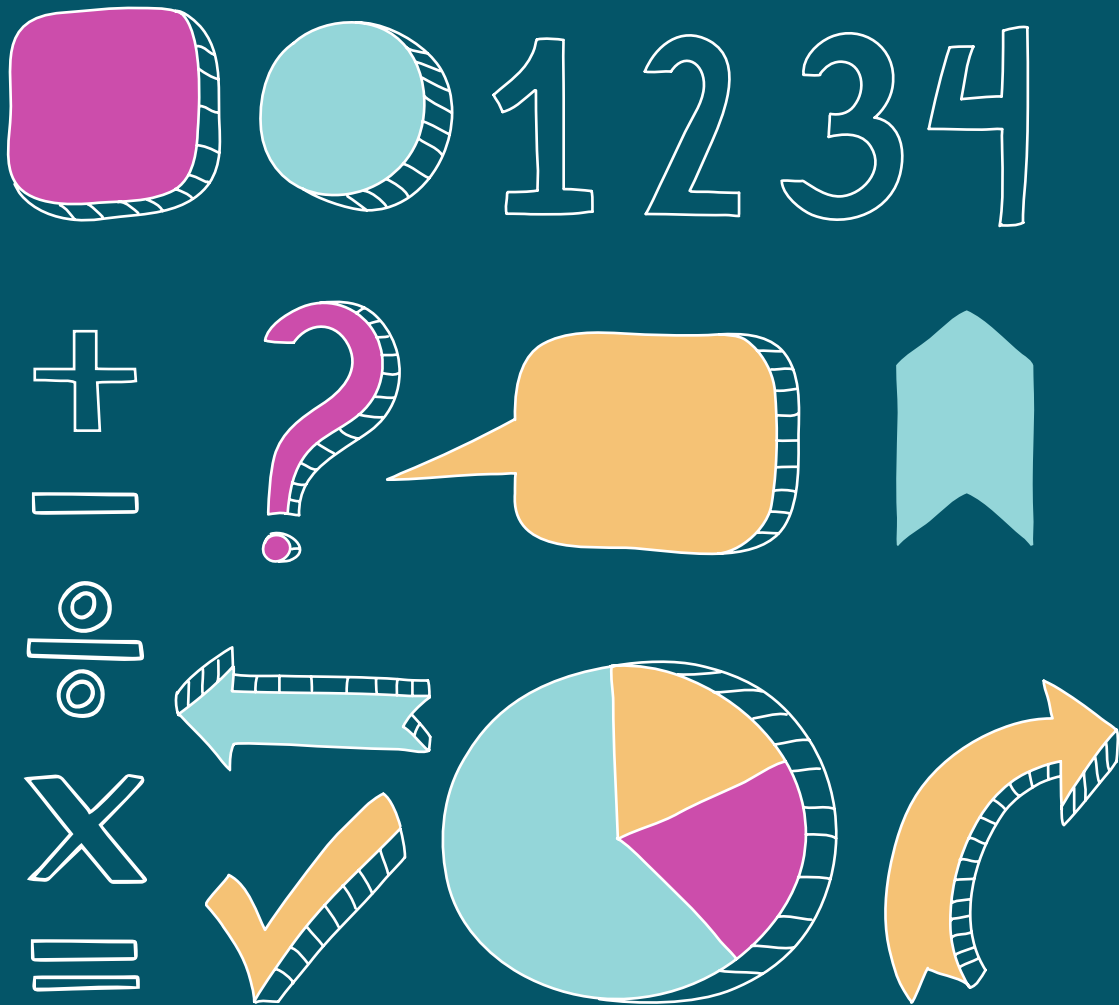


2026



An Innovation in Teaching Foundational Numeracy

Lessons from a Longitudinal Error Analysis Study in Bungoma West Sub-County.

2026

An Innovation in Teaching Foundational Numeracy:

Lessons from a Longitudinal Error Analysis Study in
Bungoma West Sub-County.

Suggested Citation: Mizizi Elimu Afrika (2026). *An Innovation in Teaching Foundational Numeracy: Lessons from a Longitudinal Error Analysis Study in Bungoma West Sub-County*. Nairobi: Mizizi Elimu Afrika.

Table of Contents

List of Tables	4
List of Figures	4
Abbreviations & Acronyms	5
Executive Summary	6
Acknowledgement	8
1. Introduction and Background	9
1.1 Error Analysis in Numeracy Instruction	9
1.2 Why Error Analysis Matters in Numeracy Learning	9
1.3 The Bungoma County Context	10
2. Objectives of the Study	11
3. Methodology	12
3.1 Study Design	12
3.2 Study Area and Participants	12
3.3 Intervention Overview	12
3.4 Assessment Structure and Process	12
3.5 Sampling and Analytical Approach	13
3.6 Data Analysis	13
4. Results	14
4.1 Sample Distribution, Demographic Characteristics, and Performance by Assessment Wave	14
4.1.1 Sample Distribution	14
4.1.2 Demographic Characteristics	14
4.1.3 Performance by Assessment Wave	15
4.2 Error Pattern by Wave and Task Type	18
4.2.1 Error Patterns by Wave, Sex, and Task Type – Addition	18
4.2.2 Error Patterns by Wave, Sex, and Task Type – Subtraction	20
4.2.3 Summary of Errors in Both Addition and Subtraction	22
4.3 Method Patterns by Wave, Sex, and Task Type	23
4.3.1 Addition	23
4.3.2 Subtraction	25
4.3.3 Summary of Methods – Both Addition and Subtraction	27
4.4 School Performance by Task – Addition and Subtraction	28
5. Conclusion	31
6. Recommendations	31
7. References	33

List of Tables

Table 1: Sample Distribution by Wave	14
Table 2a: Demographic Characteristics (Overall Proportion)	14
Table 2b: Demographic Characteristics (Number and Sex)	15
Table 3a: Learners Assessed by Sex	16
Table 3b: Performance by Wave Over Time	16
Table 4: Error Pattern by Wave and Task Type – Addition	19
Table 5: Error Pattern by Wave and Task Type – Subtraction	21
Table 6: Method Pattern by Task – Addition	24
Table 7: Method Pattern by Task – Subtraction	26
Table 8: School Performance by Task – Addition and Subtraction	28

List of Figures

Figure 1: Numeracy Performance Trends Over Time - Baseline, Follow-up 1 & Follow-up 2	16
Figure 2: Addition Error Patterns by Assessment Wave: Distribution of Error Types Over Time	13
Figure 3: Subtraction Error Patterns by Assessment Wave: Distribution of Error Types Over Time	16
Figure 4: Method Pattern by Task – Addition	20
Figure 5: Method Pattern by Task – Subtraction	23
Figure 6: School Performance by Task – Addition and Subtraction	26

Abbreviations & Acronyms

ASER	Annual Status of Education Report
CBC	Competency-Based Curriculum
CEMASTE	Centre for Mathematics, Science and Technology Education in Africa
CoPs	Communities of Practice
CRA	Concrete–Representational/Semi-Concrete–Abstract
FGDs	Focus Group Discussions
HoI	Head of Institution
J-PAL	Abdul Latif Jameel Poverty Action Lab
KNEC	Kenya National Examinations Council
LMICs	Low- and Middle-Income Countries
MoE	Ministry of Education
NAEP	National Assessment of Educational Progress
SACMEQ	Southern and Eastern Africa Consortium for Monitoring Educational Quality
STEM	Science, Technology, Engineering, and Mathematics
SCDE	Sub-County Director of Education
TaRL	Teaching at the Right Level
TSC	Teachers Service Commission
UNESCO	United Nations Educational, Scientific and Cultural Organization
UNICEF	United Nations International Children's Emergency Fund
ZAF	Zizi Afrique Foundation (now Mizizi Elimu Afrika)

Executive Summary

This report presents the findings of an innovative numeracy intervention jointly implemented by Mizizi Elimu Afrika (formerly Zizi Afrique Foundation) and the Centre for Mathematics, Science and Technology Education in Africa (CEMASTE) in partnership with the Ministry of Education in Bungoma West Sub-County. The initiative aimed to enhance foundational numeracy outcomes among early-grade learners through the integration of error analysis, a reflective teaching approach that uses learner mistakes as diagnostic tools for instructional improvement.

The study employed a mixed-methods design, combining quantitative learner assessments with qualitative teacher reflections and classroom observations. To deepen this inquiry, the study was guided by two central research questions that shaped both the data collection and analysis:

- i. *What are the common errors observed in learners' numeracy performance, and how do these errors shift over time?*
- ii. *How does targeted teacher training and ongoing instructional support influence classroom practice, and in turn, how does this impact numeracy learning outcomes over time?*

In addressing the second question, the study further examined the fidelity with which teachers applied the recommended instructional practices, drawing on evidence from the lesson observation tool. Comparisons of classroom practice before and after the structured support showed measurable improvements in teachers' use of guidance-aligned techniques, such as explicit modelling, probing questioning, structured practice routines, and error-focused feedback. These shifts in practice were positively associated with gains in learner numeracy outcomes, demonstrating that consistent use of the lesson observation tool not only strengthened the quality of instruction but also contributed to improved learner performance. Across three assessment rounds, baseline (October 2024), midline (March 2025), and endline (July 2025), a total of 194 Grade 2 learners were assessed at baseline and 292 at endline. However, this report focuses on 103 learners who were consistently tracked and matched across all three assessment points.

“Factual errors declined sharply across the three assessment waves, signalling stronger arithmetic fluency.”



Photo: CEMASTE and Mizizi team taking part in joint stakeholder dissemination of the Error Analysis study findings in Bungoma West Sub-county.



Photo: CEMASTEa and Mizizi team taking part in joint stakeholder dissemination of the Error Analysis study findings in Bungoma West Sub-county.

The findings reveal significant and sustained improvement in learners' mastery of addition and subtraction. At baseline, only 9 per cent of addition responses and 4 per cent of subtraction responses were correct, and no learner demonstrated mastery across all tasks. By the final assessment, correct responses had risen to 42 per cent and 43 per cent respectively, and 34 per cent of learners demonstrated mastery across all tasks. Factual errors, which were initially the most common, declined sharply across the three waves, indicating strengthened arithmetic fluency and conceptual understanding. Learners also transitioned from concrete problem-solving methods (such as drawings and manipulatives) to abstract reasoning and place-value strategies, consistent with the Concrete-Representational-Abstract (CRA) model of learning.

At the school level, notable differences emerged: Schools 1 and 3 demonstrated consistent and superior performance, while others exhibited stagnation or slower improvement. A closer examination of classroom practice across the schools shows that these differences were not incidental but closely tied to how fidelity to the recommended instructional practices evolved. In Schools 1 and 3, teachers progressively strengthened their practice through sustained lesson observations, which provided actionable feedback that was subsequently reflected in their instructional routines. This translated into clearer modelling of concepts, more accurate identification and correction of learner errors, and greater adherence to structured lesson sequencing.

Moreover, teachers in these schools increasingly adopted individualised learner support, using assessment data and error patterns to tailor remediation, resulting in more responsive and differentiated instruction. Classroom management also improved notably as teachers emphasised collaborative learning structures, enabling learners to engage in peer explanation, shared problem-solving, and collective error analysis. Together, these shifts reveal a visible improvement in fidelity to the instructional approach, distinguishing high-performing schools from those where practices remained inconsistent or weakly applied. The alignment between improved fidelity and rising learner performance thus provides more concrete evidence that teacher capacity, consistent instructional practice, and effective leadership meaningfully shaped the learning outcomes observed.

Overall, the intervention affirms that structured, data-driven, and reflective teaching practices can meaningfully improve numeracy performance in early grades. The findings provide actionable insights for educators and policymakers seeking scalable and equitable strategies to strengthen foundational learning within Kenya's Competency-Based Curriculum (CBC) framework.

“By the final assessment, correct responses in addition rose from 9% to 42%, and in subtraction from 4% to 43%.”

