PlayMatters

Implementation Research of PlayMatters 1.0

Baseline Report for ECD and Primary School

PlayMatters Project

The PlayMatters Consortium led by the International Rescue Committee, and includes Plan International, War Child Holland, Innovations for Poverty Action, the Behavioral Insights Team in partnership with the **LEGO Foundation.**

2023

Prepared by:

Idalia Rodriguez Morales, Education Researcher, IRC – Airbel Impact Lab
Furaha Abwe, Research Coordinator, IRC Tanzania
Joel S. Mtebe, Associate Professor, University of Dar es Salaam
Christina Raphael, Senior Lecturer, University of Dar es Salaam
Silvia Diazgranados Ferráns, Director, Education Research, IRC – Airbel Impact Lab

Tanzania

In partnership with

PlayMatters seeks to improve holistic learning outcomes and well-being for **800,000** refugee and host community children ages **3-12+** who live in refugee and host community contexts in **Ethiopia, Uganda and Tanzania** using **Learning through Play** methodologies.

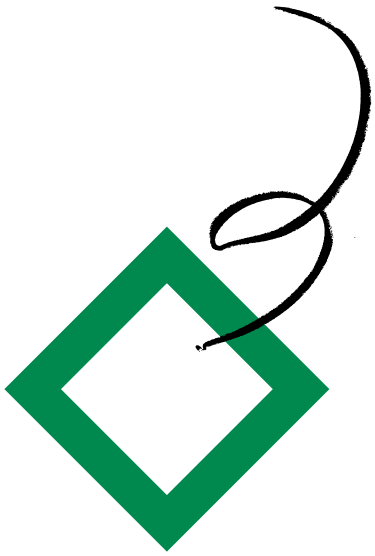
Acknowledgment

The authors of this study want to acknowledge all of the participants in this study, including the children, teachers, and Head Teachers.
We would also like to thank members of the PlayMatters Research, Monitoring, Evaluation and Learning team.

Contents

Acronyms/Abbreviations

DEAS	Developmental and Educational Activities Scale
ECCE	Early Childhood Care and Education
ECD	Early Childhood Development
GER	Gross Enrolment Rate
PM	PlayMatters
IDELA	International Development and Early Learning Assessment
IRB	Institutional Review Board
IRC	International Rescue Committee
LtP	Learning through Play
MoE	Ministry of Education
NGO	Non-governmental Organization
PHRP	Protecting Human Research Participants
PI	Principal Investigator
SEL	Socio-emotional Learning
SERAIS	Socio-Emotional Response and Information Scenarios
TIPPS	Teacher Instructional Practices and Processes System
TPD	Teacher Professional development
UNESCO	United Nations Educational, Scientific and Cultural Organization
UNHCR	United Nations Refugee Agency
UWES	Utrecht Work Engagement Scale



Executive Summary

Learning through Play (LtP) is built on the premise that play allows children to set goals, ignore obstacles, and strive for outcomes. There is growing evidence that play enhances child holistic development defined by cognitive, social, physical, emotional, and creative aspects. In conflict contexts or emergency situations, play improves self-regulation because of its link with accepting losses, managing conflicts, socializing children, and developing hope, a critical ingredient to coping and resilience.

Study Purpose and Audience

The purpose of this report is to describe the baseline status of children and educators involved in PlayMatters (PM) in Tanzania. The PM project is a Teacher Professional Development (TPD) program that works through existing education systems to provide Early Childhood Development (ECD) and primary teachers in Tanzania, Ethiopia and Uganda with the skills, motivation, and resources to integrate contextually relevant play-based methods into their teaching practice, which is understood and referred to as Learning through Play (LtP) approaches. In 2022, PM in Tanzania intended to define initial characteristics of children and educators in PlayMatters intervention refugee and host community to help guide learning and details of the subsequent interventions. The present report presents findings from PM's baseline data collection for Tanzania in February/March 2022. The main audiences for the study are: 1) PM Tanzania and Regional teams, 2) The LEGO Foundation, and 3) and stakeholders in the Tanzania government.

Study Questions

The baseline assessment of the initial characteristics of children and educators engaged in PM intervention both in refugee and host communities, is guided by the following questions:

I. Child Level

1. What are children's literacy and numeracy skills? How do they vary by displacement, sex, and region?

2. What is the status of children's socio-emotional skills and wellbeing? How do they vary by displacement, sex, and region?

3. What are the child characteristics explaining the variation in children's literacy, numeracy, and socio emotional (SEL) skills?

II. Educator Level

1. What is the perception of educators on LtP?

2. What are educators' instructional practices?

3. What are educators' work engagement, occupational wellbeing, and self-efficacy levels?

Data and Methods

The study employed systematic sampling technique to identify educators and children from the PM intervention in both refugee and host communities. Children data included the IDELA to measure early development, including literacy and numeracy skills, and Kiddy-KINDL to measure children's wellbeing. Children data collected included the Early Grade Reading and Mathematics Assessments (EGRA and EGMA) to measure learning outcomes, and multiple tools to measure socioemotional skills, including Kiddy-KINDL to measure wellbeing, a scale to measure empathy, the ACES scale to measure emotional attribution accuracy, a tool to measure bullying and victimization, and the Socio-Emotional Response and Information Scenarios (SERAIS) tool to measure hostile attribution bias, emotional regulation, and conflict resolution.

Educator data, for both primary and ECD, included a survey on their perceptions about play in the classroom, the Developmental and Educational Activities Scale (DEAS) to inquire about their teaching skills and practices, and scales to measure work engagement (Utrecht Work Engagement Scale/UWES), and self-efficacy levels. Educator data also included an instructional observation of a class thought by the educator using the Teacher Instructional Practices and Processes System (TIPPS) tool.

Major Findings

Q1 - Children's Literacy and Numeracy

1. IDELA and EGRA and EGMA, both overall and domain-specific mean scores showed low performance levels, and varying differences by sex, community type, and regions.

a. Swahili ECD learners in the refugee community score better than students in the host community in five of the six IDELA domains, a gap that is statistically significant in four domains (Emergent Literacy, Socioemotional Skills, Executive Function, and Approaches to Learning) as well as for the overall score, with Socioemotional Skills and the Executive Function domains displaying the largest differences.

b. Overall, results suggest that young children are yet developing their foundational skills: both host and refugee communities scored higher in Fine motor, followed by Executive Function, and Socioemotional skills, with higher order skills (Emergent Numeracy and Literacy) lagging with learners scoring, on average, less than one third of the domains correctly.

c. Differences by sex at the community aggregate, indicates that boy refugees have stronger skills than girl refugees, particularly for the Fine motor domain, Approaches to Learning, and the Overall IDELA score. None of the differences among girls and boys for IDELA in the host community were statistically significant.

d. For EGRA, the Swahili refugee and host learners perform similarly for the foundational skill of letter identification, but host learners perform better than refugee learners for both oral reading fluency and

reading comprehension. Kirundi refugee learners display considerably better performance reading comprehension skills than the Swahili refugee learners.

e. For EGMA's number identification, the percentage of host learners who performed at or above performance category outperforms the percentage of refugee learners. However, for the rest of the EGMA subtasks, host and refugee learners scored similar in terms of performance categories.

Q2 - Children's Socio-Emotional SKills and Wellbeing

1. For both ECD and Primary, results suggest that children have moderately favorable views of their wellbeing. For ECD, disaggregation analyses indicated better positive views of girls and host community children, though differences were not statistically significant. For primary, refugee Swahili learners reported higher wellbeing than their host Swahili learners and refugee Kirundi learners reporting similar wellbeing to the Swahili refugee learners.

2. For primary, results suggest that children have healthy emotional attribution skills, for host and refugee Swahili learners and refugee Kirundi learners. We find no statistically significant differences between girls and host community children.

3. We find that primary Swahili host learners report significantly higher attitudes of upstand perpetrators than host learners and that while both and refugee learners report similar levels of exposure to bullying, refugee learners report significantly higher victimization rates than host learners.

4. Swahili primary learners host and refugee learners report similar levels of exposure to bullying, refugee learners report significantly higher victimization rates than host learners. However, host learners also display a stronger emotional orientation towards calmness, compared with the refugee learners; and refugee learners display stronger emotional orientation towards sadness compared with the host learners. Refugee learners also lean comparatively more towards aggression than their host learners' peers.

Q3 - Children's Characteristics Explaining the Variation on Outcomes

1. Among the characteristics that explain differences in outcomes for ECD children are:

i. Language, with Swahili learners generally displaying better learning outcomes than their Kirundi peers.

ii. Community type, i.e., being a child from a refugee community has a statistically and positively relationship with higher total IDELA scores among all domains except emergent numeracy.

iii. Age is statistically and positively related with higher all learning and wellbeing outcomes. Though we found certain sex differences in the descriptive analyses, sex did not hold at the complete sample used for inferential analyses.

iv. Having a female teacher has a statistically and positively relationship with wellbeing.

v. Teachers' highest qualification (having a bachelor's degree) has a statistically and positively relationship with Emergent Literacy.

Q4 - Educators' Perception on LtP

1. Results indicate that educators generally perceived learning through play positively, implying their readiness to practice if they are provided with support.

2. Educators' engagement in LtP developmental and educational was moderate, and when analyzed by subscale, preschool educators seem to have a better mean score on the emotional support subscale.

Q5 - Educators' Instructional Practices

1. Results suggest that the learning environment in the classrooms observed is high and that most classrooms visibly displayed indicators of quality.

Q6 - Educators' Engagement, Occupational Wellbeing, and Self-Efficacy

1. The educator's mean work wellbeing and engagement score suggests that preschool educators have high levels of occupational well-being. Educators scored highest in the dedication subscale, compared with the vigor and absorption subscales.

2. Results show that educators have moderate levels of self-efficacy. Educators scored higher in their self-efficacy related to classroom management.

Conclusions and Recommendations

Learners

• Both ECD and Primary learners display low levels of learning outcomes, particularly Primary learners. Results suggest that learners for both levels have acquired basic or foundational skills, but still struggle with higher order skills.

o Emphasize foundational skills, and make sure that teacher training strategies are particularly appropriate for fostering higher order skills.

• Results depict significantly different results and classroom experiences of girls and boys, though differences are mixed by community, languages, levels, and types of outcomes.

o Revise program design and implementation to ensure the use of equity-focused principles to improve and enhance girls' performance and outcomes, particularly in terms of learning differences and bullying and victimization of Primary students.

• Language and sample sizes played a role in the limitations of analyses for certain outcomes. For example, some disaggregations were not feasible given small sample sizes (particularly for exploring differences by language). Though the research and implementation teams conducted several discussions on the reach of research activities and samples, the baseline data collection served as an exercise of confirming data previously collected (such as schools and classroom registrations) and revealed gaps in understanding and the information collected. Further, certain measures or tools did not perform as expected in terms of their reliability.

For example, though we find similar levels of overall wellbeing across samples, the tool did not allow for a further exploration of different wellbeing components. While endline data could have provided initial insights on the reasons, endline was not conducted.

- o Reconsider and adapt the tools and implementation of the learning assessments according to learnings from the baseline. For example, both ECD and Primary learning assessments can be shortened.
- o Given the finding of the relationship between teacher's sex and learners' wellbeing, it is warranted to continue exploring the measurement of wellbeing and its different components with children's outcomes.
- o Continue measurement research and analyses relevant outcomes and make sure that planning accounts for language diversity and sample sizes limitations and inaccuracies.
- o Strive for parsimony in research tools and methods and build from baseline learnings to improve tools' display, connectivity, compatibility, and deployment.

Educators

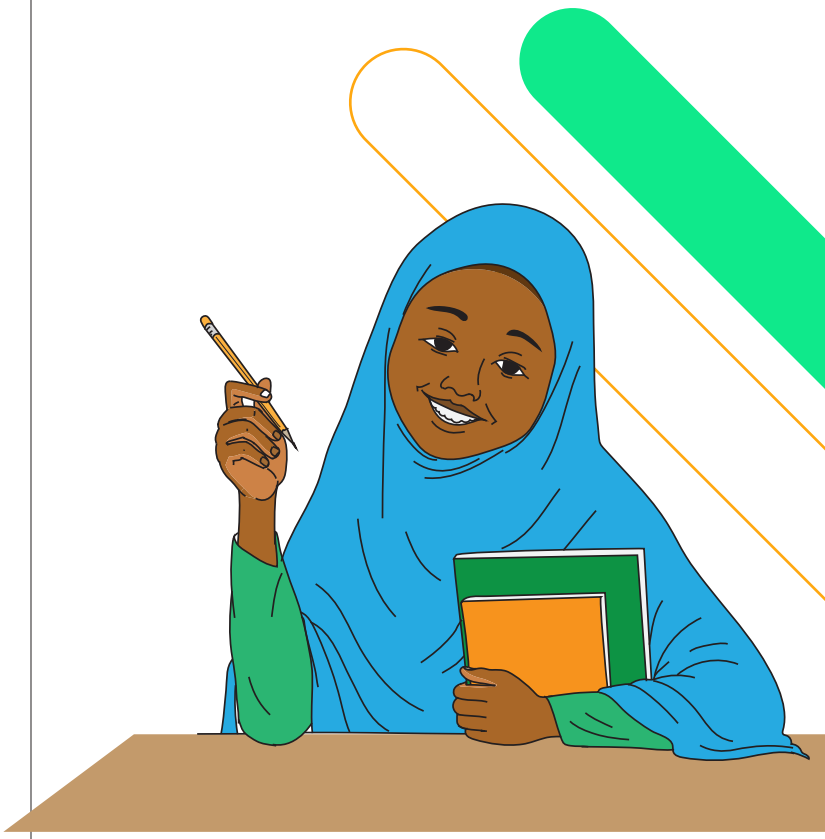
• Results indicate that educators perceive learning through play positively, but that their engagement in LtP developmental and educational was moderate, suggesting a gap in their knowledge, understanding and operationalization of positive LtP activities. Further, while quantitative results suggest that classrooms have high levels of quality, feedback from the county team suggested that there were some enumerators displaying positivity biases during the data collection.

