PlayMatters

Implementation Baseline

Primary

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:

Dr. Abraha Asfaw, Principal Investigator
Dr. Daniel Tefera, Principal Investigator
Dr. Kassa Michael, Principal Investigator
Asham Assezenew, Senior Research Manager, IRC Ethiopia
Idalia Rodriguez Morales, Researcher, IRC – Airbel Impact Lab
Dr. Silvia Diazgranados Ferráns, Associate Director, Education Research, IRC – Airbel Impact Lab

Ethiopia

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 baseline study team is indebted to the children, educators, and head teachers for their assent/consent and participation in the present study. Our appreciations also go to International Rescue Committee (IRC) and Plan International Ethiopia (PIE) project teams in Somali, Gambella, and Benishangul Gumuz (BGR) Regions for their support in enumerator selection, training and data collection and monitoring.

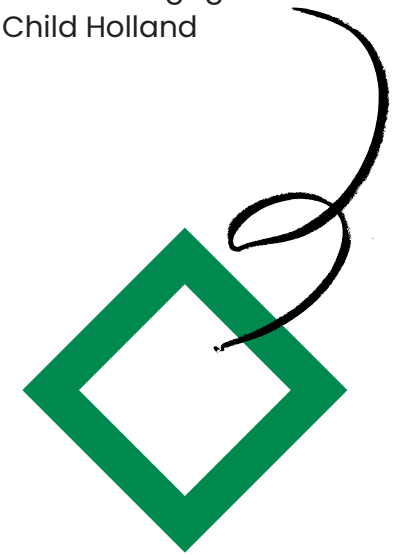
The study team would like to thank the guidance and support from PlayMatters project Research, Monitoring, Evaluation, Accountability and Learning (RMEAL) and the PlayMatters Regional RMEAL team which were remarkable in terms of providing guidance and leadership, organizing capacity building workshops, sharing documents, mobilizing resources, analyzing data and reviewing and providing inputs for draft reports throughout the study and draft report.

The data collection involved enumerators from respective sites who speak the language of the children. They played very pivotal roles in collecting quality and complete data. Thus, the team is very grateful to all.

Contents

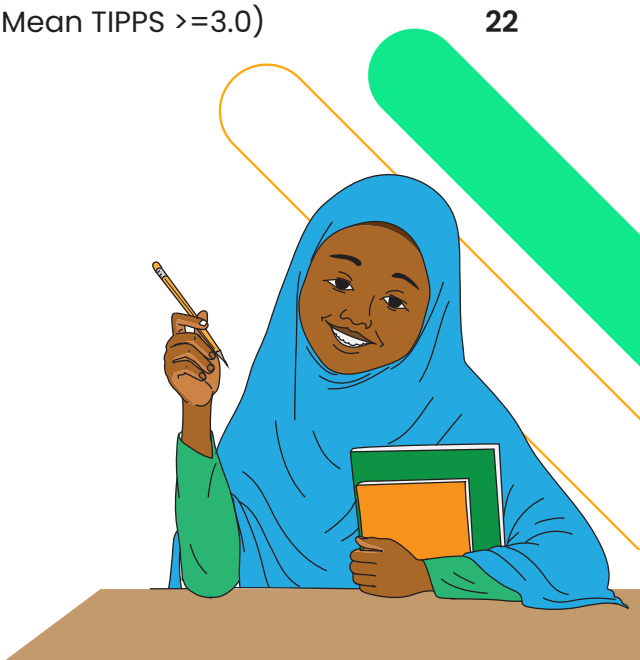
Acronyms/Abbreviations

BGR	Benishangul Gumuz Region
BIT	Behavioral Insights Team
DEAS	Developmental and Educational Activities Scale
ECCE	Early Childhood Care and Education
ECD	Early Childhood Development
EGRA	Early Grade Reading Assessment
EGMA	Early Grade Mathematics Assessment
ESSSWA	Ethiopian Society of Sociologists, Social Workers, and Anthropologists
GER	Gross Enrolment Rate
PM	Play Matters
IPA	Innovations for Poverty Action
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
PIE	Plan International Ethiopia
RMEAL	Research, Monitoring, Evaluation, Accountability and Learning
SEL	Socio-emotional Learning
SERAIS	Socio-Emotional Response and Information Scenarios
TIPPS	Teacher Instructional Practices and Processes System
ToT	Training of Trainers
UNHCR	United Nations High Commissioner for Refugees
UWES	Utrecht Work Engagement Scale
WCH	War Child Holland



List of Tables

List of Figures



Executive Summary

PlayMatters (PM) is a teacher professional development (TPD) program implemented by a consortium led by the International Rescue Committee (IRC) that includes the Behavioral Insights Team (BIT), Innovations for Poverty Action (IPA), Plan International (Plan), and War Child Holland (WCH) (together, "The Consortium"), in partnership with the LEGO Foundation. PM works through existing education systems to provide Early Childhood Development (ECD) and primary teachers in Ethiopia, Tanzania, and Uganda with the skills, motivations, 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. PM in Ethiopia intended to define initial characteristics of children and educators in PM intervention refugee and host community.

The present report presents findings from PM's baseline data collection for Ethiopia in March 2022. Thus, this study was guided by the following basic questions:

I. Child Level

1. What are learners' literacy and numeracy skills? How do they vary by displacement, sex, and language?

2. What is the status of learners' socioemotional skills and wellbeing? How do they vary by displacement, sex, and language?

3. What are the child characteristics explaining the variation in learners' literacy, numeracy, and socio-emotional SEL skills?

II. Educator Level

1. What is the perception of educators on LtP?

2. What are teachers' instructional practices?

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

The study employed systematic random sampling technique to identify educators and children from the PM intervention refugee and host communities in Somali, Gambella and BGR. A total of 290 primary children (147 girls and 143 boys) and 60 primary educators participated in this baseline study. Children's 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 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), 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.

Conclusions and Major Findings

A. Child Level

For EGRA and EGMA, results depicted struggling learners, with varying differences by of sex, community type and region. Displacement status (i.e., the type of community) is not a factor that explains the variation for EGRA, though it does for certain EGMA subtasks. Regional variations vary by subtasks, though generally children in Somali region outperform those in Gambella and BGR. The difference among girls and boys is stark and pronounced, with boys outperforming girls for all subtasks. Children display high socioemotional skills (empathy and emotional attribution accuracy) and report low instances of bullying and victimization but report low levels of wellbeing.

1. The overall EGRA and EGMA results suggest learners have challenges in both literacy and numeracy, with results suggesting they are still in early stages of their learning development. For EGRA, children scored higher in foundational subtasks, but struggled with more complex ones. Children were able to identify about two thirds of the vocabulary words depicted in images, but 48% of children were not able to read a single word correctly, and very few were able to read a short text paragraph and comprehend what they read. Similar to the EGRA results, children scored higher in the most basic numeric subtasks (number identification and number discrimination) and struggled with more complex and higher-order subtasks (missing number, addition and subtraction and word problems).

a. There are stark differences between girls and boys: for most EGRA subtasks (letter identification and oral reading) and all EGMA subtasks, with boys outperforming girls in all the cases. For the number identification, number discrimination, and Addition L2 subtasks, the refugee community outperformed the host community, though due to limited sample sizes it is not possible to disaggregate by region or country of origin. As expected, children in upper grades outperform those in lower grades.

2. Children's empathy and emotional attribution accuracy were relatively high, suggesting that children can adequately recognize and encode feelings from common situations, and have feelings of empathy for situation others experience. We found no evidence of differences in terms of sex or community type. On general terms, children rated their wellbeing unfavorably, particularly their physical wellbeing. Of all the wellbeing domains measured, children reported highest rates at school and among their friends. Boys report higher wellbeing than girls in the friends dimension, and children from host communities report higher family wellbeing than those in the refugee community. In terms of bullying, children reported positive attitudes with most of students reporting that they confront perpetrators, help the victim or report incidents to authorities rather than disengaging or joining perpetrators.

B. Educator Level

1. Educators report a high interest in learning about LtP and report moderate levels of confidence in being to use and apply LtP approaches. Generally, they also report moderate positive perceptions on experiencing a supportive environment for LtP in their schools. Results suggest that educators more regularly implement activities in the classroom designed to provide emotional support to children, than activities on pretend play or self-regulation, with some regional differences, for example, those in BGR and Somali regions implementing pretend play and self-regulation activities more regularly. Results also suggest that educators have high levels of energy and mental resilience while working, and of dedication, they find meaning and experience a sense of enthusiasm for their work. Educators from the Somali region, considerably lead the rest in terms of vigor and dedication, while educators from Afar lead in terms of absorption.

2. Through classroom observations, we found that almost half of classrooms fulfilled with high quality of instructional practices. Classrooms in the Somali region and those led by female educators, more regularly comply with the quality standards.

Introduction

Learning through Play (LtP) 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 at 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). Basically, "play is both the way that children express themselves and the means through which they resolve issues" (Hyder, 2005, p.23). 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, Learning through Play (LtP) and Crises in Ethiopia

Evidence on the quality of early grade learning (grades 1-4) in Ethiopia has documented low learning levels for over a decade. An assessment using the Early Grade Reading Assessment (EGRA) in 2010, showed that after two to three years of a literacy intervention, a considerable proportion of Grade 2 and 3 children could not read correctly a single word in their own mother tongue language (Piper, 2010). Similar results recurred in subsequent EGRA assessments conducted in 2014 (USAID-IQPEP, 2014; RTI, 2014), 2016 (USAID-READ M&E, 2016) and 2018 (USAID-READ M&E, 2018). The national learning assessments, which started in 2000, have also consistently shown low learning in early-grade mathematics (MoE, 2000, 2004, 2008, 2013; Abraha, 2015). However, to this date, there is a lack of evidence on student learning outcomes for refugee populations.

From 2016 to 2021, the refugee population at the global level grew from 22.5 to 26.4 million persons (UNHCR, 2017a; UNESCO Institute for Statistics, UNHCR, 2021). Of the global total, over 80% of the refugee population (for example, 84% in 2016 and 86% in 2021) is hosted by developing countries (Ahmed & Saleh, 2022). As a signing party of the 1951 Convention on the Status of Refugees (and its 1967 Protocol) and the 1969 Refugee Convention of the Organization of African Unity (usually referred to as the "OAU Convention"), the Government of Ethiopia (GoE) has long maintained an open-door policy for refugees and asylum seekers. Ethiopia is the second largest hosting country in Africa and the sixth worldwide; as of 2022, hosts over 880,000 refugees and asylum seekers displaced from multiple countries,

including Somalia, South Sudan, Eritrea, and Sudan (UNCHR, 2017c; UNHCR, 2021). Accordingly, as is true in similar contexts, the main support for the refugees in Ethiopia focuses on protection and the provision of shelter, education, health, food security and nutrition, and water and sanitation services. Historically, the GoE has maintained policies requiring refugees to reside in refugee camps, with some exceptions.[1] In 2019, the GoE reviewed its refugee law and adopted an all-encompassing integrative policy (FDRE, 2019), one of the most progressive refugee policies in Africa. Led by Ethiopia's Refugee and Returnee Services (RRS), the law replaces the 2004 Refugee Proclamation (which upheld the key principles of the 1951 Convention) and the 1969 OAU Convention and provides refugees with the right to reside outside of camps and access the same education and financial services as the host community.

