



Sierra Leone National Early Grade Reading and Mathematics Assessment Baseline Study

November 2021



CONTENTS

ACKNOWLEDGEMENTS	4
FORWARD	5
ABBREVIATIONS	1
EXECUTIVE SUMMARY	2
1. INTRODUCTION	5
1.1. Background to the study.....	5
1.2. Purpose of the study.....	5
1.3. Use of findings	6
2. SURVEY METHODOLOGY	6
2.1. Sampling Framework.....	6
2.1.1. Quantitative Sampling	6
2.1.2. Qualitative Sampling	7
2.2. Data collection tools.....	8
2.3. Data collection	10
2.4. Analytical framework	11
2.5. Challenges, limitations and technical considerations.....	14
2.6. Research ethics and ethical considerations due to COVID-19.....	15
3. FINDINGS ON LEARNERS	15
3.1. Learner characteristics.....	16
3.2. Early Grade Reading Assessment.....	20
3.3. Early Grade Math Assessment.....	23
3.4. Additional findings	26
4. SURVEY COHORT EQUIVALENCY AT BASELINE	28
5. LEARNER PROFILES AND PROGRESS TRACKING 2014 TO 2021	33
5.1. Learner profiles for reading	34
5.2. Learner profiles for numeracy.....	38
5.3. Progress tracking 2014 to 2021.....	39
6. FINDINGS ON TEACHERS	44
6.1. Teacher workforce profile and management.....	44
6.2. Professional development and literacy instruction	45
6.3. Teacher expectations and learner performance	47
6.4. Lesson observations.....	47
7. FINDINGS ON SCHOOLS	55
7.1. Learner enrolment and attendance	55
7.2. Teacher assignment and attendance	56
7.3. Teaching and learning materials.....	56
7.4. Water, hygiene and sanitation facilities	60
8. CONCLUSIONS AND RECOMMENDATIONS	62
8.1. Learners.....	62
8.2. Teachers	63
8.3. Schools	63
9. ANNEXES	65
9.1. Population and sample distribution per district	65
9.2. Detailed study methodology.....	66

9.2.1. Quantitative sampling – schools and learners	66
9.3. EGRA/EGMA	67
9.4. The World Bank Teach tool.....	69
9.5. Data collection tools	69
9.6. Revised work plan	70
9.7. Research ethics.....	71
9.8. Ethical considerations due to COVID-19	71
9.9. Additional tables and graphs	72
9.9.1. Data collection summary and respondent sample.....	72
9.9.2. Challenges and mitigation measures during fieldwork.....	74
9.10. Reliability analysis.....	77
9.10.1. English EGRA tool analysis.....	77
9.10.2. EGMA tool analysis	78
9.11. Additional regression analysis.....	81

ACKNOWLEDGEMENTS

This report would not have been possible without the commitment and efforts of the members of staff at the Ministry of Basic and Senior Secondary Education (MBSSE) and particularly the Technical Steering Committee led by the Teaching Service Commission (TSC) who offered ministerial oversight during the delivery of this National Early Grade Reading and Mathematics Assessment Baseline Study.

MBSSE would like to thank UNICEF (grant agent) through whom Global Partnership for Education (GPE) have provided financial support for the completion of this important study.

I would also like to thank other members of the Technical Steering Committee at the MBSSE that include the Directorate of Planning and Policy, the Directorate of Quality Assurance, the Change Unit, Pre-service Teacher Training Institutions and the GPE Coordinator.

I would also like to sincerely thank Montrose International for conducting this Early Grade Reading and Mathematics Assessment Baseline Study.

Dr Yatta Kanu
Chief Education Officer
Ministry of Basic Senior Secondary Education

FORWARD

The Ministry of Basic and Senior Secondary Education (MBSSE) has conducted this national reading and mathematics baseline assessment to identify learners' competencies in key foundational skills in literacy and numeracy as well as the school environment conditions that are likely impacting their ability to learn. The overall objective of the baseline assessment is to measure and document the current state of literacy and numeracy skills among children in early primary grades one-three.

This has been possible with a USD 17.2 million-dollar grant from GPE through a comprehensive programme in Sierra Leone that supports sector priorities as outlined in the Education Sector Plan 2018-2020. The GPE programme aims to increase equitable access to education and improved learning outcomes for boys and girls in pre-primary and primary schools. The components of the programme include: 1) Expansion of Early Childhood Development(ECD) opportunities, 2) In-service training of primary teachers to improve early grade reading and mathematics skills, provision of teaching and learning materials and the establishing a learning assessment unit, and 3) Decentralised and integrated school monitoring, data collection and management.

The GPE grant focused on improving early grade reading and math skills for approximately 900,000 children. It also aimed to increase access to early childhood development opportunities for 10,000 children, train up to 17,000 primary school teachers in early grade reading and mathematics instruction and provide teaching and learning materials.

In addition, the funding aimed to improve decentralised monitoring, data collection and management and establish a learning assessment unit within the Ministry of Basic and Senior Secondary Education (MBSSE). These proposed priority interventions were aligned to the government's free quality school education policy.

Through these interventions, we as the MBSSE will continue to progress towards developing an appropriately educated and internationally competitive citizenry for Sierra Leone by 2030.



Dr David M. Sengeh
Minister of Basic and Senior Secondary Education

ABBREVIATIONS

ASC	Annual School Census
COVID-19	Coronavirus disease of 2019
EGMA	Early Grade Mathematics Assessment
EGRA	Early Grade Reading Assessment
FDG	Focus Group Discussion
GPE	Global Partnership for Education
INGO	International Non-Government Organisation
KII	Key Informant Interview
MBSSE	Ministry of Basic and Senior Secondary Education
MoHS	Ministry of Health and Sanitation
NGO	Non-Government Organisation
PPE	Personal Protective Equipment
PTA	Parents Teachers Association
QA	Quality Assurer
SC	Technical Steering Committee
SL	Sierra Leone
SRS	Simple Random Sampling
StatsSL	Statistics Sierra Leone
TL	Team Leader
TOR	Terms of Reference
TSC	Teaching Service Commission
UNICEF	United Nations Children’s Fund
WPM	Words per minute

EXECUTIVE SUMMARY

This assessment aimed to provide baseline Global Partnership for Education (GPE) programme indicators for assessing and monitoring children's reading and mathematics skills. By using the Early Grade Reading and Mathematics Assessment (EGRA/EGMA), the study sought to establish the state of children's developmental reading skills and captured children's ability on number identification, discrimination, patterns, and subtraction. Furthermore, the assessment gathered information on factors in schools, homes and classrooms that affect children's ability to become literate and numerate.

The study was overseen by a learning assessment Technical Steering Committee (SC) led by the Teaching Service Commission (TSC). Although initially commissioned in 2019, this study was postponed and eventually conducted in 2021 to align to the GPE outcome level results framework (measuring increased learning outcomes in the early grades by assessing learners at the end of academic year 2 and 4) and the interruption of the 2020 COVID-19 pandemic shut down.

A mixed-methods approach comprising both qualitative and quantitative data collection methods was used for this study. Overall, the study collected data from 260 schools across 16 districts of Sierra Leone at the end of the current academic year (2 June to 22 July 2021)¹ to best capture evidence on the end of year reading and mathematics proficiency of children in grades 2 and 4. In line with the data collection plan 4,729 EGRA/EGMA learner assessments were administered to learners in grades 2 and 4 while 435 lessons were observed, and 84 lessons were video recorded for further analysis.

Additionally, 476 teachers (one English and one maths teacher in each school), 244 head teachers and 4,729 learners were interviewed using semi-structured survey tools. Across all schools, 260 school observation checklists, 264 grade rosters and 242 employee rosters were completed, 41 focus group discussions were held with Parent Teacher Association (PTA) members while five key informant interviews (KIIs) were held with district education officers.

The study followed the 'UNEG Ethical Guidelines for Evaluations' and the 'UNICEF Procedure for Ethical Standards in Research, Evaluation, Data Collection and Analysis'. Additionally, the study also employed guidelines and protocols aimed at preventing the spread of COVID-19 while conducting fieldwork that were in line with standard operating procedures from Ministry of Health and Sanitation (MoHS).

