Bridging Technology and Pedagogy in Education: Combining Technically Robust Offline-First Design with Formative Assessment for Early Literacy in Low-Resource Contexts

Edward Idun Amoah[1*]

[1] Nyansapo Labs Inc., State College, PA, USA
*Corresponding author: edward@nyansapolabs.org

Abstract

Background: Early-grade teachers in low-resource contexts often lack practical tools to conduct frequent formative assessment and translate results into actionable instruction. Mobile assessment platforms may address this gap, yet evidence on real-world adoption, usability for young learners, and scalable learning analytics remains limited, particularly for systems integrating offline-first engineering with differentiated-instruction design.

Objectives: This study evaluates an offline-first, mobile Foundational Literacy Platform (FLP) for early literacy in Kenyan primary schools, examining three research questions: (1) How do learners engage with mobile assessments, and what factors influence completion? (2) What teacher adoption patterns emerge, and what workflow barriers affect full-class coverage? (3) What learning analytics can the platform generate to inform differentiated instruction?

Methods: A seven-week field deployment was conducted across 20 schools in Kajiado and Uasin Gishu Counties, Kenya, encompassing 33 classes, 1,041 assessed students, and 2,428 teacher-created assessments yielding 4,926 completed student assessment results. Learners' usability was assessed via duration distributions and completion rates; teacher adoption via weekly active counts and efficiency ratios; coverage via class-level completion metrics; and learning analytics via letter-level mastery rates and between-school variation.

Results: Median assessment duration was 45 seconds with 98.2% completion, indicating minimal learner-facing friction. Active schools grew from 6 to 20 by Week 3, and weekly assessments increased 160-fold from 9 to over 1,400 at peak, demonstrating rapid teacher adoption. Teacher efficiency improved significantly over time ($\rho = 0.89$, $p = .007$). For the Letter Identification strand, 42.4% of classes achieved full-student coverage. Letter-level mastery analysis revealed systematic difficulty patterns (Y, D, N, E < 70% mastery; V, O, S, X, Z ≥ 85%) and substantial between-school variation (range: 44.4%–97.4%, $SD = 16.1\%$).

Conclusions: Offline-first mobile formative assessment can achieve high learner engagement and rapid teacher adoption while producing actionable learning analytics in low-resource settings. This study provides implementation evidence that offline-first mobile architecture paired with formative-assessment design can support classroom use and generate differentiated-instruction analytics at scale in LMIC contexts.

Keywords: formative assessment; mobile learning; early literacy; learning analytics; educational technology; Kenya; teacher adoption; differentiated instruction; offline-first architecture; LMIC

1. Introduction

1.1 The Challenge of Foundational Literacy Assessment

Foundational literacy, the ability to recognize letters, decode sounds, and read simple words, is a critical gateway skill that predicts long-term academic success (Shanahan & Lonigan, 2010; Gredler, 2002). More than half of all children in low- and middle-income countries do not learn to read with comprehension by age 10, despite the ambitions of Sustainable Development Goal 4 for inclusive and equitable quality education (World Bank, 2019; Jaime, 2023). In sub-Saharan Africa, where learning poverty rates exceed 80% (World Bank, 2019), early identification of struggling readers is essential for timely intervention. Yet teachers in these contexts face significant barriers to implementing practical formative assessment: large class sizes, limited instructional time, inadequate training in data-driven instruction, and a lack of tools for efficient evaluation and analysis (Dubeck & Gove, 2015; Piper et al., 2018).

Traditional approaches to early literacy assessment, such as the Early Grade Reading Assessment (EGRA), provide valuable diagnostic information but are resource-intensive, typically requiring trained assessors and one-on-one administration (A. K. Gove & Wetterberg, 2011). The Uwezo's annual measures have documented the extent of the learning crisis across Kenya and East Africa (Uwezo, 2016). Teachers need assessment tools that integrate seamlessly into daily instruction, provide immediate feedback, and scale across entire classrooms without specialized training (Black & Wiliam, 1998; Stiggins, 2005). Given the scale of the learning poverty challenge, resources within each country need to be directed to the most cost-effective approaches possible (Jaime, 2023).

1.2 Theoretical Foundations: Integrating Pedagogy and Technology

The pedagogical foundations for mobile-assisted learning were established through early explorations into the use of portable devices for educational purposes (Valk et al., 2010; Motlik, 2008; Amoah et al., 2022). Concurrently, the principle of differentiated instruction gained prominence, proposing that tailoring teaching to students' individual learning profiles yields superior educational outcomes (Tomlinson, 2014). A meta-analysis affirmed that technology-supported personalized learning approaches in LMICs, particularly those that integrate diagnostic and formative assessment into the instructional flow improves learning outcomes significantly (Major et al., 2021). The Global Education Evidence Advisory Panel, has identified structured pedagogy with linked materials and targeted instruction as 'Good Buys', interventions with strong evidence of cost-effectiveness for improving learning in low- and middle-income countries (Jaime, 2023).

A parallel stream of research has focused on the technical implementation of offline-first systems (Ijtihadie et al., 2010; Renz et al., 2017). However, a significant gap exists in the holistic integration of pedagogy and technology, as prior work tends to focus on either pedagogical strategies or offline technical solutions separately (Rodriguez-Segura, 2022). Despite promising evidence, two gaps persist. First, many studies evaluate pedagogical interventions without reporting system-level usability, adoption, and operational constraints under real classroom conditions. Second, technical studies of offline-first architectures often omit the instructional logic needed to support differentiated teaching decisions. This leaves limited evidence on whether integrated, offline-capable mobile formative assessment platforms can be adopted at scale while generating teacher-actionable analytics.

1.3 The Present Study

To address these gaps, this study evaluates an offline-first, mobile formative assessment system, the Foundational Literacy Platform (FLP), deployed under typical school conditions and examines student usability, teacher workflow adoption, and learning analytics outputs. Three research questions guide this investigation: RQ1 (Learners Usability), how do early-grade learners engage with mobile literacy assessments, and what factors influence assessment completion? This question is addressed by analyzing assessment duration distributions across competency levels and tracking weekly assessment completion rates to identify usability patterns and potential friction points. RQ2 (Teacher Adoption), what adoption patterns emerge among teachers implementing mobile assessment, and what workflow challenges affect assessment coverage of all learners in a classroom? This question is addressed by tracking weekly counts of active schools and classes, computing efficiency ratios (student assessment results per teacher assessment created), and calculating the proportion of classes achieving full student coverage along with the time and assessments required. RQ3 (Learning Analytics), What learning analytics can the FLP generate, and how do these inform differentiated instruction at the classroom and system levels? This question is addressed by computing letter-level mastery rates across all assessment events, and quantifying between-school variation using standard deviation.

2. Platform Design

2.1 Design Principles

The FLP was designed according to four principles informed by prior research on educational technology in developing contexts (Wagner, 2018; Warschauer & Ames, 2010) and offline-first architecture (Ijtihadie et al., 2010; Renz et al., 2017). (1) Minimal training requirements, teachers should be able to create and administer assessments within their first session, without specialized training. (2) Classroom-friendly workflow, assessments should fit naturally into existing literacy instruction blocks, requiring no more than one minute per student. (3) Offline capability, core functionality should operate without continuous internet connectivity, synchronizing data when connections are available. (4) Immediate, actionable insights, results should be presented in formats that directly inform instructional decisions.

These principles directly map to the study's research questions. RQ1 examines learner completion time and completion rates, addressing classroom-fit and usability (Principles 1-2). RQ2 investigates teacher adoption trends and assessment coverage metrics, evaluating minimal-training feasibility and workflow efficiency (Principles 1, 3). RQ3 analyzes letter-level and competency-level analytics to determine whether the system generates actionable instructional insight (Principle 4).