Acknowledgements

The successful development and completion of this report was made possible through strong government leadership, with CEMASTEА at the forefront, and the collaboration, commitment, and technical contributions of numerous institutions and individuals.

We particularly acknowledge **Ms. Virginia Ngindiru** (Mizizi Elimu Afrika), **Utibe Henshaw** (The Education Partnership (TEP) Centre), **Ingrid Saphire** (Funda Wande), and **Ambang Tatianne** (eBASE Africa) for their invaluable contributions to the conceptualisation and technical development of the project. We are especially grateful to Ms. Virginia Ngindiru for her strategic leadership, including providing insightful technical guidance and constructive feedback throughout the report development process.

We further acknowledge **Vollan O. Ochieng** for leading the design and implementation of the project, coordinating the research and evidence synthesis process, and authoring this report. We are equally grateful to **Cyrus Wakibia** and **the eBASE team** for leading the analysis of the study findings; **James Mburu** for coordinating field activities and logistics that ensured the smooth implementation of the study; and **Francis Njuguna** for facilitating the procurement and operational processes that supported project implementation. We also acknowledge **Dr. John Mugo** and **Dr. Maurice Mutisya** for their valuable advisory input in shaping and packaging this publication.

Special appreciation goes to the Communications Team for their support in designing, editing, and documenting the materials that contributed to the production of this report. We also thank all members of the Mizizi Elimu Afrika team who contributed in diverse ways through technical input, coordination, quality assurance, and feedback that strengthened both the intervention and this publication.

Our sincere appreciation goes to the **Centre for Mathematics, Science and Technology Education in Africa (CEMASTEА)**, whose leadership and technical expertise were instrumental throughout the project. CEMASTEА guided the design of the teaching strategies, supported the development of training materials, coordinated two rounds of data collection, and provided sustained professional support to participating teachers. These contributions were foundational to the rigour and success of this work.

We also acknowledge the valuable contributions of the Daara Learning Partnership, including The Education Partnership Centre (Nigeria), Funda Wande (South Africa), and eBASE Africa (Cameroon), whose shared learning, technical expertise, and cross-country collaboration enriched this report. We further appreciate the Daara Development Academy, the Daara Innovation Fund through which this work was supported, the Gates Foundation support structure, and eBASE Africa for conceptualizing and conducting the evaluation. The broader collaborative architecture, technical contributions, and Innovation Fund learning and reporting products also informed the synthesis and interpretation reflected in this publication.

Finally, this work would not have been possible without the continued support of **the Ministry of Education through the Sub-County Director of Education and the Sub-County Director of the Teachers Service Commission (TSC)**. Their leadership, technical guidance, facilitation, and human resource support were instrumental in the successful implementation of the intervention and the achievement of its intended outcomes.

To everyone who contributed their time, expertise, and commitment to this initiative, we express our heartfelt gratitude.



1. Introduction and Background

A strong foundation in numeracy during the early years of primary education (ages 6–11) is essential for lifelong learning, problem-solving, and participation in economic and social life. However, global evidence shows that many children do not acquire the basic numeracy competencies expected at their level. According to UNESCO (2021), approximately 60 per cent of learners globally lack foundational numeracy skills by the end of primary school, with the challenge being more pronounced in low- and middle-income countries (LMICs).

While boys and girls often display similar numeracy abilities in early childhood, girls sometimes perform as well as or even better than boys in the early grades. Yet, as learners progress to middle and upper primary levels, gender gaps often emerge, with boys beginning to outperform girls in numeracy-related tasks. This imbalance contributes to the higher representation of boys in Science, Technology, Engineering, and Mathematics (STEM) fields at higher education levels. In Sub-Saharan Africa, for instance, studies show that girls aged between 6 and 11 years can perform comparably or better than boys in numeracy (World Bank, 2022). However, cultural and socio-economic barriers, such as early marriage, domestic responsibilities, and school dropout, limit girls' continued progress (World Bank, 2021).

Evidence from Uwezo (2021) reveals that in Kenya and Uganda, girls in Grades 2 and 3 perform equally well or slightly better than boys in basic numeracy. Similar findings in Southern Africa show that girls in countries like Malawi and South Africa also demonstrate comparable or higher performance in early numeracy (SACMEQ, 2017; UNICEF, 2020). Despite this promising start, disparities tend to appear at higher levels of achievement, with fewer girls advancing into mathematics-intensive fields (NAEP, 2021). The same pattern is seen in South Asia, where girls in India and Pakistan perform at par with boys in early grades but later face limitations due to cultural norms and restricted access to learning opportunities (ASER, 2022; World Bank, 2021).

Bridging these gaps calls for targeted interventions that promote equitable learning outcomes by identifying and addressing the underlying challenges that affect girls and boys differently as they advance through school. One effective approach to improving numeracy learning outcomes and addressing learning gaps is error analysis, a reflective, evidence-based practice that focuses on understanding learners' mistakes as opportunities for learning.

1.1 Error Analysis in Numeracy Instruction

Error analysis refers to the systematic process of identifying, categorising, and addressing the mistakes learners make while solving mathematical problems (Brookfield, 2017). This approach provides insights into learners' thought processes and helps teachers design interventions that directly address learning difficulties. As a reflective practice, error analysis encourages both teachers and learners to critically evaluate the causes of errors and develop corrective strategies (Brookfield, 2017; Schön, 1983).

For teachers, it offers an opportunity to reflect on their instructional methods and make adjustments to enhance understanding. For learners, it builds metacognitive awareness, helping them recognise and correct their own mistakes, thereby deepening conceptual understanding and improving problem-solving skills.

1.2 Why Error Analysis Matters in Numeracy Learning

Error analysis is particularly valuable in numeracy education because mathematical knowledge builds progressively. Misconceptions or small misunderstandings at early stages can lead to persistent challenges later. Through error analysis, teachers can:

- Identify common patterns of mistakes and misconceptions in basic operations such as addition, subtraction, multiplication, and division.
- Differentiate between factual errors (lack of number facts or basic information), conceptual

errors (misunderstanding mathematical principles), and procedural errors (incorrect steps or operations).

- Design lessons and interventions that address specific learner needs before misconceptions become entrenched.

When used consistently, error analysis enhances formative assessment, promotes learner reflection, and fosters a classroom environment where mistakes are viewed as a natural part of the learning process rather than failure (Brookfield, 2017; Schön, 1983).

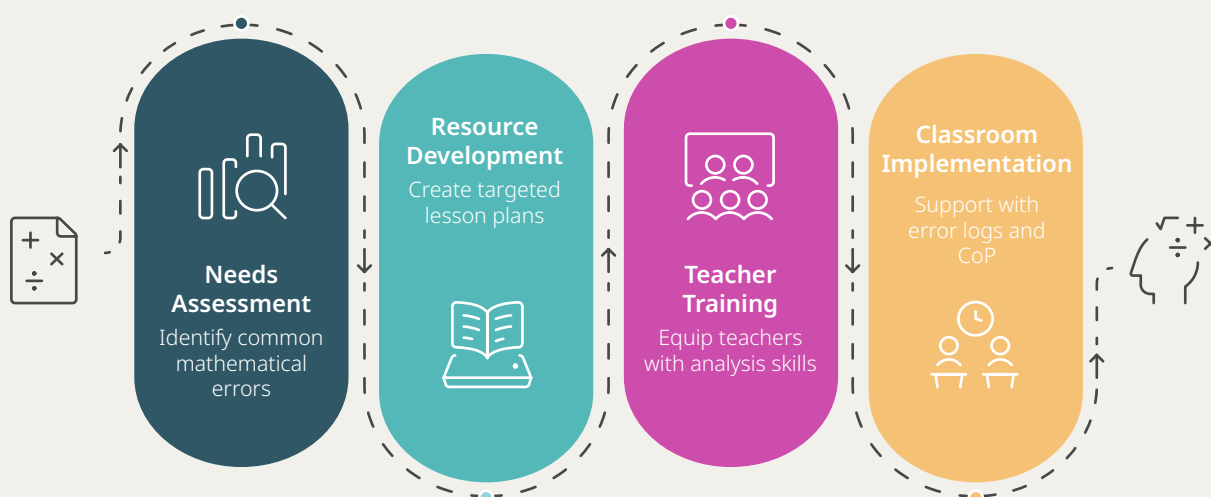
1.3 The Bungoma County Context

In 2024, the Centre for Mathematics, Science and Technology Education in Africa (CEMASTEA), together with Mizizi Elimu Afrika (then Zizi Afrique Foundation – ZAF), launched a targeted error-analysis intervention in **Bungoma County, Western Kenya**. The choice of Bungoma was far from arbitrary: despite being an agriculturally productive region, Bungoma has persistently lagged in learning outcomes. According to Uwezo’s assessments, children in Bungoma underperformed in both literacy and numeracy, placing the county among the lower-performing counties in past national surveys (Usawa Agenda, 2023). This aligns with broader equity concerns highlighted by the Usawa Agenda, whose work stems from the Uwezo citizen-led assessment methodology (Usawa Agenda, 2023). By situating the project in Bungoma, CEMASTEA and Mizizi Elimu Afrika sought to address this paradoxical underperformance in a relatively resource-rich county, underscoring a commitment to education justice and learning equity.

The project was structured around a comprehensive set of interventions designed to improve teaching and learning through error analysis. First, a needs assessment was conducted to identify the most common mathematical errors early grade learners make. Next, teacher resources and lesson plans were developed, explicitly designed to target those errors. Teachers were trained in error-identification and analysis techniques, equipping them to diagnose misconceptions and respond appropriately. Implementation in the classroom was supported through ongoing mechanisms such as error logs, individualised learner support, and a Community of Practice (CoP). Finally, there was regular monitoring, including classroom observations and structured feedback sessions, enabling continuous refinement of instructional practice. Throughout, CEMASTEA played a critical role, not only offering technical guidance and material development but also leading two rounds of data collection to track progress.

Additionally, the study design included reviewing learner scripts to categorise and map error types, conducting focus group discussions with headteachers and teachers, and carrying out in-depth classroom observations and interviews. Through this multi-faceted intervention, the project aimed to understand not just *what* errors learners made and *why*, but also *how* improved teacher practice, rooted in error awareness and fidelity to pedagogical principles, could drive meaningful, sustained gains in numeracy learning.