Findings on learners

Many learners are an average of two years older than they should be in both grade 2 and grade 4, which contributes to overcrowding and an early grade bulge that adds cost and wastage to the system. Also, only one in four learners attended nursery school before starting primary school. The study found that learners needed more practice with beginning literacy skills that precede whole word reading as the largest percentage of non-readers was registered on the letter-sound correspondence (54 per cent) and decoding subtasks (non-word reading – 66 per cent). Overall, learners performed better in listening comprehension than in reading comprehension.

For EGRA, results of interest included the average of correct letter or words per minute (WPM), broken down by grade or other factors. For EGMA, the number of right or wrong answers per minute were measured. Learners performed well in number identification and quantities discrimination with fewer than 21 per cent zero scores across both subtasks, demonstrating competence in key procedural maths skills which must be mastered in the early grades as they are critical to developing more advanced

¹ Third term begins immediately after the Easter break (19 April 2021) and runs for 13 weeks up to 16 July 2021. The mid-term break is 28-31 May 2021.

maths skills. Learner profiles for oral reading fluency indicate that learners are progressing as readers from grade 2 to grade 4, with the share of non-readers in grade 4 reduced significantly from that in grade 2. Furthermore, there was a statistically significant gender differences in performance between boys and girls with boys outperforming their female counterparts on nearly every subtask in EGRA and EGMA in both grades.

Comparing zero scores across learner cohorts, even in different geographies and languages, is possible because 'zero' has the same weight in every system. There was significant improvement in performance from the first assessment in 2014 to the one administered in 2021, with a strong reduction in non-readers between years. Subtasks in the EGMA also followed a similar pattern in zero scores, with higher numbers of zero scores in the 2014 study than in the 2021 survey, and grade level reductions in zero scores both times. When comparing zero scores internationally, Sierra Leone falls in the upper half, demonstrating fewer zero scores on the oral reading fluency test than other countries, especially those administering English language assessments.

Learning during COVID-19

A total of 87 per cent of the learners surveyed reported listening to the radio teaching programme launched by the government nationally during COVID-19, while 17 per cent tried to continue learning by watching the television teaching programme. Although most learners only did this occasionally, these findings demonstrate a potentially high level of reach from these inputs across the country that is useful to explore further when developing next generation distance or blended learning programmes.

Findings on teachers

Findings indicate that the upgrading and certifying of in-service teachers through targeted training will be critical to ensure continued supply of skilled teachers in the system with nearly 25 per cent of the teachers surveyed have only a basic education or West African senior school certificate and hence lacked certified training to become a teacher. The overall picture is that of an experienced work force with teachers found to have worked an average of 8.6 years in their current school and 70 per cent of uncertified teachers have been working for 6 years or less. Positively, close to 40 per cent of the teachers surveyed received professional development and training within the last two years however, explicit training in literacy instruction is urgently needed for teachers to improve reading of learners.

Just 35 per cent of teachers had been trained on how to teach letter sounds and could only pronounce three out of five letter sounds when asked during the interview. The study also found that teachers generally professed low expectations of learner performance in literacy and numeracy across the early grades. Additionally, teachers did not expect learners to be school-ready when they begin primary as they reported not expecting incoming P1 learners to demonstrate any emergent literacy or numeracy skills.

In the lessons observed, findings reveal that almost a quarter of the lesson time goes unused as focus gets lost over time, especially during the last 10 minutes of the class. In literacy classes, it was observed that English was spoken about 70 per cent of the time, and its use decreased as the class progressed. Although teachers displayed positive practices during their lessons, these positive actions significantly diminish in the last 10 minutes of the lesson. Also, teachers do not appear to have prepared for the lesson in about 40 per cent of cases for grade 2 classes and about 30 per cent for grade 4 classes.

Findings on schools

Enrolment was found to diminish over time by grade, for instance P6 had, on average, half of the total number of learners enrolled as in P1 (for both boys and girls). It was also observed that there was a

large discrepancy between learners enrolled, and learners present at school with boys across all grades, having an average attendance rate of 53 per cent and girls 54 per cent. Recruitment and assignment of teachers to posts should reflect the needs and size of the school as there was on average, slightly more than one teacher assigned per grade.

The study also found that teachers' attendance was high at an average of 89.9 per cent across all schools while 15.7 per cent of the numeracy classes and 14.1 per cent of literacy classes were assisted by a co-teacher. In line with government's policy, most schools had basic materials available for teaching core subjects, including maths textbooks (95 per cent) and ESP&S (letter writing and comprehension) textbooks (84 per cent). Very few schools possessed all the water, hygiene and sanitation (WASH) facilities required as per national guidelines; for instance, 22 per cent of schools had no latrines for learners while 40 per cent of schools had no latrines for teachers.

Recommendations

1. Ensure that the goal of the national literacy programme is to reduce non-readers within each grade and between grades and grow the proportion of capable readers who can comprehend text, both within and between grades. Moving clusters of learners from lower performance profiles (e.g. non-readers) to higher ones (e.g. developing readers) provides a basis for measuring true system-level change, rather than narrowly focusing achievement on singular oral reading fluency scores.
2. The EGMA results suggest that learners' experience of mathematics instruction is more about memorisation of facts and rules than development of strategies to find answers to problems. Therefore, conceptual mathematics must be explicitly taught as mastering these foundational concepts is directly predictive of future performance.
3. Upgrade and certify in-service teachers through targeted training to ensure there continues to be a consistent, immediate supply of skilled teachers in the system to meet current demand.
4. Teachers must learn how children acquire literacy and numeracy and the key competencies they must progressively demonstrate in the early grades to become fluent readers. Additional training and support in teaching early grade literacy and numeracy will ensure teachers acquire this knowledge.
5. Conduct further investigation into the national teacher workforce management strategy to ensure appropriate recruitment and assignment of teachers, and the possible reallocation and distribution of certified teachers already in the system to meet needs in underserved areas.
6. To counteract learners attending classes at the wrong age level, alternative education programmes for overage youth who leave formal basic education must be made available and accessible to all these learners. Additionally, ensuring learners start school at the right time (age 6) and with the right foundation (ideally gained through early childhood development programmes or nursery school) will reduce stress on the system and reduce repetition in the early grades.
7. Support teachers to plan and prepare their lessons and to pace their teaching to utilise the entire class period for a lesson. Ensure schools and classrooms have the teaching and learning materials required for effective instruction.

1. INTRODUCTION

1.1. Background to the study

Enrolment rates over the last few years in Sierra Leone have improved, but challenges of low and inequitable access to school persist across the system. While most children countrywide enter primary school, many of them drop out of school before completing even a basic education cycle. The COVID-19 crisis has put substantially more children at an even higher risk of remaining out of school, particularly girls, children with disabilities, and children from low-income households.

The provision of quality education is an essential factor in accelerating the future growth and development of Sierra Leone. Improving access to and completion of at least a basic education cycle for all children is essential, as is strengthening the ability of the sector to meet and fulfil the growing need for skilled labour in the workforce.

In the post-Ebola period, the Ministry of Basic and Senior Secondary Education (MBSSE) successfully demonstrated that rapid sector improvements are possible. These lessons can be applied to today's context to address the underlying challenges and constraints in the education system needing urgent attention to radically transform learning outcomes for children.

To do this, the MBSSE must generate and examine evidence on children's learning; the environments where they learn; the resources they learn with; the qualities and abilities of their teachers; and the mechanisms and functioning of the system responsible for delivering their education. This baseline survey was commissioned to respond to some of these evidence needs, particularly regarding learner learning and teacher effectiveness.

1.2. Purpose of the study

This assessment aimed to provide the Government of Sierra Leone and partners baseline Global Partnership for Education (GPE) programme indicators for assessing and monitoring children's reading and mathematics skills. It also intended to provide evidence to inform programme strategies and activities under the GPE programme, which aims to improve early grade reading and mathematics instruction through supporting in-service training for primary school teachers.