2.2 Technical Architecture

The FLP employs a modern, cloud-native architecture specifically designed to address the infrastructure constraints common in low and middle-income countries (Figure 1). The system comprises four integrated components: a cross-platform mobile application, a serverless cloud backend with Google Firebase Cloud Functions, AI services via Microsoft Foundry, and a web-based monitoring and management dashboard. Critically, the entire architecture is serverless, meaning operational costs scale with actual usage rather than requiring fixed infrastructure investments.

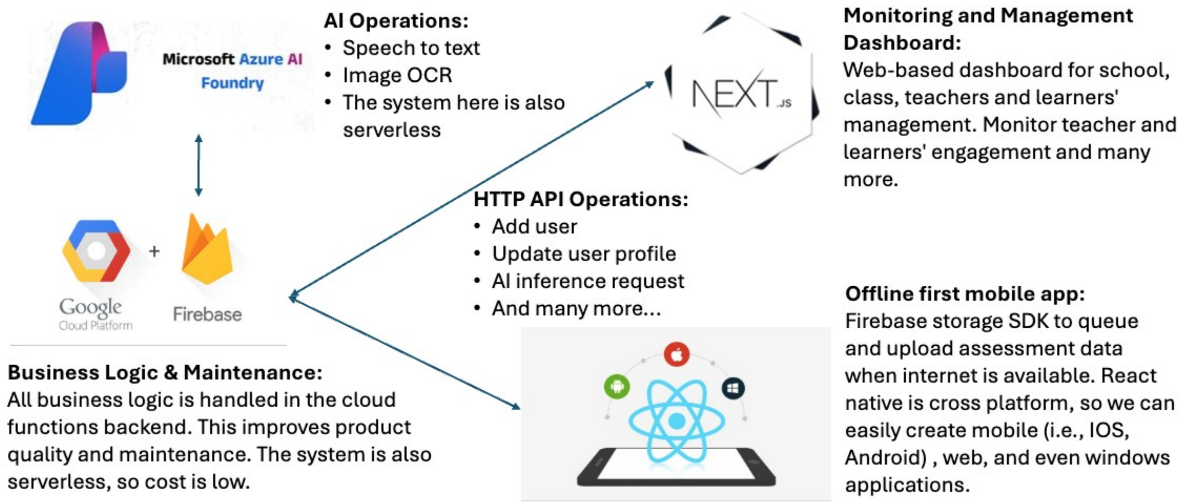

Figure 1. FLP technical architecture. The system comprises four integrated components: a cross-platform mobile application, a serverless cloud backend with Firebase Cloud Functions, AI services via Microsoft Foundry, and a web-based monitoring and management dashboard built with Next.js.

2.2.1 Mobile Application Layer

The mobile application was developed using React Native, a cross-platform framework that enables deployment across iOS, Android, web, and even Windows platforms from a single codebase. This architectural choice reduces development and maintenance costs while ensuring broad device compatibility, critical considerations for resource-constrained education systems where schools may possess heterogeneous device ecosystems. The application implements an offline-first design pattern: all core assessment functionality operates without network connectivity, with data stored locally on the device. When internet connectivity becomes available, the Firebase Storage SDK automatically queues and uploads assessment data to the cloud backend, ensuring no data loss even in environments with intermittent or unreliable network access.

2.2.2 Serverless Cloud Backend

The cloud infrastructure utilizes Google Firebase Cloud Functions, implementing a fully serverless architecture. The backend integrates three core Firebase services: Cloud Functions for business logic execution, Firebase Storage for media files (including audio recordings of learner's responses), and Firestore as the NoSQL database for storing assessment data, user profiles, and learning analytics. All business logic is centralized in cloud functions rather than distributed across client applications, improving product quality, security, and maintainability updates propagate instantly without requiring teachers to update their applications.

The serverless model operates on a pay-per-use pricing structure, where compute resources are consumed only when assessments are being processed, rather than requiring continuously running servers. This dramatically reduces operational costs compared to traditional server-based architectures, making the platform economically sustainable for education systems with limited technology budgets. During periods of low activity (nights, weekends, school holidays), infrastructure costs approach zero, while the system automatically scales to handle peak loads during active assessment periods.

2.2.3 AI-Powered Response Processing

Learners’ responses are processed using Microsoft Foundry AI services, which like the rest of the architecture operates on a serverless model. The platform supports two primary AI operations. (1) Speech-to-text transcription for verbal letter and sound responses, and optical character recognition (OCR) for written letter formation or tracing assessments. These AI capabilities enable automated scoring of student responses without requiring teachers to manually evaluate each response, reducing assessment friction and enabling immediate feedback. The AI inference requests are routed through the cloud functions backend, minimizing mobile device computational requirements and ensuring consistent accuracy across different device types. This design also enables continuous improvement of recognition accuracy through model updates without requiring client-side changes.

2.2.4 Monitoring and Management Dashboard

A web-based monitoring and management dashboard, built using Next.js, provides teachers and education administrators with comprehensive visibility into the education system. The dashboard supports school, class, teacher, and learner management functions, enabling administrators to monitor teacher and learner engagement across the deployment. For instructional purposes, the dashboard displays: (a) class-level competency distributions showing the proportion of students at each level (BE, AE, ME, EE); (b) letter-level mastery heatmaps identifying which specific letters students have mastered versus those requiring additional instruction; (c) individual student learning profiles tracking performance trajectories over time; and (d) actionable instructional recommendations suggesting which students need reinforcement, which are ready for advancement, and which specific letters require additional classroom focus.

2.3 Pedagogical Framework and Curriculum Alignment

The FLP's pedagogical design aligns directly with Kenya's Pre-Primary 1 (PP1) and Pre-Primary 2 (PP2) literacy curriculum, ensuring that assessments measure skills teachers are expected to develop. This section describes how assessment content, competency classifications, and instructional recommendations are structured to support differentiated teaching.

2.3.1 Literacy Strands and Assessment Structure

Assessments are organized around curriculum-aligned literacy strands specified in the Kenyan PP1/PP2 curriculum: letter identification (reciting letters A-Z), letter naming and letter-sound correspondence (matching letter names and letter sounds with letters), phonemic awareness (letter-sound knowledge in uppercase and lowercase letters, in PP1), and for PP2, blending three-letter words and familiar word reading.

Learner’s complete brief, game-like tasks, typically viewing a letter and providing a verbal or touch-based response, with the platform automatically recording responses and calculating accuracy using proprietary AI models fine-tuned on local dataset. Each assessment typically contains multiple letters across several attempts and requires a few minutes to complete, fitting naturally into existing literacy instruction blocks.

2.3.2 Competency-Based Assessment Rubrics

Assessment results are categorized using a two-stage competency algorithm aligned with curriculum expectations. In the first stage, the platform evaluates mastery at the letter level, a letter is classified as 'mastered' if the learners correctly responded on at least 70% of attempts for that letter within the assessment. In the second stage, the proportion of mastered letters determines the overall competency level. Exceeding Expectations (EE): 100% of assessed letters mastered and assessment completed in under 3 minutes, demonstrating both accuracy and fluency. Meeting Expectations (ME): At least 80% of assessed letters mastered, showing solid understanding with minor gaps. Approaching Expectations (AE): At least 50% but less than 80% of assessed letters mastered, indicating partial understanding. Below Expectations (BE): Less than 50% of assessed letters mastered, requiring intensive support.

