

Evaluation of the Enhancing Education for Forcedly  
Displaced and Host Community Children in a  
Protective, Climate Resilient and Gender-Sensitive  
Approach in Benishangul Gumuz Region, Ethiopia  
(2025-2027)

29 July 2025

Final Inception Report

---

Commissioned by:

United Nations Children's Fund (UNICEF) Ethiopia

Prepared by:

Economic Policy Research Institute (EPRI)



## Disclaimer:

This report is intended solely for the recipient's use and contains confidential and proprietary information. Unauthorised distribution or reproduction is strictly prohibited.

The views and opinions expressed in this report are those of the authors and do not necessarily reflect the official policy or position of the Economic Policy Research Institute or its affiliates.

---

**Authors:** Preksha Golchha, Jose Giarola

---

Title of Evaluation	Enhancing Education for Forcedly Displaced and Host Community Children in a Protective, Climate Resilient and Gender-Sensitive Approach in Benishangul Gumuz Region., Ethiopia
Region	Eastern and Southern Africa Regional Office
Year of Report Completion	2025
Office	UNICEF Ethiopia
Country	Ethiopia
Years Covered	2025 to 2027
Date of Submission	June 29, 2025

---

## Acronyms

Acronym	Full Form
ALP	Accelerated Learning Programme
BG	Benishangul-Gumuz
CCC	Community Care Coalition
CAPI	Computer-Assisted Personal Interviewing
CFM	Community Feedback Mechanism
COVID-19	Coronavirus Disease 2019
DRR	Disaster Risk Reduction
EMIS	Education Management Information System
ERG	Evaluation Reference Group
FDP	Forcibly Displaced Population
FGM	Female Genital Mutilation
FDC	Formal Displacement Camp
GPE	Global Partnership for Education
GPI	Gender Parity Index
HC	Host Community
HML	Humanitarian Mechanism for Learning (or Humanitarian Mechanism for Learning approval – depending on UNICEF's use)
IDP	Internally Displaced Person
IR	Inception Report
KII	Key Informant Interview
LtP	Learning through Play
MEL	Monitoring, Evaluation and Learning
M&E	Monitoring and Evaluation
MDS	The evaluation firm (insert full name if known)


Acronym	Full Form
MHH	Menstrual Health and Hygiene
MoE	Ministry of Education
MoWSA	Ministry of Women and Social Affairs
MHPSS	Mental Health and Psychosocial Support
NFE	Non-Formal Education
NSPC	National Social Protection Council
NSAF	National Social Assistance Framework
PTA	Parent Teacher Association
REB	Regional Education Bureau
SBC	Social and Behaviour Change
SDG	Sustainable Development Goal
SRHR	Sexual and Reproductive Health and Rights
SRGBV	School-Related Gender-Based Violence
ToC	Theory of Change
ToT	Training of Trainers
UNICEF	United Nations Children's Fund
UNGA	United Nations General Assembly
VAC	Violence Against Children
WASH	Water, Sanitation and Hygiene
WEO	Woreda Education Office
WGQ	Washington Group Questions (for disability screening)

## Table of Contents

1.	Introduction to this document.....	8
2.	Context of Operation .....	11
3.	Programme Overview .....	16
3.1.	Geographic and Population Coverage .....	19
3.2.	Implementation Modality.....	21
3.3.	Implementation Timeline .....	21
3.4.	Current Progress (as of Q2 2025).....	22
4.	Theory of Change and Integrated Results Framework .....	22
4.1.	Theory of Change (Narrative) .....	22
4.2.	Integrated Results Framework (Log Frame) .....	28
4.3.	Risks and Assumptions Behind the Theory of Change.....	33
4.3.1.	Operational Risks & Assumptions.....	33
4.3.2.	Contextual Risks & Assumptions .....	34
4.3.3.	Institutional Risks & Assumptions.....	35
4.4.	Interventions Excluded or Partially Covered .....	36
4.5.	Outcomes that are partially analysed.....	38
5.	Evaluation Objectives and Questions .....	39
5.1.	Original Evaluation Objectives and Questions.....	39
5.2.	Revised Evaluation Objectives .....	42
5.3.	Proposed Evaluation Questions.....	43
6.	Overall Evaluation Design and Approach .....	45
6.1.	Data Collection Methods .....	45
6.1.1.	Baseline and Endline Surveys.....	45
6.1.2.	Qualitative Methods .....	48
6.1.3.	Monitoring Data.....	51
6.2.	Data Analysis.....	51
6.2.1.	Effectiveness Analysis .....	51
6.2.2.	Implementation Fidelity Analysis.....	52
7.	Evaluation Plan.....	53
7.1.	Baseline Evaluation (June–Sept 2025).....	53

---

7.1.1.	Baseline Objective.....	53
7.1.2.	Draft Baseline Evaluation Questions .....	53
7.1.3.	Baseline Data Collection Methods & Sample Sizes .....	54
7.1.4.	Ethical Considerations.....	58
7.1.5.	Enumerator Training and Piloting.....	60
7.1.6.	Data Collection.....	60
7.1.7.	Analysis Approach.....	61
7.1.8.	Baseline Reporting Process.....	62
8.	Deliverables.....	65
9.	Risk Mitigation .....	65
10.	Annexures .....	68
10.1.	Terms of Reference.....	68
10.2.	Data Collection Tools .....	68
10.3.	Ethical Clearance.....	69
10.4.	Midline Evaluation .....	69
10.4.1.	Midline Objective.....	69
10.4.2.	Midline Focus Areas.....	70
10.4.3.	Draft Midline Evaluation Questions.....	71
10.4.4.	Midline Data Collection Methods and Sample Sizes .....	71
10.4.5.	Enumerator Refresher and Tool Refinement .....	77
10.4.6.	Data Collection - Fieldwork.....	77
10.4.7.	Data Management and Analysis .....	77
10.4.8.	Midline Reporting Process .....	78
10.5.	Endline Evaluation (June–Sept 2027) .....	80
10.5.1.	Endline Objective .....	80
10.5.2.	Endline Focus Areas .....	81
10.5.3.	Endline Draft Evaluation Questions .....	82
10.5.4.	Endline Data Collection Methods and Sample Sizes.....	84
10.5.5.	Enumerator Training and Preparation.....	91
10.5.6.	Data Management and Analysis .....	91
10.5.7.	Endline Reporting Process .....	92



11.	Consolidated Timeline of Evaluation Activities (2025–2027) .....	97
12.	Ethical Considerations.....	103
13.	Limitations.....	103
14.	Utilisation of Findings .....	103
15.	Annexures .....	104
15.1.	Annex 1: Qualitative and Process Evaluation Tools.....	104
15.2.	Annex 3: Sampling Approach.....	107
15.3.	Annex 4: Implementation Progress and Target Coverage by Component .....	109

---

## 1. Introduction to this document

This document presents the inception report, following an evaluability assessment for the evaluation of the Enhancing Education for Forcedly Displaced and Host Community Children in a Protective, Climate Resilient and Gender-Sensitive Approach in Benishangul Gumuz Region, Ethiopia. The programme is a multi-donor initiative supported by the LEGO Foundation (Foundation), the Danish Ministry of Foreign Affairs (Gvt), and the Grundfos Foundation (Pvt), implemented in partnership with UNICEF Ethiopia, regional government bureaus, and Plan International. Though originally conceived as three distinct interventions, the programme components have since been integrated under a common Theory of Change (ToC) to deliver a coordinated package of education, WASH, and mental health and psychosocial support (MHPSS) interventions in refugee-hosting and crisis-affected communities.

The purpose of this inception report is to establish the groundwork for the evaluation, confirming the scope, design, methods, tools, and timeline. The report outlines the programme context, defines key evaluation questions, and lays out the evaluability assessment, including data availability and methodological feasibility. It outlines a comprehensive evaluation plan encompassing baseline, midline, and endline phases, with a focus on contribution analysis, mixed-methods research, and equity-focused inquiry.

### **Activities undertaken to inform this inception report**

As part of this process, the team conducted key stakeholder consultations with UNICEF programme staff at the National level as well as the Field Office managing the programme, desk review of programme documents, population statistics and data from the Benishangul region of Ethiopia as well as data on target schools on their staffing structure and student composition. The evaluation team conducted seven group consultations with programme staff from the Child Protection, Education, Social Protection, and WASH sections of UNICEF Ethiopia, including at the Field Office level, as well as one consultation with Plan International. The team also engaged in an exploratory meeting and a presentation of potential evaluation methodology with the Donors as part of this process. Furthermore, several email consultations were held with the UNICEF programme team to clarify doubts, agree on the evaluation scope, and finalise the evaluation approach. The UNICEF team also reviewed the data collection tools. The tools were then revised and submitted to the Ethical Review Board at Assossa University for ethical clearance. Ethical clearance has been received and is attached with this inception report.

This inception report is the outcome of an evaluability assessment conducted as part of the inception phase. The Evaluability Assessment provides a solid foundation for shaping a robust and contextually grounded final evaluation. Several key insights and positive developments have emerged that can help sharpen the focus of the evaluation and ensure its strategic

alignment with programme goals and donor priorities.<sup>1</sup> The full evaluability assessment report is provided separately, but a few highlights are presented below:

### **Emerging Issues and Influencing Factors for the Impact Evaluation**

A diverse set of programme documents—including individual and integrated proposals, intervention site maps, and service mapping matrices—has been made available and reviewed. This documentation confirms strong coherence in the programme’s design and delivery modalities, as well as clear articulation of target groups, implementing partners, and geographic coverage. The consolidated intervention site mapping is particularly useful for refining sampling strategies and ensuring the evaluation reflects the geographic and demographic diversity of the intervention areas.

The evaluability assessment also highlighted the need to reconstruct the integrated Theory of Change (ToC) to articulate causal pathways from activities to outcomes and impact fully. The programme has developed a unified, high-level Theory of Change that synthesises the education, MHPSS, and WASH components. This ToC reflects a commendable effort to articulate a shared vision for inclusive, quality education in crisis-affected settings. While the ToC presently emphasises processes, there is strong potential to reconstruct it during the inception phase into a more evaluable form—one that maps clearer impact pathways, makes assumptions and risks explicit, and aligns with measurable intermediate and final outcomes. This reconstructed ToC is presented in the inception report and will provide a more rigorous foundation for contribution analysis and other theory-based evaluation methods.

The assessment concludes that experimental and quasi-experimental designs are not feasible, given full programme coverage and ethical considerations. Full programme coverage across all targeted schools, coupled with ethical obligations not to withhold services from vulnerable populations (Forcibly Displaced Persons and the host community) has made Randomised Controlled Trials (RCTs), stepped-wedge designs, and matched comparison group designs infeasible. No eligible control or phased implementation group exists, and the volatile humanitarian context further complicates the potential identification of counterfactuals. Instead, a mixed-methods contribution analysis—anchored in before/after comparisons, strong qualitative inquiry, and process tracing—offers a credible and useful approach. This is especially suited for capturing intermediate outcomes such as shifts in teacher practices,

---

<sup>1</sup> It is important to note the alignment of the programme with respect to the donor priorities. For instance, according to the 2025 Government’s Priorities for Danish Development Cooperation, Denmark focuses on greater emphasis on Africa; the green transition; humanitarian and nexus approaches. The LEGO Foundation prioritizes investments in three thematic areas: early childhood development, education, and humanitarian relief. Within these, they focus on learning through play, child protection, and supporting children in humanitarian emergencies. Finally, the Grundfos Foundation primarily focus in three thematic areas: Water & Development, Research & Innovation, and Inclusion & Community Engagement.

learner wellbeing, school climate, and community engagement in education and protection.<sup>2</sup> These are valuable signals of early progress toward the programme’s objectives, outputs, and outcomes.

The programme’s focus on forcibly displaced children and families—refugees, IDPs, and host communities—is both urgent and appropriate. These populations face layered vulnerabilities, including trauma, disrupted schooling, and protection risks. The evaluation team appreciates UNICEF’s emphasis on safeguarding, and the phased approach to data collection reflects ethical best practices. Building the final evaluation design around context-sensitive, child-safe, and inclusive methods will not only uphold ethical standards but also improve data quality and community trust. Collecting data from a diverse, vulnerable, and mobile population presents ethical, logistical, and methodological hurdles. The evaluation involves multiple age cohorts (pre-primary to secondary students), forcibly displaced populations (refugees and IDPs), and sensitive domains (psychosocial wellbeing, protection, gender-based violence). This necessitates child-sensitive tools, extensive training for enumerators, and robust safeguarding protocols. Refugee and IDP mobility, incomplete population records, linguistic diversity, and security issues may also affect sample tracking, response rates, and longitudinal consistency across evaluation rounds.

All key stakeholders—including UNICEF programme teams, regional government actors, and implementing partners—have demonstrated high interest in using the evaluation findings for accountability and learning. This creates a conducive environment for the evaluation to be both participatory and actionable. The inception process has already benefited from extensive stakeholder engagement, and this will be continued throughout the evaluation lifecycle to ensure utilisation of findings.

Despite some delays in programme rollout, the timing of baseline data collection offers an ideal entry point to capture pre-intervention conditions. Tools and protocols for monitoring and primary data collection are being strengthened, and UNICEF has indicated willingness to support these efforts, including capacity building for field teams. These investments will support evaluation credibility and support triangulation across data sources.

Building on the above, the evaluation team is prioritising:

- A refined, testable Theory of Change with measurable outcomes.
- Mapping interventions to specific schools and populations to ensure coverage and comparability.
- Clarifying intended pathways of change across components (education, MHPSS, WASH).

---

<sup>2</sup> In parallel with the evaluation, the Examination and Assessment Agency will administer an Early Grade Reading Assessment (EGRA), and an Early Grade Mathematics Assessment (EGMA) under the supervision of UNICEF. The findings will inform and be synchronized with the findings and analysis of the present evaluation.

- Identifying the added value of integrating the interventions across the Gvt, LEGO Foundation, and Grundfos Foundation projects.
- Developing indicators that reflect both outcome-level changes and equity dimensions (e.g. gender, disability, displacement status).
- Sequencing data collection to balance ethical safeguards with data needs, particularly for vulnerable groups.<sup>3</sup>

The evaluation team remains committed to a rigorous, ethical, and participatory process that supports UNICEF and its partners in delivering transformative results for forcibly displaced and host community children.

This document is structured as follows:

1. **Programme Overview and Theory of Change:** Presents an overview of the integrated programme, donor contributions, and the shared results framework.
2. **Evaluation Purpose and Scope:** Clarifies the objectives of the evaluation, its intended users, and the boundaries of the assessment.
3. **Evaluation Design and Methodology:** Describes the evaluation approach, questions, indicators, sampling strategy, and mixed-methods tools.<sup>4</sup>
4. **Evaluation Timeline and Deliverables:** Sets out the key milestones and timeline across the three phases—baseline, midline, and endline—including data collection and reporting schedules; followed by a consolidated timeline.
5. **Ethical Considerations and Limitations:** Discusses safeguarding principles and design constraints.
6. **Annexes:** Includes the detailed evaluation matrix, sampling calculations, draft data collection tools, and other supporting materials.

## 2. Context of Operation

### Socio-economic Context in Benishangul region<sup>5</sup>

The Benishangul-Gumuz (BG) region of Ethiopia is shaped by a fragile socio-political landscape and overlapping humanitarian crises, which significantly impact the wellbeing and educational participation of children, particularly those who are displaced. BG hosts a

---

<sup>3</sup> It is important to highlight that the sequencing of data collection aligns well with the school calendar. Children's data will be collected in September, while adult data will be gathered in August.

<sup>4</sup> The evaluation team conducted an Evaluability Assessment, which involved reviewing the Theory of Change, data systems, and the feasibility of the proposed evaluation design. The findings from this assessment directly informed the Inception Report, particularly the Evaluation Design and Methodology sections..

<sup>5</sup> The data reported in this section are taken from UNICEF program proposals to the three donors as well as UNICEF's communication plan for this project.

significant number of internally displaced persons (IDPs) and refugees. Humanitarian needs for those who are seeking a safe haven after escaping the conflict in their country remain high, with many requiring support to establish their lives. The conflict in Sudan since April 2023 has led to an influx of refugees, and third-country nationals into Benishangul Gumuz, particularly in the Kurmuk-Assosa Zone, where transit centers have been established. This has added pressure on local resources and services.

BG is home to multiple ethnic groups (including Gumuz, Berta, Shinasha, Komo, Amhara, Oromo, and others), and community relations have historically been strained by competition over land and resources. The region hosts a significant refugee population from neighbouring Sudan, as well as many internally displaced persons (IDPs) from local conflicts.<sup>6</sup> By early 2025, BG was hosting over 100,000 refugees (primarily Sudanese who fled conflict in Blue Nile and other areas) and tens of thousands of IDPs. These displaced families often reside alongside host communities in Assosa Zone (e.g. near Sherkole, Bambasi, and Tsore refugee camps) and rely on humanitarian assistance. Social indicators in BG lag behind national averages – access to basic services like education, clean water, and healthcare is limited which is more strained with added service demand by refugee arrivals.

According to UNICEF program documents, approximately 11,000 displaced children and refugees are out of school, out of a broader estimate of 40,000 school-age children out of school in the region. These figures highlight the severity of exclusion in education. The educational system in BG faces compounding challenges: only 29% of primary schools have access to basic water supply, and a mere 17% have access to basic sanitation. Among IDP and refugee camp settings, 40% of sites lack clean water and basic sanitation, with disproportionate effects on girls, especially during menstruation. Literacy and school completion rates are low, and child marriage is prevalent. In fact, BG has one of the highest child marriage rates in Ethiopia (around 50% of girls marry before age 18)<sup>7</sup> contributing to high female dropout rates. Overall, poverty and traditional gender norms heavily shape the social landscape, with women and children (especially girls, refugees, and those with disabilities) being among the most vulnerable groups. Assessments in Assosa (the regional capital) indicate high levels of mental health problems related to forced displacement. Nearly half of respondents reported mental health issues, and a significant portion experienced suicidal ideation.<sup>8</sup> The region faces risks such as family separation, gender-based violence, and limited access to legal identity and justice. Vulnerable groups, including women, children, and ethnic minorities, are particularly at risk.<sup>9</sup>

---

<sup>6</sup><https://www.acaps.org/en/countries/ethiopia>

<sup>7</sup> <https://share-netinternational.org/wp-content/uploads/2024/04/Policy-Brief-Ending-Child-Marriage-In-Ethiopia-Share-Net-Ethiopia.pdf>

1. <sup>8</sup> [https://globalprotectioncluster.org/sites/default/files/2024-05/pau24\\_07\\_protection\\_analysis\\_update\\_ethiopia\\_march\\_2024-final\\_28.04.2024.pdf](https://globalprotectioncluster.org/sites/default/files/2024-05/pau24_07_protection_analysis_update_ethiopia_march_2024-final_28.04.2024.pdf)

<sup>9</sup> [https://globalprotectioncluster.org/sites/default/files/2024-05/pau24\\_07\\_protection\\_analysis\\_update\\_ethiopia\\_march\\_2024-final\\_28.04.2024.pdf](https://globalprotectioncluster.org/sites/default/files/2024-05/pau24_07_protection_analysis_update_ethiopia_march_2024-final_28.04.2024.pdf)

Widespread poverty in BG means many children come to school hungry or not at all. Poor households often cannot afford indirect costs of schooling (clothes, school materials, transportation), pushing children into work or keeping them at home. During lean seasons or economic shocks, families may resort to negative coping strategies that affect children’s education – for example, pulling children (especially boys) out of school to herd livestock or work in gold panning, or marrying off adolescent daughters to ease the economic burden.<sup>10</sup> Chronic malnutrition is also a concern: surveys in 2024 found Global Acute Malnutrition (GAM) rates categorized as “serious” to “medium” in parts of BG.<sup>11</sup> Hunger and undernutrition directly impact students’ ability to concentrate and learn. Moreover, under-resourced schools struggle to provide meals or sufficient learning materials. Poverty-related migration (in search of food or work) contributes to irregular attendance and high dropout rates. Overall, economic hardship is a root cause of low enrolment and poor learning outcomes, as it exacerbates child labor, absenteeism, and early dropout.

Ongoing conflict has directly impacted schools and learners. Attacks and insecurity have led to school closures, damage to facilities, and the displacement of students and teachers. Children from conflict-affected communities often experience psychosocial trauma, loss of family members, and inconsistent schooling due to repeated displacement. In BG, more than 74,000 people remained internally displaced by conflict as of early 2025, in addition to large numbers of returnees trying to rebuild their lives.<sup>12</sup> Many displaced children have fallen behind academically or dropped out altogether. Even host community schools feel the strain – class sizes swell when IDP and refugee children enrol, and tensions can arise between groups. Safety is a major concern: in some areas, families keep children (especially girls) at home for fear of abductions or violence on the way to school. Conflict has also disrupted government oversight and the delivery of supplies (like textbooks) to remote schools. In short, conflict and displacement have caused severe interruptions in education, undermined the learning environment, and created a generation of conflict-affected youth with significant psychosocial needs.

The region is also deeply affected by climate shocks and conflict. Nationally, Ethiopia is the second largest refugee-hosting country in Africa, with over one million refugees, and is home to 4.5 million internally displaced people, driven by both conflict and climate events. In total, 21 million people are in need of humanitarian assistance, including 11 million children, and 7.6 million children (60% girls) are out of school due to emergencies. Climate-related shocks – primarily flooding and drought – have increasingly affected BG, intersecting with the other drivers. Ethiopia is highly susceptible to extreme climatic variability, and BG is no exception.<sup>13</sup>

---

<sup>10</sup> <https://share-netinternational.org/wp-content/uploads/2024/04/Policy-Brief-Ending-Child-Marriage-In-Ethiopia-Share-Net-Ethiopia.pdf>

<sup>11</sup> <https://www.unicef.org/media/166006/file/2025-HAC-Ethiopia.pdf>

<sup>12</sup> [https://globalprotectioncluster.org/sites/default/files/2024-05/pau24\\_07\\_protection\\_analysis\\_update\\_ethiopia\\_march\\_2024-final\\_28.04.2024.pdf](https://globalprotectioncluster.org/sites/default/files/2024-05/pau24_07_protection_analysis_update_ethiopia_march_2024-final_28.04.2024.pdf)

<sup>13</sup> <https://www.unocha.org/publications/report/ethiopia/ethiopia-situation-report-10-jan-2024>

The region's long rainy summer season (Kiremt) often brings floods and landslides. In 2023, unusually heavy rains led to flooding across western Ethiopia, damaging roads, bridges, and some school structures. Floods have temporarily displaced communities (including students) and turned school compounds into shelters for displaced families on occasion. Conversely, shorter dry spells or drought in BG can cause water shortages, impacting both livelihoods and school functioning (e.g. lack of water for students). Climate change is also linked to increased disease outbreaks (malaria, cholera) which keep children out of school.<sup>14</sup> Most schools in BG have very limited infrastructure: many are made of mud or wood and are ill-equipped to withstand heavy storms. They often lack clean water sources, proper toilets, or electricity. These deficiencies create an uncondusive learning environment and are exacerbated by climate stresses (for example, without water points, students must leave school to fetch water, a burden that falls disproportionately on girls). In summary, climate shocks and poor infrastructure disrupt schooling through physical destruction, health impacts, and resource scarcity, and they underscore the need for climate-resilient education facilities in BG.

The intersection of climate vulnerability and education is particularly stark: 83% of out-of-school children live in countries with a Climate Change Risk Index higher than the global median, and crisis-affected children in Ethiopia acquire foundational skills six times slower than their peers due to trauma and interrupted schooling.

