PlayMatters

Implementation Research of PlayMatters 1.0

Baseline Report for ECD School

PlayMatters Project

The PlayMatters Consortium is 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
Idalia Rodriguez Morales, Education Researcher, IRC – Airbel Impact Lab
Asham Assezenew, Senior Research Manager, IRC Ethiopia
Dr. Silvia Diazgranados Ferráns, 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 ECD 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) PlayMatters project teams in Somali, Gambella, and Benishangul Gumuz (BGR) Regions for their support in enumerator selection, training, and data collection monitoring, which was key for completing data collection in time.

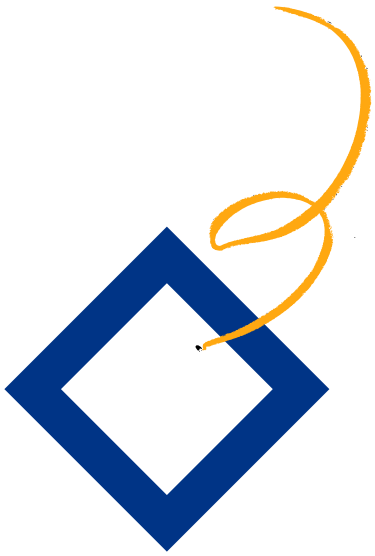
The study team would like to thank the guidance and support from the PlayMatters project Research, Monitoring, Evaluation, Accountability and Learning (RMEAL) and the PlayMatters Regional RMEAL team, which were remarkable in terms of providing guidance and leadership, organizing capacity building workshops,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

The data collection involved enumerators from respective sites who speak the children's language. They played very pivotal roles in collecting quality and complete data. Thus, the team is very grateful to all.
Last but not least, the study team would like to acknow
ledge the support from IRC and PIE field offices in the three PlayMatters project locations and the community.

Contents

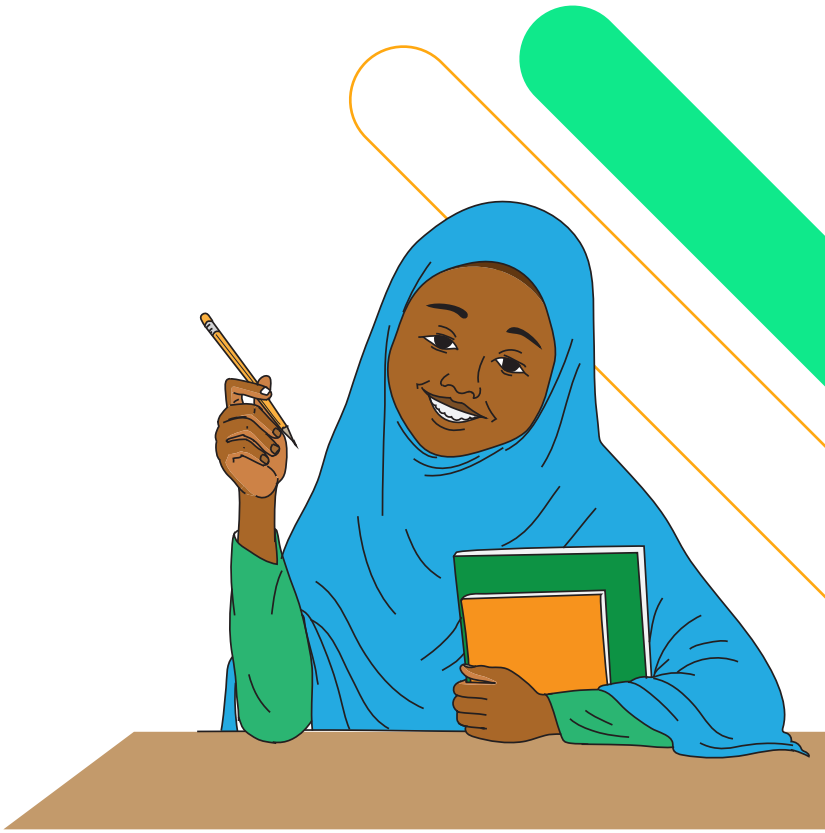
Acronyms/Abbreviations

BGR	Benishangul Gumuz Region
BIT	Behavioral Insights Team
ECCE	Early Childhood Care and Education
ECD	Early Childhood Development
ESSSWA	Ethiopian Society of Sociologists, Social Workers, and Anthropologists
GER	Gross Enrolment Rate
PM	PlayMatters
IDELA	International Development and Early Learning Assessment
IPA	Innovations for Poverty Action
IRB	Institutional Review Board
IRC	International Rescue Committee
KG	Kindergarten
LtP	Learning through Play
MoE	Ministry of Education
NGO	Non-governmental Organization
PHRP	Protecting Human Research Participants
PI	Principal Investigator
PIE	Plan International Ethiopia
SEL	Socio-emotional Learning
TIPPS	Teacher Instructional Practices and Processes System
ToT	Training of Trainers
UNESCO	United Nations Educational, Scientific and Cultural Organization
UNHCR	United Nations High Commissioner for Refugees
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.

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

Effective Learning through Play (LtP) pedagogy involves five basic principles that matter most for its implementation in classroom and evaluation of its outcomes: (a) actively engaging, (b) meaningful (relates learning to personal experiences), (c) joyful (motivates learners to get involved, face and overcome difficulties), (d) iterative – make sustained practices to learn further, prove assumption or deepen learning, and (e) socially interactive – bring two or more children to compete, contribute thoughts for common goal and strengthen relationship.

Ethiopia hosts over 800,000 refugees who have flown from neighboring countries such as South Sudan, Somalia, and Eritrea. As of the academic year ending in June 2022, a total of 52,486 were attending preschool education in ECD centers in respective areas. This enrollment figure constituted about 31% of the school-age children registered and a 43% Gross Enrollment Rate (GER) of the refugee population. Though documents show low participation rates and system efficiency, evidence on play pedagogy and learning outcomes in refugee ECDs is still thin. The PlayMatters project in Ethiopia intended to define a baseline of the characteristics of children and educators in the intervention, both in the refugee and host communities.

Thus, this study on ECD was guided by the following questions:

Child Level

1. What are children's literacy and numeracy skills? How do they vary by displacement, sex, and region?
2. What is the status of children's socio-emotional skills and wellbeing? How do they vary by displacement, sex, and region?
3. What are the child characteristics explaining the variation in children's literacy, numeracy, and socio-emotional (SEL) skills?

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 a systematic random sampling technique to identify educators and children from the PM intervention refugee and host community ECDs. Accordingly, a total of 108 ECD children (53 girls and 55 boys) and 23 ECD educators participated in this baseline study. Children's data collected included the International Development and Early Learning Assessment (IDELA) to measure early school readiness skills and Kiddy-KINDL to measure socioemotional skills. 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 surveys to assess their burnout levels and self-efficacy. Educator data also included an instructional observation of a class thought by the educator using the Teacher Instructional Practices and Processes System (TIPPS) tool.

The data collected from children and educators were quantitative in nature, so the study employed descriptive and inferential statistical techniques for the appropriate analyses. We summarize the main findings, concluding remarks, and implications of the study in the following paragraphs.

Major Findings

A. Child Level

1. The overall IDELA mean score showed that, on average, children gave correct answers to about 56% of the questions. Results varied across the different domains IDELA covers to assess children's holistic development. Children scored 58% in executive function, 63% in motor skill development, 49% in emergent literacy, 55% in emergent numeracy, and 53% in socioemotional skills. Further look at subtasks' performance by domains showed that children scored lowest in drawing a person (44%), letter identification (24%), number identification (39%), and friends (41%) from motor skill, emergent literacy, emergent numeracy and socio-emotional domains respectively.

a. We explored differences by sex, community type, and region. Girls scored 54% in the overall IDELA compared with boys, who scored 58%, but this difference was not statistically significant. Boys outperformed girls in all domain scores, though the only statistically significant difference was for the emergent numeracy score. We also find statistically significant differences between boys and girls in self-regulation (executive function domain), letter identification and letter sound (literacy domain), shape identification, number identification, number sense, operations (numeracy domain), and number of friends (socioemotional domain).

b. Children from the host community and refugee settings performed comparably on both the overall IDELA and domain scores. Children from Somali region achieved higher (59%) compared to those from Benshangul Gumuz (45%) and Gambella (50%). A post-hoc analysis found these differences to be statistically significant.
c. Analyses comparing the proportion of children who scored less than 50% in IDELA across different groups revealed that girls, children from refugee communities, and those in the Gambella and Benshangul-Gumuz regions are at disadvantage.

B. Educator Level

1. The overall mean TIPPS score (score range: 1-4) for the ECD classrooms observed is 3.32, suggesting the quality of the learning environment in the classrooms observed is high and that most of the indicators of quality classrooms were visible during the classroom observations. In fact, in 19 of 23 classrooms observed, classrooms showed most of the quality indicators assessed. Areas for improvement, however, include the use of scaffolding to promote learning and understanding, connecting concepts and learning objectives to children's everyday life experiences, promote positive relationships and between teacher and learners and amongst peers, and reduce the presence of favoritism on the classroom.
2. The mean score of educators' perception of LtP survey (score range: 0-4) is 3.28, indicating that educators generally perceived learning through play positively, implying their readiness to practice if they are provided with support.
3. The overall rating of the educators' engagement in LtP developmental and educational activities (measured by DEAS, with a score range between 0-4) was moderate (M = 2.45). When analyzed by subscale, preschool educators seem to have a better mean score on the emotional support subscale (M = 2.84), and lower scores in the play (M = 2.31) and pretend play (M = 2.27) subscales.
4. The educator's mean work wellbeing and engagement score (measured by Utrecht Work Engagement Scale, with a score range between 0-4) was 3.09, suggesting that preschool educators have high levels of occupational well-being. Educators scored highest in the dedication subscale (M = 3.31), compared with the vigor (M = 2.90) and absorption (M = 3.06) subscales.
5. The total self-efficacy (score range: 0-4) mean score for preschool educators is found to be 2.60, which shows a moderate level of educators' self-efficacy. Educators scored higher in their self-efficacy related to classroom management (M = 2.70), and lower in terms of instructional strategies (M = 2.58) and student engagement (M = 2,52).

Conclusions

Based on the findings of this study, the following conclusions can be drawn:

Child Level

Results suggest that children's literacy and numeracy skills are emerging, and the differences between girls' and boys' development markers are stark. Displacement status (type of community) does not explain the variation in children's performance. Regional variations have been apparent in the mean percent scores on all domains except for the socio-emotional domain, where children from the Somali region appear to be better. Generally, children's well-being or quality of life is found to be high. The region, sex, and types of community are not the child characteristics that explain the variation in children's socio-emotional well-being.