Improving Math Learning Through Error Analysis

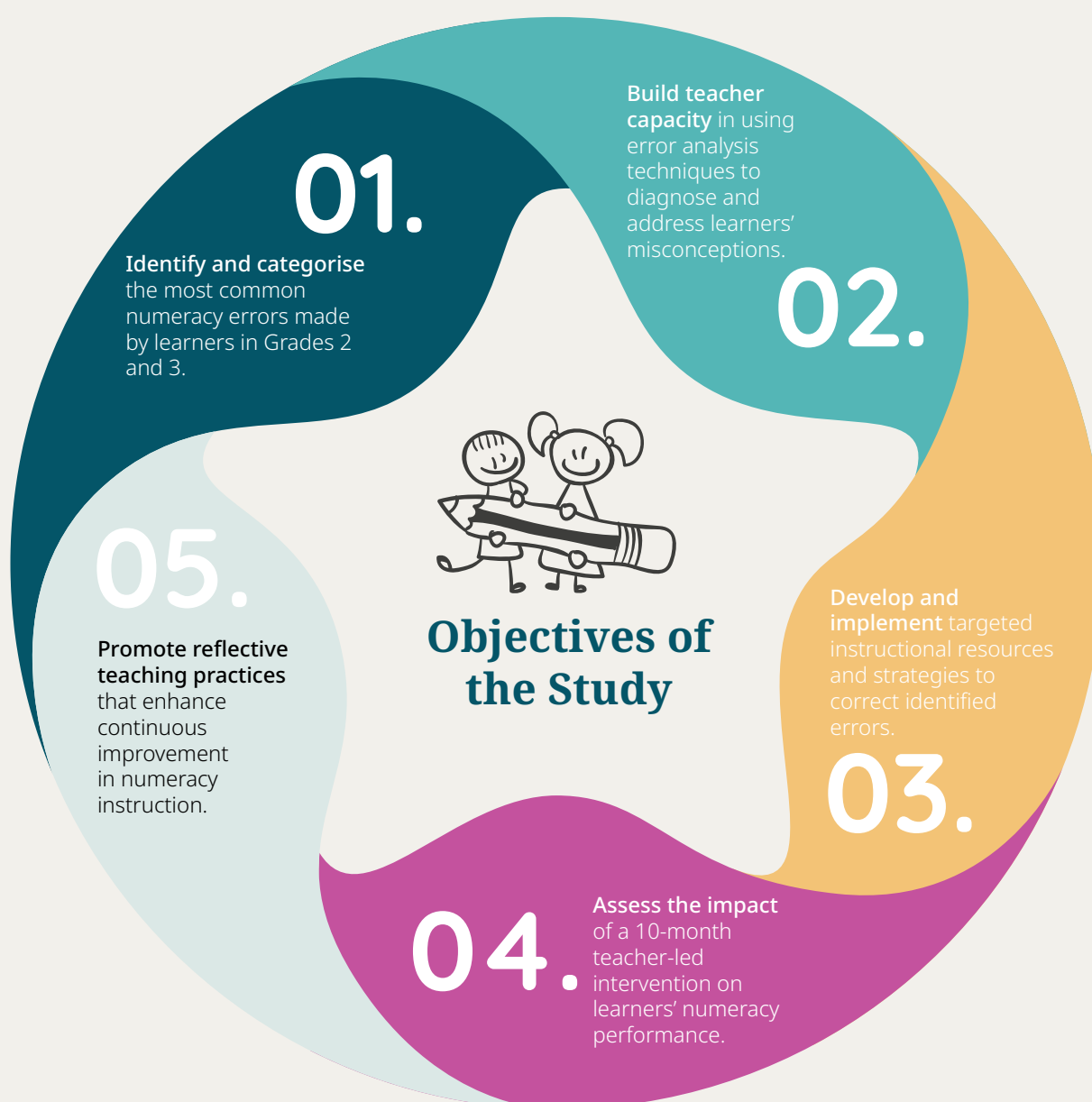


2. Objectives of the Study

The main objective of this study was to improve foundational numeracy outcomes among early-grade learners through the integration of error analysis into classroom practice.

The specific objectives were to:

1. Identify and categorise the most common numeracy errors made by learners in Grades 2 and 3.
2. Build teacher capacity in using error analysis techniques to diagnose and address learners' misconceptions.
3. Develop and implement targeted instructional resources and strategies to correct identified errors.
4. Assess the impact of a 10-month teacher-led intervention on learners' numeracy performance.
5. Promote reflective teaching practices that enhance continuous improvement in numeracy instruction.



3. Methodology

3.1 Study Design

The study adopted a mixed-methods approach, integrating both quantitative and qualitative methods to obtain a comprehensive understanding of early-grade numeracy performance and the role of error analysis in improving learning outcomes. Quantitative data were collected through learner assessments, while qualitative insights were drawn from teacher interviews, focus group discussions (FGDs), and classroom observations.

This design allowed the research team to triangulate findings and better understand both the patterns of learner errors and the instructional practices that influence numeracy learning outcomes.

3.2 Study Area and Participants

The study was conducted in Bungoma West Sub-County, selected due to its diversity of learning environments, empirical evidence that puts Bungoma County among the lower-performing counties (KNEC, 2023; Usawa Agenda, 2023), and the existing collaboration between the County and Mizizi Elimu Afrika on foundational learning initiatives.

Participants included:

- Grade 2 and Grade 3 learners from five public primary schools;
- Teachers handling Grade 2 and Grade 3 classes; and
- Headteachers from the same schools.

3.3 Intervention Overview

To determine the foundational numeracy performance levels, a baseline assessment was administered. Using the evidence from the baseline, Mizizi Elimu Afrika and CEMASTEIA, in partnership with the Bungoma West Sub-County Education Office, implemented a 10-month structured intervention during Term 1 of 2025. Teachers maintained error logs for learners throughout the intervention period. These logs captured common mistakes, teacher observations, and corrective strategies applied. The logs were used for reflective teaching, progress monitoring, and individualised learner support.

The intervention sought to strengthen teacher capacity in error analysis and reflective instructional practices to improve learner outcomes.

Key components of the intervention included:

1. Conducting a diagnostic needs assessment to identify common mathematical errors among learners.
2. Training teachers on identifying, categorising, and addressing learner errors.
3. Implementing classroom-based support mechanisms such as coaching, peer learning, and the use of error logs.
4. Continuous monitoring through classroom observations, communities of practice (CoPs), and feedback sessions.

3.4 Assessment Structure and Process

The study involved three rounds of learner assessments conducted over two academic terms to monitor learner performance over time.

- The baseline assessment was conducted in October 2024 prior to the intervention.
- The midline assessment occurred during the six-week intervention in March 2025.
- The endline assessment was conducted in July 2025 at the conclusion of the intervention.



At baseline (October 2024), a total of 194 Grade 2 learners participated in the assessment. These were followed to 2025, now at Grade 3, where a total of 292 Grade 3 learners (formerly Grade 2 in 2024) were assessed at endline. However, due to variations in school enrolments and learner attendance, the present paper focuses on 103 learners (62 female, 41 male, *see Table 1*) who were consistently tracked and matched across all three assessment points: baseline, midline, and endline.

Tracking these 103 learners provided a consistent sample for examining changes in individual learner performance over time, ensuring that observed improvements could be directly attributed to the intervention and not to differences in group composition.

The differences in learner numbers between assessment rounds can be attributed to several factors, including learner transfers to other schools or sub-counties, absenteeism during data collection periods, and new entrants or enrolments joining Grade 2 after the baseline assessment. In some cases, a combination of two or more of these factors affected learner availability. Such variations are common in longitudinal studies involving early-grade learners in public schools and were carefully considered during data cleaning and analysis to maintain the accuracy and reliability of the findings.

3.5 Sampling and Analytical Approach

Given the longitudinal nature of the study, the analysis emphasised changes in performance among the 103 matched learners who participated in all three rounds of assessment. This matched sample formed the basis for tracking learner progress and evaluating the effect of the intervention on individual learning outcomes.

Both descriptive and inferential statistical analyses were used to measure performance changes, supported by qualitative themes derived from teacher interviews, FGDs, and classroom observations. The triangulation of quantitative and qualitative data provided a more nuanced understanding of learner progress and the practical application of error analysis in classroom contexts.

3.6 Data Analysis

Data analysis combined quantitative and qualitative methods. Quantitative data from learner assessments were analysed to determine mean scores, error frequency, and performance shifts across the three rounds. Qualitative data from interviews, FGDs, and observations were analysed thematically to explain trends and contextual factors influencing learning outcomes.

By focusing on the 103 learners tracked throughout the study, the analysis offers a clear picture of learner growth, persistence of errors, and the overall effectiveness of error analysis as a reflective instructional practice in numeracy.

4. Results

4.1 Sample Distribution, Demographic Characteristics, and Performance by Assessment Wave

4.1.1 Sample Distribution

The study successfully matched 103 Grade 2 learners across all three assessment waves: October 2024, March 2025, and July 2025, ensuring a consistent longitudinal cohort for analysis. Within this matched sample, 62 learners were female (60.2%), and 41 were male (39.8%) (see Tables 1 & 2a). This stable sex distribution across the three data collection points strengthens the reliability of sex-disaggregated analyses, enabling meaningful comparisons of learner performance, progression trends, and error patterns over time. Unlike the cross-sectional data, which compares different groups of learners at each point, the matched analysis tracks the same learners over time, providing a more accurate picture of genuine learning progress.

The predominance of female learners in the matched cohort provides an opportunity to examine whether observed numeracy outcomes reflect broader enrolment patterns or emerging gender-related differences in foundational skills acquisition. Additionally, the balanced representation across waves minimises attrition bias, allowing for a more accurate interpretation of learning trajectories. Overall, the consistency of the sample enhances the validity of inferences drawn regarding how boys and girls develop foundational numeracy skills and the extent to which specific misconceptions or error types may be gender-differentiated.

Table 1: Sample Distribution by Wave

Wave	Female		Male		Overall	
	n	Percentage	n	Percentage	n	Percentage
Baseline	62	60.2	41	39.8	103	33.3
Follow-up 1	62	60.2	41	39.8	103	33.3
Follow-up 2	62	60.2	41	39.8	103	33.3

4.1.2 Demographic Characteristics

At baseline, the matched cohort of **103 Grade 2 learners** had a mean age of **8.6 years** (age recorded only once at the first assessment) (see Tables 2a & 2b). No age difference between female and male learners was observed at baseline, indicating comparable chronological maturity across sex groups in this cohort. Because age was measured only at baseline, it functions here as a descriptive demographic characteristic rather than a time-varying covariate.

The baseline age homogeneity between girls and boys reduces the likelihood that sex-disaggregated differences in numeracy outcomes are confounded by systematic age differences at the start of the study. Consequently, longitudinal comparisons of performance and error-pattern trajectories by sex are less likely to reflect simple chronological-age effects and more likely to reflect genuine differences in learning or instruction-related factors.

However, the single time-point measurement of age introduces important limitations. Birthdays that occur between assessment waves and any age-related maturation effects that unfold during the study are not captured, so age cannot be modelled as a time-varying influence on learning gains. This restricts the ability to distinguish developmental (age-driven) change from instructional or intervention effects across waves.

Table 2a: Demographic Characteristics (Overall Proportion)

Wave	Sample Size	Male	Female	Average Age	Standard Deviation (Age)
Baseline	103	40%	60%	8.57	1.09

Table 2b: Demographic Characteristics (Number and Sex)

Wave	Female		Male		Overall	
	n	Average Age	n	Average Age	n	Average Age
Baseline	62	8.6	41	8.6	103	8.6
Follow-up 1	62	8.6	41	8.6	103	8.6
Follow-up 2	62	8.6	41	8.6	103	8.6

Interpretation

The baseline finding that the matched cohort of 103 Grade 2 learners had a mean age of 8.6 years, with no observable age differences between girls and boys, indicates that the sample was chronologically comparable at the start of the study. This age homogeneity suggests that both groups began the assessment period with similar developmental maturity, reducing the likelihood that differences in numeracy scores or error patterns across the waves were attributable to unequal age-related cognitive development. Since age was collected only once at baseline, it functions strictly as a descriptive variable rather than a factor that can be tracked over time. As such, the study's interpretation of change in numeracy performance relies primarily on instructional exposure and learning progression rather than age as a changing demographic characteristic. This strengthens the internal validity of sex-disaggregated comparisons while simultaneously highlighting the limits of age-related analysis across assessment periods.

Implications

The absence of baseline age differences between male and female learners allows for a more confident interpretation of sex-related trends in numeracy performance, as these patterns are unlikely to be confounded by age disparities. This supports reliable comparison of learning trajectories and error types across groups. However, because age was not reassessed in subsequent waves, the study cannot fully isolate the influence of natural maturation on observed learning gains. Some improvements in numeracy performance may partially reflect typical developmental progress rather than instructional or intervention-driven effects. Additionally, reporting the distribution of ages (e.g., range or standard deviation) could offer further insights into whether even slight age variations within the cohort might influence numeracy acquisition patterns among early-grade learners.

4.1.3 Performance by Assessment Wave

Sex-disaggregated analysis uncovered distinct trajectories in addition and subtraction performance. At baseline, male learners outperformed females in addition tasks (15% vs 5%) (see Table 3b and Fig. 1), while both sexes displayed similarly low levels of subtraction skill. Over the course of the study, girls made proportionally larger gains. For instance, by the second follow-up, females achieved 47 per cent correct in addition compared with 34 per cent for males, and 44 per cent correct in subtraction versus 42 per cent for males. Considering overall mastery across the full set of numeracy tasks, no learner demonstrated complete mastery at baseline; by Follow-up 2, mastery rates reached **36 per cent for females** and **32 per cent for males**. These patterns indicate not only great overall improvement across waves but also a closing, and in some measures a reversal, of the initial sex gap in basic numeracy.

