Daara
Development Academy

EVALUATION REPORT

Using Assessment for Learning to Enhance Numeracy Instruction

ebase effective basic services

Funda Wande
Reading for Meaning

TEP
THE EDUCATION PARTNERSHIP
CENTRE
...Global standards for education partnerships

Zizi Afrique Foundation

EXECUTIVE SUMMARY

This report details the findings from a Daara innovation grant pilot project funded by the Gates Foundation and undertaken by a consortium led by TEP Centre in Nigeria and also comprising Zizi Afrique Foundation (ZAF) in Kenya, Funda Wande in South Africa, and eBASE Africa in Cameroon. The overarching goal of this innovation is to enhance numeracy outcomes by empowering teachers to adopt and sustain new instructional practices, with a specific focus on leveraging "error analysis" in the teaching of basic numeracy. Recognizing the persistent challenge of teachers effectively adopting and maintaining new pedagogical approaches despite investments in training and capacity development, this first six-month phase centered on a collaborative effort to co-create and support teachers in implementing instructional practices that directly address common errors encountered by children in early grade numeracy. The consortium brings together organizations with a shared commitment to improving foundational learning outcomes across sub-Saharan Africa: ZAF, TEP, and Funda Wande offer significant experience in supporting teachers and learners with effective foundational literacy and numeracy (FLN) pedagogies. Complementing this practical expertise, eBASE Africa provides crucial research, monitoring, and evaluation capabilities, specifically focusing on the adoption of evidence-based pedagogical practices across the continent.

Collectively, the project endeavors to build upon prior FLN interventions by placing a strong emphasis on strengthening teacher capacity and ultimately enhancing learning outcomes in numeracy within the region. The initiative directly tackles the dual challenges of low numeracy outcomes, and the difficulties teachers face in sustaining new instructional practices without continuous support. By promoting the use of error analysis, underpinned by a comprehensive how-to guide for teachers, the project seeks to empower early grade mathematics educators to effectively integrate formative assessment data into their numeracy instruction, thereby aiming to achieve improved numeracy outcomes for learners.

This small-scale pilot intervention targeted Grade 2 and 3 learners, identified from previous projects sites. The project assessed learner performance with the ICAN (International Common Assessment of Numeracy), with some items from addition and subtraction administered. Specifically, the assessment measured place value, 2-digit addition and subtraction with regrouping. Data was collected from 10 teachers from Nigeria (8 from intervention schools and 2 from a control school), 9 teachers from South Africa (7 from intervention and 2 control schools), 8 teachers from Kenya (Kenya didn't have a control school). Only the interventions schools received training on practical application of error analysis and formative assessment. Additionally, learners in Grade 2 and 3 from 5 schools participated in the learners' assessments in each country, ensuring representation across different school environments. The intervention lasted 5 months in South Africa and 6 months in Nigeria and Kenya and the core innovation of this project lies in its emphasis on error analysis. Rather than treating learners' errors as mere mistakes, the project promotes a shift towards viewing them as valuable sources of information about learners thinking to inform teachers' practice. Teachers were equipped to:

- Systematically identify and analyze common errors that learners make in key numeracy concepts (e.g., 2-digit addition and subtraction, place value).
- Use error analysis to inform their instructional planning, allowing them to target specific misconceptions and learning gaps.

This approach is deeply intertwined with formative assessment. The project recognizes that effective teaching relies on a continuous cycle of gathering evidence of learners' learning (through error analysis and other formative assessment techniques), interpreting that evidence to understand learners' needs, and adjusting instruction to better meet those needs. Ultimately, this project aimed to empower teachers to become more effective numeracy instructors by providing them with the tools and knowledge to utilize error analysis and formative assessment to improve learner learning.

This study employed a mixed-methods design, combining quantitative and qualitative approaches to assess the feasibility and acceptability of implementing error analysis and formative assessment. It also examined the evidence that these practices improve numeracy performance, enhance learner engagement, and strengthen teacher instructional practices. To quantify these changes, we measured learning outcomes using baseline and endline assessments of numeracy

skills and engagement. Teacher compliance was assessed by calculating the percentage of predefined instructional criteria that were met during classroom observations.

A two-arm experimental design was piloted, with schools randomly assigned (except in Kenya) to the intervention group (trained in the practical implementation of error analysis and formative assessment) or the control group (no training). Qualitative data was collected to enrich the analysis through classroom observations, teacher interviews, and learner assessments. Specifically, observations examined the incorporation of error analysis and formative assessment into lessons and the level of learner engagement. Interviews investigated teachers' experiences, challenges, and adaptations. Learner assessments measured learner performance, contributing to an understanding of engagement and comprehension. By triangulating these findings, the study offered a comprehensive evaluation of the intervention's feasibility, acceptability, and evidence, detailing both its achievements and obstacles to implementation.

The project started in August 2024 and concluded in March 2025. Learners' numeracy skills improved significantly, indicating a strong likelihood that error analysis and formative assessment are powerful tools for enhancing instruction. The focus on analyzing errors suggests a high potential to help teachers better target their support and to increase learner engagement. Nevertheless, the study identified key challenges that need to be addressed to fully realize this potential: specifically, the need for more training on how to efficiently conduct error analysis within existing classroom time constraints, the need to support teachers to systematically use assessment data to inform their planning.

Table 1: Summary of pilot findings

Research question	Findings
Is the project feasible and executable?	There is strong evidence that **the project is both feasible and executable in real classroom settings**. Teachers from all three countries demonstrated not only a clear understanding of error analysis and formative assessment practices but also a **willingness to integrate these strategies consistently into their instructional routines**. Most interviewed teachers in Kenya reported applying error analysis in nearly every lesson. Similarly, in South Africa, all eight interviewed teachers stated that they use error analysis in every lesson. Despite challenges such as overcrowding and time constraints, teachers consistently affirmed the utility of the approach. One teacher called the training an *"eye-opener,"* emphasizing the importance of understanding the causes of learner errors, not just identifying them. In Nigeria, almost all teachers reported using error analysis daily by the endline, compared to 2 teachers at baseline. The shift from incidental correction to structured, diagnostic instruction reflected growing confidence and integration of the method. As one Nigerian teacher explained, *"Now, error analysis happens in every lesson… it's part of how I teach"*. Moreover, nearly all Nigerian teachers expressed a strong commitment to continuing with the approach long-term, describing it as essential for ensuring no learner is left behind. **Stakeholder ownership was also high.** School administrations supported peer collaboration and resource sharing. In sum, the **consistent uptake across diverse contexts, combined with high levels of teacher enthusiasm and adaptability, demonstrate that the intervention is not only feasible but potentially executable at scale**. However, scaling will require sustained support through targeted training, resource provision, and system-level integration. As one teacher in Nigeria stated, *"Error analysis isn't optional now it's how I ensure no child is left behind."*
Are the teachers' practices changing?	The intervention led to significant and **measurable changes in teacher practices across all three countries** Kenya, Nigeria, and South Africa with improvements most strongly associated with structured training, feedback mechanisms, and peer collaboration. In Nigeria, teacher compliance rose from 37% at baseline to 88% at endline, supported by targeted online coaching sessions and mobile-adaptable tools. One teacher

	reported: *"Now, error analysis happens in my math lesson it's part of how I teach"*, reflecting a move from incidental correction to structured diagnostic instruction. Another described how training helped them better support struggling learners: *"When I use the error log now, I can identify patterns. It helps me change how I teach a topic."* These shifts in daily classroom routines point to a deepening of formative assessment practices. In Kenya, compliance improved from 57% to 85%, and several teachers reflected on how the training helped them reconsider their instructional approach. One noted: *"It helped me plan my lessons differently… now I check more carefully for where learners are going wrong"*, while another highlighted: *"We come together, talk to one another, and find solutions to class issues"*, highlighting the value of collaborative learning through communities of practice. In South Africa, teacher compliance rose from 16% to 68%. One teacher described the practical impact: *"This program helped me understand better where learners are making mistakes."* Another said: *"Learners are now confident to explain their answers in front of the class"*, suggesting a shift toward more learner-centered, responsive teaching. Control group data reinforces that those changes in teacher practice were not the result of general trends or observational effects, but specific to the intervention. Several operational lessons emerge. First, teachers with lower baseline compliance showed the greatest improvements, particularly where support was consistent and reflective. Second, the intensity of implementation support appears to matter. Countries with structured coaching sessions such as Nigeria achieved stronger and more consistent gains. Third, differences in the range of endline compliance scores (tight clustering in Nigeria versus wider variation in South Africa) may point to differing levels of fidelity and contextual constraints such as teacher turnover or classroom size. Overall, the evidence demonstrates that **the intervention successfully shifted instructional practices, but also highlights the need for sustained, structured support to ensure consistency and deeper pedagogical change across all schools**.
Are learners' performances improving	The evaluation demonstrated **measurable improvements in learner performance across all three countries particularly in conceptual understanding**. Gains were evident in correct responses, reductions in error types, and increased learner engagement. In Kenya, correct responses among learners in intervention schools increased from 12% at baseline to 43% at endline, while factual errors declined from 79% to 63%. These improvements suggest that learners were better able to process and apply foundational numeracy concepts, likely due to more structured instructional feedback. Though overall gains were strong, some variation in skill mastery was observed, indicating that different numeracy domains may respond differently to the intervention and may benefit from differentiated instructional strategies. In Nigeria, learner performance showed similarly strong results, with correct responses improving from 26% to 55% and a reduction in conceptual errors from 13% to 1%. These outcomes demonstrate a significant improvement in understanding underlying mathematical concepts. However, procedural errors increased from 3% to 8%, suggesting that while the intervention strengthened conceptual knowledge, procedural fluency in areas like subtraction may require additional support or instructional time. In South Africa, outcomes were mixed. Factual errors decreased slightly from 48% to 43%, but correct responses also declined from 41% to 39%. This performance dip is attributed in the report to a Grade 2 teacher being transferred, which led to Grade 2 merging with Grade 3 and disrupted learning facilities. Despite these challenges, qualitative data indicated increased learner participation and confidence, as reported by the observation checklist and by teachers during interviews. Control group results show that the gains in the intervention schools were not mirrored outside the program.