2.3.3 Differentiated Instructional Activities

The platform provides teachers with competency-matched instructional recommendations drawn from a curated activity bank. Activities are mapped to specific competency levels and designed to use minimal or locally available materials, recognizing resource constraints in participating schools. Examples include, 'Letter Hunt Wall' (i.e., for BE/AE), display letter cards on the board, say a letter aloud, and have students find and point to it, strengthening letter recognition and retrieval. 'Letter Bingo', provide cards with random letters, call out letters one at a time, and students cover matching letters, building letter-name awareness through gaming. For AE/ME students, 'Swat the Letter', place letter cards on a table, call out a letter, and students race to swat it with a pointer, building reflexive letter-name associations. 'Letter Shape Posters', draw large letter shapes and discuss their visual features (curves, straight lines), asking guiding questions like 'Do you see any straight lines in this letter?' to build visualization skills. For ME/EE students, 'Match the Letters', lay out uppercase and lowercase letter cards in a jumble, and students match corresponding pairs (A with a), building visual discrimination. 'Find the Letter in Words', display short CVC or familiar words and ask students to identify target letters within words, supporting early decoding skills.

2.3.4 Teacher Dashboard Integration

The platform translates raw assessment data into actionable instructional guidance through the teacher dashboard (Figure 2C). For each competency group, the dashboard displays (1) the specific letters or sounds that students in that group are struggling with (color-coded red for struggling) and (2) the recommended instructional activities matched to the group's competency level. The dashboard also displays learner’s names within each group to facilitate targeted small-group instruction. This design embodies the principle of 'actionable intelligence', presenting analytics in formats that directly inform instructional decisions rather than requiring teachers to interpret raw data (Wise & Shaffer, 2015).

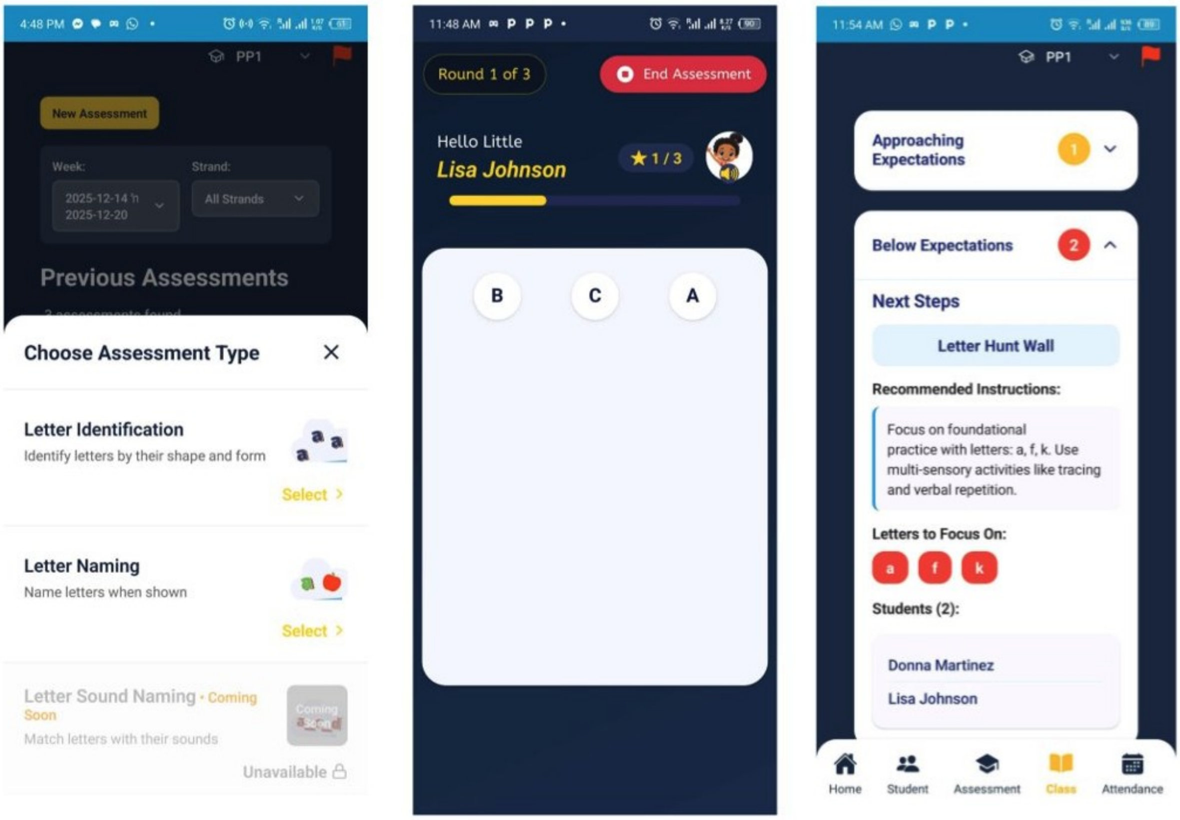

Figure 2. Curriculum-aligned formative assessments for differentiated instruction. The panel on the left shows the screen for creating new assessments. The middle panels show the learners assessment interaction for the letter identification strand. Here the learners are given audio prompts and respond by tapping on letters on the screen. The right most panel is the class strand detail screen, which provides teachers with information that directly supports differentiated instruction. Learners are grouped according to their competency levels, and the teachers are provided with recommended instructions and letters to focus on to support each group.

3. Methods

3.1 Participants and Setting

The study was conducted in Kajiado and Uasin Gishu Counties, Kenya, representing both peri-urban and rural contexts. Over the seven-week study period (September-October 2025), the FLP was deployed across 20 schools comprising 33 classes and 1,041 assessed students. Teachers received an initial in-person orientation session, with ongoing support provided through a WhatsApp group.

3.2 Data Collection

Data were collected automatically through the FLP's integrated logging system. The offline-first architecture enabled continuous data capture regardless of network availability: assessment data were stored locally on the device during administration and automatically synchronized to the cloud backend (Firestore database) when internet connectivity became available.

Three categories of data were captured: (1) Assessment metadata including timestamp, duration, student identifier, class, school, literacy strand, specific letters assessed, and completion status; (2) Response data comprising each item's target response (expected letter/sound), student response (captured via speech-to-text or touch input), response accuracy, and response time; and (3) Usage logs tracking teacher actions including assessment creation and student assignment. Learner’s identifiers were anonymized and no personally identifiable information was retained in the research dataset.

3.3 Statistica Analysis

Key measures used in the analysis are defined in Table 1. Analyses were conducted using Python (pandas, scipy, numpy). For each research question, appropriate statistical tests were selected based on data characteristics and assumptions.

RQ1 (Learners Usability): Assessment duration distributions were examined using descriptive statistics (median, percentiles). Durations were compared across competency levels using one-way ANOVA with Tukey HSD post-hoc tests. Levene's test assessed homogeneity of variances; ANOVA is robust to this violation with large sample sizes (Blanca Mena et al., 2017). Effect sizes were computed using eta-squared and Cohen's d. Completion rate trends across weeks were analyzed using Spearman correlation.

RQ2 (Teacher Adoption): Adoption growth trends were assessed using Spearman correlation. Teacher efficiency (ratio of student results to assessments created) trends were similarly analyzed using Spearman correlation. Coverage achievement rates for the Letter Identification strand were tested against a 50% benchmark using binomial tests.

RQ3 (Learning Analytics): Letter-level mastery patterns were analyzed using descriptive statistics, with letters categorized by difficulty thresholds (≥85% easy, 75–84% medium, 70–74% hard, <70% very hard). School-level letter mastery rates were computed using the same methodology to examine between-school variation. All tests used $\alpha = .05$ for significance.

