



Kingdom of the Netherlands

unicef   
for every child

EVALUATION



REPORT 3

**UNICEF Impact Feasibility Assessment of PROSPECTS**

# Uganda Case Study

© United Nations Children's Fund (UNICEF), June 2025

This report was prepared by Preksha Golchha, Jose Victor C. Giarola (Economic Policy Research Institute), with the guidance and supervision from Eduard Bonet Porqueras (Evaluation Office, UNICEF) and support from Innocent Kaba (Displacement and Migration Hub, UNICEF). The Impact Feasibility Assessment was commissioned by Tasha Gill and Rhonda Fleischer, the global lead and coordinator (Displacement and Migration Hub, UNICEF), and managed by Eduard Bonet Porqueras, with support from Innocent Kaba.

Acknowledgments: Earlier drafts of this report benefited from the valuable feedback provided by Amber Peterman and Andrew Kaiser-Tedesco, who nonetheless bear no responsibility over any flaws that this published version may have. We would like to thank all the UNICEF PROSPECTS programme staff who have kindly shared ideas and especially access to documentation; Lauren Farwell and Khaled Khaled, Finally, we would like to express our sincere appreciation to the Netherlands Ministry of Foreign Affairs for funding, facilitating and guiding the PROSPECTS innovative partnership among United Nations agencies and multilateral financial institutions to find sustainable solutions to improve the well-being of displaced and host populations.

Suggested citation: United Nations Children's Fund, 'Impact Feasibility Assessment of PROSPECTS: Uganda Case Study', UNICEF: New York, 2025.

Cover photo: © UNICEF/UNI762359/Anthony  
Design and layout: Elena Panetti

Please contact:

UNICEF

Evaluation Office

3 United Nations Plaza

New York, NY 10017, USA

Email: [evalhelp@unicef.org](mailto:evalhelp@unicef.org)

Website: [www.unicef.org/evaluation/](http://www.unicef.org/evaluation/)

# Table of Contents

01	<b>Motivation and Programme Description</b>	<b>4</b>
02	<b>Impact Pathways and Key Outcomes of the Intervention</b>	<b>8</b>
03	<b>Evaluation Questions</b>	<b>11</b>
04	<b>Evaluation Design</b>	<b>12</b>
05	<b>Key outcome indicators</b>	<b>20</b>
06	<b>Timeline</b>	<b>21</b>
07	<b>Data Collection</b>	<b>22</b>
08	<b>Analysis Methods</b>	<b>24</b>
09	<b>Estimated Resources for Data Collection</b>	<b>26</b>
10	<b>Limitations</b>	<b>27</b>
11	<b>Ethical Considerations</b>	<b>28</b>

## 01

# Motivation and Programme Description

Uganda hosts the largest refugee population in Africa – over 1.5 million refugees, 57% of whom are children. Refugee and host community adolescents face significant challenges: access to education is low (gross secondary enrolment for refugee youth is only 12%), and youth unemployment is high (estimated 40–47% unemployment among refugee youth, versus 12–15% for Ugandan youth).

These challenges underscore the need for integrated solutions that equip young people with skills for education, work, and resilience. In this context, vulnerable adolescents, including both refugees and host communities, face intersecting barriers to education, skilling, and dignified employment. PROSPECTS is a multi-agency partnership to improve education and economic inclusion outcomes for forcibly displaced persons (FDPs) and host communities. Within this framework, UNICEF leads the Youth Life and Digital Skilling intervention, which integrates education, skilling, protection, and livelihoods support.

This evaluation aims to assess the effectiveness of the Uganda intervention in contributing to improved learning, skills, and economic empowerment outcomes for adolescents by the end of 2029. The PROSPECTS 2.0 partnership—a collaboration of UNICEF, the World Bank, the ILO, UNHCR, and the IFC—aims to improve the economic and social inclusion of displaced persons and host communities by “putting decent work at the nexus of humanitarian and development cooperation.”

Within PROSPECTS’ multi-sectoral approach (spanning Education & Learning, Economic Inclusion, Protection/Social Protection, and Infrastructure), UNICEF leads a “Learning to Earning – Youth Life and Digital Skilling” intervention in Uganda, targeting

vulnerable adolescents (refugee and host community youth). This evaluation design report outlines how a rigorous impact evaluation will assess the effectiveness of this intervention by the end of 2029. It follows the structure and analytical tone of UNICEF’s Impact Feasibility Assessment (IFA) design reports, aligning the evaluation with the program’s theory of change, national results frameworks, and ethical standards. The goal is to determine *whether, how, and for whom* the Uganda youth skilling intervention delivers on PROSPECTS outcomes, and to generate evidence to inform scale-up and policy.

**Intervention Overview:** The Uganda “Youth Life and Digital Skilling” initiative is a comprehensive program that combines life skills education, digital skills training, entrepreneurial support, on-the-job learning, and social protection for adolescents<sup>1</sup>. It is implemented through both formal education systems and non-formal community platforms, with a focus on refugee settlements and host communities. Key components include:

- ▶ **Life Skills Training (Formal & Non-Formal):** UNICEF, in collaboration with the Ministry of Education and Sports and the Ministry of Gender, Labour and Social Development, is rolling out a standardised life skills curriculum for **in-school and out-of-school adolescents**. Training

<sup>1</sup> Please refer to Stage 2 Impact Feasibility Assessment Report for a discussion on the intervention mapping. The mapping included clustering and reorganising the programs and activities into broader interventions using an evaluation lens.

emphasises transferable skills (communication, teamwork, problem-solving, etc.), social competencies, and personal empowerment needed for work and life. **School clubs and co-curricular activities** are platforms for skills practice and peer learning, reinforcing positive outcomes like improved school retention and civic engagement. The content is localised and inclusive, accommodating learners with disabilities and diverse backgrounds. This component addresses the barrier that many refugee youth cannot access formal secondary education by providing alternative pathways to develop essential life competencies.

- ▶ **Digital Skills (iUPShift) Scale-Up:** Building on a successful Phase I pilot of the *Digital Skills Toolkit in 15 schools*, the intervention is **institutionalising digital literacy and innovation training (called “iUPShift”) in both schools and community centers**. Guided by a national Digital Agenda Strategy (2023–2027), UNICEF will reach **10,000 adolescents** with digital skills training in Phase II – **5,000 students** across roughly 40 secondary schools and 4 vocational training centers, and **5,000 out-of-school youth** through Youth Skills Centers and community hubs established with ILO and UNHCR). The curriculum enhances ICT literacy, problem-solving using technology, and fosters social innovation projects (modeled after UPSHIFT) to empower youth as creators of solutions. The program aims to bridge the digital divide and prepare youth for emerging “digital economy” livelihood opportunities in Uganda and beyond by integrating digital skills into education and training settings.
- ▶ **Job-Specific Skilling and Work-Based Learning:** To facilitate the transition “from learning to earning,” the initiative provides **job-specific training in green and digital economy sectors**, coupled with practical work experience. About **2,500 adolescents** will receive targeted vocational or entrepreneurship training aligned to market needs in areas like renewable energy, agri-tech, e-commerce, and fintech. Of these, **1,000 youth** will be linked to **apprenticeships or internships** with private sector companies (with priority on those in green economy and ICT sectors). These placements offer hands-on experience, mentorship,

and networking that improve employability. Youth interested in entrepreneurship receive support to develop and sustain small businesses. By 2029, the program envisions a cohort of skilled youth either gainfully employed or running enterprises – demonstrating improved livelihoods. This responds directly to the unemployment gap for refugee youth (often **30–40 percentage points higher** than nationals).

- ▶ **Environmental and Community Engagement:** The intervention integrates climate action and community service into youth skilling in partnership with ILO and local authorities. During training, adolescents develop “green skills” and participate in environmental projects (e.g., tree planting, recycling initiatives). *Youth-led participation platforms* are established, connecting young “green champions” with government officials to voice recommendations and participate in community development. This component builds youth leadership and civic engagement, aligns with Uganda’s climate resilience goals, and fosters social cohesion between refugees and hosts. It illustrates that the program sees youth as “active agents of change” in their communities, not just beneficiaries.