More specifically, the study aimed to assess the current state of early grade reading and mathematics skills among children in primary grades 2 and 4 in English. This was done to identify learners' competencies in key foundational skills in literacy and numeracy as well as the school, classroom and instructional conditions that are likely impacting – either positively or negatively – their ability to learn.

By using the Early Grade Reading and Mathematics Assessment (EGRA/EGMA), the study sought to establish the state of children's developmental reading skills – e.g., phonemic awareness, alphabetic principle, vocabulary, reading fluency and reading comprehension – through a series of timed subtasks that measured each competency. The mathematics assessment captured children's ability on number identification, discrimination, patterns, and subtraction. Furthermore, the assessment also gathered information on factors in schools, homes and classrooms that affect children's ability to become literate and numerate.

The study was overseen by a learning assessment Technical Steering Committee (SC) led by the Teaching Service Commission (TSC) and comprising 13 members from the Directorates of Planning and Policy, and Quality Assurance at MBSSE, the Change Unit, GPE Coordinator, representatives from universities and pre-service teacher training institutions and UNICEF. Although initially commissioned

in 2019, this study was postponed and eventually conducted in 2021 to align to the GPE outcome level results framework (measuring increased learning outcomes in the early grades by assessing learners at the end of academic year 2 and 4) and the interruption of the 2020 COVID-19 pandemic shut down.

This report comprises six main sections including an introduction, survey methodology, findings, conclusions, recommendations and appendices. The findings section is further categorised by learner, teacher and school environment with learners grouped into learning profiles based on their performance.

1.3. Use of findings

The EGRA and EGMA findings support achievement of priorities in Sierra Leone's Education Sector Plan 2018-2020 – *to improve literacy and numeracy at primary level*.² These priorities are supported by a GPE-funded comprehensive programme which aims to increase equitable access to education and improve learning outcomes for boys and girls in pre-primary and primary schools.³

Results from this assessment are aimed at informing MBSSE plans to improve the pre-service training of new teachers so they can effectively teach early grade literacy and numeracy, as well as inform inputs to improve classroom practices in general and teachers in particular. The assessment also aimed to develop a set of recommendations for MBSSE that can be used for a nuanced approach to teacher-support based on the specific needs and conditions of the school and the varying educational priorities of systems in different geographical settings.

The target audience for this report include UNICEF's Education section, the TSC, Directorates of Curriculum and Research, Planning and Policy, School Quality Assurance and Resource Management of the MBSSE and other government and non-governmental education sector partners.

2. SURVEY METHODOLOGY

A mixed-methods approach comprising both qualitative and quantitative data collection methods was used for this study. Qualitative data was collected through key informant interviews (KIIs), lesson observations and focus group discussions (FGDs). Qualitative data was triangulated with quantitative data collected through EGRA and EGMA and other semi-structured interviews including teacher interview, head teacher interview and learner context interview. Other quantitative data was collected using a school observation checklist, and a grade and employee roster.

2.1. Sampling Framework

2.1.1. Quantitative Sampling

School Selection

The sampling frame was generated using the 2019 Annual School Census (ASC) list of 7,154 primary schools that was provided by UNICEF/MBSSE as the 2020 ASC data set was not ready for use. For this study, 260 schools were drawn using proportion to size allocation with simple random sampling without

²Ministry of Education Science and Technology, Education Sector Plan 2018-2020, accessed from <<http://www.education.gov.sl/PDF/Slider/ESP-2018-2020-V6-ONLINE.pdf>>, p.20, on 15/09/2019.

³ EGRA/EGMA Terms of Reference

replacement in each stratum.⁴ The population comprised 16 districts and five school ownership categories (community, government, mission/religious, private and others) forming a total of 63 strata. The sample was drawn from a population of 4,597 primary schools that met the predetermined criteria of having an enrolment of at least 30 learners in grade 1 in the 2019 and an active grade 4 in 2019 (enrolment not zero). Drawing from other EGRA/EGMA studies conducted by Montrose in the past, a backup sample of up to 78 schools (30 per cent of the sample population) with which enumerator teams could replace sample schools was also drawn.

In the distribution of sampled schools by ownership, the majority of sampled schools were owned by mission/religious groups (62.7 per cent, n=163), followed by government-owned schools at 18.5 per cent (n=48). Additionally, in school distribution by district, the majority of the sampled schools (54 per cent) were found in Bo, Kambia, Kenema, Kono, Port Loko and Kailahun districts. Refer to **annex 9.1** for a detailed table of the population and sample drawn per district and to **annex 9.2.1** for more details on how the school sample was generated.

Learner Selection

Systematic random sampling was undertaken to select the required number of learners in each grade (10 learners). The enumerator team leader was required to randomly select 10 learners of proportional gender representation (5 boys and 5 girls). Pupils were drawn from 2 grades per school, grade 2 and grade 4. If multiple streams were found in the school, then both (or all three) streams were included in the sampling frame, by lining up boys and girls from all streams in a single line and then the prescribed sampling procedures followed. Once the sample was drawn, enumerators began conducting the assessment with the learners.

This brought the total sample of learners to 20 learners per school, totalling 5,200 learners in 260 schools. Even if the total sample of learners was not met due to low enrolment and attendance in grades 2 and 4 on the day of the visit, the findings would remain reliable and representative given that sampling calculations were based on the school as the unit of selection, rather than the learner. Refer to **annex 9.2.1** for more details on how learners were sampled within schools.

Teacher Selection

Purposive sampling was used to select the teachers targeted for the study such that only Math and English grade 2 and 4 teachers were selected. In each school only 2 teachers were randomly selected (one in grade 2 and the other in grade 4) to participate in the study. Randomisation of schools was used to identify the subject teacher in either grade that would be interviewed and observed by the enumerator teams during the school visit. Enumerator teams were given daily data collection plans to guide this process.

2.1.2. Qualitative Sampling

Qualitative data was collected at district and school levels and was used to triangulate findings derived from quantitative data by providing additional justification or explanations to the findings.

⁴ Since EGRA/EGMA assessments are designed to measure the national level of literacy and numeracy of school learners, large districts that serve more learners are more likely to have greater proportional influence on the overall scores. This sampling ensured the closest possible estimate of sample results to that of the total number of schools recorded during the last school census.

Key Informant Interviews

Key informants were identified purposively with participants selected from institutions with a mandate to deliver on education at the district level. These key informants included the District Deputy Director of Education and/or District School Inspector for each of the 16 districts that were visited. Interviews were conducted based on participants' availability and depending on their willingness to take part in the study. Consent was always obtained from all participants to permit the interviewer to record the sessions for proper note taking afterwards.

Focus Group Discussions

Focus group discussions (FGDs) were held with Parent Teacher Association (PTA) members of pre-selected sample schools. Each FGD comprise five participants who were selected based on convenience, their availability and willingness to participate in the study. Through a telephone call from the enumerator team leader, head-teachers were notified ahead of the school visit to mobilise five of the PTA members for the discussion. Two FGDs were to be organised in each of the 16 districts in each round of data collection, resulting in a total of 64 FGDs overall.

2.2. Data collection tools

EGRA and EGMA

Montrose maintained the subtasks used during the 2014 EGRA/EGMA study as these were previously agreed upon between UNICEF, MBSSE and the steering committee that was convened to oversee the implementation of that 2014 study. Whilst these tests were contextualised to Sierra Leone, the overall difficulty or level was designed to be consistent with the appropriate grades and best practice.

The EGRA and EGMA subtasks were designed to test reading and mathematics skills in a progressive fashion, approaching functional levels of literacy and numeracy in the final subtasks. However, it should be noted that this does not imply that learners were tested at exact grade appropriate curricular standards, as the purpose of the assessments was to measure key foundational skills that are predictive of later academic success. Performance on these skills can be grouped into achievement bands, which provide a picture of the learning profiles present among children in each grade, or across more than one grade.