Political instability and weak governance structures in BG further exacerbate these challenges. Since 2018, BG has experienced bouts of ethnic conflict and insurgency that have undermined security and governance. The region's peripheral location and historical marginalisation have fueled rebellions by Gumuz militias and other groups. A major spike in violence occurred in 2019–2021 (especially in Metekel and Kamashi zones), involving attacks on villages and ethnic killings. In 2022, the regional government reached peace agreements with two rebel groups – the Gumuz People's Democratic Movement (GPDM) in October 2022 and the Benishangul People's Liberation Movement (BPLM) in December 2022 – leading to a tentative improvement in security. However, peace remains fragile. Sporadic clashes and violence continued in 2023–2025, often along the borders with Oromia and Amhara regions.<sup>15</sup> Notably, in late May 2025, armed attacks by non-state groups in the Oromia–Benishangul border area (e.g. Kamashi Zone) displaced over 11,000 people (around 5,900 from BG alone), with incidents of killings, abductions, and arson reported.<sup>16</sup> Such insecurity has periodically forced local administrations to declare command-posts or restrict movement. Politically, BG is governed by a regional state apparatus under Ethiopia's federal system, but frequent turnover of officials and security operations have strained local governance. Humanitarian access has at times been hampered by the volatile security situation. The presence of Ethiopian military units and regional special forces is still significant in conflict-prone woredas.

---

<sup>14</sup> <https://www.unocha.org/publications/report/ethiopia/ethiopia-situation-report-10-jan-2024>

<sup>15</sup> [https://globalprotectioncluster.org/sites/default/files/2024-05/pau24\\_07\\_protection\\_analysis\\_update\\_ethiopia\\_march\\_2024-final\\_28.04.2024.pdf](https://globalprotectioncluster.org/sites/default/files/2024-05/pau24_07_protection_analysis_update_ethiopia_march_2024-final_28.04.2024.pdf)

<sup>16</sup> <https://www.acaps.org/en/countries/ethiopia>

Overall, while the political context in mid-2025 shows some progress toward peace, continued ethnic tensions and insurgent activity pose serious challenges to stability and the protection of civilians in BG.

The integrated programme, therefore, situates itself as a direct response to these compounding risks—aiming not just to deliver services, but to rebuild protective, inclusive environments where children can regain a sense of normalcy, access learning, and heal from trauma; thus, building resilience.

### **Comparative Analysis of Refugee and IDP-Hosting Regions in Ethiopia**

BG is among Ethiopia’s most heavily impacted regions by displacement-related pressures, but it is not singularly the worst across all indicators. It represents a chronic crisis setting marked by longstanding underdevelopment now compounded by a growing refugee and IDP population, particularly from Sudan. In the education sector, BG has the highest national rate of out-of-school children—approximately 55%, compared to 31% in Tigray and 21% in Amhara. It also hosts 47,463 refugee children, nearly half (49%) of whom are not enrolled in school. In contrast, contested areas such as Tselemti have experienced near-total collapse, with 100% of schools closed for over four years, depriving ~19,000 children of education. While BG’s situation is the most severe in terms of chronic access deficits, contested areas represent more acute and complete service collapse due to war and administrative vacuum.

On child protection, BG has some of the highest documented caseloads among stable regions: 15,667 children with protection concerns, including 1,092 unaccompanied minors and 7,730 separated children. School Related Gender Based Violence (SRGBV) is pervasive, and adolescent girls face elevated risks of child marriage, FGM, and sexual violence. In general, BG has overwhelmed but present protection services, especially in formal refugee camps.

In the WASH sector, BG faces severe chronic shortages. 87% of residents report water is unavailable all day, and 48% of households access only one 20L jerrycan per day—well below humanitarian standards. 298,483 returnees in BG lack clean water at home, and some refugee schools have only 2 latrines for over 800 students.

Refugee and IDP-hosting contexts across Ethiopia show varied severity. For instance, while BG hosts a significant number of Sudanese refugees (over 35,000 new arrivals expected) and high proportions of out-of-school and unprotected children, other regions—especially the contested northern areas—have experienced total institutional collapse due to conflict.

### **Alignment with Government Policies and Political Frameworks**

The integrated programme is well-aligned with national, regional, and sectoral policy frameworks, ensuring coherence with Ethiopia’s broader development priorities and legal commitments. At the national level, the programme contributes directly to the Education Sector Development Plan VI (2020/21–2024/25), Ethiopia Education Transformation Program (2023 to 2027) which both documents emphasise inclusive and equitable access to quality

education, with specific focus on refugees, girls, and children with disabilities. ESDP VI also promotes the integration of refugee students into the national education system and encourages education in emergencies approaches that are contextually adapted for regions like Benishangul-Gumuz.

Built on its commitment to the Global Compact for Refugees, Ethiopia was an early adopter of the Comprehensive Refugee Response Framework (CRRF) CRRF, which promotes a comprehensive approach to refugee situations, including solutions for refugees and support for host communities. The country has generously updated its national refugee laws, including the 2019 Refugee Proclamation, which grants refugees the right to work, access financial services, and benefit from labour protections.

In practice, this is reflected through active collaboration with the Refugees and Returnees Service (RRS) and the inclusion of refugee camp schools (e.g., Sherkole, Tsore, Bambasi) in the integrated programme. This also aligns with the Global Compact on Refugees, reinforcing Ethiopia's leadership in inclusive refugee policy.

In child protection and mental health, the programme aligns with Ethiopia's National Case Management Framework, focusing on training frontline workers and community-based structures in psychosocial support, child safeguarding, and protection from gender-based violence. The LEGO-funded component explicitly promotes Learning Through Play and MHPSS in line with this framework, recognising the long-term developmental impacts of trauma and the need for supportive learning environments. Similarly, all the WASH program implementation for instance is through the Benishangul Regional Water and Health bureaus and is aimed at enhancing the local capacity while ensuring ownership and sustainability. This is in alignment of government policies.

On climate resilience, the programme reflects priorities from the Climate Resilient Green Economy Strategy and national Disaster Risk Reduction (DRR) policies. Schools will be supported to establish DRR clubs, promote environmental awareness, and serve as community platforms for climate-related social mobilisation. These efforts respond to local government recognition of the need to enhance preparedness and social cohesion through education, as explicitly referenced in BG regional strategies.

Regionally, all key regional bureaus—Education, Water, Health, Social Affairs—are involved in co-implementation and monitoring, under a governance structure coordinated by UNICEF. Quarterly review meetings, joint monitoring visits, and planning sessions are structured to ensure multi-sectoral synergy, alignment with regional development plans and to build institutional capacity over time.

### 3. Programme Overview

In the Benishangul-Gumuz region of Ethiopia, a multi-donor integrated initiative is being implemented to address the complex and interlinked challenges facing forcibly displaced and

host community children. This initiative brings together three complementary programmes—funded by the Danish Ministry of Foreign Affairs (MFA), the LEGO Foundation, and the Grundfos Foundation—under a unified framework coordinated by UNICEF Ethiopia.<sup>17</sup> Collectively, the programmes aim to build resilience, promote inclusion, and improve access to quality services through an integrated package that spans education, mental health and psychosocial support (MHPSS), and water, sanitation, and hygiene (WASH) interventions.

The programmes, which initially had separate designs and results frameworks, have now been harmonised into a single Theory of Change and results framework. This integration was developed during the inception phase, in consultation with donor partners (through the Donor Group Plus, or DGpN), regional government authorities, and UNICEF Denmark. The revised framework forms the foundation of a joint Monitoring, Evaluation, Accountability and Learning (MEAL) system, enabling unified planning, learning, and evidence generation.

The programme’s key objectives are to:

- Improve enrolment and learning outcomes at pre-primary, primary, and secondary levels in a climate-resilient, gender-equal, and protective environment.
- Enhance emotional wellbeing and support safe return to school through age-appropriate play-based learning and psychosocial interventions.
- Expand access to inclusive and climate-resilient WASH infrastructure in schools and communities.
- Strengthen community structures to promote positive parenting and inclusive, equitable demand for education.
- Foster multisectoral partnerships that integrate education, health, protection, and WASH services for greater efficiency and impact.

A common Theory of Change (ToC) guides all components, with one overall objective: to improve access to inclusive, equitable, and quality education for refugee, IDP, and host community children, focusing on foundational learning and psychosocial well-being in a gender-responsive, climate-resilient, and protective environment. The three donors’ contributions are complementary: the MFA-funded component emphasises formal education (with gender, climate, and child protection aims), the LEGO Foundation supports play-based learning and MHPSS, and Grundfos funds school WASH infrastructure.

**Component 1 – Education (Danish MFA):** The MFA component focuses on primary and secondary education with strong gender and climate lenses. It supports gender-responsive

---

<sup>17</sup> It is important to note that, in recognition of the huge budget requirement for infrastructure development and as part of this multisectoral programming, the WASH component has secured two complementary grants—one from Grundfos and one from the U.S. Bureau of Population, Refugees, and Migration (BPRM)—to expand reach and ensure wider coverage. The latter, UNICEF has secured in the course of this evaluation, and, therefore, fall out of the scope of this evaluation, which has already started collecting baseline data in selected schools.

pedagogy, inclusive education for displaced and host children, and climate-smart schooling. Key activities include teacher training on gender-sensitive and psychosocial approaches, provision of teaching and learning materials, and establishing clubs or campaigns around gender-based violence prevention and climate awareness. The MFA funding also covers limited education infrastructure, such as constructing new classrooms and providing basic WASH supplies in schools. Additionally, a child protection element under MFA addresses GBV prevention, referral, and case management in schools and communities. These “soft” interventions (training, clubs, community awareness) create a safe and protective learning environment for girls and boys.

**Component 2 – MHPSS & Learning Through Play (LEGO Foundation):** Branded “*Resilient Minds, Playful Hearts*,” the LEGO-funded component targets younger children (pre-primary and primary) with play-based learning and mental health support. It introduces Learning Through Play curricula in classrooms and child-friendly spaces, trains teachers and community facilitators in social-emotional learning and psychosocial support, and provides play materials (e.g. LEGO play boxes) and renovate and furnish classrooms to stimulate foundational skills. It also establishes community-based satellite pre-primary centers to expand access to early childhood education for children within their local communities. MHPSS (Mental Health and Psychosocial Support) is delivered both in schools (through teachers integrating socio-emotional activities) and in the community (through parents’ groups and trained social workers), driven by the need to roll-out evidence informed and globally validated MHPSS models (e.g., TeamUp). Parenting programmes and community playgroups engage caregivers to promote children’s well-being at home. The LEGO component also emphasizes disability inclusion, planning to identify about *1,188 children with disabilities* for assessment and referral, while ensuring all interventions are disability-friendly. This component aims to improve children’s readiness to learn and recover from trauma by fostering play, emotional resilience, and inclusive early learning.

**Component 3 – WASH in Schools (Grundfos Foundation):** The Grundfos-funded component addresses critical water, sanitation and hygiene needs to create a climate-resilient learning environment. This includes construction or rehabilitation of water supply systems (installation of electromechanical equipment, solar-powered pumps, pipelines, water storage reservoirs, water points), installation of gender-segregated latrines and handwashing facilities, establishment of menstrual hygiene management (MHM) safe spaces for girls, conducting MHH awareness creation, MHH skills training, distribution of MHH materials and establishment/strengthening of water management committees/utilities. It also supports hygiene education and community engagement so that improved school WASH influences household practices. As part of the multisectoral programme, UNICEF has secured two complementary grants—one from the Grundfos Foundation and another from the U.S. Bureau of Population, Refugees, and Migration (BPRM)—to expand the reach of the WASH component and ensure broader coverage. However, the scope of this evaluation is limited to

the schools supported through the Grundfos-funded portion of the intervention.<sup>18</sup> It is noteworthy that all schools are expected to benefit from at least some WASH inputs, such as hygiene training or minor facility upgrades. The WASH component aligns with climate adaptation goals by improving school facilities' water reliability and disaster resilience. Considering that most of the water sources are located in the host community, the WASH programming design has utilized a holistic programming approach to ensure that the nearby hosting community and schools also benefits from the integrated WASH programming.

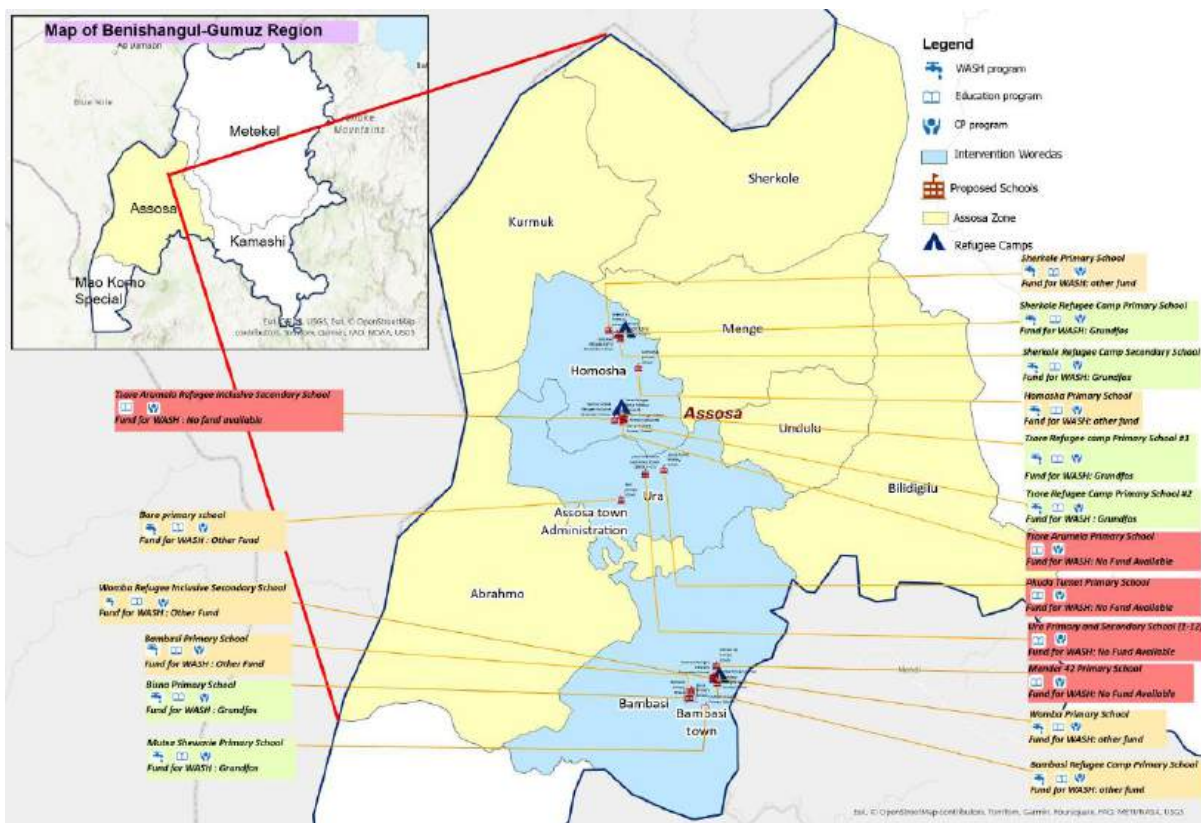
### 3.1. Geographic and Population Coverage

The programme does not cover the entire Benishangul-Gumuz region, but is focused specifically in the Assosa Zone, with implementation sites across the Ura, Homosha, and Bambasi woredas. These sites were selected based on population needs and the presence of refugee camps and underserved host communities. The initiative targets children in refugee camps, internally displaced communities, and surrounding host populations in the Assosa Zone. Implementation is concentrated in the woredas of Ura, Homosha, and Bambasi, across both camp-based and host community schools. Each donor supports a distinct pillar of the programme: the Danish MFA leads on education and protection infrastructure and capacity building of teachers to ensure learning; the LEGO Foundation supports play-based learning and MHPSS; and the Grundfos Foundation strengthens WASH infrastructure and services to enable safe, dignified access to schools—particularly for girls and the neighbouring community. The map below shows the program target schools and refugee camps. These include Tsore Refugee Camp Schools, Sherkole Refugee Camp Schools, Bambasi Refugee Camp Primary School and Host community schools such as Womba and Baro.

Figure 1: Target schools and refugee camps

---

<sup>18</sup> A total of 22 schools were identified in the implementation area, of which 19 are being reached with WASH programming: 11 through Grundfos funding and 8 through BPRM support. One school is being covered by another partner working in the area, while 2 schools remain unreached due to funding constraints.



The list of interventions schools is as follows:<sup>19</sup>

School	Woreda	Schools with Grundfos-funded WASH
Bambasi Refugee Camp Primary School	Bambasi	
Bambasi Primary School	Bambasi	
Womba Refugee Inclusive Secondary School	Bambasi	
Womba Primary School	Bambasi	
Bisna Primary School	Bambasi	
Mutsa Shewarie Primary School	Bambasi	
Tsore Refugee Camp Primary School #1	Homosha	Grundfos
Sherkole Primary School	Homosha	
Sherkole Refugee Camp Secondary School	Homosha	Grundfos
Sherkole Refugee Camp Primary School	Homosha	Grundfos
Homosha primary school	Homosha	Grundfos
Baro primary school	Ura	
Ura Primary and Secondary School (grade 1 - 12)	Ura	Grundfos
Akuda Tumet Primary School	Ura	
Tsore Arumela Refugee Inclusive Secondary School	Homosha	
Tsore Arumela Primary School	Homosha	

<sup>19</sup> UNICEF acquired additional funding from the U.S. Bureau of Population, Refugees, and Migration (BPRM). At this stage, it was not possible to revise the sampling frame. As a result, three of the four selected baseline schools previous pre-selected for in-depth investigation had to be revised. The final baseline data now covers five schools with Grundfos-funded WASH programs

### 3.2. Implementation Modality

UNICEF Ethiopia plays a central coordination and implementation role, working closely with government partners including the Bureaus of Education, Health, Social Affairs, Water, and the Refugee and Returnee Services. Plan International Ethiopia and Bureaus of Water and Health supports on-the-ground implementation, while Assossa University has been engaged to facilitate documentation, capacity-building, and evaluation. A shared governance structure—including a steering committee and quarterly coordination meetings—supports oversight and alignment across the different streams of work.

UNICEF Ethiopia leads execution in partnership with **Plan International** (the main NGO implementing partner on the ground) and with regional government bureaus (Education, Water, Health and Women & Social Affairs). All interventions are delivered as a package in each target community, rather than sequentially or in isolation. Notably, *there is no phased roll-out* – the programme launched activities in all target schools simultaneously, given the urgency and donor expectations.<sup>20</sup> Each school (and its surrounding community) receives the “software” interventions from the start: teacher trainings, MHPSS activities, child protection services, community engagement, etc., occur across all sites in parallel.

Hard infrastructure investments are more site-specific: new classrooms or WASH facilities are constructed in select locations based on needs and available funds. Out-of-school children in these areas are reached through community outreach – e.g. child protection caseworkers and community facilitators establish safe spaces/ satellite pre-primary centers and non-formal learning groups to include children who are not enrolled in the formal schools. Considering that most of the water sources are located in the host community, the WASH programming design has utilized a holistic programming approach to ensure that the nearby hosting community and schools also benefits from the integrated WASH programming.

Interventions are designed to leverage existing systems: for example, social workers from the Bureau of Women and Social Affairs (BOWSA) lead case management with UNICEF support, and there is a growing synergy between the Refugee and Returnees Service (RRS) and Regional Education Bureau to integrate education services. The joint implementation plan is closely coordinated among UNICEF’s sectoral teams to ensure education, WASH, and MHPSS activities reinforce each other on the ground.

### 3.3. Implementation Timeline

Following the joint discussion and engagement by the three donors, the implementation timelines were aligned to end in October 2027: the Danish MFA’s education programme spans 2023 to 2027; the LEGO Foundation’s MHPSS and play programme runs from 2024 to 2027;

---

<sup>20</sup> It is important to highlight that WASH has prioritize hardware activities such as water system rehabilitation or upgrading, latrine construction, and MHH safe space establishment during the initial phase. Software components, including training of water management committees/utilities, MHH awareness, and hygiene education, will follow in the second phase once infrastructure development is completed.

and the Grundfos Foundation’s WASH programme runs from 2024 to 2027. As of mid-2025, the initiative has completed its inception phase and transitioned into active implementation.

### 3.4. Current Progress (as of Q2 2025)

The integrated programme is still in early stages of implementation. Following a successful inception phase and co-creation workshop with donor partners in September 2024, UNICEF began implementation by refining the approved plan and working closely with the Bureau of Education to adapt Mental Health and Psychosocial Support (MHPSS) and Life Skills manuals to the regional context in BG. These efforts supported back-to-school readiness, with a particular emphasis on girls and accelerated learning. Teachers from targeted schools were trained to deliver the adapted Life Skills Manual. Gender Clubs were established across schools to strengthen protection systems and create safe spaces for reporting violence against children (VAC) and gender-based violence (GBV).

In partnership with the Bureau of Women and Social Affairs, UNICEF also established case management systems and provided training on the national framework and CPIMS+ to individuals from both host and internally displaced communities. Social workers were deployed to support the identification and management of child protection cases. During the reporting period, children affected by abuse, neglect, violence, or exploitation were identified and supported. Unaccompanied and separated children were placed in kinship care, with follow-up support provided by social workers to ensure their continued safety and reintegration into school.<sup>21</sup> It is important to note that the evaluation team confirmed that even though these activities are being delivered this will not have significant impact on collecting data for the baseline.

## 4. Theory of Change and Integrated Results Framework

### 4.1. Theory of Change (Narrative)

The overall goal of the integrated program is to improve access to quality education for crisis-affected children (refugees, internally displaced, and host communities) in Benishangul-Gumuz, Ethiopia, while fostering their foundational learning and socio-emotional well-being in a protective, gender-responsive, and climate-resilient environment. This goal aligns with Sustainable Development Goal 4 (inclusive, equitable quality education) and related targets on gender equality and climate resilience. The program is a collaboration between UNICEF and three donors – the Danish Ministry of Foreign Affairs (MFA Denmark), the LEGO Foundation and the Grundfos Foundation – each funding complementary interventions under one unified theory of change. The interventions are delivered in the same schools and communities, enabling a multi-pronged approach to achieve holistic outcomes in education, water/sanitation, child protection and mental health.

---

<sup>21</sup> For detailed information on the specific progresses made on each of these areas, please refer to the Inception Report and the 2024 Annual donor report.

Key target groups include pre-primary, primary and secondary school-age children (with emphasis on vulnerable groups such as refugee and IDP children, girls, and children with disabilities), teachers and school staff, parents/caregivers, and community members. Approximately 38,613 children are expected to be reached, of whom about 45% are refugees, 10% IDPs, and 45% from host communities. At least half of the child beneficiaries are girls, and a small proportion are young children (0–6 years) and children with disabilities. In addition, around 890 teachers (25% pre-primary, 75% primary) and 135 school leaders and officials will receive training, and over 15,000 parents and caregivers will be engaged through community-based activities. These groups are all critical to achieving and sustaining the results.


The integrated results framework is structured as a results chain from inputs/activities to outputs, outcomes, and impact. Financial and human inputs (e.g. donor funding of approximately US\$13 million across the three donors,<sup>22</sup> technical assistance, materials) support a set of activities grouped into key intervention packages. These activities produce outputs (tangible deliverables and immediate results, such as trained teachers, rehabilitated school facilities, established WASH services, and community outreach sessions). The outputs in turn lead to outcomes – the desired changes in behaviours, systems, or conditions, including improved psychosocial well-being, increased educational participation and learning, and strengthened systems for safe, inclusive education. Together, the outcomes contribute to the impact: crisis-affected children “thrive emotionally and academically, leading to increased learning” and progress toward SDG 4 (quality education) as well as SDG 5 (gender equality), SDG 6 (clean water and sanitation), SDG 13 (climate action) and SDG 16 (protective, inclusive societies).