- o Provide and strengthen professional development to educators: For example, activities could include providing explicit training on inquiry-based learning, cooperative learning, and hands-on activities. Activities for educators should be tailored to specific priorities and the design of PlayMatters moving forward. Such activities should be designed to deepen educators' understanding of LtP and child-centered pedagogy.
- o Equip teachers with the necessary skills and strategies to create a student-centered classroom environment will require continuous monitoring and research.
- o Reconsider the classroom observation tool design and implementation.

Programmatic and Research

• While the present study used a set of defined tools aligned with the initial design of the program, any future research tools should be aligned with the specific details of the implementation to enhance the accuracy, relevance, and comprehensiveness of the research findings. Further, the baseline tools were selected in alignment with the initial plans for educators' teacher professional development activities and the desired outcomes at the educator level.

- o Align research tools with PM's ToC and definitions: This is particularly relevant for the educators' measures since, the tools selected for the baseline were aligned with PM's early definition of LtP, which has since evolved and further delimited.
- o Refine the educator outcomes to continue to develop and define appropriate tools, while maintaining a close relationship with research plans, tools, and reach.
- o Implementation research should integrate measures and indicators that directly assess the key aspects and outcomes of the implementation. Future implementation research should also strive to improve the learning and interaction between the different methods (quantitative and qualitative) used.



Introduction

Learning through Play in Conflict and Crises

LtP is based on the premise that play is not purposeless but a process that improves brain structure and function and facilitates the process of learning by helping children to pursue goals, ignore distractions, and build resilience (Frost et al., 2012). Evidence supports that LtP can improve holistic outcomes for children more effectively than either traditional instruction or free play (Yogman et al, 2018), as it enhances cognitive, social, physical, and emotional aspects of children for it improves the level of engagement and motivation. While playing, children practice self-regulation when they take turns, accepting losses and managing conflicting interests (Vygotsky, 2016). Children also practice persistence and self-perception when they, for example, compete for better performance to win (Gaffar & Campbell, 2021).

Though the definition and delimitation of LtP is an ongoing debate, some consensus exists around key elements of LtP include children: i) being actively engaged, ii) relating new experiences to what they already know (i.e., play being meaningful), iii) enjoying a task for its own sake and the thrill of surprise, insight, or success after overcoming challenging experiences, iv) iterating (i.e., trying out possibilities, revising hypotheses and discovering new questions), and v) interacting socially (i.e., to communicate thoughts, share ideas, understand and enjoy being with others, and build stronger relationships) (Zosh et al., 2017).

Conflict and crises affect both children and adults, but their effects have further-reaching adverse effects on children than on adults. First, children lack the physical and emotional readiness to cope with the consequences of the crisis as they are still in a developing stage. Second, children and their wellbeing depend on their caregivers, who are themselves affected by the crisis. Studies on children exposed to war and separated from familiar environments and relationships, for example, show that children experience emotional stress, and the consequences become more severe when children are separated from their parents due to a crisis (Osofsky, 1999).
In recent years, LtP has emerged as a relevant and affordable pedagogical approach and/or intervention for children in crisis-affected settings as it helps them discharge emotions and develop coping mechanisms and hope. In times of crisis or difficulty, play develops hope and helps children not to jettison the problem but to develop the ability to cope (Yohani & Larsen, 2009). Though play is universal and LtP enhances holistic learning, the challenges facing educators in refugee settings are extreme (INEE, 2019). Currently, there is little evidence from low-resource contexts on how to assist educators in overcoming the challenges they face implementing LtP in humanitarian settings.

Education system and learning in Tanzania

In Tanzania, education policy is formulated by the Ministry of Education, Science, and Technology (MEST). The education system is categorized into pre-primary, primary, secondary, and higher education. However, compulsory basic education only covers pre-primary to lower secondary levels. Pre-primary education lasts one or two years for children aged 4-5 years, followed by six years of primary education, four additional years for lower secondary education, which marks the end of compulsory basic education. Students at this level sit for examinations to advance to secondary school. Successful completion of advanced secondary education and examinations allows students to pursue higher education leading to diplomas or bachelor's degrees.

Education policies in Tanzania are sustained by robust policies and curriculum reforms based on teacher competencies and evidence-based recommendations, placing access and learning outcomes at the center of the education system. For example, the government of Tanzania recognized the importance of early childhood education (ECE) for children ages 5 and 6 years under Tanzania's Education and Training Policy (ETP) in 1995 (Ministry of Education and Culture, 1995), and placed it under the direction of the Ministry of Education. However, it was formalized and made compulsory and free in the 2014 Education and Training Policy (MoEVT, 2014). Today, the ECE curriculum policies reflect the acknowledgement of its critical role in laying the foundation for a child's educational journey and overall development, and in the need for developing strong foundations for learning.

For primary, a competence-based curriculum was introduced, shifting the focus from content knowledge to developing students' competencies in 2005. The 3Rs reform implemented in 2014 aimed to strengthen reading, writing, and arithmetic skills in early grades at the primary education level. More recently, in 2021, President Samia Suluhu Hassan highlighted the improvement of education quality as a priority area and proposed a review of the 2014 Education and Training Policy and curriculum to include 21st century skills, followed by an announcement of the Minister for Education Science and Technology to review the policy and curriculum (Komba & Shukia, 2023).

The extent to which the current curriculum for basic education (pre-primary, primary, and secondary) in Tanzania fosters the development of foundational and more advanced skills remains unclear with mixed and limited research. Evidence of low achievement from limited learning assessments suggests that a greater effort for developing and monitoring in the core 3R skills (reading, writing, and mathematics) is needed, both at ECD and Primary. For primary, a report in 2012 (UWEZO) revealed that 47 percent of students in the final grade of primary were unable to read basic English stories, 26 could not read basic Kiswahili stories, and 11 could not perform simple tasks. These statistics highlight the urgent need for targeted interventions and focused attention on developing these essential skills early in the educational journey. For ECD, there is a dearth of evidence on learning outcomes, and for both ECD and Primary there is no evidence about refugee learning outcomes.

Empirical evidence suggests that despite the recognition of the importance of early childhood education (ECE) in educational policies, there has been insufficient prioritization (with efforts mainly focusing on physical infrastructure and teaching materials over teacher training), resulting limited availability of qualified teachers and inadequate supervision (Mtahabwa, 2007; Zuze & Reddy, 2016; Ndijuye et al., 2020). Qualitative studies and interviews with national education stakeholders have indicated that teachers generally lack the pedagogical skills required to effectively teach the 3Rs (reading, writing, and arithmetic) in both pre-primary and early grades. Overall, evidence suggests that more continuous professional development is crucial for equipping teachers with the necessary skills for effective instruction in the 3Rs at both foundational and more advanced skills (Komba & Shukia, 2023).

Refugee inclusion and education in Tanzania

By 2010, the global number of forcibly displaced refugees reached an estimated 43.7 million, with approximately half located in sub-Saharan Africa and the Middle East. Countries have approached the provision of services to refugee children based on three distinct categories of refugee civil status: voluntary repatriation, resettlement in a third country, or local naturalization or integration (Akaro, 2001; Warner, 1994). Tanzania, since its independence in 1962, has been recognized as one of the most hospitable nations globally (Rutinwa, 1996; Milner, 2013). Its Open-Door Policy, championed by its first President, Julius Nyerere, welcomed thousands of refugees escaping conflicts and insecurity in the region, dating back to the 1960s. However, the 1990s marked a significant shift in Tanzania's approach to refugees, prioritizing repatriation over self-sufficiency and settlement (Milner, 2003).

Currently, Tanzania hosts approximately 250,000 refugees and asylum-seekers, primarily from Burundi and the Democratic Republic of the Congo, residing in Nduta and Nyarugusu camps in the northwest region. Of the refugee population, over 55% are children under 17 years old, nearly 20% are under 4 years old, and 20% fall between the ages of 5 and 11. Refugees are confined to camps, with almost half living in overcrowded emergency shelters and facing restrictions on attending school in the host community. The coordination of education within the refugee camps falls under the responsibility of the UNHCR, which provides funding to NGO partners to deliver formal schooling directly. In line with the principle of education for repatriation, refugee schools within the camps employ the curriculum of the refugees' home countries rather than Tanzania's curriculum. The language of instruction for early childhood care and development (ECCD) and up to Grade 4 is Swahili for Congolese refugees and Kirundi for Burundian refugees. From Grade 5 onwards, French becomes the medium of instruction for both populations, while Kiswahili and English are subjects across all ECCD and primary grades.

The Present Study

Research Aims and Questions

The baseline data collection for the Implementation Research Study took place in February and March 2022. This study aim was to conduct a baseline assessment of the initial characteristics of children and educators engaged in PlayMatters' intervention in refugee and host communities in Tanzania. Results intend to inform program implementation and provide insights to improve and refine PM's package and implementation. The original research design considered baseline, midline, and endline. However, due to delays in program implementation and revisions to the project strategy and overall design, midline and endline data collection points were cancelled to save resources and strive for future alignment between research and the program design. Thus, the research questions from the original design were revised to the following:

I. Child Level

1. What are children's literacy and numeracy skills? How do they vary by displacement, sex, and region?

2. What is the status of children's socio-emotional skills and wellbeing? How do they vary by displacement, sex, and region?

3. What child characteristics explain the variation in children's literacy, numeracy, and SEL skills?

II. Educator Level

1. What is the perception of educators on LtP?

2. What are educators' instructional practices?

3. What are the levels of educators' work engagement, occupational wellbeing and self-efficacy?

Ethical Considerations

This study received ethical clearance from the International Rescue Committee's institutional review board (IRB) and from the the University of Dar es Salaam's IRB in August 2021. All the Principal Investigators (PIs) and co-PIs involved in the study completed a web-based course on protecting human research participants' online training provided by Protecting Human Research Participants (PHRP). Before collecting any data, the team obtained parental consent explaining to participants the purpose of the study and the voluntary nature of it.

While collecting data, consent from parents and educators and assent from children were obtained, anonymity was kept, and the use of the data collected was limited to the purpose of the study only. Data has since been stored in password-protected cloud services of the IRC and only accessible to the PM team.

Process

In January 2022, researchers and project teams participated in a 6-day regional training of trainers (ToT) in Tanzania which covered the rationale and implementation requirements for all the study's tools. In February and March 2022, enumerators were selected based on specified criteria such as language proficiency, experience, and qualification. Enumerators were then trained for five days covering the objectives of the assessment, assessment procedures and ethical considerations, measures, and use of tablets with both the CommCare and Tangerine software. In addition to paired and small group simulated practices, the training also included school-based practice in 3 schools in Kigoma . From February 22, 2022 – March 09, 2022, research coordinators, IRC's MEAL staff and the hired enumerators visited schools to administer the instruments (surveys and assessments) and conducted the classroom observations. Enumerators administered questionnaires to children individually in 2 local languages (Swahili and Kirundi) and spent from 1 to 2 hours of their time (varying time by children's age). School officials at the refugee camps and enumerators at host community schools also conducted 45-minute classroom observations. All the data was collected using CommCare and Tangerine, and later cleaned and analyzed by IRC's Airbel Impact Lab researchers using Stata and with input from IRC's Tanzania MEAL staff.

Method

Participants

The baseline study covered the regions of Kasulu and Kibondo and covered 23 sites (11 ECD and 12 Primary), both in refugee and host communities (11 and 23 respectively). Children were assessed in the medium of instruction: Kirundi (refugees in Kibondo, n=222) or Swahili (rest of the sites, n=412). Within schools, educators were selected in consultation with respective head teachers or by availability sampling when there was just one classroom or teacher. All educators answered the survey in English. Within each classroom, at least five children, when available due to enrollment, were selected using a sex-specific systematic random sampling technique. Table 1 shows the number of participating sites and the total sample sizes by region, community type, and sex for both children and educators.

Table I. Summary of Sample

Level/District	Level/District	# of Sites	# of Children Assessed			# of Children Observed			Medium of Instruction/ Assessment Language
			M	F	Total	M	F	Total	
ECD/ Kasulu	Host	2	8	8	16				Swahili
	Refugee	3	16	14	30				Swahili
ECD/ Kibondo	Host	3	10	11	21				Swahili
	Refugee	3	17	23	40				Kirundi
Total ECD	**Host**	**5**	**18**	**19**	**37**				**Swahili**
Total ECD	**Refugee**	**6**	**33**	**37**	**70**				**Swahili and Kirundi**
Total ECD		**11**	**51**	**56**	**107**				**Swahili**
Primary/Kasulu	Host	4	34	32	66				Swahili
	Refugee	3	100	100	200				Swahili
Primary/Kibondo	Host	3	164	158	186				Swahili
	Refugee	2	22	24	182				Kirundi
Total Primary	**Host**	**7**	**198**	**190**	**388**				**Swahili**
Total Primary	**Refugee**	**5**	**122**	**124**	**246**				**Swahili and Kirundi**
Total Primary		**12**	**320**	**314**	**634**				

Instruments

The instruments for study were selected after a careful review of the study objectives, population, and alignment with PlayMatters intended outcomes and proposed Theory of Change.