Educator Level

The perception of preschool educators toward LtP is positive and favorable, which could be taken as one of the inputs of preschool educators to implement LtP in preschool settings. The quality of the learning environment in the classrooms observed is quite high. The indicators of quality classrooms are apparently visible in most of the classrooms observed.
The occupational well-being of the preschool educators who participated in the baseline study is high, their engagement in all developmental and educational activities of children measured by DEAS is moderate, and their self-efficacy is moderate. These are important educator characteristics that should be addressed during the intervention so as to base the implementation of LtP on a strong foundation.

Recommendations

Based on the findings of this baseline study, the following recommendations are forwarded.

1. Emphasize foundational skills which will ultimately contribute to the development of higher-order skills in numeracy and literacy.
2. Target and monitor differential experiences of girls and boys to address differences in outcomes. Teacher and teacher professional development should also be designed with the objective of closing gender gaps in both experience and outcomes.
3. The professional development activities for educators should align with PlayMatters' priorities and focus on deepening their understanding of LtP and child-centered pedagogy.
4. Future research should align measures with the PM 2.0 redesign and programmatic changes to enhance the accuracy, relevance, and comprehensiveness of the research findings. Research should integrate measures aligned with key aspects and outcomes, considering the objectives and intended impact of the implementation.
5. Future research should employ a mixed-methods approaches. Additionally, increasing sample sizes for teachers and regions is recommended to enable more detailed analyses.
6. Continuous monitoring and research are necessary to equip teachers with the skills and strategies needed for a student-centered classroom. This includes assessing changes in instructional skills and in-classroom experiences, as well as providing a training package that focuses on supporting and enhancing children's play.

Background

Learning through Play in Conflict and Crises

Learning through Play (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. Evidence supports that LtP can improve holistic outcomes for children more effectively than either traditional instruction or free play (Yogman et al., 2018). LtP enhances the cognitive, social, physical, and emotional aspects of children by improving engagement and motivation. While playing, children practice self-regulation when taking turns, accepting losses, and managing conflicting interests (Vygotsky, 2016). When playing in teams, children practice communication and collaboration skills (Zosh et al., 2017). When competing, children also practice persistence and self-perception (Gaffar & Campbell, 2021).

The LEGO Foundation defines key elements of LtP as: i) active engagement, ii) meaningful learning where children relate new experiences to what they already know, iii) enjoyment of tasks and the thrill of surprise, insight, or success after overcoming challenging experiences, iv) iteration - trying out possibilities, revising hypotheses and discovering new questions, and v) social interactivity - so that children are able to communicate thoughts, share ideas, understand others and enjoy being with them and build stronger relationships (Zosh et al., 2017).

Crises differentially affect children and adults. First, children lack the physical and emotional readiness to cope with the consequences of the crisis. Second, children and their wellbeing depend on their caregivers, who are themselves being affected by the crisis. Third, violence or crisis has far-reaching negative effects on children than on adults. Studies on children exposed to war and separation from familiar environments and relationships reveal, for example, that such children are affected by emotional stress, and the consequences become more severe when children are separated from their parents due to crisis (Osofsky, 1999). LtP is a relevant pedagogical approach for children in crisis-affected settings as it helps to discharge emotions and develop hope and coping mechanisms (Yohani & Larsen, 2009). Amidst challenges, "play is both the way that children express themselves and the means through which they resolve issues" (Hyder, 2005, p.23). Play is a fundamental component of developing lasting intrinsic mechanisms and motivations.

Though play is a universal human characteristic and LtP is appropriate to enhance holistic learning, educators face extreme challenges in refugee settings (INEE, 2019). Currently, there is little evidence from low-resource contexts on how to assist educators in overcoming the challenges they face when implementing LtP in humanitarian settings.

Education Context in Ethiopia

Early Childhood Development (ECD) in Ethiopia started in 1900 in Dire Dawa for the kids of French consultants involved in the construction of a railway from Djibouti to Addis Ababa (Hoot et al., 2004). However, the scope of ECD education is still limited, an urban phenomenon, and run mostly by private investors, NGOs, and communities (Abraha, 2018; Rossiter, 2016). For instance, ECD services reached only 3,418 of the approximately 7.3 million of ECD-aged children (ages 4-6) in 2010 (MoE, 2011a). It was only after the establishment of the first sub-sector policy (MoE et al., 2010a; 2010b) that the Gross Enrolment Rate (GER) of the sub-sector jumped from 5.2% in 2010/11 to 21.6% in 2011/12 and reached 36.7% in the 2020/21 academic year (MoE, 2012; 2021). There are two types of ECD programs in Ethiopia: (a) a three-year-long formal ECD program with three levels, sometimes called KG1, KG2, & KG3; and (b) a one-year long program, called O-class, aimed at preparing age 6 children for formal primary education. O-class is annexed to primary schools where learners enroll at age 6 and stay for a year until they join primary schools (MoE, 2015). However, practice tells us that children of ages (4-7 years old) join O-class. (Tirusew & Belay, 2016). Though limited, evidence on the quality of ECD education shows more challenges than achievements in learning outcomes (Abraha, 2018; Rossiter, 2016; Rossiter et al., 2018; Dowd et al., 2016). As is the case in refugee settings globally, research on LtP in the Ethiopian education system is scant.

From 2016 to 2021, the global refugee population grew from 22.5 to 26.4 million persons (UNHCR, 2017a; UNESCO Institute for Statistics & UNHCR, 2021). Of the global total, over 80% of refugees (for example, 84% in 2016 and 86% in 2021) are hosted in developing countries (Ahmed & Saleh, 2022). Ethiopia is the third largest destination in Africa and hosts over 800,000 refugees, mostly from South Sudan, Somalia, and Eritrea, displaced because of security and political issues. Refugee support responses in Ethiopia primarily focus on protection, shelter, education, health, food security and nutrition, and water and sanitation sectors. As of the academic year ending in June 2022, 52,486 (30.6%) of the total 171,447 school children were attending pre-primary education in the refugee ECD centers in the nation (UNHCR, 2022a).

Ethiopia adopted an integrative policy in 2019 that treats refugees the same as Ethiopian nationals in access to preschool (FDRE, 2019) and developed a refugee education strategy for the period from June 2020 to 2025 (UNHCR, 2020). UNHCR (2022b) has identified that, though problems of refugee education vary by location, low participation and low internal efficiency are persistent over time. ECCE Gross Enrollment Rate (GER) among refugee children reached 50.84% in 2018/19, higher than the national average that fell to 41% at the time. In the 2020/21 academic year, GER for refugee preschool education stood at 43%, still higher than the national GER of 36.7% (UNHCR, 2022).

[1] Except in citations, ECD in this document is used as an inclusive word for formal three-year Kindergarten programs, O-classes, preschool, as well as the varying acronyms for the pre-primary child support activities.

It is unfortunate that low-income countries that cannot even adequately finance social services such as education for their own citizens disproportionately host and support refugees, frequently becoming a burden on national systems. In such situations, the global community continues to prioritize the provision of food, shelter, and safety rather than education (MacKinnon, 2014; Mccarthy, 2017; UNHCR, 2017b). As a result, evidence on the learning outcomes of preschool children remains scarce to inform scholars, humanitarian agencies, and all concerned bodies (Fransen et al., 2018). However, the low quality of ECD services in refugee camps is likely indicative of the broader education system in Ethiopia.

The Present Study

Research Aims and Questions

This study aim was to conduct a baseline assessment of the initial characteristics of children and educators engaged in PlayMatters' intervention in refugee and host community ECD centers. Results intend to inform program implementation and provide insights to improve and refine the PM package. The original research design considered baseline, midline, and endline data collection to evaluate PM's implementation. This baseline in ECD centers for the Implementation Research Study took place in February and March 2022 in Somali, BGS, and Gambella. However, delays in program implementation and revisions to the project strategy led the research team to save resources and cancel the midline and endline data collection points. Thus, the research questions were revised to the following:

Child Level

1. What are children's literacy and numeracy skills? How do they vary by displacement, sex, and region?
2. What is the status of children's socio-emotional skills and wellbeing? How do they vary by displacement, sex, and region?
3. What child characteristics explain the variation in children's literacy, numeracy, and SEL skills?

Educator Level

1. What is the perception of educators on LtP?
2. What are teachers' instructional practices?
3. What are the levels of educators' work engagement, occupational wellbeing and self-efficacy?

Ethical Considerations

This study received ethical clearance from the IRC's institutional review board (IRB) in August 2021 and from the Ethiopian Society of Sociologists, Social Workers, and Anthropologists (ESSSWA) in March 2022. All the Principal Investigators (PIs) and coPIs involved in the study completed a web-based course on protecting human research participants' online training provided by Protecting Human Research Participants (PHRP). Before collecting any data, the team obtained parental consent explaining to participants the purpose of the study and the voluntary nature of it. While collecting data, consent from parents and educators and assent from children were obtained, anonymity was kept, and the use of the data collected was limited to the purpose of the study only.

Participants

The baseline study covered three regions of Somali, Gambella, and Benishangul Gumuz (BGR). The sample included 11 refugee and host community ECDs/O-classes purposely selected because they are targeted by the PM project. Within schools, teachers were selected in consultation with respective head teachers or by availability sampling when there was just one classroom or teacher. Within each classroom, five children were selected using a gender-specific systematic random sampling technique. Table 1 shows the number of participating ECD centers and the total sample sizes by region, community type, sex for both children and educators and language of assessment or medium of instruction.

Table I. Summary of Sample

Region	Location Type	# of Sites	# of Children Assessed			# of Children Observed			Medium of Instruction
			F	M	Total	M	F	Total	
Somali	Refugee	4	29	30	59	9	4	13	Af Somali
	Host	3	7	7	14	1	2	3	Af Somali
Gambella	Refugee	1	5	5	10	2	-	2	Nuer
	Host	1	2	3	5	1	-	1	Agnuak
BGR	Refugee	1	7	8	15	1	2	3	English
	Host	1	3	2	5	1	-	1	Berta
Total		11	53	55	108	15	8	23	-

Notes:
# of Educators Observed = 23
# of Educators Surveyed = 16

Of the total 11 ECD centers, six were from refugee settings and five from host community settings functioning in Somali, BGR, and Gambella regions. No ECD centers were sampled from Afar region despite PlayMatters operation due to a lack of ECD project sites in this region. The language of instruction varies by region and type of community. In Somali, the language of instruction (and hence child testing) was Af Somali in both refugee and host community settings. In Gambella, refugee and host communities speak Neur and Agnauak, respectively. In BGR, while the medium of instruction in refugee schools is English, the refugee community speaks various tribal languages. The host community uses Berta as the language of instruction.