Table 1. Definition of measures used in the analysis

Measure	Definition	Research Question
Assessment Duration	Time in seconds from assessment start to completion	RQ1
Completion Rate	Proportion of initiated assessments that were completed, calculated weekly	RQ1
Competency Level	Learners’ performance category: Beginning Expectations (BE), Approaching Expectations (AE), Meeting Expectations (ME), or Exceeding Expectations (EE)	RQ1
Active Schools	Number of schools with at least one assessment in a given week	RQ2
Active Classes	Number of classes with at least one assessment in a given week	RQ2
Assessment Volume	Number of learners assessment results per week	RQ2
Efficiency Ratio	Learners assessment results divided by assessments created; higher values indicate more students assessed per teacher action	RQ2
Full Coverage	Achievement of 100% learners’ assessment within a class for the Letter Identification strand	RQ2

Measure	Definition	Research Question
Letter Mastery Rate	Proportion of letters covered that were mastered (≥70% accuracy on that letter)	RQ3
School-Level Letter Mastery	Letter mastery rate aggregated by school using the same methodology	RQ3

4. Results

4.1 Descriptive Overview

Over the seven-week study period, teachers created 2,428 assessments administered to 1,041 students, yielding 5,048 student assessment results (Table 2).

Table 2. Summary statistics for the seven-week pilot study.

Metric	Value
Schools	20
Classes	33
Assessed students	1,041
Assessments created	2,428
Student assessment results	5,048
Completed assessments	4,957 (98.2%)
Median assessment duration	45 seconds
Mean letter mastery rate	78.2%

4.2 RQ1: Learners Usability and Engagement

4.2.1 Assessment Duration and Completion Rate Trends

The distribution of assessment durations was strongly right skewed, with a median of 45 seconds and 90th percentile of 2.1 minutes. Most assessments (88.7%) were completed in under two minutes (Figure 3A), indicating that the brief, game-like format was manageable for early-grade learners. Assessment duration varied significantly by competency level (Figure 3B). A one-way ANOVA revealed significant differences among groups, F (3, 4922) = 108.4, $p < .001$, $\eta^2 = 0.062$. Levene's test indicated unequal variances ($p < .001$); however, ANOVA is robust to this violation with large sample sizes (Blanca Mena et al., 2017). Post-hoc Tukey HSD tests revealed that all competency levels differed significantly from each other (all $p < .001$), with Meeting Expectations (ME) students showing the longest durations. The ME vs. Exceeding Expectations (EE) comparison yielded the largest effect ($d = 0.80$, medium effect).

Weekly completion rates improved dramatically after Week 1. Week 1 showed a completion rate of 33.3%, while Weeks 2–7 averaged 98.1% (range: 95.6%–99.7%) (Figure 3C). Spearman correlation confirmed a strong positive trend over time ($\rho = 0.93$, $p = .003$), suggesting that minimal orientation was sufficient for successful platform use.

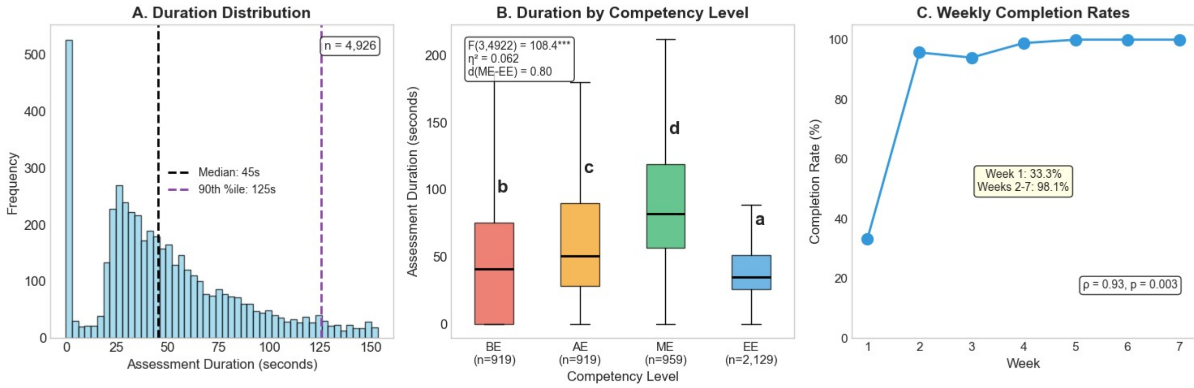

Figure 3. Learners Usability Metrics. (A) Distribution of assessment durations with median (45 seconds) and 90th percentile (125 seconds) indicated by dashed lines. Outliers removed using IQR method for visualization; n = 4,926 completed assessments. (B) Assessment duration by competency level. Box plots show median (black line), interquartile range (box), and whiskers (1.5 × IQR). Letters indicate Tukey HSD groupings; levels not sharing a letter differ significantly ($p < .05$). ANOVA statistics and Cohen's d for the largest pairwise comparison (ME vs. EE) shown. (C) Weekly completion rates across the seven-week pilot. Week 1 completion (33.3%) reflects initial orientation period; Weeks 2–7 averaged 98.1%. Spearman correlation indicates significant positive trend.

4.3 RQ2: Teacher Adoption and Workflow

4.3.1 Adoption Growth

Active schools increased from 6 in Week 1 to 20 by Week 3, remaining stable through Week 7. Active classes grew from 12 to 33 over the same period (Figure 3). Weekly student assessment volume increased dramatically from 9 assessments in Week 1 to 1,436 at peak (Week 5), representing a 160-fold increase. Spearman correlation confirmed a significant positive trend in assessment volume over Weeks 1–5 ($\rho = 1.00$, $p < .001$), before declining slightly in Weeks 6–7 as some schools began vacation periods.

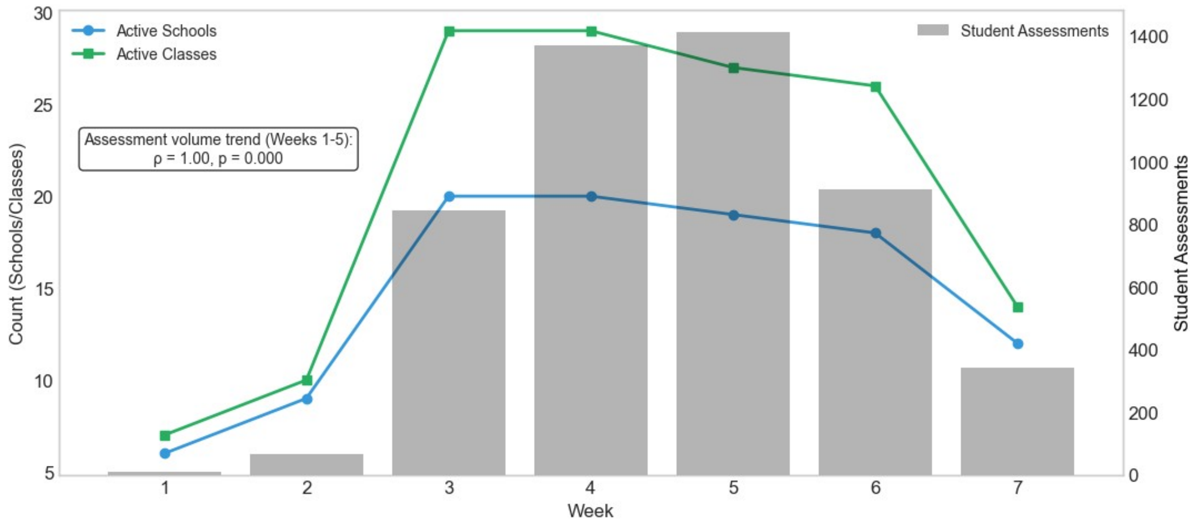

Figure 4. Weekly Platform Adoption Growth. Active schools (circles) and active classes (squares) shown on left axis; student assessment volume (gray) shown on right axis. Schools increased from 6 to 20 and classes from 12 to 33 over the first three weeks, then remained stable. Assessment volume grew from 9 in Week 1 to 1,436 at peak (Week 5), a 160-fold

increase. Spearman correlation calculated for Weeks 1–5 only; Weeks 6–7 showed reduced activity as some schools began vacation periods.