	Overall, the findings confirm that the **intervention was particularly effective in enhancing conceptual numeracy, such as number sense and error recognition. However, persistent procedural gaps notably in subtraction highlight the need for additional instructional supports to complement formative assessment and error analysis**. These results affirm the intervention's impact on foundational numeracy and identify opportunities for strengthening its implementation across all assessed domains.
Are learners engaging and participating	The intervention resulted in significant **increases in learner engagement and participation, particularly in Kenya and Nigeria, with more moderate gains observed in South Africa**. These improvements were attributed to more responsive teaching practices, the use of formative assessment, and structured error analysis that made lessons more interactive and learner centered. In Kenya, learner engagement improved from 42% at baseline to 71% at endline demonstrating the likelihood of effectiveness of the intervention in motivating learners to participate actively in numeracy lessons. Teachers described noticeable shifts in classroom dynamics. One noted: *"Most of the learners now understand better... because I have changed the way I teach,"* and another explained: *"They raise their hands now to explain their answers, even the quiet ones."* In Nigeria, engagement also improved slightly, rising from 68% to 75%. Teachers attributed this shift to their improved ability to identify and address learning gaps in math lessons. One shared: *"Now I give time for learners to explain their errors. They love being asked what they think,"* and another added: *"Learners are more confident. Before they just sat quietly."* In South Africa, engagement increased modestly, from 56% to 64%. One stated: *"Learners are now confident to explain their answers in front of the class,"* and another mentioned that *"they participate more now because they know mistakes are allowed".* In contrast, control schools saw stagnation or decline in engagement, which reinforces that the observed gains in intervention schools were not due to general improvement but directly tied to the intervention's instructional components. Overall, **the intervention created more engaging, inclusive learning environments, with teachers across contexts reporting higher learner confidence, participation, and willingness to engage with mathematics content**. These results confirm that when teachers are supported to deliver more responsive, error-aware instruction, learners become more actively involved in their learning.

TABLE OF CONTENT

1. INTRODUCTION

This evaluation report details the findings of the pilot of a project funded by the Daara Innovation Fund. The Daara Innovation Fund was established to support collaborative projects among Daara partners that effectively address critical sector needs and contribute to the improvement of foundational learning outcomes in Africa. This particular project represents a collaborative effort between four key organizations in the region: TEP LEARNigeria Initiative (Nigeria), Zizi Afrique Foundation (Kenya), Funda Wande (South Africa), and eBASE Africa (Cameroon). The principal goal of this collaborative project is to enhance numeracy outcomes in early grades by empowering teachers to adopt and sustain new, evidence-based instructional practices. Specifically, the project focuses on the implementation of “error analysis” as a key strategy for improving the teaching of basic numeracy. Error analysis, in this context, involves a systematic examination of learners' errors to inform instructional decisions and address learning gaps more effectively. The project was designed to unfold in several phases. This report specifically focuses on the evaluation of the first phase of delivery. The primary objective of this initial phase was to explore and identify lesson plans and instructional practices that are most effective in addressing the common errors children make in early grade numeracy. This foundational work is intended to pave the way for subsequent phases that will involve broader implementation and scaling of these effective practices.

1.1 Background

The intervention leverages existing research and field experiences in using error analysis and formative assessment to enhance numeracy instruction in early grades. Previous studies by Funda Wande and its partners have shown that targeted teacher training combined with structured support improves teachers’ ability to identify and address learner errors, leading to better numeracy outcomes. However, past evaluations revealed a disconnect between training and actual classroom application, with teachers struggling to sustain new teaching methods without ongoing mentorship and practical feedback. This intervention was designed to bridge this gap by equipping teachers with practical tools to systematically observe, document, and analyze learner errors, adapt instructional strategies, and engage learners through interactive tools such as the throw circle, mind mapping, and paper money games to improve place value understanding.

The intervention is in the early stages of execution, moving from pilot testing to a more structured model that can be scaled within the context of Nigeria, Kenya, and South Africa. Developed as an eight-week structured program, it integrates consultative sessions, mobile-friendly resources, and continuous teacher check-ins. The materials and teaching strategies were adapted from existing Funda Wande resources and were customized to meet the needs of Grade 2 and Grade 3 numeracy instruction, focusing on two-digit addition and subtraction with and without regrouping. This approach is particularly relevant as error analysis in early-grade numeracy is still an emerging concept in the Nigerian education system.

This intervention adopts a structured approach that embeds evidence-based teaching strategies into daily classroom practice. By focusing on improving foundational numeracy, the project has the potential to be integrated into local teacher development programs and curriculum reforms. Teachers are now using techniques aimed at fostering behavior change such as tracking learner activities, documenting error patterns, and engaging learners in discussions to explore misconceptions. These practices are not only changing teachers’ perceptions of errors but also influencing how they design their lessons and provide feedback, fostering a data-driven approach to numeracy instruction.

The evaluation of this intervention aims to contribute to the understanding of its effectiveness in helping teachers systematically integrate error analysis into their instructional practices and its impact on learner performance. The study aims to assess whether teachers can sustain these new techniques beyond the intervention period, whether learners demonstrate measurable improvements in numeracy skills, and what challenges need to be addressed to refine the approach. The evaluation employs a mixed-methods design, combining learner assessments, classroom observations, and teacher feedback. This comprehensive approach allows for an in-depth understanding of how teachers are navigating challenges such as time constraints, limited teaching materials, and learner engagement issues. Additionally, it evaluates the effectiveness of mobile-friendly resources and community-based support systems in sustaining instructional changes. These insights will inform further adaptation and scaling of the intervention

to reach more classrooms and improve numeracy outcomes for vulnerable learners in Nigeria, Kenya, and South Africa.

1.2 Key Research Questions

Key Research Questions

- Is this intervention executable?
- How might error analysis influence numeracy instruction in early grade classrooms?

Learning Areas:

- What are we learning about teachers' understanding of error analysis?
- How are classroom instructional practices shifting (focus on lesson plans, learning experiences)?

1.3 Ethical Considerations

Ethical practices ensured transparency, respect, and the protection of participants' rights. Before data collection, enumerators were trained on key ethical protocols, including how to obtain informed consent, introduce themselves professionally, and clearly communicate that participation was voluntary. Stakeholders joined the study voluntarily and provided informed consent, with clear assurances that their data would remain confidential, anonymous, and that they could withdraw at any time without penalty.

1.4 Project Team

The implementation team, comprising Zizi Afrique, TEP LEARNigeria, and Funda Wande, delivered the intervention. They designed the training materials, facilitated teacher workshops, and provided ongoing mentorship to ensure effective adoption of error analysis and formative assessment techniques in classrooms.
The evaluation team, led by eBASE Africa, conducted the baseline, midline, and endline assessments to measure changes in teacher practices and learner outcomes. They were responsible ensuring the intervention's likelihood of impact was systematically documented.

1.5 Timeline

The timeline of activities related to the evaluation and intervention delivery, including recruitment period, data collection and delivery schedule is shown in the table below.

