GEROS Evaluation Quality Assurance Tool

Version: September 2021

Evaluation of the Sanitation Technology Demonstration in Emergency Settings in Lebanon Programme (STDP)

REPORT RATING SUMMARY			
Overall Rating		53%	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 use the evaluation with caution. Substantive improvements in some areas are needed**
● ▬ ▬ ▬	Unsatisfactory (0% - 34.99%)	1	

REPORT DETAILS	
Title of the evaluation report	Evaluation of the Sanitation Technology Demonstration in Emergency Settings in Lebanon
Report sequence number	Lebanon-59-2023-18446
Region	MENA
Year of report	2022
Office	Lebanon
Coverage (countries)	Lebanon
ToRs present	Yes
Date of review (dd/mmm/yyyy)	June 20, 2023
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	No
Every child is protected from violence and exploitation	No
Every child lives in a safe and clean environment	Yes
Every child has an equitable chance in life	No
Gender equality (cross-cutting)	Yes
Humanitarian action (cross-cutting)	Yes
Evaluation object	Pilot/innovation
Evaluation type	Summative
Evaluation strategy	Mixed methods
Evaluation design *(primary method used)*	Theory-based
Evaluation level	Output & Outcome
Geographic scope	National
Primary SDG(s) covered *(number)*	3, 6, 9, 11

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 evaluation was commissioned during the last phase of implementation of the Sanitation Technology Demonstration in Emergency Settings Programme (STDP) and aimed to provide a clear evolutionary assessment to help in defining the potential for developing and implementing decentralised water and sanitation (DEWATS) solutions to displaced and host communities in Lebanon. A focus on future roll out through small and medium sized enterprises (SME's) was envisaged by UNICEF who needed a level of validation of the range of STDP models tested through 3 'batches' or phases of analysis. The purpose of the evaluation was to provide lessons about the challenges and success factors of implementing innovative sanitation solutions through an independent assessment of the project design, implementation, and performance, and to identify success/failure factors, critical barriers and enablers to adapting, testing and mainstreaming requirements. The evaluation used a mixed methods approach involving quantitative methods for the technical assessment across 13 sites (informal settlements) 3 sites were visited for qualitative data collection, including a neighbouring host community. Laboratory tests were conducted for 16 systems across the 13 sites.

Key strengths of the evaluation:

> A constructive inception phase and engagement with UNICEF and implementing partners (IPs) across sites
> A clear executive summary
> The incorporation of project, technical, innovation and replication analyses across the methodology, findings and conclusions though incomplete was an effective structure of analysis relative to the ToR
> A strong technical (sanitation, environmental) analysis to determine systems effectiveness and a review of the batch process to determine the most appropriate technology options

Areas for improvement:

> Qualitative analysis including gender and rights-related considerations appeared secondary to technical considerations. While this was a reflection of limitations in the STDP design, the evaluation could have gone further in building an understanding of gender requirements through the innovation system ("funnel") and whether there were gender related differences in sanitation systems preferences
> A lack of attention to the users of the evaluation and how the recommendations might need to be adjusted to ensure utility - e.g. roles within organisations, and attention to how IPs and government counterparts may or may not respond positively to recommendations they are expected to implement. Possible constraints to this aren't considered
> A confusing presentation of findings without numbering against clusters of EQs rather than individual sub-questions makes it difficult to track the relationship to the evaluation matrix
> A weak analysis of financial considerations in the findings undermines their ability to clearly support decision-makers as to where opportunities do or do not lie. This is only partly mitigated in the conclusions
> A weak presentation of triangulated evidence to give confidence to the findings, using visual tabulation and figures in the narrative and annex