A total of 108 children (53 girls and 55 boys) were assessed in this study, with girls representing 49.1% of the total. Of the 108 total children, a majority (n=84, 77.8%) were from refugee centers, and the rest (n=24, about 22.2%) were from host community centers with similar gender proportions. This huge difference is attributed to the absence of formal ECD centers in host communities except for a single O–class integrated in primary schools.

Furthermore, analysis of the age distribution of children who participated in the study showed the following indications: (a) the age distribution was not complete due to the fact that some enumerators skipped asking the date of birth of children or might have forgotten to record the information. (b) analysis based on the available data (N = 49) shows that the average age of the children in the host community was 5.8 years with a standard deviation of 0.42 years, while it was 5.23 years with a standard deviation of 0.81 years in the refugee camps. (c) assessed children in the host community were slightly older than children in the refugee camps. The age difference between the two groups is statistically significant ($t(47) = 2.138$, $p = 0.038$). This difference might be explained in terms of the nature of O-class. In O-class, the entry criteria stipulated in the policy limiting the age of entry to six years old (MoE, 2015) is not enforced in practice as children of different ages join O-class for a year. Children are often 5 and 6 years-old but can be as young as 4 years-old and as old as 7 years-old in some cases. There is no statistically significant difference in the average age of boys and girls.

The study also included classroom observation data of 23 educators conducted in English (15 female and 8 male). Similar to the proportion of children, 18 out of the 23 educators teach in refugee ECD centers, and the remaining five in host community contexts. Furthermore, though a total of 23 educators were observed using the Teacher Instructional Processes, Practices, and Systems (TIPPS) tool, the return rate of educator surveys was about 70% (16 out of 23), with seven dropped cases due to missing data.

A systematic random sampling technique was employed to select children and educators from the respective target ECD centers in all areas. The sampling frame considered both types of ECD services (formal ECD program and O-class) as eligible, and sampling started with a random selection of educators on the basis of 'one class from one level' (for classroom observation and educator survey) followed by identification of five children from the educator's classroom for the intended assessment. Thus, the maximum sample size from each ECD center reached three educators when the ECD center included all three preschool education levels. In cases like O-classes or when there was only one level (KG1, KG2, or KG3), even with multiple sections or streams, only one educator was picked randomly or deliberately when otherwise (per indication of the school authorities).

As shown in Table 2, the majority of educators (12, 75%) have experienced forced displacement. Out of the displaced educators, one female educator from BGR teaching in the host community reported that she is internally displaced, while the remaining 11 identified as refugees from other countries. All of the educators who were not displaced from their homes are females, and three of them are from the Somali region. The data on the educational background yielded that nearly half (7, 43.75%) of the educators have a national diploma, 6 (37.5%) educators have a primary level of education, and 3 (18.75%) educators have a certificate of education.

On average, educators served as teachers for 4.4 years with a standard deviation of 3.88 years and minimum and maximum years of experience of 1 year and 16 years, respectively. When disaggregated by region, the mean years of experience of educators in Gambella is 6.8 years with a standard deviation of 6.4 years, and that of Benishangul-Gumuz and Somali regions are 3.8 and 3.6 years, respectively. Upon analysis, only two educators in the host community could be considered 'tenured' teachers based on their greater years of experience teaching compared to their peers. Table 2 presents the results of the analysis.

Table 2. ECD Educators' Background Variables

Background Variable	Total	Sex		Community Type		Region		
		Male	Female	Refugee	Host	BGR	Gambella	Somali
Experienced Forced Displacement (n=16)	12 (75%)	5 (31.3%)	7 (43.8%)	11 (68.8%)	1 (6.3%)	4 (25.0%)	3 (18.8%)	5 (31.3%)
Qualification:								
a) Certificate	3	1	2	2	1	2	0	1
b) National Diploma	7	2	5	6	1	0	1	6
c) Primary	6	2	4	6	-	2	3	1
Experience in Years:								
Mean	4.4	3.80	4.73	3.29	12.5	3.75	6.75	3.63
SD	3.88	2.28	4.50	2.02	4.95	3.78	6.45	2.07
Min	1	1	1	1	9	1	1	1
Max	16	6	16	6	16	9	16	6

Measures

Tools for the research study were selected after a careful review of the study objectives, population, and alignment with PlayMatters intended outcomes. Prior to data collection, all tools were translated into the relevant languages, tested, and cognitively pretested (with a sample of comparable children) in the study areas to check suitability and appositeness to the local context. Table 2 summarizes the tools used to collect quantitative data, their score range, subtasks and/or constructs measured, number of items, scoring, and Cronbach alpha for establishing reliability.

We also provide a more detailed description of each measure below.

Table 3. Measures Used in the Study

Tool (Score Range)	Subtasks or Constructs	# of Items	Scoring	Reliability C. Alpha
Child Measures				
IDELA (1-100%)	Child holistic development with focus on: Executive Function (EX); Emergent Literacy (EL); Emergent Numeracy (EN); Socio-Emotional Learning (SEL); & Motor Skills (MO)	24	Total or percent correct, appropriate/inappropriate and Ratio (like in writing skill, hoping, etc.) varying by the type of subtask.	0.74
Kiddy-Kindl (0-2)	Health-related quality of life assessment (Physical health, feelings and relationships).	12	Three-point rating scale (never, sometimes, & very often).	0.79
Educator/Classroom Observation				
TIPPS (0-3)	Quality of classroom environment	20	Two-stage, two-point each rating scale (somewhat accurate & very accurate)	0.79
Educator Measures				
Educator LtP Survey (0-4)	Educator's perception on LtP and their assessment on their capacity to implement LtP strategies in the classroom		Five-point rating scale (0=Strongly disagree; 1= Disagree; 2 = Neutral; 3= Agree; 4 Strongly Agree)	LtP Perception = 0.88
DEAS (0-4)	Developmental and educational activities scale: play, pretend play, self-regulation, & emotional support	35	Five-point rating scale (0=Not at all applicable; 1= A little applicable; 2 = Applicable; 3= Strongly applicable; 4 Very strongly applicable)	DEAS = 0.94
Utrecht (0-4)	Educators' wellbeing and work engagement: vigor, dedication & absorption	9	Five-point rating scale (0=Never; 1=Seldom; 2=Sometimes; 3=Often; 4=Always)	Utrecht = 0.83
Teacher Self-Efficacy (0-4)	Efficacy for instructional strategies, efficacy for classroom management, & efficacy for student engagement	12	Five-point rating scale (0=Not at all applicable; 1= A little applicable; 2 = Applicable; 3= Strongly applicable; 4 Very strongly applicable)	Self-Efficacy = 0.89

As depicted in Table 3, all the tools used in this assessment have reliability coefficients that range between acceptable (0.7-0.8) to excellent (≥ 0.9) (George & Mallery, 2003).

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

II. Kiddy-KINDL: A self-reported questionnaire that measures wellbeing of children between 3 and 17 years of age (Bullinger et al., 1994). The version used in this study has 12 items, covering six dimensions related to children's wellbeing: physical health, feelings, and relationships with family members and friends. All items are scored on a 3-point Likert scale (0= never, 1 = sometimes, and 2 = very often), which yields one overall score and scores by domains. Reverse scoring is applied to two items in the physical well-being dimension and one item in the emotional well-being dimension. The overall scale, as well as individual domains exhibited good levels of reliability, except for the emotional wellbeing domain (0.11), which is not included in this report.

III. The Teacher Instructional Practices and Processes System (TIPPS): An observational tool that measures the quality of teacher´s classroom instructional practices using 20 items. Items are scored on a four-point Likert type scale to illustrate the "degree" to which the concept is present in the classroom. It has been used and validated in the Democratic Republic of Congo, and was adapted later to Uganda, Ghana, Pakistan, and other countries (NYU Global Ties for Learners, 2021). A previous validation study suggested three conceptually distinct constructs or domains, however, due to small sample sizes, we report only aggregate scores and analyses by items.

IV. Teacher Survey: A questionnaire to gather teacher demographic information and assess teachers' perceptions of learning through play, the support they received, and the professional development they undertook.

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

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

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

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

Procedure

In January 2022, researchers and project teams participated in a 6-day regional training of trainers (ToT) in Tanzania which covered the rationale and implementation requirements for all the study's tools. In February and March 2022, enumerators were selected based on specified criteria such as language proficiency, experience, and qualification. Enumerators were then trained for five days covering the objectives of the assessment, assessment procedures and ethical considerations, measures, and use of tablets with both the CommCare and Tangerine software. In addition to paired and small group simulated practices, the training also included school-based practices in all areas (Jigjiga, Gambella, and Asossa towns). From February 14, 2022 – March 5, 2022, researchers and enumerators visited schools to administer the questionnaires and conduct the classroom observations. Enumerators administered questionnaires to children individually in 4 local languages and English (Af Somali, Nuer, Agnuak, and Berta), and spent approximately 1 hour of their

time. They also conducted 45-minute classroom observations. All the data was collected using CommCare and Tangerine, and later cleaned and analyzed by researchers from Addis Ababa University and IRC's Airbel Impact Lab.

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 ANOVA analyses) to answer the research questions. Because of small sample sizes, statistical testing was applied on disaggregation at region level with focus on community type (refugee or host), and sex, but further disaggregation was not possible. In analyzing TIPPS items, frequency distribution or percentage of observed behaviors) employed to understand implications towards quality of classroom instruction and environment. Throughout the report we presented findings in tables, figures and charts indicating the disaggregated when appropriate. We presented statistical details, as per ANOVA findings, by this disaggregation whenever the findings are statistically significant.

Scope and Limitations of the 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 Benishngul-Gumuz, because of availability of PlayMatters project ECDs. 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.

Despite efforts to minimize limitations, the study has passed through some challenges: (a) Due to limited number of ECD centers and educators, and language difference between refugee and host communities in Gambella and Benishangul-Gumuz, disaggregating analysis by language was not possible; (b) Unlike ECDs in refugee camps that provide service for three years, host community ECDs are one-year school readiness programs called 'O' classes.

Thus, comparison of child variables by community type should be understood in view of such program variation in mind; (c) Since all educators observed did not complete the educator survey questionnaire, perception and observation data were analyzed separately; (d) Inclusion of children with disabilities was difficult because technical difficulties in identification and administration; and (e) TIPPS is not fully validated for the context of Ethiopian school system but was used to draw lessons for further practical and research endeavors.