Table 2: Timeline

#	Output Description	Status (e.g., planned, in progress, complete, cancelled)	Target Completion Date*	Actual Completion Date
1	Implementation plan Project statement MEL plan Impact statement	Completed	July 2024	September 2024
2	Tools for need assessment are developed A comprehensive list of common errors and misconceptions in early grade numeracy to be treated are identified	Completed	August 2024	October 2024
3	Summarized mobile-friendly resources for teachers	Completed	August 2024	November 2024

4	Harmonized list of common errors in early grade numeracy across the three countries	Completed	Monthly	October 2024
5	Learning report, contextualized lesson plans and instructional practices	Completed	September 2024	December 2024
6	Lesson plans and instructional practices delivered	Completed	September - October 2024	December 2024
7	3 rounds of audit and feedback report	Ongoing	December 2024	February 2025
8	Project report on monitoring data	Completed (Combined for the 3 countries)	January - February 2025	March 2025
9	Work presented at the CIES. Among them were: i) Transformative teaching: Enhancing Mathematics Instruction in Sub-Saharan Africa: Localized, evidence-based approaches.	Completed	March	March
10	ZAF endline data collection and analysis	Completed	March 2025	March 2025
11	Endline report writing	Completed	March 2025	April 2025

The project was implemented concurrently in Nigeria, South Africa, and Kenya, with each country following a similar structure while adapting to local academic timelines, systems, and institutional arrangements. The project was led by the TEP Center team (Nigeria), which developed in collaboration with the consortium an adapted version of an existing Funda Wande guide and delivered mobile-friendly resources tailored to the national curriculum on implementation of error analysis.

In Nigeria, implementation began with the development of core documents and planning tools, followed by needs assessment activities and the identification of common numeracy misconceptions. This groundwork informed the design of contextualized teacher support materials. State-level engagement was critical to align with existing structures. A strong emphasis was placed on co-implementation through government systems. Baseline data was collected in four intervention and one control schools and prior to data collection, enumerators were trained on the data collection tools to ensure consistency and quality. Teacher training and resource distribution were completed October 2024, with instructional activities commencing shortly after. Teachers implemented lesson plans using mobile-friendly resources, and instructional coaching was conducted through embedded structures. Implementation fidelity varied across schools due to differences in institutional capacity, but state-level ownership was evident, particularly where focal points were assigned. Three rounds of audit data were collected and analyzed. Feedback sessions were planned to support continuous improvement, and endline data collection began in January 2025. The Nigeria process evaluation highlighted that while challenges such as teacher transfers and coordination delays existed, the use of real-time feedback mechanisms and collaboration with state actors enabled adaptive responses.

In South Africa, the project was led by the Funda Wande team, which developed training modules tailored to the national curriculum. Engagement with provincial and district education departments ensured alignment with ongoing initiatives. Baseline data was collected in four intervention schools and one control school. To ensure consistency and quality, enumerators received training on the data collection tools prior to data collection. Teacher training and resource distribution were completed in September 2024, and instructional activities began shortly thereafter. Implementation in schools began following teacher training in late 2024. Teachers were supported through instructional coaching and feedback sessions that emphasized reflective practices. The process evaluation emphasized the importance of structured learning loops, where implementation teams engaged in regular sessions to discuss what was working, what wasn't, and why. Adaptations were made in response to contextual factors, such as school timetables and language use in the classroom. The South Africa experience was marked by strong internal learning culture and institutional anchoring, which supported consistent delivery despite some operational challenges. Endline data collection and analysis were completed in March 2025, feeding into the final report writing.

In Kenya, implementation officially started in October 2024. The program started with training of data collectors and the data collection was done at baseline in 4 schools completely and just learners' assessment was done in the 5th school due to ongoing exams. There after activities included the delivery of teacher training, the distribution of contextualized learning materials, and support for instructional delivery through regular coaching sessions. Implementation took place in 5 selected schools, where teachers applied structured lesson plans focused on addressing common numeracy misconceptions. Teachers participated in reflective feedback sessions to review early challenges and successes. Like in the other countries, all data collectors were trained beforehand on the tools and protocols for data collection. Instructional practices were adjusted based on these feedback loops, allowing for greater contextual relevance. Implementation continued through late 2024, followed by data collection activities and review meetings in early 2025. While Kenya faced challenges related to timetabling and variations in instructional quality, adaptive responses from implementing teams helped sustain delivery. Kenya's component of monitoring data collection was completed by March 2025.

In all three countries, the project followed a cycle of design, implementation, feedback, and refinement. Process evaluations revealed that success depended not only on the delivery of materials but also on relationships, institutional capacity, and adaptability. Feedback loops, contextual responsiveness, and embedded government engagement were key enablers across the three contexts.

The data collection phase involved baseline, midline and endline in Nigeria and South Africa and baseline and endline in Kenya at different times, establishing a benchmark for measuring the intervention's impact. In total, data was gathered from 13 implementing schools (which received resources and training) and 2 control schools (which did not). After analysis, findings were shared with the teams to refine training materials before rolling out teacher training.

In Kenya teachers applied error analysis and formative assessment techniques over three weeks following a one-day in-person training. Post-implementation, a week of feedback and coaching was provided, though no mobile or digital tools were leveraged to support the process. Limited structured support hindered deeper adaptation of materials, though minor adjustments were attempted. After three weeks of classroom use, endline data was collected, and the intervention concluded with a teacher reflection session. Key learning: Without embedded support mechanisms or culturally responsive content, sustained fidelity in error analysis and formative practice proved difficult. There was no control school in Kenya as all the teachers from the selected schools received training on error analysis and formative assessment.

In Nigeria, baseline data was collected from intervention and control schools before rolling out formative assessment and error analysis training activities for teachers. Teachers received structured training, followed by a four-week implementation period. Midline data revealed initial difficulties in diagnosing learner misconceptions, prompting targeted online coaching sessions to strengthen teachers' interpretive skills. A collaborative review session further reinforced their ability to link errors to instructional adjustments. One teacher noted improved capacity to *"trace wrong answers back to learner thinking,"* demonstrating formative use of assessment data. Post-coaching, implementation resumed with higher teacher confidence, and endline data showed marked progress in utilizing classroom assessments to tailor instruction. Key implementation learning: Midline feedback loops and responsive coaching were critical in building teachers' ability to act on error analysis insights.

In South Africa, following baseline assessments and in-depth training, teachers implemented error analysis and formative strategies for four weeks. Midline data highlighted increased teacher attention to error patterns but also identified gaps in applying findings instructionally. Structured online coaching sessions were introduced, deepening teachers' ability to use assessment data diagnostically. For instance, one teacher *recognized the need to reteach a concept after observing recurring mistakes*. The subsequent phase saw stronger instructional adjustments informed by learner errors. Endline data and a final review confirmed improved responsiveness and engagement. Key learning: Midline interventions that reinforce the formative use of error analysis can enhance teaching precision and learner outcomes.

MEL data suggests that continuous feedback mechanisms, paired with localized content are pivotal for sustainability.

2. EVALUATION APPROACH AND METHODOLOGY SUMMARY

This project's evaluation strategy integrates qualitative and quantitative methods, designed to assess the feasibility, effectiveness, and scalability of error analysis as a pedagogical tool to improve foundational numeracy for Grade 2 and 3 learners in Nigeria, South Africa, and Kenya. The approach is tailored to address the unique educational contexts of this country while adhering to global best practices. Continuous monitoring to ensure fidelity of implementation and identify real-time challenges. Integration of feedback loops for iterative improvement of the intervention. Rigorous endline evaluation to measure the overall effectiveness, identify lessons learned, and provide evidence for policy recommendations.

2.1 Survey Design

The mixed-methods approach is exceptionally well-suited for the evaluation of the project, as it enables a comprehensive assessment of both measurable outcomes and the contextual dynamics driving these changes. This dual methodology evaluates the extent of changes in teacher behavior and learner performance while exploring underlying factors influencing these outcomes. Quantitative data provides objective metrics on teacher compliance, learner numeracy performance, and engagement levels, essential for tracking progress against predefined indicators. Complementing this, qualitative insights from teacher interviews, focus groups, and observations reveal practical challenges, successes, and perceptions of storytelling as a pedagogical tool. Together, these methods capture both the "what" and the "why" of the intervention's impact.

The approach also enhances validity through triangulation, cross-referencing quantitative data (e.g., learner assessments and attendance records) with qualitative insights (e.g., teacher feedback and focus group discussions). This ensures a robust, well-rounded analysis of the intervention's effectiveness. Additionally, the mixed methods design addresses inclusivity by using qualitative data to explore how the intervention accommodates learners with disabilities or marginalized groups, while quantitative assessments track performance trends across diverse demographics. The methodology identifies implementation gaps, such as disparities in learner outcomes or challenges in adapting stories for specific math topics, using qualitative insights to guide targeted interventions.

Furthermore, mixed methods facilitate iterative learning, with quantitative trends highlighting areas for immediate attention and qualitative findings informing strategic adjustments, such as enhancing teacher training or providing additional resources. Leveraging digital tools like Kobo Collect ensures efficient and consistent data collection, enabling real-time analysis and timely project adaptations. By combining the objectivity of quantitative data with the depth of qualitative insights, this approach provides a nuanced evaluation of the intervention's impact, ensuring actionable recommendations for scaling and sustaining foundational numeracy improvements.