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.*

Examples of improvements that could be made:
> Strengthening the presentation of findings in line with the evaluation questions (Eqs) and sub-questions
> Integrating the retrospective theory of change and updated theory of change (ToC) into the overall methodology and analysis including an analysis of assumptions under relevant sub-questions
> Giving greater attention to the presentation of tabulated data and figures to support the background of the evaluation (e.g. IPs across sites) and findings. Ensure figures are readable. This includes the ToC, primary evaluation data (from KIIs, FGDs), quantitative sanitation and environment (technical) assessments
> Including an examination of UNICEF and IP monitoring systems and the data they generated as part of the data collection and analysis - even if this shows significant gaps
> Better developing of the financial analysis in relation to the public private partnership aspirations of the STDP design and evaluation purpose
> Using stakeholder mapping - including roles and interests - to inform the methodology, analysis and recommendations
> Giving much more consideration to the structure of the evaluation methodology and its relationship to the presentation of findings and conclusions. For example using the areas of project, technical, innovation and replication analysis developed by the evaluation team throughout the report and presenting this visually in relationship to the EQs
> Giving greater focus to the presentation of findings, conclusions and recommendations to reduce duplication. The conclusions and lessons are too long. All should include a numbered, summary statement
> Undertaking a thorough edit to ensure the presentation and grammar are clear to the reader