Importantly, factual errors¹, the most common type of mistake at the beginning, declined considerably over the three waves. This decline confirms that the learners' progress was not due to chance or differences in test groups, but real improvements in computation skills. The consistency of gains

¹ Error Categorisation

Learner errors were systematically classified into three categories:

- i. Factual errors – resulting from a lack of number facts or basic information;
- ii. Procedural errors – stemming from incorrect steps in mathematical operations; and
- iii. Conceptual errors – arising from misconceptions or misunderstandings of mathematical principles.

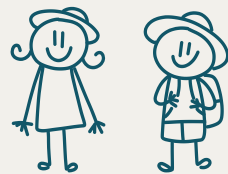
Findings showed that common numeracy errors among early-grade learners (specifically Grades 2 and 3) were factual errors (lack of number fact knowledge), procedural errors (e.g., skipped regrouping), and conceptual errors (misunderstanding place value).

These insights guided the development of a structured intervention aimed at improving early numeracy through reflective teaching and targeted learner support.

across all learners supports the reliability of the teaching and learning strategies implemented during the period.

Table 3a: Learners Assessed by Sex

Male	41
Female	62
Overall	103



Learners
Assessed
by Gender

41 Male
62 Female



Overall
Learners
Assessed by
Sex

103
Learners

Table 3b: Performance by Wave Over Time

Wave	Addition		Subtraction		All Tasks by Sex		Overall		
	Female	Male	Female	Male	Female	Male	Addition	Subtraction	All Tasks
Baseline	5%	15%	5%	2%	0%	0%	9%	4%	0%
Follow-up 1	26%	39%	21%	24%	10%	15%	31%	22%	12%
Follow-up 2	47%	34%	44%	42%	36%	32%	42%	43%	34%



Figure 1: Numeracy Performance Trends Over Time – Baseline, Follow-up 1, and Follow-up 2

The presented evidence indicates that cohort tracking across baseline, midline (*Follow-up 1*), and endline (*Follow-up 2*) demonstrated steady improvement in numeracy mastery. To support the quantitative progression observed, teacher reflections affirmed consistent learner growth over time. Teachers also noted that tracking the same learners strengthened understanding of individual learning trajectories:

“You know we’ve monitored these learners from the beginning, and you see a difference. When we started, they could not add even simple numbers... but now, most of them are getting them right” (Teacher 2_FGD_School 1);

“Those we have been following, they improve pole pole [slowly], but they get there. By end of term, even the slow ones catch up somehow because we keep repeating” (Teacher 4_FGD_School 4).

These teacher observations align with the progressive mastery pattern shown in **Table 3** and are consistent with longitudinal literacy evidence demonstrating learning gains with structured, continuous support (J-PAL & Pratham, 2022; Mix et al., 2022).

Consistent with the strong gains recorded in the quantitative data, teachers also reported visible improvement in computation accuracy, especially after the first assessment wave. They emphasised that learners first mastered addition before gaining confidence in subtraction, mirroring the sequential growth pattern in **Figure 1**:

“For addition they picked faster... subtraction came later when they understood how to break and combine numbers” (Teacher 1_FGD_School 3);

“At first subtraction was a challenge... borrowing confused them. But with time and more practice, they got it” (Teacher 7_FGD_School 2).

These perceptions confirm that subtraction mastery followed foundational strengthening in addition, consistent with developmental arithmetic trajectories (Mix et al., 2022; Byrne, 2023).

Interpretation

The baseline advantage observed for male learners on addition (15% vs 5%) suggests an early disparity in emergent numeracy skills within this cohort. However, the longitudinal data show that this gap did not persist. Instead, female learners demonstrated larger absolute and relative gains across both operations, such that by the final assessment, females exceeded males in addition and slightly in subtraction. This convergence (and partial reversal) points to greater learning momentum among girls during the period covered by the study. Several explanations are plausible: differential responsiveness to instruction, variations in classroom participation or practice opportunities, or initial floor effects for girls that allowed for steeper proportional gains. The comparable subtraction performance at baseline, followed by larger female gains, implies that instruction or other learning inputs during the study were effective in addressing early deficits, particularly for girls. The emergence of substantive mastery levels by *Follow-up 2* (36% female, 32% male) signals a meaningful shift from near-zero baseline mastery to measurable competency for a sizeable minority, indicating that the combination of exposure and time was sufficient for many learners to acquire integrated numeracy skills.

Implications

The observed sex-differentiated trajectories carry several programmatic and analytic implications. First, the fact that girls closed and, in some cases, surpassed boys on key numeracy measures suggests that classroom instruction and/or targeted supports deployed during the study period can be effective in reducing early sex gaps. Therefore, scale-up efforts should preserve whatever elements of pedagogy, practice intensity, or learner support contribute to these gains. Second, because gains were operation-specific and not uniform across sexes, assessment and remediation should remain disaggregated by task (addition vs subtraction) to identify persistent weaknesses and tailor instructional approaches. For instance, differentiated small-group practice or scaffolded tasks for learners who lag in a particular operation. Third, the attainment of nontrivial mastery for

roughly one-third of learners by *Follow-up 2* indicates substantial room for improvement. Programme designers should aim to raise the proportion achieving mastery by increasing instructional time on core concepts, using frequent formative checks, and integrating targeted error-analysis practices that address common misconceptions. Finally, from a monitoring and evaluation perspective, these findings recommend routine sex-disaggregated reporting, formal testing of gain differences (including confidence intervals and effect sizes), and exploration of mediating factors (teacher practices, attendance, home support) to determine why girls in this cohort experienced relatively larger gains, information that can inform equitable scaling and further intervention refinement.

Overall, the matched learner data provides clear, compelling evidence of genuine learning progress in foundational numeracy. Learners improved steadily in both addition and subtraction, reduced their error rates, and demonstrated stronger conceptual understanding by the end of the second follow-up (endline). The results affirm that structured, continuous instruction supported by learner tracking can significantly improve arithmetic proficiency. This pattern reinforces global research emphasising the role of sustained, evidence-based pedagogy in achieving foundational numeracy goals (de Barros et al., 2022; J-PAL & Pratham, 2022; Mix et al., 2022).

4.2 Error Pattern by Wave and Task Type

The analysis of learner performance across the three assessment waves revealed a consistent and encouraging pattern of improvement in both addition and subtraction operations.

4.2.1 Error Patterns by Wave, Sex, and Task Type – Addition²

Error patterns on addition tasks shifted markedly across the three assessment waves (*see Table 4 and Fig. 2*). At baseline, factual errors were overwhelmingly dominant, comprising 65.7 per cent of responses overall (female = 67.7%, male = 62.6%), while only 28.2 per cent of responses were correct and procedural errors were minimal (6.1% overall). By Follow-up 1, correct responses rose to 46.1 per cent overall (males 48.4%, females 44.6%), and factual errors declined to 42.9 per cent, although they remained the most frequent error type. Notably, conceptual errors³ surfaced at Follow-up 1 (≈6.2%), suggesting the beginnings of learners grappling with underlying number concepts rather than only producing surface-level mistakes. By Follow-up 2, overall accuracy had increased further to 58.6 per cent (females 62.4%, males 52.8%). Factual errors continued to fall to 30.1 per cent overall, while procedural errors rose modestly to 10.7 per cent, consistent with a shift towards attempting more complex or multi-step strategies. Conceptual errors remained rare at 0.6 per cent overall.

² The tasks on addition involved both tasks with and without regrouping, with the former presenting high level of complexity to compared to 'addition without regrouping'. Addition with regrouping (former) to a significant extent contributed to the observed addition results. The same applied for 'subtraction' tasks.

³ A strong numeracy foundation requires a balanced development of conceptual knowledge and procedural fluency, both of which play distinct yet interdependent roles in mathematical learning. Conceptual knowledge refers to a learner's understanding of mathematical ideas, operations, and the relationships between them. It enables learners to make sense of why a procedure works, not just how to carry it out. On the other hand, procedural fluency is the ability to apply mathematical procedures accurately, efficiently, and flexibly. This includes recalling steps, choosing appropriate strategies, and adapting known procedures to solve new types of problems.

It is important to recognize that conceptual knowledge and procedural fluency are:

- Dependent on a learner's developmental level, meaning both evolve as the learner progresses through different stages of understanding.
- Mutually reinforcing, with conceptual understanding supporting the meaningful application of procedures, and procedural fluency strengthening and deepening conceptual insight through repeated, purposeful practice.
- A continuum rather than separate skills, where growth in one area naturally supports the other.

However, they are not discrete skills, nor can they be reduced to memorization. Effective numeracy development moves beyond rote recall to building deep connections that allow learners to transfer knowledge to real-world situations.

Despite their interdependence, classroom instruction often leans heavily towards isolated procedures, emphasizing step-by-step methods without anchoring them in meaningful concepts. This can easily lead to confusion, as learners may appear procedurally fluent yet lack the conceptual understanding needed to adapt those procedures to unfamiliar problems. True application of mathematical knowledge, especially in new or complex situations, requires both conceptual understanding and procedural fluency working together.

Table 4: Error Pattern by Wave and Task Type – Addition

Addition		Count			Percentage		
Wave	Error Type	Female	Male	Overall	Female	Male	Overall
Baseline	Correct	45	42	87	24.2	34.1	28.2
	Factual error	126	77	203	67.7	62.6	65.7
	Procedural error	15	4	19	8.1	3.3	6.1
Follow-up 1	Correct	83	59	142	44.6	48.4	46.1
	Factual error	82	50	132	44.1	41	42.9
	Procedural error	8	7	15	4.3	5.7	4.9
	Conceptual error	13	6	19	7	4.9	6.2
Follow-up 2	Correct	116	65	181	62.4	52.8	58.6
	Factual error	49	44	93	26.3	35.8	30.1
	Procedural error	19	14	33	10.2	11.4	10.7
	Conceptual error	2	0	2	1.1	0	0.6

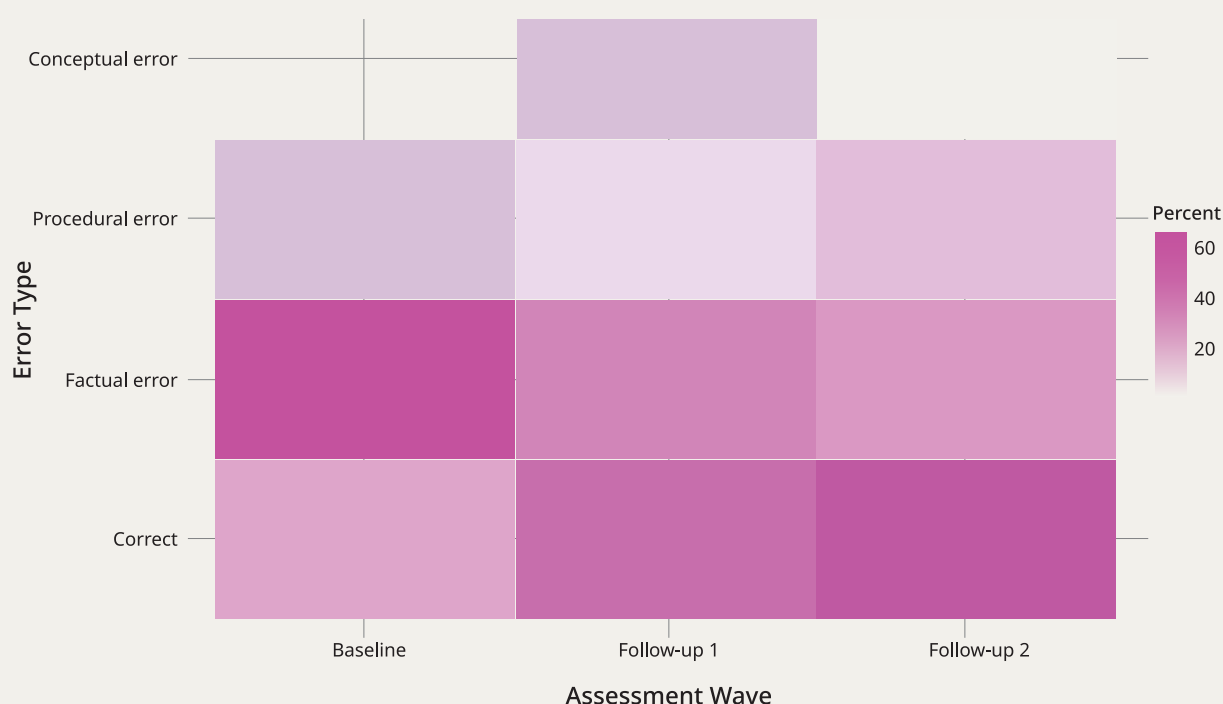


Figure 2: Addition Error Patterns by Assessment Wave: Distribution of Error Types Over Time