The GoE also reviewed their Country's Refugee Response Plan with response strategies for the period ranging from June 2020 to 2025. The strategy document indicated that, though the problems of refugee education vary by refugee location, low participation (i.e., low enrollment and transition rates) and internal efficiency were the two persistent problems of refugee education in the past years. Participation of girls decreases in the transition from lower primary (grades 1-4) to upper primary (grades 5-8), and up to 40% of primary children are overaged for the level they are attending. Primary education Gross Enrolment Rate (GER) among refugee children remained at 50.84% in 2018/19, higher than the national average, which fell at 41% at the time (UNHCR, 2020). In 2020/21 academic year, however, GER for primary education among refugee children declined to 47%, far lower than national GER that stood at 95.1% (UNHCR, 2022b). As of the academic year ending in June 2021, from a total of 171,447 school children, 102,383 (59.7%) were attending primary education in the refugee camp primary schools in the nation (UNHCR, 2022a).

It is unfortunate that low-income countries that cannot even adequately finance social services such as the education of their own citizens are taking the lion's share in hosting refugees. In such situations, the priority of the global community contributes with on provisions of food, shelter and safety more commonly than the provision of education (MacKinnon, 2014; McCarthy, 2017; UNHCR, 2017b). As a result, evidence on learning outcomes of primary education remains scarce to inform scholars, humanitarian agencies, and all concerned bodies (Fransen et al., 2018). However, it is likely that the low quality of education services in refugee camps is indicative of the broader education system in Ethiopia.

The Present Study

Research Aims and Questions

The objective of this study was to conduct a learning assessment that closely studies the status of children's holistic learning outcomes and that of educators' overall awareness and implementation of PlayMatters (PM) strategies in refugee camps and host community primary schools in Ethiopia. Accordingly, the study aimed to answer the following questions:

I. Child Level

1. What are learners' literacy and numeracy skills? How do they vary by displacement, sex, and home language?

2. What is the status of learners' socioemotional skills and wellbeing? How do they vary by displacement, sex, and home language?

3. What are the child characteristics explaining the variation in learners' literacy, numeracy, and SEL skills?

II. Educator Level

1. What is the perception of educators on LtP?

2. What are teachers' instructional practices?

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

Ethical Considerations

This study received ethical clearance from the IRC's institutional review board (IRB) in August 2021 (IRB #: 00009752 FWA #: 00022773, Protocol Number: EDU 1.00.022) and from the Ethiopian Society of Sociologists, Social Workers, and Anthropologists (ESSSWA) in March 2022. All the Principal Investigators (PIs) involved in the study completed a web-based course on protecting human research participants' online training provided by Protecting Human Research Participants (PHRP). 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.

Participants

The study employed both quantitative and qualitative methods that included assessment tests from children and interviews with educators and school principals, and classroom observations. The study employed a two-stage sampling technique. Schools in refugee camps and host communities within the Afar, BGR, Gambella and Somali regions in which the PM project operates were included in the sample. Within those schools, participant children and educators were selected randomly through a systematic sampling technique. Table 1 shows participating primary schools and total sample sizes (of children and educators) by region, gender and community type.

[1] For example, for individuals with serious protection and safety concerns and for Eritrean refugees. In 2010, the GoE issued the Out-of-Camp (OCP) policy, which provided Eritrean refugees the opportunity to live in Addis Ababa and other non-camp locations of their choice, as long as they have financial means for self-support.

Table I. Summary of Sampled Primary Schools, Children and Educators by Sex and Community Type

Region	Location Type	# of Primary Schools	# of Children Assessed			# of Educators Observed			Medium of Instruction
			G	B	Total	F	M	Total	
Somali	Refugee	2	28	30	58	11	2	13	Af Somali
	Host	4	43	47	90	13	5	18	Af Somali
Gambella	Refugee	1	13	10	23	5	0	5	Nuer
	Host	1	12	8	20	1	4	5	Agnuak
BGR	Refugee	1	13	11	24	5	0	5	English
	Host	1	14	11	25	3	1	4	Berta
Afar	Refugee	1	11	14	25	5	0	5	Afar Af
	Host	1	13	12	25	4	1	5	Afar Af
Total		12	147	143	290	47	13	60	

G= Girls; B= Boys; F = Females; M = Males; T = Total

Measures

Child Measures

A number of tools were used to collect quantitative data from children including:

(a) Early Grade Reading Assessment (EGRA) is a tool developed by RTI and adapted to Ethiopia, which intended to assess children´s early literacy skills. The items are not curriculum-based but measure competency in pre-reading. The subtasks included in the EGRA focused on assessing´s children vocabulary, ability to identify letters, oral reading fluency, and reading comprehension. EGRA is a one-on-one assessment in which an enumerator presents a child individually with the subtasks to answer and clearly marks right and wrong answers. The subtasks, Letter Identification (LI) and Oral Reading Fluency (ORF), 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 percent 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. For this, the children attending classes from grade one through five were provided with fifteen same oral vocabulary items. For the oral vocabulary, 15 words were presented to the children, one after the other where they were supposed to show the correct image representing the word. Additionally, for the timed subtasks, the correct answers per minute is computed and analyzed. Thus, the subtasks of letter identification and oral reading were measured by calculating both the percent correct and the mean score of letters identified in a grid or words in a short story that learners correctly read per minute. Reading comprehension was measured by the mean score of the correct answers to questions that are 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.

(b) Early Grade Math Assessment (EGMA): A tool developed by RTI and adapted to Ethiopia to assess children´s numeracy skills. EGMA also uses a mix of time and untimed subtasks that include number identification, missing number, addition level 1 and level 2, subtraction level 1 and level 2, and word problems. Of these subtasks, number identification, addition L1, and Subtraction L1 are timed within 60 seconds. All subtasks are scored using percent correct, and the timed subtasks are also scored using correct per minute. The tool was administered one-on-one by enumerators using the software Tangerine, built for the purpose of administering EGMA.

(c) Battery of Social Emotional Learning Tools:

(i) Empathy Scale - focused on measuring empathy – ability to understand and share another person's feeling and thoughts based on a given situation. It included 12 items with a three-rating scale.

(ii) Emotional Attribution accuracy (ACES) - 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.

(iii) 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 PM by the Consortium, 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). 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) emotional orientation: integrating how does the social information processing children make turns into/ influences children's behaviors into different types of orientation, towards calm, sadness, or anger.
(c) conflict-resolution strategies: referring to the final step of children deciding on a conflict resolution strategy as disengagement, proactive positive problem-solving, or aggression.

(iv) Witnesses´ responses to bullying (Diazgranados et al., 2016). This tool provides 5 scenarios 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.
(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 at a scale of (0 – 3).
(b) Behavioral responses to bullying: Assesses how children responded to situations of bullying in the last two week: Upstand, bystander, perpetrate. Final scores represent the average response for each of these behaviors measured as a scale of (0 – 3).
(c) Exposure to bullying: Score represents the sum of the number of instances in which children report having witnessed situations of bullying in their school in the last two weeks.
(d) 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.

(v) 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 related to children's physical health, feelings, and relationships with family members and friends. All items are scored on a 5-point Likert scale with 0= never, 1 = seldom, 2 = sometimes, 3 = often, 4 = all the time.

Educator Measures

A number of tools were used to collect quantitative data from educators including:

(a) LtP Perceptions Survey a 16-item survey using a scale of whether they agree or disagree with statements regarding LtP.[2]

(i) LtP beliefs: measuring self-perception on educators' capacity to implement LtP approaches in the classroom.

(ii) LtP interest: measuring their interest in learning or receive training for implementing LtP approaches in the classroom.

(iii) LtP supportive environment: measuring how supporting educators consider their schools and school's authorities to be supportive of LtP approaches.

(b) Developmental and Educational Activities Scale (DEAS) measures educators' engagement with developmental and educational activities, measured with a scale of (0 – 4), for four constructs: (i) play (8 items), (ii) pretend play (8 items), (iii) self-regulation (11 items), and (iv) emotional support (8 items).

(c) Utrecht Work Engagement Scale (UWES) measures educators' the affective aspect of teacher well-being at the workplace. The scale consists of 9 items with a five-point Likert-type scale (with scale values ranging from 0 = Never to 4 = Always) that measure the affective aspect of teacher well-being at the workplace through (i) vigor, (ii) dedication, and (iii) absorption (Schaufel et al., 2006).

(d) Teacher Self-Efficacy Scale measures educators' self-efficacy for instructional strategies, classroom management, and student engagement measured with a scale of (0 – 4) in three categories of self-efficacy: (i) Instructional Strategies, (ii) Classroom Management, and (iii) Student Engagement.

(e) The Teacher Instructional Practices and Processes System (TIPPS) measures the quality of classroom instructional practices through direct observation of teachers using 21 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 among other countries (NYU Global Ties for Children, 2021). A previous validation study suggests three conceptually distinct constructs or domains: Facilitating Deeper Learning (FDL), Supporting Student Expression (SSE), and Emotional Support and Behavior Management (ESBM) (Wolf et al., 2018). However, the sample size in our study was too small, limiting our ability to conduct factor analysis to confirm and/or identify these constructs or domains with our data. Thus, results are presented descriptively by items and accompanied by narratives.

(f) Demographics Questionnaire included questions to uncover the support they received within some time interval before the dates of data collection, including the support they received on other teaching methods, and encouragement and professional development.

[2] The tool was developed by IRC's PM team, based on the original conceptualization of LtP for the PM project, which has since been revised and changed.

Process

In January 2022, researchers and project teams participated in a 6-day regional training of trainers (ToT) in Tanzania (specifically Zanzibar city), which covered the rationale and implementation requirements of the study and data collection tools. Afterward, the TOTs conducted training in each site for PM project staff and education experts. The trained experts led the adaptation and translation of the tools, conducted cognitive pretesting; and ensured all the tools were appropriate to the context both culturally, linguistically and cognitively. In March 2022, enumerators were selected based on specified criteria such as language proficiency, experience, and qualification and got trained for five days with a focus on the objectives of the assessment, assessment procedures, tools, and use of Tablets with CommCare and Tangerine software. In addition to discussions and paired/whole group simulated practices, the training also included one-day school-based practices in all areas (Assayita (Afar), Jigjiga, Gambella, and Assossa towns) using the actual data collection tools to check the calibration of enumerators for actual data collection. Finally, the enumerators were deployed to field sites to collect data supervised by the PIs and PM Staff from IRC. They assessed children one by one, and interviewed educators and principals after securing consent from parents and school principals and assent from each child.

LtP is basically children's activity guided by classroom educators, with support from parents or guardians as appropriate. Accordingly, instructional observation of the respective classroom educators of these kids were conducted using Teacher Instructional Practices and Processes System (TIPPS) followed by filling educator survey questionnaire regarding LtP in practice. In this regard, while classroom observation covered a total of 60 (47 female and 13 male), the number of educators who filled the survey questionnaire was 54, reduced because of missing values.

Data collection from the sample areas was in different languages because of variations in the media of instruction in the regions, i.e, Afar, BGR, Gambella and Somali regional states. In Afar and Somali both refugee and host community schools use similar languages of instruction, i.e., Afar Af and Af Somali respectively. In BGR and Gambella, however, languages of instruction vary by type of community. Host community primary schools in Gambella use Agnuak and refugee camp school functions in Neur. In BGR too, while English serves as a medium of instruction in the refugee setting, Berta is the medium of instruction in host community primary schools. Thus, the language of data collection tools, enumerators training, and enumerator selection process followed same pattern.

Analytical Strategy

The data collected for the baseline, both from children and educators, was of quantitative nature. Thus, all data was analyzed using descriptive statistical techniques (descriptive analyses and inferential analyses) to answer the research questions. Because of small sample sizes, statistical inferential testing was applied on disaggregation at the region level, with a focus on community type (refugee or host), sex, and grade levels, but further disaggregation was not possible. Throughout the report, we presented findings in tables, figures and charts indicating the disaggregated when appropriate. We present statistical details, as per ANOVA findings, by this disaggregation whenever the findings are statistically significant.

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,[3] and at and above 50% correct. For each of the learners' socioemotional tools and educators' tools and surveys, we present mean scores according to each tool's score range. 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 (Mean TIPPS score > = 3).