SECTION RATINGS			
SECTION A:	**EXECUTIVE SUMMARY (weight 5%)**	83%	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).	Yes	> A clear summary of the evaluation findings conclusions and recommendations is provided that is accessibly written and presented and that works well as a standalone document > The summary is concise and precise in the presentation of findings which link well to the presentation of conclusions and recommendations
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	> Most necessary areas included including the evaluation context, an overview of the intervention, the evaluation purpose, methodology, main findings, conclusions recommendations > Gaps include the intended audience and lessons learnt
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 executive summary is well presented and includes the information required to understand the intervention and overall evaluation assessment. Nothing new is included. The presentation of findings is clear and coherent
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	> A description of the intervention draws on the ToR to identify the STDP's Locations, Timeline and Implementation status. However the presentation of the project Purpose is not correct. It should reflect the ToR which describe the need to 1) identify and implement innovative sanitation technological solutions to treat wastewater generated by Syrian refugees and discharged from Informal Settlements (IS) in Lebanon, which can serve as a lesson learned to improve global humanitarian responses in similar contexts. And, (2) to build a strong Public Private Partnerships (PPP) benefitting Lebanese and Refugees (originally with the same treatment). Unfortunately the evaluation report appears to confuse the project purpose of the STDP that should have been presented in the evaluation Background with the evaluation purpose > The implementation phases and status is described but no financial information is provided. This is disappointing given a key priority is to understand cost drivers behind any potential PPP
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	> Rights holders are described in terms of beneficiary numbers in informal and host communities but without disaggregation or further analysis. No summary analysis of beneficiaries in the 42 pilot sites is provided. It isn't always clear who is being referred to in the information provided. For example, it is unclear whether "8% of the population is [...] served by wastewater treatment, while only 60% of the population is connected to a sewage collection network. Poor on-site pit latrines or septic tanks are common..." refers to informal settlements, host communities or both, and/or who is excluded > Duty bearers are listed but no additional information provided. For example, analysis of the costs/budget by year by implementing agency would have been useful to set up a picture of the intervention, progress so far, UNICEF's leadership role, and the nature of its relationship with its partners
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 contextual analysis provides a summary of the ToR. While this is appropriate, other relevant areas that could have been developed further to complement the evaluation include: - Better clarity on the political context: page 15 reads "the political decision by the Government of Lebanon (GoL) not to officially recognize these settlements has consequences". The consequences are described, but not what the "decision" was and why it was made. This is important to the scale ambitions of UNICEF. Is there political appetite and support for expanded access to an effective sanitation system? - The policy arrangements that oversee this political position - The institutional arrangements under which government operate in rural areas (whether host or informal communities) and how this may affect public authority acceptance of the intervention as described in the thematic scope. This would allow the evaluation to link forward to the scaling-up and sustainability analysis
ii	Linkages drawn to the SDGs and relevant targets and indicators for the area being evaluated.	No	> No reference is made to SDG targets and indicators > Relevant water and sanitation standards and benchmarks under the Humanitarian Charter and Minimum Standards in Humanitarian Response would have provided a useful contextual reference to benchmark where the STDP is currently and what it aims to achieve. > This is especially the case where EQ5 of the evaluation focuses on opportunities to build institutional knowledge about ways to improve global humanitarian responses in similar emergency contexts. > Sphere standards are introduced in the bibliography, but it is unclear where in the narrative they relate to. This and other references should be clearly presented as footnotes and again in the list of documentation
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 the rightsholders are introduced in broad terms - e.g. % of people without sanitation services > However little analysis is provided of the beneficiaries of the STDP across its "42 Informal Settlement (IS) sites distributed between the Bekaa Valley (Mid/Eastern Lebanon) and the Akkar governorate (Northern Lebanon)". This analysis would have given precision to the presentation of the current status and needs of rights holders in the STDP target sites. > It would have also provided an opportunity to examine the current status of sanitation services against Humanitarian benchmarks. > A short examination of the power relationships that operate between host and displaced communities (and how this may vary in different localities) as well as with landlords would have helped understand the context and potential for sustainability
Question 4.	Are key stakeholders, their relationships and contributions clearly identified?		
i	Identification of implementing agency(ies), development partners, right holders, and additional duty bearers and other stakeholders; and of linkages between them (e.g., stakeholder map) (if relevant).	Partially	> The narrative briefly identifies the implementing agencies and relevant government agencies. Reference is given to beneficiary numbers - e.g. "UNICEF estimated that 1,061 sites hosting more than 98,500 Syrian refugees will be targeted by the Project". > This data isn't disaggregated, nor does the background examine the relationships between rights holders and duty bearers. > An important analysis that is absent is the relationship between duty bearers, rights holders and landlords. A map of stakeholder roles would have helped clarify relationships
ii	Identification of the specific contributions and roles of key stakeholders (financial or otherwise), including UNICEF.	No	> The evaluation does not explore the roles and responsibilities of implementing agencies, their planned or actual contributions, or the STDP's / agencies' relationships with relevant government counterparts. > A table / Annex presenting the implementing partners' roles and responsibilities, their technical and financial commitments, site coverage, and engagement with the 3 Batches would have been useful to contextualise the STDP and understand how the multi-site approach was expected to function across the partners
SECTION C:	**EVALUATION PURPOSE, OBJECTIVES AND SCOPE (weight 5%)**	50%	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 evaluation purpose is provided under Section 2. > Changes in the presentation and wording of the evaluation purpose between the TOR, IR and ER aren't explained > For example, it is unclear whether changes in the wording of the purpose between the ToR purpose (innovative sanitation technological solutions) and ER (innovative WASH solutions) meant any substantive shifts and broadening of the evaluation analysis. > It was also difficult understand why the TOR component on "building a strong Public Private Partnerships (PPP) benefitting Lebanese and Refugees" was dropped from the ER purpose and instead included as part of an "additional in the Evaluation Subject. Again this was not explained > This isn't helped by confusion in the list of additional Evaluation "Objectives" and "Specific Objectives" in the ER. The former are equivalent to learning requirements agreed in the IR. The latter map back to objectives listed in the ToR. It isn't explained how the evaluation sought to to navigate the relationships between Purpose, Objectives and areas of learning > Evaluation users are not described. While it may be assumed they include the partners, stakeholders and (potentially) beneficiaries listed under the Evaluation Subject this isn't clearly stated. > Adoption of stakeholder mapping would allow the evaluation to better clarify the relationship between the expected uses of the evaluation and its users including government, partner, donor and wider national and international actors
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).	Partially	> The evaluation provides a good picture of areas of learning the evaluation proposes to contribute under the Purpose, Objectives and Scope. A possible improvement could have been to describe how the evaluation Purpose and Objectives were designed to contribute to / or validate the success ambitions of the STDP described in the Scope and whether there were any limits to this validation. For example by showing how the evaluation Purpose and Objective responded to the project hypothesis described in the Scope: "the STDP would be considered a success if it could demonstrate improvement of sanitary conditions and environmental conditions (via discharge of treated wastewater). These improvements should consider the context of the IS in its design and its targeting, working towards reducing women's vulnerabilities and answering their needs. An additional success would be the national scaling of the technology to IS and host communities with the appropriate conditions, and replication in Syria and other humanitarian contexts in which UNICEF is responding". For example, a table could have been used to map the evaluation purpose and objectives back to the hypothesis
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	> The evaluation Scope provides an adequate description of the overall objectives of the STDP, its development hypothesis, geographical scope > No analysis of areas that would not be covered is provided or reasons for this. While the Limitations section includes information on important limitations surround sanitation data collection, secondary data availability and community participation that may have influenced the purpose and objectives, this was not done. > For example the Limitations section describes how: "...in the ISs [informal settlements] selected for qualitative data collection, the respective IPs [implementing partners] have not sufficiently engaged the host community (such as households in the vicinity of ISs) during the project implementation. Therefore, it was deemed unsuitable for the evaluators to reach out to host community members and explore their perceptions." > In response, some adjustments to the evaluation scope surround gender inclusion and community perspectives would have been useful background to the findings
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 ER provides a limited overview of the expected results chain under sections exploring evaluation Scope and presentation of a retrospective Theory of Change under the Methodology > The STDP results chain would have been a useful addition as an Annex