The Youth Life and Digital Skilling was selected as a promising intervention to develop an impact evaluation plan based on the systematic progress in Stage 2 of the Impact Feasibility Assessment, which includes the assessment of both country- and intervention-level factors (see Box 1).

## Youth Life and Digital Skilling Selection

The Youth Life and Digital Skilling was selected for Stage 3 of the Impact Feasibility Assessment via a standardized process consisting of: 1) an analysis of **country context** and 2) an **intervention context** by mapping of PROSPECTS interventions in the 8 PROSPECTS countries (with 33 interventions included in total). Key considerations in the country context were: 1) Political interest and will from government and partners to understand what works and to what extent ministries would support the system changes necessary to scale up a successful intervention; 2) the operational facility, including potential risks to a successful evaluation; 3) the prioritization based on knowledge gaps – as assessed on the IFA Stage 1 Rapid Review; and 4) the national data and evaluation capacity, including the existence of strong research institutions in the country and high-quality sources of secondary data. Key considerations in the intervention context were: 1) the scale and scalability of programming, which considers whether interventions are large enough to support rigorous impact evaluations; 2) previous or planned impact evaluations; 3) the potential for future expansion; 4) the knowledge gains, which prioritizes interventions capable of addressing knowledge gaps identified during Phase One of the IFA (Rapid Review);<sup>2</sup> and 5) the type of programming, which assessed interventions based on ToC Integration and partners integration.

- ▶ **Country Context:** Uganda has scored high in terms of country context, with high political will, operational facility and national evaluation capacity.
- ▶ **Intervention Context:** The Youth Life and Digital Skilling was ranked as the top priority in Uganda (among four interventions included), as it met all assessment criteria, including those relating to scale and scalability, plans for future expansion, no existing impact evaluation, knowledge gains and type of programming based on ToC integration and partners integration.

The Youth Life and Digital Skilling is one of three interventions shortlisted as a priority for Stage 3 (alongside interventions in Jordan and Ethiopia) and one additional intervention as secondary priority (Kenya) for which impact evaluation plan is being developed.

<sup>2</sup> It is important to note that evidence gaps identified in the rapid review were limited to PROSPECTS countries and forcibly displaced populations. A detailed explanation of this decision can be found in the Section 2.1 “Selection Criteria” of the Stage 1 Impact Feasibility Assessment Report.

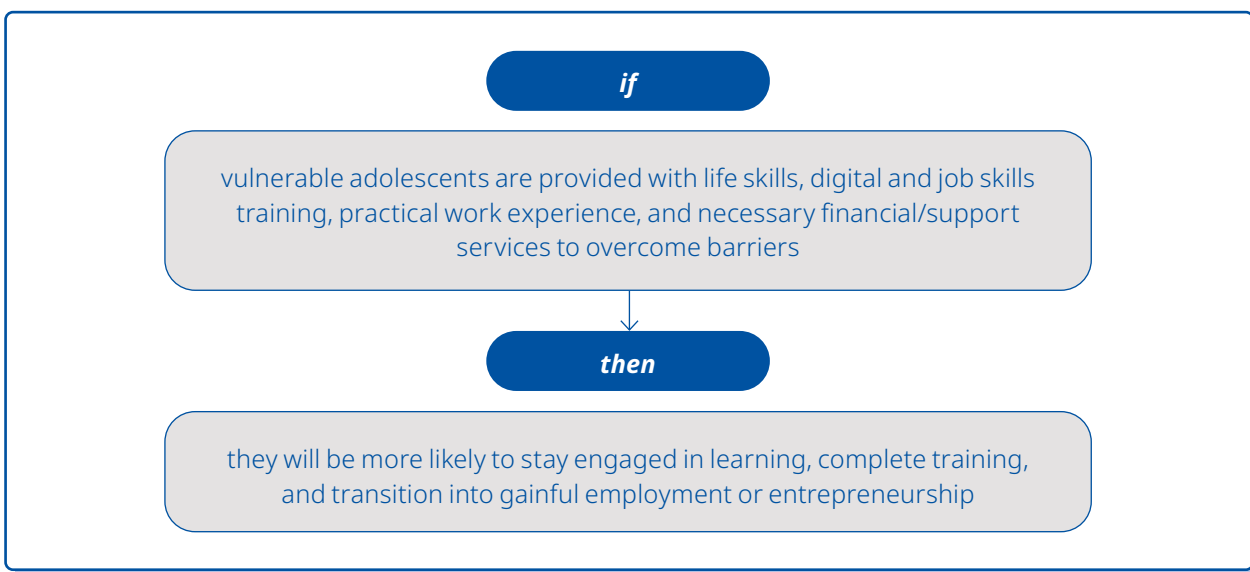
Name	Objective	Target Group	Services / Benefits Provided	Implementation Platform	Linkages with Other Interventions	Outcome Indicators (from TOC)
<b>Rolling out life skills training</b>	Equip adolescents with key transferable skills for work and life through a localized, scalable life skills programme.	36,000 in-school adolescents (180 primary, 40 secondary schools); 9,600 out-of-school adolescents	Life skills training; disability-inclusive learning resources; and co-curricular club strengthening	In-school: Primary & Secondary Schools (via school clubs); Out-of-school: UNHCR/ILO Youth Skills & Community Centers	Foundation for co-curricular models and life skills; supports social cohesion, institutionalization of life skills	1.2 Transferable Skills
<b>Scaling up digital skills training</b>	Institutionalize and expand digital literacy through formal and non-formal systems, leveraging the Digital Skills Toolkit.	5,000 in-school adolescents (40 secondary schools, 4 TVETs); 5,000 out-of-school adolescents	Digital literacy and digital economy skills; ICT infrastructure use	In-school: Secondary Schools and TVETs; Out-of-school: UNHCR/ILO Community Centers and UNICEF Youth Skills Centers	Aligned with apprenticeship and internship programmes (Pillar 1.4); leverages infrastructure set up in Phase 1	1.2 Transferable Skills
<b>i-UPSHIFT life and entrepreneurship skills</b>	Scale social innovation and entrepreneurship through youth-driven, community-led design approaches.	5,000 in-school; 5,000 out-of-school adolescents, especially from refugee and host communities	HCD workshops, mentorship, incubation, seed funding, localized content, disability inclusion	In-school: Secondary Schools and TVETs; Out-of-school: ILO/UNHCR/ UNICEF centers	Aligns with digital & green economy goals; connects to participation and climate action streams	1.2 Transferable Skills
<b>Job-specific training &amp; internships</b>	Support adolescents' transition from learning to earning through job-specific skills and access to employment pathways.	2,500 out-of-school adolescents (1,000 with internships or entrepreneurship support)	Training in green/digital economy sectors; internships; support for small business creation	Out-of-school: Centers facilitated by UNICEF and partners	Complements digital skills, i-UPSHIFT and internship/apprenticeship streams	1.4 Internship, Apprenticeship, Entrepreneurship, Employment Opportunities

# Impact Pathways and Key Outcomes of the Intervention

The Uganda youth skilling intervention is explicitly designed to contribute to PROSPECTS 2.0 Theory of Change (ToC) results under Pillar 1 (Education & Learning) and Pillar 2 (Economic Inclusion). By providing “expanded access to quality education, skills, and learning opportunities”, especially for refugee and host-community adolescents, the program feeds into the outcome that forcibly displaced persons (FDPs) and host communities gain the knowledge and skills they need to improve their livelihoods.

The theory posits that equipping young people with relevant life and digital skills (Pillar 1 outcomes: *access to basic services like education/training*) will enhance their readiness for further education or employment. In turn, connecting youth to economic opportunities (jobs, entrepreneurship, and cash support – Pillar 2 outcomes: *access to socio-economic opportunities*) is expected to increase their income, resilience, and self-reliance, thereby breaking cycles of aid dependence. This sequenced pathway – education/skills → employability → employment/income – is central to PROSPECTS’ vision of improved self-reliance and durable solutions for refugees and host communities.