Grouping and tracking learners using learning profiles provides a clearer picture of progressive gains and overall performance, given that the stated goal of this instruction is comprehension of a text (literacy) or application of a strategy to solve a problem (numeracy). Given this information, the EGRA and EGMA tools were revised in the following ways:

1. One EGRA tool and one EGMA tool were created for the baseline, that targeted Primary 2 and 4 learners.
2. The assessment was **levelled at Primary 2** across all subtasks (except level 2 subtasks in each assessment, which were levelled at Primary 4) to ensure content was accessible to lower grade learners, as well as non-readers and emerging readers who took the assessment.

Refer to **annex 9.3** that contains a detailed narrative on the objectives of the tool changes, the benefits of these changes and a summary table of the changes made to each tool.

It was estimated that it would not take longer than 20 minutes for learners to take the EGRA, EGMA and participate in the learner interview. An additional five-minute period between each test allowed for enough time for preparation and the change-over between learners that were being tested. Each enumerator team of four enumerators was therefore able to assess 20 learners per day except for when schools were found to have fewer learners enrolled in the target classes.

Learner Interview

The learner interview included a combination of closed and open-ended questions with in-built marking categories designed where possible to standardise responses. The tool covered several domains, including socio-economic demographics at home, household literacy levels, learner school history, reading at home, experiences during COVID-19 school closures and exposure to government distance learning initiatives. Learner interviews were conducted with every child immediately after they completed their learning assessments. Enumerators asked learners a series of questions and then noted responses via an electronic tablet. Data was submitted electronically every day.

Head Teacher and Teacher Interview

The head teacher and grade 2 and 4 teachers in each school who were observed were interviewed using tools that comprised open-ended questions across a variety of domains related to teaching and learning. Specific questions focused on the inputs provided to date regarding teacher training, support supervision, hardware for learning (human resources, buildings, and school environment for reading and maths lessons) and software for learning such as teacher skills, content knowledge, ability to use the curriculum and materials when teaching literacy and numeracy, active teaching and learning methods and the ability of district inspectors to mentor and train teachers in literacy and numeracy. A huge factor in teacher performance is subject area knowledge, which is often self-reported and only minimally evaluated through lesson observations. Montrose, therefore, gave teachers an abbreviated reading test to assess their knowledge of letter sounds and comprehension.

Lesson Observations

Lesson observations were conducted by enumerators in grade 2 and 4 classrooms using an observation checklist that recorded the observable actions and behaviours of teachers and learners at 10-minute intervals during English and numeracy lessons. The tool collected information related to instructional content, lesson activities and tasks, use of learner-centred methods, learner assessment and classroom management techniques. Lesson subjects for observation were randomised across grades and schools to achieve an equal balance between the types of lessons observed and so that only one lesson per teacher was watched during the school visit. Enumerators videotaped 20 per cent of the lessons observed for additional analysis post-fieldwork; these were also randomly selected across grades and subjects. A team of analysts who were specifically trained and certified in classroom observation analysis evaluated the videos and recorded findings using a contextualised version of the World Bank's *Teach* tool. Refer to **annex 9.4** for more details on the Teach tool that was used for this analysis.

School Observation Checklist

This tool was designed to capture data on the school's teaching and learning environment and the availability and accessibility of hardware and software inputs important to learners receiving a quality, inclusive education. School facilities, classrooms, teaching and learning materials, water, hygiene and sanitation facilities, school feeding, and compound safety were assessed using a rubric scale.

Grade and Employee Roster

This tool collected demographic information about all the teachers and administrators in a school, including information on their age, gender, education history, teaching certifications, employment history, current employment at the school, grades and subjects taught, presence in the school on the day of the visit and (if present) classes being taught that day. It also collected enrolment and attendance data for all the grades in the school, disaggregated by gender. This information helped to paint a deeper profile of the schools and teachers in the sample, allowing the research team to explore how the demographics of this teacher workforce, and the learner population's size and distribution, potentially – positively or negatively – affected learning. With this data, the adequacy of the school's human and material resources to provide a high quality and safe place to learn were examined.

Sampling Register, Daily Reports and Fieldwork Protocol

To facilitate data collection, enumerator teams and particularly team leaders were provided with additional field documents used to record learner specific bio data. Enumerators were equipped with a sampling register that was used to document the learners assigned unique tangerine ID, the learner age, grade, and the number of learners present in the classroom that day which assisted in the data cleaning process. Daily reporting templates and a fieldwork protocol were also provided to guide data quality documentation and checking, reporting on daily outputs, uploading of data, and overall survey management.

2.3. Data collection

The study collected data from 260 schools from all four regions across 16 districts of Sierra Leone. The assessment was conducted at the end of the current academic year (2nd June to 22nd July 2021)⁵ to best capture evidence on the end of year reading and mathematics proficiency of children in key grade-level skills. The 260 schools were visited in two rounds of data collection with 147 schools in round 1 and 113 schools in round 2. Data collection was preceded by a 6-day enumerator training that was held from the 24th to 29th May 2021 with 45 skilled enumerators in Makeni district. Based on prior experience, Montrose selected enumerators based on several key competencies including their ability to communicate well with children, use assessment tools and materials (EGRA/EGMA was an advantage), use a tablet and their fluency in English, Krio and another local language.

Pre-testing exercises that were guided and observed by the trainers, enabled enumerators to be conversant with the tools and to build their confidence in handling the instruments. At the end of each day the trainee group met to discuss experiences, lessons learnt and report problems while administering the assessments. The revised work plan the shows the two rounds of data collection is included in this report as **annex 9.6**.

Data was checked and inventoried through daily reporting, data quality and completeness checks conducted during the end-of-week data review and handover meetings between enumerators and the national project management sub-contractor (TIMAP for Justice) that supported this exercise.

Table 1: Data collection tools (planned verses actual)

Data collection tool	Planned	Actual	Percentage achieved	Reasons for variance
EGRA/EGMA	5,200 learners in grades 2 and 4	4,729*	90%	Some learners were absent on the day of the visit to their school and were unable to be assessed.
Pupil context Interview	5,200 learners in grades 2 and 4	4,729*	90%	
Lesson Observation Tool	512	435	85%	Some teachers were absent on the day of the visit to their school and were unable to be observed.
Lesson Observation Videos	104 (20% of 512 lesson observations)	84	81%	me teachers were absent on the day of the visit to their school and were unable to be videotaped.
Teacher Interviews	512 teachers (1 English and 1 Maths teacher in all schools)	476	93%	Some teachers were absent on the day of the visit to their school and were unable to be interviewed.
Head Teacher Interviews	260	244	94%	Some head teachers were absent on the day of the visit to their school and were unable to be interviewed.

⁵ Third term begins immediately after the Easter break (19 April 2021) and runs for 13 weeks up to the 16th of July 2021. The mid-term break is from the 28th to the 31st of May 2021.

Data collection tool	Planned	Actual	Percentage achieved	Reasons for variance
Focus Group Discussions with PTAs	64 (4 in each district)	41	64%	Some PTAs, although mobilised, did not come to the school on the day of the visit and were unable to be interviewed.
Key Informant Interviews	16 (1 KII per district)	5	31%	Some district officials were unavailable, while others declined to be interviewed.
School Observation Checklist	260	260	100%	Completed as expected.
Grade roster	260	260	100%	Completed as expected.
Employee roster	260	242	93%	Some employee rosters were not completed due to lack of time. Incomplete rosters were not included in the analysis.

* For the EGRA/EGMA and learner interview tools, 2,418 learners were in grade 2 and 2,311 learners were in grade 4

In the table above, less data was collected in some instances such as head teacher interview and teacher interview due to absence of the required respondent which also affected the collection of the grade and employee roster data. Furthermore, only 31 per cent of the district key informant interviews were conducted due to the unavailability of most District Deputy Directors of Education to participate in the interview. Fewer lessons were observed, and videos taken as some schools (especially round 2 of data collection) had already stopped conducting lessons in preparation for the end of term exams or had already completed the end of term exams. Refer to **annex 9.9.1** for detailed tables on the actual sample obtained by data collection tool.