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	> An infographic of the theory of change is provided that we are told is based on the project results chain. While this appears to indicate the causal relationships along the results chain, it is unfortunate that the presentation of the infographic is impossible to read. > To understand the underlying the EQA process the EQA referred to the If-Then narrative described in the IR which was less developed than the infographic > Inclusion of symbols to indicate how and where the ToC relates to the EQs and evaluation matrix would have helped build understanding of the relationship between the evaluation and the STDP ambitions > A narrative analysis of the alignment between the ToC evaluation matrix and ToC is not provided
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.	Partially	> At the request of the UNICEF country office a retrospective theory of change was developed during the inception phase through a half day workshop. It is not stated which stakeholders participated. > In terms of utility, it was good to see a revised TOC, updated on the basis of the evaluation findings, included in the evaluation conclusions. This may offer a basis to support future phases of the STDP and its scaling-up > However this emerged a little out of the blue after going through the evaluation findings. More could have been done to incorporate the ToC into an actual tool for the evaluation. For example: - Use the TOC to indicate specific areas of the results chain that were and were not covered by the evaluation - Review any underlying assumptions behind the ToC and how they were or could have been addressed during the evaluation, if at all - Show how the EQs and methods mapped onto different parts of the TOC
SECTION D:	**EVALUATION DESIGN AND METHODOLOGY (weight 20%)**	73%	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.	Yes	> Together, the Evaluation questions presented under the Purpose and Objectives section and evaluation matrix (Annex 3) are largely appropriate for meeting the objectives and purpose of the evaluation > Sensible incorporation of lessons (EQ4.5) and Partnerships (EQ8) aligns with the intended scope of the terms of reference > Some text in the narrative could have expanded on the sub-questions described in the evaluation matrix which were comprehensive, adjustments to the EQs listed in the ToR and reasons > There is clear adoption of evaluation criteria and relevance and alignment of the criteria with the EQs and TOR
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	> There is no use of indicators or benchmarks in the evaluation matrix or report. While it would have been possible to address some EQs (e.g. EQ1 relevance, EQ2 cost efficiency, EQ3 technology effectiveness and EQ4 knowledge contributions) the gaps limited the evaluation's ability to explore unintended effects (EQ5), sustainability and scale (EQ6) and gender (EQ7). > A reference to BORDA15 standards under the first Efficiency finding (footnote 20) is an example of what could have been used, alongside credible qualitative indicators > The matrix presents a list of evaluator questions used to explore individual sub-questions. These would have been better placed in a separate Annex with the evaluation matrix providing a shortlist of key indicators and benchmarks. > Underlying assumptions were not presented or used as part of the use of the results chain and TOC methodology and analysis. Again, this limited the more qualitative analysis required in later EQs (5-7). The use of assumptions is limited to technical aspects (flow measurement ratings for un-monitored sanitation units) not the overall STDP assumptions > In other respects it is a comprehensive matrix with data sources, data collection tools and analysis methods provided
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 evaluation methods are comprehensive and adequately described. Methods are also relevant to the purpose and objectives specified in the ToR. They include a strong focus on technical aspects of wastewater sampling in line with the validation requirement of different STDP options.