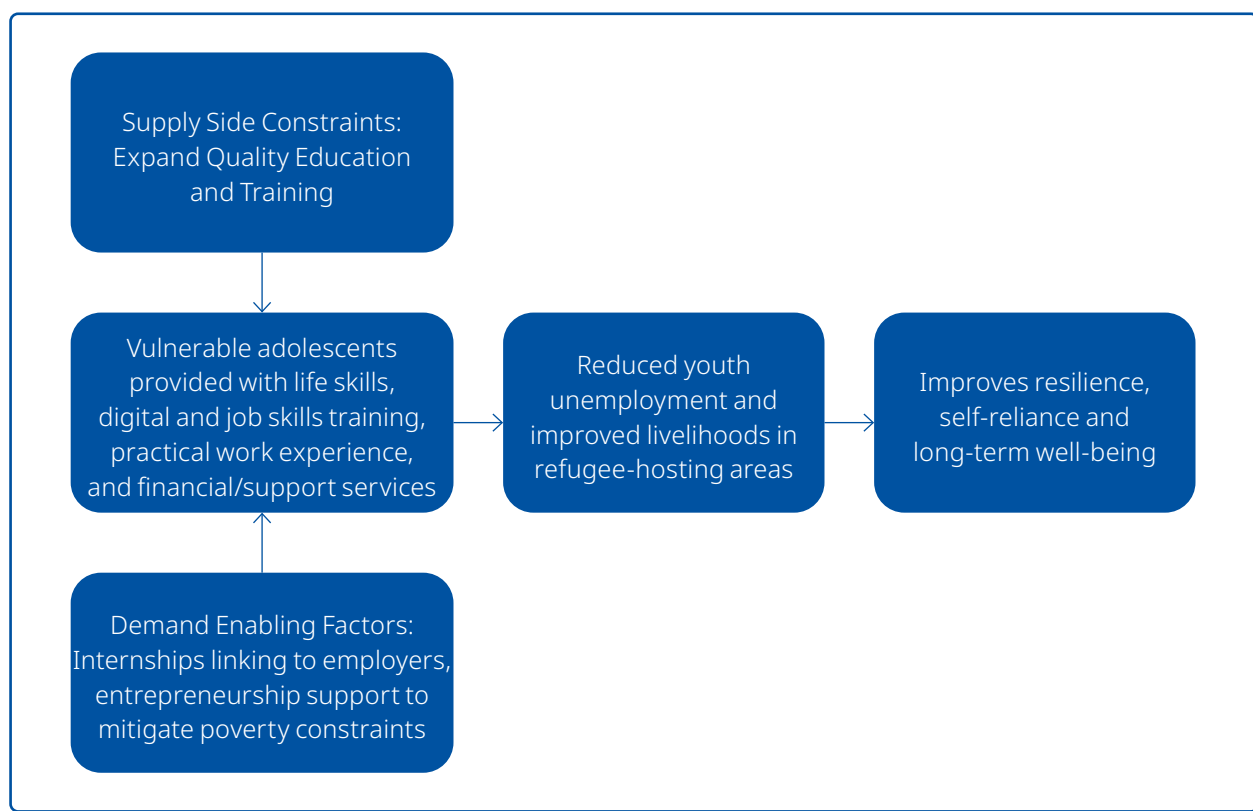
The program’s **implicit Theory of Change** can be summarised as follows: *If* vulnerable adolescents are provided with life skills (to build confidence, soft skills, and social networks), digital and job skills training (to build human capital relevant to the labour market), practical work experience (to facilitate school-to-work transition and link them with employers), and necessary financial/support services to overcome barriers, *then* they will be more likely to stay engaged in learning, complete training, and transition into gainful employment or entrepreneurship. *This is because* the intervention addresses both the supply of skills (through education and training) and the demand/



enabling factors (through internships linking to employers, entrepreneurship support to mitigate poverty constraints). Over the longer term, these outcomes are expected to contribute to reduced youth unemployment and improved livelihoods in refugee-hosting areas, aligning with national development objectives and SDG targets (e.g. SDG 4 on education and SDG 8 on decent work). The ToC assumes that local labour markets will have sufficient opportunities

(or that digital/gig economy allows remote work) to absorb skilled youth, and that complementary initiatives by other PROSPECTS partners (e.g. IFC/World Bank investments, ILO job creation programs) will bolster the demand side. It also assumes that the intervention is delivered at adequate quality and scale, and that social norms (e.g. gender norms around girls' work) do not severely impede youth from utilising their new skills.

**Figure 1.** Implicit Theory of Change Pathways



## Alignment with Uganda's Results Framework

The intervention was developed in line with the Uganda PROSPECTS Multi-Annual Country Programme (MACP 2024–2027) results framework and the Uganda UNICEF Country Programme outcomes for adolescents. In the MACP, this program corresponds to multiple output areas, including those on life skills education (MACP outputs 1.3–1.7) and socio-economic support for youth (output 3.13). It contributes to outcome indicators such as *“number of refugee and host community adolescents accessing quality skills training”* and *“percentage of trained youth gaining employment or self-employment within 12 months”*, which feature in the Uganda Integrated Results Framework 2024–2027. The program's target to reach at least 10,000 adolescents with digital training and 1,000 with internships by 2027 is directly drawn from these planning documents. By improving education completion and youth employment rates, the intervention also supports Uganda's national Education Response Plan for refugees (which

stresses secondary education and vocational training access) and the national “Skilling Uganda” strategy that seeks to boost youth entrepreneurship. In sum, there is a clear line of sight from activities to national and PROSPECTS outcomes: e.g., life skills clubs and digital courses (activities) lead to increased skills and knowledge (outputs), which lead to better educational retention and job readiness (intermediate outcomes), ultimately contributing to expanded access to education (PROSPECTS Pillar 1 outcome) and increased youth economic inclusion (Pillar 2 outcome). This alignment ensures that the evaluation will measure outcomes of high relevance to both UNICEF Uganda and the broader PROSPECTS partnership. Evidence from a recent multi-country review already indicates that such integrated approaches (combining flexible learning pathways and vocational training) can “significantly improve education and livelihoods outcomes for vulnerable populations” – a finding which this evaluation in Uganda will build upon and verify in context.

# Evaluation Questions

To assess the effectiveness, impact, and implementation of the “Youth Life and Digital Skilling” intervention in improving education, skills, and employment outcomes for refugee and host community adolescents in Uganda. The evaluation will determine the extent to which the intervention achieved its intended

results (as per the Uganda results framework and PROSPECTS ToC) by the end of 2029, and explore the reasons behind those results. It will also examine differential outcomes for sub-groups and document lessons learned for scaling and sustainability.

**Table 1.** Evaluation Questions and Sub-questions

Question	Sub-question
1. Does the program improve the long-term livelihoods and socio-economic inclusion of refugee and host community youth?	1.1 Increased participation in internships, apprenticeships, or income-generating activities
	1.2 Improved employment/self-employment in digital and green sectors
	1.3 Sustained employment and income gains
	1.4 Equitable livelihood gains for refugees and host communities
2. Does the program increase access to and participation in decent work and self-employment among targeted adolescents?	2.1 Improved life skills, digital literacy, and job-specific competencies
	2.2 Sustainability of skills outcomes across delivery modalities
	2.3 Smoother transition into vocational training/employment
3. Does the program lead to sustainable institutional change in the delivery of skills-based education and training?	3.1 Strengthened institutional capacity to deliver skilling curricula
	3.2 Sustained uptake/system integration of digital and life skills models
4. Does the program reduce structural disparities in access and outcomes across gender, disability, and displacement status?	4.1 Improved outcomes for girls and youth with disabilities
	4.2 Inclusive access to program services
	4.3 Equitable empowerment and social engagement

## 04

# Evaluation Design

## Key Considerations for Clarification Before Finalizing Evaluation Design

This section is based on currently available information about the Uganda skills programme (life skills, digital skills, and internships). However, several critical operational details remain uncertain and should be clarified with the country team before finalizing the impact evaluation design. These include fundamental questions about how and where the intervention will be delivered, who exactly will participate, and the timing of activities and expected outcomes. Addressing the following points will ensure the evaluation design is appropriately tailored:

- ▶ **Community-level delivery and integration of components:** The intervention is said to be delivered at the community level, combining life skills education, digital skills training, entrepreneurial support, and internships (on-the-job learning). It needs confirmation whether all these components will be delivered as one integrated package in each target community, or if certain components roll out separately (e.g. digital skills training at youth hubs vs. life skills in schools). Clarification is needed on whether every participating community (refugee settlement or host community area) will receive the full suite of activities simultaneously, or if different program components are introduced in phases or only in specific settings.
  - ▶ Understanding the co-location and bundling of program elements is important part of the design, since the evaluation must be designed around the actual implementation model – for example, if life skills and digital training always go hand-in-hand for the same youth, the evaluation will treat them as a unified intervention; if not, the design may need to account for it.
- ▶ **Target group overlap between refugees and host communities:** The programme targets both displaced youth (refugees) and host community youth. A key question is whether refugee and host community adolescents will be served together or through separate channels. Are the life skills clubs, digital training classes, and internship placements intentionally mixing refugees and hosts, or are there distinct cohorts for each group? Additionally, are there specific quotas or targeting criteria that differ by displacement status?
  - ▶ Clarifying any divergence or overlap in target groups will inform the sampling strategy and analysis plan. The evaluation will need to ensure comparability between the treatment and control groups in terms of refugee-host composition; if one group were to consist mostly of refugees and the other hosts, differences in outcomes might reflect demographic factors rather than program impact.
  - ▶ Understanding how participants are selected (e.g. vulnerability criteria, gender balance, age range) and whether refugees and host community youth have equal access will also help interpret results and ensure the evaluation addresses equity considerations (for example, examining if impacts differ for refugees vs. host youth).
- ▶ **Duration and intensity of training/support:** The length of each component of the intervention must be clarified. How long will adolescents participate in the life skills curriculum (e.g. a certain number of sessions or months)? How intensive is the digital skills training (hours per week and total duration)? And how long are the internship placements for those who receive them (several weeks,

months, or ongoing part-time apprenticeships)? Additionally, will these components run concurrently or sequentially for each cohort?

- ▶ The intervention length and scheduling will determine when outcomes are expected to materialize. If life skills training is front-loaded and internships come later, immediate outcomes (like improved psychosocial skills or ICT literacy) might be observable early on, whereas employment or income changes might only occur after the internship phase.
- ▶ Clear information on the duration of support each participant receives is crucial for timing the evaluation activities – it will guide when to conduct a midline survey (to capture intermediate outcomes, perhaps at the end of training periods) and when to conduct the endline (to capture longer-term outcomes, such as job placement or sustained livelihood improvements after internships).
- ▶ **Geographic coverage and rollout plan:** The program’s geographic spread and implementation schedule need to be confirmed. In Uganda, the intervention is focused on refugee-hosting areas; it will be important to know how many communities, schools, or centers are involved and where they are located. Just as importantly, will the rollout be staggered or simultaneous? A phased rollout (where some communities start earlier and others later) is often influenced by logistical or funding constraints, and if such phasing is planned, it can be leveraged for evaluation purposes.
  - ▶ Clarifying the geographic scale (e.g. targeting X districts or settlements) and whether implementation will occur in waves will directly inform the evaluation design. If a randomized phase-in (cluster-randomized rollout) is possible, the evaluation can adopt a robust experimental design. Conversely, if all target communities are set to receive the intervention at the same time (full coverage immediately), then identifying a comparison group becomes more challenging and the design would need to rely on either external comparisons or pre-post analyses.

▶ **Timing and evolution of expected impacts:** The theory of change likely anticipates a sequence in which short-term outcomes (like increased knowledge, self-confidence, digital literacy, or changed attitudes toward employment) precede longer-term outcomes (such as higher educational attainment, job placement, or income gains). We need clarity on what outcomes are realistically expected during the program versus after completion.

- ▶ Understanding the anticipated timeline of impacts will guide the evaluation’s measurement strategy. If employment is constrained by factors beyond youths’ skills (e.g. limited job availability in the local economy), the program’s ultimate goals (gainful employment, increased earnings) may not materialize until well after the intervention or not without complementary efforts.

It is important to situate this intervention in the broader evidence from similar skills training and youth employment programs. Experience in comparable humanitarian-development contexts suggests that standalone skills training or TVET initiatives often yield modest or even null impacts on employment and earnings. This is frequently because young people’s unemployment is driven more by a lack of job opportunities (weak labour demand) than by a lack of skills – even if training improves individuals’ capabilities, it may not translate into jobs if the economy cannot absorb them. Acknowledging these realities is important and the evaluation should be designed with realistic expectations and a plan to capture nuanced outcomes.

The potential evaluation design options outlined below, are seemingly feasible options with the information available at this moment. They may need to be refined or even revised once the information gaps and operational details are available, favouring whenever feasible and retaining good quality designs, those design options with lighter requirements in data collection. The designs recommended should use counterfactuals, as this element is critical in the assessment of causality in the changes observed (or not observed). Nevertheless, the following designs are not always the best ideal options in this sense, but just feasible options with the information at hand.

Even if rigorous quantitative impact estimates prove challenging, the evaluation can still add significant value through mixed-methods and a long-view qualitative approach. Documenting the program's implementation process and the longer lifecycle of support – from training through to any eventual livelihood activities – will provide rich insights. In other words, a strong qualitative component (e.g. in-depth interviews with youth over time, case studies of successes and challenges) can illuminate how the program influences young people's trajectories, even if measurable economic impacts are slow to emerge. By clarifying the above operational questions and keeping in mind the contextual challenges, we can ensure the evaluation is both feasible and capable of capturing the meaningful changes (even subtle ones) that this comprehensive skills intervention may bring about for refugee and host community adolescents in Uganda.

For instance, in evaluations where establishing a rigorous counterfactual is challenging—such as in retrospective or pre-post designs without a comparison group—a carefully embedded qualitative component can substantially strengthen the credibility and explanatory power of the evaluation. Qualitative Impact Protocol (QUIP) is an increasingly adopted and

well-documented approach to assessing contribution, particularly in development evaluations where experimental designs are not feasible. It is recognised within the evaluation community for its structured, bias-reducing methodology and alignment with causal contribution analysis.<sup>3</sup> When applied alongside a pre-post quantitative survey (particularly where sample size or comparability constraints limit causal inference), QUIP provides a complementary lens to assess contribution—enabling evaluators to assess not just whether change occurred, but why and how it occurred, and the relative role of the programme compared to other influencing factors (e.g. community dynamics, market conditions, social networks).

For this evaluation, integrating a QUIP-informed approach could enhance the robustness of findings, especially in contexts where an RCT or well-matched comparison group is not viable. It would enable triangulation of evidence and help validate or question the extent to which observed changes in youth outcomes—such as increased digital skills use, confidence, or engagement in livelihoods—can be plausibly linked to programme activities. This layered methodology ensures the evaluation maintains credible standards of rigour even under practical constraints.

## Overall Design

The evaluation will employ a mixed-methods, quasi-experimental design to rigorously assess the Uganda Youth Life and Digital Skilling programme. This approach combines quantitative impact evaluation with qualitative inquiry, capturing both the magnitude of effects and the mechanisms behind observed changes. It is well-suited to the programme's multi-component, community-based structure: quantitative methods will establish what works through statistical impact estimation, while qualitative methods (e.g. interviews, focus groups, case studies) will explore how and why it works by examining implementation processes, contextual factors, and pathways of

change. By triangulating these methods, the design not only ensures strong internal validity for attribution of outcomes to the programme, but also yields contextual insights to inform policy and practice.

**3** Developed by Bath Social and Development Research Ltd., QUIP is a structured approach that uses blindfolded interviews (interviewers unaware of the intervention) and causal contribution analysis to trace pathways of change from the perspective of programme participants, without leading or prompting them toward specific attribution. By avoiding reference to the intervention in initial questioning, QUIP elicits spontaneous accounts of changes in participants' lives and the factors driving them. These narratives are then analysed using a transparent coding framework to identify causal mechanisms and test for contribution against the programme's theory of change. QUIP differs from standard qualitative approaches in its discipline and intentionality: rather than open-ended inquiry alone, it employs rigorous methodological controls to reduce confirmation bias, support analytical transparency, and surface unintended as well as intended outcomes.