#### *Donor-funded activities vs. complementary interventions*

It is important to distinguish the interventions funded by the three donors from other complementary efforts by UNICEF, government or partners that lie outside the evaluation scope. The donor-funded activities are those described in the results framework below – for example, teacher training, mental health and psychosocial support (MHPSS) in schools, school construction and WASH facility upgrades, community mobilization for education, and evidence-generating studies. These activities are funded by LEGO Foundation, MFA Denmark, and Grundfos Foundation respectively (through a joint funding alliance). In contrast, complementary interventions – such as government provision of teachers and curricula, other UNICEF programs (e.g. an Education Cannot Wait Multi-Year Resilience Programme, or routine child protection services), and other NGO or government WASH projects in the region – are not financed by these three donors. They may still influence the context and ultimate impact (for instance, government school feeding or a parallel Accelerated Learning Program by another agency could affect enrolment and learning), but they are not part of this

---

<sup>22</sup> An additional \$902,975 was secured from the BPRM. Some schools are now funded through this as well. The list of schools and the donors funding the WASH component is presented in Section 3.1.



program's results framework or evaluation scope. The evaluation design will acknowledge such external factors and account for them qualitatively (as contextual factors or baseline conditions) but will assess contributions only of the donor-funded interventions outlined here.

Table 1: Impact and Outcome Level Results Framework

IMPACT	Improved wellbeing, resilience, and learning outcomes of refugee, IDP, and host community children through inclusive, protective, and climate-resilient education systems.								
CONTRIBUTOR TO OUTCOME	Outcome 1: Increased access to inclusive and quality learning for refugee, IDP, and host community children.			Outcome 2: Communities increasingly support safe, inclusive, and equitable education, especially for girls, displaced, and vulnerable children.			Outcome 3: Partnerships, Policy and Evidence Strengthened		
	More children, especially girls, children with disabilities, and displaced learners, attend and benefit from meaningful learning	School attendance and participation improves, especially for out-of-school children and adolescents affected by crisis	Children are more likely to attend and stay in school, especially adolescent girls and children with disabilities	Children face fewer barriers to participation, fewer harmful practices, and more support for sustained school engagement	Vulnerable children receive psychosocial and social protection support that keeps them protected from SRGBV in schools and in the community.	Increased social cohesion reduces bullying, marginalisation, and dropout risk among displaced and minority children	Schools implement risk reduction and safety plans aligned with national DRR frameworks.	Adoption and adaptation of integrated approaches at scale	Integrated, evidence-based education systems are better resourced, more accountable, and more resilient to future crises.

IMPACT									
Improved wellbeing, resilience, and learning outcomes of refugee, IDP, and host community children through inclusive, protective, and climate-resilient education systems.									
OUTCOME	Outcome 1: Increased access to inclusive and quality learning for refugee, IDP, and host community children.			Outcome 2: Communities increasingly support safe, inclusive, and equitable education, especially for girls, displaced, and vulnerable children.			Outcome 3: Partnerships, Policy and Evidence Strengthened		
	MT OUTCOME	Teachers apply new knowledge in classrooms, improving instructional quality and inclusivity	Previously excluded children are re-integrated into education via targeted outreach and support	Schools become physically safer, more accessible, and girl-friendly	Families and communities shift attitudes around girls' education, disability, child protection, and climate resilience	Case management improves; more children identified, referred, and supported	Refugee and IDP children are better accepted in host communities; tensions reduced	Schools implement and maintain inclusive, gender-responsive, and risk-informed safety and disaster preparedness measures in line with national DRR frameworks.	Programme insights shape discourse and influence donor/government investment decisions
ST OUTCOME	Teachers and education leaders capacitated in inclusive, gender-responsive, and climate-sensitive pedagogy	Increased enrolment of girls and boys in formal and non-formal education (FDC and host communities)	Improved WASH and climate-resilient school infrastructure	Parents and community leaders engaged in SBC and positive parenting for inclusion and equity	Social workers and community coalitions capacitated in protection and child well-being	Increased acceptance of inclusive education and refugee integration	School leaders and teachers trained in school safety and DRR protocols and protocols developed and endorsed	Evidence and learning disseminated in regional/global forums	Data and research systems support programme learning and scale-up

IMPACT									
Improved wellbeing, resilience, and learning outcomes of refugee, IDP, and host community children through inclusive, protective, and climate-resilient education systems.									
ACTIVITY	Outcome 1: Increased access to inclusive and quality learning for refugee, IDP, and host community children.			Outcome 2: Communities increasingly support safe, inclusive, and equitable education, especially for girls, displaced, and vulnerable children.			Outcome 3: Partnerships, Policy and Evidence Strengthened		
		Train educators in Learning through Play (LtP), MHPSS, TaRL, inclusive education, climate change, and gender-sensitive methodologies	Support school re-entry, ALP/NFE, enrolment campaigns; provide learning materials; mobilise families to re-enrol displaced and vulnerable children	Rehabilitate/construct gender-segregated sanitation facilities (incl. MHH cubicles, handwashing stations, teachers' toilets)	Conduct structured community conversations and SBC interventions addressing harmful norms (VAC, child marriage, FGM), promoting inclusive education and climate action	Train CCCs and social workers in the National Case Management System (identification, MHPSS, care planning), with attention to climate-induced vulnerabilities	Facilitate peer learning platforms and community dialogues on refugee and disability inclusion, girls' education, and social cohesion	School leaders and head teachers trained on DRR policies and protocols for the school.	Present programme insights in advocacy and learning platforms (e.g. DGpN, GPE, UNGA); contribute to global learning on integrated programming

The framework above demonstrates how MFA Denmark’s focus on gender-responsive, climate-resilient education LEGO Foundation’s focus on learning through play and MHPSS, , and Grundfos Foundation’s focus on WASH in schools jointly contribute to an integrated set of outcomes. Each output is linked to specific donor-funded activities, while also leveraging government and UNICEF baseline programs (e.g. trained teachers deployed by the government, existing community structures) to achieve the results.

#### 4.2. Integrated Results Framework (Log Frame)

The following tables present the integrated results framework in a log frame format, linking inputs/activities to outputs, outcomes, and impact. The table below presents the logframe. The indicators will be disaggregated by gender, and whether host communities, refugee or internally displaced persons.

Level	Result Statement	Indicator	Means of Verification	Assumptions/Risks
Impact	Improved wellbeing, resilience, and learning outcomes of refugee, IDP, and host community children	Proportion of female and male FD and HC learners who report feeling safe, included, and protected (free of SRGBV)	Student surveys, FGDs	Safe school practices in place; students feel secure enough to report
		In-school disaster risk reduction strategy in place and implemented	School DRR plans, implementation monitoring reports, FGDs	Schools have capacity and support to implement DRR plans
		Gender Parity Index (GPI) for primary, middle, and secondary schools in target areas	EMIS, school enrolment data	Families and systems support equal access for girls

Level	Result Statement	Indicator	Means of Verification	Assumptions/Risks
		Gross Enrolment Ratios at pre-primary, primary and secondary in targeted woredas	Enrolment records (school data)	Learning centres are functional and accessible
		Proportion of learners achieving grade 3-level reading and arithmetic proficiency	National learning assessment data - Early Grade Reading Assessment (EGRA), and Early Grade Mathematics Assessment (EGMA) <sup>23</sup>	Children have consistent access to quality instruction, teachers are trained on literacy and numeracy skills and apply them consistently. Learners' classroom attendance is at a minimum of 85%
Outcome 1	Education Access and Quality Improved	% of children previously out of school who are now regularly attending	Attendance tracking, school records (monitoring)  FGDs	Learning environment remains stable, incentives effective

<sup>23</sup> UNICEF is planning to engage with the National Assessment Agency to collect the data in baseline and endline.

Level	Result Statement	Indicator	Means of Verification	Assumptions/Risks
Outcome 2	Community and Social Norms Strengthened	Increased share of community members who believe education is equally important for boys and girls	FGDs	SBC interventions are well-received
Outcome 3	Partnerships, Policy and Evidence Strengthened	# of regional or national education policies influenced by programme learning	Stakeholder interviews	Policy space remains open and responsive to data
Intermediate outcome 1	Teachers are trained and applying inclusive and gender-sensitive pedagogy	% of observed classrooms where teachers apply inclusive methods	Observation tools, lesson plans, coaching logs	Teachers are retained and supported
Intermediate outcome 2	Improved infrastructure and safety measures in schools	# of schools with operational WASH, MHH, and DRR aligned infrastructure and safety measures.	School observation reports	Infrastructure investments are sustained

Level	Result Statement	Indicator	Means of Verification	Assumptions/Risks
Intermediate outcome 3	Communities actively engaged in promoting education	# of SBC sessions held  % community participation in SBC sessions	SBC reports, event logs (monitoring)	Community leaders champion inclusion
Contributory Outcome 1	Social workers and CCCs manage child protection risks	# of cases identified, referred, and resolved  Existence of protection systems, processes and protocols in schools	Case management logs (monitoring)  CPIMS+ reports  FGDs	Case management systems are functional and funded
Contributory Outcome 2	Learners attend school regularly and feel safe	% of students attending >80% of school days  % of students report feeling safe	Student survey , school attendance records	Protection structures active in schools
Contributory Outcome 3	Teachers support learning through inclusive practices	% of teachers demonstrating pedagogical inclusivity	Classroom observation and teacher survey	Training translates into improvement in teacher practises

Level	Result Statement	Indicator	Means of Verification	Assumptions/Risks
Contributory Outcome 4	Schools are gender-responsive	% of schools implementing DRR	Observations, FGDs, KIIs	DRR tools and resources are adapted locally
Contributory Outcome 4	Schools are gender-responsive	% of schools implementing gender plans	Observations, FGDs, KIIs	Resources are adapted locally

x

### 4.3. Risks and Assumptions Behind the Theory of Change

Below is a detailed analysis of the key assumptions and risks for this integrated program, grouped by category, along with their implications for evaluation design and implementation (notably on evaluation scope, cost, and ability to attribute outcomes to the program).

#### 4.3.1. Operational Risks & Assumptions

- **Assumption:** Sufficient coordination and capacity among implementing partners (UNICEF staff, local education and WASH authorities, NGO sub-partners) to deliver the multi-sector activities effectively.
  - **Risk:** If operational coordination fails or local partner capacity is weaker than expected, some outputs (e.g. infrastructure works or training programs) could be delayed or of lower quality. This could lead to uneven implementation across sites.
  - **Evaluation implications:** The evaluation team should be prepared for variability in implementation fidelity across different woredas or schools. Additional process evaluation may be necessary to document the effectiveness of activities on the ground. In terms of scope, evaluators might need to sample sites with both strong and weak implementation to understand outcome differences at endline. If coordination issues cause delays, the timing of endline evaluation might need adjustment (there are cost implications if extending evaluation timeline).
- **Assumption:** Resources and inputs are delivered as planned – e.g. materials procured on time, funds disbursed without major shortfalls.
  - **Risk:** Logistical challenges (procurement delays, inflation, or currency fluctuations) could reduce the quantity or quality of inputs. For example, rising fuel and commodity prices could limit construction materials or the number of dignity kits procured.
  - **Evaluation implications:** If inputs are less than planned, outcomes may not be achieved to the expected degree, complicating attribution (the evaluation would have to note that lower-than-planned scale of the intervention might explain weaker results). If inflation significantly altered what was delivered, evaluators must contextualize outcome findings in terms of “what was implemented” versus “what was planned.”
- **Assumption:** Monitoring data (baseline, targets) are available and reliable. Several of the indicators require monitoring data from the school – e.g. attendance, learning outcomes, implementation and distribution of kits etc.

- **Risk:** Weak data systems could lead to gaps or inaccuracies in monitoring indicators (e.g. school attendance records or well-being survey data might be incomplete).
- **Evaluation implications:** The evaluation design relies on the monitoring data to limit the cost of primary data collection to supplement monitoring data. Relying on primary data collection for these data will significantly raise costs (for surveys, travel).
- **Risk: Multi-donor coordination** – While the three projects are designed under one framework, reporting requirements differ. There is a risk of misalignment in indicators or timelines between donors’ expectations.
  - **Evaluation implications:** The evaluation team has harmonised the results framework and will rely on the common evaluation TOR (to be agreed by all UNICEF program teams and donors) to avoid scope creep. Coordination risks mostly affect program management, but for evaluation it means clear scope definition is needed to isolate the donor-funded components.

#### 4.3.2. Contextual Risks & Assumptions

- **Assumption:** A relatively stable security situation in Benishangul-Gumuz.
  - **Risk:** The region has experienced conflict and inter-ethnic violence in recent years. A resurgence of conflict or new insecurity could disrupt activities – schools might close, communities might flee, or evaluators might be unable to access certain areas. Additionally, a large influx of new refugees (e.g. due to conflicts in neighbouring countries) could strain resources and alter the target population significantly.
  - **Evaluation implications:** Security issues pose scope and cost challenges. The evaluation might have to limit geographic coverage or use remote data collection (phone surveys, local enumerators) for unsafe areas, which can reduce the breadth of evidence. If conflict causes interruptions, outcomes may not be measurable in certain areas (attrition of sample, missing data), introducing potential bias that the evaluators need to acknowledge.
- **Assumption:** No major climate shock or natural disaster beyond the project’s coping capacity.
  - **Risk:** Benishangul-Gumuz is vulnerable to climate-induced events like floods or drought. A severe drought could dry water sources, undermining WASH outputs; a flood could damage school infrastructure; extreme weather might also shift community priorities (e.g. hunger reducing school attendance).

- **Evaluation implications:** A climate shock during the implementation period would act as a confounding factor. The evaluation may need to adjust its counterfactual analysis – for example, if both intervention and comparison areas were hit by a flood, outcomes might dip regardless of the program.
- **Assumption: No overlapping crisis** (e.g. epidemics, nationwide political crisis).
  - **Risk:** Another emergency like a pandemic or national conflict could divert government and UNICEF attention away from this program. Schools might close (as seen in COVID-19), halting the interventions.
  - **Evaluation implications:** Similar to security/climate shocks, a major crisis would force a reevaluation of evaluation scope (perhaps pausing or redesigning the study). Cost could increase due to needed flexibility (extra rounds of data collection, PPE for health safety, etc.). Attribution would be very challenging if outcome changes are dominated by the larger crisis effects rather than the program.

#### 4.3.3. Institutional Risks & Assumptions

- **Risk:** Overlap or interference from other programs. Institutionally, UNICEF and other NGOs might run parallel initiatives in the same region (e.g. Education Cannot Wait or government programs). While complementarity is intended, there's a risk that other interventions (training, supplies, cash transfers, etc.) might intervene. This could enhance results but also make it more challenging to isolate the contribution of these interventions.
- **Evaluation implications:** The evaluation will gather information on other interventions in the area. If some schools received similar support from elsewhere, this will be factored into the analysis.
- **Assumption:** Policy environment remains enabling. For example, the Refugee Education Integration policy (allowing refugees into national schools) needs to continue, and national curricula/policies support the introduced components (like Life Skills curriculum, inclusion of play, etc.).
  - **Risk:** If a policy were reversed or if bureaucratic approvals for new curricula or approaches are not granted, some planned activities (teacher trainings, new content in classrooms) might not fully materialize.
  - **Evaluation implications:** If an external policy decision hampered the program, the evaluation must note this as an external factor. It might narrow the evaluation scope (certain intended outcomes might be dropped from assessment if the activity was not allowed to proceed). However, this is largely

outside project control, so the evaluation would treat it as a limitation, not as a failure of the project design per se.

#### 4.4. Interventions Excluded or Partially Covered

**Exclusion/Partial Coverage of Community WASH Systems:** The Grundfos-funded component includes upgrading water systems that serve both schools *and* surrounding communities.<sup>24</sup> While school WASH outputs (water points on school grounds, school latrines, etc.) are directly linked to educational outcomes (e.g. attendance, girls' comfort), the broader community water supply improvements – such as extended pipelines or communal water points for villages – go beyond the school context. The integrated evaluation will primarily assess WASH improvements as they pertain to schools (e.g. whether target schools have functioning water and toilets, and how that affects students). It will not fully evaluate community-level outcomes like household water access rates, reduction in water-borne diseases in the community, or time saved fetching water; however, these will be discussed in community-level FGDs.

- **Rationale:** The evaluation's purpose is to see how combined interventions influence education and child well-being outcomes for school-age children. Evaluating community WASH would entail a public health evaluation approach – surveying households, measuring water quality, etc. – which is outside the core scope of this education-focused evaluation and would broaden the scope considerably. Moreover, isolating the impact of community water systems on education (indirect pathway) would be methodologically complex without significant cost extension. Therefore, community WASH outputs are considered means to strengthen the school environment, and the evaluation will stop short of surveying general community impact.
- **Implications for Results:** This exclusion means some benefits of the Grundfos contribution might not be fully quantified in the evaluation. For example, if a new solar-powered water pump now provides clean water to 1,000 community members, that health/time-saving impact is not captured, even though it's a significant outcome from Grundfos's perspective. However, any indirect effects on schools (say, healthier children, or improved community support for schools due to water access) might surface qualitatively. The evaluation will mostly reflect Grundfos's impact in terms of schools having water and sanitation, rather than broader community development outcomes.
- **Mitigation Strategies:** The evaluation report will clearly acknowledge these community benefits qualitatively, possibly including a few community informant

---

<sup>24</sup> in recognition of the huge budget requirement for infrastructure development and as part of this multisectoral programming, the WASH component has secured two complementary grants—one from Grundfos and one from the U.S. Bureau of Population, Refugees, and Migration (BPRM)—to expand reach and ensure wider coverage. As a result, in some targeted schools, WASH activities are funded by PRM, while Education and MHPSS components are supported by MFA and LEGO.

interviews to illustrate how the school-community water systems affected people's lives (stories or quotes). Grundfos Foundation may also rely on UNICEF's regular WASH monitoring for those community indicators. Importantly, the evaluation's scope statement will clarify that community-level outcomes were beyond scope, to manage expectations.

**Accelerated Learning Programs (ALPs) for Out-of-School Youth:** Both LEGO and MFA contributions mention accelerated learning for out-of-school children (allowing them to catch up and re-enter formal school). These ALP classes – often held in community or non-formal settings – are a bridge to formal education. The integrated evaluation will track how many previously out-of-school children eventually enrol in formal primary schools (an outcome of interest), which captures the result of ALP. However, the evaluation will not deeply assess the quality or effectiveness of the ALP classes themselves (e.g. literacy gains during the 10-month course, teaching quality in ALP, etc.). In other words, ALP is treated as an output leading to enrolment outcome, rather than a separate object of evaluation.

- *Reason:* Evaluating ALP impact properly would require pre- and post-tests for ALP learners, possibly a control group of similar out-of-school kids not in ALP – essentially a sub-study. This is a significant undertaking and diverges from the main evaluation questions (which focus on integrated outcomes in the school system). Many ALP participants are expected to join the regular schools, so their progress can be indirectly measured by whether they stay and succeed in school afterward. Thus, the evaluation will include them in overall enrolment and retention statistics but won't have a dedicated analysis of ALP pedagogy or curriculum efficacy.
- *Mitigation:* The evaluation's qualitative fieldwork can include a few interviews with ALP instructors or graduates, to capture success stories or challenges, giving a flavour of how ALP contributed to the project. This way, the contribution of ALP is not entirely lost. The evaluation will make it clear that while the ALP was an important strategy to reach the most marginalized, detailed evaluation of ALP outcomes was outside its scope.

**Community-based initiatives and peripheral activities:** A few other interventions from the proposals are only tangentially covered:

- For example, the "Girl Effect – Yegna" multimedia campaign (a radio drama to empower girls) is part of MFA Output 3. The evaluation will note exposure and some effects (perhaps via interviews or girls' focus groups about messaging) but will not attempt to measure the campaign's impact rigorously (since it's mass media and attribution would be difficult in one region alone).
- Likewise, school feeding or nutrition programs and health programs in schools are external but complementary. They can influence attendance and health, thus impacting outcomes. The evaluation will treat these as contextual factors. They are

not funded by the three donors, so they aren't evaluated, but the team will document their presence to explain outcome patterns (e.g. if one district had a feeding program leading to better attendance everywhere, including control schools, that will be noted).

- Any government or UNICEF-funded initiatives running parallel (such as an ongoing Education Cannot Wait program improving schools in nearby areas) will also be described but not evaluated. The aim is to clearly demarcate what changes we attribute to the LEGO/MFA/Grundfos package versus what is outside intervention.

#### 4.5. Outcomes that are partially analysed

- **Learning Outcomes Measurement:** One notable partial gap is the evaluation of academic learning outcomes for students through primary data. Improving learning (literacy, numeracy) is a goal, especially under MFA's learning recovery component. However, as acknowledged in the evaluation design, directly testing children's learning achievements was deemed costly and difficult in this context. Implementing standardized tests across thousands of children in multiple languages with reliability is a major exercise. Instead, the evaluation will use aggregate exam results as a proxy for learning gains. This means that fine-grained improvements in, say, reading fluency or math skills due to the programme's inputs might not be captured with precision. If exam data is incomplete or if many target children are in early grades (where formal exams aren't standardized), the evaluation might not fully reflect learning progress. Thus, the "learning recovery" aspect is only indirectly measured, leaving a partial evidence gap that can be filled through monitoring data on national examination scores. In this evaluation, the focus will be more on enrolment/attendance and psychosocial outcomes, and reported data from schools which are easier to survey, rather than test scores from separately implemented activities. Learning outcomes measurements will take place if the National Statistics Bureau can be brought on board to collect the necessary data.
- **Climate Resilience Outcomes:** Activities like environmental clubs and DRR training (supported by MFA's climate-responsive angle) are relatively small-scale and their outcomes (e.g. school preparedness for disasters) are not straightforward to quantify in a two-year evaluation. The evaluation may record that clubs were formed and, perhaps qualitatively, that students gained knowledge, but it will likely not have specific indicators for "climate resilience" improvements. This is another area that will be mentioned narratively more than measured, due to feasibility – assessing resilience would require long-term observation (beyond project timeline).
- **Breadth vs. Depth Trade-off:** Overall, given finite resources, the evaluation has chosen breadth across sectors but will not probe deeply into each micro-intervention. For instance, disability inclusion efforts (LEGO-supported area) will be noted (numbers of children identified, devices provided), but the evaluation may not fully follow up on each child's educational outcome. These are conscious scoping decisions to keep the

evaluation feasible and provide quality evidence on the areas included in the evaluation.

## 5. Evaluation Objectives and Questions

### 5.1. Original Evaluation Objectives and Questions

UNICEF developed a preliminary Terms of Reference (attached separately) for this evaluation with the aim of estimating the impact of the combined set of interventions and identifying the added value of integrating the interventions across the MFA , LEGO, and Grundfos projects. However, due to the integrated and complex nature of the package of interventions, an evaluability assessment was expected to inform the final evaluation objectives and questions. The preliminary evaluation questions are presented below:

Criteria	Questions	
Relevance	1. How relevant are the integrated protective, climate resilient and gender-sensitive approaches and services to the needs of forcedly displaced children and children living in host communities?	Included at baseline (rephrased)
	2. Have services been informed by needs and situational assessments and fully adapted to meet the needs of different groups and in different target woredas as well as evolving circumstances?	Included (rephrased)
	3. To what extent do beneficiaries feel the programme addresses their needs?	Will be included at endline
Coherence	4. Has the integration of approaches been synergistic? What has been the difference between schools which received all component interventions and those where the full package	Excluded – only two schools excluded from WASH and that will not allow for sufficient sample for comparison, given the number of other interventions to be considered.