2.4. Analytical framework

Enumerators recorded learning assessments, lesson observations, learner, and teacher/head teacher interviews directly on tablets. The data was uploaded every evening to the server. The Quantitative EGRA/EGMA Expert reviewed the uploaded data to ensure consistency with the detailed field report and that assessments were not missing. In case of any inconsistencies, teams were required to provide a valid explanation, to repeat assessments or to return to schools to complete any missing assessments. These quality assurance checks were run daily, and any issues reconciled at the end of the day or week; some schools received a call back to collect or confirm data. At the end of the data collection process once all the data had been uploaded and reviewed again before data analysis took place. Enumerators worked to a strict data management plan to ensure data could be linked effectively during analysis. This involved the use of unique learner and school ID numbers that were then replicated on all interviews.

Standard Analysis

For EGRA, results of interest included the average of correct letter or words per minute (WPM), broken down by grade or other factors. For EGMA, the number of right or wrong answers per minute were measured. These results allow for some comparability across contexts, considering contextual factors like language of instruction and the grade-appropriateness of content. This WPM and the composite score were analysed against factors such as school, socio-economic status, attendance, and attitudes and perceptions to assess whether there were significant trends in learning outcomes along these variables.

This report indicates the performance of the learners in each of the subtasks administered. In addition, reporting was done by grade and gender to assess whether there are significant increases between

grades, and whether learners were progressing in key skills. The impact of school closures and reduced economic activities due to COVID-19, which disrupted learning and education services, were also considered when documenting and analysing findings.

Reliability analysis

In this study, we carried out reliability analyses on the English EGRA and EGMA subtasks. To do this, Pearson correlations was first used to compute the reliability coefficients among each tool's subtasks. Strong correlations among subtasks are desired because they typically indicate consistency in the performance of sampled pupils across subtasks. Additionally, the Cronbach's alpha for all the subtasks in each tool was computed to measure the internal consistency of the tool. When combined with item-level analyses, internal consistency offers insights on subtask functioning.

The highest coefficient of 0.80 was found in reading comprehension 1 and listening comprehension. The lowest coefficient was 0.69 for familiar word fluency and oral reading fluency 1. These results show that the English EGRA tool was reliable for assessing the baseline study group of pupils, with an overall reliability coefficient of 0.78.

The highest coefficient of 0.93 was found in subtraction level 2, while the lowest coefficient was 0.92 for all the remaining EGMA subtasks. These results show that the EGMA tool was highly reliable for assessing the Sierra Leone baseline study group of pupils, with an overall reliability coefficient of 0.93. Please refer to **annex 9.10** for a detailed explanation for how the reliability analysis was conducted.

Comparative Analysis

The last national early grade reading, and maths assessments were conducted in Sierra Leone in 2014. Findings indicated that very few primary learners were acquiring the basic reading and maths skills necessary for their future academic – and life – success. Most children were non-readers and unable to complete grade level phonemic awareness, fluency and reading comprehension tasks. In maths, children lacked computational skills and strategies; many were unable to complete even simple addition tasks.

In response, UNICEF, with GPE funding, developed and implemented a set of interventions with the government to improve literacy and numeracy outcomes for children in grades 1-3 through improved instruction. Building strong foundations for early learning through training early grade teachers in reading and mathematics is also a priority in the Education Sector Plan, as is providing resources to schools to improve instruction.

Direct, statistically significant comparisons between the 2021 learner data and the 2014 learner data are not possible due to differences in the school and learner sample, test design and administration process. However, the report does provide some relevant analysis of learning outcomes across these cohorts, notably by examining one level of performance that both categories of learners share in every subtask – zero scores. Score distributions for each cohort are compared for selected reading and maths subtasks to provide a basis for reflection and discussion, though they cannot be utilized as key findings beyond this.

Learner Profiles

The EGRA and EGMA are used to evaluate studies and monitor projects that address reading and numeracy skills in low- and middle-income countries. Results are often described solely in terms of performance on a high-level subtask, such as passage-reading or word problems, thereby overlooking learners' progress in related skills. Using data from the study, the research team applied a methodology that uses oral passage-reading results from the EGRA to create five learner profiles: Non-reader,

Beginner, Instructional⁶, Fluent and Next-Level Ready. The team also created five learner profiles for the EGMA using findings from the word problems subtask: Non-numerate, Beginner, Instructional, Functional and Next-Level Ready.

In each assessment, the learner profiles were compared with learners' results on the other subtasks to identify the skills they needed to develop to progress from one profile to another. Regression models were used to determine whether the learner profiles were related to and consistent with the learners' results on the other subtasks. This aided in identifying membership within the categories across all learners and determined whether there was a shift in the distribution of learner profiles from grade 2 to 4, which is useful for exploring their progressive achievement of critical literacy and numeracy skills in the early grades.

By describing learners using profiles and by the number of learners that shift from one profile to another, we can more accurately and clearly describe learners' abilities and their corresponding instructional needs – and show changes in their performance in a more meaningful, actionable way. Ultimately, the profiles can guide classroom instruction to meet the needs of learners so that those who lack a particular skill receive appropriate practice and instruction regardless of grade level. Furthermore, they help to shift the interpretation of results from a deficit perspective, meaning what learners cannot do on one subtask, to a recognition of their abilities, helping us determine how to best address their different needs within the classroom.

Lesson Observations

Videotaped lessons of literacy and numeracy lessons were post-scored using the World Bank's Teach tool. Lesson observation tools documenting teacher and learner actions, instructional methods and materials used, teacher time on task, assessment practices, and time spent developing key foundational skills were analysed and results presented by grade and subject. The research team attempted to examine whether the current tools and instructional methods used by teachers align with the instructional needs of their learners, which was identified in the learning profile analysis described above. The goal is to identify what teachers are currently doing, versus what they should be doing, to develop learner's foundational skills. This provides an additional level of evidence to support recommendations for improving teaching and learning that connects classroom practice with learner achievement.

Qualitative Data

Where possible, qualitative data was uploaded on the software server where it was retrieved by the National Research Specialist/Technical Team Leader for further in-depth analysis. Where internet challenges hindered this process, qualitative data was stored on the tablet and transferred to the National Research Assistant on a weekly basis. Scanned copies of the paper-based tools were also sent to the National Research Specialist/Technical Team Leader for further reference.

Other qualitative data that was collected from KIIs with district education authorities was transcribed from audio files into notes stored in Microsoft Word by the National Research Specialist/Technical Team Leader and Montrose hired data scribes. Data was categorised by source i.e., teacher or head teacher and then further divided by the headings found in the data collection tool such as use of teaching and learning materials, experience teaching, among others.

⁶ Instructional reading level is the highest level at which a reader is not independent, but has adequate background knowledge for a topic, and can access text quickly and with no or few errors. Think of independent level as the highest level you would ask a child to read with only a small amount of assistance.

All qualitative data was analysed using manual content analysis to allow for the identification of common patterns and themes. Quantitative and qualitative data was triangulated to observe trends and important features that may be common among schools, for instance, the same performance bracket, district, school type. This triangulation helped to understand the characteristics of schools and teachers that appear to be more successful than others and to identify the scope and reforms that are needed to improve literacy and numeracy.

2.5. Challenges, limitations and technical considerations

There are challenges and limitations associated with this methodology and the following list is by no means exhaustive.

1. The study does not specifically measure personal factors that inhibit learning. This study does not account for other underlying factors e.g., disabilities that might affect children's ability to learn numeracy and literacy.
2. Phonics is critical to most methods of teaching young children to read. The EGRA tools have been developed in English. Generally standard English sounds are set as the base. During the training workshop, a set of standard sounds were agreed, and enumerators were trained to listen for these and to score accordingly on the EGRA tests. Where possible, enumerators were assigned to collect data in areas where they spoke the local language; this helped reduce the negative effects accent and dialect may have had on children's pronunciation and on the enumerator's ability to listen for the correct letter sound.
3. Enumerators may not have been taught letter sounds while in school. The EGRA tool requires the enumerator to be conversant with the different letter sounds. Enumerators were trained to familiarise themselves with letter sounds and the standard that will apply in Sierra Leone.