## Quantitative Evaluation Designs

To rigorously measure the programme’s impact while accommodating real-world constraints, we propose two complementary quantitative evaluation designs: (i) a cluster Randomized Controlled Trial (RCT) as the primary option, and (ii) a retrospective matching approach as an alternative. The RCT offers the highest level of methodological rigor by randomly assigning the Youth Life and Digital Skilling intervention to some communities (or schools/centers) and not to others, thereby creating a clear counterfactual for comparison. This design’s strength lies in its internal validity – any differences in outcomes between the treatment and control groups can be confidently attributed to the programme. However, it hinges on several key assumptions: that programme coverage can be partial or staggered (i.e. not every eligible community is reached immediately), that a baseline survey can be conducted before implementation begins, and that stakeholders consent to the randomization process as ethically and politically acceptable. It also requires practical safeguards like ensuring control communities are sufficiently distinct to avoid spillovers, so that contamination of the control group is minimized. Under these conditions, a cluster-RCT with baseline and endline measurements would provide robust evidence of impact, directly aligning with PROSPECTS’ emphasis on learning “what works” through rigorous evaluation.

If an experimental rollout is not feasible – for instance, if the intervention is deployed simultaneously in all target areas or if random assignment is not acceptable – the evaluation will adopt the retrospective matching design. The Retrospective Matching is presented as a last resort review option. It is considered a design of last resort when prospective methods are not possible, as its causal inference strength is significantly lower. The absence of a true baseline and randomization means we have to make stronger assumptions. Recall bias is a concern: respondents may not accurately remember past conditions, which could blur true changes. Unobserved differences can also remain – we might not capture all how program communities differed from non-program ones initially. As a result, impact estimates from this design are interpreted more cautiously, indicating potential effects rather than definitive proof.

Each of these approaches is described below in detail, with particular attention to their strengths, assumptions, and feasibility in the Ethiopian context.

### Primary Design – Randomized Controlled Trial (RCT)

Given the intent to deliver the Uganda adolescent skills program at the community level, the primary evaluation design proposed is a cluster-based Randomized Controlled Trial. This design assumes it is feasible and acceptable to randomly assign some communities to receive the intervention early while others serve as a comparison group for a period of time. By randomising rollout to communities, we ensure that any differences in outcomes between participants and non-participants can be attributed to the program with a high degree of confidence. The RCT would operate as a phased or “wait-list” trial: all target communities would eventually receive the program, but the order and timing of implementation is randomized. This approach not only addresses ethical considerations (since no community is permanently denied the intervention) but also creates a clear window for impact evaluation. The structure of the RCT would mirror that of similar evaluations in other PROSPECTS countries (e.g. Ethiopia, Egypt), adapted to Uganda’s context:

**Baseline (Pre-intervention):** All selected communities – both those slated for immediate intervention and those designated as initial controls – would be surveyed before the program begins. This baseline establishes the pre-program conditions of adolescents and their communities, capturing metrics such as demographic characteristics, education level, prior training or work experience, and baseline measures of skills (using standardized life skills and digital literacy assessments), as well as initial outcomes like current schooling status, employment situation (if out of school), and aspirations. Crucially, the baseline allows us to verify that the randomized groups are comparable on key characteristics. It also provides a point of reference to measure change over time.

**Year 1 Implementation (Treatment vs. Control):** Immediately following the baseline, the program will be launched in the treatment communities, while the other half (the control communities) will temporarily not receive the new intervention. Treatment communities

will begin delivering the full package of activities. Control communities during this phase continue “business as usual”. This creates a clear contrast between areas with the program and those without it for the duration of Year 1. After roughly one year of implementation, a midline survey will be conducted to measure short-term outcomes. At this midline point, we expect to capture intermediate results such as improvements in knowledge and skills (e.g. scores on life skills and digital literacy tests, IT proficiency), changes in attitudes and aspirations (confidence, job seeking self-efficacy, higher educational aspirations), and possibly early indicators of behavioural change (for example, increased school attendance or retention for those in school, or greater engagement in job searching or entrepreneurship for those who finished school). By comparing midline outcomes between treatment and control groups, we can estimate the program’s short-term impact after one year. It’s worth noting that by the one-year mark, many participants in treatment communities will have completed the training components and some may have entered internships or started income-generating activities, whereas control community youth would not have had those new opportunities – thus we might even see initial differences in employment rates or livelihoods engagement, though likely modest in magnitude at this stage.

**Endline (Year 2):** After midline data collection, the intervention could be introduced to control communities using a wait-list design. At the two-year mark, an endline survey will re-assess outcomes for all communities, comparing groups based on their duration of exposure. Key outcomes measured will include educational progression, employment indicators, and psychosocial well-being. This phased approach allows the evaluation to examine sustained impacts and whether initial advantages in early-treated communities are maintained over time, while also assessing if later-treated communities catch up once exposed. Alternatively, control communities could remain untreated throughout the evaluation period for a pure control-treatment comparison, though the wait-list design described here is likely preferable from an ethical and operational standpoint.

The validity of the RCT hinges on the assumption of no contamination between treatment and control areas – that is, youth in control communities should not inadvertently receive similar training or support during the control period, and those in treatment communities should indeed receive the intended services. In a community-level skills intervention, possible threats to this assumption could include: participants crossing over (e.g. a motivated youth from a control village traveling to a neighbouring treatment village’s training sessions), or the program expanding services to control sites earlier than planned. The evaluation team should mitigate these risks by establishing clear geographic and administrative boundaries for the intervention. Communication protocols should be put in place so that all stakeholders understand the implementation plan. Maintaining fidelity to the randomized design is crucial for unbiased impact estimates. The team may also consider physical spacing – selecting control communities that are not immediate neighbours of treatment communities – to reduce the likelihood of spillover.

### Sample Size and Power

The cluster RCT approach will inform the required sample size in terms of number of communities and households. In general, to detect moderate impacts on household-level outcomes, we would aim for a substantial number of clusters per arm to improve precision. For illustrative purposes, suppose the pilot could engage on the order of 80 communities in total. These might be split into roughly 40 treatment and 40 control communities. If we plan to survey, say, 20–30 households per community, this yields approximately 800–1,200 households in each arm, which should be sufficient to detect an effect size of practical interest.<sup>4</sup>

<sup>4</sup> These numbers should be refined with formal power calculations once more details are known about outcome variability and ICC in the target population. Key parameters for the power analysis include the expected impact, the ICC, and the baseline-to-endline correlation of outcomes. If the ICC is low and outcomes are volatile, fewer clusters might suffice; if ICC is high or anticipated impacts are small, more clusters or households would be needed.

**Figure 2.** Randomized Controlled Trial

**Key Assumptions and Limitations:** The success of this RCT design rests on several assumptions. First, it assumes that random assignment is feasible both politically and ethically – i.e. that stakeholders agree some communities will start later and that this delay is acceptable in the short run. Close collaboration with government and community leaders is needed to manage expectations and explain the value of the phased approach. Second, the design assumes no major external shocks or policy changes will differentially affect treatment or control communities during the study window; unforeseen events (e.g. refugee influx surges, economic crises, or new similar youth programs introduced by other agencies) could complicate the attribution of impact. Third, we assume that attrition (loss of respondents by endline) will be low and similar across groups – since the target population is adolescents and young adults, there is some risk of migration or drop-out, so the evaluation team should plan measures to track participants over time. In terms of limitations, one is that the two-year timeframe may be short to capture some ultimate outcomes like stable employment or income gains – many youths might still be continuing education or only recently completed internships by the endline. Thus, the evaluation might understate long-term impacts if they only materialize after several years. The evaluation team can mitigate this by focusing on intermediate indicators (like improved skills, self-employment initiatives started, etc.) as proxies for longer-run potential. Despite these considerations, a cluster RCT – with robust implementation monitoring and a mixed-methods complement – is the strongest design to provide credible evidence on the program’s effectiveness.

### Alternative Option 1 – Retrospective Matching

If a RCT is not feasible—such as when the program has already been implemented at scale or when a baseline was not collected—an ex post impact evaluation using a retrospective matching design remains a possible option. In this approach, individuals who participated in the program are identified and matched, after the fact, to non-participants with similar observable characteristics. The intention is to simulate the counterfactual by constructing a statistically comparable group, allowing us to estimate the program’s impact. Typically, this design involves a single round of data collection conducted after program implementation, drawing on participant recall or available records to reconstruct baseline conditions. While this method is inherently less robust than prospective designs, it provides a practical and analytically sound alternative when randomization and baseline data are unavailable.