Table 3: Key Components of the Evaluation Strategy

Mixed method methodology	Quantitative: ICAN test (standardized numeracy assessments aligned with the local curriculum to measure learning outcomes in intervention and control schools) Audit checklist to explore the number of criteria teachers can adhere to for their numeracy lesson
	Qualitative Methods: In-depth interviews, focus group discussions, and classroom observations to capture contextual factors influencing teaching and learning
Data phase	Baseline, Midline, and Endline Surveys: Data collected at three points to compare progress and identify changes in teacher practices, learner outcomes, and engagement.
Counterfactual evaluation	Two -Armed Study Design: Intervention Group: teachers were provided with training and mobile friendly resources Control Group: Teachers didn't receive training or mobile friendly resources
Difference-in-Difference analysis	Used to assess net impact by comparing changes in intervention groups against the control group over time

Focus on teacher training and compliance	Teachers receive professional development on integrating storytelling, direct explicit instruction, and formative assessment into their pedagogy. Regular feedback and coaching sessions, facilitated by local education officials and consortium members, ensure ongoing support and adaptation to identified challenges
Meta-synthesis of evaluation data	Comprehensive review and synthesis of baseline, midline, and endline data, including evaluation reports, learning assessments, and feedback from teachers and learners. Comparative analysis was conducted to identify changes in learner engagement, performance, and conceptual understanding of numeracy across intervention and control schools.

2.2 Sampling Techniques

The sampling strategy for the project was designed to ensure diversity and representativeness across teachers, classrooms, and learners in the targeted regions of Kenya, Nigeria and South Africa. This approach aimed to provide comprehensive data to inform interventions and monitor progress while capturing the nuanced dynamics of early-grade numeracy education. By employing a carefully structured design, the project sought to ensure inclusivity, equity, and actionable insights for enhancing foundational learning outcomes.

2.3 Selection of Schools

Schools were purposely selected to ensure familiarity with previous or existing education interventions, allowing for better contextual integration and sustainability. The project targeted four intervention schools and one control school in Nigeria and four intervention schools and one control school in South Africa and five schools in Kenya (no control school), prioritizing those with Grade 2 and 3 classrooms, as this stage is critical for developing foundational numeracy skills in addition and subtraction. Schools were evaluated based on their readiness to implement error analysis and formative assessment strategies. Assessments focused on identifying gaps in teachers' ability to diagnose learners' misconceptions, the availability of tools for practical feedback (such as diagnostic tasks or visual aids), and existing practices that could support responsive instruction.

2.4 Selection of Teachers

To ensure comprehensive insights into early-grade numeracy instruction, the study adopted a systematic sampling strategy: selecting one Grade 2 teacher and one Grade 3 teacher from each participating school across Kenya, South Africa, and Nigeria. This approach allowed for targeted implementation of error analysis and formative assessment techniques while enabling valuable cross-grade comparisons.

In Kenya, five schools were initially selected, with the goal of engaging ten teachers (one per grade level at each school). However, when baseline data collection coincided with examination periods at one school, the starting sample adjusted to eight teachers across four schools. The study's endline assessment arrived in a new academic year, bringing both continuity and change. While the intervention had initially engaged both Grade 2 and Grade 3 teachers, only Grade 3 teachers were assessed at endline, ensuring we tracked the same cohort of teachers and learners from baseline. Teacher transfers and school-year transitions reshaped participation: Kenya, which began with eight teachers, retained just two consistent Grade 3 teachers from baseline to endline due to transfers.

The South African cohort demonstrated challenges of longitudinal education research. Beginning with eight teachers across four schools (two teachers per school), the sample was reduced to seven when a Grade 2 teacher transferred mid-intervention without replacement. This necessitated merging the affected class with a Grade 3 section, yet the remaining teachers maintained rigorous implementation of the error analysis protocols. Two control teachers (one per grade level) provided essential comparative data throughout the study period.

Nigeria's implementation followed the planned structure most closely, with four intervention schools contributing eight teachers (one Grade 2 and one Grade 3 teacher per school), complemented by two control teachers (one per grade) from separate institutions. This clean

sampling frame allowed for particularly robust analysis of grade-level differences in pedagogical adoption.

Across all countries, 27 participating teachers provided informed consent after thorough orientation about the study's objectives, expected commitments, and professional development opportunities.

2.5 Selection of Learners

In Kenya the cohort began with 405 Grade 2 and 3 learners across five intervention schools at baseline. Due to the longitudinal design tracking the original Grade 3 cohort into the new academic year, the endline assessment captured data from 288 learners - representing a 28.9% decrease. The total number traced from baseline to endline was 130.

In South Africa, the study commenced with 386 learners (301 intervention, 85 control) across eight schools. Demonstrating strong retention, the endline assessment maintained 390 learners (297 intervention, 93 control) a slight increase of 1.0%. The control group numbers changed from 85 baseline to 93 endline.

In Nigeria the implementation started with 131 learners (101 intervention, 30 control) across four schools. By endline, the sample decreased to 91 learners (77 intervention, 14 control), representing a 30.5% attrition rate. This reduction was more pronounced in the control group (53.3% decrease versus 23.8% in intervention schools).

Across all three countries, the varying attrition patterns highlight both the challenges of multi-country educational research and the importance of context-specific retention strategies. While South Africa's near-perfect retention demonstrates what's possible with effective tracking systems, Kenya and Nigeria's experiences reflect the reality of learner mobility in many African education systems. These differences emphasize the need for research designs that anticipate and account for expected attrition while maintaining methodological rigor in cross-country comparisons. The consistent ethical framework - requiring institutional approvals, parental consent, and learner assent - proved adaptable across contexts, though its effectiveness in minimizing attrition varied by national implementation environment.

2.6 Data Collection Tools

The study employed a mixed-methods approach to ensure a comprehensive and rigorous evaluation of the project. This approach was chosen because it allows triangulation of data, enhancing the reliability of findings by combining quantitative performance metrics with qualitative insights from teachers and learners (Creswell & Plano Clark, 2018). A combination of structured observations, standardized assessments, and stakeholder interviews ensured that both objective learning outcomes and subjective classroom experiences were captured.

2.6.1 Observation Checklist for Learners

The observation checklist was developed to systematically document learners' engagement and participation during classroom instruction. Utilizing a binary coding system ("observed" or "not observed"), this instrument recorded qualitative data on learners' interactions with instructional content, learning tasks, and peer collaborations. To facilitate direct comparison between teaching approaches, the same checklist was implemented uniformly across both intervention and control schools. This standardized methodology enabled researchers to analyze variations in learner involvement patterns across different pedagogical environments, providing valuable insights into classroom dynamics under contrasting instructional conditions.

2.6.2 Audit Checklist for the Teacher

To evaluate teachers' implementation of error analysis best practices, we developed a 20-criteria audit checklist based on the JBI model. These criteria were organized into five critical domains: classroom environment, identification and documentation of errors, diagnosis of errors, utilization of error analysis for instruction, and teacher's content and pedagogical knowledge. The tool systematically measured how consistently teachers created a positive learning environment for error discussion, identified and recorded learner mistakes, diagnosed underlying misconceptions, and adapted instruction accordingly. We deployed this checklist across intervention and control schools, enabling structured comparison of how different approaches affected error analysis

practices in classrooms. Like the learner observation tool, this instrument provided measurable, standardized data on teaching quality while capturing differences in pedagogical implementation.

2.6.3 Learners Assessment

To assess learner performance, the ICAN (International Common Assessment of Numeracy) was adopted with some aspects of the ICAN from addition and subtraction administered. The test measured place value, 2-digit addition and subtraction with regrouping.

2.6.4 Learner Pulse Survey

Learners' attitudes toward numeracy were also captured through a Learners' Pulse Survey, which provided insights into their motivation, engagement, and classroom experiences. Capturing learner perceptions is essential, as research indicates that positive attitudes toward math are strong predictors of long-term numeracy success.

2.6.5 Key Informant Interviews (KIIs)

KII with teachers were conducted using a semi-structured format, allowing deeper exploration of the challenges and successes of integrating error analysis and formative assessment into numeracy instruction. KIIs enabled researchers to collect rich, qualitative data on teachers' experiences, which provided valuable context to the quantitative findings.

2.6.6 Error Log

It a is a systematic method for teachers to record and analyze learner mistakes during classroom evaluations or exercises. The primary purpose of the error log is to identify learning gaps, track learners progress, and inform targeted teaching strategies. Teachers document errors by noting their type, context, and frequency, often linking them to specific learners or identifying common trends across the class. This data enables evidence-based instruction, personalized support, and adjustments to the curriculum (Hattie, 2009).

2.7 Changes in the Evaluation design and Methodology

The evaluation framework maintained its core longitudinal design from baseline through endline, tracking the same cohort of teachers and learners across all three countries. While the research methodology remained consistent, we refined the data collection instruments (audits and observation checklist) after South Africa had collected baseline to better capture error analysis implementation. The teacher audit tool evolved to focus specifically on the five critical domains of error analysis practice, and the classroom observation checklist was adapted to measure learner engagement and participation. These targeted modifications allowed us to maintain methodological continuity while deepening our investigation of how teachers identify, diagnose, and utilize learner errors in mathematics instruction. All original participants from the baseline sample were retained where possible, though natural attrition occurred as expected in longitudinal studies, particularly in tracking the progression of Grade 3 learners into the new academic year. The consistent application of these tools across both intervention and control groups preserved our ability to compare teaching practices and learning outcomes.