Criteria	Questions	
	wasn't implemented (e.g. WASH)?	
	5. Has the project design adequately considered national, regional and UNICEF priorities and integrated these in its design to ensure continuity and ownership?	Included
Effectiveness	6. As activities implemented as per the project plan? When variances occurred, what strategies are employed to mitigate them?	At endline
	7. What has been the increase in the enrolment of students to target schools? (Disaggregated by gender, disability, and vulnerabilities)	Rephrased and included
	8. What is the number of students with access to school based MHPSS and, functional and accessible WASH facilities, and playful learning?	This is a monitoring question, and has been excluded
	9. What percentage of children report feeling safe coming to school and included in various activities?	
	10. How many students have achieved minimum literacy and numeracy outcomes during the execution of the project?	If learning outcomes assessment data is collected by the National Statistics Agency, this will be measured as part of the impact question (currently in discussion with UNICEF on the way forward with this.

Criteria	Questions	
Efficiency	11. To what extent is the programming approach efficient in the achievement of desired results in terms of resource utilization (human, technical, financial) and timely delivery?	Included
Impact	12. To what extent has the program contributed towards impact level change? How and why?	Included as part of the impact/effectiveness question
	13. What is the difference in impact indicator achievement in the different schools and woredas?	Included. This comparison will be provided in the equity component – for all quantitative indicators – diaggregation of data.
	14. Were there external factors that contributed to the changes observed? How are they related to the project? How can they be harnessed for enhanced project success?	Included
Sustainability	15. Are refugee inclusive schools or ALP schools targeting FDC officially absorbed into the regional education system including in financing and resource allocation?	This is not an evaluative question. Monitoring can provide an answer.
	16. Are community and social systems strengthening mechanisms in the project, capacitating the community in skills required to maintain the WASH infrastructure built and other gains made? What maintenance are in place?	Not included. As explained, the evaluation has to trade off breadth and depth of investigation – focusing on the package providing a conducive environment for learning outcomes, the evaluation focuses on building the infrastructure.

Criteria	Questions	
	17. Are community structures actively engaging in the current management of project activities and provided the appropriate tools to continue to do so beyond the project?	Not included in the evaluation scope.
	18. Are coordination mechanisms functioning independently?	Not included in the evaluation scope but systems strengthening interventions are included.
Equity, disability and inclusion	19. Has the project brought about changes in the enrolment rate of girls, FDC, CWD and other children facing vulnerabilities?	Included in the impact question
	20. To what extent has the gender parity index (GPI) changed during the project implementation?	Included in the impact question
	21. What mechanisms are utilised to target these marginalised and vulnerable children?	Not included
	22. Once major barriers in schools (SRGBV, discrimination, lack of inclusive learning materials, community barriers, etc.) are addressed, will these children be able to continue attending school?	Beyond the scope of this evaluation.

## 5.2. Revised Evaluation Objectives

Considering the evaluability assessment<sup>25</sup>, the evaluation's purpose is refined to emphasise learning, contribution, and process insight. The evaluation will document what changes occurred during the programme and explore *how and why* those changes happened, rather

<sup>25</sup> A separate evaluability assessment details the findings.

than attempting to isolate a precise impact estimate. It will seek to validate the programme's Theory of Change – testing whether the expected outcomes and intermediate results materialised and identifying factors that influenced success or shortfalls.

The revised objectives are thus:

(1) **Learning:** Generate lessons on effective strategies for integrated education/MHPSS/WASH programming in humanitarian contexts (what worked well, what didn't, and why).

(2) **Adaptive improvement:** Provide feedback to UNICEF and partners on implementation processes, coordination, and any needed mid-course corrections (though summative, it can inform similar future interventions).

(3) **Accountability (Contribution to Results):** Assess how the programme contributed to observed outcomes in education access, learning environment quality, child well-being, and social cohesion, acknowledging other influencing factors. The evaluation will not make definitive attribution claims. Still, it will present evidence of changes consistent with programme influence based on contribution stories and the strength of the evidence from qualitative methods. It will also address each donor's interests in their focus area, ensuring visibility on results in WASH, MHPSS, etc., within the integrated framework. Ultimately, the evaluation aims to strengthen UNICEF's evidence base on integrated multi-sector programming for refugees and IDPs, rather than to produce a narrow impact coefficient.

### 5.3. Proposed Evaluation Questions

The evaluation will be structured around a set of 6–8 major questions that reflect the above objectives. These questions blend effectiveness, process, and relevance criteria, and are tailored to a contribution-focused inquiry.

**Proposed evaluation questions include:**

1. **Relevance:** *Do the protective, climate-resilient, and gender-sensitive services meet the real needs of displaced children and those in host communities? Have these services been designed using up-to-date needs assessments and adapted for different groups and changing situations in each target area? To what extent do beneficiaries believe the project activities and interventions addressed their needs?*
2. **Effectiveness of Outputs & Outcomes:** *To what extent has the programme achieved its intended outputs and outcomes across education, WASH, and MHPSS components? To what extent can observed changes in the target schools and communities be plausibly linked to the programme interventions? (e.g. increased enrolment/attendance, improved psychosocial well-being, improved access to water/sanitation, etc., as observed from baseline to endline.)*

3. **Process and Fidelity:** *How well was the integrated package implemented as planned?* This looks at implementation fidelity – were activities delivered on schedule and to quality standards (e.g. number of teachers trained, infrastructure completed) – and what challenges were encountered in delivery.
4. **Integrated Approach Value-Add:** *To what extent is the programming approach efficient in the achievement of desired results in terms of resource utilisation (human, technical, financial) and timely delivery? What value did the integrated multi-sector approach provide, compared to a sector-specific approach?* Here the evaluation will explore synergies (for example, did the presence of WASH and MHPSS services enhance educational outcomes or child protection uptake?) and any inefficiencies. Since quantitative comparison is not possible, this will rely on stakeholder perceptions and before/after comparisons, possibly including a retrospective contrast with previous single-sector projects.
5. **Contribution to Observed Changes:** *To what extent can observed changes in the target schools/communities be plausibly linked to the programme interventions? Were there external factors that contributed to the changes observed? How are they related to the project? How can they be harnessed for enhanced project success?* This question embraces contribution analysis – examining evidence for the programme’s role in changes like improved safety perceptions, enrolment, etc., while accounting for other factors (e.g. government policy changes, other aid). It will use qualitative attribution techniques (such as participatory ranking or the QUIP method) to gauge if beneficiaries and local officials attribute improvements to the programme.
6. **Beneficiary Experience and Equity:** *How did different groups (refugee vs host, girls vs boys, younger vs older, children with disabilities) experience the programme, and were any groups left behind?* This question ensures analysis of equity and inclusion. It will seek to understand if the programme was relevant and responsive to each group’s needs – for instance, did refugee girls feel safer and more included? Were there barriers that still hindered IDP children’s participation?
7. **Sustainability and Integration into Systems:** *Does the programme introduce the outcomes and practices likely to continue or be sustained by local systems after programme end?* This covers aspects like local capacity built (teachers’ skills, water management committees), government ownership (integration of refugee education into regular systems), and community buy-in. It also touches on how the multi-partner coordination functioned and if it built a durable model.

*(Additional questions may cover coordination/management, such as: How effectively did UNICEF and partners coordinate the three components at all levels? What were the critical success factors or bottlenecks in the multi-sector partnership?)*

## 6. Overall Evaluation Design and Approach

The evaluation will use a **pre–and post-mixed-methods design in a repeated cross-sectional design**. Quantitative data will be collected before the intervention (baseline), midline and at endline from the same target population, and changes will be analysed descriptively. Qualitative methods will support the effectiveness analysis and the process evaluation. These will run alongside the surveys to interpret the *how and why* of those changes.

### 6.1. Data Collection Methods

In light of ethical imperatives, technical considerations, and programmatic data needs (requisite to collect some baseline data before the end of the school year), the evaluation team has adopted a phased approach to the baseline data collection for the Integrated Programme. This approach is designed to uphold UNICEF’s safeguarding commitments—particularly in relation to engaging with children—while ensuring timely, useful, and ethically sound baseline evidence generation to inform the programme’s implementation and learning agenda.

#### Rationale and Ethical Imperatives

The evaluation team fully aligns with UNICEF’s commitment to “do no harm” and affirms that the wellbeing of children supersedes all operational concerns. Given the risks associated with administering lengthy or cognitively demanding survey tools to young children—especially in fragile contexts—direct engagement with children will only be undertaken after completing the necessary ethical reviews, obtaining guardian consent, and providing enumerators with robust, context-appropriate training. This includes methods for child-sensitive data collection, recognition of distress signals, and adherence to child safeguarding protocols.

##### 6.1.1. Baseline and Endline Surveys

A quantitative survey will be administered in selected intervention schools at baseline (just as activities commence) and repeated at endline (after 24 months of implementation). These surveys will include student surveys (secondary schools) and teacher surveys in primary and secondary schools to measure selected *education outcomes* (enrolment status, attendance, study time), *child psychosocial well-being questionnaires* (for example, SDQ or other context-appropriate tools to gauge mental health outcomes), and *knowledge/behaviour questions* on WASH and protection (e.g. handwashing practices, safety perception in school). Learning outcomes (literacy/numeracy) is tougher to measure and given the scope of data collection, it may be difficult to extend the surveys to include metrics. However, aggregate school level examination results may serve as a proxy. The sample will cover a representative subset of beneficiaries – for instance, a random sample of students in upper primary and secondary for outcomes like attendance and learning.

### Sampling strategy for student surveys<sup>26</sup>

The evaluation will use a stratified random sampling of respondents with each of the four programme-supported secondary schools serving as a primary sampling unit. The evaluation will draw a sample of roughly 80-90 students per school, resulting in a total sample size of 320-360 students, statistically significant at 5% with a 5% margin of error. To estimate progress with greater precision (4% margin of error), the sample size will increase to 522 students. Our proposed sample balances statistical rigor with feasibility and resource constraints.

Within each school, a simple random sample of students will be drawn using class rosters, with stratification by grade and gender where possible. Sampling will aim for a balanced gender distribution (e.g., 50% girls, 50% boys) and representation across all four grades. Substitute students will be identified in advance to mitigate the effects of non-response due to absenteeism or refusals. To ensure methodological rigour and equitable representation across participant groups, the evaluation team will adopt a gender-stratified random sampling approach for both student and teacher surveys. This approach enables the random selection of participants while actively maintaining gender balance—a critical consideration in assessing inclusive outcomes related to gender equity, participation, and wellbeing.

For the survey population, a comprehensive sampling frame will first be established. In the case of students, this will typically be drawn from official enrolment registers, attendance records, or class rosters provided by school administrators. Where such lists are incomplete or unavailable, efforts will be made to construct reliable sampling frames through collaboration with school management committees and education officers.

Once the sampling frame is established, participants will be stratified by gender, separating male and female participants into distinct strata. In cases where learners identify outside the binary, their inclusion will be handled sensitively and in consultation with UNICEF safeguarding protocols, ensuring both privacy and voluntary participation.

Within each gender stratum, a random selection process will be applied. This will include the use of simple random sampling methods such as random number generators. In contexts with gender-skewed student populations, the evaluation team may either oversample the underrepresented group to ensure adequate comparative power or adjust the gender distribution proportionally based on population characteristics, ensuring such decisions are transparently documented.

Selected participants must meet basic inclusion criteria (e.g., currently enrolled in the school or employed as a teacher) and must provide informed assent (students), with guardian consent obtained for children under the age of 18. In cases where selected individuals are

---

<sup>26</sup> Detailed estimation techniques are presented in the annex.

unavailable or decline to participate, replacement participants will be randomly selected from the same gender stratum to maintain both the randomisation integrity and the gender balance of the sample.

Field enumerators will be trained on the ethical dimensions of sampling and gender-sensitive engagement, and wherever appropriate—particularly for interviews on sensitive topics—same-gender enumerators will be assigned to ensure comfort, cultural appropriateness, and data quality.

The survey will be administered at baseline, midline, and endline using repeated cross-sections, rather than panels, given the high likelihood of student turnover and grade progression. Within each school, student respondents will be sampled from relevant grades will be surveyed.

#### [Sampling strategy for teacher surveys](#)

The evaluation will use a stratified random sampling of respondents with all 16 programme-supported primary and secondary schools serving as the primary sampling units. The evaluation will draw a sample of roughly 15-18 teachers per school, resulting in a total sample size of 255 (192 primary and 63 secondary teachers), statistically significant at 5% with a 5% margin of error. Our proposed sample balances statistical rigor with feasibility and resource constraints.

Within each school, a simple random sample of teachers will be drawn using rosters, with stratification by grade where possible. Sampling will aim for a balanced gender distribution where the composition of teachers in the school permits. Substitute students will be identified in advance to mitigate the effects of non-response due to absenteeism or refusals. Where fewer female or male teachers are available than required to achieve 50%, all teachers in that group will be included in the surveys. The evaluation team will adopt a gender-stratified random sampling approach for teacher surveys. This approach enables the random selection of participants while actively maintaining gender balance—a critical consideration in assessing inclusive outcomes related to gender equity, participation, and wellbeing.

For teachers, a comprehensive sampling frame will first be established. A school-level staff lists will serve as the primary frame. Where such lists are incomplete or unavailable, efforts will be made to construct reliable sampling frames through collaboration with school management committees and education officers.

Once the sampling frame is established, participants will be stratified by gender, separating male and female participants into distinct strata. In cases where teachers identify outside the binary, their inclusion will be handled sensitively and in consultation with UNICEF safeguarding protocols, ensuring both privacy and voluntary participation.

Within each gender stratum, a random selection process will be applied using a simple random sampling method such as random number generators. In contexts with low female teacher representation, the evaluation team may either oversample the under-represented group to ensure adequate comparative power or adjust the gender distribution proportionally based on population characteristics, ensuring such decisions are transparently documented.

Selected participants must meet basic inclusion criteria (e.g., currently enrolled in the school or employed as a teacher) and must provide informed consent (teachers) or assent (students), with guardian consent obtained for children under the age of 18. In cases where selected individuals are unavailable or decline to participate, replacement participants will be randomly selected from the same gender stratum to maintain both the randomisation integrity and the gender balance of the sample.

### 6.1.2. Qualitative Methods

A robust qualitative component will delve deeper into why changes did or did not occur and how the integrated approach functioned. **Qualitative methods – including narrative techniques like the MSC approach** – will run alongside the surveys to interpret *how* and *why* those observed changes took place. This mixed-method design balances measurement of what changed with in-depth inquiry into the change processes and drivers, which is crucial for evaluating an integrated programme. By triangulating survey results with MSC stories and other qualitative evidence, the evaluation will build a nuanced understanding of the programme's effects and the extent to which observed changes can be credibly linked to UNICEF's multi-component intervention.

This will include **Focus Group Discussions (FGDs)** with various participant groups: students (separately by gender/age), parents/caregivers, and teachers. FGDs will explore perceptions of the programme's effects (e.g. "Do you feel school is safer or more child-friendly now? If so, what made it so?"), changes in attitudes (towards play, toward girls' education, etc.), and any persisting challenges. As part of each FGD, an MSC storytelling exercise will be incorporated: participants will be invited to recount *"Since this programme started, what is the most significant change you have experienced (or observed) in your child's schooling or community, and why is it important to you?"* Through this exercise, beneficiaries themselves identify the changes they value most since the programme began and discuss what or who contributed to those changes. This MSC technique is especially valuable in an integrated programme context – participants' stories can reveal how multiple interventions (education, WASH, protection/MHPSS) may have collectively produced a meaningful change. The team will document these MSC stories and later analyse them to understand common themes and causal pathways, which will enrich the contribution analysis by illustrating *how and why* key changes happened from the beneficiaries' perspective. Focus group discussions (FGDs) will be held separately for teachers, students (segregated by gender), and parents/caregivers

(segregated by status), using participatory tools such as ranking exercises, timelines, and perception mapping. These discussions will explore how services were experienced, perceived quality and accessibility, and any challenges encountered. FGDs will be facilitated by trained moderators in local languages and will last 60 to 90 minutes.

**Key Informant Interviews (KIIs)** will be conducted with stakeholders such as school principals, social workers, Plan International field staff, local government officials (Education Bureau, Water Bureau, etc.), and community leaders. KIIs will gather insights on implementation fidelity, coordination, and context (for example, how did government involvement affect the programme, what external events influenced results). The qualitative sample will be purposive to ensure all site types (refugee camp schools vs host schools, well-performing vs struggling sites) are represented. Interviews will be semi-structured following an interview guide (see Annex 2 for a summary of qualitative tools).

KIIs will gather insights on implementation quality, coordination among the three components, and contextual factors influencing results. KII respondents will be asked about any significant changes they have observed and their views on what influenced those changes – complementing the community perspectives captured through MSC. For the implementation fidelity exercise, semi-structured key informant interviews (KIIs) will be conducted with UNICEF programme officers, implementing partner staff (including Plan International), local government representatives, school leadership, social workers, and WASH technicians. These interviews will explore timelines, delivery processes, coordination practices, adaptations, and feedback from the field. Each interview will last approximately 45 to 60 minutes.

**Direct observations** will be conducted at each school site using structured checklists. These observations will assess the status and use of WASH facilities (construction, accessibility, functionality), the learning environment (use of play materials, safety features), and the extent to which integrated delivery is visible (e.g., coordinated messaging, sectoral co-location). Field teams will document infrastructure quality, use patterns, and upkeep.

#### [Sampling Strategy for qualitative methods](#)

A purposive sample of 4 programme schools (2 primary and 2 secondary) will be selected to ensure diversity across several key dimensions. The sampling strategy is purposive, meaning participants are selected based on their relevance to specific programme activities and pathways in the Theory of Change, rather than for statistical representativeness. These include geographic context (covering refugee camps, host communities, and IDP settings), the presence or absence of completed WASH infrastructure, variation in implementation status (i.e., early versus delayed roll-out), and differences in student population characteristics, such as the proportion of refugee learners or gender parity. This stratification will help the evaluation capture contextual differences in implementation across the region. Due to a lack

of clarity on the implementation timeline across the programme schools, the team have the following primary and secondary schools for qualitative data collection, as the highest number of interventions are planned for these schools with representation from all three camp sites - Womba Primary School,<sup>27</sup> Bisna Primary School, Sherkole Refugee Camp Secondary School and Baro primary school. The full list of schools and number of interventions is presented below.

TOTAL STUDENTS	School	Woreda	Number of interventions
4777	Bambasi Refugee Camp Primary School	Bambasi	15
2794	Bambasi Primary School	Bambasi	16
1160	Womba Refugee Inclusive Secondary School	Bambasi	14
650	<b>Womba Primary School</b>	<b>Bambasi</b>	<b>16</b>
995	<b>Bisna Primary School</b>	<b>Bambasi</b>	<b>17</b>
307	Mutsa Shewarie Primary School	Bambasi	17
7407	Tsore Refugee Camp Primary School #1	Homosha	13
1192	Sherkole Primary School	Homosha	16
853	<b>Sherkole Refugee Camp Secondary School</b>	<b>Homosha</b>	<b>15</b>
4211	Sherkole Refugee Camp Primary School	Homosha	16
1128	Homosha primary school	Homosha	16
808	<b>Baro primary school</b>	<b>Ura</b>	<b>16</b>
413	Ura Primary and Secondary School (grade 1 - 12)	Ura	13
815	Akuda Tumet Primary School	Ura	14
1160	Tsore Arumela Refugee Inclusive Secondary School	Homosha	13
567	Tsore Arumela Primary School	Homosha	15

From each selected site, a range of participants will be engaged using appropriate qualitative methods. These will include 1 to 2 principals or head teachers (via key informant interviews), 4 to 5 teachers involved in education or MHPSS components (via focus group discussions), and 1 to 2 community facilitators or social workers (also via interviews). For students aged 12 to 17, focus group discussions will be conducted in gender-segregated groups of 6 to 8 participants. Similarly, 6 to 8 parents or caregivers per site will participate in FGDs, grouped by their displacement or community status (host, refugee, IDP). Adolescent girls aged 12 to 17 will engage in participatory FGDs focusing on menstrual hygiene management and safety. Additional key informant interviews will be conducted with 6 to 8 local government officials (from education, WASH, and women's affairs offices) and approximately 12 UNICEF and implementing partner field staff. The composition varies by baseline, midline and endline, and

<sup>27</sup> All primary schools also have pre-primary schools.

is detailed in the respective evaluation plan. The summary table presented at the beginning of this section provides the numbers of each stakeholder for each type of data collection tool per school.

### 6.1.3. Monitoring Data

In the first six months of implementation, the evaluators will gather monitoring data from the programme implementation team to document what was implemented when and how each month. Following that, this data will be collected in the form of reports and monitoring databases as maintained by UNICEF during midline and endline data collection. This includes reviewing Plan International’s progress reports, UNICEF’s output monitoring (e.g. number of teachers trained, infrastructure constructed). A timeline of interventions (by school) will be compiled to correlate inputs with outcomes. This process data is crucial to explain outcome findings – for example, if a school saw little change in attendance, was it because some inputs (e.g. a promised water point) were delayed?

## 6.2. Data Analysis

### 6.2.1. Effectiveness Analysis

#### Contribution Analysis Framework

All evidence will be synthesized using a contribution analysis approach. We will begin by articulating the programme’s results chain and assumptions (from the ToC) and then systematically examine the extent to which each link in the chain occurred, using both quantitative indicators and qualitative findings. For outcomes like “improved attendance” or “children feel safer,” we will look for the presence of expected intermediate changes (e.g. new WASH facilities built, psychosocial activities delivered) and gather stakeholder confirmation that these were important. We will also investigate alternative explanations for observed changes (for instance, a new government policy or another NGO project) and assess their influence relative to the programme. The end result will be a reasoned judgment of the programme’s contribution to each outcome area, with evidence tracing the narrative from activities to outcomes (see Annex 1 – Evaluation Matrix, which integrates this logic).

Quantitative data will be analysed using **before-after comparisons** – primarily paired statistical tests or difference in proportions over time within the sample. We will use disaggregation to see if outcomes differed by sub-groups (e.g. did host community children’s enrolment increase more than refugees’?). While causality cannot be confirmed, a significant improvement across most sites in girls’ attendance would be taken as evidence supporting the programme’s contribution, especially if triangulated with qualitative feedback like “the new latrines made it easier for girls to attend regularly.”

Qualitative data from FGDs/KIIs will be transcribed and coded to identify common themes and outliers. We will use qualitative techniques, **including participatory ranking and the Most Significant Change (MSC)** storytelling approach. These methods will gauge whether

beneficiaries and local officials attribute improvements to the programme and identify which changes they perceive as most significant. In particular, integrating the MSC technique allows beneficiaries to define the most important outcomes and explain how and why those changes occurred, providing rich, beneficiary-defined evidence of the programme's contribution within the broader integrated approach. Any contradictions (e.g., if surveys show improvement but some participants say nothing changed) will also be explored and explained. By using mixed methods, the evaluation aims to build a compelling evidence base for the programme's results narrative, even without experimental rigor.

### 6.2.2. Implementation Fidelity Analysis

The analytical framework for assessing implementation fidelity draws on established implementation science principles and consists of four core dimensions: adherence, dosage, quality, and adaptation. Adherence refers to the extent to which interventions are delivered as designed, while dosage assesses the frequency, duration, and coverage of activities relative to what was planned. Quality examines the skill, consistency, and contextual appropriateness with which services are delivered. Adaptation considers whether deviations from the original design occurred, and if so, whether they enhanced or undermined the intended results.