Scope and Limitations of Study

This is a descriptive study focused on documenting learning outcomes and instructional practices of students and teachers in refugee centers and host communities in three conflict-affected regions of Ethiopia, namely Somali, Gambella and BGR, because of the availability of PlayMatters project ECDs. Thus, the findings of this study are relevant for conflict-affected settings similar to that of Ethiopia, where refugees are allowed to learn from the same curriculum of the host community members but in separate schools managed by INGOs and NGOs as opposed to government schools. The study is also limited to the samples of children and educators who teach those children as sampled per the design of the study.

It is important to note that the results presented here are largely descriptive and do not cover relationships among the explained variables (unless stated), nor do they imply causality. Adequacy of the sample size for some of the groups, especially the number of educators, didn't allow to employ analysis of associations and correlations and limited certain data disaggregation. The results of the study are also subject to language variations between the children and those of refugee and host community contexts.

[3] While there are established benchmarks for certain mother tongue languages in Ethiopia, 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 chose the 50% as a target/cut point for ease of interpretation and presentation of results for each language (given language-varying differences), but this is not indicative of a standardized process for establishing benchmarks for either language or EGRA or EGMA subtasks.

Results

Child Data

Literacy and Numeracy (EGRA and EGMA)

a. Fluency and Percent Correct Scores

Initially, children develop their vocabulary from their parents and their immediate community, which includes schools. The use of vocabularies at schools that the children were exposed can enhance learning and better understanding of the world surrounding them. Results of the oral vocabulary subtask, indicate that the children were able to identify about two third of the simple words included in the subtasks' image items. There was slight variation across languages with English and Agnuak learners scoring lowest at 62% percent correct, and Somali learners at the highest end with 82% ***(Table 2).*** About 90% of the children were able to identify more than fifty percent of the oral vocabulary items correctly and they got 76.3% of the items correct.

For letter identification, the aggregate mean percent correct was 31.88%, and English and Somali being the highest scores (42.93% and 39.01%, respectively), and considerably lower mean scores for the rest of the languages (25.6% for Afar, 12.4 for Agnuak, 20.4% for Nuer, and 13.4% for Berta), showing weak capability in letter identification. If a child faces problems with letter identification, they will likely face several challenges in reading words and comprehending what they read. As the students encompassed children from grades one through five, failing to read letters correctly is a critical problem that seeks immediate action.
For the oral reading fluency, results indicate that the maximum average number of words correctly read per minute is 19.9 for Af Somali and the minimum is 0.90 for Berta, with an aggregate value across languages of 13.64 correct words read per minute. If the children cannot read words correctly, it is difficult for them to comprehend the ideas and develop understanding, which is an indication of the difficulty in learning.
Results in terms of percent correct for each language also show weak oral reading capability. Though among the 286 children assessed, about 48% of them were not able to read a single word (zero readers), the percentage decreases as the grade level increases ***(Table 3).*** Only 19.23% were also able to get more than half correct scores on the oral reading; the minimum mean score was 3.1 on grade 1 and maximum was 30.5 in grade 5 which is minimal. In addition, observing zero readers at grades 4 and 5 where they are expected to shift from 'learning to read' to 'reading to learn' is a serious problem.

Reading comprehension was assessed by the number of questions asked, according to the number of words read in the oral reading passage. The children who could not read any word (zero readers) were not asked any reading comprehension questions. From the 291 children who took the test, only 97 (32.9%) were asked items to measure reading comprehension.
For reading comprehension minimum score was 0 and the maximum 5, with Mean = 1.90 and SD = 1.43. Having only 97 from 291 children who could read and got assessed for reading comprehension is a critical challenge for learning to happen. From among those 97 who got assessed for reading comprehension only 28 (9.6%) were able to score above average out of 5 point score. This is an indication that the children are facing a serious challenge to learn through reading.

The mean score of percent correct for EGMA as a total of the eight subtasks was found to be 38.69% which is below the target of 50%. Similar to the EGRA results, children scored higher in the most basic numeric subtasks (number identification and number discrimination) and struggled with more complex and higher-order subtasks (missing number, addition and subtraction and word problems). For number identification, the aggregate mean score was 67.69% correct, with learners assessed in Berta with the lowest mean score (31%) and the rest of the learners scoring higher than 65% percent correct ***(Table 2).*** A similar picture emerges from the number discrimination subtask, though with lower scores, with an aggregate score of 63.44%, learners assessed in Berta with the lowest percent correct 40.5%, and the rest of the learners scoring higher than 58% correct. For both number identification and number discrimination subtasks, learners assessed in Somali and Agnuak scored higher than the rest of the learners. For missing number, the aggregate mean score was 34.51% correct, with slight differences in learners assessed in each language, ranging from 23.9% for Berta as the lowest score, and 36.7% for Somali as the highest score.

Each Addition and Subtraction subtask included Level 1 (L1) and Level 2 (L2) items, with Level 2 involving more complex operations. As expected, children scored higher for both Addition and Subtraction L1 subtasks than in L2 in all languages ***(Table 2),*** except for learners assessed in English who scored slightly higher in Subtraction Level 2 than in Subtraction Level 1 (44.5% and 33.1% correct respectively). The aggregate mean scores, however, depict children struggling with most of these higher-end numeracy subtasks, with Addition L1 being the highest aggregate score with 40.87% correct (with learners assessed in Berta and Somali below the aggregate mean with 36.7% and 38.8% correct respectively), and Subtraction L2 with the lowers aggregate score with 20.94% correct (with learners assessed in English scoring significantly higher with 44.5% correct and the rest of learners scoring closer to the aggregate score). Fluency scores for Addition L1 and Subtraction L1, report similar levels of low achievement with 8.56 and 6.07 correct answers per minute, respectively, and similar to the aggregate results across all languages, with learners assessed in English being the only group scoring higher than the aggregate (11.3% and 7.6% respectively). The aggregate mean score for the Word Problems subtask was 34.89% correct, along the lines of Addition L1 and Subtraction L1, with similar results across all languages, all with mean scores higher than 31% correct.

Results disaggregated with respect to sex, community type, and class are presented below ***(Table 3).***
Analyses found stark and statistically significant differences between girls and boys and for all and all EGMA subtasks, boys outperformed girls. In terms of differences between community types, we found statistically significant differences only for number identification, number discrimination, and Addition

Table 2. Mean Score of Percent Correct and Correct Per Minute of EGRA and EGMA Results

		Aggregate			Afar	Gambella		BGR		Somali
		Mean	S.D	Mini	Max	Afar	Aguank	Nuer	Berta	English	Somali
EGRA	Oral vocabulary % correct	76.36	18.07	20.0	100	78.5	62.7	65.2	71.7	62.8	82.6
	Letter ident. % correct	31.88	31.09	.0	100	25.6	12.4	20.4	13.4	42.9	39.0
	Correct Letter per minute	32.72	32.49	.0	159	25.6	12.4	20.6	17.9	42.9	39.9
	Oral reading % correct	20.95	31.92	.0	100	12.1	5.1	16.8	1.3	17.0	30.6
	Oral Reading Fluency	13.64	20.99	.0	98.8	6.4	4.2	10.9	0.9	12.7	19.9
	Reading comp. % correct	37.94	28.64	.0	100	35.4	80.0	5.0	40.0	36.0	42.2
EGMA	Number ident. % correct	67.69	35.30	.0	100	57.3	70.2	66.4	31.0	67.3	75.9
	Correct number per minute	22.71	17.64	.0	85.7	17.0	16.9	19.5	9.3	18.1	28.6
	Number discr. % correct	63.44	29.73	.0	100	59.8	66.4	65.6	40.5	58.3	68.0
	Missing number % correct	34.51	21.52	.0	100	30.8	28.8	30.6	23.9	45.2	36.7
	Addition L1 % correct	40.87	25.84	.0	100	42.3	40.6	41.9	36.7	51.4	38.8
	Add. L1 Correct per Minute	8.56	5.65	.0	31.0	8.6	8.1	8.5	8.7	11.3	8.1
	Addition L2 % correct	26.57	32.35	.0	100	28.8	19.2	28.9	13.3	37.1	26.4
	Subtraction L1 % correct	25.59	24.53	.0	100	23.9	18.4	29.2	28.0	33.1	25.1
	Subt. L1 Correct per Minute	6.07	7.77	.0	60.0	4.9	3.7	6.2	6.9	7.6	6.5
	Subtraction L2 % correct	20.95	29.85	.0	100	25.1	12.5	17.8	21.5	44.5	17.6
	Word problems % correct	34.89	24.81	.0	100	31.9	33.3	39.6	35.0	34.6	35.6

L2, with refugee community outperforming the host community, though due to limited sample sizes it is not possible to disaggregate by region or country of origin.

Overall and as expected, scores increase as the class progresses for all subtasks in both EGRA and EGMA, with statistically significant differences ***(Table 3).*** For the more elemental subtasks, i.e., oral vocabulary, number identification, and number discrimination, learners in class 1 score at least 30% correct, while expectedly, learners in lower grades score significantly lower than those in higher grades for the most complex subtasks.

Regardless, the performance across grades depicts struggling learners with, for example, low mean scores for both letter identification and oral reading (both in terms of fluency and percent correct) in classes 1, 2, and 3. As with the full sample (all grades), children scored higher for both Addition and Subtraction L1 subtasks than in L2 across classes, though the performance in Subtraction L1 is considerably lower across all classes (ranging from 4.2% for class 1 to 7.4% percent correct for class 5). However, differences across classes needs to be interpreted with caution given the reduced sample sizes.

Table 3. EGRA and EGMA Disaggregated by Sex, Community Type, and Class

	EGRA/EGMA Subtasks	Sex		Community Type		Class				
		Girl	Boy	Refugee	Host	1	2	3	4	5
EGRA	Oral vocabulary % correct	75.1	77.6	77.3	75.5	70.3	72.6	80.7	78.8	80.3**
	Letter ident. % correct	25.6	38.3***	33.8	30.0	9.6	18.9	35.4	42.7	58.8***
	Correct Letter per minute	26.7	38.**	34.6	30.9	10.0	19.2	35.4	44.8	59.9***
	Oral reading % correct	14.1	28***	21.9	19.6	3.6	5.3	22.1	31.7	47.3***
	Oral Reading Fluency	9.7	17.7***	14.5	12.6	3.1	4.8	13.3	19.9	30.5***
	Reading comp. % correct	37.1	38.4	37.0	38.4	8.0	13.3	33.3	35.3	51.8***
EGMA	Number ident. % correct	57.9	77.6***	72.2	63.9*	38.0	63.3	71.9	83.9	85.5***
	Correct number per minute	17.4	28.1***	24.7	21.0	9.9	16.9	23.7	30.2	35.2***
	Number Disc. % correct	57.6	69.4***	68.0	59.6*	46.5	50.3	73.3	73.3	76.7***
	Missing number % correct	32.2	37.8*	37.5	32.6	22.9	30.8	32.9	46.7	41.6***
	Addition L1 % correct	35.6	48.2***	43.5	38.4	21.0	29.3	45.0	54.6	56.0***
	Add. L1 Correct per Minute	7.7	9.4**	9.0	8.1	4.6	6.0	9.2	11.9	11.3***
	Addition L2 % correct	20.9	32.4**	32.1	21.8*	12.9	18.7	27.3	34.0	36.3***
	Subtraction L1 % correct	22.4	28.9*	26.5	24.4	4.2	4.9	6.2	7.7	7.4***
	Subtraction L1 Correct per Minute	5.6	6.6	6.7	5.5	4.4	16.9	32.5	31.9	49.0***
	Subtraction L2 % correct	16.1	25.8**	23.6	18.3	5.7	21.3	23.0	19.3	36.6***
	Word problems % correct	30.8	39**	34.7	35.1	24.7	29.8	37.3	38.9	45.7***