Interpretation

The longitudinal pattern indicates a clear movement from predominantly factual mistakes towards higher correctness and greater engagement with procedural strategies. The large baseline share of factual errors suggests many learners began with limited automatic recall of number facts or arithmetic facts, which constrained their ability to solve addition items correctly. The reduction in factual errors and simultaneous rise in correct responses across waves implies effective gains in fact fluency and/or greater practice with the tasks. The modest increase in procedural errors by Follow-up

2, alongside rising accuracy, likely reflects that learners were attempting more sophisticated solution paths (for example, carrying or decomposing numbers), a developmentally positive sign even when such attempts sometimes produced procedural mistakes. The transient appearance of conceptual errors at Follow-up 1 (and their low frequency at Follow-up 2) warrants careful interpretation: it may indicate that some learners moved beyond rote procedures and, while beginning to reason about underlying concepts, occasionally exposed conceptual gaps; alternatively, the small percentage suggests conceptual misunderstandings were not widespread. Sex-disaggregated differences, notably higher female accuracy at Follow-up 2 (62.4% vs 52.8%), mirror the broader performance trends in the study and indicate that females in this cohort made relatively larger gains in converting previously factual/guessing errors into correct or procedurally attempted responses.

Implications

These error-pattern dynamics have several practical implications for instruction, assessment, and programme design. First, the dominance of factual errors at baseline points to an immediate need for interventions that build arithmetic fact fluency (e.g., short daily practice, retrieval practice, number talks, and targeted drills), especially in the early phase of instruction. Second, the observed shift towards procedural attempts suggests that once fluency improves, teachers should scaffold procedural strategy development, explicitly teaching systematic methods (decomposition, carrying, estimation), modelling solution steps, and providing corrective feedback that targets common procedural missteps. Third, the emergence, albeit small, of conceptual errors underscores the importance of complementing fluency and procedure with activities that build conceptual understanding (manipulatives, visual representations, guided number-sense tasks) so that procedural competence is supported by robust conceptual foundations. Fourth, sex-disaggregated monitoring of error types is warranted because the pace and nature of improvements differed by sex. Tailored small-group instruction or differentiated practice could address operation-specific weaknesses. Finally, for monitoring and further research, it is recommended to construct error-transition matrices to track how individual learners move between error categories across waves, evidence that will better inform targeted remediation and teacher professional development focused on the most persistent and instructional-responsive error types.

4.2.2 Error Patterns by Wave, Sex, and Task Type – Subtraction

Subtraction tasks were consistently more challenging than addition, exhibiting similar improvement trajectories over time but with lower overall accuracy at each wave (see Table 5 and Fig. 3). At baseline, factual errors overwhelmingly dominated subtraction responses, comprising 77.3 per cent of items overall (females 76.9%, males 78.0%), while only 12.6 per cent of responses were correct. Conceptual errors were slightly more prevalent in subtraction at baseline (5.5%) compared with addition, indicating early weaknesses in underlying number relationships and subtraction reasoning.

By Follow-up 1, accuracy improved substantially to 40.5 per cent correct overall (females 40.9%, males 39.8%). Despite this progress, factual errors remained the primary impediment, composing 43.7 per cent of responses, while conceptual errors increased to 7.8 per cent, suggesting greater engagement with subtraction concepts even as many learners continued to struggle to apply them accurately.

At Follow-up 2, overall correctness rose further to 58.1 per cent, approaching parity with addition performance. Females again showed somewhat higher accuracy (59.5%) than males (56.1%). Factual errors declined to 32.8 per cent overall, but the sex-disaggregated pattern reveals a persistent male disadvantage in fact recall or simple computation (males 42.3% vs females 26.5% factual errors). Procedural errors at Follow-up 2 displayed a divergent pattern by sex. Males showed no procedural errors at this wave, whereas females exhibited 8.6 per cent procedural errors, a pattern that may reflect sex-differentiated shifts in problem-solving strategies (for example, females attempting more multi-step procedures while males relied more on basic recall or alternative approaches).



Table 5: Error Pattern by Wave and Task Type – Subtraction

Subtraction		Count			Percentage		
Wave	Error Type	Female	Male	Overall	Female	Male	Overall
Baseline	Correct	24	15	39	12.9	12.2	12.6
	Factual error	143	96	239	76.9	78	77.3
	Procedural error	11	3	14	5.9	2.4	4.5
	Conceptual error	8	9	17	4.3	7.3	5.5
Follow-up 1	Correct	76	49	125	40.9	39.8	40.5
	Factual error	79	56	135	42.5	45.5	43.7
	Procedural error	14	11	25	7.5	8.9	8.1
	Conceptual error	17	7	24	9.1	5.7	7.8
Follow-up 2	Correct	110	69	179	59.5	56.1	58.1
	Factual error	49	52	101	26.5	42.3	32.8
	Procedural error	16	0	16	8.6	0	5.2
	Conceptual error	10	2	12	5.4	1.6	3.9

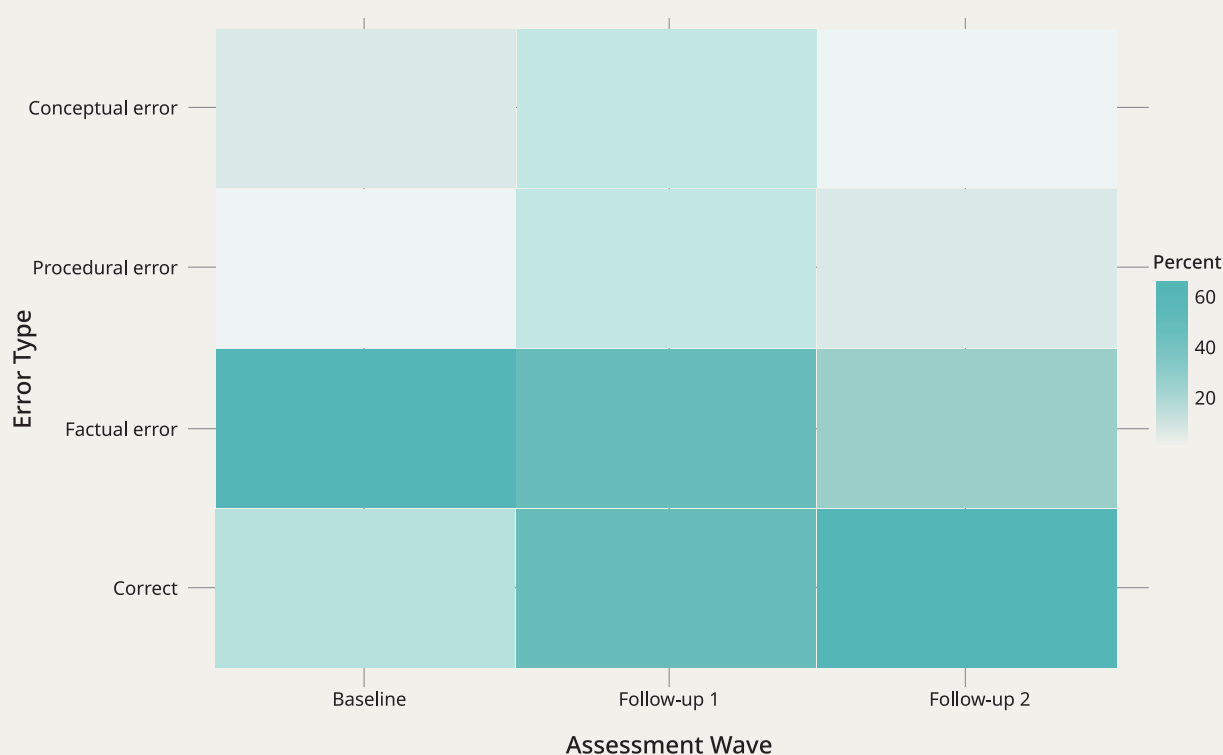


Figure 3: Subtraction Error Patterns by Assessment Wave: Distribution of Error Types Over Time

As factual accuracy improved, teachers also noticed fewer careless mistakes, validating the reduction in factual errors observed. They attributed this to repeated practice, drills, and teacher guidance:

“Nowadays they don’t guess...they know the facts. We do number bonds every morning and it has helped” (Teacher 8_FGD_School 5);

“At the start they were doing many small mistakes. Sasa, mistakes zimepungua [now, mistakes have reduced], especially the obvious ones” (Teacher 4_FGD_School 2).

These qualitative insights reinforce the downward trend in factual errors reported in Tables 4 and 5, aligning with research showing that routine fluency practice reduces computation errors (Harnett, 2018).

Interpretation

The subtraction error profile shows a clear developmental progression. A baseline dominated by factual deficits, followed by increasing correctness and a gradual shift towards more varied error types as learners attempted more sophisticated strategies. The higher baseline rate of conceptual errors in subtraction implies that learners initially had a weaker understanding of the relationships and operations that underpin subtraction (e.g., part-whole concepts, borrowing/renaming). The rising share of conceptual errors at Follow-up 1, concurrent with improved accuracy, indicates that some learners began to move beyond rote counting or guessing and were beginning to reason about subtraction, sometimes revealing conceptual gaps in the process. By Follow-up 2, the reduction in factual errors, alongside a small increase in procedural errors (for females), suggests that many learners progressed from attempting simple recall to trying procedural strategies. Procedural mistakes at this stage are developmentally expected and often easier to remediate through targeted feedback than pervasive conceptual misunderstandings. The persistent sex differences, notably the higher factual error rate among males at Follow-up 2 and the appearance of procedural errors primarily among females, point to differential learning pathways. Males may still rely more on incomplete fact fluency, while females more frequently engage with multi-step strategies that expose procedural weaknesses. These descriptive patterns are informative but require confirmatory inferential analyses (e.g., interaction tests between sex and time) to determine whether observed differences are statistically robust and to estimate effect sizes.

Implications

Programmatically, the subtraction findings call for a two-pronged instructional focus. First, reinforce numeric/arithmetic fact fluency through short, frequent practice and retrieval activities targeted at learners, particularly boys, who continue to show high factual error rates. Improving automaticity will free cognitive resources for procedural reasoning. Second, explicitly teach and scaffold subtraction procedures and the underlying concepts using manipulatives, number lines, and visual representations to address conceptual misunderstandings and reduce procedural errors. This is especially important for girls in this cohort who are attempting, but sometimes executing multi-step strategies incorrectly. Teachers should use diagnostic formative assessments that disaggregate error types (factual vs procedural vs conceptual) so that small-group instruction can be precisely targeted. For monitoring and evaluation, we recommend sex-disaggregated reporting of error transitions (e.g., error-transition matrices showing how individual learners move between error categories across waves), routine calculation of effect sizes and confidence intervals for gains, and qualitative follow-up (classroom observation or teacher interviews) to identify instructional practices associated with different learner trajectories. Finally, professional development should emphasise formative error analysis, scaffolded modelling of procedures, and techniques for building conceptual understanding – interventions that together are most likely to convert factual deficits into robust procedural competence and conceptual clarity.