Challenges experienced during data collection and their corresponding mitigation measures were extracted from the field report and included in this report as **annex 9.9.2** with a few key challenges included below.

4. The available national list of schools provided for sampling was outdated, resulting in a substantial number of schools being replaced because the schools were either not found or not accessible by road/vehicle. To mitigate this, schools' accessibility was verified during pre-data collection phase when letters were being dropped off at school and inaccessible schools replaced from the list of buffer schools. Where the schools on the buffer list were also not found, district officials and head teachers were approached to support the identification of alternative schools.
5. Delays in finalising contract discussions for the additional 110 schools resulted in round 2 of data collection starting later than planned, with a 3-day gap between each round. As a result, some schools had finished conducting lessons and taken their end of term exams by the time they were visited. The goal is to ensure equivalency between the two cohorts in rounds 1 and 2 to prove the data is comparable despite being collected at different times. While schools were selected using the same metrics, the timing of the sample's data collection differed, with cohort 2 fieldwork running later in the school term (e.g., after additional weeks of learning, studying and final exams). Therefore, the research team aimed to prove that there were no inherent differences between the two samples of teachers and learners due to these variances. Refer to **section 4** of this report for the analysis that was conducted to check the equivalency of the two cohorts.
6. Even with the delivery of the letter from the Ministry and a follow-up phone call from the enumerator team leader, some teachers and head teachers were still not found in the school on the

day of the assessment. This was mitigated, where possible, with repeated contacts to the head teachers and rescheduling where necessary. Where possible the teams mobilised teachers to the school once they arrived if some of the teachers were staying nearby.

7. There was a need for in-country technical support for the enumerators which was mitigated through the recruitment of a national Research Specialist (with extensive EGRA/EGMA experience) who visited all teams and supervised the intake process at the end of each data collection week.

2.6. Research ethics and ethical considerations due to COVID-19

Montrose followed the UNEG Ethical Guidelines for Evaluations' and the 'UNICEF Procedure for Ethical Standards in Research, Evaluation, Data Collection and Analysis'. Additionally, Montrose paid heed to the ethical guidelines on conducting research with children.⁷

Montrose expected no physical, psychological, social, or legal risks to respondents. The main risk was that of a breach of confidentiality which was mitigated by storing all identifiable data securely using encrypted, password-protected files, and by anonymising data (removing participant names) prior to analysis. Additionally, one of the main priorities for the enumerators, prior to beginning the survey, was to ensure that they conducted the survey in a safe and secure environment during daytime where the respondent feels comfortable answering questions. Furthermore, a designated member of the senior Montrose team was responsible for data security and only the enumerators and Montrose staff had access to respondent information. More elaborate research ethics employed during this study can be found in **annex 9.7**.

In addition to the above research standards, Montrose also employed the following guidelines while conducting fieldwork due to the COVID-19 pandemic. Throughout this study, Montrose followed guidelines and standard operating procedures from Ministry of Health and Sanitation through selected COVID-19 pre-data collection and during data collection protocols. Some of these include:

- Ensuring enumerators worked/collected data in the districts/communities where they already lived.
- Requiring enumerators to be trained on COVID-19 preventative measures.
- Equipping enumerators with personal protective equipment (PPE) during training and fieldwork.
- Providing each enumerator team with an infra-red thermometer and a record sheet to track body temperature.

Additional details on the ethical considerations put in place due to the COVID-19 pandemic can be found in **annex 9.8**.

3. FINDINGS ON LEARNERS

The following section details key findings for learners on the EGRA and EGMA. Data is presented first on learners who were unable to complete any item of a subtask and thus received zero scores. There were non-readers on every EGRA subtask, and innumerate learners on every EGMA subtask, though less so than in the reading assessment. This information is important for determining future interventions, as learners with non-existent or extremely low skill levels must receive specific, and often intensive, literacy and numeracy interventions to catch up.

⁷ Berman, Gabrielle (2020). Ethical Considerations for Evidence Generation Involving Children on the COVID-19 Pandemic, *Innocenti Discussion Papers* no. 2020-01, UNICEF Office of Research - Innocenti, Florence

Next, data is presented on mean scores for each subtask for learners who completed at least one item. Data is disaggregated by grade and gender. These scores are important to determine the average performance of learners as they progress in school to inform instructional and pedagogical interventions that will raise literacy and numeracy outcomes in line with grade level expectations.

3.1. Learner characteristics

A total of 96 per cent of the projected learner sample was reached in round 1 of data collection; just over 90 per cent was reached in round 2. On average across both rounds, nearly 94 per cent of the targeted learner sample was reached, totalling 4,729 learners. The final sample represents a significant enough percentage of the learner sample to be representative and allow us to identify statistically sufficient findings. Gender and grade distribution across the learner sample were largely balanced.

Learner Characteristics Finding 1 – Age, Grade Level and Links to Early Learning: Many learners are an average of two years older than they should be in both grade 2 and grade 4, and only 1 in 4 learners attended nursery school before starting primary school. A substantial number of grade 2 learners (about 16-18 per cent) are aged 9-10 when they should be 7-8 years, while many grade 4 learners (about 22-24 per cent) are aged 11-13 (when they should be 9-10 years). Additionally, many underage learners are also enrolled, notably children of 6-7 years who should still be in grade 1. This contributes to overcrowding and an early grade bulge that adds cost and wastage to the system.

Table 2: Learner distribution by grade and by gender

Grade	Boys	Girls	Total
P2	1,224	1,194	2,418
P4	1,157	1,154	2,311
Total	2,381	2,348	4,729

Table 3: Learner distribution by age group

Age	P2				P4			
	Boy	%	Girl	%	Boy	%	Girl	%
6 years old	59	4.8	69	5.8	0	0.0	0	0.0
7 years old	287	23.4	281	23.5	4	0.3	8	0.7
8 years old	350	28.6	374	31.3	35	3.0	45	3.9
9 years old	248	20.3	230	19.3	175	15.1	176	15.3
10 years old	193	15.8	154	12.9	360	31.1	319	27.6
11 years old	45	3.7	45	3.8	214	18.5	264	22.9
12 years old	32	2.6	32	2.7	209	18.1	225	19.5
13 years old	7	0.6	6	0.5	100	8.6	74	6.4
14 years old	2	0.2	0	0.0	41	3.5	34	2.9
15 years old	1	0.1	2	0.2	12	1.0	8	0.7
Over 15 years old	0	0.0	1	0.1	7	0.6	1	0.1

Learner Characteristics Finding 2 – Language: For many learners, English is their second or even third language. Literacy interventions must consider this in their design to help learners transfer their knowledge and skills from one language to another. English and Krio are largely spoken in fairly equal measure in school during instruction, although English is the only reported language of instruction in national documents and the only language found in textbooks. About 30 per cent of learners reported

speaking their local language in school during class instruction, meaning their exposure to English listening and speaking tasks is likely diminished, as is their ability to read and write in the language. Speaking English at school demonstrates a positive correlation with learning achievement, while speaking Krio or another local language has the opposite effect. A total of 75 per cent of learners reported speaking their local language at home with family, while another 50 per cent use Krio. Very few learners reported speaking English at home, meaning that the only exposure most learners get to this language of instruction is at school. Speaking English or Krio at home both have a positive correspondence to learner performance, demonstrating they are a driver of better results. This has an impact on literacy development, as less time spent interacting with a language slows children's acquisition of key foundational skills.