Table 4: Change in the Evaluation design and Methodology

Data collection tool	Baseline	Midline	Endline
Standardized Math Assessment (ICAN)	No modification of the test at baseline.	No modification	No modification of the test in South Africa and Nigeria but in Kenya the test structure was maintained but the numbers changed
Audit checklist	Same for Nigeria and Kenya	Modification to include evidence-based criteria and then South Africa continued with the tool.	Same

Observation Checklist		In South Africa, the tool was revised following baseline data collection to improve validity and ensure consistency with instruments used in Kenya and Nigeria. As the initial baseline data were not fully comparable, the midline assessment conducted with the revised tool was designated as the analytic baseline for South Africa. However, as the midline was collected after the intervention had commenced, this limits the ability to fully isolate pre-intervention engagement levels in this context. Consequently, changes in learner engagement in South Africa may not reflect both initial uptake effects and subsequent intervention impact.	Same
KII	Same	Questions were modified to capture the challenges teachers are facing in implementation and provision of feedback based on challenges	Modification of questions based on the specific needs of teachers and tailored feedback
Error log		Same	Same
Learners pulse survey	Same	Same	Same

2.8 Development of the Theory of Change

The theory of change was developed create a clear pathway from inputs to outcomes. It began with a needs assessment identifying gaps in numeracy instruction, particularly in error analysis. The framework then mapped how teacher training and instructional resources would lead to improved classroom practices like structured error diagnosis and formative feedback. These activities were linked to measurable outputs, including increased teacher use of error analysis and higher learner engagement, ultimately leading to improved numeracy outcomes. This structured approach ensured alignment with national education goals while providing a robust monitoring and evaluation framework.

Table 5: Theory of change (inputs, activities, outputs, and outcomes)

Inputs	Activities	Outputs	Outcomes
Teacher training resources	Develop and pilot assessment tools	Tools for need assessment are developed	Immediate: Teachers master error analysis and formative assessment.
Research on common errors (e.g., TIMSS 2015)	Scoping common errors in early grade inventory	Identified common errors and misconceptions	Short-term: 50-80% of teachers improve formative assessment practices.
Funda Wande guide on error analysis	Review and adapt the Funda Wande's guide	Mobile-friendly, summarized error analysis guide for teachers	Immediate: Teachers adopt positive attitudes toward learner errors.
Monitoring and evaluation frameworks	Provide continuous support and M&E	Data on teachers' use of error analysis and learner engagement	Short-term: 70% of learners engage more actively in error discussions.
Lesson plans and instructional materials	Test treatment plans for addressing errors	Feedback-driven improvements to intervention strategies	Ultimate: Sustained integration of error analysis into teaching practices.

2.9 Data Collection Process and Mitigation of Bias

The data collection process was designed to ensure methodological consistency while minimizing potential biases. A team of trained enumerators, supervised by field coordinators, conducted assessments using standardized protocols to maintain consistency across all research sites. Prior to data collection, enumerators underwent comprehensive training on assessment tools, ethical guidelines, and bias mitigation strategies, including techniques to reduce observer effects and ensure uniform administration of surveys.

To address social desirability bias, participants were assured of confidentiality, and responses were anonymized before analysis. The study employed a controlled experimental design, with clear separation between intervention and control groups to prevent cross-contamination. Additionally, quantitative data were triangulated with qualitative insights from classroom observations and teacher interviews, providing a comprehensive understanding of outcomes while minimizing single-method biases. These measures ensured the integrity of the data collection process, supporting robust and credible findings.

Table 6: Data sources

Categories	Evaluation Questions	Data sources	Purpose of the data
Effectiveness	Is the project executable/feasible	Need assessment report KII from teachers and FGD	Assess intervention acceptability and feasibility Identify implementation challenges and enablers
Impact	How has the intervention influenced teachers' instructional practices and attitudes?	Audit checklists Classroom observations	Assess the adoption and integration of error analysis techniques Evaluate shifts in

		Coaching session reports	teacher behaviors and attitudes
Impact	What changes in learner, Performance engagement, participation, and critical thinking are observable?	Classroom observation	Evaluate improvements in learner participation and engagement
Efficiency	To what extend does the project take into consideration inclusive and accessible for children with disabilities?	Classroom observations	Identify disparities in learning outcomes - Assess the inclusivity of storytelling materials and methods
Sustainability	What evidence supports the scalability and sustainability of the intervention?	Endline interviews Progress report analysis	Gather institutional support Determine resource requirements for expansion Identify factors critical to scaling the intervention

We use quantitative data from the project surveys as well as the project datasets and outcome spreadsheets to assess the magnitude of change in education outcomes and educational barriers that projects across the schools have achieved since baseline and midline. We therefore draw on the analysis presented in the projects' endline evaluation reports to unpack how changes have come about and how effective different activities have been at creating these changes.

3. FINDINGS

The pilot was conducted across 14 implementing schools, 4 in Nigeria, 4 in South Africa and 5in Kenya, that received training and mobile friendly resources, and 2 control schools (1 in Nigeria and 1 in South Africa) that did not receive these interventions. Below is a summary of the characteristics of the participating school

Table 7: Study Participants

Grade 2	**Teacher**									
	Baseline					**Endline**				
Variable	**Intervention**		**Control**		**Total**	**Intervention**		**Control**		**Totals**
Gender	**Female**	**Male**	**Female**	**Male**		**Female**	**Male**	**Female**	**Male**	
Kenya	8	2	N/A	N/A	8	3	1	N/A	N/A	2
South Africa	6	0	2	0	8	7	0	2	0	8
Nigeria	7	1	2		10	6	0	1	0	7
Sub-total	**21**	**1**	**4**	**0**	**26**	**15**	**0**	**3**	**0**	**17**
Learners										
Grade 2	**Intervention**		**Control**		**Totals**	**Intervention**		**Control**		**Totals**
Kenya	405		N/A		405	288		N/A		288
South Africa	301		85		386	297		93		390
Nigeria	101		30		131	77		14		91
Sub-total	**807**		**115**		**922**	**662**		**107**		**769**

The teacher demographics reveal significant gender disparities across all three countries. In Kenya, the intervention group consisted of 8 female and 2 male teachers (8 at baseline, 4 at endline), and no control group data recorded (N/A). South Africa showed a similar pattern, with 6 female teachers in the intervention group at baseline (increasing to 7 at endline) and only 2 female teachers in the control group no male teachers participated. Nigeria had slightly more

diversity, with 7 female and 1 male teacher in the intervention group at baseline (dropping to 6 females at endline) and 2 female teachers in the control group. The overwhelming predominance of female teachers (21 out of 26 at baseline, 15 out of 17 at endline) suggests that foundational-level teaching in these contexts is highly feminized, potentially reflecting broader societal trends where early education is viewed as "women's work."

The learner data reveals notable attrition trends across countries, with varying patterns between intervention and control groups. Kenya's intervention group significantly declined from 405 to 288 learners (29% attrition), while its control group had no recorded data (N/A). South Africa showed relative stability, with the intervention group decreasing slightly from 301 to 297 learners (1% attrition) and the control group increasing from 85 to 93 learners (9% growth). Nigeria experienced the most severe attrition, with the intervention group dropping from 101 to 77 learners (24% loss) and the control group plummeting from 30 to 14 learners (53% loss). These disparities suggest potential differences in program implementation, contextual challenges (e.g., Nigeria's higher dropout rates), or data collection inconsistencies. The lack of gender-disaggregated data ("Male/Female" combined reporting) severely limits insights into whether attrition affected boys and girls differently, a critical gap for equity analysis.

The apparent learner attrition reflects expected grade progression rather than true dropout, as the study tracked the original Grade 2 cohort into Grade 3 while excluding the new Grade 2 entrants at endline. Kenya retained 71% of intervention learners (405→288), Nigeria showed concerning retention gaps (76% intervention vs 47% control), and South Africa demonstrated near-perfect promotion (99% intervention, with control schools potentially enrolling new learners). These patterns reveal critical system dynamics: South Africa's stable transitions suggest effective grade promotion policies, while Nigeria's sharp control-group decline (47%) may indicate systemic dropout or repetition challenges during this transitional period. For accurate interpretation, future reporting should explicitly label cohort progression (Grade 2→3), present retention rates rather than raw numbers, and distinguish between natural grade movement and true attrition needing intervention. This revised understanding transforms the data from seemingly problematic attrition to valuable insights about educational flows, with Nigeria's control-group results particularly warranting investigation into grade-transition barriers

3.1 Teacher Practices

Learning areas: How are classroom instructional practices shifting? What are we learning about teachers' understanding of error analysis?