4.3.2 Efficiency Metrics

Teacher efficiency, measured as the ratio of student results to assessments created, improved progressively throughout the pilot (Figure 5). Efficiency increased from 0.25 in Week 1 to a peak of 4.53 in Week 6, before declining slightly in Week 7 as some schools began vacation periods. Spearman correlation confirmed a significant positive trend ($\rho = 0.89$, $p = .007$). The mean efficiency for Weeks 2–7 was 2.41, indicating teachers assessed approximately 2–3 students per assessment created, demonstrating effective adoption of the batch assessment workflow.

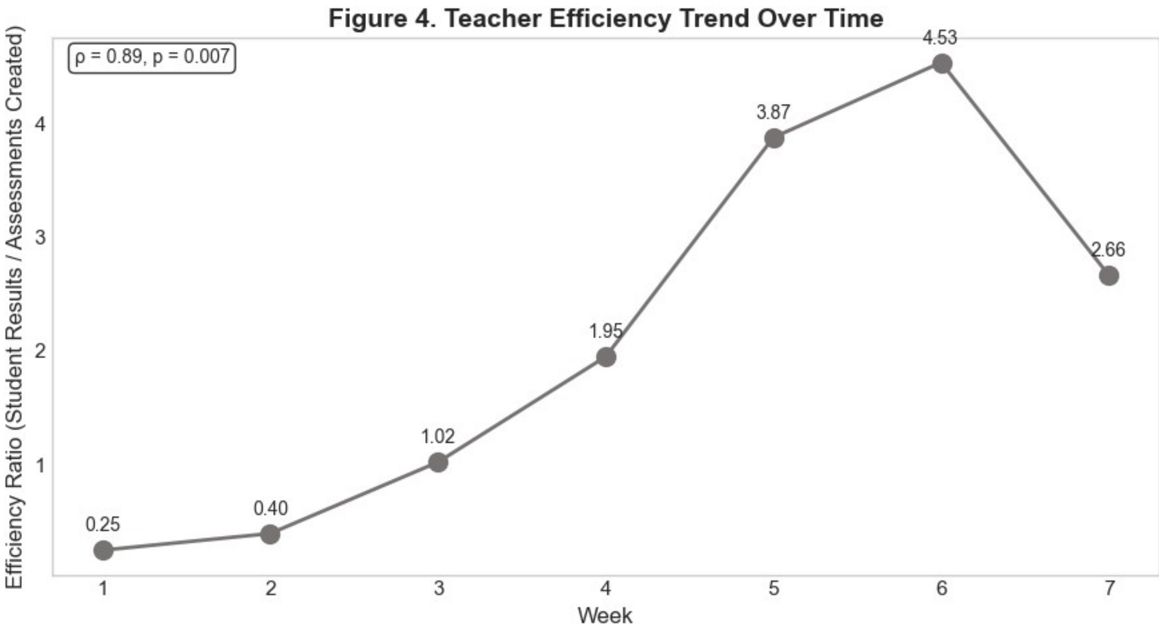

Figure 5. Teacher Efficiency Trend Over Time. Efficiency ratio calculated as student assessment results divided by assessments created. Efficiency increased progressively from 0.25 in Week 1 to a peak of 4.53 in Week 6, before declining slightly in Week 7 as some schools began vacation periods (Weeks 2–7 mean = 2.41). Spearman correlation confirms a significant positive trend ($\rho = 0.89$, $p = .007$), indicating teachers became increasingly proficient at batch-assessing students over time.

4.3.3 Coverage Metrics

Coverage metrics examined teachers' ability to assess all enrolled students for a class strand (e.g., Letter Identification). For the Letter Identification strand of the 33 classes, 14 (42.4%) achieved full coverage during the seven-week pilot (Figure 6). This rate did not differ significantly from 50% (binomial test, $p = .243$), indicating that a substantial proportion of teachers successfully assessed their entire class within the study period. Among classes achieving full coverage, the median time was 2.5 weeks, and the median number of assessments required was 14 (range: 2–99). The wide range reflects variation in class sizes and assessment strategies, with some teachers achieving coverage efficiently through 2–3 batch assessments while others used more inefficient iterative approaches.

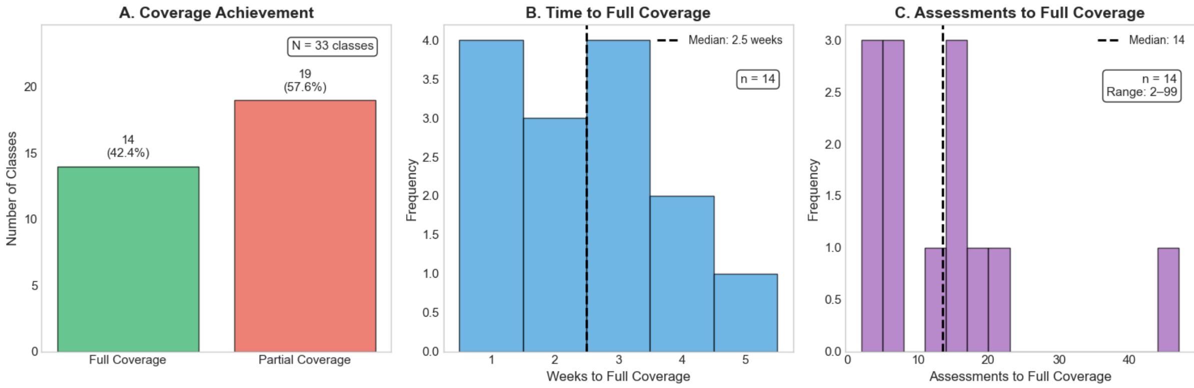

Figure 6. Class Coverage Analysis for Letter Identification. (A) Proportion of classes achieving full coverage (100% of enrolled students assessed) during the seven-week pilot. Of 33 classes, 14 (42.4%) achieved full coverage. (B) Distribution of time to achieve full coverage among successful classes (median = 2.5 weeks). (C) Distribution of assessments required to achieve full coverage (median = 14 assessments; range: 2–99). Dashed lines indicate medians.

4.4 RQ3: Learning Analytics

4.4.1 Letter-Level Mastery Patterns and Between-School Variation

The FLP generated detailed letter-level analytics revealing substantial variation in mastery rates (Figure 7A). Mean letter mastery was 78.2%, ranging from 50.0% (letter Y) to 100.0% (letters X and Z). Letters were categorized by difficulty: 5 letters were classified as easy (≥85% mastery), 11 as medium (75–84%), 6 as hard (70–74%), and 4 as very hard (<70%). The hardest letters were Y, D, N, and E, while the easiest included V, O, S, X, and Z. School-level letter mastery rates revealed substantial between-school variation (Figure 7B), ranging from 44.4% to 97.4% (M = 75.7%, SD = 16.1%). Seven schools achieved easy-level mastery (≥85%), 5 were classified as medium (75–84%), 1 as hard (70–74%), and 7 as very hard (<70%). This granular data enables targeted interventions at both letter and school levels, allowing education administrators to identify schools requiring additional support.