Table 4: Learners distribution by language spoken at home and at school

Language	Spoken at School	Spoken at Home
English	65%	1%
Krio	56%	49%
Local language	28%	75%

Table 5: Significance of school language on scores

Subtask	School Language - English			School Language - Krio			School Language - Other language		
	Coef.	Std. Err.	p-value	Coef.	Std. Err.	p-value	Coef.	Std. Err.	p-value
Letter sounds	3.1	0.60	0.001**	-2.6	0.54	0.001**	-2.9	0.64	0.001**
Non-words	4.0	0.55	0.001**	-2.4	0.46	0.001**	-3.4	0.62	0.001**
Familiar words	6.3	0.46	0.001**	-2.7	0.42	0.001**	-5.1	0.51	0.001**
Oral reading story 1	8.1	0.70	0.001**	-3.2	0.63	0.001**	-7.0	0.78	0.001**
Reading comprehension 1	0.0	0.03	0.21	0.0	0.03	0.11	-0.1	0.03	0.021*
Oral reading story 2	6.9	1.23	0.001**	-3.6	1.03	0.001**	-5.9	1.32	0.001**
Reading comprehension 2	0.0	0.03	0.31	0.0	0.02	0.19	0.0	0.03	0.57
Listening comprehension	-0.2	0.03	0.001**	0.0	0.03	0.57	0.2	0.03	0.001**
Number identification	4.1	0.20	0.001**	-0.5	0.20	0.021*	-3.7	0.22	0.001**
Quantity discrimination	1.7	0.11	0.001**	-0.5	0.10	0.001**	-1.7	0.12	0.001**
Missing number	1.4	0.09	0.001**	-0.5	0.08	0.001**	-1.2	0.10	0.001**
Addition level 1	2.2	0.15	0.001**	-0.5	0.15	0.001**	-1.6	0.17	0.001**
Addition level 2	0.5	0.08	0.001**	0.0	0.07	0.90	-0.3	0.08	0.001**
Subtraction level 1	1.6	0.17	0.001**	-0.2	0.16	0.12	-0.8	0.19	0.001**
Subtraction level 2	0.4	0.10	0.001**	-0.1	0.08	0.32	-0.4	0.10	0.001**
Word problems 1	0.4	0.03	0.001**	0.0	0.03	0.29	-0.4	0.04	0.001**
Word problems 2	0.2	0.03	0.001**	0.0	0.03	0.31	-0.1	0.03	0.011*

Table 6: Significance of home language on scores

Subtask	Home Language - English			Home Language - Krio			Home Language - Other language		
	Coef.	Std. Err.	p-value	Coef.	Std. Err.	p-value	Coef.	Std. Err.	p-value
Letter sounds	-1.1	2.01	0.59	1.2	0.54	0.031*	0.7	0.62	0.29
Non-words	9.3	1.68	0.001**	2.2	0.46	0.001**	-2.3	0.50	0.001**
Familiar words	14.9	1.76	0.001**	3.6	0.42	0.001**	-3.8	0.46	0.001**
Oral reading story 1	18.3	2.42	0.001**	5.0	0.63	0.001**	-6.4	0.70	0.001**
Reading comprehension 1	0.1	0.10	0.26	0.0	0.03	0.26	-0.1	0.03	0.001**
Oral reading story 2	24.3	4.04	0.001**	5.0	1.03	0.001**	-6.9	1.14	0.001**
Reading comprehension 2	-0.1	0.10	0.47	0.0	0.02	0.89	0.0	0.03	0.26
Listening comprehension	0.0	0.15	0.93	-0.2	0.03	0.001**	0.2	0.04	0.001**
Number identification	3.4	0.88	0.001**	2.2	0.20	0.001**	-1.4	0.23	0.001**
Quantity discrimination	1.8	0.43	0.001**	0.9	0.10	0.001**	-0.4	0.12	0.001**
Missing number	1.2	0.35	0.001**	1.1	0.08	0.001**	-0.3	0.09	0.001**
Addition level 1	1.9	0.60	0.001**	1.3	0.14	0.001**	-0.4	0.16	0.031*
Addition level 2	0.7	0.27	0.011**	0.2	0.07	0.001**	-0.2	0.07	0.011**
Subtraction level 1	2.0	0.59	0.001**	1.1	0.16	0.001**	-0.4	0.18	0.041*
Subtraction level 2	0.5	0.30	0.07	0.2	0.08	0.001**	-0.1	0.09	0.42
Word problems 1	0.1	0.14	0.40	0.3	0.03	0.001**	-0.1	0.04	0.051*
Word problems 2	-0.1	0.11	0.63	0.1	0.03	0.001**	0.0	0.03	0.78

Learner Characteristics Finding 3 – Reading at Home: About 50 per cent of learners' fathers and 40 per cent of their mothers are literate. 75 per cent of learners reported reading and studying at home with a range of other family members and friends. Most learners read with their siblings. Just over 30 per cent take books home from school to read – mostly maths (47 per cent) and letter writing and composition (54 per cent), while over 50 per cent reported having other reading books at home already, namely storybooks (40 per cent). Reading at home had a positive correlation with better performance on the EGRA and EGMA. Interestingly, analysis also showed that mothers reading at home with their children had the greatest statistically significant impact on learner performance, followed by reading with fathers; reading with a sibling corresponded to negative impacts on learning.

Table 7: Significance of reading at home

Subtask	Walking			Bicycle		
	Coef.	Std. Err.	p-value	Coef.	Std. Err.	p-value
Letter sounds	1.3	1.17	0.26	-2.0	2.98	0.51
Non-words	-5.6	0.92	0.001**	4.7	2.34	0.051*
Familiar words	-7.5	0.89	0.001**	6.2	2.50	0.011*
Oral reading story 1	-10.6	1.34	0.001**	7.0	3.71	0.06
Reading comprehension 1	0.0	0.06	0.51	0.1	0.15	0.69
Oral reading story 2	-11.5	2.37	0.00	10.7	5.75	0.06
Reading comprehension 2	0.0	0.06	0.42	-0.1	0.16	0.67
Listening comprehension	0.0	0.09	0.98	0.0	0.17	0.84

Subtask	Walking			Bicycle		
	Coef.	Std. Err.	p-value	Coef.	Std. Err.	p-value
Number identification	-2.0	0.47	0.001**	-0.5	1.13	0.66
Quantity discrimination	-0.8	0.23	0.001**	-0.4	0.55	0.48
Missing number	-0.5	0.19	0.011*	-0.4	0.45	0.37
Addition level 1	-0.9	0.33	0.011**	-0.4	0.79	0.64
Addition level 2	-0.3	0.15	0.051*	-0.2	0.39	0.64
Subtraction level 1	-0.8	0.34	0.021*	-0.3	0.81	0.68
Subtraction level 2	-0.1	0.18	0.59	-0.5	0.43	0.30
Word problems 1	-0.1	0.07	0.41	-0.1	0.19	0.65
Word problems 2	0.0	0.06	0.50	0.0	0.14	0.80

Table 8: Distribution of learners by who reads with them at home

Who Learner Reads With	Percentage
Reads with mother	9%
Reads with father	13%
Reads with sibling	47%
Reads with adult relative at home	12%
Reads with friends	4%
Reads with no one (reads alone)	10%
Reads with another companion	9%

Table 9: Significance of reading with family members

Subtask	Reads with mother			Reads with father			Reads with sibling		
	Coef.	Std. Err.	p-value	Coef.	Std. Err.	p-value	Coef.	Std. Err.	p-value
Letter sounds	3.8	1.01	0.00	2.7	0.82	0.00	-3.1	0.58	0.00
Non-words	2.7	0.86	0.00	2.1	0.69	0.00	-1.7	0.50	0.00
Familiar words	2.2	0.83	0.011**	0.9	0.70	0.19	-1.0	0.49	0.041*
Oral reading story 1	3.4	1.22	0.011**	0.8	1.03	0.44	-0.4	0.71	0.55
Reading comprehension 1	0.0	0.06	0.45	0.0	0.05	0.56	0.0	0.03	0.50
Oral reading story 2	6.2	2.10	0.00	1.0	1.67	0.54	1.0	1.13	0.38
Reading comprehension 2	0.0	0.05	0.99	0.0	0.04	0.48	0.1	0.03	0.021*
Listening comprehension	0.0	0.07	0.86	0.0	0.06	0.76	-0.1	0.04	0.09
Number identification	0.6	0.39	0.10	0.7	0.33	0.031*	-0.7	0.22	0.00
Quantity discrimination	0.5	0.20	0.011**	0.5	0.17	0.00	-0.5	0.11	0.00
Missing number	0.6	0.17	0.00	0.6	0.14	0.00	-0.5	0.10	0.00
Addition level 1	0.5	0.28	0.07	0.9	0.24	0.00	-0.5	0.16	0.00
Addition level 2	0.3	0.13	0.02	0.1	0.11	0.50	0.0	0.07	0.83
Subtraction level 1	0.1	0.28	0.61	0.7	0.24	0.011**	-0.5	0.17	0.00
Subtraction level 2	0.3	0.15	0.021*	0.2	0.13	0.12	0.0	0.09	0.71
Word problems 1	0.3	0.07	0.00	0.1	0.05	0.011*	-0.2	0.04	0.00