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	Data sources are adequately described and appropriate. They include a combination of qualitative and quantitative data sources and primary and secondary data sources.
iii	Sampling strategy is provided - it should include a description of how diverse perspectives are captured (or, if not, provide reasons for this).	Yes	The sampling strategy is adequately described and responds to the demands of the ToR including the site selection criteria. A range of stakeholder perspectives and data needs are captured in the outline table of data sources.
iv	Clear and complete description of the methods of analysis.	Partially	The description of the analysis methods is moderately well developed and the combination of methods make for a coherent methodology. > two case studies compare an informal settlement (IS) in the Bekaa and one in a host community in Akkar. However these should be considered as methods of analysis rather than data collection. > the evaluation compares STDP sites against systems in non-target communities that depend on latrines with on-site containment (septic tanks, holding tanks, covered and uncovered pits, cesspits) or direct discharge to water bodies > The presentation of analysis methods would nevertheless have benefited from indicating clearly how the different data sources and analysis contributed to findings across the project, technical and innovation levels. A simple table could have been used
v	Methodology allows for drawing causal connections between outputs and expected outcomes.	No	> It is not clear how the different evaluation methods led to the examination of causal relationships along the results chain. While it is possible to surmise this from a review of the tools adopted, this element is not described by the methodology. It would have been made clearer if the evaluation TOC had been introduced as part of the methodological approach with the different evaluation questions, data collection and analysis methods mapped onto it
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	> Limitations are well described and coherent. They introduce realistic evidence gaps mitigation actions. > The limitations section includes an exploration of possible bias is addressed > It is unclear to what extent the inability to undertake sludge assessment for Batch 3 systems due to recent instalment may have been a barrier to the findings and conclusions surrounding sustainability and scale > The Limitations section would have benefited from an overall assessment as to whether they created a significant barrier to resolving any of the EQs
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).	Yes	> The evaluation ethical concerns are well described and cross reference to UNEG obligations identified within the text. > In community FGDs site invitation were obtained from community WASH committees and IS Shawish. Informed consent was obtained from all participants according to a template agreed with UNICEF
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.	Yes	> A UNICEF ethical review process is described that indicates the evaluation involved no children, included tailored measures when interviewing members of vulnerable groups and addressed risks to the safety, privacy and wellbeing of the evaluation participants. Informed consent forms and protocols were adopted for the protection of participants' identity and for their safety, as well as for data protection.
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 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 STDP focus on innovation is a central theme in the evaluation. Both the IR and ER describe the alignment of the Batch processes and an innovation funnel. The introduction of project, technical and innovation levels to the analysis was an innovative approach. Mapping the EQs onto these dimensions as part of the methodology would have been helpful to the evaluation's ability to build the sustainability case > The integration of properly developed cost assessments (capital and operational expenses) of the different systems alternatives across sites and IPs could have allowed the evaluation to explore opportunities for scale and Public Private Partnerships. Couped with the qualitative assessment of preferences and user and IP experiences this would have represented an innovative component of the evaluation methodology. Unfortunately the financial analysis was not sufficient to sustain this
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	> Depending on the EQ, the findings use of evidence ranges from unfocussed and difficult to verify from the information provided (qualitative assessments of relevance, unintended effects and sustainability) to clear and systematic (quantitative areas of technology effectiveness). This is not helped by gaps in presentation and use of evaluation data sources (KIIs, secondary literature, FGDs) to support the evaluation's qualitative reasoning. The level of detail provided in the technical analysis (e.g. the first findings under section 3.3 effectiveness) should have been mirrored across all EQs > Note that the findings aren't numbered and there is no clear relationship mapped or presented to show which EQ or sub-EQ they respond to > The lack of explicit linkage between the questions addressed in the findings section and the EQ sub-questions in the Evaluation Matrix makes this hard to answer. No numbering is provided in the Findings section. The findings statements in Bold do not correspond with the sub-questions. > This leads to gaps in analysis. For example, the Relevance findings (users, national plans) respond to EQ1.1 (needs) they do not address sub-question EQ1.2 (alignment with UNICEF outcomes) isn't discussed. Whilst the ToR indicate this was likely not a problem, this needs to be said, or the sub-question dropped in the inception phase
ii	Explicit use of the intervention's results framework/ToC in the formulation of the findings.	No	> There is no explicit use of the intervention logic or retrospective ToC to support the presentation of the evaluation findings > The presentation of multiple Findings clustered under multiple EQs presented together under single section headings (criteria) makes the assessment of the evaluation analysis against the ToR and matrix extremely difficult and confusing > The findings aren't numbered. There is no clear presentation of their relationship to the EQs. This makes navigation and utility difficult > It is unclear why the evaluation team went to the effort of developing the matrix and evaluation sub-questions and then ignoring them. These could have provided a perfectly adequate basis for responding to the ToR and structuring the report in a clear and concise way > the numbering of EQs in the report Findings is incorrect. EQ8 (partnerships) has been introduced under Efficiency as EQ3 placing all the following EQs out of there numerical order. Greater attention to editing was needed > Given it was known by UNICEF and the evaluation team that the cost basis of public sanitation systems wasn't known and could not be ascertained, why was this element (EQ3.1) retained in the Inception report? No reason is given as to why the analysis of basic installation and running costs wasn't undertaken properly an why the evaluation team relied on "back of envelope calculations [to] demonstrate that assuming an investment cost of around US $10,000 per system and an operation cost under US $1,000 per system per annum, there is potential for scale-up in host communities and ISs." A more robust presentation of the data may at have provided a reference for decision-makers
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	> The use of mixed data is evident across findings and sometimes within. > The presentation of technical aspects (Sanitation and environmental analysis) is significantly stronger than qualitative areas where the presentation of the findings and findings statements is quite long and discursive. This isn't to say they aren't correct - but that it would help to know how may IS witnessed vandalism and problems with landlords. Such information can easily be tabulated to help clarify what is being said and improve the overall presentation of findings > There is little explicit use of data and analysis triangulation, or triangulation of methods to improve validity > It is a shame that the innovation systems analyse (funnel) is not used as a point of analysis in the findings > Other gaps include private sector engagement (EQ6.3) materials and supplies (6.4), extension to other contexts (6.5), and meeting UNICEF GEEW requirements (7.1) Innovation systems analysis PPP Output and outcome level data Triangulation