Results

Child Data

IDELA Results

In this baseline assessment, IDELA total score represents the average performance of children in the core domains and provides basic information about baseline levels of holistic child development. In general, both IDELA overall and domain-specific mean scores by sex and community type ranged from 46-64%. The overall mean percent correct score of all children is 56% showing that, on average, children provided 56% of correct responses for all items included in IDELA. This means that nearly half of the IDELA items were difficult for children to answer or attempt them correctly. Disaggregation by sex shows that, boys, as a group, performed better (58%) than girls (54%) in the overall IDELA, but the difference is negligible and not statistically significant. Table 4 provides details on the pattern of overall and domain-specific achievements by sex and community type.

In terms of domain-specific mean scores, children achieved the lowest and highest mean scores in emergent literacy (49%) and motor skill domain (63%), respectively. The mean score for the Executive Function domain score was 58%, suggesting children are still their skills. The difference between girls and boys was statistically significant for self-regulation (53% and 66% respectively). In the area of motor skills, we observe the highest levels of development, with average scores of 63% correct. On average, children exhibited high gross motor development with hopping scores of 83% correct, and lower fine motor development, with average scores as high as 63% correct for copying shapes, and as low as 44% correct for drawing a person. We observe that girls exhibited slightly higher fine motor skills than boys, while boys slightly higher in gross motor skills than girls, but overall, their motor skills are comparable.

The literacy domain mean score was 49% correct, with varying results across subtasks: oral comprehension (66% correct), print awareness (65%) and first letter sound (44%), and lower scores in literacy tasks such as letter identification (24% correct). Compared to boys, girls appear to be disadvantaged as they obtained lower scores in all literacy subtasks, and the difference is statistically significant for letter identification and letter sounds. This finding is concerning because identification of symbols, letters and numbers set the foundation for literacy and numeracy development in particular and further learning competency in general (Glass, 2002; French, 2013; Antoni & Heineck, 2012; Shomos, 2010).

The numeracy domain mean score was 55%, with the lowest score observed for identification of numbers (up to two digits) and the highest score (73%) for simple operations (of numbers between 1 and 3). Looking at sex differences, we observe that girls are significantly behind boys in the emergent numeracy domain, scoring 61% and 50% respectively (t (df = 99) = -2.921, p = 0.004), and in the shape identification, number identification, and number sense subtasks.

With regards to the social emotional domain, the mean score of 55% percent corrects. Children exhibited the highest score for their self-awareness which reflects their knowledge of personal information (73% correct), and their lowest score for the friends subtask, which reflects their ability to name up to 10 friends (44%). We observe that, on average, boys have higher overall SEL score than girls, but this is mainly driven by the fact that boys are able to name a significantly higher number of friends than girls (47% for boys vs 37% for girls).

Analysis of the overall IDELA mean percent correct score by type of community also reveals similar performances between children in the host communities and refugee camps. The larger performance is on the motor domain while the lower performance is on emergent literacy for both communities. Although the difference was not statistically significant, it seems that children from the host community performed slightly better than their counterparts in all domains. For example, the mean percent correct score on emergent numeracy for the host community is 57% and for the refugee camps is 55%.

According to the Ethiopian education policy directive, 50% achievement in each subject is the cutoff point for promotion from grade to grade (MoE & USAID-AED, 2008). Analysis based on this policy framework showed that, of the children assessed, over one-third of them (37.9%) achieved below the requirement, i.e., 50%. Accordingly, we identified the proportion of children who obtain scores below 50% scores to understand which group is at a disadvantage. Results showed that the proportion of girls obtaining scores below the standard is 41.3% while boys not reaching the target is 34.7%. Similarly, the proportion of children from refugee settings not meeting the target is 38.7%, compared to 35% for host community children (35.0%). We conclude that girls and children from refugee context are more disadvantaged than boys and host community children as they showed lower performances in the IDELA overall and core domains.

Table 4. IDELA Mean Percent Correct by Domain, Region and Sex

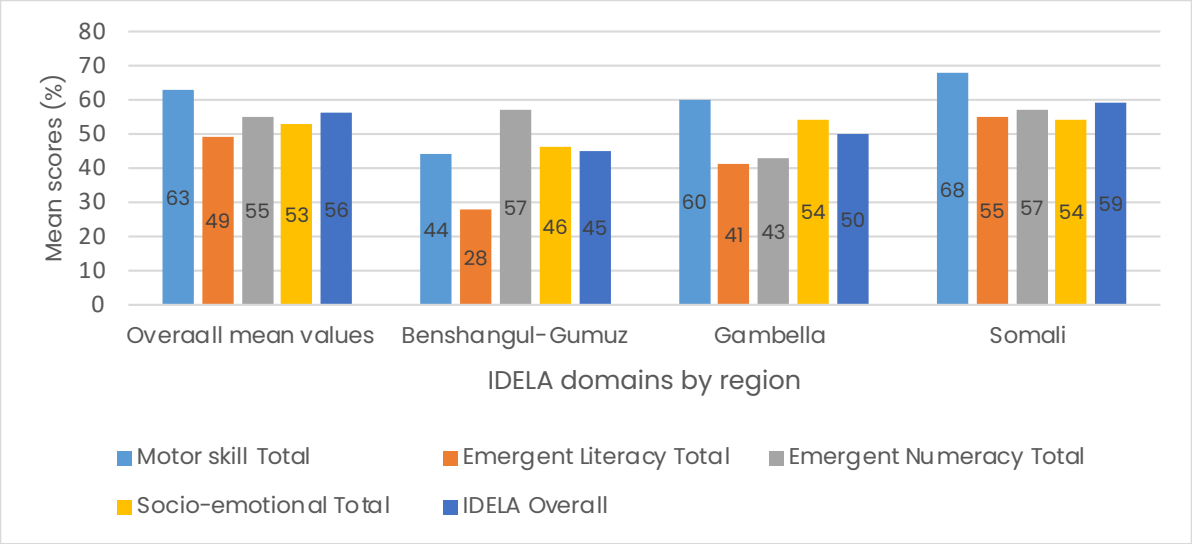
Domains and Overall	Total (N=108)		Sex [a]		Community Type	
	Mean	SD	Boys (n=48-53)	Girls (n=41-51)	Refugee (n=84)	Host (n=24)
Executive Function Domain	**0.58**	**0.26**	**0.62**	**0.54**	**0.57**	**0.62**
Memory	0.57	0.30	0.58	0.56	0.57	0.57
Self-regulation	0.60	0.39	0.66*	0.53	0.58	0.66
Motor Skill Domain	**0.63**	**0.25**	**0.64**	**0.63**	**0.63**	**0.64**
Copying Shape	0.63	0.39	0.63	0.63	0.64	0.62
Drawing a Person	0.44	0.39	0.46	0.43	0.45	0.42
Folding	0.58	0.27	0.57	0.59	0.60	0.53
Hopping	0.83	0.30	0.84	0.82	0.82	0.87
Emergent Literacy Domain	**0.49**	**0.21**	**0.52**	**0.46**	**0.50**	**0.50**
Oral Vocabulary	0.34	0.20	0.35	0.33	0.34	0.32
Print Awareness	0.65	0.38	0.67	0.64	0.66	0.63
Letter Identification	0.24	0.33	0.32**	0.16	0.26	0.18
First Letter Sound	0.49	0.41	0.57**	0.41	0.52	0.39
Oral Comprehension	0.66	0.34	0.63	0.69	0.64	0.74
Emergent Writing	0.43	0.38	0.45	0.40	0.42	0.45
Emergent Numeracy Domain	**0.55**	**0.20**	**0.61*** **	**0.50**	**0.55**	**0.57**
Comparing by Size and Shape	0.59	0.16	0.59	0.60	0.58	0.65
Sorting and Classification	0.60	0.36	0.62	0.59	0.61	0.58
Shape Identification	0.44	0.29	0.53***	0.36	0.48	0.33
Number Identification	0.39	0.38	0.50***	0.27	0.36	0.47
Number Sense	0.59	0.40	0.67**	0.52	0.59	0.60
Simple Operations	0.73	0.34	0.78	0.67	0.70	0.81
Puzzle Completion	0.43	0.32	0.45	0.40	0.43	0.43
Socio-emotional Domain	**0.53**	**0.22**	**0.54**	**0.52**	**0.54**	**0.52**
Self-awareness	0.73	0.24	0.71	0.74	0.72	0.73
Friends	0.41	0.22	0.47***	0.34	0.40	0.44
Emotional Awareness	0.46	0.38	0.46	0.46	0.48	0.41
Empathy/Perspective Taking	0.51	0.37	0.54	0.47	0.51	0.50
Conflict Resolution	0.49	0.45	0.47	0.50	0.49	0.46
IDELA Overall	**0.56**	**0.16**	**0.58**	**0.54**	**0.56**	**0.57**
< the 50% scorers - %	**37.9**	**-**	**(SD=0.19) 34.7**	**(SD=0.13) 41.3**	**(SD= 0.16) 38.7**	**(SD= 0.16) 35.0**

Notes: [a] The number of boys and girls assessed in each of the subtasks the domains are not the same and varies between 48 and 53 for boys and between 41 and 51 for girls. *** p<0.01, ** p<0.05, * p<0.1

Another interesting analysis for intervention is examining the performance of children on the four domains across regions, but it is worth noting that the small sample sizes and variation across regions demand caution with the interpretation and use of these findings. Figure 1 depicts resulting figures in this regard. Children in the Somali region, on average, answered 59% of the overall IDELA questions correctly while children in Gambella and Benishangul-Gumz answered correctly 50% and 45% of the questions, respectively.

When analyzed by region, that the larger mean correct score (68%) is observed in the motor domain for the Somali region and the lower score (31%) is observed in the emergent literacy domain for the Benishangul-Gumuz region. On average, children in the Benishangul-Gumuz region were able to provide only 31% of correct answers to the questions presented in the emergent literacy domain. In both Gambella and Somali regions, higher mean correct scores were observed for motor domains, while a higher emergent numeracy score is observed for the Benishangul-Gumuz region.