In the context of the BG programme, the implementation fidelity framework is particularly relevant given the integrated and multisectoral nature of the interventions. The dimension of adherence will be assessed by comparing what was delivered on the ground with what was planned in the design documents—for example, teacher training sessions under the education component, delivery of play-based learning activities under the LEGO-supported stream, construction and usability of WASH facilities, and provision of community-based MHPSS and child protection services. The dosage dimension will focus on the intensity and reach of each intervention. This includes whether target groups—especially girls, refugee children, and children with disabilities—received the intended number and frequency of services such as sessions, materials, or visits. For quality, the evaluation will examine whether services were delivered with consistency, appropriateness, and skill, drawing on direct observations, user feedback, and interviews with teachers, students, and caregivers. Finally, the adaptation component will capture any changes made during implementation, exploring whether deviations from the original design were necessary responses to contextual barriers, such as access, staffing limitations, or infrastructure delays, and whether these changes enhanced the programme's relevance or undermined fidelity to its core design. This framework application will allow the evaluation to capture a nuanced picture of how well the programme is functioning in practice and where mid-course corrections may be required. This external study will serve as a critical early review to strengthen implementation quality, course-correct weaknesses, and reinforce the integrated delivery model for improved outcomes by programme end.

## 7. Evaluation Plan

### 7.1. Baseline Evaluation (June–Sept 2025)

#### 7.1.1. Baseline Objective

The baseline evaluation's primary objective is to establish a comprehensive pre-intervention snapshot of the programme's context and key indicators across education, WASH, and MHPSS components. This phase will document initial conditions and needs in target schools/communities and provide reference values for all outcome indicators before full implementation. It will also validate the programme's theory of change assumptions against on-the-ground realities (ensuring a "clean" baseline since no significant services have been delivered prior). The baseline findings will inform programme implementation by identifying critical gaps and setting a benchmark for measuring change at midline and endline.

#### 7.1.2. Draft Baseline Evaluation Questions

1. **Baseline Status of Key Outcomes:** *What is the current status (baseline values) of the programme's intended outcomes and outputs in target schools and communities? – e.g. current enrolment and attendance rates, existing WASH access (water availability, functional latrines), and children's psychosocial well-being indicators before the intervention.*
2. **Needs, Barriers, and Context Relevance:** *What are the main education, protection, and WASH needs and barriers faced by children (refugee, IDP, host) and schools at the start of the programme? How relevant are the programme's planned interventions to these needs? (This explores whether the programme is addressing the most pressing gaps identified by communities and stakeholders at baseline.)*
3. **Baseline Disparities and Inclusion:** *How do baseline conditions differ across various groups and locations?*
4. **Initial Implementation Readiness:** *What is the status of programme roll-out at baseline and are systems in place for implementation? (While full implementation may not have started, this question documents any early activities or preparatory steps by June 2025 – such as training of some teachers or initial community meetings – and assesses the readiness of schools and partners to absorb the upcoming interventions.)*
5. **Ownership:** *Has the project design adequately considered national, regional and UNICEF priorities and integrated these in its design to ensure continuity and ownership?*

*(These baseline questions are aligned with the broader evaluation questions and theory of change but framed to gather preliminary data rather than assess performance. They lay the foundation for answering effectiveness and process questions in later phases.)*

### 7.1.3. Baseline Data Collection Methods & Sample Sizes

Four schools have been selected for in-depth data collection. The selection criteria involved shortlisting schools which had Grundfos-funded WASH components. From those, both secondary schools were selected - Sherkole Refugee Camp Secondary School and Ura Primary and Secondary School (grade 1 - 12). Of the three primary schools, Homosha primary school and Tsore Refugee Camp Primary School #1 were selected to get a larger number of program locations. The table below presents a snapshot of all baseline data collection activities. Details on each data collection method and sampling is provided in the sub-sections that follow.

PHASE	Method	Target Group	Participants per Group	Sampling Strategy	Estimated Total Sample Size	Modality
2	Student Surveys (Secondary)	Secondary school students (boys and girls)	85-90 per school	Clustered (4 secondary schools)	352	Enumerator-led
1	Teacher Surveys (Primary)	Primary school teachers	15-18 per school	Clustered (12 primary schools)	192	Self-led
1	Teacher Surveys (Secondary)	Secondary school teachers	15-20 per school	Clustered (4 secondary schools)	63	Self-led
1	Teacher Survey (Pre-primary)			Clustered (12 Pre-primary schools)		Self-led
2	FGDs - Refugee Parents	Parents of refugee children	1 FGD per school	Purposive (refugee parents across 2 primary and 2 secondary schools)	4	Facilitated
2		Parents of refugee children	1 FGD per school	Purposive (refugee parents across 2 pre-primary schools)	2	Facilitated
2	FGDs - Host Community Parents	Parents from host communities	1 FGD per school	Purposive (host parents across 2 primary and 2 secondary schools)	4	Facilitated
2		Parents from host communities	1 FGD per school	Purposive (host parents across 2 pre-primary schools)	2	Facilitated
2	FGDs - IDP Parents	Parents of IDP children	1 FGD per school	Purposive (IDP parents across 2 primary and 2 secondary schools)	4	Facilitated

PHASE	Method	Target Group	Participants per Group	Sampling Strategy	Estimated Total Sample Size	Modality
<b>2</b>		Parents of IDP children	1 FGD per school	Purposive (IDP parents across 2 pre-primary schools)	2	Facilitated
<b>1</b>	FGDs - Teachers	Primary or secondary teachers	1 per school	Purposive (across across 2 primary and 2 secondary schools)	4	Facilitated
<b>1</b>		Pre-primary teachers	1 per school	Purposive (across 2 pre-primary schools)	2	Facilitated
<b>1</b>	FGDs - Community Leaders	Local leaders, community elders, SMC/PTA	2 FGD per Woreda	Purposive	6	Facilitated
<b>1</b>	FGDs / KIIs - UNICEF/IP Staff	Programme field implementers	1 FGD per Woreda	Purposive	3	Facilitated
<b>1</b>	FGDs - Social Workers	Frontline social workers/community facilitators	1 FGD per Woreda	Purposive	3	Facilitated
<b>2</b>	Case Study - Children with Disabilities	Children with disabilities + caregiver	2-3 case studies	Purposive (identified through schools)	3	Facilitated
<b>1</b>	KIIs - School Principals	Head teachers of selected schools	1 per school	Purposive (complement to survey)	4	Facilitated
<b>1</b>	KIIs - Community Leaders	Influential elders, religious or camp leaders	4-6 total	Purposive	6	Facilitated
<b>1</b>	KIIs - local government		6-8 total	Purposive	8	Facilitated
<b>1</b>	Observation checklist (School level)	Primary and secondary	12 primary schools + 4	All	16	Enumerator-led

PHASE	Method	Target Group	Participants per Group	Sampling Strategy	Estimated Total Sample Size	Modality
			secondary schools			
<b>2</b>	Classroom observations	Primary and Pre-primary school	4 pre-primary schools (2 classrooms each)	purposive	8	Observation by facilitator

#### 7.1.4. Ethical Considerations

The evaluation team recognises that safeguarding and ethics are not ancillary concerns but foundational principles in conducting responsible evaluation, particularly in humanitarian and displacement-affected settings such as Benishangul-Gumuz. The team will adopt a “do no harm” approach throughout, guided by MDS’s Safeguarding Policy and aligned with UNICEF’s Ethical Guidelines for research and evaluation, the Ethical Research Involving Children (ERIC) guidance, and UNEG principles. Our detailed ethical protocols and guidelines to be followed by us and our partners.

**Core Ethical Framework:** Our ethical model is grounded in three interdependent commitments:

1. **Anticipate** potential harms—whether physical, emotional, or social—arising from participation in the evaluation;
2. **Mitigate** these risks through preventive measures, inclusive methods, and tailored protocols; and
3. **Address** any harms swiftly, through robust, survivor-centred reporting, referral, and redress systems.

**Duty of Care:** We hold a dual duty of care—to our evaluation team and to all stakeholders participating in the baseline and evaluation:

- **To the team:** All team members will undergo risk assessment prior to field deployment. They will be trained on safeguarding, business ethics, and, where relevant, hostile environment protocols. A risk register will be maintained throughout, and appropriate duty-of-care procedures will be in place to prevent or manage operational risks.
- **To stakeholders:** Ethical reciprocity underpins our engagement, especially with vulnerable populations. We will ensure participants—including children and marginalised groups—are treated with dignity and agency. Reciprocity will be contextually defined in collaboration with participants to maximise mutual benefit and minimise burden or extraction.

**Alignment with Ethical Standards:** The evaluation will fully comply with UNICEF’s ethical and safeguarding standards. Table X below outlines how the evaluation operationalises the UNEG principles:

UNEG Principle	Implementation in the Evaluation
Integrity	The team will operate independently from programme implementation actors and will maintain objectivity. Any conflicts of interest will be declared. Multiple viewpoints will be presented transparently if findings diverge.

Accountability	Stakeholders will be fully informed of the evaluation approach through consultations (e.g., with the ERG). A safeguarding referral protocol will be activated immediately upon any reported violation. Regular updates will be shared with the Evaluation Manager.
Respect	Gender-sensitive methods will be used, including same-gender enumerators where appropriate. Separate interviews will be conducted with women and men. Tools will be tailored for children’s age and comprehension.
Beneficence	All voices—including those of vulnerable groups—will be reflected in findings. An updated risk register will be used during each data collection round to ensure ongoing adherence to “do no harm” principles.

**Ethical Engagement with Children:** Given that children are primary stakeholders in the integrated programme, they will be included in the evaluation using child-friendly, rights-based approaches. Ethical protocols for child engagement include:

- **Risk and Benefit Assessment:** Risks of child participation will be assessed before each data collection round. Inaccessible or insecure contexts may warrant excluding children to protect their safety.
- **Privacy and Confidentiality:** All interviews with children will ensure privacy without compromising child protection. Identities will be anonymised in all data and reporting.
- **Informed Consent and Assent:** Consent will be obtained from caregivers, and assent from children, with a minimum two-day window prior to data collection. Children will be informed of their rights, including the right to withdraw at any point.
- **Compensation:** Participation will not involve payment. However, basic refreshments and breaks will be provided. Interviews will be limited to a maximum of 1.5 hours to minimise disruption.

These safeguards will also apply, with appropriate adjustments, to adult respondents, particularly those who are vulnerable or from marginalised groups.

**Team Composition and Training:** All evaluation team members will be trained on research ethics, child protection, and cultural sensitivity. They will be required to sign MDS’s safeguarding code of conduct. A gender-balanced team will facilitate culturally appropriate engagements, especially when discussing sensitive topics.

**Data Protection and Ethical Approval:** All personal data will be anonymised, encrypted, and securely stored. Data access will be limited to authorised personnel, and all data will be deleted from MDS systems upon submission to UNICEF. The evaluation will be reviewed by the UNICEF Ethical Review Board, with additional national-level ethical clearance sought as required.

### 7.1.5. Enumerator Training and Piloting

In June 2025, the evaluation team will pilot test all baseline instruments in a non-sampled school or community near Assosa. This piloting (conducted in collaboration with Assosa University) checks question clarity, cultural appropriateness, and logistical flow. Enumerators and facilitators (many of whom were recruited via Assosa University’s network of trained graduates) will undergo a thorough training workshop before data collection. The pre-testing of the survey tool has been completed, and the pre-test report is attached. The training covers:

- Survey administration techniques (sampling within schools, standardising interview protocols).
- Ethical considerations when interviewing children and vulnerable populations (in line with UNICEF procedures on consent, confidentiality, and child protection referrals in case sensitive issues arise).
- Role-play exercises and practice with the tools in the local language to ensure enumerators can comfortably handle both quantitative surveys and FGD facilitation. This close partnership with Assosa University not only builds local capacity (using faculty or students as enumerators) but also grounds the evaluation in local cultural context. Adjustments from the pilot (e.g. rephrasing ambiguous questions) will be finalized by early July.

### 7.1.6. Data Collection

Phase 1 data collection will take place before schools close in July 2025. Phase 2 data collection will take place in September 2025, once schools reopen.

Phase	Timing	Focus	Responsible Activities
<b>Phase 1: Institutional and Adult Respondents</b>	<b>Late June – July 2025</b>	Collection of data for baseline indicators that do not involve direct interaction with children. This includes school infrastructure, WASH conditions, teacher and school leadership capacities, and administrative data from school and woreda officials.	Tool finalisation and piloting; data collection by trained enumerators; initial data cleaning and analysis.
<b>Inter-Phase Preparation</b>	<b>August 2025</b>	Tool refinement, enumerator training, and securing ethical clearance	Evaluation Team Leader to travel to Ethiopia to lead in-person enumerator training

		from the HML National Review Board for child engagement.	and tool finalisation, incorporating lessons from Phase 1. Preparation of revised tools and ethics submission. Capacity-building sessions to be conducted for Assosa-based partners.
<b>Phase 2: Child-Centred Data Collection</b>	<b>Early September 2025</b>	Data collection from children on key outcome and experience indicators, including learning outcomes, perceptions of safety and inclusion, and school participation. Tools will be simplified and child-friendly.	Supervised fieldwork; real-time quality assurance; strict adherence to safeguarding and child protection

### 7.1.7. Analysis Approach

Baseline analysis will be largely **descriptive** and diagnostic:

- *Quantitative:* The evaluation team will compute baseline values for all key indicators (mean or frequency as appropriate). This includes, for example, current gross enrollment rates in target schools, the percentage of students who feel safe at school (based on survey responses), the average number of school days attended in the last month, teacher-to-student ratios, the functionality of existing water points (yes/no), and other relevant factors. Disaggregation will be done by gender, refugee/host status, and other relevant categories to highlight any baseline inequities. No impact inference is made at this stage; instead, these metrics set the benchmark for subsequent comparisons.
- *Qualitative:* FGD and KII transcripts will be coded by theme (using the evaluation questions and focus areas as a guide – e.g. themes like “barriers to girls’ education”, “perceptions of WASH needs”, “community expectations”). The analysis will identify common patterns as well as unique insights or divergent opinions. For example, parents’ FGDs might reveal *safety of girls on the way to school* as a major concern; teacher discussions might highlight *lack of psychosocial training*. These findings will contextualize the quantitative data and help explain *why* certain baseline conditions exist (e.g. cultural factors behind enrolment gaps). They will also inform any refinement of the theory of change or implementation priorities.

- *Triangulation*: The team will triangulate across data sources – comparing survey findings with qualitative narratives and with any existing secondary data (such as EMIS statistics or prior assessments in the region). If discrepancies arise (for instance, if school records show higher enrolment than our survey found attending), these will be noted and, if possible, clarified through follow-up queries or acknowledging as limitations. The inception report’s evaluability assessment emphasised triangulation to strengthen validity in this pre-post design.

#### 7.1.8. Baseline Reporting Process

As data analysis concludes in late August, the evaluators will synthesise findings into a baseline report. A preliminary results briefing may be held with UNICEF Ethiopia and partners to validate findings and gather initial feedback (this could be an informal presentation of key statistics and insights, ensuring no major surprises in the report). Stakeholder validation at this stage helps ensure the baseline interpretations are accurate and buy-in is built for using the data.

#### Outputs and Reporting


- **Baseline Evaluation Report – Draft:** A comprehensive draft report will be delivered in **October 2025**. This report will detail the baseline findings across all focus areas, with data visualisations (tables, graphs of baseline indicators) and qualitative summaries. It will not draw conclusions about programme effect (since interventions just started) but will include *observations and recommendations* relevant to programme planning (for example, highlighting which needs are most urgent or if any aspects of the intervention need acceleration based on baseline gaps).
- **Baseline Evaluation Report – Final:** After incorporating feedback from UNICEF, donors, and implementing partners (who will review the draft), the final baseline report will be finalised. This final report will be a reference document for all stakeholders, serving as the officially agreed-upon baseline values for later comparison.
- **Baseline Presentation:** A presentation (PowerPoint slide deck) summarizing key baseline findings and recommendations will be prepared for a stakeholder dissemination meeting. The evaluation team will present the baseline results to UNICEF, donor representatives (Danish MFA, LEGO Foundation, Grundfos), implementing partners (e.g. Plan International), and regional government officials. The presentation will focus on actionable insights (e.g. “Areas of concern identified at baseline”) so that the programme team can immediately use this information for course-setting.

#### Baseline Evaluation Report Outline

- **Introduction:** Background on the integrated programme (objectives, components) and purpose of the baseline evaluation. This section will link the baseline’s role to the

overall evaluation framework and spell out the baseline evaluation questions and scope.

- **Methodology:** Description of the baseline design and methods – sampling approach, tools used (surveys, FGDs, KIIs), data collection process (with dates and any limitations encountered), and analysis methods. Ethical considerations and limitations of the baseline (e.g. no control group, security constraints affecting data collection) will also be noted.
- **Findings:** Organized by major theme or sector for clarity:
  - *Education Outcomes (Baseline):* e.g. enrolment, attendance, gender parity, any learning assessment results, current teaching practices and school climate.
  - *MHPSS and Play-based Learning:* e.g. baseline psychosocial well-being of children, availability of play activities, community attitudes towards play and mental health support.
  - *WASH in Schools:* e.g. baseline status of water supply (functionality, reliability), sanitation facilities (number and sex-segregation of latrines, hygiene practices), and any related findings on menstrual hygiene management and climate resilience of infrastructure.
  - *Protection and Inclusion:* e.g. baseline incidence (or perception) of violence or bullying in schools, existence of child protection mechanisms, inclusion of children with disabilities or marginalized groups at baseline, and community perspectives on safety.
  - Each sub-section will present quantitative indicators (with baseline values) and qualitative insights side by side, to paint a complete picture for that domain.
- **Discussion:** An interpretation of what the baseline findings mean for the programme. This will revisit each baseline evaluation question:
  - Relevance of the programme design given the needs identified (e.g. confirming that the programme’s focus areas align with what communities see as priorities, or highlighting any misalignments).
  - Key disparities to address (ensuring equity considerations from baseline are explicit).
  - Any immediate recommendations for implementation drawn from the baseline (for example, if the baseline shows very low knowledge of hygiene among students, the team might recommend prioritising hygiene education early on).
  - This section will also highlight any contextual factors observed that might affect implementation (e.g. if baseline found that some schools are in very



remote areas with access issues, this could affect delivery and is noted for programme management).

- **Conclusion:** A brief summary of the baseline situation and a forward-looking statement on how the programme will use this data. It will emphasise that subsequent evaluations (midline, endline) will measure changes from this baseline.
- **Annexes:** Detailed data tables of baseline indicators, the evaluation matrix linking baseline data to future questions, copies of data collection instruments, and any further technical details (e.g. sampling calculations).

## 8. Deliverables

Deliverables	Deadline
Inception Report	10 July 2025
Phase 1 Data Collection Preliminary Findings PPT	20 August 2025
Phase 2 Data Collection Preliminary Findings PPT	30 September 2025
Draft Baseline Report	20 October 2025
Final Baseline Report	5 November 2025

## 9. Risk Mitigation

Risk Category	Description of Risk	Likelihood	Impact	Mitigation Measures
Ethical Risks (Children)	Emotional distress or fatigue among children due to inappropriate tools or poor facilitation.	Medium	High	Use child-friendly, brief tools; train enumerators in child safeguarding; secure ethical clearance.
Sampling and Access	Inability to reach target participants due to absenteeism, school closures, or displacement.	High	High	Coordinate with schools and local officials in advance; oversample in planning; use replacements.
Security and Instability	Heightened risk of data collection disruption in conflict-affected or fragile areas.	Medium	High	Conduct risk assessments prior to deployment; avoid high-risk zones; partner with local actors.
Enumerator Capacity	Enumerators may lack the necessary training to engage ethically and sensitively with children.	Medium	Medium	Deliver intensive in-person training with practical roleplays and child engagement protocols.

Risk Category	Description of Risk	Likelihood	Impact	Mitigation Measures
Language and Translation	Misinterpretation of questions or concepts due to dialect/language mismatch.	Medium	Medium	Hire local-language speakers; pre-test tools in multiple dialects; review translations carefully.
Data Quality (Quantitative)	Measurement errors, non-responses, or social desirability bias affecting validity.	Medium	Medium	Pilot surveys; build rapport; include quality checks and daily field team debriefings.
Data Quality (Qualitative)	Power dynamics and cultural norms may suppress honest views in FGDs or KIIs.	Medium	Medium	Use skilled facilitators; separate FGDs by age/gender; triangulate with other methods.
Timing and School Calendar	Misalignment of data collection phases with academic calendar or seasonal migration.	High	Medium	Align with school calendar; phase data collection to avoid exam periods and migration cycles.
Stakeholder Engagement	Limited cooperation or buy-in from local government, schools, or community members.	Medium	Medium	Engage REBs, WEOs, and school leadership early; explain value of the evaluation.
Data Storage and Privacy	Breaches in confidentiality or unauthorised data access affecting participant safety.	Low	High	Encrypt data, anonymise identifiers, restrict access to authorised team members only.
Inclusion and Representation	Exclusion of specific subgroups (e.g. minorities, children with disabilities) may bias results and cause community tensions.	Medium	High	Map all vulnerable groups during sampling design; ensure intentional inclusion in survey and qualitative tools.
Ethical Clearance Delays	Delays in securing required ethical and administrative clearances from national or local authorities.	Low	High	Prepare ethical protocols in advance; engage with authorities early; use UNICEF facilitation for clearances.
Stakeholder Review Delays	Slow review of the inception report and evaluation tools by stakeholders delays start of fieldwork.	Medium	Medium	Agree on timelines and non-objection protocols with stakeholders; engage ERG early.

Risk Category	Description of Risk	Likelihood	Impact	Mitigation Measures
Programme Staff Turnover	Organisational changes limit access to institutional memory or knowledgeable staff.	High	Medium	Use programme documentation to fill gaps; triangulate with newer staff and administrative records.
Cultural Sensitivity (Difference Blindness)	Lack of awareness of local customs, norms or sensitivities among external staff causes distrust or offence.	Low	Medium	Deploy local enumerators; sensitise external team members; provide cross-cultural orientation.
Informed Consent in Children with Disabilities	Some children may not be cognitively or legally capable of providing informed consent independently.	Low	High	Obtain guardian and child assent; use community workers for children with disabilities.
Confidentiality for Vulnerable Populations	Sensitive information from vulnerable children may lead to identification or unintended harm even with anonymisation.	Low	High	Use neutral pseudonyms and aggregate reporting; assess harm risk per case study before publication.
Pandemic-Related Disruptions	New waves of COVID-19 or other infectious diseases may interrupt fieldwork.	Low	Medium	Plan for remote/phone-based data collection; follow public health protocols; stagger fieldwork if needed.

## 10. Annexures

### 10.1. Terms of Reference

Attached separately.

### 10.2. Data Collection Tools

These tools have also been attached separately.