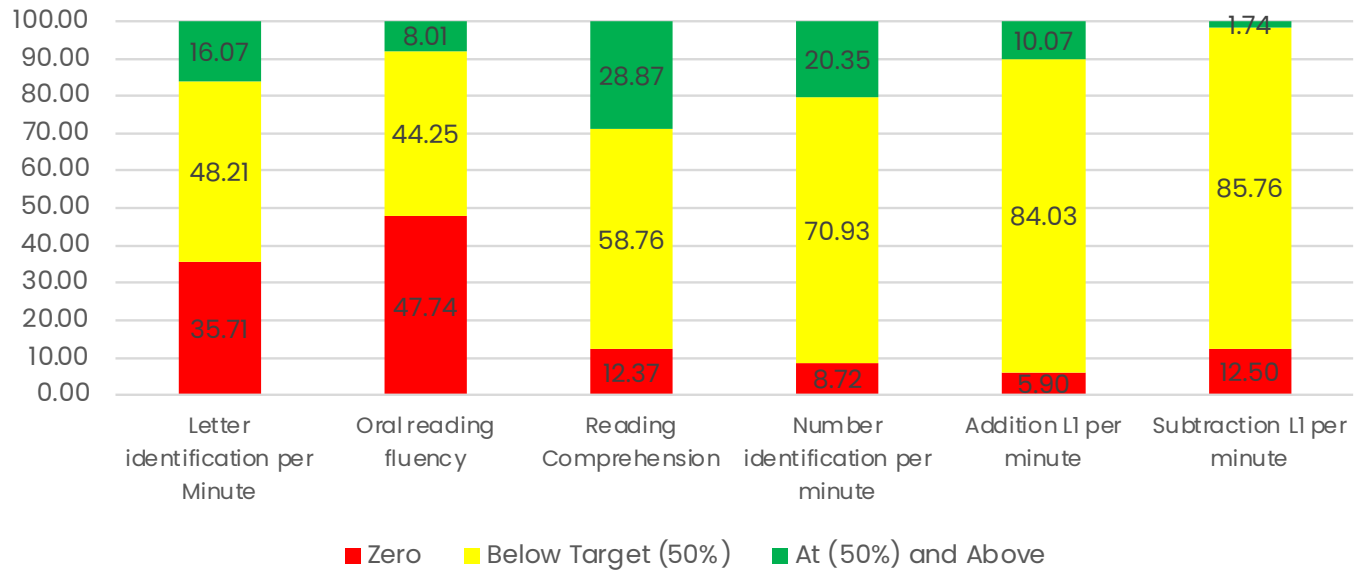
Note: Statistical significance marked as *p<.05 **p<.01 *** p<.001.

b. Performance categories
For EGRA and EGMA timed subtasks and reading comprehension, we 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 the target of 50% correct, and at and above 50% correct. Children generally showed better results in EGMA than in EGRA, with a higher number of children at or above target for number identification and Addition L1 (Figure 1). Fig 1 further indicates that children are facing the most difficulties in oral reading fluency and letter identification, and while there is a high percentage of children at or above target for reading comprehension, this only includes the children who were able to read the oral reading passage in the first place, which is only a small percentage of the full sample.

For letter identification fluency, 35.7% were zero scorers who were unable to identify a single letter, with regional variations, as Berta was found to have a significant proportion of learners who are zero scorers. In terms of oral reading fluency, results indicate that nearly half of the student population is a zero reader (Fig. 1). Children in the host community were found to have a higher number of zero readers. In terms of reading more than half of the words provided, the proportion of girls (12.33%) was found to be lesser than boys (26.95%), but the proportion was more or less the same with community type (in Annex).

Though it seems that the children with zero scores is less in numeracy as compared to literacy, a significant proportion of them are still below the target (Fig. 1). For EGMA, the percent of zero scorers increases as the task increases in its cognitive demand from number identification (less difficult) to subtraction (comparatively demanding task), with the percentage of students who got zero scores being 8.72% and 12.5% respectively. The correct number identification per minute was 28.6 (above the 20 numbers provided) for Af Somali, while it is below 20 for all other languages (See Table 2). These show problems associated with foundational numeracy skills that must be addressed at earlier grade levels to help the children develop their competence in later grades.

Figure 1. EGRA and EGMA Timed Subtasks - Percentage of Children by Performance Categories

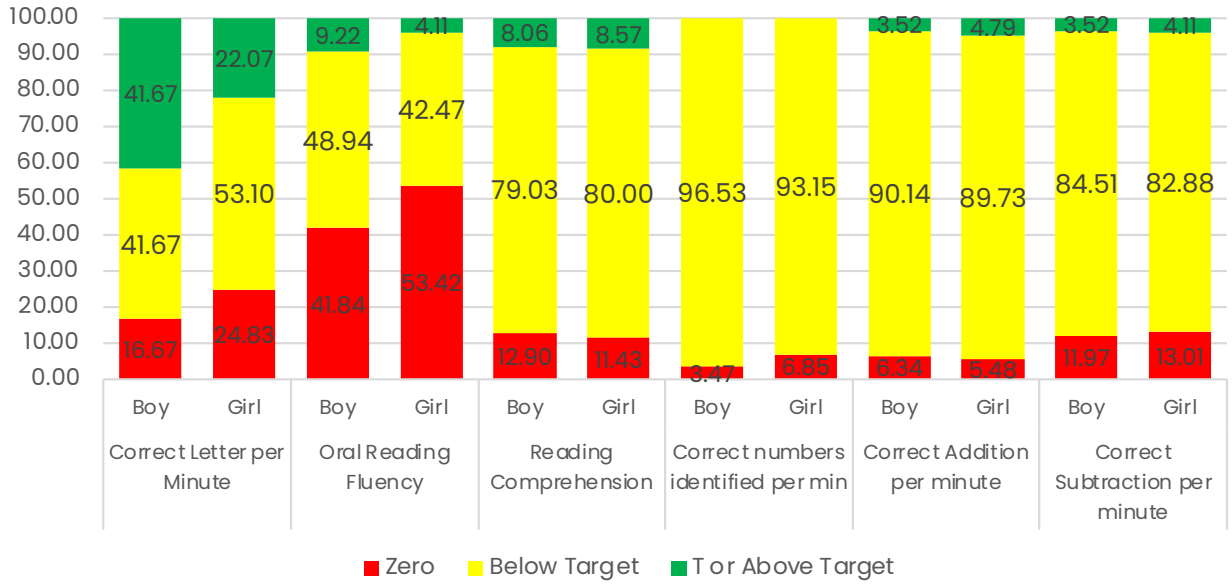


The results by performancezcategories are further uncovered disaggregated with respect to sex, community type, and class, each presented in Fig.2, Fig.3, and Fig. 4, respectively.

The results indicate that there are more girls with zero scores than boys for literacy. However, for numeracy the results are more or less the same (Fig. 2), which is what is driving the statistically significant differences observed between boys and girls in the majority of the subtasks in terms of mean scores (See Table 3 and previous section).
While the percentages of children in each category for reading comprehension are similar between girls and boys, this is again just depicting the reduced sample of children who attempted the reading comprehension subtask.

[4] For untimed EGRA subtasks (not shown in the Figure), the percentage of children in the zero score category was largest for the addition and subtraction L2, followed by Word problem and subtraction L1.

Figure 2. Percentage of children's EGRA and EGMA results correct per minute and fluency by Category of Sex



Differences among host and refugee communities are varied across EGRA and EGMA fluency subtasks (Fig. 4). For EGRA's correct letters identified and read (ORF) per minute, the host community has a larger proportion of children in the zero scores categories and the refugee community a higher proportion of children in the at or above target category. For reading comprehension, the refugee community a higher proportion of children in the at or above target category but also a larger proportion of children in the zero categories. For EGMA subtasks overall, there are higher percentages of children in the below target and lower percentages of students in the zero scores compared to EGRA, but also a limited number of children at or above target in comparison with EGRA subtasks. The refugee community performed slightly better than the host community, as there is a higher percentage of zero scores in the host community than in the refugee in the correct numbers per minute (7.5% host and 2.3% refugee), and with a higher percentage of refugee community at or above target for Subtraction L1 per minute (5.47%) compared with the host community (1.89%).

Differences across grades for EGMA are as expected with a decreasing percentage of children in the zero category and an increasing percentage of children in the at or above target category, as grades increase for letter identification and oral reading fluency (Fig. 5). The pattern is not as clear for the reading comprehension subtask, most likely due to the limited number of students attempting the subtask in each class category. For the three EGMA subtasks, across grades, most students are in the below target category, and only a limited percentage in the at or above target category; however, compared with EGRA, there is also a lower percentage of students in the zero category across grades with subtraction being the subtask with a higher percentage of zero scores, as it is the more complex subtask. However, differences among classes need to be interpreted with caution due to varying and small sample sizes per class.

Figure 3. Percentage of Children's EGRA and EGMA Results Correct Per Minute and Fluency by Category of Community Type

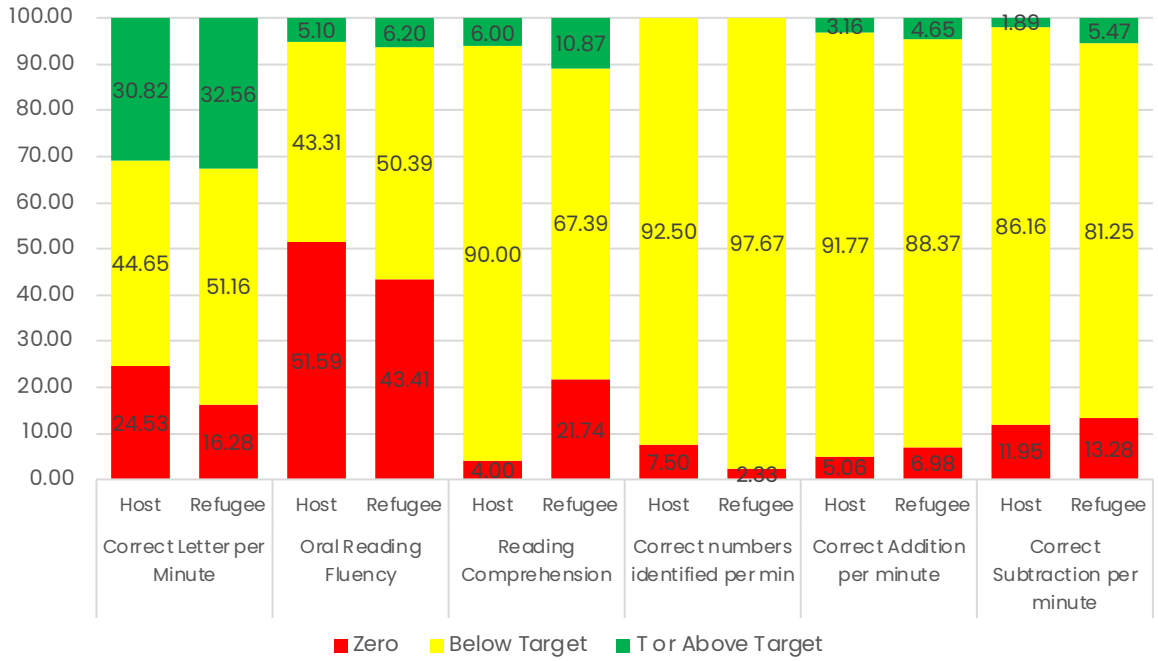
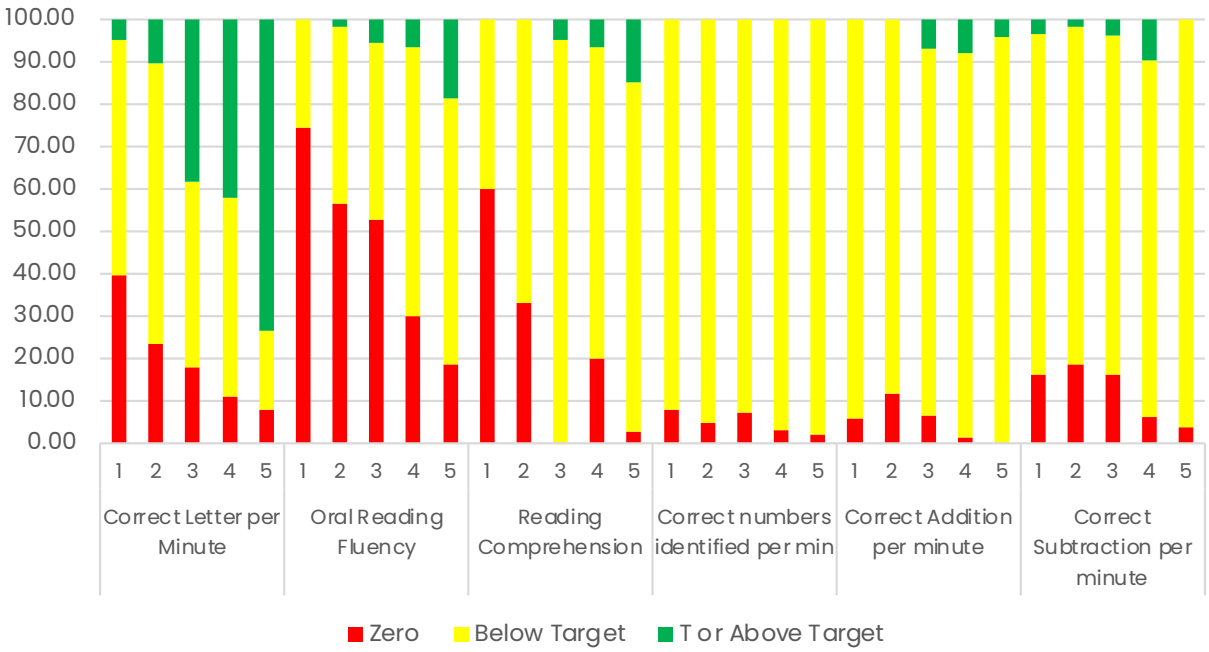