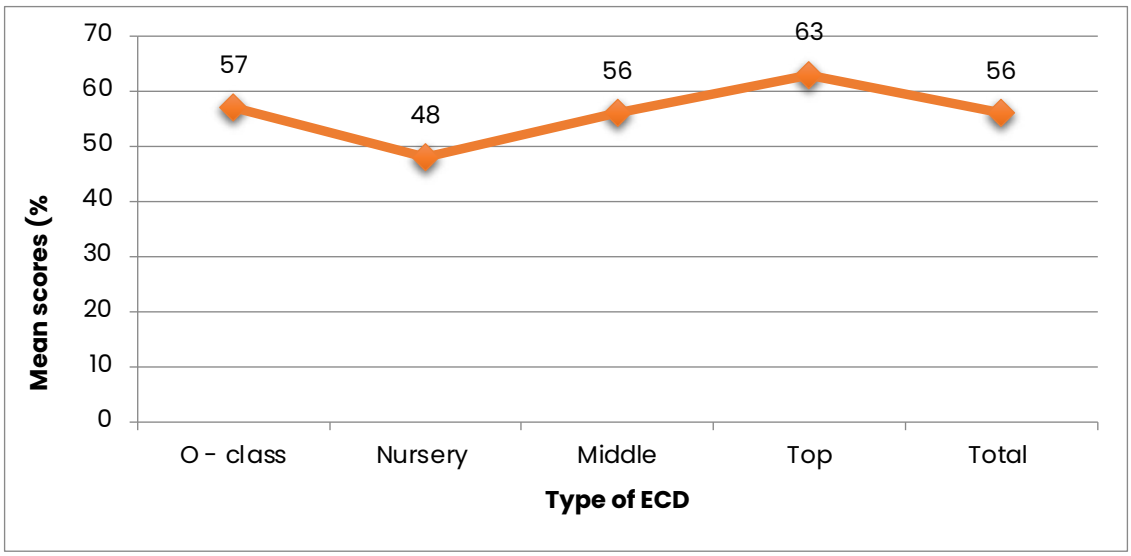
Figure 1. IDELA Domain Specific Mean Scores (%) by Region



Note: There is large variation in sample sizes among regions, and therefore these analyses are of exploratory nature. Benishangul-Gumuz = 9, Gambella = 15, and Somali = 72.

IDELA data was collected in both the host community and refugee camps. In host communities, data was collected in O-classes, formal centers attached to public primary schools serving mostly 6-year-old children, though younger children can also attend. In refugee camps, data was collected in ECD centers, which often have three levels of homogeneous age groups (Nursery = 4-year group, Middle KG = 5-year group, and Upper KG or Top = 6-year group). Thus, the analysis here is limited because, basically, refugee vs. host community comparison is same as comparing scores of ECD vs. 'O' class children. Figure 2 shows that the mean percent correct score of children of O-class is 57%, and the three levels of ECD are 48%, 56%, and 63%, respectively. We can also see that the performances of the three levels on IDELA increase as the levels move from nursery to upper KG or top level.

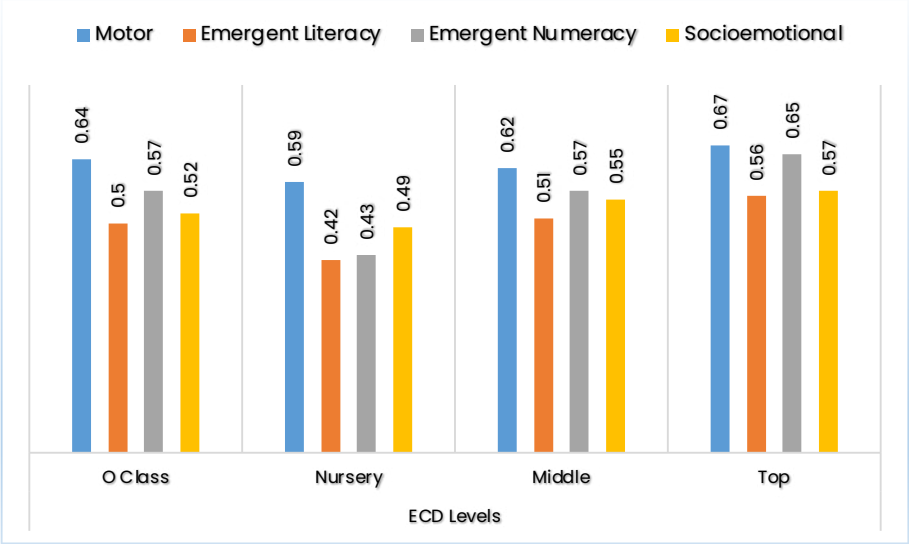
Figure 2. Descriptive Statistics of IDELA Total Score by ECD Level



Note: Sample sizes for different groups are very small, and therefore, these analyses are exploratory. O-class = 20, Nursery = 30, Middle = 18, Top = 28.

Further analysis of the data by subdomains (Figure 3) shows that a higher mean percent correct score is observed on the motor domain across the four groups. Emergent numeracy stands second for all groups except for the nursery where the socioemotional domain stands third. The mean percent correct scores on emergent literacy for O class, nursery, middle KG, and upper KG are 50%, 42%, 51%, 56%, respectively. Children in the nursery level performed lower, as expected, given their maturity in learning.

Figure 3. IDELA Domains Mean Percent Correct Score by ECD Levels



Note: There is large variation in sample sizes among regions, and therefore these analyses are of exploratory nature. Benishangul-Gumuz = 9, Gambella = 15, and Somali = 72.

Children's Wellbeing

Children's perceptions of physical and psychological well-being and their self-esteem may depend on a number of factors such as the nature of relationships with adult figures at home (a secure home environment, responsive care, and early learning opportunities), in the community, and in the schools, the presence of disabling situations including mental health, the characteristics of the social environment (socioeconomic conditions), physical and psychological abuse, emotional neglect, etc.

Table 5 summarizes overall findings on children's wellbeing, including disaggregation by sex, community type and region. The overall Kiddy-KINDL mean score of children is found to be 1.38 with a standard deviation of 0.34, suggesting that, on average, children rated their wellbeing favorably. The disaggregation of data by sex and regions, indicates that girls and those children in Somali rated their wellbeing the highest compared with their peers. However, the difference between girls (1.42, SD = 0.34) and boys (1.35, SD = 0.34) is not statistically significant. Similarly, the difference between host communities (M = 1.44, SD = 0.38) and refugee camps (M = 1.36, SD = 0.33), yielded no statistically significant difference.

Further analysis of the overall Kiddy-KINDL score revealed that a little over half of the children assessed (61.3%) rated their quality of life below the midpoint score while 38.7% of children rated at the midpoint or above. Only 11 (10.2%) children rated their quality of life as unfavorable. As Table 5 illustrates, 7.5% of girls as compared to 12.7% of boys scored below the midpoint on the Kiddy-KINDL scale. Although the difference appears to be negligible the trend shows that more girls seemed to rate themselves better than their counterparts. Similarly, more host children (as compared to refugee children) rated themselves below the midpoint of the scale on the overall Kiddy-KINDL score. The disaggregated data by region reveals that more children from the Somali region (11.0%) and BGR (10.0%) scored below the midpoint of the scale as compared to children from the Gambella region (6.7%). However, the interpretation of the results should take into account the large variation in sample size between regions (N = 15, 20, and 73, respectively).

Table 5. IKiddy KINDL Overall and Subscales Score by Sex, Community Type, and Region

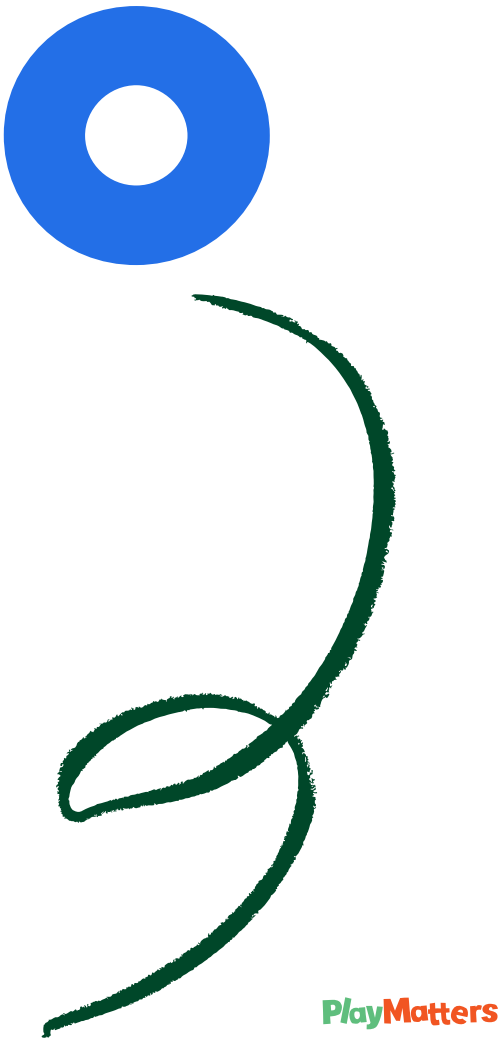
Kiddy KINDL Score	Overall Mean	Sex		Community Type		Region		
		Boys	Girls	Refugee	Host	BGR	G	S
Physical well-being	1.31	1.25	1.36	1.27	1.44	1.35	1.40	1.27
Self-esteem	1.27	1.22	1.33	1.26	1.33	1.38	0.83	1.34
Family	1.58	1.55	1.62	1.58	1.60	1.57	1.53	1.60
Friends	1.59	1.53	1.65	1.58	1.60	1.60	1.40	1.62
School	1.32	1.35	1.30	1.30	1.42	0.70	1.27	1.51
Mean	**1.38**	**1.35**	**1.42**	**1.36**	**1.44**	**1.29**	**1.25**	**1.43**
SD	0.34	0.34	0.34	0.33	0.38	0.24	0.28	0.37
Total		100.0	100.0	100.0	100.0	100.0	100.0	100.0
< the midpoint (1) -%	10.2	12.7	7.5	8.3	16.7	10.0	6.7	11.0
≥ the midpoint (1) -%	89.8	87.3	92.5	91.7	83.3	90.0	93.3	89.0

Note: In the Kiddy-KINDL tools, scores between 0 and 1 reflect unfavorable rating and scores between 1 and 2 reflect favorable rating of the health-related quality of life. The midpoint of the scale is 1.

Consistent with the total score, the Kiddy-KINDL subscales scores indicate that children rated their quality of life in all dimensions favorably, with children rating their relationship with family members and friends more favorably than in other dimensions. Relatively, a lower rating is observed in self-esteem, physical well-being, and daily functioning in preschools, respectively.

The disaggregation of data by regions shows that the Kiddy-KINDL mean score for the self-esteem subscale for children from Gambella is the lowest (0.83) which indicates an unfavorable rating for the group. A similar low rating (0.70) is observed among children from Benishangul-Gumuz on daily functioning in the preschool subscale. However, these results have to be viewed cautiously and should be used for descriptive purposes as there is a huge difference in sample size among regions where the sample sizes appear to be insufficient for Gambella and Benishangul-Gumuz. No regional difference has been observed on the Kiddy-KINDL mean scores obtained for physical well-being, family, and friends subscales. For example, the mean scores for the family subscale for Gambella, Benishangul-Gumz, and Somali are 1.53, 1.57, and 1.60, respectively.