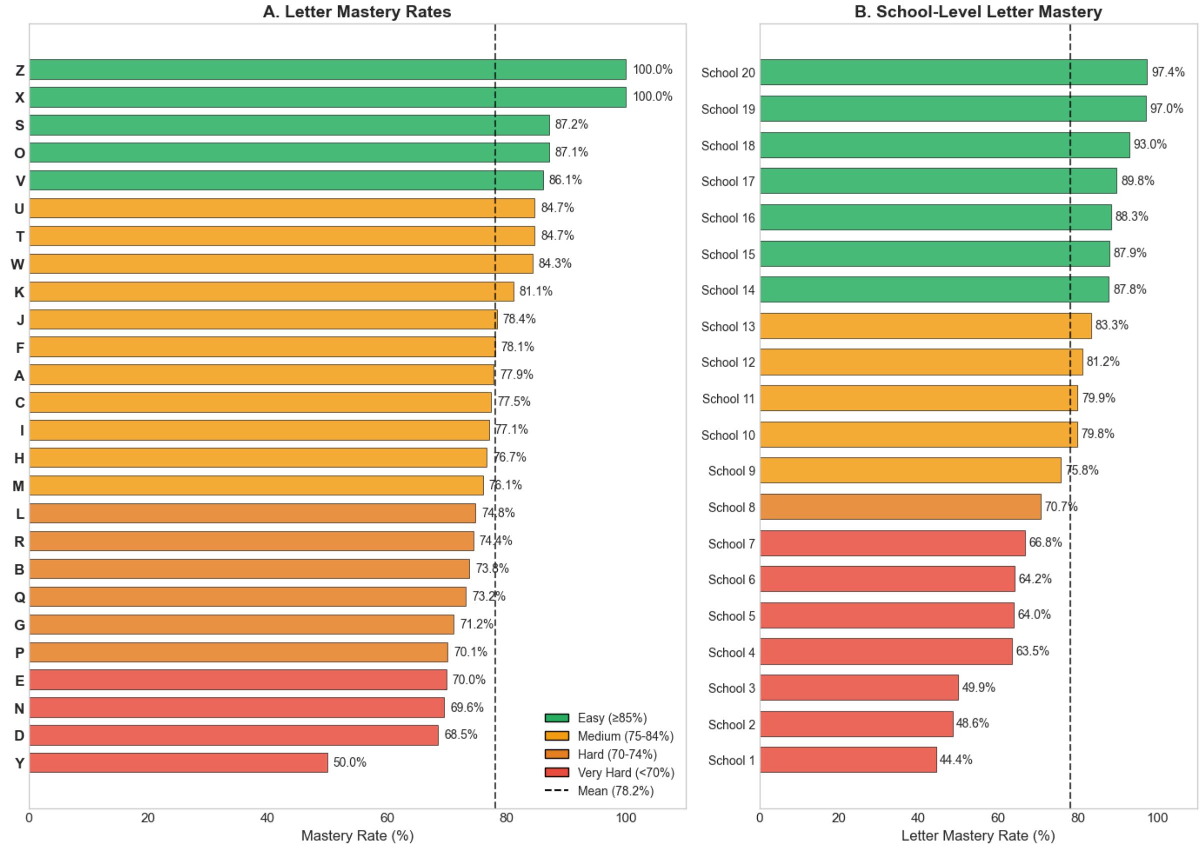

Figure 7. Letter-Level Mastery Patterns and School Variation. (A) Overall mastery rates for each letter, sorted from lowest to highest. A letter was considered mastered if the student achieved ≥70% accuracy on that letter. Bar colors indicate difficulty categories: green (≥85%, easy), yellow-orange (75–84%, medium), dark orange (70–74%, hard), and red (<70%, very hard). Dashed line indicates overall mean (78.2%). (B) School-level letter mastery rates using the same methodology, with schools anonymized and sorted from lowest to highest. Schools show substantial variation (range: 44.4%–97.4%), colored using the same difficulty thresholds.

5. Discussion

5.1 Student Engagement with Mobile Assessment

The findings provide strong evidence that mobile-based literacy assessments can achieve high engagement among early-grade learners in low-resource contexts. The 45-second median completion time compares favorably to traditional individual reading assessments such as EGRA, which require approximately 15 minutes per learner (A. Gove et al., 2023). The dramatic improvement in completion rates from Week 1 (33.3%) to subsequent weeks (98.1%), supported by a strong Spearman correlation ($\rho = 0.93$, $p = .003$), suggests that minimal orientation is sufficient for successful platform use. The game-like format, brief duration, and simple interaction patterns appear well-suited to the attention spans and technical familiarity of 4–6-year-old learners.

The statistically significant relationship between competency level and assessment duration ($F(3, 4922) = 108.4$, $p < .001$, $\eta^2 = 0.062$) offers insights into developing literacy skills. While the overall effect size was small, the ME-EE comparison yielded a medium effect ($d = 0.80$), indicating practical significance for this key distinction. Meeting Expectations (ME) learners took significantly longer than both lower-performing (BE, AE) and higher-performing (EE) students, suggesting that partially skilled learners experience greater cognitive processing demands than either low or high performers during assessments (Anderson, 1982; Sweller, 1988).

5.2 Teacher Adoption: Successes and Challenges

The rapid scaling of FLP adoption, from 6 to 20 schools and from 9 to over 1,400 weekly assessments within five weeks, demonstrates that teachers in low-resource contexts can quickly integrate mobile assessment tools into classroom practice. The perfect monotonic relationship in assessment volume growth ($\rho = 1.00$, $p < .001$ for Weeks 1–5) and the significant improvement in efficiency ratios over time ($\rho = 0.89$, $p = .007$) challenge assumptions that technology adoption in developing countries necessarily requires extended training (Warschauer & Ames, 2010). Teacher efficiency increased progressively from 0.25 students per assessment in Week 1 to a peak of 4.53 in Week 6, with a mean of 2.41 for Weeks 2–7, indicating teachers became increasingly proficient at batch-assessing learners.

However, the finding that only 42.4% of classes achieved full learner coverage for Letter Identification reveals workflow challenges that warrant attention. While this rate did not differ significantly from 50% (binomial test, $p = .243$), suggesting reasonable adoption, the wide range in assessments required (2–99, median = 14) indicates substantial variation in teacher assessment strategies. Some teachers achieved coverage efficiently through a few batch assessments, while others used more iterative approaches requiring many more assessment sessions. This finding indicates that workflow design and teacher guidance, rather than platform capability alone, determine coverage success.

5.3 Learning Analytics: Actionable Insights at Letter and School Levels

The FLP's learning analytics capabilities demonstrate the potential for mobile assessment platforms to generate instructionally relevant data at multiple levels of granularity. At the letter level, mastery rates ranged from 50.0% (letter Y) to 100.0% (letters X and Z), with a mean of 78.2%. The identification of specific difficult letters, Y, D, N, and E, all below 70% mastery, provides teachers with concrete targets for instructional focus.

Between-school variation was substantial, with school-level letter mastery rates ranging from 44.4% to 97.4% (*M* = 75.7%, *SD* = 16.1%). Seven schools achieved mastery rates above 85%, while seven schools fell below 70%. This variation, nearly 53 percentage points between the highest and lowest performing schools, underscores the importance of school-level monitoring and targeted support. Education administrators can use such data to identify schools requiring intervention and to investigate factors contributing to high performance at successful schools.

The granularity of these analytics distinguishes the FLP from traditional assessment approaches. Rather than providing aggregate scores that obscure specific learning gaps, the platform identifies precisely which letters each student has and has not mastered, enabling the differentiated instruction emphasized in the pedagogical framework. This aligns with research demonstrating that actionable, specific feedback is more likely to influence teacher practice than general performance summaries (Wise & Shaffer, 2015).

5.4 Bridging the Pedagogy-Technology Gap

This study contributes to addressing a significant gap in the literature: the holistic integration of pedagogy and technology for early literacy in LMICs. Prior research has tended to focus on either pedagogical strategies or offline technical solutions separately (Rodriguez-Segura, 2022). This study provides implementation evidence that an offline-first architecture can be paired with formative-assessment design to support classroom use while generating differentiated-instruction analytics at scale in LMIC contexts.