Figure 4. Summary of EGRA and EGMA Results Correct Per Minute and Fluency by Category of Class



Socio-Emotional Learning and Wellbeing

a. Empathy and Emotional Responses

Both the Empathy score and the ACES score, measuring emotional attribution accuracy, were relatively high (0.72 and 0.63 respectively, both from a 0 to 1 scale), suggesting that children can adequately recognize and encode feelings from common situations, and have feelings of empathy for situation others experience. The results in Table 4 also indicate no statistically significant difference between sex and community type, except for learners in class 3 reporting higher ACES. However, differences among classes need to be interpreted with caution due to varying and small sample sizes per class.

Table 4. Social-Emotional Learning: Empathy, ACES, and SERAIS Disaggregated by Sex, Community Type, and Class

	Aggregate		Sex		Community Type		Class				
SEL (Scales)	**M**	**SD**	**Girl**	**Boy**	**Refugee**	**Host**	**1**	**2**	**3**	**4**	**5**
Empathy score (0-1)	0.72	0.27	0.73	0.73	0.72	0.74	0.69	0.71	0.73	0.74	0.79
ACES score (0-1)	0.63	0.28	0.61	0.67	0.61	0.67	0.54	0.56	0.75	0.69	0.67***
SERAIS: Hostile attribution bias score (0-1)	0.48	0.28	0.47	0.49	0.46	0.50	0.48	0.47	0.51	0.45	0.49
SERAIS: Emotional orientation-calmness (0-1)	0.49	0.34	0.50	0.49	0.51	0.48	0.47	0.44	0.48	0.51	0.60
SERAIS: Emotional orientation-sadness (0-1)	0.26	0.28	0.26	0.27	0.26	0.27	0.26	0.27	0.33	0.20	0.26*
SERAIS: Emotional orientation-angry (0-1)	0.23	0.27	0.23	0.24	0.21	0.25	0.21	0.29	0.27	0.27	0.11*
SERAIS: Conflict resolution -disengagement (0-1)	0.21	0.27	0.22	0.20	0.24	0.19	0.27	0.28	0.17	0.15	0.18*
SERAIS: Conflict resolution -problem solving (0-1)	0.66	0.34	0.65	0.68	0.63	0.69	0.61	0.54	0.74	0.69	0.75*
SERAIS: Conflict resolution -aggression (0-1)	0.13	0.21	0.13	0.12	0.14	0.12	0.12	0.17	0.09	0.16	0.06

Note: Statistical significance marked as *p<.05 **p<.01 *** p<.001.

For the first SERAIS sub-domain, we found that about half of children interpreted certain social situations as intentional and hostile provocations, with similar results by sex, community types, and classes (Table 4). For the second SERAIS sub-domain of emotional orientation, we find that children tend to report a moderate degree of calm feelings (0.49), and very low levels of sadness (0.26) or anger (0.23), with consistent results across sex, community type, and classes. For the third SERAIS sub-domain, children majoritarily endorsed problem-solving strategies (0.66), over disengagement (0.21), or aggression (0.13), with very similar results by sex, community types, and classes (Table 4).

b. Wellbeing, Bullying, and Victimization

For wellbeing, the Kiddy-KINDL measure uses wellbeing across different domains: physical, emotional, self-esteem, family, friends, school, and a total aggregate score. The mean score of the Kiddy-KINDL tool, which measures wellbeing, is 1.70 with a standard deviation of .57 (Table 5). The result shows that on average, children rated their wellbeing unfavorably, as the distance from the possible limits of maximum (4) and minimum (0) achievable values can give a first indication of a respondent's self-assessment.

Table 5. Cross-Tabulation of Mean Scores and Standard Deviations of Wellbeing Dimensions by Region

	Physical Wellbeing**	Emotional Wellbeing	Self-Esteem	Family***	Friends	School**	Total
Afar	1.22 (.86)	1.03 (.74)	1.63 (1.44)	1.80 (.66)	2.40 (1.08)	2.45 (.87)	1.74(.64)
Agnuak	.84 (.59)	.99 (.41)	1.65 (.93)	1.73 (.59)	2.11 (.66)	1.58 (.96)	1.49(.47)
Berta	1.5 (1.01)	1.41 (1.11)	2.11 (1.08)	2.94 (.66)	2.71 (.81)	2.76 (1.03)	1.98(.78)
English	1.37 (.66)	1.41 (.67)	1.84 (1.10)	1.69 (.73)	2.18 (.75)	2.00 (.91)	1.72(.46)
Nuer	1.22 (.76)	1.5 (.67)	1.54 (.79)	1.80 (.75)	2.02 (.68)	1.93 (.77)	1.69(.56)
Somali	.91 (.70)	1.22 (.84)	1.79 (.99)	1.98 (.77)	2.14 (.89)	2.13 (.91)	1.65(.27)
Total	1.05(.77)	1.21(.79)	1.75(1.08)	1.93(.78)	2.20(.89)	2.13(.93)	1.70(.57)

Note: Statistical significance marked as *p<.05 **p<.01 *** p<.001.

Comparisons indicated significant differences between sex and community types, indicate that boys report higher wellbeing than girls in the Friends dimension, while children from host communities report higher wellbeing than those in the refugee community (Table 6) with statistically significant differences.

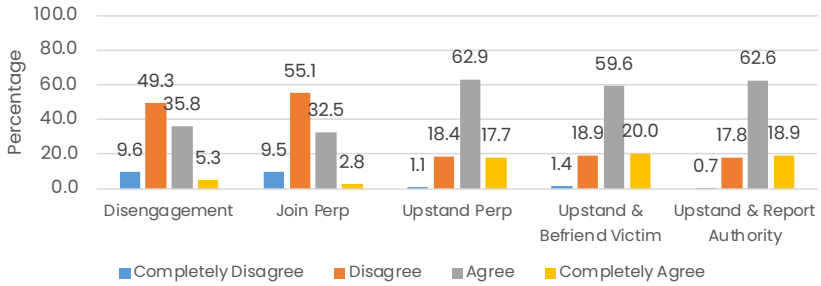
Table 6. Cross-Tabulation of Mean Scores and Standard Deviations of Wellbeing Dimensions by Region

	Category	Physical	Emotional	Self-Esteem	Family	Friends	School
Sex	Boy	1.07 (.74)	1.29 (.84)	1.78 (1.07)	1.95 (.74)	2.33 (.90)*	2.15 (.91)
	Girl	1.04 (.78)	1.13 (.72)	1.74 (1.09)	1.93 (.80)	2.10 (.85)	2.14 (.96)
Community	Host	1.06 (.76)	1.16 (.78)	1.83 (1.10)	2.04 (.79)**	2.28 (.87)	2.15 (.97)
Type	Refugee	1.05 (.76)	1.28 (.80)	1.66 (1.05)	1.82 (.72)	2.13 (.90)	2.14 (.89)
Class	1	1.01 (.83)	1.13 (.75)*	1.62 (1.17)	1.74 (.97)**	2.19 (.85)*	1.85 (.97)
	2	1.03 (.70)	1.26 (.80)	1.89 (.92)	1.92 (.49)	2.05 (.84)	2.17 (.90)
	3	1.00 (.71)	.94 (.68)	1.72 (1.17)	2.16 (.80)	2.48 (.82)	2.24 (.83)
	4	1.26 (.87)	1.37 (.90)	1.84 (1.07)	2.03 (.69)	2.22 (.95)	2.31 (.98)
	5	.93 (.60)	1.38 (.70)	1.70 (1.05)	1.89 (.71)	2.16 (.89)	2.17 (.93)

Note: Statistical significance marked as *p<.05 **p<.01 *** p<.001.

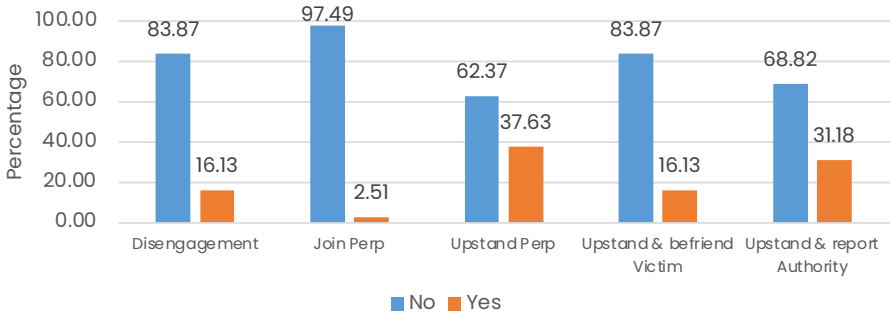
Results in terms of the percentage of learners' attitudes towards bullying and responses to it (disengagement, joining perpetrators, upstanding perpetrators, upstanding and befriending victims, and upstanding by reporting to authorities) are presented in Figures 5 and 6. Results indicate that around 60% of learners tend to have attitudes of response to bullying by upstanding towards the bullying (as well as befriend and report to authority), and approximately half disagree on responding by disengagement or joining perpetrators (Figure 5).

Figure 5. Attitudes towards Bullying - Percentage of Children on Scale of Agreement to Bullying Scenarios



Cognizant of the level of bulling attitude and behavior, an attempt was made to check whether, in the past two weeks, they have observed situations like the ones presented in the scenarios (exclusion, verbal aggression or physical aggression) with yes or no questions. The percentage of learners described by whether they observed bullying behavior suggests that children have considerably more exposure to upstand actions (37.63%), including upstand and report to authority (31.18%) and upstand and befriend victim (16.13%). On the other end, 16.13% reported exposure to disengagement and only 2.51% exposure to experiences of joining perpetrators (Figure 6).