The disaggregation of the subscales' mean scores by child sex yielded that girls scored higher on the friends subscale while boys scored higher on the family subscale. Although there is no significant gender difference, the results depicted in Table 5 indicate that girls rated their quality of life in all dimensions except their daily functioning in the preschool in a relatively better way than boys. Looking at the data by community type, the analysis shows that higher Kiddy-KINDL mean scores were observed in family and friends subscales in both groups. The mean score for self-esteem subscale is 1.33 for host community children and 1.26 for children from the refugee camps. No substantial difference has been observed between the two groups across all subscales.



Educator Data

Overall Results of Teacher Instructional Practices and Processes System (TIPPS)

The ratings of individual items included in the TIPPS made by the observers are presented in Table 6, providing total and item-level rating findings of classroom instruction in the sample schools. The analysis of each of these ratings allows us to give a better picture of the different classroom behaviors that contribute to the quality of the learning environment in the classroom. Using a mean TIPPS score ≥ 3 as an indicator of a quality classroom environment, the data shows that 82.61% of the 23 ECD classrooms observed meet the criteria for high-quality classrooms, while the remaining 17.39% need support to improve the classroom instructional quality.

On average, in 83.9% of the classrooms, the observers marked the presence of the behaviors and rated them at least somewhat accurate, while in 16.1% of the classrooms, the observers were unable or slightly able to detect classroom behaviors that contribute to the quality of the learning environment. It appears from the data that in 30.4% of the classrooms observed teachers seem to have difficulty using scaffolding to promote child learning and understanding of subject matter. Similarly, in 17.3% and 21.7% of the classroom teachers were not observed trying to connect activities and subject matter to key instructional concepts or learning objectives and to connect children's studies to their everyday life experiences, respectively. Besides, the observers observed the presence of behavioral indications of negative environment between the teacher and children and amongst peers in 60.8% of the classrooms observed. Showing favoritism towards some children was a problem observed in 56.5% of the classrooms observed.

No.	TIPPS ECD Items	VA - A		SA - A		VA - A		SA - A	
		N	%	N	%	N	%	N	%
1.	Teacher supports learners' development through the use of free playtime.	2	8.7	1	4.3	7	30.4	13	56.5
2.	Teacher structures activities in a way that learners learn to work, play and share with others.	1	4.3	3	13.0	4	17.4	15	65.2
3.	Teacher incorporates learners' ideas and interests to inform activities and assignments.	1	4.3	0	0	10	43.5	12	52.2
4.	Teacher encourages learners to use language to reason and problem solve.	1	4.3	1	4.3	6	26.1	15	65.2
5.	Teacher uses instructional materials to facilitate learning.	2	8.7	1	4.3	6	26.1	14	60.9
6.	Teacher connects activities and subject matter to a key instructional concept or learning objective.	1	4.3	3	13.0	8	34.8	11	47.8
7.	Teacher connects learners' studies to their everyday life experiences, showing the relevance of lessons outside the classroom.	3	13.0	2	8.7	6	26.1	12	52.2
8.	Teacher provides learners with specific feedback to facilitate learning rather than just getting the correct answer or finishing an activity.	1	4.3	0	0	7	30.4	15	65.2
9.	Teacher models quality language expression to advance learners' understanding and use of language.	0	0	0	0	6	26.1	17	73.9
10.	Teacher uses scaffolding to promote child learning and understanding of subject matter.	4	17.4	3	13.0	9	39.1	7	30.4
11.	There are behavioral indications of positive environment between teacher and learners and amongst peers.	-	-	1	4.3	8	34.8	14	60.9
12.	Teacher monitors and is responsive to learners' academic and emotional needs.	3	13.0	0	0	9	39.1	11	47.8
13.	There are behavioral indications of negative environment between the teacher and learners and amongst peers.	7	30.4	2	8.7	7	30.4	7	30.4
14.	Teacher tone of voice to encourages learners.	1	4.3	1	4.3	6	26.1	15	65.2
15.	Teacher actively employs gender responsive strategies	1	4.3	0	0	7	30.4	15	65.2
16.	Teacher actively employs responsive strategies for diverse learners.	1	4.3	1	4.3	7	30.4	14	60.9
17.	Teacher employs behavior management to create an environment that is conducive to learning.	2	8.7	2	8.7	6	26.1	13	56.5
18.	Teacher establishes classroom routines to create an environment that is conducive to learning.	1	4.3	1	4.3	5	21.7	16	69.6
19.	Learners are engaged in learning activities.	0	0	2	8.7	8	34.8	13	56.5
20.	Teacher shows favoritism towards some learners.	8	34.8	2	8.7	7	30.4	6	26.1
	Total	38	8.3	36	7.8	129	28.0	257	55.9

Note: Very Accurate – A (VA – A) represents the complete absence of the behavior observed; Somewhat Accurate – A (SA – A) represents the slight presence of the behavior; Somewhat Accurate – B (SA – B) represents noticeable presence of the behavior; and Very Accurate – B (VA – B) represents the complete and highly visible presence of the behavior observed.

Further analysis was made by disaggregating data into the region, sex of the educators, community type, and classroom type. The disaggregation of data by O-class and formal KG yielded identical results with community type since KGs and O classes are functional in refugee and host community settings, respectively. Data in Table 7 shows that the classrooms observed in the Somali region, on average, quite exceedingly reached the minimum quality indicator (Mean score ≥ 3.00) considered in this analysis while Gambella and Benishangul-Gumuz regions on average scored approximately at the minimum quality indicator showing that these regions need support for effective implementation of instructional practices and systems in the classroom. However, the interpretation of the mean scores obtained by Gambella and Benishangul-Gumuz should take into account the small numbers (3 and 4, respectively) of classrooms observed.

In terms of gender, the data shows that both female and male educators reached the minimum quality indicators of classroom instructional practices and systems. Nevertheless, females appeared to be better in their instructional practices and systems. The analysis by community type also reveals that classrooms observed in the host community looked better in terms of quality instructional practices and process systems. The three types of types of classrooms (open outdoors, covered with open sides, and covered with four walls), attained similar quality scores.

Table 7. Descriptive Statistics of TIPPS ECD Total Mean Score by Selected Variables

Variable	Level	N	Mean	N	Mean	Mean
Region	Gambella	3	2.93	0.465	2.40	3.25
	Benishangul-Gumuz	4	3.01	0.350	2.55	3.35
	Somali	16	3.48	0.282	2.70	3.85
Sex of Educators	Female	14	3.42	0.274	2.70	3.80
	Male	9	3.17	0.485	2.40	3.85
Community Type	Host	5	3.52	0.273	3.15	3.85
	Refugee	18	3.27	0.396	2.40	3.80
Classroom Type	Open Outdoor	7	3.37	0.364	2.70	3.85
	Covered with Open Sides	1	3.40	0.000	3.40	3.40
	Covered with Four Walls	15	3.30	0.412	2.40	3.80

Results of Educator Survey

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 socio-emotional skills. Onditi et al. (2018), for example, indicated in their study that many preschool teachers have positive perceptions of the importance of play-based activities in helping children to develop social, emotional, physical, and motor skills necessary in their life. The same authors specified that there are some teachers who hold the belief that play-based activities waste valuable time for academic work and believe that it is not very easy to integrate it into the teaching and learning process.

In this section, the results of the educators' survey explored their background, perceptions of regarding the importance of learning through play, engagement with developmental and educational activities through four constructs, occupational well-being, and self-efficacy are presented. Table 8 presents the summary of descriptive statistics on these measures.

A. Mean Perception Scores by Region, Sex, and Community Type

The analysis of the data yielded that the total mean perception score for educators who participated in the study is 3.28 with a standard deviation of 0.58. The minimum and maximum total mean perception scores are 1.75 and 4.0, respectively. About 10 (62.5%) educators scored above the mean and the rest (6, 37.5%) scored below the mean. The result shows that preschool educators generally perceived learning through play positively implying their readiness to practice if given the opportunity. The disaggregation of the data by region revealed that educators from Gambella region seemed to have lower perception scores than educators from the other two region. On the other hand, the disaggregation by sex and community type yielded similar results between groups.

B. Mean DEAS Scores by Region, Sex, and Community Type

Educators were also asked to complete a self-reported questionnaire known as the developmental and educational activities scale (DEAS). The mean total DEAS score is 2.45 (maximum is 4 points) with a standard deviation of 0.42. The total mean score shows that the overall rating of the educators' engagement in all developmental and educational activities measured by DEAS, on average, is at the middle of the measurement scale.

Items in the Play subscale assess the degree to which the educator makes suggestions that enrich children's play, provides children with opportunities for free and

Table 8. Descriptive Summary Statistics for the Results of Educators' Survey

Dimensions		Region			Sex		Community Type	
	Aggregate (N=16)	BGR (N=4)	Gambella (N=4)	Somali (N=8)	F (N=11)	M (N=5)	Refugee (N=14)	Host (N=2)
LtP Beliefs	3.28	3.40	2.65	3.53	3.20	3.43	3.26	3.42
	(0.58)	(0.04)	(0.71)	(0.43)	(0.65)	(0.39)	(0.61)	(0)*
DEAS	2.45	2.50	2.21	2.54	2.44	2.46	2.42	2.63
	(0.42)	(0.12)	(0.72)	(0.33)	(0.48)	(0.31)	(0.45)	(0.12)
Play	2.31	2.66	1.84	2.39	2.32	2.30	2.26	2.69
	(0.71)	(0.39)	(0.53)	(0.83)	(0.65)	(0.90)	(0.74)	(0.83)
Pretend Play	2.27	2.28	2.22	2.30	2.27	2.28	2.29	2.19
	(0.37)	(0.12)	(0.41)	(0.46)	(0.32)	(0.51)	(0.39)	(0.09)
Self-Regulation	2.38	2.48	2.18	2.43	2,34	2.47	2.34	2.59
	(0.56)	(0.26)	(1.01)	(0.43)	(0.66)	(0.27)	(0.59)	(0.45)
Emotional Support	2.84	2.59	2.59	3.09	2.86	2.80	2.81	3.06
	(0.68)	(0.48)	(1.10)	(0.48)	(0.70)	(0.70)	(0.72)	(0.27)
Educators' Occupational Wellbeing	3.09	3.06	3.17	3.07	3.19	2.87	3.0	3.72
	(0.70)	(0.83)	(0.93)	(0.61)	(0.75)	(0.56)	(0.70)	(0.08)
Vigor	2.90	2.83	2.92	2.92	2.91	2.87	2.79	3.67
	(0.89)	(0.79)	(1.10)	(0.96)	(0.98)	(0.77)	(0.90)	(0)
Dedication	3.31	3.58	3.25	3.21	3.27	3.4	3.24	3.83
	(0.90)	(0.50)	(0.88)	(1.11)	(1.08)	(0.28)	(0.94)	(0.24)
Absorption	3.06	2.75	3.33	3.0	3.39	2.33	2.98	3.67
	(0.87)	(1.29)	(0.90)	(0.68)	(0.66)	(0.88)	(0.90)	(0)
Teacher Self-efficacy	2.60	2.46	2.33	2.80	2.52	2.78	2.58	2.75
	(0.78)	(0.34)	(0.89)	(0.90)	(0.80)	(0.77)	(0.82)	(0.47)
Efficacy in Instructional Strategies	2.58	2.44	2.13	2.88	2.43	2.90	2.57	2.63
	(0.75)	(0.43)	(0.52)	(0.89)	(0.74)	(0.74)	(0.81)	(0.18)
Efficacy in Classroom Management	2.70	2.50	2.56	2.88	2.59	2.95	2.70	2.75
	(0.85)	(0.50)	(1.07)	(0.94)	(0.86)	(0.86)	(0.89)	(0.71)
Efficacy in Student Engagement	2.52	2.44	2.31	2.66	2.52	2.50	2.46	2.88
	(0.98)	(0.59)	(1.39)	(1.01)	(1.02)	(1.0)	(1.03)	(0.53)