Children

I. Children Survey: A questionnaire to gather demographic information such as age, sex, displacement status, etc. Applied for both ECD and Primary children.

II. International Development and Early Learning Assessment (IDELA): A one-on-one test designed to measure young children's learning and development (Pisani et al., 2018). Trained enumerators administered the assessment providing children the required stimuli and materials when required and recording responses using a tablet. The assessment included 22 subtasks, each measuring specific constructs of early development, across four domains: gross and fine motor, emergent literacy, emergent numeracy and socioemotional. Each subtask is measured in terms of percent correct or appropriate response, i.e., how many of the total items were answered correctly or appropriately, and the total score for each domain is calculated by adding the weighted score of all the subtasks in the domain. The aggregated IDELA score is then calculated by aggregating the four main domains and it reflects holistic child development and is regularly used to measure program effectiveness. IDELA was developed by Save the Children and its psychometric properties have been evaluated in multiple studies and countries in the Global South. Applied for ECD children.

III. Early Grade Reading and Mathematics Assessments (EGRA and EGMA): A one-on-one assessment designed to measure and evaluate foundational reading skills that are critical for literacy development, such as letter recognition, phonemic awareness, decoding, fluency, and comprehension for learners in primary grades. Certain subtasks of EGRA and EGMA are administered within a time restriction of 1 minute, while the rest are untimed. All subtasks are analyzed in terms of percent correct, i.e., the percentage of items a child scored as correct out of the total number of items per subtask. For example, Oral vocabulary was measured by calculating the number of correct identifications of a picture that represents a word. Additionally, for the timed subtasks, the correct answers per minute is computed and analyzed. For example, EGRA's letter identification and oral reading subtasks were measured by calculating both the percent correct and the mean score of letters identified in a grid or correctly read words in a short story. For EGRA's reading comprehension, students were only asked questions aligned with the portion that learners were able to read in the short story provided in the oral reading subtask. The tool was administered by enumerators using the software Tangerine, built for the purpose of administering EGRA and EGMA. Applied for Primary children.

IV. Kiddy-KINDL: A self-reported questionnaire that measures wellbeing of children between 3 and 17 years of age (Bullinger et al., 1994). The version used in this study has 12 items, covering six dimensions related to children's wellbeing: physical health, feelings, and relationships with family members and friends. All items are scored on a 3-point Likert scale (0= never, 1 = sometimes, and 2 = very often) for ECD and on a 4-point Likert scale (0= never, 1 = seldom, 2 = sometimes, 3= often, and 4=always) for Primary which yields one overall. Reverse scoring is applied to two items in the physical well-being dimension and one item in the emotional well-being dimension. Applied for both ECD and Primary children. For both ECD and Primary, the overall tool exhibited good levels of reliability, but reliability for each of the different subtasks varied. Thus, we focus on the estimates at the tool level in this report.

V. Empathy scale: A self-reported measure of empathy i.e., the ability to understand and share another person's feelings and thoughts based on a given situation. It included 12 items with a three-rating scale. Applied for Primary children.

VI. Emotional Attribution Accuracy (ACES): Self-reported scale that uses 10 vignettes that aim to measure children´s ability to identify the emotions (happy, sad, scared, and non-feeling) that others would feel in a given situation. Answers are scored as correct (1) or incorrect (0) and the final score reflects the total correct. Applied for Primary children.

VII. Socio-Emotional Response and Information Scenarios (SERAIS): A scenario-based tool, which has been used in contexts like Lebanon, Colombia and Nigeria and adapted for PlayMatters by the IRC, where children are asked to account for what they would feel and do if they were in the variety of different social situations to measure hostile attribution bias, emotional regulation, and conflict resolution (Kim & Dolan, 2019), all reported in a scale of (0 - 3). Applied for Primary children. The tool captures information about different social, emotional, and cognitive sub-domains following the multi-step information process children undertake in social situations: first, the internal encoding of social cues, then, formulate a goal for the interaction and possible responses, and finally, evaluate the possible responses to resolve a conflict and select one to enact. In this line, the three SERAIS sub-domain measure:

a. Hostile attribution bias indicates the extent to which a child encodes ambiguous social cues (for example, another child cutting him/her in a waiting line) as hostile.

b. Conflict-resolution strategies: referring to the final step of children deciding on a conflict resolution strategy such as disengagement, proactive positive problem-solving, or aggression.

VIII. Witnesses' responses to bullying: Self-reported scenario base-tool in in which children are asked to imagine that they observe a situation in which a peer is being bullied at school and then indicate their attitudes and behaviors towards the situation, as well as the degree to which they have experienced similar situation in their own school (Diazgranados et al, 2016). Applied for Primary children.

a. Attitudes toward bullying: Assesses how children feel about different responses to bullying: 1) Upstanding (by confronting perpetrators, helping the victim, or reporting to an authority); 2) acting as

bystander or doing nothing, and 3) joining perpetrators. Scoring indicates the level of agreement or disagreement with different responses measured on a scale of (0 – 3).

b. Exposure to bullying: Reports the number of instances in which children report having witnessed situations of bullying in their school in the last two weeks.

c. Experiences of victimization: Assesses the degree to which children have been victims of different types of bullying in the last two weeks. Responses are provided in terms of frequency of the experience (never, once, twice, three or more times). The final score represents the average frequency.

Table II. Child Measures

Tool (Score Range)	Subtasks or Constructs	# of Items	Scoring
IDELA (1-100%)	Child holistic development with focus on: Executive Function (EX); Emergent Literacy (EL); Emergent Numeracy (EN); Socio-Emotional Learning (SEL); & Motor Skills (MO)	24	Total or percent correct, appropriate/inappropriate and Ratio (like in writing skill, hoping, etc.) varying by the type of subtask.
EGRA	Timed subtasks: • Letter identification • Oral reading (fluency) Untimed subtasks: • Vocabulary • Reading comprehension	Varying by subtask	Percent correct, score adjusted by time (timed subtasks), percentage of sample with zero scores, and percentage of sample at or above the performance category
EGMA	Timed subtasks: • Number identification • Addition L1 • Subtraction L1 Untimed subtasks: • Number discrimination • Missing number • Addition L2 • Subtraction L2 • Word problems	Varying by subtask	Percent correct, score adjusted by time (timed subtasks), percentage of sample with zero scores, and percentage of sample at or above the performance category
Kiddy-KINDL (0-2 and 0-4)	Health-related quality of life assessment (Physical health, feelings, and relationships).	12	Three-point (never, sometimes, and very often) for ECD and five rating scale (never, seldom, sometimes, often, all the time) for Primary

Educators

I. Educator Survey: A questionnaire to gather demographic information, such as age, displacement status, educational background, and professional development received.

II. The Teacher Instructional Practices and Processes System (TIPPS): An observational tool that measures the quality of educators' classroom instructional practices using 20 items. Items are scored on a four-point Likert type scale to illustrate the "degree" to which the concept is present in the classroom. It has been used and validated in the Democratic Republic of Congo, and was adapted later to Uganda, Ghana, Pakistan, and other countries (Seidman et al., 2018). A previous validation study suggested three conceptually distinct constructs or domains, however, due to small sample sizes, we report an overall aggregate score.

a) **LtP Perceptions:** A 16-item survey, adapted for the present study from the work of Brackett et al. (2012) to identify whether they agree or disagree with statements regarding their ability to teach learning through play on a five-point scale (0 = strongly disagree; 1 = Disagree; 2 = Neutral; 3 = Agree; 4 = Strongly Agree). Specifically, the scale assesses: (a) pedagogical comfort with implementing LtP, (b) commitment to LtP, and (c) perceived support from the school, each measured by 4 items.

b) **Developmental and Educational Activities Scale (DEAS):** A self-report tool to assess educators' engagement with developmental and educational activities of children in ECD settings (Slot et al., 2014; 2015) using a five-point Likert type scale (with scale values ranging from 0 – 4) focusing on four dimensions viz. during play (8 items), pretend play (8 items), self-regulation (11 items), and emotional support (8 items). A score close to 4 indicates a high level of engagement, and a mean score close to 0 shows a low level of engagement.

c) **Utrecht Work Engagement Scale (UWES):** A self-reported questionnaire that measures the affective aspect of teacher well-being at the workplace through (a) vigor, (b) dedication, and (c) absorption (Schaufeli et al., 2006). The tool consists of 9 items that measure the educator's well-being at the ECD settings using a five-point Likert-type scale (with scale values ranging from 0 = Never to 4 = Always) focusing on vigor, dedication, and absorption. A mean score close to 4 indicates high well-being and a mean score close to 0 shows low well-being. The total UWES mean score specifies the overall occupational well-being of educators in the workplace.

d) **Teacher Self-Efficacy Scale (TSES):** A self-report questionnaire created by Tschannen-Moran and Woolfolk Hoy (2001). The tool measures educators' beliefs in their capability to make a difference in student learning through three constructs: teacher efficacy in (a) instructional strategies, (b) classroom management, and (c) student engagement. The scale consists of 12 items that measure the educator's self-efficacy in the ECD settings using a five-point Likert-type scale (with scale values ranging from 0 = Not at all applicable to 4 = Very strongly applicable). A mean score close to 4 indicates high self-efficacy and a mean score close to 0 shows low self-efficacy.

Table III. Educator Measures

Tool (Score Range)	Subtasks or Constructs	# of Items	Scoring
TIPPS (0-3)	Quality of classroom environment.	20	Two-stage, two-point each rating scale (somewhat accurate & very accurate)
Educator LtP Survey (0-4)	Educator's perception on LtP and their assessment on their capacity to implement LtP strategies in the classroom.	15	Five-point rating scale (0=Strongly disagree; 1= Disagree; 2 = Neutral; 3= Agree; 4 Strongly Agree)
DEAS (0-4)	Developmental and educational activities scale: play, pretend play, self-regulation, & emotional support.	35	Five-point rating scale (0=Not at all applicable; 1= A little applicable; 2 = Applicable; 3= Strongly applicable; 4 Very strongly applicable)
UWES (0-4)	Educators' wellbeing and work engagement: vigor, dedication & absorption.	9	Five-point rating scale (0=Never; 1= Seldom; 2 = Sometimes; 3= Often; 4 Always)
TSES (0-4)	Efficacy for instructional strategies, efficacy for classroom management, & efficacy for student engagement.	12	Five-point rating scale (0=Not at all applicable; 1= A little applicable; 2 = Applicable; 3= Strongly applicable; 4 Very strongly applicable)

Note. See Annex 2 for the reliability estimates for each tool/subtask.

Instruments Validity and Reliability

Validity was sought through the selection of previously validated measures to the extent possible. Prior to data collection, all the instruments were translated into the relevant languages by expert in-country translators. Once translated, instruments were cognitively pretested with a sample of comparable children in the study areas to check suitability and appositeness to the local context. Reliability was tested using the psychometric tests is Cronbach's alpha, which estimates the internal consistency reliability of an instrument, indicating the extent to which subtasks or items deliver consistent scores.

The range for Cronbach's alpha is 0.00 to 1.00, with higher values indicating better (or more desirable) reliability. Before conducting analyses, we calculated Cronbach's alpha separately for each instrument and language, for each of the study groups (ECD and Primary children and educators). For analyses, we mostly omit on reporting sub-scales that did not comply with the threshold unless they are considered critical skills to be reported and for cases in which reliability varies by language (such as, number operations, and empathy). Unless noted otherwise, the estimates included in the reporting throughout have reliability coefficients that range between acceptable (0.7-0.8) to excellent (≥ 0.9). Detailed information can be found in Annex (2). For learning assessments, we also calculate Pearson correlation coefficients among each assessment's subtasks to indicate the consistency of performance by the subtasks on the test. Subtask's correlation for each assessment and language can be found in Annex (3).

Analytical Strategy

The baseline data collected from both children and educators was quantitative in nature, thus descriptive and inferential statistical methods were used to analyze data. In order to address the research questions, we also conducted simple inferential analyses of differences by relevant groups. For children's questions 1 and 2, we use descriptive analyses to explore baseline literacy, numeracy, socio-emotional skills, and wellbeing. Due to the small learner sample sizes, statistical inferential testing included only differences by community type (refugee or host), separately for each language. Though sex disaggregation by language and community was not feasible due to small sample sizes and unbalanced groups, we include in the Annex differences by children's sex at the community level (refugee or host) across languages, these results are illustrative and should be interpreted with caution.