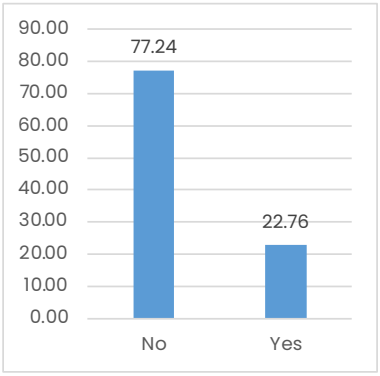
Figure 6. Exposure to Bullying - Percentage of Children Observed Bullying Behavior in Past 2 weeks



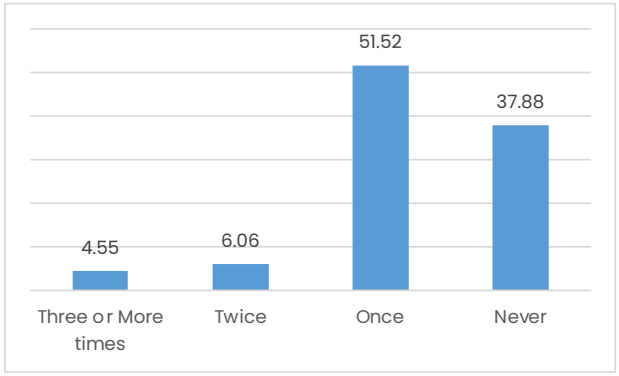
Children were also asked if and how often they had personally been victims of bullying in the past two weeks (never, once, twice, and three or more times). The results of these are presented in Figure 7 below. The chart indicates that 77.1% did not observe situations of bullying in the past two weeks, and 28.3% were victims of bullying, which demands to be addressed. Likewise, about 71.7% of the children had never been victims of bullying, which is fine, but that 23.1% reported having been victims.

Figure 7. Percent of Exposure to Bullying Behavior and Extent of Victimization

% of children who observed bullying behavior (n = 290).



% of children who experienced victimization from among those who observed bullying behavior. i.e, from the 22.76%



Child Sample Demographics

Beyond sex and community type which are discussed above other child characteristics such as age, time they spent during the day, extent of repetition, availability of resources at home, and involvement in household chores were explored. The mean age is 10.49 with SD = 2.41; and minimum of 5 and maximum of 17. Significant number, 260 (88.1%), have repeated once. The number of people living in home ranges from 1 to 23 with an average of seven people. Number of people to sleep at home ranges from 1 through 5 with an average size of three.

The proportion of those who have reading resources at home is 192 (65.1%). For reading resources (textbooks) at home it was noted that 160 (54.2%) have textbooks but 32 (10.8%) do not. 163 (55.3%) have reported that they have reading resources (religious books) at home while 29 (9.8%) do not. Likewise, 187 (63.4%) do not have newspaper as reading resources, only 5 (1.7%) have said they have newspaper as reading resources. 185 (62.7%) do not have children's stories as reading resources. Those who reported to have children's stories are only 7 (2.4%). 180 (61.0%) have reported to have household chores; 64 (21.7%) additional work; (13.6%) domestic work; and (8.8%) deliveries. For the time spent doing some tasks the following indicates some results.

As literacy develops through the exposure to different resources, the children were asked to tell their experiences the time, they spent doing the tasks mentioned in the table above. Significant proportion, above 60% of the children never spent time listening to music, community groups, watching TV, chatting online, or texting. These could have contributed to developing literacy, but for various reasons, the children are not spending time with these tasks, which could be related to the children's literacy capability and low levels of learning. In terms of disability status, 14.2% (n=42) of the children were also found to have at least 1 domain with difficulty (some, a lot, or can't do at all), which could also be a contributing factor to the literacy and numeracy skills.

Table 7. Percentage of Children who spent Time in Various Engagements

	Sport	Reading	Playing	Music, Art, Dance	Community Group	Religious Activity	Watch TV	Hang with Friend	Chat, Online, Texting
None	27.6	14.8	17.2	68.3	60.0	10.0	60.7	17.6	65.9
< hour	23.1	23.1	24.1	12.1	17.2	11.4	10.0	32.4	13.8
1 hour	22.4	29.0	26.2	11.4	16.2	22.1	12.1	23.4	11.0
2 hours	19.3	22.8	21.0	5.5	3.4	24.1	8.6	16.2	5.9
> 2 hours	7.6	10.3	11.4	2.8	3.1	32.4	8.6	10.3	3.4
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

As literacy develops through the exposure to different resources, the children were asked to tell their experiences the time, they spent doing the tasks mentioned in the table above. Significant proportion, above 60% of the children never spent time listening to music, community groups, watching TV, chatting online, or texting. These could have contributed to developing literacy, but for various reasons, the children are not spending time with these tasks, which could be related to the children's literacy capability and low levels of learning. In terms of disability status, 14.2% (n=42) of the children were also found to have at least 1 domain with difficulty (some, a lot, or can't do at all), which could also be a contributing factor to the literacy and numeracy skills.

Educator Data

Educators' data collection included tools inquiring about their perceptions of learning through play; developmental and educational activities that include play, pretend play, self-regulation, & emotional support; their work engagement via vigor, dedication & absorption; and their self- efficacy for instructional strategies, for classroom management, & for student engagement. The results are presented below.

Perceptions of LtP

It is believed that play-based learning is important for educators to shift from the accustomed teacher-centered instructional approach to student-centered learning, and this shift contributes towards enhancing the learning abilities of children and the development of socioemotional skills. Onditi et al. (2018), in their study on influence of teachers' perceptions on Play-Based Activities in pre-school curriculum implementation in Homabay County stated that "Many pre-school teachers held the perception that play-based activities are important because they enable children to develop social, emotional, physical and motor skills necessary in their life. However, although many of the preschool teachers had a general perception that they were adequately trained on the use of play-based activities in teaching and learning, some of them felt they were not adequately trained on the integration of play in teaching/learning. Equally, it came out that some of the teachers held a general belief that; play-based activities waste valuable time for academic work and that it is not very easy to incorporate it in the teaching and learning process." (p.71)

Educators' perceptions of LtP are provided in Table 8, and suggest that overall, interest of educators on learning about LtP is high, and with minor differences across regions, with Afar region scoring highest and Gambella the lowest. Educators' beliefs on their capacity to implement LtP approaches, is also relatively high, with educators in BGR scoring higher than educators in other regions. Perceptions on supportive environments for LtP is moderate, and similar and consistent across regions and community types. Female educators report higher scores than their men counterparts in all three categories of LtP beliefs, interest in learning LtP approaches, and perceptions on a supportive environment for LtP.

Table 8. Perceptions of Educators on LtP

	Aggregate	Region				Sex		Community Type	
		Afar	BGR	Gambella	Somali	F	M	Refugee	Host
LtP Beliefs	2.79 (0.98)	2.79 (0.98)	2.94 (0.85)	2.42 (0.95)	2.9 (1.06)	3.02 (0.86)	2.68 (1.05)	2.72 (0.97)	2.87 (1)
LtP Learning	3.24 (0.7)	3.64 (0.36)	3.38 (0.51)	3.12 (0.28)	3.14 (0.88)	3.41 (0.39)	3.17 (0.81)	3.26 (0.65)	3.22 (0.76)
LtP Supportive Environment	2.56 (1.07)	2.28 (0.93)	2.69 (0.87)	2.04 (1.19)	2.77 (1.07)	2.75 (1.05)	2.53 (1.03)	2.61 (1.07)	2.50 (1.07)

Developmental and Educational Activities Scale (DEAS)

DEAS reports on educators' engagement with developmental and educational activities and results (Table 9) suggest that educators more regularly implement activities in the classroom designed to provide emotional support to children, than activities on pretend play or self-regulation. Educators in the Gambella region, however, report using emotional support activities considerably less than their peers in other regions. Overall, educators reported using pretend play activities with less regularity than other types of activities. However, differences emerge by regions, as results suggest that educators in the BGR and Somali regions implement pretend play and self-regulation activities more regularly than other regions.

Table 9. Mean Scores of Educators' Developmental and Educational Activities Scale (DEAS)

	Aggregate	Region				Sex		Community Type	
		Afar	BGR	Gambella	Somali	F	M	Refugee	Host
DEAS Aggregate	1.99 (.70)	1.81 (.53)	2.16 (.55)	1.64 (.68)	2.12 (.76)	2.15 (.70)	1.99 (.71)	2.01 (.54)	1.98 (.79)
Pretend Play	1.71 (.90)	1.63 (.86)	2.02 (.78)	1.49 (.86)	1.73 (.98)	1.89 (1.13)	1.72 (.86)	1.80 (.64)	1.65 (1.06)
Self-Regulation	2.01 (.77)	1.73 (.45)	2.20 (.65)	1.74 (.62)	2.13 (.9)	2.09 (.84)	2.02 (.77)	1.96 (.57)	2.04 (.89)
Emotional Support	2.46 (.84)	2.56 (.70)	2.55 (.85)	1.83 (.81)	2.61 (.83)	2.68 (.92)	2.44 (.83)	2.34 (.75)	2.54 (.90)

Utrecht Work Engagement Scale (UWES)

Educators were asked to describe the extent of their engagement in development and educational activities to measure their professional wellbeing. Results (Table 10) suggest that educators have high levels of vigor, i.e., a high level of energy and mental resilience while working, and of dedication, i.e., being involved in one's work, finding meaning, being challenged, and experiencing a sense of enthusiasm, and absorption, i.e., being fully concentrated and engrossed in one's work. Educators from the Somali region, considerably lead the rest in terms of vigor and dedication, while educators from Afar lead in terms of absorption. In terms of differences by community, educators from host communities report moderately higher levels of vigor and dedication than those from the refugee communities, while refugee and host educators report similar levels of absorption. Results also suggest limited differences among female and male educators.

Table 10. Mean Scores of Educators' Utrecht Work Engagement Scale (UWES)

	Aggregate	Region				Sex		Community Type	
		Afar	BGR	Gambella	Somali	F	M	Refugee	Host
Wellbeing Aggregate	2.86 (.77)	2.96 (.78)	2.68 (1.00)	2.36 (.51)	3.05 (.71)	2.84 (.79)	2.89 (.78)	2.78 (.80)	2.92 (.75)
Vigor	2.94 (.87)	2.85 (.88)	2.91 (1.12)	2.37 (.63)	3.15 (.80)	2.88 (.89)	2.98 (.90)	2.83 (.82)	3.01 (.91)
Dedication	2.93 (.98)	3.04 (1.03)	2.46 (1.19)	2.41 (.91)	3.20 (.84)	2.91 (1.16)	2.96 (.95)	2.76 (1.03)	3.05 (.93)
Absorption	2.72 (.88)	3.00 (.69)	2.67 (1.11)	2.30 (.65)	2.79 (.91)	2.73 (.89)	2.74 (.89)	2.76 (.92)	2.70 (.86)

Self-Efficacy

Results for the self-efficacy tool (Table 11), indicate high percentage of educators have self-efficacy believes for instructional strategy, classroom management and student engagement, suggesting educators feel confidence in implementing adequate instructional strategies, use variety of assessment strategies, provide alternative explanation when students are in need, can craft good questions for their students, and implement classroom management to control disruptive behavior in the classroom. Educators from Somali region, particularly report high perception on their self-efficacy in implementing classroom management strategies.