self-managed play for instance by asking questions or providing materials for richer play. The analysis revealed that the DEAS mean play score for educators who participated in the study is 2.31 with a standard deviation of 0.71. The minimum and maximum DEAS mean play scores are 1 and 3.5, respectively. Half of the educators obtained the DEAS mean play score above the mean for the subscale and the other half scored below the mean. The DEAS mean play score shows that, on average, educators rated the degree to which they provide opportunities for children to engage in free and self-managed play as applicable. The disaggregated data depicted above shows that the educators in the Gambella region seemed to rate their roles during children's free and self-managed play, on average, as a little applicable.

The pretend play subscale measures the extent to which the educators stimulate cognitive distancing, symbolizing, and pretending in children by modeling behavior and encouraging children to participate in symbolic and pretend play. The analysis yielded that the DEAS mean pretend play score is 2.27 with a standard deviation of 0.37. The minimum and maximum DEAS mean pretend play scores are 1.5 and 2.88, respectively. Further analysis of the data yielded that 9 (56.3%) educators scored below the mean for the subscale. Generally, the result divulges that on average educators sometimes engaged in modeling behaviors and encouraging children to participate in symbolic and pretend play. This shows that educators seem to lack the skill to engage in pretend play while practicing LtP often or always. The disaggregation of data by region, sex, and community type yielded similar results across different groups with regard to educators' engagement in pretend play.

Self-regulation is an important skill teachers should support children to learn to develop to express and manage a range of feelings and get along with adults and peers. The roles educators play in this regard are fundamental. Items in this subscale assess the extent to which the educator uses play, care routines and other activities to enhance children's behavioral self-regulation, for instance talking about feelings and emotions, helping them resolve peer conflicts or

playing games in which children have to take turns. The analysis of the data uncovered that the DEAS mean self-regulation score is 2.38 with a standard deviation of 0.56. The minimum and maximum DEAS mean self-regulation scores are 0.73 and 3.18, respectively. Further analysis of the data yielded that 9 out of 16 (56.3%) educators scored below the mean for the subscale. The disaggregation of the data by region, sex, and community type yielded the following results. The above results show that on average educators seemed to support children to develop self-regulation on sometimes basis. This indicates educators' limited initiation to help children develop self-regulation skills when implementing teaching through play.

Emotional support in preschools represents intentional verbal and nonverbal care and affection that educators show to children. By providing emotional support to young children, educators may help children feel accepted, encouraged, and cared for, and make them feel important. The subscale contains items that assess the degree to which the educator provides emotional support to children and shows care and affection. The analysis of the data uncovered that the DEAS mean emotional support score is 2.84 with a standard deviation of 0.68. The minimum and maximum DEAS mean emotional support scores are 1 and 3.88, respectively. Further analysis of the data yielded that 10 (62.5%) educators out of 16 scored above the mean for the subscale. As a whole, the mean score for emotional support indicates that the emotional support provided by educators is close to often. The disaggregation of data presented in Table 15 shows that the educators in the Somali region seemed to rate the emotional support they provided to children, on average, as often. That is, educators in the Somali region provided emotional support largely often.

C. Educators' Occupational Well-being

Teachers' occupational well-being is a complex construct and can be conceptualized in different ways (Cumming, 2017). According to Rinne et al. (2022), the occupational well-being of educators can be defined as a balance between resources and workload factors as seen from four aspects of working life: (i) individual, (ii) working conditions, (iii) professional competence and (iv) work community. According to Schaufel et al., (2006) teachers' well-being is seen in terms of (a) vigor, (b) dedication, and (c) absorption.

The total Utrecht Work Engagement Scale (UWES) to measure wellbeing mean score is found to be 3.09 with a standard deviation of 0.70. The minimum and maximum Utrecht mean total scores are 1.89 and 4.0, respectively. About 9 (56.3%) educators have above the mean of the total Utrecht Work Engagement mean score. The mean score generally shows that on average preschool educators who participated in the baseline study have close to high occupational well-being. It appears from the data that preschool educators seemed to be better in their dedication well-being score.

Vigor, in this scale, refers to a high level of energy and mental resilience while working, the willingness to invest effort in one's work, and persistence in the face of difficulties. The mean UWES-Vigor score is 2.90 with a standard deviation of 0.89. The minimum and maximum UWES-Vigor mean scores are 1.33 and 4.0, respectively. About 13 (81.3%) educators out of 16 scored above the mean for the subscale indicating that most of the educators have shown the attributes of vigor with a close to high well-being score. The disaggregation of the data shows that the groups appear to be the same except two educators in the host community seemed to score high on this subscale.

Dedication refers to being involved in one's work, finding meaning in one's work, being challenged, and experiencing a sense of enthusiasm, inspiration, and pride. The mean UWES - Dedication score is 3.31 with a standard deviation of 0.90. The minimum and maximum UWES-Dedication mean scores are 1.33 and 4.0, respectively. Out of 16 preschool educators who participated in the study, 12 (75%) educators scored above the mean for the subscale specifying that three–quarters of educators are highly dedicated to their tasks. The disaggregation of data by region, sex, and community type yielded similar results across different groups. All groups seem to have high well-being mean scores on UWES – Dedication subscale.

Absorption refers to being fully concentrated and engrossed in one's work, whereby time passes quickly and one has difficulties detaching oneself from work. The mean UWES-Absorption score is 3.06 with a standard deviation of 0.87. The minimum and maximum UWES-Absorption mean scores are 1.33 and 4.0, respectively. Teaching is not only an activity conducted in the classroom. About 9 out of 16 (56.3%) educators, on average, scored above the UWES-Absorption mean score. On average, educators seemed to have worked intensely, immersed in their work, and carried away by their activities most often. Although it is difficult to make statistical tests with this sample size, the disaggregated data shows that educators from Benishangul-Gumuz and male educators seemed to have scored lower than their counterparts on the UWES-Absorption subscale.

D. Teacher Self-Efficacy Scale (TSES)

The total self-efficacy mean score is found to be 2.60 with a standard deviation of 0.78. The minimum and maximum self-efficacy mean total scores are 1.33 and 4.0, respectively. Only 7 out of 16 (43.7%) educators scored above the total self-efficacy mean score. The mean score generally shows that on average preschool educators who participated in the baseline study rated their self-efficacy at the middle of the scale (between applicable and strongly applicable). It appears from the data that preschool educators seemed to be better in their self-efficacy related to classroom management.

The mean efficacy score for instructional strategies is 2.58 with a standard deviation of 0.75. This subscale measures how well educators believe that they have the capacity to utilize various teaching strategies and assessment methods according to learner needs and level of understanding. The minimum and maximum mean efficacy scores for instructional strategies are 1.5 and 4.0, respectively. Further analysis of the data revealed that only 6 out of 16 educators (37.5%) scored above the mean of the subscale. This result shows that most of the educators seemed to perceive their capability to use various instructional and assessment strategies as applicable or below.

Educators' belief in their own capability to provide proper instruction has an effect on the way they deliver instruction and make assessments, thereby affecting student learning and learning outcomes. Although it is difficult to make statistical tests with this sample size, the disaggregated data shows that educators from the Somali region and male educators seemed to perceive their capability to use various instructional and assessment strategies in a better way than their counterparts.

Teacher self-efficacy is likely to play a significant role in the use of classroom management strategies that would create a safe and conducive learning environment and has implications for the quality of teaching. Items included in this subscale measure how well teachers manage classroom rules and student behavior. The analysis yielded that the mean efficacy score for classroom management is 2.70 with a standard deviation of 0.85. The minimum and maximum mean efficacy scores for classroom management are 1 and 4.0, respectively. Half of the educators scored above the mean of the subscale. This result may imply that 50% of the educators perceived their capability to control disruptive behavior and get children to follow classroom rules between applicable and strongly applicable. As Table 8 shows the disaggregation of data by region, sex, and community type yielded similar results across different groups.

Learning occurs when children engage in learning activities. Teachers' beliefs about their capability to make a difference in student learning have the potential to influence students' engagement and achievement. Educators were asked to rate their ability to engage children in the teaching-learning process. The study revealed that the mean efficacy score for student engagement is 2.52 with a standard deviation of 0.98. The minimum and maximum mean efficacy scores for student engagement are 1 and 4.0, respectively. About 7 out of 16 (43.7%) educators scored above the mean of the subscale.

Conclusions and Recommendations

Conclusions

Based on the findings of this study the following conclusions can be drawn:

I. Child Level

1. The overall IDELA mean score showed that on average children's likelihood to give correct answer was about 56%, showing that on average 44% of the items were not answered correctly, demonstrating overall low levels of performance and highlighting the need for early childhood development support across all areas. Basically, the means scores of the IDELA overall and each of the core domains ranged from 46-64%. Specifically:

a. We observe the highest levels of children' performance on motor skills (63%) followed by numeracy skills (55%), followed by SEL skills (53%), and the lowest for literacy skills (49%).

b. By Sex: Girls exhibit lower scores than boys on literacy, numeracy and SEL skills. IDELA overall mean score for girls fell at 54%, a little lower than that of boys (58%). While the difference between the two sexes was not statistically significant, the consistent pattern underscores the importance of including a gender approach to support the development of both girls and boys.

c. By refugee status: Children from host community and refugee settings performed comparable on both the overall and the domain specific tasks.

d. By region: Children from Somali achieved higher (59%) compared to those from Benishangul-Gumuz (45%) and Gambella (50%). Post-hoc analysis supported the significance of the regional differences in IDELA mean scores.