This evaluation assessed changes in teacher practices through 20 criteria measuring classroom management, identification and documentation of errors, diagnosis of errors and utilization of errors to inform instruction in mathematics. The intervention engaged 26 teachers at baseline and 17 at endline and teachers across Kenya, Nigeria, and South Africa in the intervention schools, including 2 control teachers, revealing significant improvements in teachers' compliance in the intervention schools as opposed to the control schools. The evaluation data demonstrates significant improvements in teacher practices across all three intervention countries, though with distinct patterns emerging in each context. Kenya exhibited strong progress, with average compliance scores increasing from 57% to 85%, representing a 28-percentage-point improvement. Nigeria showed the most dramatic transformation, soaring from 37% to 88% compliance - a remarkable 51-point gain that placed Nigerian teachers at the highest endline performance level. South Africa's intervention group, while starting from the lowest baseline at just 16%, achieved the greatest absolute improvement by reaching 68% compliance, a 52-point increase. These substantial gains suggest the intervention's effectiveness in enhancing teacher practices, particularly in the documentation and utilization of learner errors to inform mathematics instruction.

Several noteworthy patterns emerge from the individual teacher data. Nigeria's intervention teachers not only achieved the highest final scores despite beginning with the lowest baseline, but also included standout cases like Teacher 4, who improved from 8 to 18 points. Similarly, South Africa's Teacher 3 made exceptional progress, advancing from a baseline score of 2 to 17 at endline. These individual success stories, combined with the strong average improvements, indicate the intervention's potential to transform teaching practices when implemented effectively.

However, the study's limitations must be acknowledged when interpreting these positive results. The control group data proved particularly problematic, with Kenya's control teachers completely unavailable for endline assessment due to school transfers. The remaining control groups -

consisting of only two teachers each in Nigeria and South Africa - showed minimal improvement (Nigeria: +9 points; South Africa's endline average of 35% suggesting underperformance). These control group challenges, combined with South Africa's unusually low baseline scores that may reflect stricter initial assessment criteria, necessitate cautious interpretation of the results. While the intervention appears highly successful, especially in Nigeria and South Africa, the limited control data and small sample sizes prevent definitive claims about the program's exclusive responsibility for all observed improvements. Future research should address these methodological constraints by implementing stronger teacher retention strategies and expanding control group sizes to enable more robust comparative analysis.

Table 8: Teachers' compliance level

Variables	Teacher compliance level across countries					
Score /20	20 criteria					
Country	Kenya		Nigeria		South Africa	
Teachers (Tr)	Baseline	Endline	Baseline	Endline	Baseline	Endline
Tr1	15	18	7		3	14
Tr2	14		7		5	10
Tr3	12		8	16	2	17
Tr4	6		8	18	3	15
Tr5	14		7	18	3	11
Tr 6	6		7	19	2	17
Tr7	10	16	7	17		11
Tr8	14		7	17		
Average compliance score	11.4	17	7.4	17.5	3	15.7
Average compliance %	57%	85%	37%	88%	16%	68%
Control teachers						
TR	**Baseline**	**Endline**	**Baseline**	**Endline**	**Baseline**	**Endline**
Tr 1	N/A	N/A	7		4	7
Tr2	N/A	N/A	6	6	2	8
Average Compliance %	N/A	N/A	28%	37%	16%	38%

3.2 Learners' Performance in Numeracy

This evaluation critically examines whether the intervention's focus on identifying and addressing learner errors led to measurable improvements in teaching practices and learner outcomes. The analysis compares baseline versus endline error rates across intervention and control groups, assessing whether teachers' increased compliance in error documentation and instructional adaptation (as shown in Table 9) correlated with reduced learner mistakes, particularly in mathematics. Key considerations include: the types of errors (conceptual, procedural, or Factual errors

Table 9: Learner performance across the 3 countries

Countries	Kenya		South Africa		Nigeria		Kenya		Nigeria		South A
	Intervention Schools						Control School				
Stages	Baseline	Endline	Baseline	Endline	Baseline	Endline	Baseline	Endline	Baseline	Endline	Baseline
Conceptual errors	0%	7%	3%	2%	13%	1%	N/A	N/A	0%	1%	10%
Factual errors	79%	63%	48%	43%	59%	34%	N/A	N/A	49%	46%	30%
Procedural errors	4%	6%	18%	18%	3%	8%	N/A	N/A	6%	8%	19%
Correct response	12%	43%	37%	37%	26%	55%	N/A	N/A	63%	46%	41%

The data reveals significant differences in how the intervention impacted learner performance across the three countries. Nigeria's intervention schools demonstrated the most improvements, with conceptual errors moving from 13% to just 1% and correct responses more than doubling from 26% to 55%. This suggests the program was particularly effective at addressing fundamental misunderstandings and strengthening core knowledge in Nigerian classrooms. However, the slight increase in procedural errors (3% to 8%) indicates learners may need additional support in applying their knowledge to problem-solving.

South Africa's results were the least promising, with intervention schools showing minimal progress. Factual errors only decreased from 48% to 43%, while procedural errors remained stubbornly at 18%

and conceptual errors showed negligible improvement. Surprisingly, South Africa's control schools actually outperformed intervention schools in reducing factual errors (30% to 21%) and increasing correct responses (41% to 49%). This counterintuitive pattern raises serious questions about either the intervention's design or its execution in the South African context.

Kenya's intervention school showed mixed but promising results, with factual errors decreasing significantly from 79% to 63% and correct responses improving dramatically from 12% to 43%, suggesting the intervention helped strengthen foundational knowledge and overall performance. However, the emergence of conceptual errors (rising from 0% to 7%) and a slight increase in procedural errors (4% to 6%) raise concerns about potential gaps in learners' deeper understanding and problem-solving skills, possibly indicating that while the intervention enhanced rote learning, it may have overlooked higher-order thinking. Without control group data, it's difficult to determine whether these changes were directly caused by the intervention or influenced

There are some valuable divergent trends across these 3 countries, intervention schools Nigeria emerged as the strongest success case, with dramatic reductions in conceptual errors (13% → 1%) and factual errors (59% → 34%), alongside a striking increase in correct responses (26% → 55%). These gains suggest the intervention effectively addressed both foundational knowledge and deeper understanding, though the rise in procedural errors (3% → 8%) may reflect learners grappling with more complex applications of their improved skills. Kenya showed partial progress, with factual errors declining significantly (79% → 63%) and correct responses surging (12% → 43%). However, the new appearance of conceptual errors (0% → 7%) and minimal change in procedural errors (4% → 6%) imply the intervention prioritized rote learning over conceptual mastery, potentially leaving gaps in critical thinking. Effectiveness varied by error type: The intervention consistently reduced factual errors (most in Nigeria, least in South Africa) but had mixed results on conceptual/procedural skills. Correct responses improved most where conceptual errors dropped sharply (Nigeria), suggesting conceptual clarity drives overall performance. Implementation gaps likely exist: South Africa's stagnant results and Kenya's conceptual regression may reflect mismatches between the intervention and local needs

Whereas the control school comparisons in Nigeria and South Africa reveal important highlights. While Nigeria's control school saw a concerning drop in correct responses (63% to 46%), South Africa's control group improved despite no intervention. These divergent control group trajectories highlight how local factors like teaching quality or learner motivation may significantly influence outcomes independent of the program. The most robust finding comes from Nigeria, where the intervention school's gains far outpaced its control group, providing strong evidence of program effectiveness in that specific context.

3.5 Learners' Engagement and Participation

Table 10: Comparison of learners' engagement of baseline and endline

Category	Intervention schools			Control schools		
	Kenya	South Africa	Nigeria	Kenya	South Africa	Nigeria
Baseline	42%	56%	68%	N/A	73%	52%
Endline	71%	64%	75%	N/A	64%	64%

The engagement data reveals significant improvements in Kenya's intervention schools, with engagement rates moving from 42% at baseline to 71% at endline, a 29-percentage point increase that suggests the intervention had a high likelihood of success in boosting learner participation. Nigeria's intervention schools maintained consistently high engagement (68% to 75%), reinforcing the program's strong likelihood of success in sustaining already-good participation levels. South Africa's intervention schools showed more modest growth (56% to 64%), indicating a lower likelihood of transformative success where midline engagement was already moderate.

The control school comparison reveals Nigeria's control group improved from 52% to 64%, implying some natural progression or external factors may have contributed to engagement gains, highlighting the need to contextualize the intervention's true likelihood of success. Kenya's lack of control data (N/A) makes it difficult to isolate the program's independent likelihood of success, though the dramatic

improvement strongly suggests effectiveness. Notably, while Kenya's intervention schools started with the lowest engagement (42%), they surpassed South Africa's by endline (71% vs. 64%), demonstrating the intervention's highest likelihood of success in low-engagement settings.

These results suggest the program had the greatest likelihood of success in low-engagement contexts like Kenya, a steady likelihood of maintaining success in already-engaged Nigeria, and a more limited likelihood of achieving major gains in moderate-engagement South Africa.