UNICEFEthiopiaTeach UNICEFEthiopiaKilwit UNICEFEthiopiaKilwit UNICEFEthiopiaKIIHe UNICEFEthiopiaFGDS UNICEFEthiopiaFGDR

erSurvey20240612aPhIPs20240608aPGFIN hGovt20240608aPGF adTeacher20240608aecondaryTeachers202efugeeParents202506



UNICEFEthiopiaFGDP UNICEFEthiopiaFGDI UNICEFEthiopiaFGDHSchoolStructuredObs

primaryTeachers20250DPParents20250606zostParents20250606ervation20250602aP



### 10.3. Ethical Clearance



በኢትዮጵያ ፌዴራላዊ ዲሞክራሲያዊ ሪፐብሊክ  
ትምህርት ሚኒስቴር  
አሶላ ዩኒቨርሲቲ  
Federal Democratic Republic of Ethiopia  
Ministry of Education  
Assosa University

Ref.No/ቁጥር:- ASU/IRB/031/19  
Date/ቀን:- 07/10/2017 EC

↖ To Whom It May Concern

Subject: Ethical Clearance Approval

Title: *Baseline Evaluation for the Integrated MFA-LEGO-Grundfos Funded Project for FDPs and HCCs*

Dear Sir/Mr/s/Dr

Assosa University Institutional Research Ethics Review Board (IRB) has reviewed the aforementioned research proposal with special emphasis on the following points

- 2. Are all ethical principles considered?
  - 2.1. Respect for person Yes  No
  - 2.2. Beneficence Yes  No
  - 2.3. Justice Yes  No
- 2., Are the objectives of the study ethically achievable?
  - 1.2. Are/ is the methods ethically sound? Yes  No

Finally, we would like to take this opportunity to request your good office to maintain the highest ethical standards in the execution of the program and to monitor the ethical implementation of the research as stipulated in the proposal document.



With Best Regards  
*በአሙኒካ ነጋሪ*  
Aminu Lekasa Nagari  
Chairman, IRB

CC

- ✓ To R/T/T/V/President Office
- ✓ To Research Director
- ✓ To Dr. Abdumuhsin Hassen

☎:- +251-057-775-0692  
Fax:- +251-057-775-0752

መልስ ሰጪት ላይ የእኛን ቁጥር ይጥቀሱ  
In Replying Please Quote our Ref. No.

Website:- www.asu.edu.et  
☎ 18

y.3

### 10.4. Midline Evaluation

#### 10.4.1. Midline Objective

The midline evaluation aims to assess interim progress and implementation quality about halfway through the programme, providing an opportunity for course correction and learning. The objectives at midline are two-fold: (1) Effectiveness (Early Outcomes): Determine what initial changes or trends are observable approximately 6–9 months into implementation, compared to baseline, in key outcome areas (education, WASH, MHPSS). (2) Process Evaluation: Examine how well the programme has been implemented thus far – focusing on fidelity to the plan, the functionality of the integrated approach, and any challenges that have arisen. In essence, midline serves as a “health check” for the programme, verifying if it is on

track to achieve its goals and providing actionable feedback to managers for the remaining implementation period.

#### 10.4.2. Midline Focus Areas

- **Early Outcome Changes:** Identify any measurable improvements or changes since baseline in the programme’s target outcomes. For example, have student enrolment or attendance rates started to improve in intervention schools? Are children participating in play-based learning and showing signs of improved psychosocial well-being (even if modest)? Are new water points functional and increasing water availability at schools? Midline will focus on detecting these early outcome “signals” even if ultimate targets are not yet met.
- **Implementation Fidelity and Quality:** Assess the extent to which activities have been delivered as planned across all components. This includes coverage (e.g. number of schools that have received teacher training or WASH facilities by Jan 2026 versus the plan), timeliness (whether activities like training, club formation, construction happened on schedule), and quality (stakeholder perceptions of how well these were done). This focus area addresses whether the programme is being implemented *according to design* and identifies any gaps (for instance, if certain schools or groups have not been reached as expected).
- **Integrated Approach in Practice:** Investigate how the integration of education, WASH, and MHPSS interventions is working on the ground. Are there signs of synergy (such as teachers noting that new WASH facilities improve attendance or that psychosocial training makes their teaching more effective)? How are the different components coordinating at the school and community level? The midline will pay special attention to the “value-add” of the multi-sector approach at this stage – capturing any qualitative evidence of positive interaction between components, or conversely any operational inefficiencies or coordination issues that need adjusting.
- **Beneficiary Feedback and Equity (Mid-course):** Gather feedback from various beneficiary groups on their experience with the programme so far. Do refugee children and host community children feel differences in benefit at this stage? Are girls able to participate equally (e.g. are the menstrual hygiene facilities making a difference for adolescent girls’ attendance)? The midline will check whether any group is being left behind in the initial roll-out and whether adjustments are needed to ensure inclusiveness in the second half of the programme.
- **Emerging Outcomes vs. External Factors:** Consider any external influences or context changes since baseline that might be affecting results. For example, has there been an influx of new IDPs, a change in government policy, or other programmes entering the area that could influence education and WASH outcomes? Midline will contextualize early outcome findings by noting such factors, which is part of preparing for a contribution analysis at endline (identifying other contributors to change early on).

#### 10.4.3. Draft Midline Evaluation Questions

1. **Progress Toward Outputs and Outcomes:** *What progress has the programme made towards its intended outputs and outcomes in the first year of implementation? Which planned activities have been completed (e.g. number of teachers trained, WASH facilities built, play materials distributed), and what early changes are evident in the schools/communities as a result (such as changes in attendance, engagement in classrooms, or hygiene practices)?*
2. **Implementation Fidelity and Challenges:** *To what extent has implementation followed the plan up to midline, and what challenges or deviations have occurred? Are all components (education, MHPSS, WASH) reaching the intended schools and participants on schedule? What operational bottlenecks (e.g. delays in procurement, security issues, staffing or coordination problems) have been encountered, and how have these affected delivery?*
3. **Integrated Delivery and Coordination:** *How effectively are the different components of the programme working together at the mid-point? Do field staff, schools, and communities observe any added value from the integrated approach (such as improved outcomes or efficiency), or are there areas where integration is weak? (For instance, are WASH hardware installations being complemented by hygiene education from the education component? Are psychosocial support activities reinforcing educational participation?)*
4. **Interim Outcomes for Different Groups:** *How are various beneficiary groups experiencing the programme so far, and are initial benefits equitable? For example, are refugee children seeing improvements in their learning environment similar to host-community children? Are girls benefitting (in terms of safety, participation) as much as boys? This question checks for any mid-course inequities or unintended exclusion.*
5. **Likelihood of Outcome Achievement:** *Based on the midline status, is the programme on track to achieve its end-of-programme outcomes? What do stakeholders (teachers, parents, officials) perceive about the momentum of change – do they expect improvements to continue or accelerate? This forward-looking question allows identification of any adjustments needed: if some outcomes lag at midline, what strategies could improve progress?*

#### 10.4.4. Midline Data Collection Methods and Sample Sizes

The midline will employ a lighter-touch mixed-method approach than baseline, with a mix of repeat quantitative measures (to track changes in key indicators) and deep qualitative inquiry (to understand process and early outcomes):

- **Quantitative Follow-up Surveys:** A subset of the baseline survey instruments will be repeated at midline (Jan 2026) to capture trend data. Specifically:

- *Student and Teacher Surveys:* The same schools and, where possible, the same respondents (students and teachers) from baseline will be targeted for a follow-up survey. The surveys will include core “baseline indicators” again—such as attendance rates, perceptions of safety, etc. – to measure change. Since only ~6 months have passed since baseline, large outcome shifts may not yet occur, but these data will indicate direction of change (e.g. slight uptick in attendance or none at all). Repeating the surveys also helps maintain engagement and provides additional data points for the eventual time-series analysis.
- *Head Teacher/Principal Survey:* Again, each school’s head teacher will provide updated information on school metrics and implementation status (e.g. current enrolment vs baseline, facilities changes, any incidents or significant events). This functions as a progress inventory at each school.
- **Qualitative Methods – FGDs and KIIs:** The qualitative component at midline is crucial for process evaluation:
  - *FGDs with Students:* By midline, students in secondary schools will have experienced some programme activities (e.g. life-skills clubs, new play materials, WASH improvements). We will conduct a few FGDs with student groups (segregated by gender – one FGD of girls and one of boys in a couple of secondary schools. These discussions will explore how they perceive changes in their school environment: Are they noticing differences in teaching methods? Do they feel safer or more supported? What do they think of the new play activities or facilities? Content focus is on “early progress pathways and implementation fidelity”, meaning we want to hear if the programme’s presence is being felt and how.
  - *FGDs with Parents:* Similar FGDs as baseline (refugee, host, IDP parents) will be held, but now the topic is their *perception of changes*: Do parents observe their children benefiting (are children more eager to attend school, do they talk about new things at school)? Also, parents can comment on community-level changes (for instance, any improvements in community attitudes or inclusion).
  - *FGDs with Teachers:* A couple of FGDs with teachers (maybe one with primary teachers, one with secondary teachers across the sample schools) will delve into their experience implementing the programme: Have they received the training and materials promised? Are they using new techniques (learning-through-play, psychosocial support) and what has been the result? This also covers how teachers feel about the integration – e.g. “Has having better water at school or play materials made teaching easier?” – addressing integrated approach value-add qualitatively.

- Each FGD at midline will also probe fidelity issues: for example, in a teachers' FGD we might ask if all planned training happened and was it on time, or in a parents' FGD, whether certain activities (like community sessions) took place as they were supposed to.
- *Key Informant Interviews:*
  - **Implementing Partners/UNICEF Field Staff:** Interviews with programme managers (UNICEF, Plan International, etc.) in Benishangul-Gumuz will gather an internal perspective on implementation progress. Topics include coordination among the three components, management of any delays or issues, and reflections on what is working well or not.
  - **Local Officials/School Management:** Interviews with officials from the education bureau or school cluster supervisors to get an external perspective on the programme at mid-point – do they see improvements in schools? How is government involvement (are they participating in monitoring, providing support)?
  - **Donor Perspective (if available):** A brief check-in KII with a donor representative or UNICEF management might be included to understand expectations and any strategic adjustments by midline.
- *Direct Observations:* Where feasible, the midline will include field observations of programme activities in action. For example, if a teacher training or a school club session coincides with the midline visit, the evaluators will observe it using a checklist (noting participant engagement, quality of facilitation, integration of themes). Similarly, evaluators may visit a sample of WASH facilities constructed to observe their usage and maintenance. These observations enrich the process evaluation by providing firsthand evidence of fidelity and quality.

## Sample Size

Method	Target Group	Participants per Group	Sampling Strategy	Estimated Total Sample Size	Content type
Student Surveys (Secondary)	Secondary school students (boys and girls)	85-90 per school	Clustered (4 secondary schools)	352	Midline indicators
Teacher Surveys (Primary)	Primary school teachers	15-18 per school	Clustered (12 primary schools)	192	Midline indicators
Teacher Surveys (Secondary)	Secondary school teachers	15-20 per school	Clustered (4 secondary schools)	63	Midline indicators
Head Teacher Surveys	Principals of all primary and secondary schools	1 FGD per school	4 primary + 4 secondary schools	8	Midline indicators
FGDs - Secondary School Boys	Boys in secondary schools	1 FGD per school	Purposive by gender/grade (4 secondary schools)	4	Early progress pathways + implementation fidelity
FGDs - Secondary School Girls	Girls in secondary schools	1 FGD per school	Purposive by gender/grade (4 secondary schools)	4	Early progress pathways + implementation fidelity
FGDs - Refugee Parents	Parents of refugee children	1 FGD per school	Purposive (refugee parents across 2 primary and 2 secondary schools)	4	Early progress pathways + implementation fidelity
FGDs - Host Community Parents	Parents from host communities	1 FGD per school	Purposive (host parents across 2 primary and 2 secondary schools)	4	Early progress pathways + implementation fidelity
FGDs - IDP Parents	Parents of IDP children	1 FGD per school	Purposive (IDP parents across 2 primary and 2 secondary schools)	4	Early progress pathways + implementation fidelity
FGDs - Teachers	Primary or secondary teachers	1 FGD per school	Purposive (across 2 primary and 2 secondary schools)	4	Early progress pathways + implementation fidelity

Method	Target Group	Participants per Group	Sampling Strategy	Estimated Total Sample Size	Content type
FGDs - Community Leaders	Local leaders, community elders, SMC/PTA	1 FGD per Woreda	Purposive	6	Early progress pathways + implementation fidelity
FGDs - UNICEF/IP Staff	Programme field implementers	1 FGD per Woreda	Purposive	6	Early progress pathways + implementation fidelity
FGDs - Social Workers	Frontline social workers/community facilitators	1 FGD per Woreda	Purposive	6	Early progress pathways + implementation fidelity
FGDs - Host Community Members	Non-beneficiary residents of host areas	1 FGD per Woreda	Purposive	3	Early progress pathways + implementation fidelity
FGDs - Local Govt Officials	Education, Water, Social Affairs officials	1 per Woreda	Purposive	3	Implementation fidelity
Case Study - Children with Disabilities	Children with disabilities + caregiver	2-3 case studies	Purposive (identified through schools)	3	Implementation fidelity
KIIs - School Principals	Head teachers of selected schools	1 per school	Purposive (complement to survey)	4	Implementation fidelity
KIIs - Plan International Staff	Field-level implementing staff	2-3 per Woreda	Purposive	8	Implementation fidelity
KIIs - Community Leaders	Influential elders, religious or camp leaders	4-6 total	Purposive	6	Implementation fidelity
Observation checklist	Primary and secondary	4 primary + 4 secondary schools	Purposive	16	For wash and other infrastructure. Checklist of teaching materials in primary school

Method	Target Group	Participants per Group	Sampling Strategy	Estimated Total Sample Size	Content type
Classroom observations	Primary school teachers	4 primary schools (2 classrooms each)	purposive	8	classroom conditions and teaching quality assessment

#### 10.4.5. Enumerator Refresher and Tool Refinement

Prior to midline data collection (early January 2026), the evaluation team will update the data collection tools based on lessons from baseline and the monitoring data. Some new questions may be added to capture progress and some baseline questions may be simplified if they were difficult. A **refresher training** for enumerators (likely many of the same team from baseline, ideally) will be conducted, by Assosa University. This shorter training (2–3 days) will reorient the team to any new protocols, emphasize reliability in re-measuring indicators, and reinforce ethical practices. Enumerators will be briefed on changes since baseline (for instance, if security or cultural context has shifted) and any new consent needed for follow-up. A quick pilot or pre-test of the midline tools will be done in one school to ensure the flow is smooth, but a full pilot is not needed as these are largely repeat measures.

#### 10.4.6. Data Collection - Fieldwork

The midline data collection is slated for **January 2026** (possibly spilling into early February), with an efficient schedule since many tools are repeats:

- Survey teams will revisit the target schools to administer follow-up surveys. Because familiarity is established, this should proceed faster than baseline. Data will again be captured electronically where possible for speed.
- FGD/KII teams will conduct qualitative sessions largely in parallel, coordinating with schools and communities to gather participants (with help from school principals or community leaders to mobilize parents).
- The evaluation team will ensure that midline fieldwork does not overly disrupt classes or activities – e.g. scheduling after-school FGDs or during free periods.
- Given the early timing (Jan is typically part of the school year in Ethiopia), care will be taken to not clash with exam schedules or holidays. If schools are on break in January, the plan will adjust accordingly (possibly doing community-based follow-ups).
- Ongoing quality control will include supervisors observing a sample of interviews/FGDs and daily data checks as done in baseline.

#### 10.4.7. Data Management and Analysis

Data handling at midline will mirror baseline:

- Quantitative data will be cleaned promptly (by mid-February 2026). The team will specifically create a baseline vs midline dataset, linking respondents across time where possible. Change metrics (e.g. difference in attendance, difference in % feeling safe) will be computed for the panel of respondents. Given the short timeframe, statistical significance tests may not be very meaningful, but basic comparisons and any notable shifts will be highlighted.
- Qualitative data from midline will be analyzed with a focus on *process themes and emerging outcomes*. Coding will incorporate new themes such as “initial

results/changes” and “implementation challenge”. The analysis will pay attention to whether comments corroborate or explain the quantitative trends. For instance, if attendance did not improve by midline, qualitative data might reveal ongoing barriers (like “water point broke down again” or “continued fear of conflict”).

- Triangulation is again key: midline results will be compared against programme monitoring data reported by UNICEF/partners in that period. If, for example, monitoring reports claim 10 boreholes built but midline found only 8 functioning, this discrepancy will be noted and investigated. The midline analysis will thus integrate multiple evidence streams (survey, FGD, KII, monitoring reports, observation checklists) to present a coherent story of progress and issues.

#### 10.4.8. Midline Reporting Process

A critical aspect of midline is ensuring findings are immediately useful. Therefore, the evaluators will convene a midline debrief workshop in early March 2026 (with UNICEF and key partners) *before* finalizing the report. In this session, preliminary findings on what’s working and not working will be discussed, and stakeholders can provide interpretations or additional context (for example, explaining a challenge that the data flagged). This participatory review helps validate the findings and directly feeds into recommendations for programme adjustments in 2026–2027.

#### Outputs and Reporting

- **Midline Evaluation Report – Draft:** A draft midline report will be produced by **early March 2026**. It will detail progress since baseline, covering both quantitative midline findings and qualitative insights on implementation. Importantly, this draft will include a set of *practical recommendations* for the remaining programme period (e.g. suggestions to intensify certain activities, address any lagging areas, or improve integration/coordination). The draft will be shared with UNICEF, donors, and implementing partners for review and feedback.
- **Midline Evaluation Report – Final:** The final midline report, incorporating stakeholder feedback, will be completed by **late March 2026**. This report will serve both accountability and learning purposes – documenting the state of the project at mid-term and explicitly pointing out any course corrections agreed upon. It will likely be shorter than the endline report, focused on interim findings and recommendations, but sufficiently detailed to justify the recommendations made.
- **Midline Presentation:** The evaluation team will prepare a concise presentation of midline results for a broad stakeholder meeting in late March 2026. The presentation will highlight key achievements to date (to reassure donors of progress), as well as challenges and recommended actions (to spur decision-makers towards necessary changes). UNICEF, government officials, and partner NGOs in the region will be invited to this presentation to ensure a common understanding of midline learnings. The

presentation can also be used internally by UNICEF to brief their senior management or the donor (MFA, LEGO Foundation, Grundfos) on progress.

### Midline Evaluation Report Outline

- **Introduction:** Recap of the programme and evaluation design, and specific purpose of the midline evaluation. This sets the context that this is an interim assessment, building on baseline and aimed at informing ongoing implementation.
- **Methodology:** Explanation of midline methods – including what was repeated from baseline and what new qualitative inquiries were added. It will note any limitations (e.g. shorter timeframe, some respondents lost to follow-up, etc.) and how these were mitigated.
- **Findings:** Structured in a way that answers the midline evaluation questions:
  - *Output Delivery and Activity Progress:* A factual reporting of what has been delivered by midline vs targets (e.g. X teachers trained out of Y, Z latrines constructed out of N planned, etc.), often presented in tables against plan. This addresses “what’s been done so far.”
  - *Initial Outcomes:* A comparison of baseline vs midline key indicators. Graphs or tables will show any changes (even minor ones) in measures like attendance, water availability, student well-being scores, etc. Each subsection will correspond to major outcome domains (education, WASH, MHPSS) and note whether trends are positive, negative, or unchanged.
  - *Implementation Fidelity and Quality:* A narrative subsection describing how implementation has unfolded. This will use evidence from KIIs, FGDs, and observations, organized by component or cross-cutting themes:
    - Education component implementation (training delivered? materials distributed? any issues like teacher turnover affecting it),
    - MHPSS/play implementation (status of play activities, psychosocial support roll-out, challenges like cultural acceptance of new methods),
    - WASH implementation (infrastructure installation progress, any maintenance issues, community engagement in WASH).
    - Integration/Coordination (how well are these activities synchronized at school/community level and management level).
    - This section will highlight successes (e.g. “stakeholders report high satisfaction with teacher training quality”) and challenges (“construction delays due to late procurement”).
  - *Beneficiary Feedback:* Qualitative findings from students, teachers, parents on their perceptions of change. Quoted anecdotes or summary of attitudes (e.g.

“Girls in refugee school X reported feeling more comfortable at school thanks to new latrines”). This complements the numeric outcome data.

- **Discussion:** An analysis of what the midline findings mean:
  - Are we on track? This explicitly states which outcomes seem on track and which are lagging, with reasoning.
  - Factors affecting implementation: discuss any systemic issues (such as security, staffing, or coordination challenges) that need addressing.
  - The value of integration so far: provide any evidence or lack thereof for the integrated approach’s added value at midline (for instance, noting if schools receiving all components show better early improvements than those that might not have full WASH yet, qualitatively).
  - Unintended outcomes: mention if any positive or negative unintended effects have emerged (e.g. community tension reduced due to joint activities, or unforeseen burden on teachers due to multiple trainings).
- **Recommendations:** This is a key part of the midline report. Based on the findings, a set of clear, actionable recommendations for the next 12–18 months will be listed. For example, “Accelerate WASH facility maintenance training for schools to ensure infrastructure remains usable,” or “Increase outreach to IDP girls who still show low attendance, possibly through community mentors.” Each recommendation will indicate who should take action (UNICEF, implementing partner, government) and ideally a timeframe (immediate vs within next 6 months).
- **Conclusion:** A brief summary reaffirming the programme’s potential to achieve its objectives, with midline data indicating where efforts should be concentrated. It will note that the endline will ultimately measure final impact, but actions taken after midline will be critical to success.
- **Annexes:** Detailed tables (e.g. a table comparing each indicator at baseline vs midline), updated evaluation matrix if needed, list of participants consulted, and technical annexes (like data collection tools used at midline, any deviations from inception plan).

## 10.5. Endline Evaluation (June–Sept 2027)

### 10.5.1. Endline Objective

The endline evaluation’s objective is to **assess the overall performance, outcomes, and lessons of the programme upon its completion**. It is a summative evaluation that will determine the extent to which the integrated programme achieved its intended results and the degree to which observed changes can be attributed to the programme interventions. The endline also aims to evaluate sustainability of outcomes and extract learnings for future programming. In summary, the endline will provide accountability to donors and stakeholders

by measuring results (effectiveness and impact), while also analysing *how* those results were achieved (process and contribution) and *what* can be sustained or replicated. A contribution analysis will be applied to evidence the programme's role in observed changes, and a thorough process documentation will explain the implementation journey. This final evaluation phase will consolidate findings to answer all key evaluation questions derived from the theory of change and inception report.

#### 10.5.2. Endline Focus Areas

- **Achievement of Outputs and Outcomes:** Verify which planned outputs were delivered and the level of outcome achievement across all programme components by 2027. This includes final outcomes in education (e.g. improved enrolment/attendance rates, learning outcomes, psychosocial well-being of students), final WASH outcomes (functional water and sanitation facilities, improved hygiene behavior), and child protection outcomes (safer school environments, reduced GBV incidents or increased referrals). The focus is on quantifying and describing the changes from baseline to endline for each result area.
- **Contribution and Attribution:** Examine the extent to which the programme contributed to the observed changes, distinguishing programme influence from other external factors. This focus area uses contribution analysis to build a credible narrative of causality: linking outputs to outcomes and checking if outcomes occurred in the absence of other major drivers. The evaluation will probe questions like “Would these changes have likely happened without the programme, or did the programme make a crucial difference?” Evidence for contribution will come from both quantitative change data and qualitative stakeholder attributions (e.g. what do beneficiaries themselves credit for the changes in their lives?).
- **Process Learnings and Implementation Performance:** Reflect on how the programme was implemented over the full duration – what worked well operationally and what challenges were encountered. This includes assessing the integrated delivery model: how effectively did UNICEF and partners manage the multi-sector partnership and coordination throughout? Were there adaptations made along the way (perhaps as a result of the midline findings) and did they improve results? The endline will document the implementation process, highlighting critical success factors and bottlenecks of the integrated approach over the two years.
- **Impact on Equity and Inclusion:** Determine how the programme impacted different groups. Did refugee and IDP children attain the same outcome improvements as host community children? Did girls benefit as much as boys in terms of education and safety? How did children with disabilities fare? This focus area ensures the evaluation addresses equity by analysing disaggregated outcomes and gathering perspectives from those sub-groups about their experience. It will identify if any group was left behind and why, thereby informing future inclusive strategies.

