presentation here.
Question 10.	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:		

i	Explicit and contextualized reference to the UNEG obligations of evaluators (independence, impartiality, credibility, conflicts of interest, accountability) and/or the principles in the 2020 revised UNEG Ethical Guidelines (integrity, accountability, respect, beneficence).	Partially	No explicit reference to UNEG obligations and associated protocols is provided, or exploration of UNEG evaluation principles. Relevant and related concepts are nevertheless captured indirectly under Annex 6.5.9 which described how the evaluation follows UNICEF's Procedure for Ethical Standards in Research, Evaluation, Data Collection and Analysis. These aspects were to some extent better covered in the IR but not referenced in the ER.
ii	Description of ethical safeguards for participants appropriate for the issues relevant to methodology and how they are applied (respect for dignity and diversity, right to self-determination, fair representation, compliance with codes for vulnerable groups, confidentiality, and avoidance of harm). For those cases where the evaluation involves interviewing children, explicit reference is made to the UNICEF procedures for Ethical Research Involving Children.	Partially	While relevant examples of principles attended to in the evaluation are described under Annex 6.5.9, the evaluation does not attend specifically to UNICEF safeguarding principles. The annex is not referenced in the main narrative or associated editing to indicate efforts to describe its use in the evaluation process.
Question 11.	Does the evaluation incorporate innovative practice that adds value to the evaluation process?		
i	Innovation practice is used to improve the quality of evaluation process. This could be evident in several ways such as the design of the methodology (i.e. use of technology for data gathering, extensive participatory processes, systematic analysis processes such as collaborative outcomes reporting and incorporation of big data, specific strategies to address complexity such as outcome harvesting, strong child rights focus), or ways of sharing of evaluation results.	Partially	The evaluation's extensive use of self-assessments against guided questions in association with the analysis of programme results was to some extent innovative and, although not described as such, did to some extent mitigate the evaluation's data limitations.
SECTION E:	**EVALUATION FINDINGS (weight 25%)**	33%	Comments on Rating
Question 12.	Do the findings clearly address all evaluation objectives and scope?		
i	Findings marshal sufficient levels of evidence to systematically address all of the evaluation's questions, sub-questions and criteria.	Partially	While the evidence base for the findings involves different data sources to address the EQs, it is not systematically presented and lacks clarity and depth. For example, when exploring the Hakaya GIL and Girls Got IT initiatives (finding 3.1.2.3) the evaluation appears to triangulate data from different training attendee feedback, reports, key informant sources. However, the text that presents these sources is not clear. Take for example the following: "Till date, GGIT supported more than 3,536 girls who were introduced to the world of STEAM. Across the events over 170 workshops have been organized. Lastly, the events involved the support of 40 entrepreneurs across Lebanon and were assisted by 120 volunteers from universities and NGOs". As with other areas of analysis under this and multiple other findings, it is very difficult to build a picture of what exactly is meant here, why this data is being used, or how it links with the other data sources to support the finding.
ii	Explicit use of the intervention's results framework/ToC in the formulation of the findings.	No	The basis of the presentation of the evaluation findings is against the criteria and their associated evaluation questions (line of enquiry) rather than against any programme results framework. See above (question 7) responses to the gap in provision of any ToC or change pathway / results framework in the evaluation approach.
Question 13.	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.		
i	Evaluation uses credible forms of qualitative and quantitative data. It presents both output and outcome-level data as relevant to the evaluation framework. Triangulation is evident through the use of multiple data sources.	Partially	While triangulation is to some extent evident in the use of different primary and secondary sources, mixed data types and stakeholder responses, it is unfortunate that the evidence base is not systematically presented in a way that allows the reader to understand how the different sources correlate with, or counter one another, in relation to the evaluation question. No findings statement is provided against which to understand and relate the different evidence examples. This makes their selection to at times appear random. Some better use of graphics and/or tabulated summaries and a shortening of the findings narrative would have been helpful to the report's clarity. For example, in the evaluation's analysis of the quality and effectiveness of the GIL's courses in building youth entrepreneurship skills (3.2.1.2), the finding starts with a heavy reliance on post-assessment graduate satisfaction ratings and attendeee feedback before describing other stakeholder perspectives (in which empolyers are mentioned in passing) and reviewing post-assessment results, but does not sufficiently draw the different sources together as part of a clear examination of the evaluation question.