The retrospective matching approach depends on several assumptions. Most critically, it assumes that all important differences between participants and non-participants are observable and can be accounted for in the data. Because we cannot rely on prospective baseline data, we must assume that any unobserved differences—such as motivation, family support, or personal drive—are either negligible or correlated with observed traits. The design also assumes that an untreated comparison group is available and sufficiently similar to those who received the intervention. For instance, youth from comparable communities or districts who did not participate in the program may serve as controls. Another key

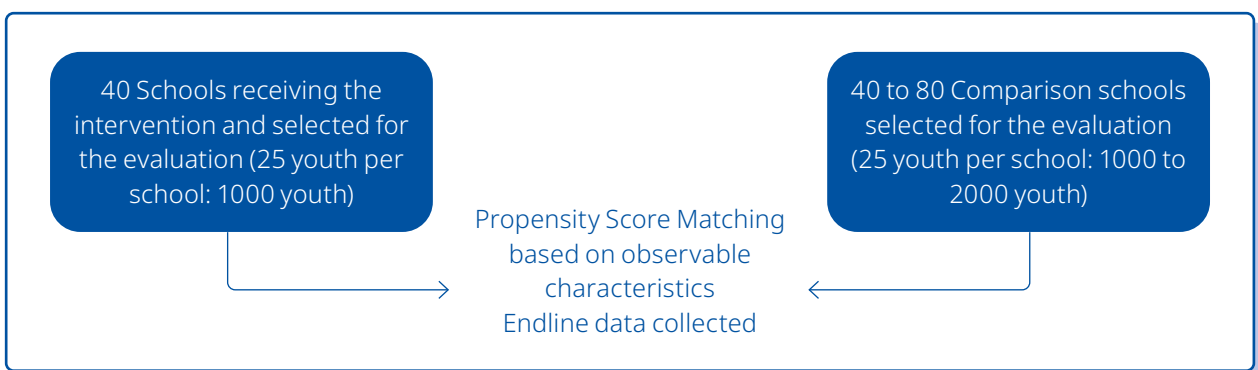
assumption is that both participants and non-participants can accurately report pre-program characteristics such as educational attainment or employment history. While recall bias is a known risk—people often remember past conditions as more similar to the present, thereby underestimating change—reasonable accuracy can be achieved through careful survey design and triangulation. Lastly, it is assumed that data quality will be high in the post-program survey, since there will be no baseline to fall back on.

Our preliminary sampling strategy starts by obtaining the full list of youth who participated in the intervention, potentially 10,000 individuals in Phase II, and selecting a representative sample for the survey, possibly around 1,000 participants, depending on resources. This sample would be stratified by relevant characteristics such as gender, age, refugee or host status, and type of training received. The comparison group would then be sampled from non-participants in the same or similar communities. This could involve identifying untreated youth from within the same geographic areas or, if saturation is high, from comparable non-program districts. Matching will be conducted on observable characteristics, likely using propensity score or stratified matching approaches. These characteristics may include age, sex, educational attainment, location, and other socio-demographic indicators. If feasible, we will also incorporate proxies for baseline status, such as the highest grade completed or pre-program employment, as reported through recall or records, to ensure that matched controls were not systematically better or worse off prior to the intervention.

The counterfactual will be constructed through statistical matching techniques that pair each treated individual with one or more non-participants who have a similar estimated probability of program participation. The objective is to replicate, post hoc, the balance of a randomized design. Depending on data quality and available covariates, we may use caliper matching or exact matching on key variables such as refugee status to strengthen the validity of comparisons. If reliable recalled baseline data can be gathered, we may explore a difference-in-differences specification as a robustness check, though the core estimates will be derived from matched comparisons. As with any non-experimental design, unobservable differences between participants and controls may remain, and these limitations will be transparently acknowledged in interpreting results.

Given the lower statistical power of a design without baseline or panel data, larger sample sizes are typically required. Subject to budget constraints, we propose surveying approximately 1,000 treated and up to 2,000 non-treated individuals, anticipating that not all controls will yield good matches. This approach improves the likelihood of generating high-quality matches and allows for one-to-many matching where appropriate. Power calculations will guide the final sample size, but preliminary estimates suggest this scale should detect moderate program effects with sufficient precision. Additional sampling may be required if subgroup analysis is planned—for example, to estimate separate effects for refugee and host community youth.

**Figure 3.** Retrospective Matching



## Summary of Research Designs

Each of the three approaches balances methodological rigor with field feasibility. The evaluation is designed to be adaptive – implementing the strongest possible design where conditions allow but prepared

to shift to alternative methods in other contexts. Table 1 summarizes the core features of the primary and alternative designs, highlighting their requirements and best-use scenarios:

**Table 2.** Evaluation Design Options

	Design A: RCT	Design C: Retrospective Matching
<b>Randomization</b>	Yes – cluster-level ()	No
<b>Baseline Data</b>	Yes	No
<b>Matching Timing</b>	Prospective – comparison groups defined before implementation	After implementation
<b>Matching Method</b>	Random assignment ensures balance	Statistical matching on reconstructed baseline
<b>Causal Inference Strength</b>	<b>High</b> – Strong internal validity due to randomization	<b>Low-Moderate</b> – Indicative results; higher uncertainty
<b>Best Use Case</b>	Randomization is feasible and ethical, and baseline data can be collected.	Programme already implemented without baseline

## Key outcome indicators

To answer the above questions, the evaluation will track various **indicators** at outcome and output level, using mixed methods for data collection.

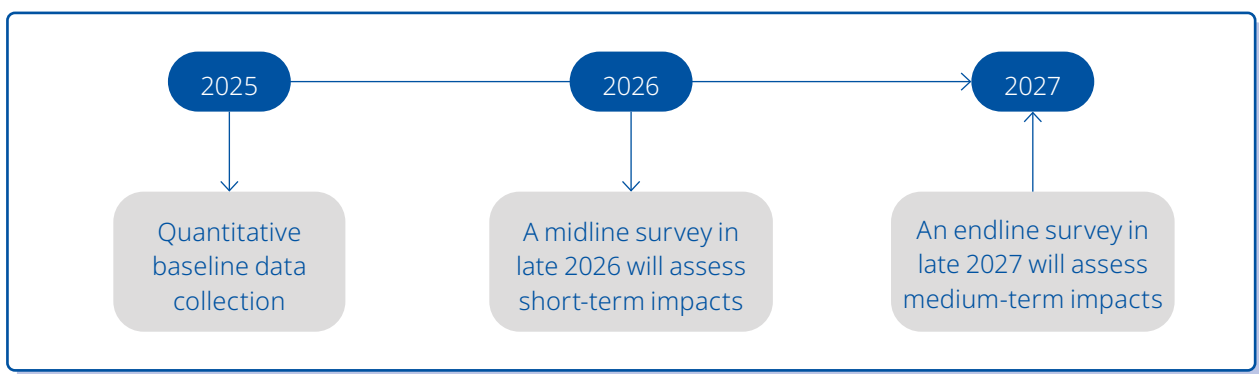
Question	PROSPECTS Indicator	Programmatic Indicator
1. Does the program improve the long-term livelihoods and socio-economic inclusion of refugee and host community youth?	1.4.1: # youth accessing internship/apprenticeship/job placement opportunities	% of youth participating in livelihood activity using program-acquired skills
	2.1.3: % youth with access to decent work	% of youth in digital/green sector jobs or businesses
	2.1.4: % youth reporting sustained earnings post-program	% with diversified income sources (e.g. wage + self-employment)
	All above indicators disaggregated by status	% of youth involved in household livelihood decision-making
2. Does the program increase access to and participation in decent work and self-employment among targeted adolescents?	1.2.1: % demonstrating improved life/digital skills	Average skill score improvement (pre-post assessment)
	1.2.1 (disaggregated)	% youth reporting confidence in job-seeking or entrepreneurship
	1.3.1: Transition rate to TVET/employment	% youth who received job offers or launched enterprises
3. Does the program lead to sustainable institutional change in the delivery of skills-based education and training?	1.1.2: # institutions implementing adapted curricula	% of trained facilitators retained and delivering training
	1.1.3: Evidence of policy/system integration	# institutions budgeting for sustained rollout
4. Does the program reduce structural disparities in access and outcomes across gender, disability, and displacement status?	1.2.2: % girls/disabled with improved competencies	Reduction in gender/disability outcome gaps (baseline to endline)
	1.1.1: % institutions with inclusive delivery models	% of targeted marginalized youth completing the program
	2.2.2: % youth in civic/leadership platforms (by status)	# youth reporting reduced exclusion or discrimination

*Note:* The above is an illustrative selection of indicators. During the evaluation planning phase, a detailed **Indicator Matrix** will be refined, drawing on the **Uganda Integrated Results Framework (IRF)** indicators for 2024–2027. For example, IRF indicators such as “% of adolescents demonstrating improved life skills (as per standardized scale)” or “# of youth (F/M) transitioned to employment or further education” will be incorporated to ensure consistency with program monitoring. Wherever possible, **baseline values and targets** from program documents (e.g., targeting 50% female participation, 80% course completion rate, etc.) will be used to benchmark progress. Data will be disaggregated by sex, age cohort (e.g., younger adolescents 14–17 vs older youth 18–24), refugee/host, and disability status to analyse equity outcomes.