For question 3, we use regression analysis to analyze the relationship between children's characteristics and key children's outcomes. For the analyses, we select variables, based on both theoretical and previous empirical evidence and research, as well as on baseline data availability, specifically we use (varying by ECD and Primary children): child sex, community type (refugee and host), child age, times s/he repeated any grade, the availability of reading resources at home, teacher's sex and teachers' highest degree of education. All regression analysis cluster standard errors at the school level.
For IDELA, we present mean scores for each domain and subtasks. For EGRA and EGMA, we present mean percent correct scores for all subtasks and correct per minute (fluency scores) for timed subtasks. For timed subtasks and reading comprehension, we also present the percentages of students that fall within three categories: zero scores (indicating the percent of children that were unable to score a single correct answer), below a target of 50% correct, and at and above 50% correct.

In analyzing TIPPS items, we analyze the frequency distribution or percentage of observed behaviors employed to understand each of the items included to understand quality of classroom instruction and environment, and we also analyze the percentage of 'quality classrooms' a category that includes categorizing indicator of quality classroom (a mean TIPPS score > = 3).

For each of the learners' socioemotional tools and educators' tools and surveys, we present mean scores according to each tool's score range. The report presents findings using tables and figures as appropriate, indicating disaggregation and statistical significance when applicable.

Scope and Limitations of the Study

This descriptive study focuses on documenting the learning outcomes and instructional practices of students, teachers and educators in refugee centers and host communities in Kasulu and Kibondo. The study is particularly relevant for similar conflict-affected settings where refugees and host communities in separate schools managed by either INGOs and NGOs or the Tanzania government.

It is important to note that the presented results are predominantly descriptive and do not explore relationships among variables (unless specified), nor do they imply causality. Due to limitations in sample sizes, particularly regarding the number of educators and different languages assessed, it was not possible to analyze certain disaggregations, associations, or correlations.

Though this report includes side-by-side results for both assessed languages (Swahili and Kirundi), it is noteworthy that language groups and samples are different, thus results are not meant to be compared across languages. Learning assessments' results are influenced by variations in language and children characteristics that differ across refugee and host community contexts. Finally, due to data entry errors during the data collection, disaggregations by disability and socioeconomic status are not possible.

[2] While there are established benchmarks for certain languages (including Swahili) in Tanzania, there is no standardized benchmarks for children assessed in a language that is not their first language, such as the refugee children in our sample. We thus, for each language, set 45 correct letters and correct word per minute for the Letter identification and Oral Reading EGRA subtasks as the target/cut point for meeting the performance category for ease of interpretation and presentation of results. For EGRA and EGMA untimed subtasks, we chose 80% correct as the target/cut point for meeting the performance category. This approach is not indicative or a replacement for a standardized process for establishing benchmarks for either language, for both EGRA and EGMA subtasks.

Results

Child - ECD

In this section, we present the findings for children in the ECD level for learning and development outcomes (IDELA overall and specific-domains scores) and socioemotional skills for both the host and refugee communities (details on IDELA Domain's subtasks in Annex 5). We highlight the differences in scores between children in each community for each of the IDELA domains. Then, we describe the illustrative differences between girls and boys for each of the Swahili and Kirundi samples.

Children's Development (IDELA)

Table 4 includes IDELA results for host refugee communities by language (Swahili and Kirundi). For Swahili (columns 1, 2, and 3) we found students in the refugee community to score better than students in the host community in five of the six IDELA domains, a gap that is statistically significant in four of the domains (Emergent Literacy, Socioemotional Skills, Executive Function, and Approaches to Learning) as well as for the overall score. While some gaps are small others are substantial. For example, students in the refugee community score, on average, 49% correct answers in the socioemotional domain while students in the host community score, on average 32% correct answers. The second largest gap across groups is the Executive Function domain with children in the refugee and host community score, on average, 51% and 38%, respectively.

Overall, results suggest that young children are yet developing their foundational skills: both host and refugee communities scored higher in Fine motor, followed by Executive Function, and Socioemotional skills, with higher order skills (Emergent Numeracy and Literacy) lagging (with learners scoring, on average, less than one third of the domains correctly). At the subtask level and samples aggregated by language (Annex 4), we find that refugee children assessed in Swahili outperform the children from the host communities in Memory, Shape identification, vocabulary, oral comprehension, writing and all the socioemotional subtasks (Emotional Attribution, Empathy, and Conflict Resolution). Results of children assessed in Kirundi follow a similar pattern with higher scores for Executive Function, Fine Motor, Socioemotional, and Emergent Numeracy domains.

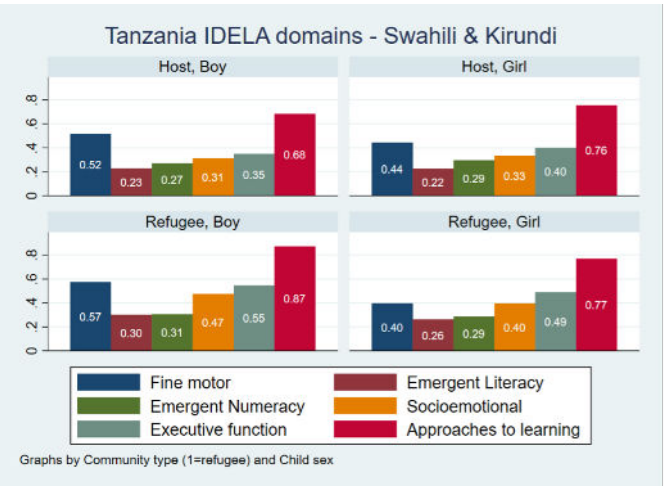
Table IV. IDELA Domains by Language and Community

	Swahili					Kirundi	
		(1)		(2)	(3) t-test		(4)
	N	Host Mean/SE	N	Refugee Mean/SE	Difference (1)-(2)	N	Refugee Mean/SE
IDELA Total Score	30	0.345 (0.024)	26	0.418 (0.032)	-0.073*	36	0.356 (0.028)
IDELA Domain Score: Fine Motor	37	0.479 (0.035)	29	0.552 (0.047)	-0.073	37	0.418 (0.041)
IDELA Domain Score: Emergent Literacy	34	0.226 (0.023)	28	0.340 (0.030)	-0.114***	40	0.237 (0.025)
IDELA Domain Score: Emergent Numeracy	32	0.282 (0.028)	28	0.255 (0.026)	0.027	38	0.323 (0.031)
IDELA Domain Score: Socioemotional	37	0.323 (0.032)	29	0.491 (0.045)	-0.168***	40	0.393 (0.037)
IDELA Domain Score: Executive Function	37	0.376 (0.040)	30	0.513 (0.049)	-0.137**	40	0.517 (0.044)
IDELA Domain Score: Approaches to Learning	27	0.719 (0.040)	25	0.837 (0.037)	-0.118**	28	0.804 (0.036)

Notes: The value displayed for t-tests are the differences in the means across the groups, sample permitting.
*** p<0.01, ** p<0.05, * p<0.1

Comparing students by sex and community, pooled by language, we find mixed results. Figure 1 (detailed statistics summary in Annex 5) shows that in both communities, boys slightly outperform girls in the Fine Motor domain, Approaches to Learning, and overall IDELA total score with all differences being statistically significant. For the rest of domains and communities, girls and boys have similar performance, mainly Emergent Literacy, and Numeracy, though differences are not statistically significant.

Figure I. IDELA Domains by Community Type and Sex

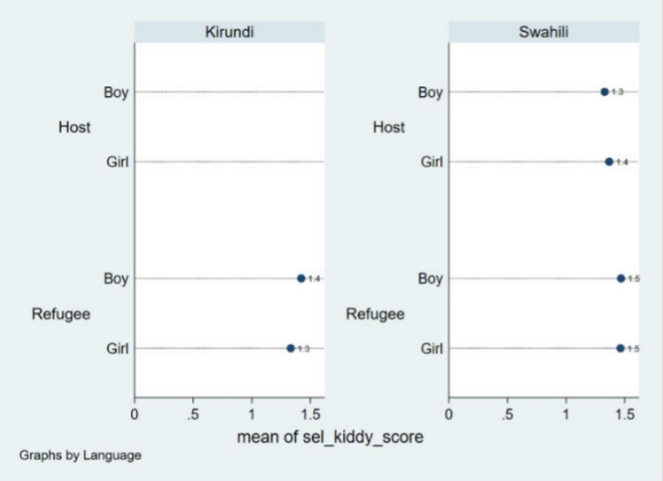


Notes: Given the small sample sizes, disaggregation by community type, sex, and language is not possible.

Wellbeing

Children across all groups (language, community type, and sex) display relatively high and positive levels of wellbeing with all scoring at or higher than 1.3 in a scale from 0 to 2 (Figure 2) and with no statistically significant difference for any of the groups compared.

Figure I. IDELA Domains by Community Type and Sex



Variation in ECD Child Outcomes

To answer research question 3, we use inferential analyses using OLS regression modelling (Table 5 below) exploring the relationship between child (sex, community type and age), and teacher characteristics (sex and highest qualification obtained) with both learning outcomes and socioemotional skills.

In terms of learning outcomes, results confirm the descriptive statistics in previous sections and indicate that students in the refugee community outperform those in the host (as indicated in columns 1-4 and 6), being a child from a refugee community is statistically and positively related with total IDELA scores and further domains of fine motor, executive function, emergent literacy and socioemotional.

Both child sex and interactions between child sex and type of community do not hold a significant relationship with any of the analyzed outcomes. However, and as expected, child age holds a positive and significant relationship with the learning outcomes analyzed, indicating that older children have, on average, higher scores than their younger peers. In terms of the teacher characteristics in relation to children's learning outcomes, we find that educators' highest degree has a significantly positive relationship with emergent literacy.

In term of socioemotional skills, none of the selected child characteristics have a significant relationship with the measure of children wellbeing (Kiddy-KINDL), though the relationship between child's age and having a female educator does seem to have a positive and significant relationship with children's wellbeing.

In this section, we present the findings for children in the Primary level for all learning outcomes (EGRA and EGMA) and socioemotional skills for both the host and refugee communities. We highlight the differences in scores between children in each community for each of the EGR/MA domains or subtasks. Then, we describe illustrative differences between girls and boys for each of the Swahili and Kirundi samples.

Table V. Determinants of ECD Child Outcomes

VARIABLES	(1) IDELA Total Score	(2) IDELA Domain Score: Fine Motor	(3) IDELA Domain Score: Executive Function	(4) IDELA Domain Score: Emergent Literacy	(5) IDELA Domain Score: Emergent Numeracy	(6) IDELA Domain Score: Socio-Emotional	(7) Kiddy-KINDL: Score (1-2)
Community type (1=refugee)	0.14***	0.15**	0.26***	0.13***	0.07	0.19***	0.11
Child sex (1=Female)	(0.05) 0.02	(0.07) -0.05	(0.08) 0.06	(0.05) -0.00	(0.05) 0.03	(0.07) 0.03	(0.09) 0.06
Community type X Child sex	(0.05) -0.07	(0.07) -0.08	(0.09) -0.09	(0.05) -0.01	(0.06) -0.03	(0.07) -0.09	(0.10) -0.11
Child age	(0.07) 0.06***	(0.09) 0.09***	(0.11) 0.05*	(0.06) 0.06***	(0.07) 0.05**	(0.09) 0.05**	(0.12) 0.06*
Teacher sex (1=Female)	(0.02) -0.02	(0.03) -0.07	(0.03) -0.06	(0.02) -0.01	(0.02) 0.01	(0.03) 0.03	(0.03) 0.15**
Teacher highest degree (1=Bachelor or diploma)	(0.03) 0.05	(0.05) 0.07	(0.05) 0.06	(0.03) 0.08**	(0.04) -0.02	(0.05) 0.05	(0.06) -0.09
Constant	(0.04) -0.00 (0.11)	(0.06) 0.04 (0.15)	(0.07) 0.05 (0.17)	(0.04) -0.13 (0.10)	(0.04) 0.03 (0.12)	(0.06) -0.01 (0.15)	(0.07) 1.04*** (0.20)
Observations	103	103	103	103	103	103	103
R-squared	0.18	0.22	0.12	0.17	0.07	0.12	0.11

Robust standard errors in parentheses, clustered at school level. *** p<0.01, ** p<0.05, * p<0.1

In this section, we present the findings for children in the Primary level for all learning outcomes (EGRA and EGMA) and socioemotional skills for both the host and refugee communities. We highlight the differences in scores between children in each community for each of the EGR/MA domains or subtasks. Then, we describe illustrative differences between girls and boys for each of the Swahili and Kirundi samples.

Learning Assessments (EGRA and EGMA)

Each row in Table 6 (below) shows the mean score, score adjusted by time (for timed subtasks) and percentage of zero scores for all EGR/MA measures or subtasks for learners of each community and the differences between both. For Swahili learners in EGRA, we found significant differences among host and refugee communities vocabulary (in terms of percent correct and percentage of zero scores), oral reading (in terms of percent correct, fluency, and percentage of zero scores) and in reading comprehension (in terms of percent correct and percentage of zero scores).

However, differences vary by community, according to foundational skills and more complex skills. For example, while refugees outperform their peers in the host community for the vocabulary subtask (both in terms of percent correct and percentage of zero scores), host community children outperform those in refugee settings for both oral reading passage and reading comprehension (both in terms of percent correct and percentage of zero scores). The overall results suggest, however, that while children in both communities are still building on their early literacy skills like letter identification, they are still struggling to perform in higher-order skills, as percentages of children not being able to respond a single question right is of high percentage for both oral reading and reading comprehension.