The pedagogical framework, with its curriculum-aligned competency rubrics and differentiated instructional activities, translated raw assessment data into actionable teaching guidance. Rather than presenting teachers with abstract percentages, the platform identified specific students and specific letters requiring attention, along with recommended activities matched to each competency level. This embodied the principle of "actionable intelligence" that educational technology research has identified as critical for teacher adoption (Wise & Shaffer, 2015).

5.5 The Platform as a Foundation for AI-Powered Foundational Literacy Platform

The present study demonstrates that the FLP can serve as a foundation for a more comprehensive AI-powered foundational literacy platform. The demonstrated capabilities, high student engagement, rapid teacher adoption, and actionable learning analytics, represent the essential data infrastructure required for adaptive, personalized learning systems. The letter-level mastery data, competency progression tracking, and between-school variation analytics provide the granular information necessary for AI-driven instructional recommendations and targeted intervention strategies.

These findings align with the Global Education Evidence Advisory Panel's identification of adaptive software that targets learning to the level of an individual child as a promising approach for improving learning outcomes (Jaime, 2023). The FLP's ability to generate real-time, item-level performance data positions it as the assessment backbone for such adaptive systems, enabling the kind of "teaching at the right level" that has been shown to dramatically improve learning outcomes in similar contexts (Banerjee et al., 2016; Muralidharan et al., 2019)

5.6 Future Directions: Toward a Unified AI-Powered Foundational Literacy and Numeracy Platform

Building on the evidence from this study, future development will integrate foundational numeracy assessment and instruction into the platform, creating a unified, low-cost, offline-first, AI-powered Foundational Literacy and Numeracy (AI-FLN) platform. Research on essential ingredients to literacy and numeracy improvement has demonstrated that similar instructional principles, structured pedagogy, formative assessment, and differentiated instruction, apply across both domains (Piper et al., 2018).

The potential impact of such platforms is substantial. Piper and colleagues showed that for every US$100 spent on programs combining professional development, coaching, books, and teachers' guides, an additional 14.7 learners were able to read at benchmark in English. A low-cost, technology-enabled platform that embeds these evidence-based ingredients could further improve cost-effectiveness by reducing per-learners costs while maintaining instructional quality.

Such a unified AI-FLN platform has significant potential beyond the Kenyan context. Foundational skills such as letter identification, letter naming, letter-sound correspondence, and basic number sense are universal developmental milestones that every child must acquire regardless of geographic or cultural context. The learning poverty crisis, with more than half of children in low- and middle-income countries unable to read with comprehension by age 10 (World Bank, 2019), is particularly acute in sub-Saharan Africa, where learning outcomes have remained low despite increased school enrollment. A low-cost, offline-first platform that generates actionable learning analytics could provide education systems across low- and middle-income countries with data-driven tools to identify struggling learners early and target instruction appropriately.

The platform's offline-first architecture and serverless cost model are particularly relevant for infrastructure-constrained environments common in sub-Saharan Africa and other developing regions. Technology interventions that do not account for implementation context often fail to deliver promised benefits (Jaime, 2023). By designing for intermittent connectivity and pay-per-use pricing from the outset, the AI-FLN ensures that core functionality remains available regardless of network conditions while keeping operational costs sustainable for education systems with limited technology budgets.

5.7 Limitations

Several limitations should be noted. First, this was an observational implementation study rather than a randomized controlled trial; thus, causal claims about learning gains cannot be made. Second, the seven-week duration may not capture longer-term sustainability or seasonal variations in platform use. Third, the study relied entirely on FLP-generated data; qualitative insights from teacher interviews or classroom observations were not included in the analysis. Fourth, the study focused primarily on Letter Identification, the strand with the most complete data, limiting generalizability to other literacy strands.

6. Conclusions

This study examined the implementation of an offline-first, mobile formative assessment platform for early literacy across 20 Kenyan primary schools. Three key findings emerged, each addressing a specific research question:

First, addressing RQ1 (Student Usability), mobile literacy assessments achieved remarkably high student engagement, with a 45-second median completion time and 98.2% completion rate, demonstrating that brief, game-like digital assessments are well-suited to early-grade learners in low-resource contexts.

Second, addressing RQ2 (Teacher Adoption), teacher adoption scaled rapidly, with weekly assessment volume increasing 160-fold within five weeks and teacher efficiency improving significantly over time ($\rho = 0.89$, $p = .007$). However, only 42.4% of classes achieved full student coverage for Letter Identification, indicating that teacher workflow optimization remains an area for improvement.

Third, addressing RQ3 (Learning Analytics), the FLP generated granular learning analytics with direct instructional relevance, including letter-level difficulty rankings (Y, D, N, E < 70%; V, O, S, X, Z ≥ 85%) and substantial between-school variation ($M = 75.7\%$, $SD = 16.1\%$, range: 44.4%–97.4%) that enables targeted intervention at both letter and school levels.

The FLP's technical architecture, featuring offline-first mobile design, serverless cloud infrastructure with pay-per-use pricing, and AI-powered response processing, proved well-suited to the infrastructure and budget constraints of LMIC education systems. The pedagogical framework, with curriculum-aligned competency rubrics and differentiated instructional recommendations, translated raw data into actionable teaching guidance.

Together, these findings demonstrate that offline-first mobile formative assessment platforms can serve as effective tools for data-driven literacy instruction in developing contexts, provided attention is paid to optimizing teacher workflows and translating analytics into actionable instructional guidance. The platform and the insights it generate establish a foundation for AI-powered foundational literacy and numeracy (AI-FLN) tools that can help address the learning poverty crisis affecting millions of children in sub-Saharan Africa and other low- and middle-income regions. Future randomized controlled trials should test whether these implementation successes translate to superior literacy and numeracy outcomes and examine the platform's effectiveness across diverse national contexts.

Acknowledgement

The author extends gratitude to the Board of Directors of Nyansapo Labs Inc. and the leadership of CcHUB for their continued support and guidance. This work was supported in part by the Sub-Saharan Africa EdTech Fund (SEF) Venture Studio at CcHUB, funded by the Bill & Melinda Gates Foundation, which supports AI-powered EdTech solutions for foundational literacy and numeracy in the region. Special thanks to Microsoft for Startups for providing cloud infrastructure support to Pixel Design Labs LLC, enabling the technical foundation of this work. The author also acknowledges the local Kenyan partners and educational experts whose contributions to curriculum alignment, pedagogical design, and field implementation made this research possible.

Declaration of Interest

The author declares the following potential conflicts of interest: President of Nyansapo Labs Inc., which developed the FLP mobile client, and Founder and CEO of Pixel Design Labs LLC, which operates the cloud infrastructure. The mobile client was co-developed with CcHUB, educational experts and local Kenyan partners.

Data Availability

The anonymized dataset supporting this study's findings is available from the corresponding author upon reasonable request.