- **Sustainability and Transition:** Assess the likelihood that the benefits of the programme will continue after funding ends. This includes looking at institutionalization (e.g. are trained teachers now part of the regular system and continuing new practices? Have local authorities taken ownership of maintaining WASH facilities? Are community structures like parent groups or child clubs likely to continue?). The evaluation will also review any plans for scale-up or handover: for instance, how UNICEF or the government plan to maintain or expand the interventions using other resources.
- **Policy and System Influence:** (Related to sustainability) Consider whether the integrated programme influenced broader policy or practice. For example, did lessons from this initiative inform the regional education bureau's approach to refugee education or psychosocial support in schools? Did it strengthen any systems (like data monitoring, referral mechanisms for protection)? This focus area captures higher-level impacts that go beyond the immediate target schools.

### 10.5.3. Endline Draft Evaluation Questions

The endline will address the full set of overall evaluation questions, ensuring all OECD/DAC criteria (where relevant) are covered through a tailored lens. These questions presented below repeat those evaluations questions:

1. **Effectiveness – Achievement of Results:** *To what extent has the programme achieved its intended outputs and outcomes across the education, WASH, and MHPSS components? What observable changes occurred in the target schools and communities from baseline to endline (e.g. improved student attendance or learning, enhanced psychosocial well-being, increased access to clean water and sanitation, safer and more inclusive school environments)? This question will be answered by comparing endline data to baseline and targets, and analyzing whether objectives were met or exceeded.*
2. **Impact and Contribution:** *To what extent can the changes observed by endline be plausibly attributed to the programme interventions? In other words, what is the programme's contribution to those outcomes, and how do we know? This will involve examining evidence of causality (timing and sequencing of changes relative to activities, stakeholders' attribution of changes to the programme, absence of other explanations) and may include participatory exercises where beneficiaries rank or explain drivers of change. The question also covers any broader impacts (intended or unintended) of the programme on the community or systems.*
3. **Process – Implementation Quality and Fidelity:** *How well was the integrated programme implemented, and did it reach beneficiaries as intended? This looks at the entirety of implementation: Were activities delivered on schedule and to the expected quality standards by endline (e.g. number of teachers actually trained vs planned, infrastructure completed and functional, psychosocial sessions delivered regularly)?*

What challenges were encountered and how were they addressed over the life of the programme? This question also reviews the management and coordination aspect: *How effectively did UNICEF and partners coordinate the three components at all levels (planning, field implementation, monitoring)?* (added per inception report suggestions).

4. **Integrated Approach Value-Add:** *What value did the integrated multi-sector approach provide compared to a sector-specific approach?* Did the combination of education, WASH, and MHPSS interventions produce synergistic effects on children's well-being and education that would not likely have occurred with single-sector projects? The evaluation will explore this through qualitative insights and any measurable differences – for example, schools that received comprehensive inputs showing better overall improvement. It will also consider efficiency or trade-offs: did integration make implementation more efficient or more complex? Stakeholder perceptions (teachers, students, community leaders) will be key in answering this, as well as any retrospective comparisons to past single-sector efforts in the region.
5. **Equity and Inclusion:** *How did different groups benefit from the programme, and were any groups left behind?* This question ensures analysis of who benefited and who did not. It will examine outcomes by gender (girls vs boys), status (refugee, IDP, host), age (younger vs older students), and disability. It asks whether the programme was inclusive and responsive to each group's needs – e.g. Did refugee girls see gains in attendance and feel safer? Did children with disabilities get better support and access as intended? If disparities are found, the evaluation will investigate causes (cultural, operational, etc.) and note lessons for targeting.
6. **Sustainability:** *Are the outcomes and practices introduced by the programme likely to continue or be sustained after programme completion?* This question looks at signs of lasting change: for instance, are local institutions (schools, PTAs, government offices) prepared to maintain the WASH facilities and continue child-friendly teaching practices? Have capacities been built (teachers' skills, water management committees, child protection structures) that will endure? It will also assess the degree of ownership by government and communities – e.g. integration of refugee education into the regional education plan, or community commitment to keep girls in school – as proxies for sustainability.
7. **Lessons Learned and Scalability:** *What key lessons have emerged from the design and implementation of this integrated programme, and how can they inform future programming or policy?* While not a classic evaluation question from inception, this is usually included to ensure the evaluation captures practical insights. This covers what worked particularly well (best practices) and what pitfalls to avoid in similar interventions. Additionally, it asks if this model is recommended for scale-up or replication in other regions: under what conditions would it work or not work?

(These questions encapsulate the full scope of the evaluation as envisioned in the inception report. They will be answered with evidence collected throughout the evaluation phases, culminating in the endline analysis. The endline, being summative, provides definitive answers to effectiveness, impact, integration, equity, and sustainability.)

#### 10.5.4. Endline Data Collection Methods and Sample Sizes

The endline will deploy the most extensive data collection effort of all phases, integrating both the quantitative pre-post comparison and deep qualitative/participatory methods to perform contribution analysis and process review.

- **Quantitative Endline Surveys:** In **June–July 2027**, the evaluation team will carry out comprehensive endline surveys in the same target schools and communities as the baseline:
  - *Student Survey:* The structured student questionnaire from baseline will be administered again to measure endline values for all key indicators (enrolment/attendance, perception of safety and inclusion, psychosocial well-being indicators, etc.).
  - *Teacher Survey:* Teachers in the target schools will be surveyed, repeating questions on training received, current teaching practices, perceptions of changes in student behavior or learning, and their own job satisfaction and capacity. This captures the programme’s effect on teachers and teaching quality. It also checks the uptake of training content (e.g. do teachers report using gender-responsive methods, socio-emotional learning techniques, etc., which would indicate institutionalization of those practices).
  - *Head Teacher/School Survey:* School principals will provide final data on school metrics (final enrolment figures, any changes in infrastructure, the status of clubs or committees established by the programme, etc.). They will also be asked qualitative questions on what changes the school experienced and plans for sustaining activities.
  - *Community Survey (if applicable):* In addition to school-based surveys, a brief household or community survey could be administered to a small sample of parents to gauge changes in community-level outcomes (such as attitudes toward education, child well-being at home, or hygiene practices).
- **Qualitative and Participatory Methods:**
  - *Focus Group Discussions:* A series of FGDs will be held with all key stakeholder groups to capture in-depth reflections:
    - **Students:** Separate FGDs with boys and girls in the upper primary/secondary level (now having experienced the full programme). These discussions will explore how the programme affected them: Did they feel more motivated to attend or learn? Did the play-based

approaches and psychosocial support help them? Did WASH facilities make a difference for them (especially girls)? Students will be encouraged to share *stories of change* – for instance, “Tell us about a time in the past two years when school felt different or better for you.” Such narratives feed into Most Significant Change techniques.


- **Parents/Caregivers:** FGDs with refugee parents, host community parents, and IDP parents (as done in baseline/midline) to discuss what changes they have seen in their children and community. Have educational outcomes or behaviours changed? Are children happier or safer? Parents can also speak to any changes in their own engagement (like PTAs or parenting practices introduced by the programme). This will also surface whether they attribute changes to the programme (e.g. “We got water at school because of the programme” or “More girls are going to school now thanks to the programme’s efforts”).
- **Teachers:** FGDs with teachers will allow collective reflection on changes in teaching practice and student outcomes. Teachers can discuss, for example, improvements in student participation or challenges that remain. They will also address whether the support/training they received was adequate and how the integrated aspects (like having WASH facilities or psychosocial training) influenced their ability to teach effectively.

Each FGD at endline will also include forward-looking questions about sustainability (e.g. asking parents “Will you continue these practices after the programme?” or asking teachers “How will you keep using what you learned going forward?”).

- *Key Informant Interviews:* Individual interviews will be critical for gathering detailed and sensitive insights:
  - **Government & School Officials:** Interviews with officials from the Regional Education Bureau, WASH bureau, and possibly school principals (beyond the survey) to assess government ownership and prospects for continuation. For instance, “Is the Education Bureau planning to budget for maintenance of these facilities or continuation of teacher training?” Their perspective on how the project aligned with government policies and what policy changes or adoptions occurred will be noted.
  - **UNICEF and Implementing Partners:** Endline KIIs with UNICEF programme managers, donor representatives (e.g. Danish MFA, LEGO, Grundfos if they have local presence or via UNICEF), and implementing partner leads (Plan International, etc.) will gather reflections on overall

success, challenges, and partnership dynamics. This includes internal views on coordination among donors and any management lessons.

- **Assosa University/Academia:** If the local university was a key partner (for enumerators or technical support), an interview with university focal points could capture their perspective on capacity built and future collaboration, tying into sustainability of local evaluation capacity.
  - **Other Stakeholders:** Possibly interview representatives of refugee camp administration, community leaders, or children’s representatives (like club leaders) for additional insights.
- *Observation and Verification:* The endline team will visit a sample of schools to observe the end results: verifying the presence and condition of infrastructure (are the water systems and latrines built and in use?), possibly observing a class in session to see if play-based learning is happening, or attending a community meeting if available. They will use structured observation checklists to systematically record these observations. Additionally, monitoring data verification will occur: cross-checking programme records (output counts) with reality and stakeholder recall (as noted in inception). For instance, if records show X children in psychosocial clubs, we may ask teachers to confirm roughly how many participated and see if club materials are present at the school.
  - *Contribution Analysis Exercises:* To strengthen attribution, the evaluation may conduct a **participatory contribution analysis workshop** or use tools like the **Most Significant Change (MSC)** stories:
    - **MSC Stories:** Collect stories from beneficiaries about the most significant change they experienced due to the programme. This might involve asking several students, teachers, or parents to share personal stories, which are then reviewed to identify common outcomes attributed to the programme (e.g. “My daughter could return to school because a latrine was built”).
    - **Participatory Ranking:** In a workshop setting (perhaps one with community representatives or local officials), ask participants to list all factors that they think contributed to, say, improved school attendance, then rank these factors. If the programme’s interventions (e.g. school feeding, WASH improvements, etc.) rank top, it indicates strong perceived contribution. If external factors (like a new government policy) rank equally, those are documented as alternative contributors.
    - **Outcome Harvesting/Contribution Workshop:** As suggested in the inception report, the evaluators will bring together data and



stakeholders in a structured workshop at endline. In this session, they will map each key outcome and discuss evidence of what caused it. They will verify the sequence from outputs to outcomes and examine alternative explanations in a collaborative manner (for example, discussing “Enrolment went up – was it the new classrooms and outreach, or also influenced by a concurrent NGO project?”). This helps validate the evaluation’s conclusions on contribution with those who implemented or observed the programme.

## Sample Size

Method	Target Group	Participants per Group	Sampling Strategy	Estimated Total Sample Size	Content type
Student Surveys (Secondary)	Secondary school students (boys and girls)	85-90 per school	Clustered (4 secondary schools)	352	Endline indicators
Teacher Surveys (Primary)	Primary school teachers	15-18 per school	Clustered (12 primary schools)	192	Endline indicators
Teacher Surveys (Secondary)	Secondary school teachers	15-20 per school	Clustered (4 secondary schools)	63	Endline indicators
Head Teacher Surveys	Principals of all primary and secondary schools	1 per school	4 primary + 4 secondary schools	8	Endline indicators
FGDs - Secondary School Boys	Boys in secondary schools	2 per school	Purposive by gender/grade (4 secondary schools)	8	Impact, contribution, limitations/opportunities
FGDs - Secondary School Girls	Girls in secondary schools	2 per school	Purposive by gender/grade (4 secondary schools)	8	Impact, contribution, limitations/opportunities
FGDs - Refugee Parents	Parents of refugee children	1 FGD per school	Purposive (refugee parents across 2 primary and 2 secondary schools)	4	Impact, contribution, limitations/opportunities
FGDs - Host Community Parents	Parents from host communities	1 FGD per school	Purposive (host parents across 2 primary and 2 secondary schools)	4	Impact, contribution, limitations/opportunities
FGDs - IDP Parents	Parents of IDP children	1 FGD per school	Purposive (IDP parents across 2 primary and 2 secondary schools)	4	Impact, contribution, limitations/opportunities
FGDs - Teachers	Primary or secondary teachers	1 per school	Purposive (across 2 primary and 2 secondary schools)	4	Impact, contribution, limitations/opportunities

Method	Target Group	Participants per Group	Sampling Strategy	Estimated Total Sample Size	Content type
FGDs - Community Leaders	Local leaders, community elders, SMC/PTA	1 FGD per Woreda	Purposive	3	Impact, contribution, limitations/opportunities
FGDs - UNICEF/IP Staff	Programme field implementers	1 FGD per Woreda	Purposive	3	Impact, contribution, limitations/opportunities
FGDs - Social Workers	Frontline social workers/community facilitators	1 FGD per Woreda	Purposive	3	Impact, contribution, limitations/opportunities
FGDs - Host Community Members	Non-beneficiary residents of host areas	1 FGD per Woreda	Purposive	3	Impact, contribution, limitations/opportunities - cohesion, unintended impact focus
FGDs - Local Govt Officials	Education, Water, Social Affairs officials	1 per Woreda	Purposive	3	Impact, contribution, limitations/opportunities
Case Study - Children with Disabilities	Children with disabilities + caregiver	2-3 case studies	Purposive (identified through schools)	3	Impact, contribution, limitations/opportunities
KIIs - School Principals	Head teachers of selected schools	1 per school	Purposive (complement to survey)	4	Impact, contribution, limitations/opportunities
KIIs - Plan International Staff	Field-level implementing staff	2-3 per Woreda	Purposive	8	Impact, contribution, limitations/opportunities
KIIs - Community Leaders	Influential elders, religious or camp leaders	4-6 total	Purposive	6	Impact, contribution, limitations/opportunities
Observation checklist	Primary and secondary	4 primary + 4 secondary schools	Purposive	16	For wash and other infrastructure. Checklist of teaching materials in primary school

Method	Target Group	Participants per Group	Sampling Strategy	Estimated Total Sample Size	Content type
Classroom observations	Primary school teachers	4 primary schools (2 classrooms each)	purposive	8	classroom conditions and teaching quality assessment

### 10.5.5. Enumerator Training and Preparation

By mid-2027, it's likely some original enumerators may not be available, so the team will recruit/train a fresh (or refreshed) cohort. A comprehensive training similar to baseline will be conducted in June 2027 by Assosa University. This covers all new and revised tools (including training on any participatory methods like MSC story collection, which requires good facilitation skills). The training will emphasize consistency with baseline measures to ensure data comparability. A pilot test of the endline instruments will be done in one school to ensure clarity and timing (especially if new sections were added to capture final outcomes like learning tests or detailed perception questions). Ethical protocols will be reiterated, particularly since some sensitive questions (e.g. on protection or psychosocial well-being) might be included in endline tools to assess changes.

### 10.5.6. Data Management and Analysis

By **August 2027**, the focus shifts to turning raw data into findings:

- *Quantitative Data Analysis:* The cleaned endline dataset will be merged with baseline (and midline) data for repeated cross section analysis. Descriptive statistics will show the change over time for each indicator. The analysis will test for significant changes where possible (e.g. using t-tests or chi-square tests comparing baseline and endline values for key outcomes, acknowledging that without a control group these are not causal tests but still show magnitude of change). Graphs will be prepared illustrating before-and-after comparisons for major indicators (with disaggregation, for example a bar chart of girls' vs boys' attendance at baseline and endline). If an outcome improved markedly (say attendance rose by 15 percentage points), the team will highlight that and examine consistency across schools.
- *Qualitative Data Analysis:* All endline qualitative data (FGDs, KIIs, stories) will be coded and analyzed thematically. Given the volume, a qualitative analysis software (like NVivo or Atlas.ti) might be used to organize quotes by evaluation question (e.g. all quotes related to sustainability, all quotes about integrated approach). The analysis will look for convergence (e.g. many stakeholders credit teacher training for better student engagement) and divergence (e.g. one community experienced an unintended conflict that others did not). The Most Significant Change stories will be reviewed to identify which changes are commonly valued by beneficiaries. The evaluation team will document these stories (possibly as vignettes in the report to bring the data to life).
- *Contribution Analysis Synthesis:* Using the contribution framework, the team will systematically examine each outcome:
  - Did it change from baseline? (from quantitative data)
  - When did change occur relative to programme activities? (from timeline of outputs and monitoring data)

- What do stakeholders say caused the change? (from KIIs/FGDs, including any participatory ranking results)
  - Were there external factors present? (from context info, KIIs with officials, etc.)
  - The team will then assess the strength of evidence for the programme’s contribution for that outcome (strong, moderate, or weak) and note any alternative explanations. This will form the basis of the conclusions on attribution.
- *Triangulation:* The endline analysis will heavily triangulate between data sources. For example, if enrollment improved, the team will cross-check school records, survey data, and interview statements for consistency. Discrepancies will be investigated or clearly reported if unresolved. The evaluators will also compare findings with any external data (e.g. government reports, UNICEF’s routine monitoring, or even evaluations by other agencies in similar contexts) to ensure plausibility.

#### 10.5.7. Endline Reporting Process

##### Validation and Interpretation

Before finalizing the report, a stakeholder validation workshop will be conducted in late August or early September 2027. Here, the evaluation team will present preliminary endline findings to UNICEF, government representatives, and implementing partners. The purpose is to:

- Verify factual accuracy (ensure no major output or event was missed or misrepresented).
- Discuss the interpretation of results, especially for contribution and sustainability. Stakeholders can confirm or challenge the evaluators’ emerging conclusions (for instance, if the evaluation attributes improved attendance to the programme, stakeholders might add insight like “yes, and also the government’s free meal program helped” which will be incorporated).
- Build consensus on the lessons and recommendations, increasing the likelihood they will be acted upon.

The validation feedback will be used to refine the analysis and craft well-founded recommendations.

##### Outputs and Reporting

- **Endline Evaluation Report – Draft:** A comprehensive draft endline report will be prepared by **early September 2027**. This report will be the most extensive, covering the entire evaluation scope. It will include an executive summary for high-level audiences, a full description of the evaluation approach, detailed findings for each question, and evidence-backed conclusions and recommendations. The draft will be

shared with UNICEF, all donors (MFA, LEGO Foundation, Grundfos Foundation), implementing partners, and key government stakeholders for review. Given the broad audience, the draft will be written in clear, non-technical language (with technical annexes for detailed methods and data).

- **Endline Evaluation Report – Final:** The final endline report will be completed by **late September 2027** after incorporating feedback. This final report will serve as the definitive record of the programme’s performance and lessons. It will likely be formally published or at least disseminated among all partners. The final version will ensure that all stakeholder comments are addressed and that the report is balanced and evidence-based (any disagreements or differing interpretations can be noted in the report if needed).
- **Endline Presentation:** The evaluation team will deliver a presentation of the endline results to stakeholders around the time of the final report release (Sept 2027). This presentation will cater to both technical and non-technical audiences, summarizing key findings, success stories, and critical recommendations. It will be used in a dissemination workshop or closing meeting with UNICEF, donors, government ministries (Education, Water, etc.), and community representatives. The aim is to communicate what was achieved and what should happen next (post-program).
- **Policy Brief (Endline):** In addition to the report and presentation, a concise Policy Brief (2-4 pages) will be produced at endline. This brief will distil the most important findings and recommendations relevant to policy makers and high-level stakeholders. It will focus on actionable insights: for example, policy implications for integrating services for refugee education, or recommendations for government on sustaining school WASH. The language will be non-technical, highlighting achievements (to garner support for similar approaches) and critical needs for sustainability (to influence policy/resource allocation). This brief will be shared widely – including with regional and federal government officials, donor headquarters, and potentially made public to inform the broader development community about the integrated approach’s results.

#### [Endline Evaluation Report Outline](#)

- **Executive Summary:** A 3-5 page summary covering the background, evaluation purpose, methods, key findings for each major question, conclusions, and top-line recommendations. This will be written last to encapsulate the most critical information for quick readership by senior stakeholders.
- **Introduction:** Introduction to the programme (brief recap of its objectives, components, theory of change) and the purpose of the endline evaluation. It will outline the evaluation questions (the ones listed above) and criteria, and describe the structure of the report.

- **Evaluation Methodology:** Detailed description of the evaluation design and methods at endline (and overall). This includes the mixed-method pre-post design, justification for not having a control group (and how contribution analysis compensates), data collection methods used at endline (surveys, FGDs, KIIs, MSC, etc.), sampling strategy, and limitations. It will also describe any changes from the inception plan (for transparency) and how quality and ethics were ensured.
- **Findings:** This core section will likely be organized by evaluation question or thematic area, ensuring each of the key questions is answered:
  - *Effectiveness/Outcome Achievement:* Presentation of results for outputs and outcomes. This may be subdivided by component for clarity (Education outcomes, WASH outcomes, MHPSS outcomes) or by outcome level (outputs -> intermediate outcomes -> impact). For each, there will be a comparison of baseline vs endline values and an analysis of progress made. Graphs and tables, will be used to illustrate changes. Accompanying qualitative evidence (quotes or case studies) will illustrate what these changes meant for people (e.g. a text box story of a particular school's journey).
  - *Process & Implementation:* A narrative of how the programme was delivered over the two years. It will detail fidelity to plan, major deviations or adaptations, and coordination mechanisms. Successes and challenges in implementation will be documented, possibly with a timeline of key milestones. This section will draw from monitoring data and stakeholder testimonies.
  - *Integrated Approach Analysis:* Findings related specifically to the multi-sector integration question – were synergies realized?
  - *Equity and Inclusion Outcomes:* This subsection will present disaggregated results. Tables or charts showing outcomes for sub-groups (e.g. attendance improvement for girls vs boys, psychosocial well-being scores for refugees vs hosts, etc.) will be included. The narrative will explain which groups benefited most/least and why, referencing the qualitative data.
  - *Sustainability and Transition:* Findings on how likely the changes are to continue. This will include evidence like: the status of institutionalization, policies influenced, community buy-in. If any early signs of backsliding or risks to sustainability are present, they will be noted here.
  - *Unintended Effects:* A brief coverage of any other significant findings not captured above, such as unintended positive outcomes (e.g. social cohesion improved between refugees and hosts) or negative ones (e.g. initial tension over resource allocation).

- **Analysis/Discussion:** This section synthesizes the findings to directly answer each evaluation question:
  - Effectiveness: summarizing overall achievement (were objectives met? which were exceeded or unmet?),
  - Contribution: providing the analysis of how changes are linked to the programme (this is where the contribution analysis findings are distilled, possibly in narrative form for each major outcome domain, and an overall statement like “The evaluation concludes the programme made a substantial contribution to improved school attendance, while other factors played a role in psychosocial outcomes”),
  - Process: discussing the quality of implementation and what it meant for results (e.g. how certain implementation challenges limited effectiveness or how strong management facilitated success),
  - Integration value: giving a considered answer on whether the integrated approach was beneficial and in what ways,
  - Equity: analysing whether the programme was equitable in reach and effect,
  - Sustainability: giving a prognosis for each key outcome – likely sustainable vs at risk – and the factors influencing that.
  - This section will interpret what the evidence means in a broader context (for example, comparing to other programmes or expected norms).
- **Conclusions:** A set of concise statements that conclude the evaluation. These will flow from the analysis, highlighting overall judgment. This will be an overarching summary of success, impact, and remaining issues.
- **Recommendations:** A list of actionable recommendations derived from the conclusions and lessons. These will likely be divided by target audience:
  - For UNICEF and implementing partners: recommendations for future programming (e.g. how to improve integrated programme design, or specific actions if a next phase were planned, such as scaling up play-based learning to all refugee-host schools in Ethiopia).
  - For Government: recommendations on policy or system improvements (e.g. incorporate psychosocial support training into regular teacher professional development, allocate budget for maintenance of school WASH facilities, etc.).
  - For Donors: any recommendations on funding approach or integrated funding models (e.g. the benefit of flexible funding to address multiple needs together, based on what was learned).