For Swahili learners, host community children outperform children in the refugee communities for most of EGMA subtasks. For example, host community children outperform children in refugee communities in the percent correct scores for the number identification, missing number, addition L2, subtraction L1, and subtraction L2. However, because significant differences mostly relate to percentage correct rather than in the fluency and zero scores results, results might suggest that refugee learners are in the path to making progress from early mathematical skills to

[3] We do not include "Approaches to learning" domain as, per design, the domain is not included in the computing of the total IDELA score.

higher-order skills.

For Kirundi learners (Table 7), we found that, compared with their Swahili speaking counterparts, they hold comparatively higher performances in both EGRA and EGMA results. Though we do not perform inferential analyses of the differences among languages due to the inherent differences and complexities of each language, it is noteworthy, for example, that Kirundi learners report more subtasks with a zero percentage of children being unable to answer a single question correctly (zero scores), than the Swahili learners.

Table VI. Summary of EGRA and EGMA Results – Swahili

	Swahili								
	Host			Refugee			Differences (H-R)		
Subtask	**% Correct**	**Correct Per Min**	**% Zero Scores**	**% Correct**	**Correct Per Min**	**% Zero Scores**	**% Correct**	**Correct Per Min**	**% Zero Scores**
EGRA									
Vocabulary	47.3		2.1	53.9		0.0	-6.608***		0.021**
Letter Identification	21.9	22.0	14.5	20.3	22.0	11.9	1.570	1.641	0.026
Oral Reading Passage (ORF)	47.4	30.6	24.2	21.2	30.6	38.0	26.188***	17.399***	-0.138***
Reading Comprehension	41.5		39.7	15.6		68.0	25.946***		-0.283***
EGMA									
Number Identification	69.7	23.7	4.1	62.2	23.7	4.0	7.513**	8.251***	0.002
Number Discrimination	62.1		11.3	63.2		10.9	-1.055		0.004
Missing Number	37.7		15.5	26.1		22.8	11.542***		-0.073**
Addition L1	43.6	9.3	12.9	39.7	9.3	12.9	3.929	0.991	0.000
Addition L2	38.5		41.8	31.1		45.5	7.364**		-0.038
Subtraction L1	27.8	6.3	22.2	22.9	6.3	27.7	4.964**	0.855	-0.056
Subtraction L2	29.4		57.7	20.8		60.4	8.589***		-0.027
Word Problems	34.6		17.0	31.5		18.8	3.104		-0.018

*** p<0.01, ** p<0.05, * p<0.1

Table VII. Summary of EGRA and EGMA Results – Kirundi

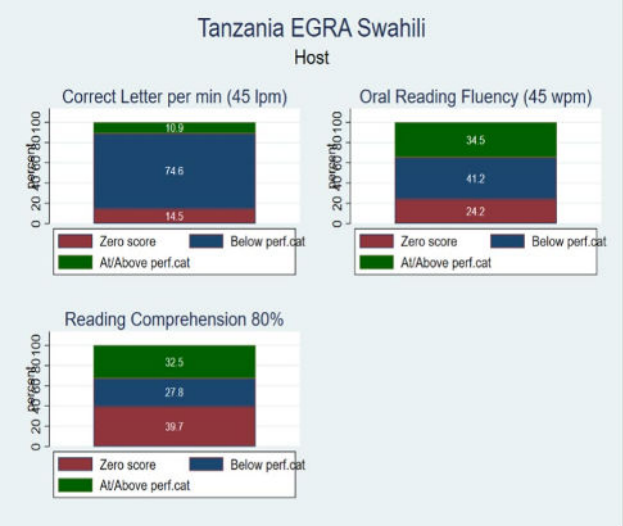
		Kirundi (n=44)		
		Refugee		
Tool	**Subtask**	**% Correct**	**Correct Per Min**	**% Zero Scores**
EGRA	Vocabulary	70.8		0.0
	Letter Identification	73.6	82.7	0.0
	Oral Reading Passage (ORF)	73.0	43.3	0.0
	Reading Comprehension	54.5		15
EGMA	Number Identification	93.6	34.7	0.0
	Number Discrimination	86.4		0.0
	Missing Number	64.8		2.0
	Addition L1	81.0	22.1	0.0
	Addition L2	75.0		7.0
	Subtraction L1	74.4	18.6	0.0
	Subtraction L2	72.7		7.0
	Word Problems	56.8		14

In order to grasp the patterns in the results of the learning assessments, we also analyze the percentage of learners that reached a stablished performance category. Results indicate that learners are achieving 'at/above' levels of performance at different percentages for each EGRA subtask, suggesting a diversity of learners' learning profiles. For example (Figure 3), in Swahili a large percentage (75%) of host learners are below the performance category established and 15% could not score a single correct letter correctly.

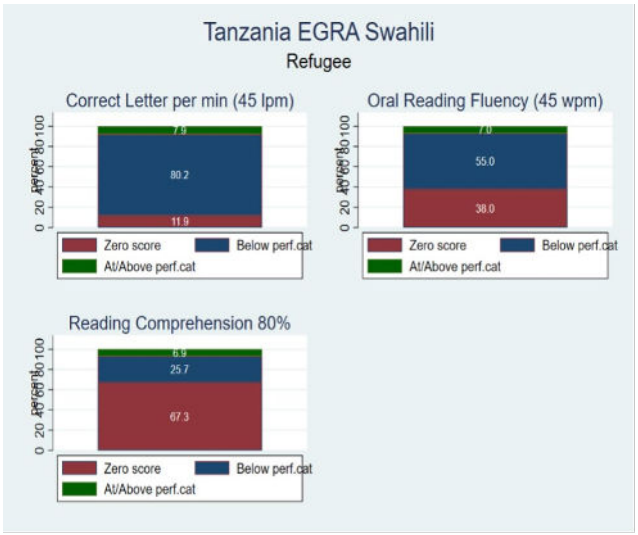
However, 35% of learners are at or above the performance category for oral reading fluency, which could suggest that while learners experience difficulties with specific letter sounds or phonics rules, they compensate for these challenges by relying on other reading strategies that allow them to perform in subtasks as reading fluency. Only 9% and 7% of refugee learners scores at or above performance categories for correct letters per minute and oral reading fluency respectively, indicating that they are considerably behind their host community peers. This difference is also reflected in the percentages of learners at or above performance for reading comprehension, with 32% and 7% of host and refugee learners respectively.

While we do not compare Swahili and Kirundi refugee children in terms of the correct works or letters read per minute, because each language could have different cut-off scores when gone through a language- and expert- specific validation workshop, we do compare the percentage of children scoring at least 80% correct in the reading comprehension, as this is a standardized measure for which languages characteristics do not play a role since enumerators read the questions out loud. Results indicate that Kirundi refugee learners display considerably better performance levels than the Swahili refugee learners in reading comprehension (with 43% and 7% of each population scoring at or above the performance categories).

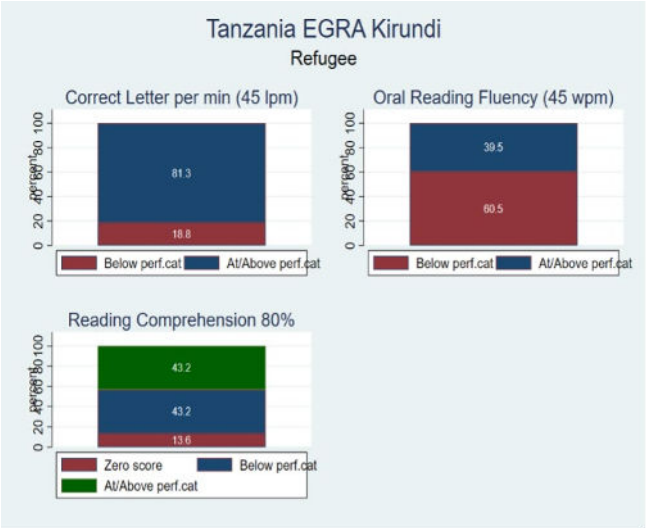
FigureIII. EGRA Performance Categories by Community – Swahili



FigureIII. EGRA Performance Categories by Community – Swahili



FigureIII. EGRA Performance Categories by Community – Swahili



Results for EGMA show that both the host and refugee communities show the best performance in the number identification subtask. In the host community, 61% of learners perform above the performance category, while in the refugee community, 53% of learners perform at that level.
Learners display a similar distribution in the addition and word problem subtasks of the EGMA. In both tasks, approximately 70% of the learners, both in refugee and host communities, fall into the below performance level category.

This result differs for the subtraction subtask in which 72% and 63% of learners in the host and refugee communities, respectively, fall into the above performance category. Furthermore, the subtraction subtask has the larger shares of zero scores for both communities amongst the EGMA subtasks, at roughly 22% for host and refugee learners.

Wellbeing and Socioemotional Skills

Kiddy-KINDL scores suggest that children have moderately favorable wellbeing with Swahili refugee learners reporting higher wellbeing than their Swahili host peers (2.51 and 2.44, respectively) and refugee Kirundi learners reporting similar wellbeing to the

Swahili refugee learners. In terms of empathy, we find that Swahili refugee children report higher levels of empathy than their host peers (0.09 and 0.87 respectively). Results suggest that children have healthy emotional attribution skills with ACES scores of 0.67, and 0.70 for host and refugee Swahili learners respectively, and 0.76 for refugee Kirundi learners. We find no statistically significant differences between refugees and host community children.
For all the measures of bullying, we find statistically and significant differences for host and refugee Swahili learners. In terms of attitudes towards bullying, we find that host learners report significantly higher attitudes of upstand perpetrators than host learners (for all upstand, upstand and befriend victim, and upstand and report to authorities). While Swahili host and refugee learners report similar levels of exposure to bullying, refugee learners report significantly higher victimization rates than host learners. Results suggest that Swahili host learners display higher levels of attribution bias compared with their refugee peers. However, host learners also display a stronger emotional orientation towards calmness compared with the refugee learners. On the other hand, refugee learners display stronger emotional orientation towards sadness compared with the host learners. In terms of conflict resolution strategies, Swahili refugee learners lean comparatively more towards aggression than their host learners' peers.

Table VIII. Primary SEL by Language and Community

	Swahili (n=44)			Kirundi (n=44)
Tool: Subtask	**Host (n=388)**	**Refugee (n=202)**	**Difference H-R**	**Refugee**
Kiddy-KINDL: Score (0-4)	2.51	2.44	0.077**	2.49
Empathy Score (0-1)	0.87	0.92	-0.048***	0.90
ACES Score (0-1)	0.67	0.70	-0.028	0.76
Bullying: Attitude: Disengagement (0-3)	1.31	1.31	-0.001	1.01
Bullying: Attitude: Join Perp. (0-3)	1.16	1.16	-0.004	0.78
Bullying: Attitude: Upstand Perp. (0-3)	2.06	1.84	0.219***	2.46
Bullying: Attitude: Upstand & Befriend Victim (0-3)	2.11	1.99	0.128***	2.49
Bullying: Attitude: Upstand Report Auth. (0-3)	2.10	1.91	0.183***	2.51
Bullying: Exposure to (0-1)	0.39	0.41	-0.023	0.31
Bullying: Victimization (0-1)	0.57	0.85	-0.282***	0.18
SERAIS: Hostile Attribution Bias Score (0-1)	0.59	0.53	0.052**	0.48
SERAIS: Emotional Orientation-Calmness (0-1)	0.36	0.25	0.106***	0.42
SERAIS: Emotional Orientation-Sadness (0-1)	0.38	0.48	-0.102***	0.54
SERAIS: Emotional Orientation-Angry (0-1)	0.29	0.30	-0.008	0.18
SERAIS: Conflict Resolution-Disengagement (0-1)	0.23	0.19	0.040	0.05
SERAIS: Conflict Resolution-Problem Solving (0-1)	0.70	0.70	-0.005	0.89
SERAIS: Conflict Resolution-Aggression (0-1)	0.08	0.11	-0.035**	0.06

*** p<0.01, ** p<0.05, * p<0.1

For the wellbeing and SEL measures (in Annex), we also analyzed within-language differences and while we find no differences for Kirundi learners in any SEL measure, we find that for Swahili learners, host boys report significantly higher attitudes of upstand perpetrators than host girls, while the difference is opposite for boys and girls in the refugee community.

For the refugee community we also find a statistically significant difference between boys and girls, with more boys reporting joining perpetrators behavior upon bullying instances. We find differences in the rates of observing or being exposed to bullying in terms of sex.

For the host community, girls reported higher and statistically significant differences compared with boys, while in the refugee community, we find the opposite with boys reporting higher and statistically significant differences compared with girls. In terms of experiencing bullying children, in the refugee community boys report higher victimization rates than girls. We also find differences in terms of the conflict resolution approaches of Swahili learners in terms of sex and community.

First, we find that for both refugee and host communities, girls report higher levels of resolving conflict through aggression. Second, we find that girls in the host community report higher levels of resolving conflict through problem solving compared with boys, while the difference is opposite for boys and girls in the refugee community, with more boys using problem solving compared than girls. Third, we find that for the host community boys report higher levels of resolving conflict by disengaging than girls.