Table 11. Mean Scores of Educators' Self-Efficacy

	Aggregate	Region				Sex		Community Type	
		Afar	BGR	Gambella	Somali	F	M	Refugee	Host
Self-Efficacy Aggregate	2.42 (.73)	2.25 (.71)	2.77 (1.02)	1.91 (.52)	2.53 (.66)	2.52 (.62)	2.39 (.78)	2.50 (.76)	2.36 (.74)
Instructional Strategies	2.29 (.81)	2.03 (.61)	2.47 (1.24)	1.84 (.48)	2.41 (.79)	2.52 (.73)	2.19 (.84)	2.23 (.75)	2.29 (.89)
Classroom Management	2.52 (.87)	2.61 (.82)	2.88 (1.0)	1.97 (.78)	2.56 (.86)	2.68 (.72)	2.48 (.93)	2.70 (.86)	2.39 (.89)
Student Engagement	2.45 (.86)	2.11 (.90)	2.97 (.99)	1.91 (.52)	2.63* (.80)	2.36 (.82)	2.51 (.89)	2.58 (.92)	2.40 (.83)

Note: Statistical significance marked as *p<.05 **p<.01 *** p<.001.

Classroom Quality (TIPPS)

The overall mean TIPPS score for the primary classrooms observed by a scale with a maximum possible score of 4 and minimum score of 1 is 2.81 with a standard deviation of 0.55. The score shows that the quality of the learning environment in the classrooms observed tends to get close to a score of 3. The minimum mean score for each of the items was 2.23 and the maximum was 3.35, suggesting there are no classes with poor learning environment.

As TIPPS is for quality classroom, the 60 classrooms observed were examined for quality instructional practices and systems, and the results by taking the quality indicator of quality classroom (Mean TIPPS score > = 3) shows that 43.3% of the classes observed meet the criteria, while the remaining 56.7% need support to improve the classroom instructional practices and processes systems (Figure 8).

Figure 8. Percent of Classrooms Fulfilling Quality Standard (Mean TIPPS >=3.0)

Further analysis was made by disaggregating data into region, sex of educators, and community type, and class level (Table 10). Results suggest that, on average, only the classrooms in the Somali region fulfil standards of quality compared with other regions. While classrooms led by female instructors comply with standards of quality more than those led by male instructors, the difference was not statistically significant. Differences by type of community, was also not significant. Results should be interpreted with caution due to the small sample sizes. Further, since teachers are assigned to teach multi-grade classes critical investigation of grade level differences may not be feasible.

Table 12. Descriptive Statistics of Classroom Quality by Region, Sex, Community Type, Classroom Type and Class

Category		N	Mean	SD	Min	Max
Region	Afar	10	2.74	.48	2	4
	Benshangul-Gumuz	9	2.45	.54	1	3
	Gambella	10	2.57	.49	2	3
	Somali	31	3.02	.53	2	4
Sex of educators	F	11	3.00	.48	2	4
	M	48	2.76	.57	1	4
Community Type	Host	31	2.93	.49	2	4
	Refugee	28	2.66	.6	1	4
Class	1	10	2.87	.41	2	3
	2	9	2.84	.62	2	4
	3	12	2.83	.69	1	4
	4	8	2.65	.40	2	3
	5	10	2.85	.53	2	3

To help describe further, the descriptive statistics for each of the TIPPS concepts is provided in Table 13. Results indicate that among the items 61.7% of the observed educators were found to have positive practices from which 31.7% were very accurate at providing opportunities for students to work together. Nonetheless, 38.4% of the observed educators never had such a practice in their class, with 21.7% of them severely failing to practice it in any form. These indicate that significant proportion of educators need support to help them create opportunities for learners to cooperatively learn, and this could benefit from implementing LtP approach.

For the implementation of LtP, it is essential to design class activities and assignments that take into consideration students' interests and needs to be aligned with the scaffolding of learning. The result revealed that about 71.7% of the observed classes demonstrated that the educators either somewhat or accurately considered students' interest in their classes. In the same way, about 61.7 % of the educators used scaffolding to promote pupil learning and understanding of subject matter.

As noted above, implementing LtP not only helps to create opportunities for cooperative learning and to use students' ideas and interests to inform class activities and assignments and use scaffolding but also supports educators in implementing instructional strategies to aid pupils in critical thinking and

connecting activities and subject matter to a central instructional concept or learning objective. In relation to these, the data revealed that 51.7% of educators tend to have a positive implementation of useful instructional strategies. 63.3% were observed extending responses for deeper learning and linking concepts. Despite these, about 48.3% of the educators were found to have failed to implement instructional strategies to aid pupils in critical thinking and 36.7% to extend responses for deeper learning and link concepts. It is vivid to conjecture how much pupils attending those classes would be influenced in developing critical thinking and linking concepts.

An investigation was made with the intent of supporting student expressions and use of students' ideas and interests to inform class activities and assignments; instructional strategies to aid pupils in critical thinking; models quality language expression to advance pupil understanding and use of language; asks open-ended questions and closed-ended questions to facilitate deeper learning; extends pupil responses to promote deeper understanding and learning of a concept; does not show favoritism towards some pupils over others; employs gender-responsive strategies; and employs responsive strategies for diverse learners. In this regard, about 71.7% of the teachers observed used students' ideas and interests to inform class activities and assignments, with 51.7% using instructional strategies to aid pupils in critical thinking. 76.7% modeled quality language expressions to advance pupil understanding and use of language enforced by 66.7% asking open-ended questions and closed-ended questions to facilitate deeper learning. About 61.7% were also extending pupil responses to promote deeper understanding and learning of a concept, with 83.3% not showing favoritism towards some pupils over others.

Despite the points noted above, some of the items were helpful to indicate the emotional support and behavior management and the instructional strategies teachers use to support students emotionally (e.g., sensitivity and responsiveness, tone of voice) and to promote positive student behavior (e.g., providing a consistent routine and promoting active student engagement). In this regard, 71.6% of the teachers implemented behavioral indications of positive environment between teacher and pupils and amongst peers. 56.7% were monitoring and were responsive to pupils' academic and emotional needs with 86.6% of them using tone of voice that encourages pupils. Beyond these, 66.6% of the teachers employed behavior management to create an environment that is conducive to learning. 63.3% also established classroom routines to create an environment that is conducive to learning and 71.7% were encouraging pupils in learning activities. In addition, about 66.7% of the teachers employed gender responsive strategies and 60.3% employed responsive strategies for diverse learners.

Table 13. Observed Teacher Instructional Practices and Processes Systems

TIPPS Items	A		B	
	Very Accurate	Somewhat Accurate	Somewhat Accurate	Very Accurate
Teacher connects activities to a key concept or learning objective	9 (15.0%)	13 (21.7%)	20 (33.3%)	18 (30.0%)
Teacher provides specific feedback	4 (6.7%)	17 (28.3%)	20 (33.3%)	19 (31.7%)
Teacher extends pupil responses to promote deeper und	11 (18.3%)	12 (20.0%)	19 (31.7%)	18 (30.0%)
Teacher uses scaffolding	12 (20.0%)	11 (18.3%)	19 (31.7%)	18 (30.0%)
Teacher creates opportunities for cooperative learning	13 (21.7%)	10 (16.7%)	18 (30.0%)	19 (31.7%)
Teacher incorporates pupils ideas and interests to inform activities	6 (10.0%)	11 (18.3%)	24 (40.0%)	19 (31.7%)
Teacher uses instructional strategies to aid pupils in critical thinking and problem solve	9 (15.0%)	20 (33.3%)	16 (26.7%)	15 (25.0%)
Teacher connects studies to everyday life experiences	13 (21.7%)	16 (26.7%)	20 (33.3%)	11 (18.3%)
Teacher models quality language	8 (13.3%)	6 (10.0%)	22 (36.7%)	24 (40.0%)
Teacher asks open-ended and close-ended questions	11 (18.3%)	9 (15.0%)	18 (30.0%)	22 (36.7%)
Teachers use instructional materials that facilitate learning	11 (18.3%)	18 (30.0%)	23 (38.3%)	8 (13.3%)
Behavioral indications of positive environment	5 (8.3%)	12 (20.0%)	20 (33.3%)	23 (38.3%)
Teacher monitors/is responsive to academic and emotional needs	11 (18.3%)	15 (25.0%)	15 (25.0%)	19 (31.7%)
Behavioral indications of negative environment	18 (30.0%)	21 (35.0%)	10 (16.7%)	11 (18.3%)
Teacher tone of voice influences children	6 (10.0%)	2 (3.3%)	17 (28.3%)	35 (58.3%)
Teacher employs behavior management	7 (11.7%)	13 (21.7%)	23 (38.3%)	17 (28.3%)
Teacher establishes classroom routines	7 (11.7%)	15 (25.0%)	27 (45.0%)	11 (18.3%)
Teacher shows favoritism towards some children	5 (8.3%)	5 (8.3%)	18 (30.0%)	32 (53.3%)
Children are engaged in learning activities	5 (8.3%)	12 (20.0%)	24 (40.0%)	19 (31.7%)
Actively employs gender responsive strategies	7 (11.7%)	13 (21.7%)	18 (30.0%)	22 (36.7%)
Actively employs responsive strategies for diverse learners	8 (13.3%)	14 (23.3%)	20 (33.3%)	18 (30.0%)

Conclusions and Recommendations

Conclusions

The Government of Ethiopia (GoE) has long maintained an open-door policy for refugees and asylum seekers and is currently the second largest refugee-hosting country in Africa and the sixth worldwide. In 2018, Ethiopia made nine pledges to respond comprehensively to refugee needs and formulated a National Comprehensive Refugee Response Strategy (NCRRS). One of the pledges is to increase the enrolment of refugee children in preschool, primary, secondary, and tertiary education without discrimination and within available resources. In 2019, the GoE adopted revisions to its existing refugee law and made it one of the most progressive refugee policies in Africa by allowing refugees to access primary education, obtain work permits drivers' licenses, and access health and financial services. Furthermore, the GoE is one of the signatories of the Djibouti Declaration (2017), which includes an agreement to integrate the education of refugees and returnees into the national education sector plans by 2020. However, these efforts to expand refugee's access to education is met with a lack of evidence and data regarding education outcomes and interventions. Regarding LtP approaches in education, multiple studies from high-income countries have found a positive impact of high-quality, playful pedagogies on holistic learning outcomes, but evidence about the relationship between playful learning approaches and children's learning outcomes in low- and middle-income countries is still emerging. In addition, there is limited research with older groups of children, with displaced children and/or those living in refugee contexts or, more generally, in contexts of crises and conflict.

Education in crisis and conflict contexts has been calling for action to respond to the learning crisis, and providing policymakers, governments, and stakeholders with relevant and actionable data is the first step. This report fills a gap in the evidence by providing a comprehensive snapshot of education with data from learners and educators in both refugee and host communities, while reporting baseline data for PlayMatters, an innovative project, one of the first large-scale programs to advance Learning through Play (LtP) approaches in refugee settings, and one of very few to focus on the integration of LtP into primary classrooms (most focus on pre-primary or out-of-school settings). In the following paragraphs, we summarize the conclusions and recommendations emerging from the baseline results described in this report.

Literacy and Numeracy: Children have strong achievement in early literacy skills, like oral vocabulary and letter, but they face difficulties in higher-order skills and struggle to read and comprehend what they read.

It is important to tune learning by using vocabulary the children know around their communities, and track learning and progression of learning competencies aligned with performance-based support; and strengthen the professional development of the educators in teaching reading.

The performance of children for numeracy declines as the difficulty level increases, and while children have strong performance in number identification and discrimination, they struggle to perform mathematical operations of addition and subtraction and word problems.

It is useful to track learning progression and implement a tested approach, as LtP, to support learners in concretizing the abstract form of math and support educators in relating their engagement with the support they receive to teach math in schools.