- Each recommendation will be linked to evidence from the findings and aimed at ensuring the sustainability and replication of successes or addressing any gaps identified.
- **Annexes:** Extensive annexes will accompany the endline report, including the evaluation matrix (questions, indicators, data sources – showing how each question was answered), detailed technical tables (complete list of indicators baseline vs endline with statistical details), methodology details (sampling full details, data collection instruments, list of stakeholders consulted, schedules), and any supplementary analyses (for example, full description of the MSC stories collected, or detailed case studies). These annexes provide transparency and allow interested technical readers to delve deeper.

The endline report will be structured to ensure that readers who only need high-level findings can get them (via the executive summary and conclusion/recommendations), while those interested in the full story and data can find it in the main text and annexes.

## 11. Consolidated Timeline of Evaluation Activities (2025–2027)

The table below maps the key activities for each evaluation phase across the timeline. It ensures coordination of tasks such as tool development, data collection, training (in collaboration with Assosa University), analysis, and reporting milestones.

Timeframe	Baseline Phase (2025)	Midline Phase (2026)	Endline Phase (2027)
<b>June 2025</b>	<ul style="list-style-type: none"> <li>– Finalize evaluation framework and tools for baseline (surveys, FGD/KII guides).</li> <li>– Pilot test instruments in target area (Assosa) for cultural relevance.</li> <li>– Train enumerators &amp; facilitators (with Assosa University’s support) on baseline tools, ethics, and logistics.</li> </ul>	<i>(No midline activities yet – programme implementation period).</i>	<i>(No endline activities yet – programme ongoing).</i>
<b>July 2025</b>	<ul style="list-style-type: none"> <li>– Conduct <b>baseline data collection</b> in the field: surveys in ~8-9 schools; FGDs with parents/teachers; KIIs with stakeholders.</li> <li>– Monitor data quality daily; address any field issues (replacement schools/respondents if needed).</li> </ul>	<i>(No midline activities).</i>	<i>(No endline activities).</i>

Timeframe	Baseline Phase (2025)	Midline Phase (2026)	Endline Phase (2027)
<b>August 2025</b>	<ul style="list-style-type: none"> <li>– Complete baseline fieldwork by early August.</li> <li>– Perform data cleaning and preliminary analysis of baseline quantitative and qualitative data.</li> <li>– Internal team review of initial findings; prepare summary statistics and transcripts.</li> </ul>	<i>(No midline activities).</i>	<i>(No endline activities).</i>
<b>September 2025</b>	<ul style="list-style-type: none"> <li>– <b>Draft Baseline Report</b> writing (early Sept) – compile findings, baseline values, and initial recommendations.</li> <li>– Submit draft baseline report for UNICEF/partner review.</li> <li>– <b>Finalize Baseline Report</b> (late Sept) after feedback.</li> <li>– Deliver <b>Baseline Presentation</b> to UNICEF, donors, government in a dissemination meeting (late Sept).</li> </ul>	<i>(No midline activities).</i>	<i>(No endline activities).</i>
<b>Oct–Dec 2025</b>	– <i>No evaluation field activities (programme implementation)</i>	<i>(No midline activities; monitoring of programme continues).</i>	<i>(No endline activities; monitoring continues).</i>

Timeframe	Baseline Phase (2025)	Midline Phase (2026)	Endline Phase (2027)
	<i>continues; ongoing monitoring by programme).</i>		
<b>January 2026</b>	<i>(Baseline phase complete; programme ongoing).</i>	<ul style="list-style-type: none"> <li>– <b>Midline planning:</b> Review and refine midline data collection instruments (update surveys, design FGD guides focusing on progress/fidelity).</li> <li>– <b>Enumerator refresher training</b> with Assosa University (early Jan) to prepare for midline.</li> <li>– <b>Midline data collection</b> begins: field teams revisit schools to conduct follow-up surveys and new FGDs/KIIs (progress and process focus).</li> </ul>	<i>(No endline activities).</i>
<b>February 2026</b>	<i>– (No baseline activities).</i>	<ul style="list-style-type: none"> <li>– Complete <b>midline data collection</b> by mid-Feb.</li> <li>– Data cleaning and analysis of midline surveys (baseline vs midline comparisons) and qualitative process findings.</li> <li>– Begin drafting midline findings</li> </ul>	<i>(No endline activities).</i>

Timeframe	Baseline Phase (2025)	Midline Phase (2026)	Endline Phase (2027)
		and identifying recommendations.	
<b>March 2026</b>	– <i>(No baseline activities).</i>	<ul style="list-style-type: none"> <li>– <b>Draft Midline Report</b> delivered (early Mar) detailing interim findings and proposed course corrections.</li> <li>– Stakeholder review and feedback on draft.</li> <li>– <b>Finalize Midline Report</b> (late Mar) incorporating comments.</li> <li>– <b>Midline Presentation</b> held with UNICEF, donors, partners to discuss progress and agreed adjustments (late Mar).</li> </ul>	<i>(No endline activities).</i>
<b>Apr–Dec 2026</b>	– <i>No baseline activities (baseline data being used as reference).</i>	– <i>No midline activities (programme implements adjustments from midline, continuous monitoring).</i>	– <b>Preparatory monitoring:</b> Evaluation team keeps track of programme monitoring reports and context changes to inform endline design. Initial endline planning may start in late 2026 (revisiting evaluation matrix, deciding on any new tools like MSC stories).*
<b>January–May 2027</b>	– <i>No baseline activities.</i>	– <i>No midline activities.</i>	– <b>Endline planning:</b> Re-engage with UNICEF and partners to refine endline approach. Update evaluation matrix and develop endline data collection tools (surveys, KII/FGD guides, MSC story protocols, etc.) by

Timeframe	Baseline Phase (2025)	Midline Phase (2026)	Endline Phase (2027)
			April 2027. – Logistical arrangements for endline (reconfirm school samples, recruit any new enumerators) in May 2027.
<b>June 2027</b>	– <i>No baseline activities.</i>	– <i>No midline activities.</i>	– <b>Enumerator training &amp; pilot (endline):</b> Train field teams (with Assosa University’s involvement) on endline instruments, including participatory methods, and refresh ethics training (early June). – Pilot test endline questionnaires and discussion guides in one community; adjust tools as needed (mid-June). – <b>Begin endline data collection</b> in late June 2027.
<b>July 2027</b>	– <i>No baseline activities.</i>	– <i>No midline activities.</i>	– <b>Endline fieldwork ongoing:</b> Conduct comprehensive endline data collection through July: repeat surveys in all target schools, FGDs with students/parents/teachers, KIIs with stakeholders, direct observations. – Ongoing data quality checks and daily team debriefs to ensure completeness and reliability.
<b>August 2027</b>	– <i>No baseline activities.</i>	– <i>No midline activities.</i>	– Complete endline fieldwork by early Aug. – <b>Data cleaning</b> (Aug): Verify and clean survey datasets, transcribe and translate qualitative data. – <b>Analysis phase:</b> Quantitative analysis of changes (baseline vs endline comparisons), qualitative coding

Timeframe	Baseline Phase (2025)	Midline Phase (2026)	Endline Phase (2027)
			<p>and synthesis, and integrated contribution analysis.</p> <ul style="list-style-type: none"> <li>– Team drafts initial findings and begins writing endline report sections.</li> <li>– Conduct a <b>stakeholder validation workshop</b> (late Aug) to review preliminary results and discuss interpretations (feedback incorporated into analysis).</li> </ul>
<b>September 2027</b>	– <i>No baseline activities.</i>	– <i>No midline activities.</i>	<ul style="list-style-type: none"> <li>– <b>Draft Endline Report</b> completed and circulated for feedback (early Sept).</li> <li>– UNICEF, donors, and stakeholders review draft; provide comments.</li> <li>– <b>Finalize Endline Evaluation Report</b> by mid/late Sept, addressing all inputs.</li> <li>– Prepare and disseminate <b>Policy Brief</b> summarizing key findings and recommendations (by late Sept).</li> <li>– <b>Endline Presentation</b> and dissemination event in late Sept 2027, sharing final results and lessons with UNICEF, government, donors, and community representatives.</li> </ul>

## 12. Ethical Considerations

The evaluation will adhere to UNEG and UNICEF ethical guidelines. Given the vulnerable populations involved (children, refugees), all data collection will ensure informed consent (and assent for children), confidentiality, and do-no-harm protocols. Local research assistants will be trained in child-sensitive and trauma-informed interviewing, especially for MHPSS-related discussions. The design has been shaped to avoid burdening or excluding beneficiaries (as noted, no one is denied services for a control group). Instead, the evaluation activities (surveys, discussions) will be framed as an opportunity for communities to reflect on progress and voice their opinions, which can be empowering if done respectfully.

EPRI's detailed ethical protocol and procedures as they apply to EPRI and its partners is provided separately.

## 13. Limitations

The team acknowledges that without a control group, attributing changes solely to the programme is a limitation. We cannot quantitatively separate programme effects from broader trends (e.g., national education improvements). The before-after design is susceptible to history effects (other events coinciding with the programme) and maturation (children growing older, etc.). The contribution analysis will try to account for these, but some uncertainty will remain. Another limitation is potential response bias – by endline, communities know the programme, so in qualitative discussions they might over-credit it for improvements. Using tools like QUIP and maintaining some interviewer independence (possibly hiring an external firm for qualitative data collection) will mitigate this. Despite these limitations, the chosen approach is considered the most feasible and will yield valuable insights into programme performance and context.

## 14. Utilisation of Findings

The evaluation will result in a report and actionable recommendations. These will be shared with all stakeholders (UNICEF, government partners, donors). Because the design emphasises learning, the findings will highlight practical lessons (e.g. on integrating WASH and education, on engaging communities) that can inform scale-up or replication. Donors are particularly interested in whether the *integration* approach yields added benefits, so the evaluation will attempt to answer that through qualitative comparisons (Annex 2 outlines some case study ideas to compare current integrated schools with how things were before or in non-integrated settings). The insights on process will help adjust similar programmes in other regions or in a next phase. Overall, the evaluability assessment has steered the evaluation towards a realistic, context-appropriate design – prioritising the collection of credible evidence and meaningful stakeholder input over rigid experimental validity.

## 15. Annexures

### 15.1. Annex 1: Qualitative and Process Evaluation Tools

This annex summarizes the key qualitative and process-oriented tools that will be utilised in the evaluation:

- **Focus Group Discussion (FGD) Guides:** We will develop structured FGD guides for different groups: (a) *Children (students)* – separate guides for upper primary/secondary students and for younger children (using child-friendly techniques, like drawing or games, to elicit responses about how they feel about school). (b) *Parents/Caregivers* – focusing on changes in their children and their own involvement (e.g. “Have you noticed any differences in your child’s behaviour or learning? What do you think caused it?” and “How have you been involved through parenting sessions or school committees?”). (c) *Teachers* – discussing training received, changes in teaching practice, perceptions of student outcomes, and integrated approach feedback (“Did having water and play materials make teaching easier?”). Each FGD will be facilitated by a trained moderator in the local language; sessions will last roughly one hour, recorded with consent, and include participatory exercises (like ranking the most significant changes, or problem tree analysis for remaining issues).
- **Key Informant Interview (KII) Protocols:** Semi-structured interview checklists will be tailored for different stakeholders:
  - *UNICEF Programme Managers and Technical Specialists:* covering strategic questions on design rationale, coordination among donor components, and any adaptation during implementation.
  - *Implementing Partner Staff (Plan International field coordinators, MHPSS facilitators, etc.):* focusing on operational lessons, challenges faced in the field (logistical, cultural), and observations of outcomes.
  - *Government Officials (Education Bureau, Water Bureau, Refugee agency RRS local reps):* covering government ownership, policy context (e.g. how refugee integration is proceeding), and plans to sustain or scale the approach.
  - *School Principals and Teachers-in-Charge:* focusing on school-level integration, e.g. managing multiple new activities, any timetable issues, and effects on school management (was it burden or benefit?).
  - *Social Workers/Case Managers:* to understand the child protection side – e.g. referral systems, cases handled, community response to services.

- *Community Leaders and PTA members*: capturing community perspective, support or resistance encountered, changes in community attitudes (towards education, gender, etc.).

The KII guides will ensure core questions are asked for consistency (like “What changes have you seen since the programme began?” and “What factors helped or hindered these changes?”) but allow flexibility for informants to elaborate on areas of their expertise. Interviews will typically be one-on-one, 30–60 minutes, held in a private setting to encourage openness.

- **Direct Observation Checklists**: Evaluators will use checklists during site visits to observe tangible outputs and context. This includes:
  - A **school observation form** to note infrastructure status (e.g. are the new latrines functional and being used? Is there water in the taps? Are play materials visible in classrooms?), classroom environment (presence of learning materials, student engagement, any observable inclusion of MHPSS activities like a relaxation corner), and overall school climate (evidence of gender inclusion, e.g. girls actively participating, any protection messages on walls, etc.).
  - Observation of a sample **intervention activities** if timing allows – e.g. sitting in on a teacher training session, a life-skills club meeting, or a community awareness session. The checklist would note participant engagement, relevance of content, and any integration (does the session incorporate multiple themes like WASH and protection together?).
  - **Community observation**: visiting a few surrounding communities to see if, for example, home water points or community safe spaces were established, and how they are maintained or used.

Observational data will complement reported data, helping validate whether outputs are actually operational and giving context (for instance, an installed water point but non-functional would be a key finding).

- **Most Significant Change (MSC) technique**: As part of FGDs, we will employ MSC storytelling. We will ask participants to recount “Since this programme started, what is the most significant change you have experienced (or seen in your community) and why is it important to you?” We will collect a range of stories (ideally at least 1–2 per community) and later analyze which domains of change they pertain to (e.g. “children are happier” = psychosocial well-being domain, “girls go to school more” = access/gender equity domain). This will provide qualitative evidence of outcomes that

matter to beneficiaries, possibly highlighting changes the formal indicators miss. We will also verify which of these changes are attributed to the programme.

- **Outcome Harvesting Workshops:** Near the end of data collection, the evaluation team will conduct one or two participatory workshops with key stakeholders (e.g. UNICEF staff, implementing partner, government reps, and some community representatives if feasible). In these sessions, we will:
  - Present preliminary findings (outcomes observed).
  - Ask participants to identify any additional outcomes they've noticed.
  - For each outcome, discuss and “harvest” the contributing factors – was it the programme? If so, which activities? If not, what else?
  - Check the ToC assumptions against reality – did the assumed pathways hold? (e.g. “We assumed training teachers in MHPSS would improve children’s well-being – did it happen, and how?”).

This process both validates the findings and generates nuanced insights on contribution. It also increases utilization, as stakeholders engage with the results directly.

- **Verification of Monitoring Data:** The evaluators will also audit or verify a sample of programme monitoring data for accuracy. For example, if reports say “500 children attended play sessions,” during FGDs or KIIs we might ask how many typically attend and cross-check with attendance sheets if available. Similarly, physical verification (count the new classrooms built, see the actual log of case management referrals) will be done for a reality check. This ensures the evaluation isn’t solely reliant on self-reported data.
- **Documentation Review:** As a tool, we will thoroughly review all relevant documents, not limited to proposals: training manuals, curricula used (to understand content), any needs assessments (like the BG needs comparison to contextualise baseline conditions), and previous evaluation reports of similar programmes. This background review enriches our qualitative inquiry (interviewers will be well-informed and can ask more pointed questions) and helps triangulate findings (e.g. comparing our findings with known lessons or statistics).

Each of these tools will be piloted during the inception phase of the evaluation to refine questions for clarity and cultural appropriateness. Enumerators and facilitators will be trained technically and on ethical considerations for interviewing children and vulnerable groups (e.g. obtaining consent, confidentiality, referral procedures if a protection issue is disclosed during an interview). By combining FGDs, KIIs, observations, and participatory techniques, the evaluation ensures a comprehensive qualitative understanding that complements the

numbers. These tools are vital for unpacking the “how” and “why” behind the programme’s performance – which is at the heart of a utilisation-focused, learning-oriented evaluative exercise.

### 15.2. Annex 3: Sampling Approach

In this study there are total of 16 schools among these 7 refugee and 9 host community schools. Based on this, these all 16 primary and secondary schools will be selected as a sample using comprehensive/holistic sampling technique, because the number of schools are easily manageable and the concerned bodies wants to obtain relevant information from the whole schools.

To determine a representative sample size from the target population, the sample size for students will be determined from the given population by using Kothari (2004, p. 179) formula by taking into accounts 0.05 (5%) standard error or significant level.

According to Kothari (2004, p. 179), the formula is 
$$n = \frac{z^2 \cdot p \cdot q \cdot N}{e^2 \cdot (N-1) + z^2 \cdot p \cdot q}$$

Where, n = the desired sample size

z = the value of the standard variation at a given confidence level (to be read from the table giving the areas under normal curve)

p = the proportion of target population estimated (50%), q = 1-p, e = acceptable error (the precision)

N = population size

Representative sample of population will be determined at 95% degree of confidence.

Therefore, at 95% degree of confidence, Z=1.96 p=0.5, q=1-p, e=5% (0.05);

by substitution;

z=1.96, z<sup>2</sup>=3.8416; e= 0.05, e<sup>2</sup>=0.0025

n=  $\frac{3.8416 \times 0.5 \times 0.5 \times 4151}{0.0025(4151-1) + 3.8416 \times 0.25}$

n= 3986.6204/11.3379= 352. So, the total sample sizes of students that will be included as respondents in this study are 352.

After the sample size is determine the proportional allocation of the sample for each stratum is drawn as:

$$n_i = \frac{nN_i}{N}$$

Where,

n= sample size

n<sub>i</sub>= Population size of the i<sup>th</sup>strata

N= the population size

Accordingly, the following sample representatives are drawn as shown on the table below stratified sampling technique will be used to select sample teachers and students by keeping their proportionality. The reason why we selected high students for obtaining data from questionnaire is that they can read, understand and give data than primary school students.

S. N	Name of Schools	Total Number of students grade 9-12	%	Sample students by their stratum
1	Sherkole Refugee Camp Secondary school	1018	8.47	86
2	Oura Secondary School (host)	413	8.47	35
3	TsoreArumela Refugee Secondary School	1560	8.47	132
4	Womba Refugee Inclusive Secondary School	1160	8.47	99
	Total	4151	8.47	352

Sample number of student in each sample school is based on stratified random Sampling technique.

Moreover, using the above formula the total sample sizes of teachers that will be included as respondents in this study are 251 which were determined as follows:

$$n = \frac{3.8416 \times 0.5 \times 0.5 \times 725}{0.0025(725-1) + 3.8416 \times 0.25}$$

$n = 696.29 / 2.7704 = 251$ . The detail number of respondents that will be included from each school is described in the table below.

S. N	Name of Schools	Total number of refugee and host community teachers who teach grade 1-12	%	Sample teachers by their stratum
1	AkudaTumet primary school	10	35	5
2	Bambasi primary school	89	35	31
3	Bambasi refugee camp primary school	94	35	32
4	Baro primary school	22	35	8

5	Besina primary school	25	35	9
6	Homosha Joiner primary school	22	35	8
7	Mender 42 primary school	4	35	1
8	Musa Shewarie/Dabus primary school	7	35	2
9	Sherkole primary school	25	35	9
10	Sherkole refugee camp primary school	101	35	35
11	TsoreArumela primary school	16	35	5
12	Tsore refugee camp primary school # 1	128	35	44
13	Tsore refugee camp primary school # 2	43	35	15
14	Womba primary school	19	35	6
15	Sherkole Refugee Camp Secondary school	32	35	11
16	Oura Secondary School (host)	18	35	6
17	TsoreArumela Refugee Secondary School	35	35	12
18	Womba Refugee Inclusive Secondary School	35	35	12
<b>Total</b>		<b>725</b>	<b>35</b>	<b>251</b>

Sample number of teachers in each sample school is based on stratified random Sampling technique.

### 15.3. Annex 4: Implementation Progress and Target Coverage by Component

#### 1. Programme Inception and Planning

- **Inception Phase Completed:** The inception phase (April–October 2024) included context assessments, refinement of activities, and updating the MEAL framework to ensure alignment with local needs and donor priorities.
- **Stakeholder Engagement:** Two rounds of workshops with government and CSO partners established a framework for integrated activity implementation and cross-sectoral collaboration.

## 2. Education

- **Teacher Training:** 211 teachers (81 female, 130 male), including 20 refugee teachers, were trained in teaching methodology, gender-responsive pedagogy, inclusive practices, and life skills education.
- **Gender Clubs:** 15 school gender clubs were established or strengthened, serving as platforms for protection, menstrual health and hygiene (MHH), and reporting of violence.
- **Environmental Protection Clubs:** Refugee-hosting woreda education offices supported the creation of Environmental Protection Clubs to promote disaster risk reduction (DRR) and climate adaptation in schools.
- **Localized Learning Materials:** The national pre-primary education standard was translated into four local languages to ensure accessibility and relevance.

## 3. Child Protection

- **Case Management Systems:** Eight social workers were deployed, and 42 individuals from key sectors were trained in case management and CPIMS+ (Child Protection Information Management System+).
- **Case Identification and Support:** 27 children (20 girls, 7 boys) experiencing abuse, neglect, or exploitation were identified and supported; 29 unaccompanied and separated children were placed in kinship care with ongoing follow-up.
- **GBV and PSEA Training:** 26 participants received training on gender-based violence (GBV) prevention and protection from sexual exploitation and abuse (PSEA), with community awareness efforts and hotline dissemination.

## 4. WASH (Water, Sanitation, and Hygiene)

- **Infrastructure Progress:** Construction and rehabilitation of WASH facilities in targeted schools are underway, with 75% of physical infrastructure completed, including water supply systems, latrines, and MHH safe spaces.

- **Community Water Supply:** Significant progress in water supply projects for both schools and host communities, though some activities face funding gaps due to external factors.

#### 5. Community Engagement

- **Outreach Activities:** Community outreach reached 520 children and 620 adults, focusing on education, protection, gender, climate, and WASH themes.
- **Feedback and Buy-In:** Preliminary feedback indicates strong community support and willingness to contribute to programme goals.

#### 6. Challenges and Mitigation

- **Funding Gaps:** Delays and shortfalls, particularly in WASH infrastructure, due to external funding suspensions.
- **Coordination Delays:** CSO engagement and contracting took longer than planned, affecting the rollout of some activities.
- **Resource Constraints:** Limited water sources and high costs for new infrastructure in some schools.
- **Mitigation Measures:** Additional emergency funds, joint acceleration plans, and capacity building for local CSOs have been implemented to address these challenges.

#### 7. Monitoring and Partnerships

- **Joint Monitoring:** Regular field monitoring and supportive supervision by UNICEF and government partners ensure adherence to plans and quality standards.
- **MEAL Framework:** Revised and strengthened to capture integrated results and learning, with baseline and impact evaluation planned in partnership with Assosa University and EPRI.
- **Inter-Agency Collaboration:** Strong partnerships with government bureaus, UNHCR, and Plan International support multi-sectoral delivery and systems strengthening.

The integrated programme in Benishangul Gumuz has made substantial progress in establishing systems, building capacity, and delivering services across education, protection, and WASH sectors. While challenges such as funding gaps and coordination delays persist, mitigation measures and strong partnerships have enabled continued advancement toward programme goals. The initiative demonstrates a robust, adaptive approach to meeting the needs of forcibly displaced and host community children in a complex and evolving context