3.6 FEASIBILITY

The intervention demonstrated that the systematic use of error analysis and formative assessment techniques is feasible to implement in Grade 2 and Grade 3 classrooms in Kenya, Nigeria, and South Africa. Implementation across four schools each in Nigeria and South Africa, and five schools in Kenya, showed that teachers were able to adopt these practices within existing instructional routines. The intervention was well-received by both teachers and learners, with observable improvements in teacher practices (table 8) and learner engagement (table 10), suggesting that identifying and addressing learner misconceptions can be practically integrated into diverse educational contexts. Despite these successes, significant barriers such as teacher transfers, limited instructional time, inconsistent teacher compliance, and variations in local contextual factors hindered full implementation fidelity. Nevertheless, targeted interventions like ongoing teacher coaching, structured feedback, and mobile-friendly resources significantly enhanced teacher capacity and adaptation, demonstrating that structured support mechanisms are critical to the intervention's feasibility. The marked improvements in learner performance, particularly in reducing conceptual and factual errors, indicate a strong potential for scaling this approach, provided that adequate adjustments and sustained support systems are incorporated to address identified challenges.

3.6.1 Implementation Practicality

The intervention demonstrated strong feasibility regarding practical implementation across diverse school contexts in Kenya, Nigeria, and South Africa. Despite initial challenges related to teacher transfers, limited instructional time, and variations in teacher compliance (table 8), educators effectively integrated error analysis and formative assessment into numeracy instruction. Significant improvements in teacher compliance were evident, with Kenya's teachers moving from 57% to 85%, Nigeria's from 37% to 88%, and South Africa's from 16% to 68%. The structured use of mobile-friendly resources and ongoing teacher coaching substantially reduced logistical barriers, enhancing implementation fidelity. However, findings indicated that continuous, structured training and consistent feedback mechanisms, as exemplified by Nigeria's targeted online coaching sessions following midline assessments, notably improved teacher capacity and instructional responsiveness, underscoring that sustained support significantly enhances feasibility in these contexts

3.6.2 Time and Resource Efficiency

The intervention demonstrated time and resource efficiency, achieving measurable gains within relatively short implementation periods; five months in South Africa and six months in Kenya and Nigeria. Longer, structured implementation correlated positively with learner engagement improvements; Kenya's learner engagement increased substantially from 42% to 71%, Nigeria's from 68% to 75%, while South Africa showed a modest improvement from 56% to 64%. These findings indicate that feasibility of the pedagogical shift particularly the systematic use of error analysis and formative assessment is enhanced when sufficient time is allocated for teacher adaptation and ongoing support. Additionally, the intervention's use of low-cost, accessible resources, such as mobile-friendly instructional guides adapted from existing materials, significantly improved financial viability and scalability across diverse, resource-constrained educational contexts

3.6.3 Teacher Adoption and Compliance (table 8)

High teacher compliance rates in the intervention schools (85% in Kenya, 88% in Nigeria, and 68% in South Africa) confirm that the error analysis and formative assessment methodology was accessible and acceptable to educators. In comparison, control schools had notably lower compliance, with Nigeria's control teachers moving only from 28% to 37%, and South Africa's from 16% to 38%, highlighting the intervention's effectiveness in driving pedagogical change. Stakeholder support was particularly strong among intervention teachers, who benefited from structured training sessions and targeted coaching based on a 20-criteria audit checklist, significantly enhancing instructional clarity, and reducing implementation ambiguity. Nigeria's remarkable compliance improvement from 37% to

88% underscores that intensive, structured support mechanisms greatly improve teachers' adoption of new practices. Nonetheless, variability in learner outcomes, such as modest learner engagement improvements in South African intervention schools (56% to 64%) compared to a slight decrease in their control group (73% to 64%), indicates that consistently meeting learning targets across contexts may require more nuanced and tailored training approaches.

3.6.4 Learner Engagement and Participation (table 10)

The intervention's design likely addressed learner engagement barriers, particularly in Kenya, where engagement increased from 42% at baseline to 71% at endline, a 29-percentage-point improvement. But with no control school, we can't determine if this is solely a success from the intervention or from other factors. However, there is likelihood of success of the intervention's structured use of formative assessment and error analysis in promoting active participation. Similarly, Nigeria showed a steady rise in engagement from 68% to 75%, while South Africa recorded a more modest gain from 56% to 64%. Notably, these increases were higher in contexts where teachers demonstrated stronger compliance with intervention practices such as Nigeria, where teacher compliance improved from 37% to 88%. However, the comparatively smaller improvement in South Africa suggests that contextual factors, including classroom size, language of instruction, and local instructional practices, influence feasibility. These findings highlight the importance of tailoring implementation to local environments to maximize learner participation and instructional impact.

In control schools, learner engagement trends remained inconsistent or declined, highlighting the impact of the intervention in comparison. In South Africa, engagement in the control school decreased from 73% at baseline to 64% at endline, a 9-percentage-point drop, suggesting reduced learner participation in the absence of structured pedagogical support. Conversely, Nigeria's control school showed an increase in engagement from 52% to 64%, a 12-point gain, though still lower than the intervention group's performance (which reached 75%). Kenya had no control school, making comparative analysis within that country impossible. These mixed results in control settings indicate that improvements in engagement observed in intervention schools were likely not part of broader trends but rather directly tied to the structured training, coaching, and pedagogical tools introduced through the program. The variability also reflects how, without consistent support or instructional innovation, learner engagement may stagnate or decline due to existing systemic constraints.

3.6.5 Impact on Learning Outcomes (table 9)

The intervention's effectiveness in improving numeracy skills particularly in conceptual domains is evident in the substantial learning gains observed in Kenya and Nigeria. In Kenya, correct responses increased from 12% to 43%, with a notable reduction in factual errors from 79% to 63%, while in Nigeria, conceptual errors dropped from 13% to 1%, and correct responses rose from 26% to 55%. These improvements reflect enhanced learner understanding when teachers systematically applied error analysis and formative assessment strategies. By contrast, South Africa's intervention group saw correct responses decline slightly from 41% to 39%, despite a reduction in factual errors (48% to 43%). This dip in correct responses may be attributed to the mid-intervention transfer of a Grade 2 teacher, after which that class was integrated into a Grade 3 section without replacement likely affecting instructional focus and learner consistency.

Control schools, meanwhile, showed minimal or inconsistent gains. In Nigeria, the control group's correct responses declined sharply from 63% to 46%, and factual errors only slightly decreased (49% to 46%). South Africa's control school saw a moderate improvement in correct responses (41% to 49%) and a reduction in factual errors (30% to 21%), but still underperformed compared to the intervention's intended impact. These contrasts affirm that observed gains in intervention schools particularly in Kenya and Nigeria were not part of broader trends, but were driven by the structured, data-informed instructional shifts introduced by the program. However, persistent challenges in operational domains were noted: for instance, procedural errors in South Africa's intervention group increased from 19% to 25%, and in Nigeria from 3% to 8%, suggesting that additional strategies, such as manipulatives or focused practice, may be needed to strengthen learners' procedural fluency.

3.6.6 Sustainability and Scalability

The intervention's low cost, minimal infrastructure demands, and alignment with existing teaching practices positioned it as a feasible model for broader adoption across Kenya, Nigeria, and South Africa. The successful use of mobile-friendly and print-adaptable materials allowed for integration into resource-constrained settings without reliance on expensive technology. Teacher compliance

improvements in intervention schools rising from 57% to 85% in Kenya, 37% to 88% in Nigeria, and 16% to 68% in South Africa demonstrated strong teacher adoption when accompanied by structured support and feedback mechanisms. In contrast, control schools showed markedly lower gains: Nigeria's control teachers improved only from 28% to 37%, and South Africa's from 16% to 38%, reinforcing the conclusion that structured training and ongoing coaching were likely key to successful implementation.

However, the variation in outcomes across contexts highlights important feasibility considerations. For example, in South Africa, the transfer of a Grade 2 teacher without replacement and the merging of Grade 2 with Grade 3 introduced instructional disruptions, likely contributing to the slight decline in correct responses from 41% to 39%, despite overall improvement in teacher compliance. In Kenya, strong engagement gains (from 42% to 71%) were not always matched by mastery in skill-specific areas, particularly subtraction, pointing to the need for targeted instructional strategies alongside the error analysis and formative assessment approach. These contextual differences indicate that long-term feasibility requires not only low-cost materials but also structural support and adaptation to local realities.

To ensure sustainability, the intervention must be institutionalized within existing teacher development frameworks. Embedding training on error analysis and formative assessment into national programs such as Kenya's competency-based curriculum implementation or Nigeria's state-led teacher capacity initiatives would support consistent application at scale. A phased approach to scaling, with pilot expansion across diverse school types, increased teacher training coverage, and strengthened monitoring systems, would help address structural gaps identified during the pilot. Future adaptations should include teacher-led content development to increase contextual relevance, as well as mechanisms for addressing equity issues, such as uneven resource distribution. With these refinements, the intervention has demonstrated both the feasibility and scalability needed to strengthen foundational numeracy in varied sub-Saharan African education contexts

3.6.7 Readiness for trial

The intervention demonstrated clear potential for scaling, evidenced by substantial improvements in teacher instructional practices, learner engagement, and numeracy performance in Kenya, Nigeria, and South Africa. The positive reception from teachers and education stakeholders suggests that systematic use of error analysis and formative assessment can effectively enhance foundational numeracy. High compliance rates, 88% in Nigeria, 85% in Kenya, and 68% in South Africa, highlighted strong adoption by teachers, underpinned by structured, criterion-based training and ongoing coaching sessions. Learner engagement similarly improved, most notably in Kenya (from 42% to 71%) and Nigeria (68% to 75%), though more moderate gains in South Africa (56% to 64%) indicated contextual differences in effectiveness.