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 partially based on performance indicators, standards or benchmarks. More could have been done to develop this analytical framework in the evaluation matrix during the inception phase > A good technical (sanitation, environmental) analysis is used to determine systems effectiveness and a review of the batch process to determine the most appropriate technology options > However, while some reference to benchmarks is provided (e.g. BORDA) there is very limited development of the benchmark analysis despite it being a recognised area of analysis in the methodology > Qualitative analysis including gender and rights-related considerations were not sufficiently explored. While this was to an extent a reflection of limitations in the STDP design the evaluation could have done more to build an understanding of how qualitative factors (social, political, gender) and financial elements influenced the innovation funnel. As examples > To what extent did gender related differences influence sanitation systems preferences? How did that / or might that understanding have fed back into STDP design adjustments? > What were the main cost drivers behind STDP development that need greater consideration / mitigation in future
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	> Causal factors behind the achievement or non-achievement of results are examined including contextual, organizational, social and partnership factors. the identification of vandalism and thefts relating to landlords is particularly important to sustainability and relates in turn to land ownership rights for Informal settlements. This area could have been more clearly presented if it had been identified as a critical assumption during development of the evaluation theory of change in the inception phase > It is unfortunate that the structure and presentation of the findings bears no clear relationship to the logical chain between implementation and results or the ToC
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	> While the evaluation does refer to elements of UNICEF's monitoring system that were found to be absent (for instance gender and cost factors), it does not undertake a review of this and the extent to which this may have held back overall implementation of the STDP and the evaluation's evaluability assessment. > No attempt was made to present or map the availability of UNICEF or IP data in the ER or IR. Both must have played at least a part of the overall STDP monitoring framework
SECTION F:	**EVALUATION CONCLUSIONS & LESSONS LEARNED (weight 10%)**	63%	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).	Partially	> The presentation of conclusions focusing on "innovation, technology and replication" responds to the framing of the evaluation purpose and questions and, to an extent, the analysis of "project, technical and innovation" considerations in the methodology > It is unclear whether this was intentional. The evaluation report would have been more easy to follow and accessible if it this alignment had been developed at the outset - for example by indicating it in the methodology, using the same terminology and presenting the same categories throughout the ER (project, technology, innovation, replication), referencing these areas throughout the findings, and returning to the components in the structure of the conclusions > The redevelopment and presentation of a revised theory of change in the conclusions is positive, insofar it develops the analysis, but is not sufficiently explored in the methodology, findings or conclusions sections. This was an opportunity missed > The presentation of project management considerations as a dedicated subsection of the conclusions would have been helpful to utility (for example adoption of a monitoring framework, agreement of SOPs for handover and gender considerations)
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.	Partially	> The presentation of conclusions is nevertheless forward looking (replication) and builds on the Findings. They offer some support to any future project phase by exploring areas for the future STDP to consider such as energy, land rights and IP handover processes > The use of short Conclusions statements would have helped the clarity of presentation > The conclusions could also be significantly shorter. This is primarily because new information is included (for instance the cost analysis table - un-numbered). No new Findings information should have been presented under the conclusions.
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	> The evaluation lessons respond to the ToR emphasis on learning and map across to the original ToR and IR objectives > There is also a clear relationship with the findings and conclusions and provide addition support to UNICEF > the introduction of the analysis of legal arrangements is an example of where the evaluation lessons go beyond the findings
ii	Lessons are clearly and concisely presented, yet have sufficient detail to be useful for intended audience.	Partially	> The presentation of summary lessons statements is helpful and provide a level of detail that should help utility for future STDP > Adoption of lessons against the project, technology, innovation and replication areas described above may have helped simplify the presentation and minimise duplication
SECTION G:	**RECOMMENDATIONS (weight 15%)**	63%	
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.	Yes	> The recommendations are clearly presented and respond well to the evaluation purpose. They derive and develop the findings and conclusions into concrete actions for future follow-up
ii	Recommendations are useful and actionable for primary intended users and uses (relevant to the intervention); guidance is given for implementation, as appropriate.	Yes	> The recommendations are clearly actionable. They are categorised in a way that can be understood easily relative to the findings and conclusions > While intended users are identified in broad terms, utility could have been helped by being more precise - at the level of specific roles within organisations rather than organisation in general > Guidance for their implementation is not provided, however this is sufficiently covered under the Lessons section preceding the recommendations