A statistically significant difference was observed between boys and girls for most of the sub-components of EGRA and EGMA, while this was not the case for the community type.

It is essential to devise ways and implement necessary support for girls to enhance their learning.

SEL: The socio-emotional skills of all the children were of medium, with no statistically significant difference between boys and girls and refugee and host community children. But differences were observed between classes.

It is important to provide support to improve the overall socio-emotional skills and help the children to develop empathy, reduce aggressions and conflicts, and promote living together harmoniously.

Wellbeing: The health-related quality of life was unfavorably rated for physical and emotional wellbeing, while they are favorable for family, friends, and school.

It is wise to develop a system where school and other partners such as family work together to enhance children wellbeing. It could also be wise to explore cultural characteristics that are directly linked to child wellbeing.

Child Characteristics: A number of child characteristics and home background have their contributions to enhance child learning and wellbeing, but the children seem to be lacking resources, spend time in non-educational activities, and some are forced to engage in several chores that could distract learning.

It is useful to have a database of those children and plan accordingly to support a child.

Educators Perceptions and Wellbeing: Educators seemed to perceive Learning through Play positively which is important, but they have limited capability to engage in developmental and educational activities and help children develop self-regulation skills when implementing teaching.

Educators need professional development support to improve their competence in implementing LtP.

Educators were found to have close to high wellbeing and their self-efficacy was found to be above the average that stands out to be between applicable and

strongly applicable.

It will be useful to utilize these and build their competence to integrate play-based instruction into daily lessons.

Classroom Quality: The educators' classroom practices were found to be positive and at a level of somewhat accurate, with only 43% of them fulfilling quality standards. However, the literacy and numeracy of the children were not also up to the level, including those schools.

It is good to have a formative nature of observing teachers' classroom practices and provide feedback to help them develop self-reflection and improved actions that could improve children's performance by integrating play-based instruction into daily lessons.

Feeback from data collection pointed out that enumerators' competence and quality of data could be considered as a limitation on the measurement and use of the TIPPS tool, suggesting revisions to the training, and possibly materials, is warranted.

References

Abraha, A. (2018). "Quality of Early Years Education (EYE) in Ethiopia: A meta –Analysis." *Ethiopian Journal of Behavioral Studies.* 1 (1), pp. 31-46.

Ahmed, M., &Saleh, A. (2022). Educational inequality in the Kebribeyah Somali Refugee Camp in Ethiopia: an auto-ethnography. *Journal of international humanitarian action,* 7 (1), 1-13.

Bullinger, M., Mackensen, S., Kirchberger, I. (1994): KINDL - ein Fragebogen zur gesundheitsbezogenen Lebensqualität von Kindern. Zeitschrift für Gesundheitspsychologie 2: 64-67

Fransen, S., Vargas-Silva, C. & Siegel, M. (2018). The impact of refugee experiences on education: evidence from Burundi. IZA Journal of Development and Migration. 8 (6), 1-20.

Frost, J. L., Wortham, S. C., & Reifel, S. (2012). Play and child development. Pearson.

FDRE. (2019). Refugees Proclamation NO. 1110 /2019. Federal Negarit Gazette of the *Federal Democratic Republic of Ethiopia.* Addis Ababa.

Gaffar, A., & Campbell, A. (2021). Learning through Play: A review of evidence related to play for children's education and development in humanitarian and low-resource contexts. PlayMatters.

Hyder, T. (2005). War, conflict and play. Maidenhead: Open University Press.

Inter-Agency Network for Education in Emergencies (INEE). (2019). 2019 Annual Report. https://inee.org/sites/default/files/resources

MacKinnon, H. (2014). *Education in Emergencies: The Case of the Dadaab Refugee Camps,* CIGI, Canada.

McCarthy, A. (2017). Non-state actors and education as a humanitarian response: role of faith-based organizations in education for Syrian refugees in Turkey. *Journal of International Humanitarian Action,* 2(13). https://doi.org/10.1186/s41018-017-0028-x

National Organization for Examinations (NOE/MoE) (2000). Ethiopia First National Learning Assessment of grade 4 students, Addis Ababa, Ethiopia.

National Organization for Examinations (NOE/MoE) (2004). Ethiopia Second National Learning Assessment of grade 4 students, Addis Ababa, Ethiopia.

General Education Quality Assurance and Examinations Agency (GEQAEA/MoE) (2008). Ethiopian Third National Learning Assessment of Grade Four Students, Addis Ababa. Ethiopia.

National Educational Assessment and Examination Agency (NEAEA/MoE) (2013). Forth National Learning Assessment on Grades 4 And 8 Students: Analytic Report, Addis Ababa, Ethiopia.

MoE, MoH, & MWA (2010a). ***National Policy Framework for Early Childhood Care and Education (ECCE) in Ethiopia.*** Addis Ababa.

New York University (NYU) (2021). Global Ties for Children. Retrieved from https://steinhardt.nyu.edu/ihdsc/global-ties/research/early-childhood-development-and-education

Onditi S. A., Otengah W.A., & Odongo, B.C. (2018). Influence of Teachers Perceptions on Play-Based Activities in Pre-School Curriculum Implementation in Homabay County. *International Journal of Humanities and Social Science Invention (IJHSSI).* Vol. 7 Issue 09, PP 62-72

Piper, B. (2010) Ethiopia Early Grade Reading Assessment. (RTI International)

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

UNESCO Institute for Statistics and UNHCR. (2021). *Refugee Education Statistics: Status, Challenges and Limitations.* Montreal and Copenhagen, UIS and UNHCR.

UNHCR. (2017a). *Global Trends: Forced Displacement in 2016,* http://www.unhcr.org/statistics/unhccrstats/5943e8a34/global-trends-forced-displacement2016.html

UNHCR (2017b) *Biennial programme budget 2018-2019 of the Ofce of the United Nations High Commissioner for Refugees.* https://www.unhcr.org/en-us/excom/excomrep/59c276a27/biennial-programme-budget-2018-2019ofce-united-nations-high-commissioner.html

UNHCR (2017c) Working Towards Inclusion: Refugees Within the National Systems of Ethiopia

UNHCR. (2020). Refugee Education Strategy: Towards Inclusion. Addis A

UNHCR. (2021). Ethiopia December 2021 Fact Sheet. Addis Ababa

UNHCR. (2022a). *Fact Sheet February 2022.* Ethiopia.

UNHCR. (2022b). Education Fact Sheet (Ethiopia). Addis Ababa.

USAID-IQPEP (2014). Early Grade Reading Assessment (EGRA) (Improving Quality Primary Education Program-IQPEP), Addis Ababa

USAID READ-TA (2014). Ethiopia 2014, Early Grade Reading Assessment; Report of Findings; RTI

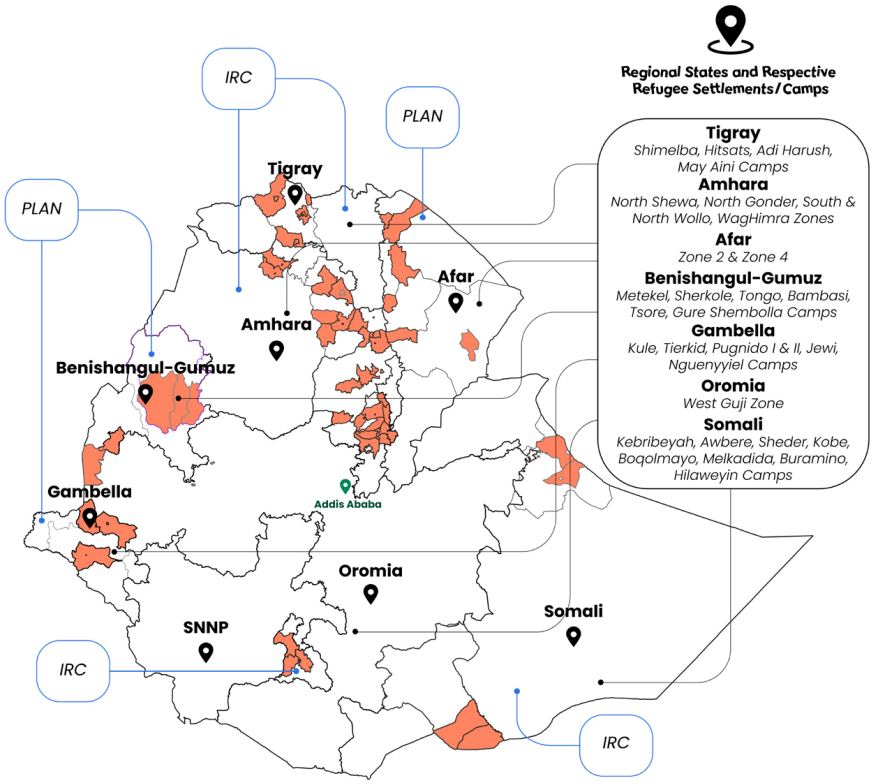
International, September, 2014.
USAID-READ M&E (2016). Early Grade Reading Assessment (EGRA): Midterm Report. USAID Reading for Ethiopia's Achievement Developed Monitoring and Evaluation (Read M&E), Addis Ababa
USAID-READ M&E (2018). Early Grade Reading Assessment (EGRA): Endline Report. USAID Reading for Ethiopia's Achievement Developed Monitoring and Evaluation (Read M&E), Addis Ababa
Vygotsky, I. (2016). Play and its role in mental development of the child. *International Research in Early Childhood Education.* 7 (2), 3 -25. (A translation of the First publication: Vygotsky, L. (1966). Igra i ee rol v umstvennom razvitii rebenka, Voprosy psihologii [Problems of psychology], 12(6), 62–76).
Wolf, S., Raza, M., Kim, S., Aber, J. L., Behrman, J., and Seidman, E. (2018). Measuring and predicting process quality in Ghanaian pre-primary classrooms using the Teacher Instructional Practices and Processes System (TIPPS). *Early Childhood Research Quarterly,* 45, 18–30. https://doi.org/10.1016/j.ecresq.2018.05.003
Yogman, M., Garner, A., Hutchinson, J., Hirsh-Pasek, K., Golinkoff, R. M., Committee on Psychosocial Aspects of Child and Family Health, & Council on Communications and Media. (2018). The power of play: A pediatric role in enhancing development in young children. *Pediatrics,* 142(3), 1–17
Yohani, S. C., & Larsen, D. J. (2009). Hope lives in the heart: Refugee and Immigrant children's 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.

Annex

Summary of EGRA and EGMA Results Percent Correct and Fluency by Category of Sex

		Zero	Below Target	Target or Above Target
Oral Vocabulary	Boys	0.00	7.64	92.36
	Girls	0.00	12.24	87.76
Letter Identification	Boys	16.67	41.67	41.67
	Girls	25.34	52.74	21.92
Oral Reading	Boys	41.84	30.50	27.66
	Girls	53.42	33.56	13.01
RC	Boys	12.31	53.85	33.85
	Girls	12.50	59.38	28.13
Number identification	Boys	3.47	14.58	81.94
	Girls	6.85	32.88	60.27
Number discrimination	Boys	3.47	17.36	79.17
	Girls	9.52	21.09	69.39
Missing numbers	Boys	9.03	55.56	35.42
	Girls	8.84	66.67	24.49
Addition L1	Boys	6.34	43.66	50.00
	Girls	5.48	65.07	29.45
Addition L2	Boys	40.00	30.37	29.63
	Girls	58.99	23.74	17.27
Subtraction L1	Boys	11.97	68.31	19.72
	Girls	13.61	71.43	14.97
Subtraction L2	Boys	45.67	31.50	22.83
	Girls	64.57	22.83	12.60
Word Problem	Boys	9.35	46.76	43.88
	Girls	17.39	56.52	26.09

Where We are Working



Ethiopia

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