ii	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 as relevant for each question.	Partially	The findings are presented in line with the GIL programme process and do not follow the order or presentation of the evaluation matrix. They are presented as discursive narratives with diagrams against sub-headings that capture 1 or more EQ (line of enquiry). The report would have benefited from being structured against the evaluation matrix order, starting with an explicit relevance EQ with associated sub-questions and lines of enquiry. A summary of the specific finding(s) relevant to each EQ (line of enquiry) would have helped make it absolutely clear what each finding was before moving into the descriptive text and data analysis that provides relevant evidence to support the finding. Whilst helpful, the use of bold sentences in most paragraphs to emphasise different points does not provide a coherent picture or compensate for the lack of any clear and specific findings statements.
iii	Causal factors (contextual, organizational, managerial, etc.) leading to achievement or non-achievement of results are clearly identified. For theory-based evaluations, findings also analyse the logical chain (progression -or not- from implementation to results).	Partially	While causal factors are described, there are significant gaps. For example, the differences in uptake and support for Innovation Hubs across Governorates is not explained.
Question 14.	Does the evaluation assess and use the intervention's Results Based Management elements?		
i	Assessment of the adequacy 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) to support decision-making.	No	Although multiple tables refer to UNICEF monitoring data as their source, there is no clear examination of the performance framework, or M&E tools and their use excepting for in the analysis of data limitations in Section 2.3.
SECTION F:	**EVALUATION CONCLUSIONS & LESSONS LEARNED (weight 10%)**	88%	Comments on Rating
Question 15.	Do the conclusions clearly present an objective overall assessment of the intervention?		
i	Conclusions are clearly formulated and reflect the purpose and objectives of the evaluation. They are sufficiently forward looking (if a formative evaluation or if the implementation is expected to continue or have additional phase).	Yes	The evaluation conclusions provide an appropriate consolidation of the findings. They are clearly presented and forward looking.
ii	Conclusions are derived appropriately from findings, and present a picture of the strengths and limitations of the intervention that adds insight and analysis beyond the findings.	Yes	The conclusions illustrate a good degree of coherence with the findings and examine both the strengths and weaknesses of the GIL programme with useful insights for future organisation of the programme.
Question 16.	Are logical and informative lessons learned identified? [N/A if lessons are not presented and not requested in ToR]		
i	Identified lessons stem logically from the findings, have wider applicability and relevance beyond the object of the evaluation.	Yes	Learning is identified in the TOR and as the main objective of the IR and ER which "aims to generate constructive learning and lessons for the future of the intervention". A good section on lessons learnt is provided (Section 4.2) after the conclusions that captures a many of the lessons identified under the findings.
ii	Lessons are clearly and concisely presented, yet have sufficient detail to be useful for intended audience.	Partially	In addition to Section 4.2 (lessons learnt), each of the findings sections ends with a summary of recommendations associated with the finding. While it is useful to have some form of 'so what' summary to the findings, these would have been better constructed in relation to lessons and learning that fed into the conclusions and recommendations rather than as recommendations (as well as being summarised under Section 4.2). The lessons information is not missing, but the reader often has to look for it under the recommendations summaries which is confusing.
SECTION G:	**RECOMMENDATIONS (weight 15%)**	25%	Comments on Rating
Question 17.	Are recommendations well grounded in the evaluation?		
i	Recommendations align with the evaluation purpose, are clearly formulated and logically derived from the findings and/or conclusions.	Partially	The evaluation recommendations derive from the findings and are clearly presented. Unfortunately the presentation of recommendations to align with each of the findings has led to a long-list that is not fully coherent with the conclusions and is confusing.
ii	Recommendations are useful and actionable for primary intended users and uses (relevant to the intervention); guidance is given for implementation, as appropriate.	Partially	While it is likely the recommendations could be interpreted by users and appear relevant, they do not provide adequate guidance to implementation and appear superficial. This is not helped by the presentation of a long-list which makes it what exactly should be done by whom, by when. Many could have been clustered against the key themes (e.g. conclusions) emerging from the evaluation findings, rather than listed against the findings. For example, several relate to efforts to improve employment and self-employment opportunities and could have been developed further as a whole.
iii	Process for developing the recommendations is described, and includes the Involvement of duty-bearers, as well as rights holders when feasible (or explanation given for why they were not involved).	No	The processes for developing the recommendations are not described. It cannot be assumed that they were developed in a participatory manner during the validation workshop with UNICEF, Content Partners and Lab managers (identified in the methodology).
Question 18.	Are recommendations clearly presented?		