Variation in Primary Child Outcomes

To answer research question 3, we use inferential analyses using OLS regression modelling (Tables 9 and 10) to explore the relationship between child (sex, community type, age, socioeconomic status, reading resources at home, and repeated grade). Due to limitations during data collection, we could not match child data with teacher characteristics, for ECD learners. Results indicate speaking Swahili remains a strong predictor of lower performance for reading and mathematics outcomes, but positive performance for empathy. Similarly, being a refugee has a negative relationship with learning outcomes, particularly with oral reading fluency, reading comprehension,

number identification, and addition, but a positive relationship with empathy. Being a girl has a negative relationship with most learning outcomes, particularly correct letters per minute and addition. Repeating grade(s) also has a negative relationship with most outcomes, particularly oral reading fluency, reading comprehension, subtraction, and wellbeing. For all outcomes, age is positively related, i.e., older students perform better in all subtasks. Finally, having reading resources at home has a positive relationship with oral reding fluency, and higher scores of socioeconomic status has a positive relationship with oral comprehension.

Table IX. Determinants of Primary Child Outcomes – EGRA, Empathy and Wellbeing

Variables	(1) clpm	(2) orf	(3) Read Comp	(4) Empathy	(5) KiddyKindl:
Language (1=Swahili)	-0.70***	-0.50***	-0.75***	0.04*	-0.07
	(0.04)	(0.08)	(0.09)	(0.02)	(0.06)
1=Girl	-0.11**	0.09	0.11	0.02	-0.01
	(0.04)	(0.08)	(0.06)	(0.02)	(0.03)
Community (1=Refugee)	0.01	-0.37***	-0.50***	0.05*	-0.06
	(0.04)	(0.07)	(0.10)	(0.02)	(0.08)
Pupil Age	0.08***	0.15***	0.17***	0.02***	0.04***
	(0.01)	(0.02)	(0.02)	(0.00)	(0.01)
Times Repeated Grade	-0.03	-0.15***	-0.15***	0.00	-0.08***
	(0.04)	(0.03)	(0.03)	(0.01)	(0.01)
Reading Resources at Home (1=yes)	0.04	0.23*	0.22	0.01	0.09*
	(0.03)	(0.10)	(0.13)	(0.03)	(0.04)
Scores for Factor 1	0.00	0.02	0.02**	-0.01	-0.01
	(0.03)	(0.02)	(0.01)	(0.01)	(0.01)
Constant	0.95***	0.18	0.02	0.62***	2.28***
	(0.16)	(0.26)	(0.29)	(0.06)	(0.15)
Observations	620	626	634	634	634
R-Squared	0.31	0.40	0.43	0.10	0.10

Table X. Determinants of Primary Child Outcomes - EGMA

Variables	(1) clpm	(2) orf	(3) Read Comp	(4) Empathy
Language (1=Swahili)	-0.40***	-0.64***	-0.80***	-0.27***
	(0.03)	(0.03)	(0.09)	(0.04)
1=Girl	0.02	-0.10***	-0.01	-0.04
	(0.05)	(0.03)	(0.06)	(0.05)
Community (1=Refugee)	-0.14**	-0.07*	-0.05	-0.06
	(0.05)	(0.03)	(0.08)	(0.03)
Pupil Age	0.12***	0.09***	0.06***	0.09***
	(0.01)	(0.02)	(0.01)	(0.01)
Times Repeated Grade	-0.04	-0.04	-0.09**	-0.02
	(0.03)	(0.03)	(0.03)	(0.04)
Reading Resources at Home (yes/no)	0.12	0.01	0.05	0.04
	(0.08)	(0.06)	(0.10)	(0.07)
Scores for Factor 1	-0.01	-0.00	-0.01	-0.02
	(0.01)	(0.01)	(0.02)	(0.01)
Constant	0.72***	0.88***	1.10***	0.32***
	(0.18)	(0.23)	(0.22)	(0.09)
Observations	634	634	634	634
R-Squared	0.39	0.31	0.24	0.23

For Tables 9 and 10 robust standard errors in parentheses, clustered at school level. *** p<0.01, ** p<0.05, * p<0.1

Conclusions and Recommendations

Conclusions

1. Both ECD and Primary learners display low levels of learning outcomes, perform better at foundational skills, but struggle with higher order skills, suggesting children are still developing the foundational skills for each education level.

a. For IDELA, overall, results suggest that: both host and refugee communities scored higher in Fine motor, followed by Executive Function, and Socioemotional skills, with higher order skills (Emergent Numeracy and Literacy) lagging with learners scoring, on average, less than one third of the domains correctly.

b. For EGRA, Swahili refugee and host learners perform similarly for the foundational skill of letter identification, but host learners perform better than refugee learners for both oral reading fluency and reading comprehension. Kirundi refugee learners display considerably better performance reading comprehension skills than the Swahili refugee learners.

c. For EGMA, both refugee and host communities score better in foundational skills like number identification than in higher-order skills of subtraction and addition. For EGMA's number identification, the percentage of host learners who performed at or above performance category outperforms the percentage of refugee learners. However, for the rest of the EGMA subtasks, host and refugee learners scored similar in terms of performance categories.

2. For both ECD and Primary, results suggest that children have moderately favorable views of their wellbeing. Primary refugee Swahili learners reported higher wellbeing than their host Swahili learners and refugee Kirundi learners reporting similar wellbeing to the Swahili refugee learners.

3. For primary, results suggest that both host and refugee Swahili learners' have healthy emotional attribution. Primary Swahili host learners report significantly higher attitudes of upstand perpetrators than host learners and that while both and refugee learners report similar levels of exposure to bullying, refugee learners report significantly higher victimization rates than host learners.

4. Swahili primary learners host and refugee learners report similar levels of exposure to bullying, refugee learners report significantly higher victimization rates than host learners. However, host learners also display a stronger emotional orientation towards calmness, compared with the refugee learners; and refugee learners display stronger emotional orientation towards sadness compared with the host learners. Refugee learners also lean comparatively more towards aggression than their host learners' peers.

5. Results depict significantly different results and classroom experiences of girls and boys, though differences are mixed by community, languages, levels, and types of outcomes.

a. Differences by sex at the community aggregate for IDELA, indicate that boy refugees have stronger skills than refugee girls, particularly for the Fine motor domain, Approaches to Learning, and the Overall IDELA score. None of the differences among girls and boys for IDELA in the host community were statistically significant.

b. For ECD, disaggregation analyses indicated more positive wellbeing views of girls and host community children, though differences were not statistically significant, nor were for emotional attribution accuracy.

Recommendations

Based on the findings of this baseline study the following can be suggested as the way forward.

Learners

1. Emphasize foundational skills:
Learners', both ECD and Primary, performance is considerably low. If the project intends to target children's academic development, the program should focus on learning activities that strengthen foundational skills to build towards a holistic development of higher-order skills, such as numeracy and literacy.

2. Target and monitor differential experiences and outcomes of girls and boys:
Given the gender gap differences uncovered by this reporting, program design and implementation should follow equity-focused principles to improve and enhance girls' performance and outcomes, particularly in terms of learning differences. The plausible different experiences and differential outcomes of girls and boys should continue to be monitored closely, particularly during future implementation activities and further research.

3. Reconsider measurement of wellbeing:
Though levels of overall wellbeing were similar across subgroups, the tool selected to measure wellbeing (Kiddy-KINDL) did not reach the psychometric requirements to be analyzed at the sub-domain level. Thus, we are only able to explore results at the overall score (with acceptable validity and reliability estimates) but unable to explore nuances in the different types of wellbeing. However, given the finding of the relationship between teacher sex and learners' wellbeing, it is warranted to continue exploring the measurement of wellbeing and its different components with children's outcomes.

4. Continue measurement research and analyses relevant outcomes:
Certain measurement analyses of both learning outcomes and particular socioemotional skills were not feasible given small sample sizes (particularly for exploring differences by language) and the lack of endline data collection. Further measurement analysis is warranted for the next stages of the program.

Educators

5. Provide and strengthen professional development to educators:
The design and implementation of professional development activities for educators should be tailored to specific priorities and the design of PlayMatters moving forward. Such activities should be designed to deepen educators' understanding of LtP and child-centered pedagogy. For example, activities could include providing explicit training on inquiry-based learning, cooperative learning, and hands-on activities.

6. Support educators' development of specific LtP activities:
Equipping teachers with the necessary skills and strategies to create a student-centered classroom environment will require continuous monitoring and research. Particular attention should be placed on assessing changes in educator's instructional skills and other in-classroom experiences.

Programmatic Research

7. Align research tools with the PlayMatters' ToC and definitions:
While the present study used a set of defined tools aligned with the initial design of the program, any future research tools should be aligned with the specific details of the implementation to enhance the accuracy, relevance, and comprehensiveness of the research findings. This is particularly relevant for the educators' measures since the tools selected for the baseline were aligned with PM's early definition of LtP, which has since, evolved and further delimited.

8. Define reach of PM educators' outcomes:
At the time of the baseline planning, research educators' tools were selected in alignment with the initial plans for educators' teacher professional development activities and the desired outcomes at the educator level. However, since the planning stage, more refined and targeted educators' outcomes have been set within a more explicit framework of educators' skill development. Within the programmatic design and activities, the refinement of the educator outcomes is necessary to continue to develop and be refined, while maintaining a close and aligned relationship with research plans, tools, and reach.

9. Conduct careful monitoring of educators' instructional skills through implementation research and M&E:
Implementation research should integrate measures and indicators that directly assess the key aspects and outcomes of the implementation. Future implementation research should also strive to be of mixed nature so that teachers' voices and experiences are not missed.

10. Refine research samples and targets, in alignment with the deployment of program implementation:
Both measurement research and disaggregations of certain key outcomes were not feasible given the reached sample sizes. Though the research and implementation teams conducted several discussions on the reach of research activities and samples, the baseline data collection served as an exercise of confirming data previously collected (such as schools and classroom registrations) and revealed gaps in understanding and in the information collected. Research activities can, have, and will continue to help with providing relevant information to program activities in terms of reach, scale, and limits. As the program moves to a larger scale, aligning the implemented activities with research activities, for example in terms of cohorts of educators and children in the program, will be not only necessary but particularly relevant to ensure the validity and relevance of the research activities.

11. Strive for parsimony in research tools and methods:
During data collection planning and activities, teams encountered unexpected technical limitations of software and data collection tools in terms of display, connectivity, compatibility, and deployment. The baseline data collection served as a learning opportunity to explore such technical considerations which need to be considered for future research activities.

References

Adiba, G., & Campbell, A. (2019). Learning Through Play: A review of evidence related to play for learners' education and development in humanitarian and low-resource contexts. PlayMatters.

Bullinger, M., Brütt, A. L., Erhart, M., & Ravens-Sieberer, U.; BELLA Study Group. (2008). Psychometric properties of the KINDL-R questionnaire: results of the BELLA study. European Child & Adolescent Psychiatry, 17(Suppl 1), 125-132. doi: 10.1007/s00787-008-1014-z

Diazgranados, S., Selman, R. L., & Dionne, M. M. (2016). Acts of Social Perspective Taking: A Functional Construct and the Validation of a Performance Measure for Early Adolescents. Social Development, 25, 572-601.

Inter-Agency Network for Education in Emergencies (INEE). (2019). 2019 Annual Report. Retrieved from https://inee.org/resources/inee-2019-annual-report.

Kim, H. Y., & Tubbs Dolan, C. (2019, September). Social emotional response and information scenarios: Evidence on construct validity, measurement invariance, and reliability in use with Syrian refugee children in Lebanon. Technical working paper. New York, NY: NYU Global TIES for Children.

Komba, A., & Shukia, R. (2023). An Analysis of the Basic Education Curriculum in Tanzania: The Integration, Scope, and Sequence of 21st Century Skills. RISE Working Paper Series 23/129. doi: 10.35489/BSG-RISE-WP_2023/129.

Ministry of Education and Culture. (1995). Education and Training Policy. MoEC.

Ministry of Education, Science and Technology. (2019). National Curriculum Framework for Basic and Teacher Education. Tanzania Institute of Education.

Ministry of Education and Vocational Training (MoEVT). (2014). Education and training policy. MoEVT.

Ndijuye, L. G., Mligo, I. R., & Maregesi, A. (2020). Early childhood education in Tanzania: Views and beliefs of stakeholders on its status and development. Global Education Review, 7(3), 22-39.

Osofsky, J. D. (1999). The impact of violence on learners. The Future of Learners Domestic Violence and Learners, 9(3), 33-49. Retrieved from https://www.jstor.org/stable/1602780.

Pisani, L., Borisova, I., & Dowd, A. J. (2018). Developing and validating the International Development and Early Learning Assessment (IDELA). International Journal of Educational Research, 91, 1-15. doi: 10.1016/j.ijer.2018.06.007.

Seidman, E., Kim, S., Raza, M., Ishihara, M., & Halpin, P. F. (2018). Assessment of pedagogical practices and processes in low and middle-income countries: Findings from secondary school classrooms in Uganda. Teaching and Teacher Education, 71, 283-296. https://doi.org/10.1016/j.tate.2017.12.017.

Schaufeli, W. B., Bakker, A. B., & Salanova, M. (2005). The Measurement of Work Engagement With a Short Questionnaire: A Cross-National Study. European Journal of Psychological Assessment, 21(3), 161-170. https://doi.org/10.1177/0013164405282471.

Tschannen-Moran, M., & Woolfolk Hoy, A. (2001). Teacher Efficacy: Capturing an Elusive Construct. Teaching and Teacher Education, 17, 783-805. https://psycnet.apa.org/doi/10.1016/S0742-051X(01)00036-1.