# Timeline

The evaluation will be conducted over approximately two years (2025–2027), aligned with the phased rollout of the intervention and allowing for baseline, midline, and endline measurements. This compressed timeline focuses on capturing short- to medium-term impacts within the project’s implementation period. Figure 4 illustrates the schedule, and the key phases are outlined below.

**Figure 4.** Evaluation Timeline



**Baseline (2025):** A comprehensive baseline survey will be conducted in all selected communities prior to program implementation. This baseline will capture demographics, educational history, initial life and digital skills levels, employment status, livelihood activities, and psychosocial indicators. Where possible, data collection will align with existing PROSPECTS Phase II baseline efforts by UNICEF. Baseline results will confirm comparability between treatment and control groups, inform power calculations, and provide context for program refinement and evaluation planning.

**Implementation & Midline (Year 1 – 2025/26):** Following baseline, the intervention will be rolled out in treatment communities. Implementation will be monitored through routine administrative data on program participation, training completion, internships, and support provided. Brief qualitative assessments will document fidelity of implementation and immediate feedback. After approximately one year (mid-2026), a midline survey with baseline

respondents will measure early changes in life and digital skills, attitudes, behaviours, and initial educational or employment outcomes. Midline findings will enable timely identification of impacts and inform potential adjustments to implementation.

**Endline (Year 2 – 2027):** By Year 2, control communities will also receive the intervention, ensuring all participants have at least one year of exposure. An endline survey will reassess the full cohort, measuring educational outcomes (e.g., school retention, training completion), employment and livelihoods impacts, and psychosocial indicators. Analysis will compare outcomes based on intervention duration. Qualitative research—including interviews and focus groups—will capture detailed community perspectives on program effects and contextual factors influencing outcomes. The final evaluation report will integrate quantitative and qualitative findings to provide robust evidence on the program’s overall effectiveness, informing future scaling and implementation decisions.

## Data Collection Methods

- ▶ **Quantitative Surveys:** Two main rounds (baseline and endline) of **household/individual surveys** will be administered. These will capture demographic data, schooling status, employment status, income (using recall of earnings or consumption proxies), and psychosocial indicators. The survey will incorporate standardised modules such as UNICEF's life skills measurement toolkit and digital literacy assessment items. For accuracy, surveys will be conducted face-to-face by trained enumerators, with digital data capture (tablets). Translators will be used as needed for refugee languages. Where possible, surveys will be harmonised with any existing data collection (e.g., leveraging questions from national labour force surveys or UNHCR's surveys to allow calibration/comparison). The evaluation will also collect contact information at baseline for tracking (phone numbers, alternate contacts), anticipating some migration over 5 years. A short tracer survey at 6-12 months post-program (for the employment outcomes) may be done by phone to supplement the endline, especially to catch data on those who move for jobs.
- ▶ **Administrative and Monitoring Data:** The evaluation will make use of program monitoring data including: lists of participants (with unique IDs, allowing linkage to survey data), attendance logs, completion records for training, internship placement records, and any assessment results from the program (e.g., scores on life skills quizzes or digital modules). Education Management Information System (EMIS) data from MoES (on school enrolment/retention in target districts) and labour market data from partners (e.g., ILO on job placements) will also be utilised to triangulate results or as secondary data sources. The Integrated Results Framework reports for the MACP will provide target vs. actual numbers on some indicators annually, which the evaluation can use to gauge progress and context (though not for impact, they indicate performance trends).
- ▶ **Qualitative Methods:** A rich qualitative component will accompany the quantitative evaluation. This includes:
  - ▶ **Key Informant Interviews (KII):** with stakeholders such as UNICEF program managers, implementing partner staff (trainers, youth center facilitators), government officials from MoES/MoGLSD, and PROSPECTS partner agencies (ILO, UNHCR field staff involved in complementary activities). KIIs will explore perceptions of the program's design, implementation challenges, partner coordination, and perceived outcomes.
  - ▶ **Focus Group Discussions (FGD):** with groups of participating youth (separately with young women and men to ensure open dialogue), as well as with non-participant youth in the community (to understand counterfactual experiences). The evaluation will also hold some FGDs with parents/caregivers and community leaders, to capture community-level effects (e.g., changes in attitudes towards girls working, or community cohesion between refugee and host youth).
  - ▶ **Case Studies and Testimonies:** A few in-depth case studies will be documented – e.g., a success story of a refugee girl who completed training and started a business, or a contrasting case where a youth dropped out despite support, to understand why. These narratives can illustrate mechanisms and humanise the statistical findings.

- ▶ **Direct Observation:** Evaluators will observe a sample of intervention activities (such as a life skills club session at a school, or a digital skills class in a youth center) using a structured observation checklist to assess fidelity to the curriculum and engagement level. Observation of cash distribution processes or mentorship sessions might also be included to assess quality and transparency.

All qualitative data will be recorded (with consent), transcribed, and analysed using qualitative analysis software. During analysis, triangulation between qualitative and quantitative findings will ensure a comprehensive understanding.

# Analysis Methods

## Quantitative – Option 1 (RCT)

**RCT Design:** For areas implementing the RCT, the analysis will estimate the average treatment effect by comparing outcomes across randomized community clusters with control units. The core analytical model will follow a simple OLS specification, comparing outcome changes in treatment arms relative to control communities.

The preferred regression model is:

$$Y_{it} = \alpha + \beta_1 (\text{Treatment}_i) + \gamma X_{i0} + \epsilon_{it}$$

Where:

- ▶  $Y_{it}$  is the outcome for household or individual  $i$  at time  $t$ ,
- ▶  $\text{Treatment}_i$  indicates assignment to treatment status,
- ▶  $X_{i0}$  are pre-intervention covariates, including baseline values of the outcome where available.

Standard errors will be **clustered at the level of random assignment** (i.e., community or enumeration area) to account for intra-cluster correlation. Where multiple follow-up points are available (e.g., baseline, interim, endline), fixed effects models will be used to leverage within-unit variation, improving statistical power and adjusting for unobserved time-invariant heterogeneity.

## Quantitative – Option 2 (Retrospective Matching)

In case no baseline data is collected, a retrospective matched design will reconstruct pre-intervention conditions using recall-based indicators and secondary sources (e.g., administrative data). Treated and untreated communities will be matched post hoc on pre-treatment characteristics, and post-intervention outcomes will be compared using cross-sectional regressions or adjusted comparisons. Cross-sectional regression models with extensive covariate adjustment will be used.

### Qualitative Methods

Qualitative data will be analysed thematically. The evaluation will use the evaluation questions as a guiding framework to code data (e.g., codes for “perceived outcomes – employment”, “implementation challenge – X”, “suggestions for improvement”, etc.). Patterns will be identified, and representative quotes will be used in the report to illustrate findings. The qualitative analysis will particularly inform questions 6

(implementation and context) and help explain any unexpected or nuanced results for questions 1–5. For example, if the quantitative data show a smaller impact on female employment, qualitative interviews might reveal cultural barriers or childcare responsibilities that women face, which the program didn’t fully address.