i	Clear identification of groups or duty-bearers responsible for action for each recommendation (or clearly clustered group of recommendations). Clear prioritization and/or classification of recommendations to support use.	No	No information is provided as to the duty bearer(s) or the timeline for any of the recommendations, nor is there a review of the potential for engaging youth stakeholders (rights bearers) in their development. Clustering and developing the recommendations against key themes with the engagement of stakeholders would likely have been beneficial to the evaluation process.
SECTION H:	**REPORT STRUCTURE AND PRESENTATION (weight 5%)**	79%	Comments on Rating
Question 19.	Does the evaluation report include all relevant information?		
i	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	The introduction is clearly presented and contains all relevant information.
ii	Annexes include: terms of reference, evaluation matrix, list of interviewees, results chain/ToC/logical framework (unless included in report body), 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, information about the evaluator(s), etc.	Partially	Annexes include the terms of reference, evaluation matrix, data collection guidance (semi-structured KIIs) and bibliography. Annexes that would have also been useful to support the evaluation process, methodology and findings but which are missing include a GIL results chain/ToC/ or logical framework, a summary of the sampling approach and some level of stakeholder analysis.
Question 20.	Is the report logically structured?		
i	Structure is easy to identify and navigate (for instance, with numbered sections, clear titles and sub-titles, well formatted).	Yes	The structure and presentation is clear and well formatted
ii	Structure accords to UNICEF guidelines for evaluation reports: context, purpose and methodology would normally precede findings, which would normally be followed by conclusions, lessons learned and recommendations.	Yes	The structure is in accordance with UNICEF guidelines with context and purpose followed by methodology, findings, conclusions and lessons learnt, and finally the recommendations.
Question 21.	Is the report well presented?		
i	Report is of reasonable length; it does not exceed number of pages that may be specified in ToR.	No	The report could have been edited down by a clearer presentation of the findings against individual findings statements. Without this, the findings adopt a fairly lengthy narrative format that could have been more precisely presented. As such, it stands at 102 pages long, excluding annexes.
ii	Report is easy to understand (written in accessible way for intended audience) and generally free from grammar, spelling and punctuation errors.	Yes	The report is well written, has few errors and is easy to understand.
iii	Frequent use of visual aids (such as infographics, maps, tables, figures, photos) to convey key information. These are clearly presented, labelled, and referenced in text.	Yes	There is a good use of visual aids throught that break up the narrative to some extent.
SECTION I:	**EVALUATION PRINCIPLES (weight 10%)**	52%	Comments on Rating
Question 22.	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?		
i	Reference and use of rights-based framework, and/or CRC, CCC, CEDAW and/or other rights related benchmarks in the design of the evaluation.	No	The evaluation does not reference rights-based frameworks or benchmarks, adequately consider human rights aspects, or gender equality and equity. The methodology fails to explore the use of an appropriate rights-based framework in the evaluation design.
ii	Clear description of the level of participation of key rights holders and duty bearers in the conduct of the evaluation (for example, a reference group is established, stakeholders are involved as informants or in data gathering).	Partially	Within the scope of the evaluation, the participation of duty bearers is appropriate but attention to describing duty bearers and their responsibilities is limited.
iii	Language is empowering and inclusive, avoiding gender, heterosexual, age, cultural and religious bias, among others; use terminology of rights holders and duty bearers; data is disaggregated by marginalized group; differential results are assessed (distribution of results across different groups).	Partially	While the language and use of rights-based terminology and analysis is limited, the use and analysis of disaggregated data by sex and disability allows for a reasonable level of analysis of who benefitted from the different aspects of the GIL programme.
iv	Evaluation assesses the extent to which the implementation of the intervention addresses child rights and Leave No-one Behind (gender and other excluded and marginalized groups). It is disability inclusive.	Yes	The evaluation explores the distribution of results across different youth groups, including girls, boys and people with disabilities, and emphasises the importance of better addressing the needs of geographically marginalized and disadvantaged youth.
Question 23.	Does the evaluation meet UN SWAP evaluation performance indicators? (Note: this question will be rated according to UN SWAP standards with detail provided below)	5	