Uwezo. (2013). Are our children learning? Uwezo Tanzania Report 2012. Retrieved from http://www.twaweza.org/go/uwezo-tanzania-2012ala.

Vygotsky, I. (2016). Play and its role in mental development of the child. International Research in Early Childhood Education, 7(2), 3-25.

Yogman, M., Garner, A., Hutchinson, J., Hirsh-Pasek, K., & Golinkoff, R. M. (2018). The Power of Play: A Pediatric Role in Enhancing Development in Young Children. Pediatrics, 142(3), e20182058. https://doi.org/10.1542/peds.2018-2058

Yohani, S. C., & Larsen, D. J. (2009). Hope lives in the heart: Refugee and Immigrant learners' perceptions of hope and hope-engendering sources during early years of adjustment. Canadian Journal of Counselling, 43(4), 246-264.

Zosh, J. M., Hopkins, E. J., Jensen, H., Liu, C., Neale, D., Hirsh-Pasek, K., Solis, S. L., & Whitebread, D. (2017). Learning through play: A review of the evidence. LEGO Foundation.

Annexes

(1) Instrument's Reliability Summary, by Language.

a) Children - ECD

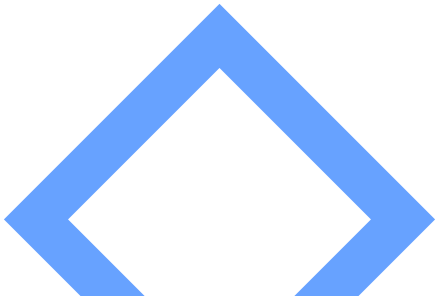
Tool	Variables	Swahili	Kirundi
IDELA	**IDELA Total: Fine Motor, SES, Literacy, Numeracy**	**0.78**	**0.81**
	IDELA Domain: Executive Function	**0.76**	**0.86**
	IDELA: Memory	0.54	0.67
	IDELA: Self-Regulation	0.96	0.83
	IDELA Domain: Fine Motor	**0.76**	**0.85**
	IDELA: Copying Shape	0.73	0.73
	IDELA: Drawing Human	0.84	0.83
	IDELA: Fold	n/a	n/a
	IDELA: Hopp	n/a	n/a
	IDELA Domain: Emergent Literacy	**0.72**	**0.82**
	IDELA: Vocabulary	n/a	n/a
	IDELA: Print Awareness	0.84	0.82
	IDELA: Letter ID	0.78	0.88
	IDELA: Oral Comprehension	0.79	0.62
	IDELA: Writing	n/a	n/a
	IDELA Domain: Emergent Numeracy	**0.82**	**0.91**
	IDELA: Size	0.60	0.35
	IDELA: Number ID	0.79	0.92
	IDELA: Operations	0.54	0.68
	IDELA: One to One Correspondence	n/a	n/a
	IDELA: Puzzle	n/a	n/a
	IDELA Domain: Socioemotional	**0.67**	**0.61**
	IDELA: Friends	n/a	n/a
	IDELA: Emotional Attribution	0.51	0.60
	IDELA: Empathy	0.55	0.60
	IDELA: Conflict Resolution	0.69	0.70
	IDELA Domain: Persistence	0.79	0.81
Socioemotional	**Kiddy-KINDL: Score (1-2)**	**0.76**	**0.57**

Notes: Reliability is not estimated for subtasks with open ended questions or non-numeric scores and re not included in the Domain's reliability estimates but are shown for clarity.

b) Children - Primary

Tool	Subtask	Swahili	Kirundi
EGRA	All EGRA	0.72	0.72
	Vocabulary	0.81	0.62
	Letter Sound Identification	0.93	0.80
	Oral Reading Passage	0.91	0.74
	Reading Comprehension	0.49	0.50
EGMA	All EGMA	0.91	0.87
	Number Identification	0.97	0.72
	Number Discrimination	0.85	0.56
	Missing Number	0.74	0.58
	Addition L1	0.96	0.61
	Addition L2	0.80	0.74
	Subtraction L1	0.96	0.67
	Subtraction L2	0.84	0.76
	Word Problems	0.69	0.60
Socioemotional	Kiddy-KINDL: Score (1-4)	0.73	0.69
	Empathy Score (0-1)	0.82	0.64
	ACES Score (0-1)	0.61	0.67
	Bullying: Attitude: Disengagement (0-3)	0.66	0.81
	Bullying: Attitude: Join Perp. (0-3)	0.66	0.51
	Bullying: Attitude: Upstand Perp. (0-3)	0.61	0.65
	Bullying: Attitude: Upstand & Befriend Victim (0-3)	0.45	0.64
	Bullying: Attitude: Upstand Report Auth. (0-3)	0.62	0.52
	Bullying: Exposure To (0-1)	0.71	0.75
	Bullying: Victimization (0-1)	0.67	0.32
	SERAIS: Hostile Attribution Bias Score (0-1)	0.46	0.42
	SERAIS: Emotional Orientation-Calmness (0-1)	0.68	0.59
	SERAIS: Emotional Orientation-Sadness (0-1)	0.59	0.35
	SERAIS: Emotional Orientation-Angry (0-1)	0.62	0.60
	SERAIS: Conflict Resolution-Disengagement (0-1)	0.72	0.30
	SERAIS: Conflict Resolution-Problem Solving (0-1)	0.75	0.39
	SERAIS: Conflict Resolution-Agression (0-1)	0.65	0.58

Notes: Reliability is not estimated for subtasks with open ended questions or non-numeric scores and re not included in the Domain's reliability estimates but are shown for clarity.



(2) Learning Assessments' Subtask Correlation

a) ECD – IDELA – Kirundi

		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
(1)	Memory	1	-0.05	0.10	-0.24	-0.24	0.14	-0.12	0.12	0.14	-0.20	0.17
(2)	Self-Regulation	-0.05	1	0.21	0.31	0.27	0.07	0.10	0.39	0.17	0.13	0.17
(3)	Copying Shape	0.10	0.21	1	0.68	0.30	0.48	0.37	0.35	0.36	0.40	0.59
(4)	Drawing Human	-0.24	0.31	0.68	1	0.51	0.42	0.60	0.33	0.21	0.56	0.46
(5)	Fold	-0.24	0.27	0.30	0.51	1	0.20	0.41	0.24	0.17	0.39	0.38
(6)	Hop	0.14	0.07	0.48	0.42	0.20	1	0.47	-0.01	0.23	0.34	0.20
(7)	Vocabulary	-0.12	0.10	0.37	0.60	0.41	0.47	1	0.28	0.24	0.61	0.26
(8)	Print Awareness	0.12	0.39	0.35	0.33	0.24	-0.01	0.28	1	0.52	0.16	0.28
(9)	Letter ID	0.14	0.17	0.36	0.21	0.17	0.23	0.24	0.52	1	0.45	0.15
(10)	Oral Comprehension	-0.20	0.13	0.40	0.56	0.39	0.34	0.61	0.16	0.45	1	0.23
(11)	Writing	0.17	0.17	0.59	0.46	0.38	0.20	0.26	0.28	0.15	0.23	1

		(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)
(12)	Size	1	0.24	0.55	0.31	0.51	0.15	0.42	0.20	0.29
(13)	Number ID	0.24	1	0.38	0.35	0.20	-0.06	0.16	0.17	0.03
(14)	Operations	0.55	0.38	1	0.44	0.36	0.19	0.51	0.38	0.36
(15)	One to One Correspondence	0.31	0.35	0.44	1	0.30	0.02	0.44	0.48	0.19
(16)	Puzzle	0.51	0.20	0.36	0.30	1	-0.01	0.35	0.18	0.08
(17)	Friends	0.15	-0.06	0.19	0.02	-0.01	1	0.20	0.02	0.22
(18)	Emotional Attribution	0.42	0.16	0.51	0.44	0.35	0.20	1	0.60	0.31
(19)	Empathy	0.20	0.17	0.38	0.48	0.18	0.02	0.60	1	0.43
(20)	Conflict Resolution	0.29	0.03	0.36	0.19	0.08	0.22	0.31	0.43	1

b) ECD – IDELA – Swahili

		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
(1)	Memory	1	0.18	0.40	0.26	0.20	0.27	0.21	0.19	0.32	0.28	0.08
(2)	Self-Regulation	0.18	1	0.35	-0.01	0.34	0.12	0.23	0.04	0.20	0.19	0.28
(3)	Copying Shape	0.40	0.35	1	0.32	0.51	0.29	0.37	0.41	0.27	0.30	0.43
(4)	Drawing Human	0.26	-0.01	0.32	1	0.29	0.25	0.31	0.23	0.17	0.32	0.15
(5)	Fold	0.20	0.34	0.51	0.29	1	0.14	0.39	0.24	0.01	0.40	0.20
(6)	Hop	0.27	0.12	0.29	0.25	0.14	1	0.06	0.04	0.07	0.34	0.26
(7)	Vocabulary	0.21	0.23	0.37	0.31	0.39	0.06	1	0.14	0.12	0.47	0.18
(8)	Print Awareness	0.19	0.04	0.41	0.23	0.24	0.04	0.14	1	0.19	0.15	0.12
(9)	Letter ID	0.32	0.20	0.27	0.17	0.01	0.07	0.12	0.19	1	-0.02	0.01
(10)	Oral Comprehension	0.28	0.19	0.30	0.32	0.40	0.34	0.47	0.15	-0.02	1	0.36
(11)	Writing	0.08	0.28	0.43	0.15	0.20	0.26	0.18	0.12	0.01	0.36	1

		(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)
(12)	Size	1	0.23	-0.04	0.18	0.05	0.19	0.01	-0.01	-0.11
(13)	Number ID	0.23	1	0.39	0.54	0.26	0.13	0.12	0.01	0.25
(14)	Operations	-0.04	0.39	1	0.45	0.36	0.18	0.27	0.17	0.27
(15)	One to One Correspondence	0.18	0.54	0.45	1	0.10	0.07	0.01	0.01	0.31
(16)	Puzzle	0.05	0.26	0.36	0.10	1	0.09	0.02	0.18	0.12
(17)	Friends	0.19	0.13	0.18	0.07	0.09	1	0.25	0.15	0.27
(18)	Emotional Attribution	0.01	0.12	0.27	0.01	0.02	0.25	1	0.47	0.39
(19)	Empathy	-0.01	0.01	0.17	0.01	0.18	0.15	0.47	1	0.46
(20)	Conflict Resolution	-0.11	0.25	0.27	0.31	0.12	0.27	0.39	0.46	1

c) Primary - EGRA and EGMA - Kirundi

		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
(1)	Vocabulary	1	0.38	0.46	0.65	0.70	0.61	0.51	0.40	0.63	0.32	0.63	0.62
(2)	Letter Sound Identification	0.38	1	0.65	0.62	0.73	0.71	0.48	0.62	0.62	0.60	0.59	0.55
(3)	Oral Reading Passage	0.46	0.65	1	0.62	0.60	0.60	0.36	0.55	0.55	0.45	0.47	0.62
(4)	Reading Comprehension	0.65	0.62	0.62	1	0.74	0.67	0.44	0.57	0.74	0.66	0.65	0.73
(5)	Number Identification	0.70	0.73	0.60	0.74	1	0.77	0.40	0.51	0.84	0.38	0.65	0.76
(6)	Number Discrimination	0.61	0.71	0.60	0.67	0.77	1	0.57	0.56	0.83	0.48	0.70	0.70
(7)	Missing Number	0.51	0.48	0.36	0.44	0.40	0.57	1	0.46	0.45	0.33	0.41	0.42
(8)	Addition L1	0.40	0.62	0.55	0.57	0.51	0.56	0.46	1	0.55	0.65	0.42	0.54
(9)	Addition L2	0.63	0.62	0.55	0.74	0.84	0.83	0.45	0.55	1	0.42	0.81	0.73
(10)	Subtraction L1	0.32	0.60	0.45	0.66	0.38	0.48	0.33	0.65	0.42	1	0.44	0.35
(11)	Subtraction L2	0.63	0.59	0.47	0.65	0.65	0.70	0.41	0.42	0.81	0.44	1	0.68
(12)	Word Problems	0.62	0.55	0.62	0.73	0.76	0.70	0.42	0.54	0.73	0.35	0.68	1

d) Primary - EGRA and EGMA - Swahili

		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
(1)	Vocabulary	1	0.46	0.36	0.36	0.51	0.53	0.41	0.48	0.44	0.45	0.39	0.56
(2)	Letter Sound Identification	0.46	1	0.62	0.61	0.67	0.64	0.52	0.63	0.54	0.52	0.46	0.50
(3)	Oral Reading Passage	0.36	0.62	1	0.93	0.73	0.67	0.67	0.71	0.65	0.55	0.53	0.47
(4)	Reading Comprehension	0.36	0.61	0.93	1	0.70	0.66	0.66	0.68	0.62	0.53	0.55	0.50
(5)	Number Identification	0.51	0.67	0.73	0.70	1	0.82	0.64	0.75	0.64	0.50	0.48	0.51
(6)	Number Discrimination	0.53	0.64	0.67	0.66	0.82	1	0.65	0.75	0.68	0.56	0.53	0.52
(7)	Missing Number	0.41	0.52	0.67	0.66	0.64	0.65	1	0.67	0.60	0.52	0.55	0.52
(8)	Addition L1	0.48	0.63	0.71	0.68	0.75	0.75	0.67	1	0.70	0.70	0.53	0.60
(9)	Addition L2	0.44	0.54	0.65	0.62	0.64	0.68	0.60	0.70	1	0.62	0.78	0.47
(10)	Subtraction L1	0.45	0.52	0.55	0.53	0.50	0.56	0.52	0.70	0.62	1	0.63	0.49
(11)	Subtraction L2	0.39	0.46	0.53	0.55	0.48	0.53	0.55	0.53	0.78	0.63	1	0.39
(12)	Word Problems	0.56	0.50	0.47	0.50	0.51	0.52	0.52	0.60	0.47	0.49	0.39	1