2. The overall Kiddy-KINDL mean score of children was found to be 1.38 (on a tree-point scale with a possible maximum of 2 points). This shows favorable view of children regarding their quality of life. The detailed indicators also confirm this result.

a. About 89.8% of the children assessed rated their quality of life above or equal to the midpoint of the scale (i.e., 1). Practically, this result indicates abundant occurrences of good life indicators in the life situation of majority of the children.

b. By sex: Though Kiddy-KINDL mean scores of girls (1.42) and that of boys (1.35) were different, the statistical significance analysis did not provide supportive evidence to consider sex as contributing factor.

c. By refugee status: The Kiddy-KINDL mean scores of children from the host community is slightly higher than children from refugee camps. However, displacement is not a factor that contributed to difference in scores as the difference between scores is not statistically significant.

d. By region: Children from the Somali Regional State (1.43) seemed to rating their quality of life favorably than their counter parts, though the difference is not statistically significant.

e. Subscale-based findings indicated that while girls scored higher on the friends subscale, boys scored higher on the family subscale. In all subscale other than daily functioning in preschools, girls provided relatively higher positive view than boys. On the other hand, children from the Benishangul-Gumuz region provided unfavorable view of their daily functioning in preschools while those from Gambella region rated their self-esteem unfavorably.

3. In conclusion, the mean values in the overall scores and subscales imply the need for intervention to improve the status of performances. There are variations in both overall and subscale performances too, whatever their level of statistical significance, had to do with sex, community type, and region. Besides, though no statistical significance variation as such, girls and children from host community and Somali region provided relatively higher quality of life ratings than their counter parts.

II. Educator Level

The conclusions at this level focused on four critical elements (motivation, skills, attitudes and behaviors) that have clear relation with the LtP theory of change. Based on the findings identified, therefore, the following conclusions look feasible:

1. At least eight out of 10 observed educators were viewed as having somewhat accurate competence or skill to ensure quality of the learning environment in their classrooms. The proportion of very accurate performance, however, was negligible. Besides, it appears from the data that in 30.4% of the classrooms observed teachers seem to have difficulty using scaffolding to promote child learning and understanding of subject matter. Similarly, in 17.3% and 21.7% of the classroom teachers having skill gaps to connect activities and subject matter to key instructional concepts or learning objectives and to connect children's studies to their everyday life experiences, respectively.

2. The perception of preschool educators toward LtP is positive. The preschool educators who participated in this study generally perceived learning through play favorably. This could be taken as one of the inputs from the part of preschool educators to implement learning through play in preschool settings.

3. Teachers' instructional skills are found to be up to the standard. The quality of the learning environment in the classrooms observed is quite high. Most of the indicators of quality classrooms are apparently visible during classroom observations. However, having good instructional skills did not imply capacity of using participatory approach such as LtP.

4. The study reveals that the occupational well-being of the preschool educators who participated in the baseline study is quite high, their engagement in all developmental and educational activities of children measured by DEAS is moderate, and their self-efficacy is moderate.

5. These are important educator characteristics that should be addressed during the intervention so as to base the implementation of LtP on a strong foundation.

Recommendations

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

6. Emphasize foundational skills: Although IDELA scores are almost average in all domains, the performance of children in the emergent literacy and numeracy domains is relatively low. If the project intends to target early development, the program should identify the sub-tasks where children scored low (e.g., letter identification and number identification) and focus on the learning activities that can foster foundational skills to build towards a holistic development of higher-order skills, such as numeracy and literacy.

7. Target and monitor differential experiences of girls and boys: Given the gender gap differences uncovered by this reporting, learning activities and skills should be carefully selected and organized to improve and enhance girls' performance and education experience. The different experiences and differential outcomes should continue to be monitored closely, particularly during future implementation activities. Teacher professional development activities should also be designed with the consideration and objective of closing gender gaps in both experience and outcomes.

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

9. Align research tools with PM's ToC and definitions: While the present study used a set of defined tools aligned with the initial design of the program, any future research should integrate measures that are aligned with PM 2.0 redesign and any other programmatic changes. Tools' alignment with PM's ToC will enhance the accuracy, relevance, and comprehensiveness of the research findings.

10. Conduct careful monitoring of educators' instructional skills through implementation research and M&E: Implementation research should integrate measures and indicators that directly assess the key aspects and outcomes of the implementation. These measures should align with the objectives, activities, and intended impact of the implementation. Future implementation research should also strive to be of mixed nature so that teachers' voices and experiences are not missed. Future data collection should also consider increasing the sample sizes for teachers and regions in order to be able to make disaggregations and other sub-group analyses.

11. Support educators' development of specific LtP activities: Equipping teachers with the necessary skills and strategies to create a student-centered classroom environment will require continuous monitoring and research. Particular attention should be placed on assessing changes in educator's' instructional skills and other in-classroom experiences. The intervention package for educators should also include a training package that focuses on improving particularly their skills that:

- Support and enhance children's free and self-managed play,
- Model the desired behavior and encouraging children to participate in symbolic and pretend play,
- Enhance the level of energy and mental resilience while working,
- Raise their ability to fully concentrate on their work,
- Increase the capacity to utilize various teaching strategies and assessment methods, and
- Engage children in learning activities.

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.

Alzahrani, M., Alharbi, M., & Alodwani, A. (2019). The Effect of Social-Emotional Competence on Learners Academic Achievement and Behavioral Development. International Education Studies, 12 (12), 141-149. https://doi.org/10.5539/ies.v12n12p141

Antoni, M. & Heineck, G. (2012). Do literacy and numeracy pay off? On the relationship between basic skills and earnings, *BERG Working Paper Series,* No. 86, ISBN 978-3-943153-00-2, Bamberg University, Bamberg Economic Research Group (BERG), Bamberg.

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

Cohen, L., Manion, L., & Morrison, K. (2008). *Research methods in education (6th ed.).* London & New York: Routledge Taylor & Francis Group

Realizing Capabilities in Ethiopia: Maximizing Early Childhood Investment for Impact and Equity. *Journal of Human Development and Capabilities,* 17 (4), 477–493, http://dx.doi.org/10.1080/19452829.2016.1225702

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.

French, G. (2013). Early literacy and numeracy matters. *From Journal of early childhood studies, OMEP,* Vol. 7, pp. 31-49.

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 learners' education and development in humanitarian and low-resource contexts.* PlayMatters.

Glass, P. (2002). Development of the visual system and implications for early intervention. *Infants and Young Children* 15: 1-10.

Hoot, J. L., Szente, J., & Mebratu, B. (2004). Early Education in Ethiopia: Progress and Prospects. *Early Childhood Education Journal,* 32 (1), 3-8.

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.

MoE, Ethiopia. (2011). *Education Statistics Annual Abstract 2003 E.C (2010/11).* Addis Ababa.

MoE, Ethiopia. (2021). *Education Statistics Annual Abstract 2013 E.C (2020/21).* Addis Ababa.

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

MoE, Ethiopia and USAID-AED/EQUIP II Project (2008). *Review of the Ethiopian Education & Training Policy and its Implementation.* Addis Ababa.

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

Rossiter, J., Hagos, B., Rose, P., Teferra, T., & Woldehanna, T. (2018). Early Learning in Ethiopia: Equitable access and Learning. System Diagnostic Report for World Bank Early Learning Program. https://doi.org/10.5281/zenodo.3371317

Rossiter, J. (2016). "Scaling up access to quality early education in Ethiopia: Guidance from international experience." Policy Paper 8. Young Lives: An International Study of Childhood Poverty.

Shomos, A. (2010). Links between Literacy and Numeracy Skills and Labour Market Outcomes, Productivity Commission Staff Working Paper, Melbourne, August.

Schaufeli, W. B., Bakker, A. B., & Salanova, M. (2006). The Measurement of Work Engagement with a Short Questionnaire. A Cross-National Study. *Educational and Psychological Measurement,* 66, 701-716. https://doi.org/10.1177/0013164405282471

Tirusew, T., & Hagos, B. (2016). *Assessment of the Status of O-Class in Four Regional States of Ethiopia.* Institute of Education, Health and Development & the World Bank Group, Addis Ababa, Ethiopia.

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 program 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. (2020). *Refugee Education Strategy: Towards Inclusion.* Addis Ababa, Ethiopia.

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

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

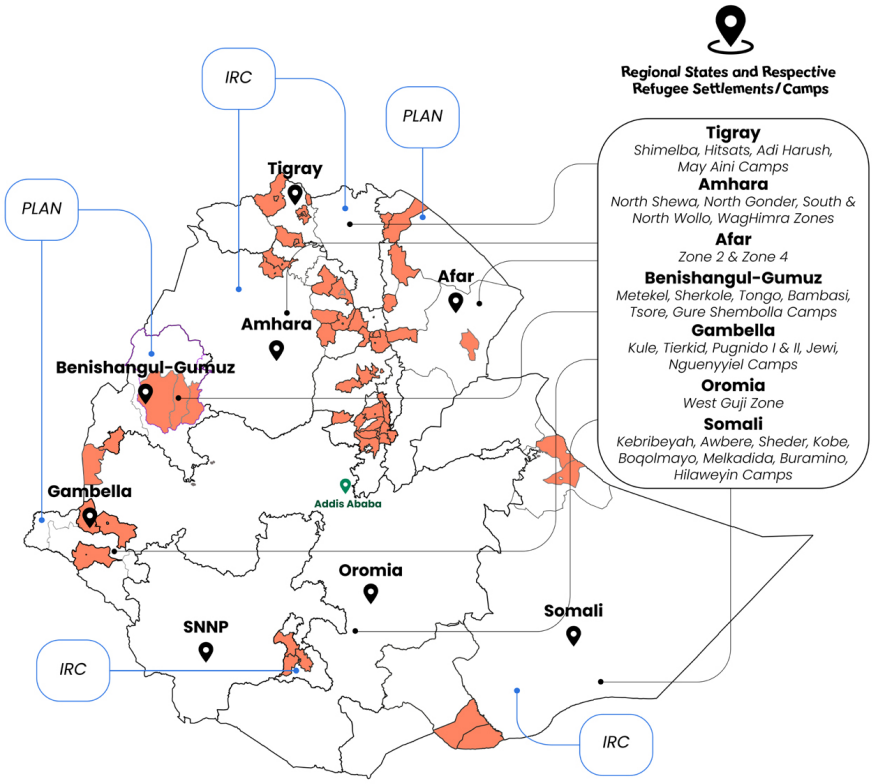
Vygotseky, I. (1978). *Mind in Society: The Development of Higher Mental Processes.* Cambridge: Harvard University press.

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).

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

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

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