4.2.3 Summary of Errors in Both Addition and Subtraction

Overall, error-pattern analysis across addition and subtraction tasks revealed clear learning gains over time alongside persistent areas of difficulty. At baseline, factual errors were the predominant challenge in both operations, accounting for 65.7 per cent of responses in addition and 77.3 per cent in subtraction, underscoring learners’ limited recall of basic number facts. Subtraction was notably more difficult at

the outset, with only 12.6 per cent of responses answered correctly compared with 28.2 per cent in addition.

Progress across the assessment waves was substantial. By Follow-up 2, correct responses had risen to 58.6 per cent in addition and 58.1 per cent in subtraction, reflecting a convergence in performance and significant overall improvement in learners' computational proficiency. Although factual errors declined markedly, they remained the most common error type at the final assessment (30.1% in addition; 32.8% in subtraction), indicating that fact fluency continues to be a foundational gap for many learners.

Sex-disaggregated trends further highlighted differentiated progress. By Follow-up 2, female learners outperformed males in both addition (62.4% vs 52.8%) and subtraction (59.5% vs 56.1%), consistent with broader patterns observed across the study. Over the same period, procedural errors increased modestly (to 10.7% in addition and 5.2% in subtraction), suggesting that learners were increasingly attempting, and at times misapplying, more advanced or multi-step solution strategies as their understanding deepened. Conceptual errors remained relatively infrequent throughout, but their continued presence points to lingering gaps in foundational mathematical reasoning despite the overall upward trajectory in performance.

Summary Insight



The overall pattern of results for both addition and subtraction paints a picture of meaningful and sustained learning. Learners not only improved their computational accuracy but also demonstrated deeper conceptual understanding, particularly in subtraction. The slight procedural rise represents cognitive growth rather than regression, as learners transitioned towards more abstract strategies. Collectively, these outcomes reinforce that consistent, feedback-driven, and conceptually rich instruction leads to significant improvements in foundational numeracy, findings echoed in global education studies emphasising evidence-based approaches to early mathematics (Mix et al., 2022; J-PAL & Pratham, 2022; World Bank, 2025).

4.3 Method Patterns by Wave, Sex, and Task Type

4.3.1 Addition

Solution strategies for addition tasks evolved substantially across the three assessment waves, indicating a clear shift from concrete counting methods towards abstract place-value reasoning (see Table 6 and Fig. 4). At baseline, learners predominantly relied on concrete representations, with 67.3 per cent of recorded strategies used being dots/rings/strokes (females 70.4%, males 62.6%). Place-value recognition was infrequently observed (5.8% overall), and 11.3 per cent of attempts were not classifiable into any identifiable method, reflecting unclear or idiosyncratic approaches.

By Follow-up 1, strategy use shifted markedly. Place-value recognition became the primary strategy for the cohort (67.2% overall; females 66.1%, males 68.9%), while reliance on dots/rings/strokes declined to 19.5 per cent. The proportion of unclassifiable attempts also fell, to 8.4 per cent, indicating greater consistency in method use and clearer evidence of taught strategies being adopted.

At Follow-up 2, place-value reasoning consolidated as the dominant approach (81.2% overall; females 84.9%, males 75.6%), and the use of concrete representations had nearly disappeared (7.8% overall). Nevertheless, 9.4 per cent of attempts remained unclassifiable, suggesting a small subset of learners continued to apply unclear or unconventional methods despite the cohort-wide shift towards more sophisticated, formalized strategies.

Table 6: Method Pattern by Task – Addition

Addition		Count			Percentage		
Wave	Method Type	Female	Male	Overall	Female	Male	Overall
Baseline	Use of manipulatives	15	15	30	8.1	12.2	9.7
	Use of dots/rings/ strokes	131	77	208	70.4	62.6	67.3
	Recognising place value	13	5	18	7	4.1	5.8
	Mastery learning	9	9	18	4.8	7.3	5.8
	Cannot make sense of method/working	18	17	35	9.7	13.8	11.3
Follow-up 1	Use of dots/rings/ strokes	46	14	60	24.7	11.5	19.5
	Recognising place value	123	84	207	66.1	68.9	67.2
	Mastery learning	6	9	15	3.2	7.4	4.9
	Cannot make sense of method/working	11	15	26	5.9	12.3	8.4
Follow-up 2	Use of dots/rings/ strokes	15	9	24	8.1	7.3	7.8
	Recognising place value	158	93	251	84.9	75.6	81.2
	Mastery learning	0	5	5	0	4.1	1.6
	Cannot make sense of method/working	13	16	29	7	13	9.4

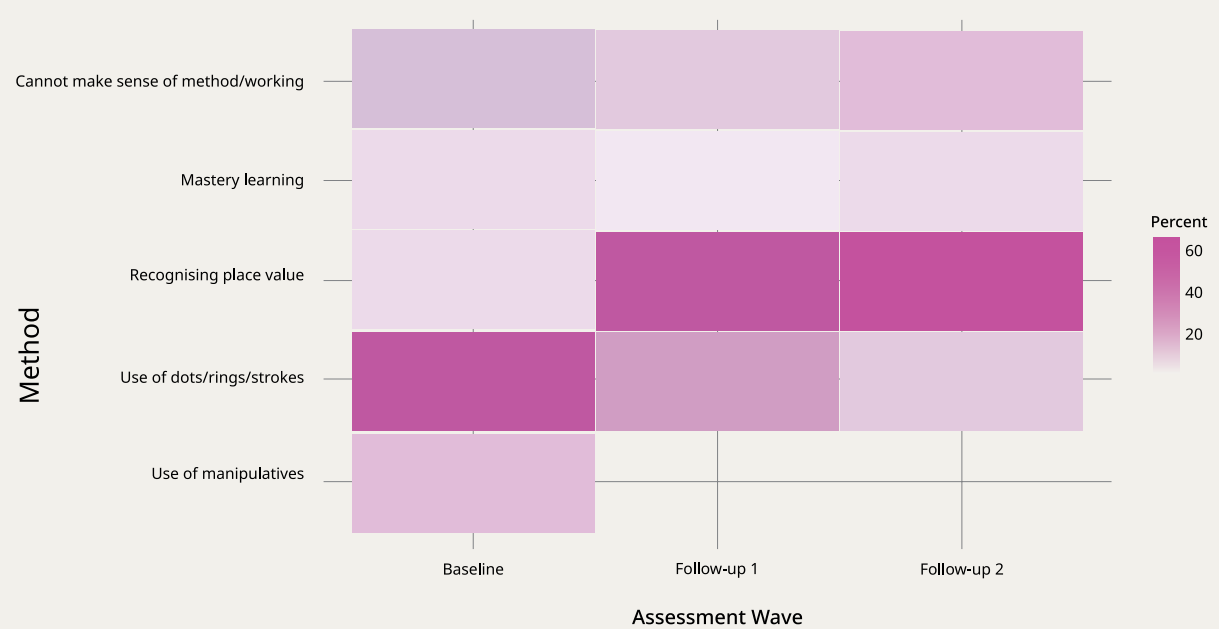


Figure 4: Method Pattern by Task – Addition

Interpretation

The longitudinal pattern demonstrates a developmentally appropriate progression from concrete counting strategies to increasingly abstract, place-value-based reasoning. The large baseline reliance on dots/rings/strokes suggests initial dependence on manipulatives and one-to-one correspondence, common at early stages of arithmetic learning, whereas the rapid and sustained rise in place-value recognition by Follow-up 1 and its consolidation by Follow-up 2 indicate effective translation of instruction into internalised strategy use for most learners. Sex-disaggregated differences at baseline (higher female reliance on concrete methods) narrowed as both sexes adopted place-value approaches, although females showed marginally greater adoption by Follow-up 2. The reduction in unclassifiable methods through Follow-up 1 suggests improved strategy fidelity following instruction. The modest rebound to 9.4 per cent unclassifiable at Follow-up 2 may reflect either occasional off-task responses, attempts at informal strategies not captured by the coding scheme, or variations in item difficulty that elicited nonstandard methods. Overall, the shift towards place-value strategies aligns with the observed increases in accuracy and the rise in procedural attempts; learners are using more advanced, formal strategies even when those strategies are not yet executed flawlessly.

Implications

These findings carry clear instructional and monitoring implications. First, the strong shift towards place-value reasoning validates curriculum and pedagogy that emphasise place-value understanding and suggests that such approaches should be maintained and scaled. Teachers should explicitly reinforce place-value concepts through a combination of manipulatives, pictorial representations, and guided transition to symbolic notation to ensure conceptual depth rather than rote procedure. Second, the near-elimination of concrete counting methods indicates that most learners can progress to abstract strategies within the study timeframe. However, the persistent minority using unclassifiable methods implies a need for targeted diagnostic assessment and small-group remediation to surface and correct idiosyncratic approaches. Third, sex-disaggregated monitoring of strategy adoption is warranted because baseline differences closed but did not disappear entirely. Differentiated instruction (for example, pairing, modelling, or scaffolded practice) may accelerate adoption among lagging subgroups. Fourth, the observed relationship between increased place-value use and rises in procedural errors suggests that professional development should equip teachers to scaffold procedural fluency after conceptual grounding, emphasising error analysis and formative feedback that converts attempted strategies into reliable performance. Finally, for future evaluation, the study recommends refining the strategy-coding scheme to capture nonstandard but systematic methods, reporting both transition matrices (to show individual-level shifts in strategy use) and links between strategy adoption and accuracy, and conducting classroom observations to identify teacher practices that most effectively promote the move from concrete to place-value reasoning.

4.3.2 Subtraction

Solution strategies for subtraction tasks followed a trajectory parallel to that observed for addition, reflecting a cohort-wide progression from concrete counting methods to abstract, place-value reasoning (*Table 7 and Fig. 5*). At baseline, concrete representations, dots, rings, or strokes, were the predominant strategy, accounting for 68.0 per cent of recorded methods overall (females 69.9%, males 65.0%). Place-value recognition was rare at baseline (6.1% overall), and 12.6 per cent of attempts were unclassifiable, indicating a non-trivial share of unclear or idiosyncratic approaches.

By Follow-up 1, the pattern shifted markedly. Place-value recognition rose to 68.3 per cent overall, with identical adoption rates across sexes (68.3% for both females and males). The use of concrete representations declined to 22.3 per cent, and the proportion of unclassifiable methods fell to 6.1 per cent, suggesting increased consistency in taught strategies and clearer evidence of conceptual uptake.

At Follow-up 2, place-value reasoning consolidated as the dominant subtraction strategy (82.5% overall; females 85.9%, males 77.2%), while concrete counting methods nearly disappeared (8.1% overall). Nevertheless, 8.4 per cent of attempts remained unclassifiable, mirroring the residual ambiguity observed in addition tasks and indicating that a minority of learners continued to rely on unclear or unconventional methods despite the cohort's overall strategic advancement. The

parallel evolution of strategies across addition and subtraction indicates a coherent shift in learners' mathematical thinking from concrete manipulation towards place-value understanding.