(3) ECD – IDELA Subtasks: Differences by Language and Community

		Swahili						Kirundi	
Variable	**N**	**(1) Host Mean/SE**	**N**	**(2) Refugee Mean/SE**	**N**	**(3) Total Mean/SE**	**T-Test Difference (1)-(2)**	**N**	**(1) Refugee Mean/SE**
Memory % Correct	37	0.385 [0.037]	30	0.533 [0.043]	67	0.451 [0.029]	-0.148**	40	0.575 [0.050]
Self Regulation % Correct	37	0.368 [0.069]	30	0.493 [0.077]	67	0.424 [0.051]	-0.126	40	0.460 [0.074]
Shape % Correct	37	0.345 [0.072]	30	0.567 [0.086]	67	0.444 [0.056]	-0.222**	40	0.313 [0.063]
Human drawing % Correct	37	0.338 [0.045]	30	0.396 [0.058]	67	0.364 [0.036]	-0.058	40	0.297 [0.045]
Fold % Correct	37	0.486 [0.037]	30	0.533 [0.057]	67	0.507 [0.033]	-0.047	39	0.372 [0.045]
Hopp % Correct	37	0.749 [0.049]	29	0.669 [0.058]	66	0.714 [0.037]	0.080	38	0.645 [0.052]
Vocab % Correct	37	0.262 [0.026]	30	0.395 [0.029]	67	0.322 [0.021]	-0.133***	40	0.278 [0.029]
Print Awareness % Correct	37	0.342 [0.064]	30	0.444 [0.080]	67	0.388 [0.050]	-0.102	40	0.258 [0.058]
Letter ID % Correct	37	0.024 [0.009]	30	0.032 [0.014]	67	0.028 [0.008]	-0.007	40	0.055 [0.017]
Oral Comp % Correct	37	0.281 [0.052]	30	0.467 [0.060]	67	0.364 [0.041]	-0.186**	40	0.315 [0.047]
Writing % Correct	34	0.206 [0.036]	28	0.357 [0.042]	62	0.274 [0.029]	-0.151***	40	0.281 [0.030]
Size Information % Correct	37	0.464 [0.032]	30	0.467 [0.039]	67	0.465 [0.025]	-0.003	40	0.446 [0.028]
Number ID % Correct	37	0.061 [0.016]	30	0.080 [0.018]	67	0.069 [0.012]	-0.019	40	0.078 [0.026]
Operations % Correct	37	0.432 [0.055]	30	0.422 [0.053]	67	0.428 [0.038]	0.010	40	0.417 [0.058]
One2One % Correct	37	0.261 [0.057]	30	0.156 [0.052]	67	0.214 [0.039]	0.106	40	0.383 [0.057]
Puzzle % Correct	32	0.119 [0.032]	28	0.136 [0.033]	60	0.127 [0.023]	-0.017	38	0.258 [0.043]
Friends % Correct	37	0.462 [0.038]	29	0.441 [0.043]	66	0.453 [0.028]	0.021	40	0.445 [0.036]
Emotional Attribution % Correct	37	0.155 [0.033]	30	0.267 [0.061]	67	0.205 [0.033]	-0.111*	40	0.287 [0.053]
Empathy % Correct	37	0.324 [0.054]	30	0.622 [0.065]	67	0.458 [0.045]	-0.298***	40	0.400 [0.055]
Conflict Resolution % Correct	37	0.351 [0.061]	30	0.583 [0.076]	67	0.455 [0.050]	-0.232**	40	0.438 [0.067]

Notes: The value displayed for t-tests are the differences in the means across the groups, sample permitting ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

(4) ECD – IDELA Domains: Differences by Community and Sex

	Host							Refugee						
Variable	**N**	**(1) Boy Mean/SE**	**N**	**(2) Girl Mean/SE**	**N**	**(3) Total Mean/SE**	**T-Test Difference (1)-(2)**	**N**	**(1) Boy Mean/SE**	**N**	**(2) Girl Mean/SE**	**N**	**(3) Total Mean/SE**	**T-Test Difference (1)-(2)**
IDELA Total Score	17	0.335 [0.026]	13	0.357 [0.045]	30	0.345 [0.024]	-0.022	29	0.425 [0.031]	33	0.344 [0.027]	62	0.382 [0.021]	0.082*
IDELA Domain Score: Fine Motor	18	0.516 [0.050]	19	0.444 [0.048]	37	0.479 [0.035]	0.072	30	0.574 [0.045]	36	0.397 [0.040]	66	0.477 [0.032]	0.177***
IDELA Domain Score: Emergent Literacy	18	0.228 [0.032]	16	0.225 [0.034]	34	0.226 [0.023]	0.003	31	0.301 [0.028]	37	0.262 [0.028]	68	0.280 [0.020]	0.039
IDELA Domain Score: Emergent Numeracy	17	0.271 [0.030]	15	0.295 [0.049]	32	0.282 [0.028]	-0.024	31	0.305 [0.030]	35	0.285 [0.031]	66	0.294 [0.021]	0.020
IDELA Domain Score: Socio-Emotional	18	0.312 [0.031]	19	0.334 [0.056]	37	0.323 [0.032]	-0.022	33	0.475 [0.047]	36	0.396 [0.035]	69	0.434 [0.029]	0.078
IDELA Domain Score: Executive Function	18	0.350 [0.057]	19	0.401 [0.056]	37	0.376 [0.040]	-0.051	33	0.546 [0.044]	37	0.489 [0.048]	70	0.516 [0.032]	0.058
IDELA Domain Score: Approaches to Learning (persist)	14	0.685 [0.066]	13	0.756 [0.043]	27	0.719 [0.040]	-0.072	26	0.872 [0.025]	27	0.769 [0.043]	53	0.819 [0.026]	0.103**

Notes: The value displayed for t-tests are the differences in the means across the groups, sample permitting ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

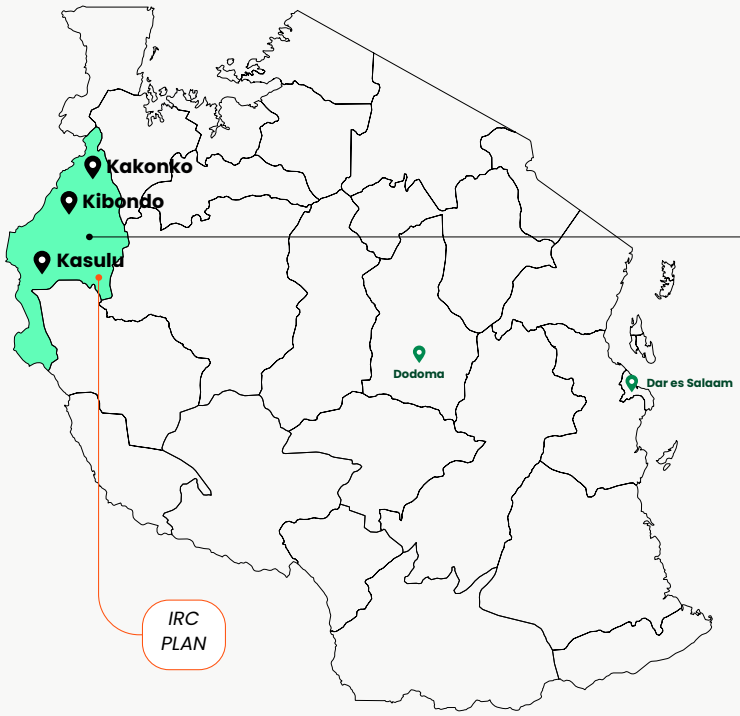
(5) Wellbeing and SEL – Primary by Community and Sex – Swahili

	Swahili (n=590)								
	Host (n=388)			Refugee (n=202)			Overall		
Tool: Subtask	**Boys (n=198)**	**Girls (n=190)**	**Difference**	**Boys (n=198)**	**Girls (n=190)**	**Difference**	**Boys (n=198)**	**Girls (n=190)**	**Difference**
Kiddy-KINDL: Score (0-4)	2.51	2.52	-0.014	2.47	2.40	0.070	2.49	2.48	0.015
Empathy Score (0-1)	0.86	0.88	-0.014	0.90	0.94	-0.034	0.88	0.90	-0.022
ACES Score (0-1)	0.67	0.68	-0.009	0.72	0.68	0.034	0.68	0.68	0.005
Bullying: Attitude: Disengagement (0-3)	1.30	1.31	-0.012	1.35	1.27	0.077	1.32	1.30	0.019
Bullying: Attitude: Join Perp. (0-3)	1.16	1.15	0.010	1.27	1.06	0.203***	1.20	1.12	0.076
Bullying: Attitude: Upstand Perp. (0-3)	2.11	1.99	0.121**	1.76	1.91	-0.150**	2.00	1.96	0.031
Bullying: Attitude: Upstand & Befriend Victim (0-3)	2.09	2.14	-0.048	1.96	2.01	-0.050	2.05	2.09	-0.047
Bullying: Attitude: Upstand Report Auth. (0-3)	2.10	2.09	0.009	1.91	1.92	-0.005	2.04	2.03	0.007
Bullying: Exposure to (0-1)	0.34	0.44	-0.101***	0.49	0.33	0.157***	0.39	0.40	-0.013
Bullying: Victimization (0-1)	0.59	0.54	0.049	1.04	0.66	0.378***	0.74	0.58	0.158**
SERAIS: Hostile Attribution Bias Score (0-1)	0.60	0.57	0.036	0.52	0.54	-0.017	0.58	0.56	0.018
SERAIS: Emotional Orientation-Calmness (0-1)	0.35	0.37	-0.017	0.27	0.24	0.029	0.32	0.32	0.000
SERAIS: Emotional Orientation-Sadness (0-1)	0.39	0.37	0.026	0.46	0.51	-0.046	0.42	0.42	-0.000
SERAIS: Emotional Orientation-Angry (0-1)	0.31	0.27	0.033	0.30	0.29	0.010	0.31	0.28	0.025
SERAIS: Conflict Resolution-Disengagement (0-1)	0.28	0.16	0.119***	0.18	0.19	-0.007	0.25	0.17	0.076***
SERAIS: Conflict Resolution-Problem Solving (0-1)	0.67	0.73	-0.066*	0.77	0.64	0.128***	0.70	0.70	0.002
SERAIS: Conflict Resolution-Aggression (0-1)	0.05	0.10	-0.053***	0.05	0.17	-0.122***	0.05	0.13	-0.078***

(6) Wellbeing and SEL – Primary by Community and Sex – Kirundi

	Kirundi (n=44)		
	Refugee (n=202)		
Tool: Subtask	**Boys (n=22)**	**Girls (n=22)**	**Difference**
Kiddy-KINDL: Score (0-4)	2.54	2.44	0.09
Empathy Score (0-1)	0.91	0.89	0.02
ACES Score (0-1)	0.72	0.79	-0.07
Bullying: Attitude: Disengagement (0-3)	0.96	1.06	-0.11
Bullying: Attitude: Join Perp. (0-3)	0.74	0.82	-0.08
Bullying: Attitude: Upstand Perp. (0-3)	2.40	2.52	-0.12
Bullying: Attitude: Upstand & Befriend Victim (0-3)	2.45	2.53	-0.08
Bullying: Attitude: Upstand Report Auth. (0-3)	2.50	2.52	-0.03
Bullying: Exposure to (0-1)	0.36	0.26	0.10
Bullying: Victimization (0-1)	0.23	0.13	0.10
SERAIS: Hostile Attribution Bias Score (0-1)	0.44	0.52	-0.08
SERAIS: Emotional Orientation-Calmness (0-1)	0.40	0.44	-0.04
SERAIS: Emotional Orientation-Sadness (0-1)	0.56	0.51	0.06
SERAIS: Emotional Orientation-Angry (0-1)	0.17	0.19	-0.02
SERAIS: Conflict Resolution-Disengagement (0-1)	0.06	0.04	0.02
SERAIS: Conflict Resolution-Problem Solving (0-1)	0.89	0.90	-0.01
SERAIS: Conflict Resolution-Aggression (0-1)	0.06	0.06	-0.01

Where We are Working



Districts and Respective Refugee Settlements/Camps

Kasulu
Nyarugusu Camp

Kibondo
Nduta Camp

Kakonko
Mtendeli Camp

International Rescue Committee (IRC)
Plan International (PLAN)

Tanzania

PlayMatters Regional Office

Plot 8, Lower Naguru East Road, Kampala, Uganda
+256 (0) 394 822 224
+256 (0) 200 900 697

www.playmatters.org
www.rescue.org/playmatters