References

Amoah, E., Mwangangi, M., Rastogi, T., Snow, E., Werner, L., & Gershenson, J. (2022). Technology Applications in Teaching at the Right Level Programs. *2022 IEEE Global Humanitarian Technology Conference (GHTC)*, 285–291. https://doi.org/10.1109/GHTC55712.2022.9910613

Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review*, *89*(4), 369–406. https://doi.org/10.1037/0033-295X.89.4.369

Banerjee, A., Banerji, R., Berry, J., Duflo, E., Kannan, H., Mukherji, S., Shotland, M., & Walton, M. (2016). *Mainstreaming an Effective Intervention: Evidence from Randomized Evaluations of "Teaching at the Right Level" in India* (No. w22746). National Bureau of Economic Research. https://doi.org/10.3386/w22746

Black, P., & Wiliam, D. (1998). Assessment and Classroom Learning. *Assessment in Education: Principles, Policy & Practice*, *5*(1), 7–74. https://doi.org/10.1080/0969595980050102

Blanca Mena, M. J., Alarcón Postigo, R., Arnau Gras, J., Bono Cabré, R., & Bendayan, R. (2017). *Non-normal data: Is ANOVA still a valid option?* https://hdl.handle.net/2445/122126

Dubeck, M. M., & Gove, A. (2015). The early grade reading assessment (EGRA): Its theoretical foundation, purpose, and limitations. *International Journal of Educational Development*, *40*, 315–322. https://doi.org/10.1016/j.ijedudev.2014.11.004

Gove, A., Harden, K., King, S., Ryan, J. P., Sou, S., & Edwards, S. (2023). 7 Literacy Development and Language of Instruction in Sub-Saharan Africa. *Global Variation in Literacy Development*, 133. https://books.google.com/books?hl=en&lr=&id=RU7XEAAAQBAJ&oi=fnd&pg=PA133&dq=RTI+International.+(2016).+Early+Grade+Reading+Assessment+(EGRA)+toolkit+(2nd+ed.).+USAID.+https://shared.rti.org/content/early-grade-reading-assessment-egra-toolkit-second-edition&ots=phs-LX3CMb&sig=XUw_g1gddGuED2-DkwHBoPoyivY

Gove, A. K., & Wetterberg, A. (2011). *The early grade reading assessment: Applications and interventions to improve basic literacy*. rti Press. https://books.google.com/books?hl=en&lr=&id=fvgTEAAAQBAJ&oi=fnd&pg=PP2&dq=Gove,+A.,+%26+Wetterberg,+A.+(Eds.).+(2011).+The+Early+Grade+Reading+Assessment:+Applications+and+interventions+to+improve+basic+literacy.+RTI+International.&ots=ZuAaXMMmZF&sig=_FVo-YRVaL-RgymXFXn4GMgA1VU

Gredler, G. R. (2002). Snow, CE, Burns, MS, & Griffin, P.(eds.)(1998). Preventing reading difficulties in young children. Washington, DC: National Academy Press, 432 pp., $35.95. *Psychology in the Schools*, *39*(3), 343–344. https://cir.nii.ac.jp/crid/1360016867492803840

Ijtihadie, R. M., Chisaki, Y., Usagawa, T., Cahyo, H. B., & Affandi, A. (2010). Offline web application and quiz synchronization for e-learning activity for mobile browser. *TENCON 2010 - 2010 IEEE Region 10 Conference*, 2402–2405. https://doi.org/10.1109/TENCON.2010.5685899

Jaime, S., Sylvia,Yoshikawa, Hirokazu,Akyeampong, Kwame,Andrabi, Tahir,Banerjee, Abhijit,Banerji, Rukmini,Dynarski, Susan,Glennerster, Rachel,Grantham-McGregor, Sally,Muralidharan, Karthik,Piper, Benjamin,Ruto, Sara,Saavedra. (2023). *2023 Cost-effective Approaches to Improve Global Learning: What does Recent Evidence Tell Us are Smart Buys for Improving Learning in Low- and Middle-income Countries?* [Text/HTML]. World Bank. https://documents.worldbank.org/en/publication/documents-reports/documentdetail/099420106132331608

Major, L., Francis, G. A., & Tsapali, M. (2021). The effectiveness of technology-supported personalised learning in low- and middle-income countries: A meta-analysis. *British Journal of Educational Technology*, *52*(5), 1935–1964. https://doi.org/10.1111/bjet.13116

Motlik, S. (2008). Mobile learning in developing nations. *The International Review of Research in Open and Distributed Learning*, *9*(2). http://www.irrodl.org/index.php/irrodl/article/download/564/1071

Muralidharan, K., Singh, A., & Ganimian, A. J. (2019). Disrupting Education? Experimental Evidence on Technology-Aided Instruction in India. *American Economic Review*, *109*(4), 1426–1460. https://doi.org/10.1257/aer.20171112

Piper, B., Simmons Zuilkowski, S., Dubeck, M., Jepkemei, E., & King, S. J. (2018). Identifying the essential ingredients to literacy and numeracy improvement: Teacher professional development and coaching, student textbooks, and structured teachers’ guides. *World Development*, *106*, 324–336. https://doi.org/10.1016/j.worlddev.2018.01.018

Renz, J., Shams, A., & Meinel, C. (2017). Offline-Enabled Web-based E-Learning for Improved User Experience in Africa. *2017 IEEE AFRICON*, 736–742. https://doi.org/10.1109/AFRCON.2017.8095574

Rodriguez-Segura, D. (2022). EdTech in developing countries: A review of the evidence. *The World Bank Research Observer*, *37*(2), 171–203. https://academic.oup.com/wbro/article-abstract/37/2/171/6333790

Shanahan, T., & Lonigan, C. (2010). The National Early Literacy Panel: A Summary of the Process and the Report. *Educational Researcher*, *39*, 279–285. https://doi.org/10.3102/0013189X10369172

Stiggins, R. (2005). From Formative Assessment to Assessment for Learning: A Path to Success in Standards-Based Schools. *Phi Delta Kappan*, *87*(4), 324–328. https://doi.org/10.1177/003172170508700414

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, *12*(2), 257–285. https://doi.org/10.1016/0364-0213(88)90023-7

Tomlinson, C. A. (2014). *The differentiated classroom: Responding to the needs of all learners*. Ascd. https://books.google.com/books?hl=en&lr=&id=CLigAwAAQBAJ&oi=fnd&pg=PP1&dq=Tomlinson,+C.+A.+(2014).+The+differentiated+classroom:+Responding+to+the+needs+of+all+learners+(2nd+ed.).+ASCD.&ots=AhporLeemr&sig=CI9YkP296v0HhOaYCRksTrFdq-I

Uwezo. (2016). *Are our children learning? Uwezo Kenya sixth learning assessment report*. Twaweza East Africa Nairobi.

Valk, J.-H., Rashid, A. T., & Elder, L. (2010). Using Mobile Phones to Improve Educational Outcomes: An Analysis of Evidence from Asia. *International Review of Research in*

Open and Distributed Learning, *11*(1), 117–140. https://doi.org/10.19173/irrodl.v11i1.794

Wagner, D. A. (2018). Technology for Education in Low-Income Countries: Supporting the UN Sustainable Development Goals. In I. A. Lubin (Ed.), *ICT-Supported Innovations in Small Countries and Developing Regions* (pp. 51–74). Springer International Publishing. https://doi.org/10.1007/978-3-319-67657-9_3

Warschauer, M., & Ames, M. (2010). Can one laptop per child save the world's poor? *Journal of International Affairs*, 33–51. https://www.jstor.org/stable/24385184?casa_token=VyuWtFMzj5AAAAAA:bUveajP_9-jdtjp9pXxMQL139yPz6LLmnu-6sFGLLlJl00k1sLzgAsVkIZ2UT0n272iO1YomRCHnqe4HOd-FPJH8MTaNYyfF0TQ5qxeX1j6NpdZwsA

Wise, A. F., & Shaffer, D. W. (2015). Why Theory Matters More than Ever in the Age of Big Data. *Journal of Learning Analytics*, *2*(2), 5–13. https://eric.ed.gov/?id=EJ1127063

World Bank. (2019, October 16). *Ending Learning Poverty: What Will It Take?* https://openknowledge.worldbank.org/entities/publication/af5e123e-27e9-5d15-b534-ec22ebe42d96