However, readiness for larger-scale implementation requires addressing several refinements. Teacher training emerged as critical, particularly in aligning formative assessment practices closely with curricular requirements. The variability in learner outcomes, especially persistent challenges in procedural skills, with Nigeria's procedural errors slightly increasing from 3% to 8% despite overall improvement underscores the need for further training on how teachers can effectively use error analysis insights in daily instruction. Structured mentorship and regular feedback cycles, exemplified by Nigeria's intensive coaching model introduced after midline assessments, appear essential for maintaining teacher fidelity at scale.

Contextual barriers must also be systematically addressed for broader implementation. Resource constraints significantly limited instructional quality in classrooms, especially due to overcrowding and inconsistent access to mobile-friendly teaching resources. Teachers faced difficulties fully integrating provided materials into daily instruction, indicating the need for more extensive involvement of educators in adapting materials to local contexts. Given these challenges, scaling the intervention effectively will require context-specific, culturally relevant instructional resources and locally adaptable lesson plans.

Methodologically, the pilot highlighted the need for improved design features in future larger trials. The minimal control groups, just one each in Nigeria and South Africa, with none in Kenya significantly limited the evaluation's capacity to confidently attribute observed improvements solely to the intervention. Strengthening future evaluations with larger, randomized control groups and stronger sampling methods would greatly enhance external validity and causal claims.

The low-cost nature of error analysis and formative assessment methods, reliant on accessible, mobile-friendly materials rather than expensive technological investments, enhances feasibility for

large-scale adoption. However, technological constraints such as inconsistent teacher access to reliable mobile devices. The scalable solutions, such as offline-accessible resources or printed supplementary materials are highly recommended for widespread implementation.

Institutional integration also emerged as critical for sustainability. Strong stakeholder ownership, particularly evident through active participation from local education officials and consistent teacher engagement across the pilot countries, establishes a solid foundation for embedding these pedagogical techniques into national education systems. Future scale-up will benefit from institutionalizing formative assessment and error analysis within teacher professional development frameworks, ideally in collaboration with Ministries of Education, teacher training institutions, and district education offices to ensure long-term adoption and policy alignment.

Overall, the intervention demonstrated promising readiness for broader implementation, conditional on strategic refinements in training, resource alignment, methodological rigor, and systemic integration. These insights offer a clear pathway toward sustainable scale-up and long-term improvements in foundational numeracy across diverse educational contexts in Kenya, Nigeria, and South Africa.

4. CONCLUSION

4.1 Formative findings

The error analysis and formative assessment-based numeracy intervention has shown strong potential in improving teacher instructional practices, learner engagement, and numeracy performance in Kenya, Nigeria, and South Africa. Comparative data revealed notable differences among the countries, with implementation duration, structured support, and teacher compliance emerging as critical factors.

Teacher support structures played a critical role in implementation quality. Nigeria’s use of structured online coaching after midline assessments notably improved teachers’ ability to apply error analysis effectively. The structured audit checklist (20 criteria) provided explicit guidance, significantly reducing ambiguity in instructional methods, and enhancing classroom practice consistency.

Contextual factors mediated intervention effectiveness substantially. Schools experiencing high teacher transfers or class sizes exceeding 40 learners faced challenges that constrained the intervention's effectiveness. The presence of existing instructional aids correlated positively with numeracy outcomes; classrooms with adequate basic materials and resources demonstrated better learner performance, particularly in foundational numeracy skills.

Implementation intensity was pivotal in achieving threshold effects. Nigeria and Kenya’s extended six-month implementation periods facilitated deeper pedagogical shifts and higher learner engagement compared to South Africa’s shorter five-month duration. The critical role of sustained training and iterative feedback was underscored by Nigeria’s marked improvement in instructional practices post-midline.

General constraints, such as overcrowded classrooms and limited instructional time, created clear ceilings on performance gains. For instance, large class sizes diluted learner engagement and teacher effectiveness, particularly where physical teaching resources were scarce. Moreover, contextual adaptability emerged as essential; rural and urban classrooms required different support strategies and culturally responsive materials to ensure learner engagement.

For scalability and sustainability, integrating error analysis and formative assessment techniques into national teacher training policies and curricula is crucial. Collaboration with ministries of education and local education authorities will be necessary to institutionalize the intervention. Future research should adopt larger sample sizes, stronger control group designs, and contextually tailored instructional materials to enhance the robustness and generalizability of findings.

In summary, while the numeracy intervention demonstrates significant promise across diverse contexts, its effectiveness depends fundamentally on tailored implementation strategies, sustained teacher support, and addressing systemic classroom constraints.

4.2 Interpretation

The findings show important lessons regarding both the feasibility and the limitations of the intervention across diverse contexts. Substantial improvements in teacher compliance and learner engagement demonstrate that structured pedagogical interventions using error analysis and formative assessment can be successfully adopted in resource-constrained settings. However, the translation of

improved instructional practices into consistent mastery of numeracy skills, especially procedural competencies, remained uneven. These results highlight the need for targeted instructional support to strengthen procedural fluency.

Contextual and general factors significantly mediated implementation success. Teacher transfers, class size challenges, and tool revisions influenced both implementation fidelity and outcome interpretation.

The critical role of ongoing teacher support emerged strongly. Higher compliance and engagement in intervention schools compared to controls clearly demonstrate that structured coaching and feedback cycles drive pedagogical change. However, persistent procedural skill gaps across contexts suggest that future interventions must integrate explicit instructional strategies for complex numeracy domains.

4.3 Limitations and Future Research

As a small-scale pilot designed to explore the feasibility of implementing error analysis and formative assessment practices in foundational numeracy, this evaluation was not intended to generate definitive conclusions. The limited sample size four intervention schools each in Nigeria and South Africa, and five in Kenya reflects the pilot nature of the study and naturally constrains generalizability. Similarly, the small control group (one control school in Nigeria and South Africa, and none in Kenya) was appropriate for pilot purposes. Future larger-scale studies should adopt more representative samples and cluster-randomized controlled designs to strengthen external validity.

The pilot's short-term scope spanning a single academic year was appropriate for an initial feasibility study but does not capture longer-term sustainability of outcomes. Future longitudinal research is needed to assess whether improvements in numeracy performance, instructional practices, and learner engagement are sustained over time. Examining teacher fidelity post-training, particularly in contexts with teacher mobility or grade transitions, is also an important priority.

Future research should further incorporate disaggregated analyses to explore differential impacts on specific learner subgroups, such as low-performing students, learners with disabilities. Given the prevalence of overcrowded classrooms in participating contexts (often exceeding 60–90 learners), scalable instructional strategies such as peer-assisted learning and differentiated formative assessment warrant further investigation.

Finally, this pilot provides a strong foundation for both academic dissemination and policy engagement. Comparative analyses across the three countries, as well as case studies of teacher adaptation, could inform the broader evidence base. Policy briefs and cross-country learning exchanges could support Ministries of Education in integrating these practices into national numeracy strategies and curricular reforms. Dissemination through international forums such as CIES and national education sector reviews would further promote the scalability and sustainability of the intervention.

By building on the promising findings of this pilot and addressing these methodological priorities, future research can contribute to the development of sustainable, scalable models for improving foundational numeracy in diverse educational contexts.

REFERENCES

Black, P., & Wiliam, D. (1998). Inside the Black Box: Raising Standards Through Classroom Assessment. Phi Delta Kappan, 80(2), 139–148.

APPENDICES

Appendix 1: Learner Standardized Test Description and Sample Items
https://docs.google.com/document/d/1m8c8q0C6PUqmkXPSTKF6aYnV8YBbYQiM/edit?usp=drive_link&ouid=103468004330621141040&rtpof=true&sd=true

Appendix 2 Observation checklist https://ee-eu.kobotoolbox.org/x/vkNnXqa2

Appendix 3: Perception of Math's https://ee-eu.kobotoolbox.org/x/DoapHPv8

Appendix 4: Audit checklist https://ee-eu.kobotoolbox.org/x/gO43U3Vj

Appendix 5: Key informant interviews for teachers https://forms.gle/CjvMottZbcGkbbdY9

Appendix 6: Error Log
https://docs.google.com/spreadsheets/d/1PvaztvMvogXzkNHqIYz6JGJ9OaGJtuP0/edit?usp=drive_link&ouid=103468004330621141040&rtpof=true&sd=true

LIST OF TABLES