Table 7: Method Pattern by Task – Subtraction

Subtraction		Count			Percentage		
Wave	Method Type	Female	Male	Overall	Female	Male	Overall
Baseline	Number line	1	0	1	0.5	0	0.3
	Use of manipulatives	15	15	30	8.1	12.2	9.7
	Use of dots/rings/strokes	130	80	210	69.9	65	68
	Recognising place value	13	6	19	7	4.9	6.1
	Mastery learning	5	5	10	2.7	4.1	3.2
	Cannot make sense of method/working	22	17	39	11.8	13.8	12.6
Follow-up 1	Use of dots/rings/strokes	47	22	69	25.3	17.9	22.3
	Recognising place value	127	84	211	68.3	68.3	68.3
	Mastery learning	3	7	10	1.6	5.7	3.2
	Cannot make sense of method/working	9	10	19	4.8	8.1	6.1
Follow-up 2	Use of dots/rings/strokes	15	10	25	8.1	8.1	8.1
	Recognising place value	159	95	254	85.9	77.2	82.5
	Mastery learning	0	3	3	0	2.4	1
	Cannot make sense of method/working	11	15	26	5.9	12.2	8.4

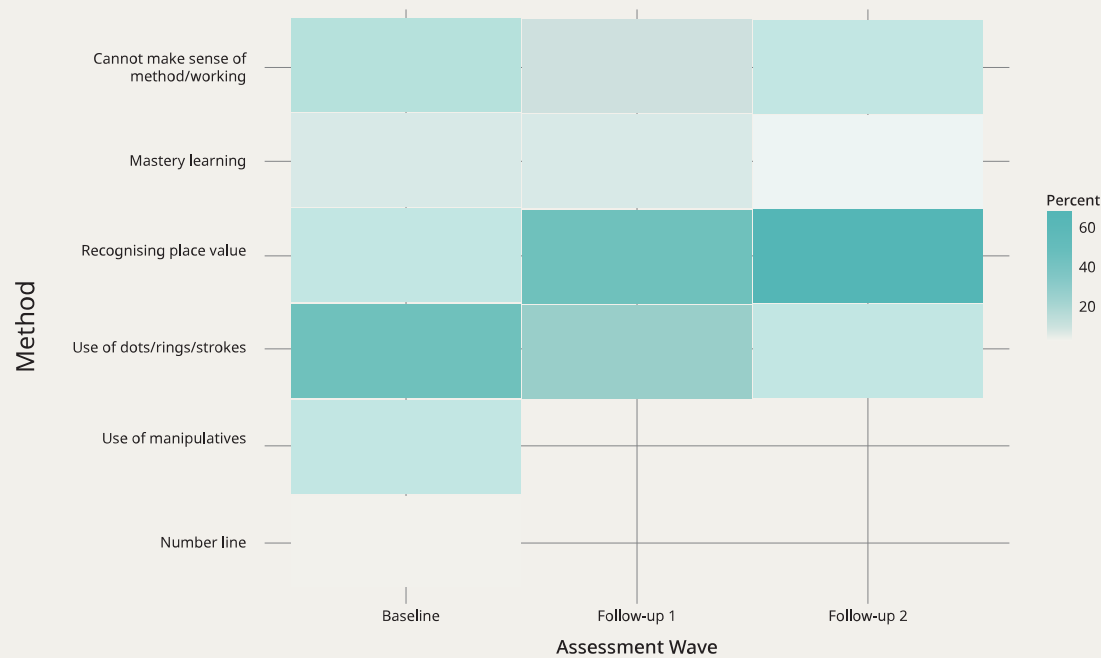


Figure 5: Method Pattern by Task – Subtraction

The shift from counting and drawing strategies to place-value-based methods is well supported by teacher observations. Educators described a visible evolution from heavy manipulative use to mental and written computation:

“Before, they draw sticks everywhere. Now they just write the place values and add or subtract” (Teacher 2_FGD_School 2);

“They began with counting one-by-one, but by end term they can regroup tens and ones. They understand the value” (Teacher 4_FGD_School 2).

These insights triangulate with Tables 6 and 7 and affirm the CRA learning trajectory documented in foundational numeracy pedagogy (Mix et al., 2022; World Bank, 2025).

Interpretation

The longitudinal shift from concrete to place-value strategies in subtraction reflects expected, developmentally appropriate learning progress. Learners moved from one-to-one counting supports to more abstract representations that capture the structure of two-digit subtraction. The near-identical, rapid adoption of place-value reasoning by Follow-up 1 suggests that instruction or practice activities during the study were effective in promoting conceptual change across both sexes. The slightly higher female adoption at Follow-up 2 (85.9% vs 77.2%) aligns with broader performance patterns in the cohort and may reflect differences in responsiveness to the instructional approach, engagement, or prior skill levels. The persistent, albeit reduced, proportion of unclassifiable attempts (8.4%) signals that some learners either resorted to idiosyncratic strategies not captured by the coding scheme or experienced difficulties transferring taught methods to specific item formats. Overall, the emergence of place-value strategies in tandem with increased correctness and modest rises in procedural attempts indicates that learners were attempting more sophisticated solution methods, a positive sign of deepening mathematical cognition even where execution was imperfect.

Implications

These findings suggest several actionable priorities for instruction, assessment, and teacher development. First, the strong shift to place-value reasoning validates continued emphasis on explicit teaching of place-value concepts using layered representations (that is, manipulatives to pictorial to symbolic) and guided transition activities that bridge counting to abstract computation. Second, the remaining minority of unclassifiable methods highlights the need for targeted diagnostic assessment and small-group remediation to surface and correct idiosyncratic or non-transferable approaches. Third, sex-disaggregated monitoring of strategy adoption is warranted. While both sexes adopted place-value strategies rapidly, the slightly lower male consolidation at Follow-up 2 indicates opportunities for differentiated supports (for example, additional scaffolded practice, paired peer modelling, or teacher prompts) to ensure equitable mastery. Fourth, professional development should equip teachers to recognise and respond to the specific strategy trajectories described here, emphasising formative error analysis, modelling of procedural steps, and feedback that connects conceptual understanding to correct execution. Finally, for programme monitoring and research, the study recommends: (a) using transition matrices to track individual-level shifts in strategy use; (b) refining coding schemes to capture systematic but nonstandard methods; and (c) linking strategy adoption data to accuracy and error-type changes to identify which instructional practices most reliably convert strategic sophistication into consistent correct performance.

4.3.3 Summary of Methods – Both Addition and Subtraction

The methods learners employed across addition and subtraction tasks exhibited a pronounced and coherent progression from concrete counting to abstract place-value reasoning. At baseline, concrete representations, dots, rings, or strokes, were the dominant approach in both operations (Addition: 67.3%; Subtraction: 68.0%), while explicit evidence of place-value recognition was minimal (Addition: 5.8%; Subtraction: 6.1%). Between baseline and Follow-up 1, a rapid transition occurred: place-value recognition became the primary strategy for both addition (67.2%) and subtraction (68.3%), with adoption rates roughly equivalent across sexes. This trend continued through Follow-up 2, where place-value strategies consolidated as the overwhelmingly dominant method (Addition: 81.2%; Subtraction: 82.5%). Females showed modestly higher rates of adoption at the final assessment (Addition: 84.9%

females vs 75.6% males; Subtraction: 85.9% females vs 77.2% males). Correspondingly, the use of concrete counting methods declined sharply to below 8 per cent for both operations by Follow-up 2. Despite this widespread shift towards more sophisticated strategies, approximately 9 per cent of attempts remained unclassifiable at the final wave, indicating a persistent minority of learners who continued to rely on unclear or nonstandard approaches.

Summary Insight

In summary, the findings demonstrate a profound and positive developmental shift from concrete to abstract problem-solving strategies in both addition and subtraction. This evolution signifies not just procedural improvement but deepening conceptual understanding of number structure and place value. Learners’ growing ability to apply symbolic reasoning reflects the effectiveness of structured, sequential teaching methods grounded in the CRA model. Sustaining such approaches and ensuring targeted support for those who lag in abstraction can significantly strengthen foundational numeracy outcomes, equipping learners with the cognitive tools necessary for advanced mathematical thinking (Mix et al., 2022; Byrne, 2023; World Bank, 2025).

4.4 School Performance by Task – Addition and Subtraction

School-level performance varied markedly across the study period, revealing divergent improvement trajectories despite uniformly low starting points. At baseline, proportion correct scores were low across all schools: addition ranged from 5 per cent to 16 per cent, and subtraction from 0 per cent to 9 per cent. School 3, the largest participating site (n = 40), recorded modest baseline accuracy (addition 8%, subtraction 5%).

By Follow-up 2, outcomes diverged substantially. School 1 achieved the highest overall gains, reaching 58 per cent correct on both addition and subtraction; this school also exhibited pronounced sex differences in final performance, with females substantially outperforming males (addition 73% vs 38%; subtraction 73% vs 38%). School 3 sustained steady improvement, attaining 55 per cent correct in addition and 53 per cent in subtraction at the final wave (See [Table 8](#) and [Figure 6](#)). In contrast, School 2 registered the weakest progress: at Follow-up 2, addition accuracy remained at 14 per cent and subtraction at 21 per cent, a pattern driven in part by persistently low female performance (reported as 0% correct in addition across waves). School 4 and School 5 demonstrated moderate overall gains; of note, School 5 female learners showed meaningful improvement (addition 50%, subtraction 40%), while male learners in that school lagged (addition 11%, subtraction 22%). These inter-school differences point to heterogeneity in learning gains that likely reflect variation in implementation fidelity, classroom practices, learner composition, and contextual factors.

Table 8: School Performance by Task – Addition and Subtraction

Wave / School	Addition Correct		Subtraction Correct		Overall Correct	
	Female	Male	Female	Male	Addition	Subtraction
Baseline – School 1	0%	38%	0%	0%	16%	0%
Baseline – School 2	0%	25%	10%	0%	7%	7%
Baseline – School 3	8%	7%	4%	7%	8%	5%
Baseline – School 4	0%	20%	17%	0%	9%	9%
Baseline – School 5	10%	0%	0%	0%	5%	0%
Follow-up 1 – School 1	36%	50%	36%	25%	42%	32%
Follow-up 1 – School 2	0%	50%	10%	0%	14%	7%

Wave / School	Addition Correct		Subtraction Correct		Overall Correct	
	Female	Male	Female	Male	Addition	Subtraction
Follow-up 1 – School 3	32%	40%	20%	33%	35%	25%
Follow-up 1 – School 4	17%	40%	0%	20%	27%	9%
Follow-up 1 – School 5	30%	22%	30%	22%	26%	26%
Follow-up 2 – School 1	73%	38%	73%	38%	58%	58%
Follow-up 2 – School 2	0%	50%	10%	50%	14%	21%
Follow-up 2 – School 3	60%	47%	52%	53%	55%	53%
Follow-up 2 – School 4	17%	20%	17%	40%	18%	27%
Follow-up 2 – School 5	50%	11%	40%	22%	32%	32%

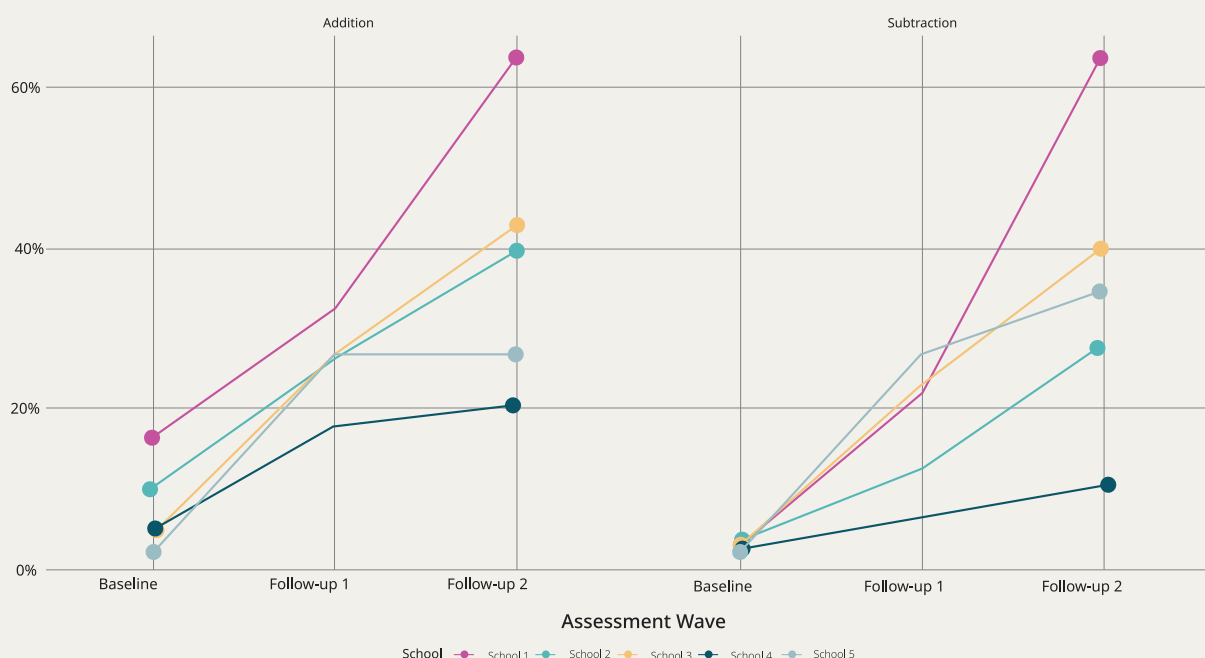


Figure 6: School Performance by Task – Addition and Subtraction

Differences across schools were also reflected in teacher discussions. While some schools reported consistent instructional routines, others indicated challenges with pacing, time, or support, helping explain varied school-level performance trends:

“Some schools are moving faster because teachers are very consistent. Others, they skip practice when there are other activities” (HoI 1_KII_School 2);

“Where the head supports and checks, results are better. If not, progress is slow” (HoI 3_KII_School 5).