### Validation and Triangulation

The evaluation will use a triangulation matrix to compare findings from different sources for each question. For instance, if quantitative data show improved life skills scores, the evaluation will check if FGDs with youth also report greater confidence or behaviour change. If employers report high satisfaction with interns, we expect to see increased hiring of program graduates in data, etc. Any discrepancies will be investigated (perhaps indicating areas for further study or limitations).

Throughout the evaluation, interim findings (from monitoring data or midline) will be shared with program managers in a utilization-focused approach, so that evidence can inform ongoing program improvements (without waiting until 2029). However, the integrity of the impact evaluation will be maintained (e.g., not expanding to control areas early due to positive interim results, until endline data is collected, to preserve the counterfactual).

## 09

# Estimated Resources for Data Collection

This section will be further refined in collaboration with UNICEF's Uganda Country Office, as additional information becomes available regarding the number of participating communities, total beneficiary targets, and operational coverage of the program components. The full cost estimate for data collection will depend on final sample sizes, unit of analysis (household or individual), agreed number of survey waves, and geographic dispersion. Once KIIs with Country Office staff are complete, a more precise costing—covering quantitative surveys, qualitative fieldwork, and administrative data acquisition—will be provided.

## 10

# Limitations

## Potential Limitations and Mitigation

**Selection Bias:** If there is no random assignment, there's a chance that the youth in the programme are different from those in the comparison group—maybe they're more motivated or live closer to services. This can be reduced by carefully matching on observable traits and using baseline data to control for starting differences.

**Attrition:** Tracking young people over five years will be tough. Many may move for work, go back to their home countries, or relocate within Uganda. If the evaluation loses too many, or if those who drop out differ from those who stay, it could bias the results. We'll plan ahead by over-sampling, collecting detailed contact info, using social media and phones to follow up, and doing phone interviews where needed. We'll also check whether attrition introduces bias and adjust for it statistically if necessary.

**Spillover / Contamination:** There's a risk that comparison youth may be indirectly exposed to the intervention, for example, through friends participating or because another NGO runs a similar programme nearby. The team will need to keep an eye on this through fieldwork and by tracking other programmes. If it's happening, adjustments to the analysis may be necessary where possible. If the control group benefits in some way, the impact estimates will be conservative.

**Measurement Challenges:** Some outcomes—like life skills or job quality—are hard to measure. The team will have to rely on self-reports for things like income, which can be prone to recall errors or bias. To address this, they should use multiple data points (e.g. triangulate with consumption or assets), ensure interviews are private, and use simple tasks to measure skills when we can objectively. For life skills, they can use a validated index, and it may be possible to verify employment outcomes through follow-up with employers or community validation when feasible.

**Contextual Changes:** Over five years, things will change—economically, politically, or due to shocks like COVID-19. These shifts could influence outcomes in ways that have nothing to do with the programme. Our design accounts for general trends across both groups, but if a shock hits one group more than the other, attribution gets tricky. The team should document key events and, where needed, adjust the analysis—or at least be upfront about how external factors may have influenced results.

**Complexity of the Intervention:** This is a multi-component programme, so it's hard to know which part is driving results—unless we've designed for that. The design focuses on the overall package. If employment improves, it'll attribute it to the integrated approach, rather than trying to isolate one element. That said, qualitative interviews will give insight into what participants found most useful. The team could consider a factorial design to isolate effects, but that's often not feasible in these settings.

# Ethical Considerations

This evaluation will adhere strictly to **UNICEF’s ethical guidelines for research and evaluation** and international standards for conducting research with human subjects, especially children and vulnerable populations. Key ethical considerations include:

- ▶ **Informed Consent:** Participation in the study (surveys, interviews, etc.) will be fully voluntary. For respondents under 18 years old (since the intervention targets adolescents, some will be minors), the evaluation will obtain *informed assent* from the youth and *consent from a parent or guardian*. For those 18 and above, their own informed consent is sufficient. Consent forms will explain in the appropriate language the purpose of the evaluation, the procedures, potential risks and benefits, and the respondent’s rights (including the right to refuse or withdraw at any time without any consequences on their involvement in the program). Special care will be taken to assure participants that declining to participate in the evaluation will not affect their access to the intervention services or any aid – to avoid any coercion.
- ▶ **Do No Harm:** The evaluation instruments will be designed to minimize any risk of harm or distress. Questions will be phrased sensitively, avoiding any triggering language especially for vulnerable youth (some may have trauma backgrounds). Enumerators and facilitators will be trained on child protection and gender-sensitive approaches. If during the course of data collection a respondent shows signs of distress or reveals a serious issue (e.g., abuse, self-harm thoughts), the team will have protocols to respond – such as pausing the interview, and making referrals to professional support services (UNICEF and partners have protection mechanisms in these communities). The evaluation will coordinate with the program’s safeguarding officers to handle any such cases. Discussions
- ▶ **Privacy and Confidentiality:** The evaluation will ensure strict confidentiality of respondents’ data. All survey and interview responses will be anonymized – identified by codes rather than names. Personal identifiers collected for panel tracking will be stored securely and separately from survey responses. Data will be encrypted and access limited to the core evaluation team. In reporting, no names or identifying details will be revealed; any quotes used will be generic or attributed to anonymized descriptors (e.g., “17-year-old female participant from Yumbe district”). Prior to stakeholder workshops or dissemination, the evaluation will also seek consent if we plan to share any photos or video documentation (though primarily this evaluation uses written data). All data handling will comply with relevant data protection laws and UNICEF’s data privacy standards.
- ▶ **Ethical Approval:** The evaluation will obtain ethical clearance from a recognized Institutional Review Board (IRB) in Uganda (for example, a university IRB or the Uganda National Council for Science and Technology if required for research with refugees) as well as from UNICEF’s own Ethics Review Panel. The research protocols (survey instruments, consent scripts, etc.) will be submitted for review to ensure they meet ethical standards. Given that refugees are a potentially vulnerable population, the IRB will scrutinize plans to ensure no exploitation or undue intrusion into their lives. The evaluation will also coordinate

with UNHCR for approval to conduct research in refugee settlements, following their guidelines for research with persons of concern.

- ▶ **Equitable Treatment and Preventing Exploitation:** In randomized components like the Cash Plus RCT, ethical justification is needed for withholding an intervention from a control group. The evaluation will employ “equipoise” – i.e., there is genuine uncertainty about the effectiveness of the new approach (cash plus), which justifies testing it. Additionally, the evaluation will use wait-list controls where possible, meaning the control group can receive the intervention later or receive an alternative support if proven effective, ensuring they are not permanently denied benefits. All participants (treatment or control) will continue to receive any standard services available; no one will get less than the status quo. The evaluation will transparently communicate this to community leaders and participants to maintain trust.
- ▶ **Cultural Sensitivity:** The evaluation team will be diverse and include members who speak the local languages (e.g., Kiswahili, Arabic, local dialects in West Nile) and understand the cultural norms of both refugee communities (South Sudanese, Congolese, etc.) and Ugandan host communities. Instruments will be translated and back-translated to ensure meaning is preserved. The evaluation will involve community leaders or youth representatives in reviewing the questions for cultural appropriateness. For example, concepts of “life skills” or certain sensitive questions will be vetted so as not to offend or confuse. By engaging local enumerators and perhaps youth peer researchers (trained youth from the community), we empower the community and ease any mistrust.

- ▶ **Avoiding Conflict of Interest:** The evaluation team will operate independently of the program implementation team. While the evaluation will collaborate for data access and logistics, an independent evaluation consultant firm (procured through UNICEF’s Evaluation Office) will lead the study to ensure impartiality. All findings – positive or negative – will be reported objectively. The evaluation will establish a reference group including stakeholders (government, donors, youth reps) to oversee the evaluation process transparently.

- ▶ **Beneficence and Feedback:** The evaluation is designed not just to extract data, but to ultimately benefit the communities by improving programs. The evaluation will practice reciprocity by feeding back results to the communities and participants in an accessible format (e.g., community meetings or brief summaries in local language) after the study, so they can see what was learned and how it will be used. This respects participants’ contribution and ensures accountability.

The evaluation will protect participants’ rights and welfare throughout the study by upholding these ethical standards. Ethical diligence is particularly paramount given that many participants are minors and refugees who have experienced vulnerabilities. The evaluation’s conduct will reflect UNICEF’s core principles of humanity, respect, and integrity.