i	GEEW is integrated in the Evaluation Scope of analysis, and evaluation criteria and questions are designed in a way that ensures GEEW-related data will be collected.	Partially integrated	The extent to which the programme strategy and interventions integrated gender in their design, implementation, and monitoring is partially included in four of the evaluation questions and as part of the lessons learnt.
ii	A gender-responsive methodology, methods and tools, and data analysis techniques are selected.	Satisfactorily integrated	Gender analysis was integrated into the data collection instruments and the examination of output to outcome data.
iii	The evaluation Findings, Conclusions and Recommendations reflect a gender analysis.	Satisfactorily integrated	A satisfactory rather than in-depth level of gender analysis is integrated that includes the examination of programme inclusion and support for girls.

SWAP Rating Guidance

i GEEW is integrated in the Evaluation Scope of analysis, and evaluation criteria and questions are designed in a way that ensures GEEW-related data will be collected.
a. Does the evaluation assess whether sufficient information was collected during the implementation period on specific result indicators to measure progress on human rights and gender equality results?
b. Does the evaluation include an objective specific to assessment of human rights and gender equality considerations or was it mainstreamed in other objectives?
c. Was a standalone criterion on gender and/or human rights included in the evaluation framework or mainstreamed into other evaluation criteria?
d. Is there a dedicated evaluation question or sub-question regarding how GEEW was integrated into the subject of the evaluation?

ii A gender-responsive methodology, methods and tools, and data analysis techniques are selected.
a. Does the evaluation specify how gender issues are addressed in the methodology, including: how data collection and analysis methods integrate gender considerations and ensure data collected is disaggregated by sex?
b. Does the evaluation methodology employ a mixed-methods approach, appropriate to evaluating GEWE considerations?
c. Are a diverse range of data sources and processes employed (i.e. triangulation, validation) to guarantee inclusion, accuracy and credibility?
d. Does the evaluation methods and sampling frame address the diversity of stakeholders affected by the intervention, particularly the most vulnerable, where appropriate?
e. Were ethical standards considered throughout the evaluation and were all stakeholder groups treated with integrity and respect for confidentiality?

iii The evaluation Findings, Conclusions and Recommendations reflect a gender analysis.
a. Does the evaluation have a background section that includes an intersectional analysis of the specific social groups affected by the issue or spell out the relevant normative instruments or policies related to human rights and gender equality?
b. Do the findings include data analysis that explicitly and transparently triangulates the voices of different social role groups, and/or disaggregates quantitative data, where applicable?
c. Are unanticipated effects of the intervention on human rights and gender equality described?
d. Does the evaluation report provide specific recommendations addressing GEWE issues, and priorities for action to improve GEWE or the intervention or future initiatives in this area?