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 process for developing the recommendations is not described. It is unclear whether or how duty bearers (or end-users) were included in the development of recommendations. This may pose a barrier to follow-up in cases where implementation is to be the responsibility of IPs or government and thereby significantly beyond UNICEF control
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.	Partially	> While a level of attention has been given to the utility of the recommendations which involve a classification / grouping that is reasonably coherent, they are not prioritised > Required timelines, roles and responsibilities are nevertheless provided
SECTION H:	**REPORT STRUCTURE AND PRESENTATION (weight 5%)**	36%	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.	Partially	Areas included in the evaluation report include: - Name of evaluated object - Table of contents - A list of acronyms/abbreviations - Timeframe of the evaluation - Date of the report - Location of evaluation object - Evaluation company name - UNICEF is named as the commissioner for the evaluation - Three Annexes are listed in the table of contents - Page numbers included Areas of weakness include: - The use of tables, graphs, figures and annexes relevant to the methodology and data used to reach the findings - No list of tables and figures is provided in the table of contents
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.	No	Only three annexes are included: - Technical report - Evaluation case studies - Evaluation matrix - List of documentation - however relevant documents in the list are not always referenced in the text Significant and important annexes that are missing include: - Terms of reference, - List of interviewees (roles, not necessarily names), - UNICEF results framework - The STDP M&E framework and position of IPs within it - A list of site visits - A list of survey or interview questionnaires (as provided indirectly in the evaluation matrix) - Inclusion of indicators and benchmarks within the evaluation matrix
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).	No	> While the broad section headings respond adequately to the ToR, large parts of the report are un-numbered and do not follow a clear structure and formatting isn't easy to follow > In particular the structure of the Findings, conclusions, lessons and recommendations could have followed a clearer structure presented up-front in the introduction and methodology. > The clustering of Findings and EQs without any numbering association to the EQ structure and sub questions in the evaluation matrix and ToR make the report very difficult to navigate > Figure and table references are not provided and are inadequately used to present the data on which the Findings are based
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 report structure follows that outlined in the ToR and UNICEF guidelines
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.	Partially	> At 65 pages the report length is a bit long relative to the evaluation subject. However the required report length is not stipulated in the TOR. > The presentation of the methodology and findings in particular are often quite wordy. This could have been mitigation by closer attention to how the report would be structured in a way to minimise duplication by using the findings to build the evidence base for a set of short conclusions and lessons learnt
ii	Report is easy to understand (written in accessible way for intended audience) and generally free from grammar, spelling and punctuation errors.	Partially	> While the report is largely accessible, it needed a thorough review and edit. Some important areas lack clarity due to poor grammar. For example, the introduction to Section 2 reads: "It is a technical and socio-economic summative evaluation of the technologies and models produced was conducted by comparing their results, outcomes, and their effects on the community to sanitation and environmental conditions before their installation." > Confusion arises from changes in the presentation of purpose and objectives between the TOR, IR and ER. > Areas of duplication should have been identified and addressed across the findings, conclusions and lessons learnt
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.	No	The use of visuals and data presentation tools is a weakness of the evaluation report. The original theory of change cannot be read. There are far too few tables of evidence to support the findings which appear discursive and imprecise. > Good examples of the use of visual tools are provided in the Conclusions (revised ToC and costs analysis)
SECTION I:	**EVALUATION PRINCIPLES (weight 10%)**	43%	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 incorporation of rights-related benchmarks or frameworks is not included
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).	Yes	> The participation and engagement of host and displaced communities is described in the methodology and ethical sections > Appropriate sub-groups were identified for focus group discussions, and the findings explore a level of gender disaggregated analysis of FGDs
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	> The language used is broadly positive and empowering and avoids unnecessary criticism. The evaluation reads as a constructive engagement that extended well during the data collection and analysis processes > At times it would have helped to have been clearer about the factors behind missing data that were within WFP control, for example, by examining the robustness of its M&E framework and attention to disaggregated data collection and analysis across its IPs > No evidence of bias was identified during the EQA process > More attention needed to be given to attending to rights-based analysis and terminology. The technical focus of the STDP project design identified in the Conclusions is to some extent mirrored by the evaluation > Weak data presentation in the evaluation report make it unclear how well data use involved disaggregated analysis
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.	Partially	> This is identified as an area of weakness in the original STDP design (Finding against EQ7), however the analysis is limited > Attention to children's engagement is not included in the methodology. It is unclear whether this may have been a mistake. Children, especially girls and adolescent girls, represent important users of sanitation services > No attention was given to wider aspect of disability and inclusion in order to leave no-one behind.
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)	3	