These comments correspond with Table 8 outcomes, echoing global evidence that school-level instructional quality and leadership influence learner performance (de Barros et al., 2022; J-PAL & Pratham, 2022).

Interpretation

The school-level patterns suggest that, while the intervention or instructional exposure produced substantial gains in some contexts, these effects were unevenly distributed. Schools 1 and 3 appear to have achieved great and consistent improvements, implying effective translation of instruction into learner outcomes in those sites. Conversely, the persistently low performance observed in School 2, especially the reported 0 per cent female addition score across waves, is anomalous and raises concerns about localised barriers (for example, problems with implementation fidelity, teacher absenteeism, atypical learner composition, or data-recording errors). The pronounced sex-disaggregated contrasts within some schools (notably School 1 and School 5) indicate that site-specific factors interact with learner sex to shape outcomes. These could include differences in classroom dynamics, teacher expectations, remedial supports, or out-of-school influences. Because sample sizes vary by school (School 3 is the largest), observed school-level estimates may be influenced by sampling variability. Thus, these descriptive patterns should be confirmed with appropriate statistical tests and adjustments for clustering, baseline differences, and sample size. Finally, the range of school trajectories demonstrates that programme effectiveness is context-dependent and that aggregated cohort-level gains can obscure substantial within-project heterogeneity.

Implications

These findings have several practical and monitoring implications. First, they call for targeted, school-level diagnostics. There is a need to conduct rapid audits in underperforming sites (especially School 2) to verify data quality and identify barriers, examine attendance, teacher presence and practice, access to materials, and learner demographics. Second, where high-performing schools (Schools 1 and 3) exist, carry out structured case studies or classroom observations to document effective practices and viable elements for replication (teacher strategies, scheduling, coaching, community engagement). Third, implement differentiated support. This entails allocating additional coaching, remedial groups, and resources to low-performing schools and considering sex-responsive strategies where female or male learners are disproportionately disadvantaged. Fourth, strengthening monitoring and evaluation by applying multilevel (hierarchical) analyses that account for clustering by school, report confidence intervals or standard errors for school estimates, and use transition matrices to track individual-level change. This will provide more robust evidence of where and why gains occurred. Finally, build a feedback loop between monitoring and implementation. This may be using school-level findings to inform in-service training, revise instructional materials where needed, and pilot context-specific adaptations with subsequent rapid re-assessment to confirm impact. Addressing both data integrity issues and contextual drivers will be critical to ensuring equitable learning gains across all schools.

In summary, while all schools recorded varying levels of progress, the patterns indicate that strong instructional leadership, teacher competence, and consistent feedback loops contribute significantly to learner success. High-performing schools illustrate what is achievable under effective implementation, while underperforming ones reveal areas needing focused intervention. These findings reaffirm global evidence that equitable learning outcomes in foundational numeracy require both pedagogical quality and systematic support structures at the school level (de Barros et al., 2022; World Bank, 2025; J-PAL & Pratham, 2022).

“Learner mistakes, when used as diagnostic tools, can drive instructional improvement and strengthen learning outcomes.”



5. Conclusion

The overall analysis across schools and assessment waves paints a compelling picture of genuine and progressive learning in foundational numeracy. Learners demonstrated consistent growth in both addition and subtraction, with dramatic declines in factual errors and parallel increases in correct responses. These improvements indicate that structured teaching, repeated practice, and focused feedback were effective in deepening computational fluency and conceptual understanding. The transition from concrete methods, such as manipulatives and drawings, to abstract, place-value-based reasoning further reflects the learners' cognitive development and mastery of core arithmetic concepts. By the end of the learning period, most learners were confidently applying mental and symbolic strategies, demonstrating readiness for more advanced mathematical thinking.

Despite this encouraging progress, the findings also revealed notable disparities in performance across schools. Some schools, particularly Schools 1 and 3, showed steady and sustained growth, while others exhibited inconsistent trajectories, stagnation, or even declines in certain operations. These differences highlight the critical role of instructional quality, teacher preparedness, and school-level support systems in sustaining learning gains. The data reinforce the idea that effective pedagogy—supported by consistent monitoring, teacher feedback, and targeted coaching—can close learning gaps and ensure more equitable outcomes. Overall, the results confirm that systematic, data-driven, and conceptually grounded instruction significantly enhances foundational numeracy when implemented with fidelity across classrooms and schools.

6. Recommendations

Drawing from the findings and conclusion, the following recommendations are proposed to consolidate gains, address gaps, and promote sustainable improvement in foundational numeracy outcomes. The recommendations are presented chronologically, following the logical flow of learning, teaching practice, and system-level strengthening:

A. Strengthening Learner-Level Competence

1. **Reinforce Foundational Arithmetic Fluency:** Continue daily structured practice sessions focusing on number bonds, addition, and subtraction fluency. Such activities help learners internalise basic facts and reduce factual errors, freeing cognitive capacity for higher-order reasoning (Byrne, 2023).
2. **Sustain the Concrete-Representational-Abstract (CRA) Progression:** Teachers should consistently use the CRA model to guide learners from hands-on manipulation to mental and symbolic reasoning. This approach ensures smooth cognitive transition and prevents procedural errors during abstraction (Mix et al., 2022).
3. **Provide Individualised Support for Struggling Learners:** Learners whose work remains unclear or unscorable should receive targeted remediation. Teachers can use diagnostic tasks and visual aids to strengthen conceptual understanding before expecting full abstraction (Harnett, 2018).

B. Enhancing Instructional Quality and Teacher Support

4. **Deepen Teacher Professional Development:** Continuous capacity-building programmes should focus on numeracy pedagogy, error analysis, and the use of formative assessment data. Empowering teachers to identify and address learner misconceptions early promotes lasting improvement (de Barros et al., 2022).
5. **Institutionalise Peer Learning and Coaching:** High-performing schools, such as Schools 1 and 3, should be leveraged as demonstration centres. Facilitated inter-school exchanges and peer observations can spread effective instructional practices and reduce performance disparities (J-PAL & Pratham, 2022).

6. **Integrate Continuous Data Monitoring and Feedback:** Schools should conduct regular progress reviews to analyse learner outcomes, track growth patterns, and adjust instructional plans accordingly. Such iterative use of data fosters accountability and timely intervention (World Bank, 2025).

C. Promoting Equity and Sustainability at System Level

7. **Address School-Level Performance Gaps:** Education officers should prioritise targeted support to underperforming schools through structured coaching visits, provision of instructional materials, and strengthening school leadership for learning (de Barros et al., 2022).
8. **Develop and Implement Longitudinal Tracking Systems:** Sustained learner tracking across terms and years is vital for assessing true learning progress. Adoption of simple learner identifiers can help monitor continuity and reduce data fragmentation (Zhan et al., 2020).
9. **Ensure Consistency Beyond Mid-Year Gains:** Learning momentum should be maintained throughout the academic year. Mid-year refresher training, peer lesson study groups, and learner re-engagement activities after breaks can help sustain improvement (Byrne, 2023; World Bank, 2025).

D. Policy and Programmatic Integration

10. **Mainstream Data-Driven Instructional Planning:** The use of detailed error analysis and performance tracking should be institutionalised at the county and national levels. Embedding these practices within continuous professional development frameworks will strengthen the Competency-Based Curriculum (CBC) implementation and promote foundational numeracy mastery (World Bank, 2025).

In essence, the evidence points to a learning system that is improving but still uneven. Learners are advancing from surface-level computation to conceptual reasoning, yet sustained gains depend on teacher competence, consistent monitoring, and equitable support across schools. Implementing these recommendations will not only consolidate the current progress but also create a strong foundation for scaling effective practices in early-grade numeracy nationwide.

7. References

- ASER. (2022). *Annual Status of Education Report (Rural) 2022*. ASER Centre.
- Brookfield, S. D. (2017). *Becoming a critically reflective teacher* (2nd ed.). Jossey-Bass.
- Byrne, E. M. (2023). *Educational interventions involving physical manipulatives for early maths learning: A review*. BERA Reviews.
- Centre for Mathematics, Science and Technology Education in Africa (CEMASTEA) & Zizi Afrique Foundation (ZAF). (2025). *Error analysis in foundational numeracy: A teacher-led intervention in Bungoma County*. Unpublished report.
- de Barros, A., et al. (2022). *Large-scale efforts to improve teaching and child learning: Evaluation insights*.
- Harnett, J. (2018). *Teaching place-value: A teacher knowledge perspective*. Educational Research Archive.
- J-PAL & Pratham. (2022). *Teaching at the Right Level: Evidence and case studies*. Abdul Latif Jameel Poverty Action Lab.
- Kenya National Examinations Council. (2023, January). *2022 KCSE Examination Essential Statistics*. Nairobi: Kenya National Examinations Council. <https://knec.ac.ke/wp-content/uploads/2023/01/2022-KCSE-EXAMINATION-ESSENTIAL-STATISTICS.pdf>
- Mix, K. S., et al. (2022). *The development of place value concepts: Longitudinal evidence*. Frontiers in Psychology.
- National Assessment of Educational Progress (NAEP). (2021). *The Nation's Report Card: Mathematics 2021*. U.S. Department of Education, Institute of Education Sciences.
- SACMEQ. (2017). *The Southern and Eastern Africa Consortium for Monitoring Educational Quality IV Project Results: Pupil achievement levels in literacy and numeracy*. SACMEQ Secretariat.
- Schön, D. A. (1983). *The reflective practitioner: How professionals think in action*. Basic Books.
- UNESCO. (2021). *Global education monitoring report 2021/2: Non-state actors in education – Who chooses? Who loses?* UNESCO Publishing.
- UNICEF. (2020). *Learning outcomes in Southern and Eastern Africa: A regional perspective*. UNICEF Regional Office for Eastern and Southern Africa.
- Usawa Agenda. (2023). *Are ALL our children learning? 1st Foundational Literacy and Numeracy Assessment (FLANA) Report 2023*. Nairobi: Usawa Agenda. https://usawaagenda.org/wp-content/uploads/2024/05/UA_Are-ALL-Our-Children-Learning_Report-03-2023-1.pdf
- Uwezo. (2021). *Are our children learning? Uwezo learning assessment report for Kenya and Uganda*. Twaweza East Africa.
- World Bank. (2021). *The state of global learning poverty: 2021 update*. World Bank Group.
- World Bank. (2022). *World Development Report 2022: Learning recovery to accelerate progress in education*. World Bank Group.
- World Bank. (2025). *Foundational learning: Core numeracy and literacy as essential skills*. World Bank Education Brief.
- Zhan, P., et al. (2020). *Longitudinal learning diagnosis: Mini-review and future directions*. Frontiers in Psychology.



Connect
With Us



Le'Mac Building, 1st Floor,
Nairobi - Kenya



info@mizizielimu.org



+254 734 620 181