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	> GEEW is included in the evaluation scope but does not provide sufficient clarity as to the required scope of analysis > There is no clear inclusion of human rights considerations - while this might be limited in a pilot technology project it does require analysis > EQ7 sub-questions did clearly integrate GEEW considerations and confirms to the use of appropriate evaluation criteria.
ii	A gender-responsive methodology, methods and tools, and data analysis techniques are selected.	Partially integrated	- The methodology incorporates GEEW through gender disaggregated FGDs and KIIs and provides a limited analysis of gender-related concerns. > An absence of differences between men and women FGD results is presented. This may be to be expected given the limited gender scope of the analysis in the STDP design but the Evaluation could have explored this further. For example, there is no clear examination of gender preferences surround the use of different technologies during Batch 1 and 2 assessments.
iii	The evaluation Findings, Conclusions and Recommendations reflect a gender analysis.	Partially integrated	> Gender analysis is incorporated in the findings and shows a broadly positive perspective among women and men. Some advantages of improved sanitation for women are explored as well as women's engagement in IP WASH committees and monitoring teams > There is no clear reflection of a gender analysis or role of women in the conclusions and recommendations, nor how the SDTP (given its recognised weaknesses in this area) needs to adopt a more effective gender approach in future > Opportunities to improve the incorporation of gender considerations included adoption of a gender based analysis of women and men's engagement in the future innovation funnel process or role in future replication.

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?