Subtask	Reads with mother			Reads with father			Reads with sibling		
	Coef.	Std. Err.	p-value	Coef.	Std. Err.	p-value	Coef.	Std. Err.	p-value
Word problems 2	0.1	0.05	0.011**	0.1	0.04	0.11	-0.1	0.03	0.031*

Learner Characteristics Finding 4 – Distance Learning: 87 per cent of learners reported listening to the radio teaching programme launched nationally during COVID, while 17 per cent tried to continue learning by watching the television teaching programme. While most learners only did this occasionally, these findings demonstrate a potentially high level of reach from these inputs across the country that is useful to explore further when developing next generation distance or blended learning programmes.

Access to a radio (51 per cent) and electricity at home (69 per cent) were key in expanding the reach of these inputs to learners. Positively, the majority of learners (88 per cent) reported returning to school and attending most of the year through in-person learning post COVID-19 school closures. For those that did not return quickly, 18 per cent reported continuing to listen to the radio or watch the television teaching programmes until they started school.

3.2. Early Grade Reading Assessment

EGRA Finding 1: Notably, the largest percentage of non-readers was registered on the letter-sound correspondence and decoding subtasks; of those who answered at least one item correctly, the average learner identified 11-12 letters or non-words, or between 10-20 per cent of all the items. This indicates learners need more practice with beginning literacy skills that precede whole word reading, including activities that develop their letter-sound knowledge, phonological awareness and application of the alphabetic principle. This will improve their decoding skills, overall word-reading accuracy and reading comprehension.

EGRA Finding 2: A notable percentage of learners demonstrated basic whole language competencies like vocabulary knowledge and oral reading fluency, evidenced by far fewer zero scores for the familiar word and story reading subtasks, and a clear progression in skill development – based on improved performance – from grade to grade. Overall, learners answered about 25-35 per cent of the items correctly in those subtasks. But, despite these relative achievements in oral reading fluency, most learners did not comprehend what they read, as evidenced by the large number of zero scores on the reading comprehension subtasks and the average mean score of only 1 question answered correctly.

Figure 1: Zero Scores by EGRA Subtask

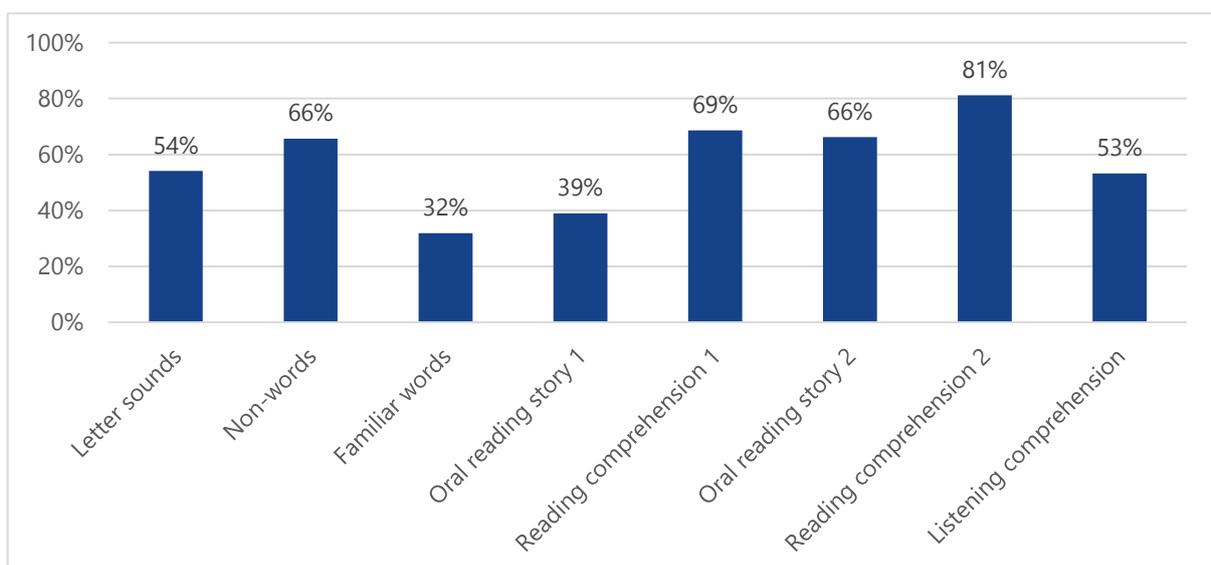


Figure 2: Zero Scores by EGRA Subtask and Gender

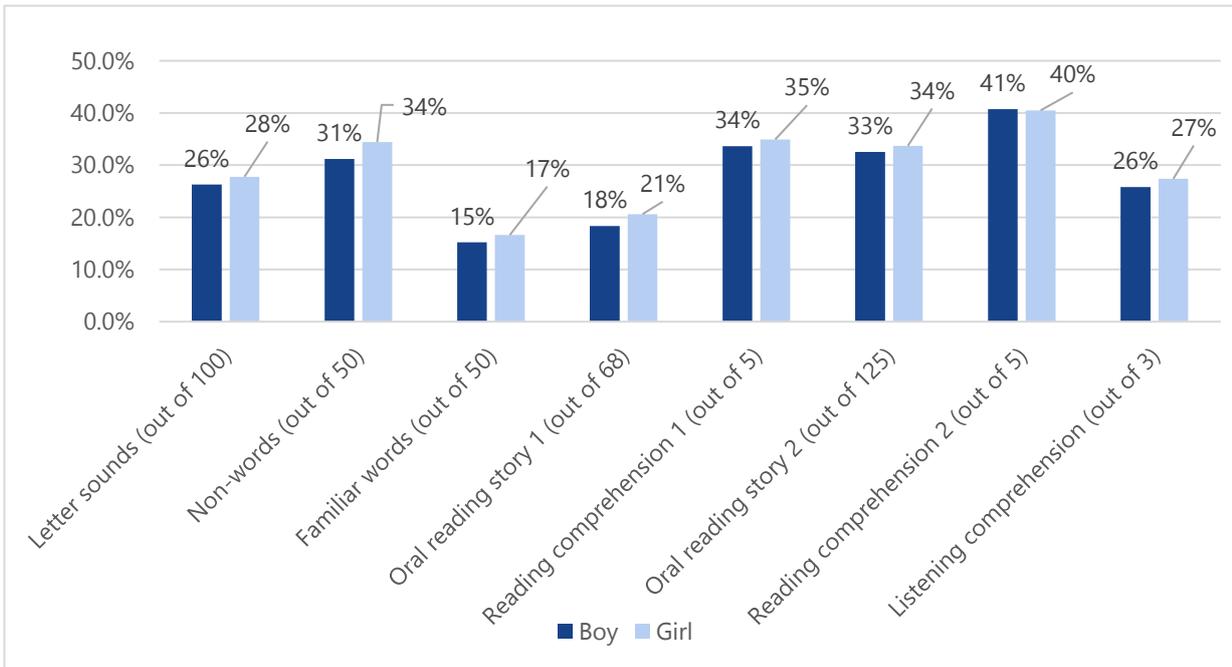
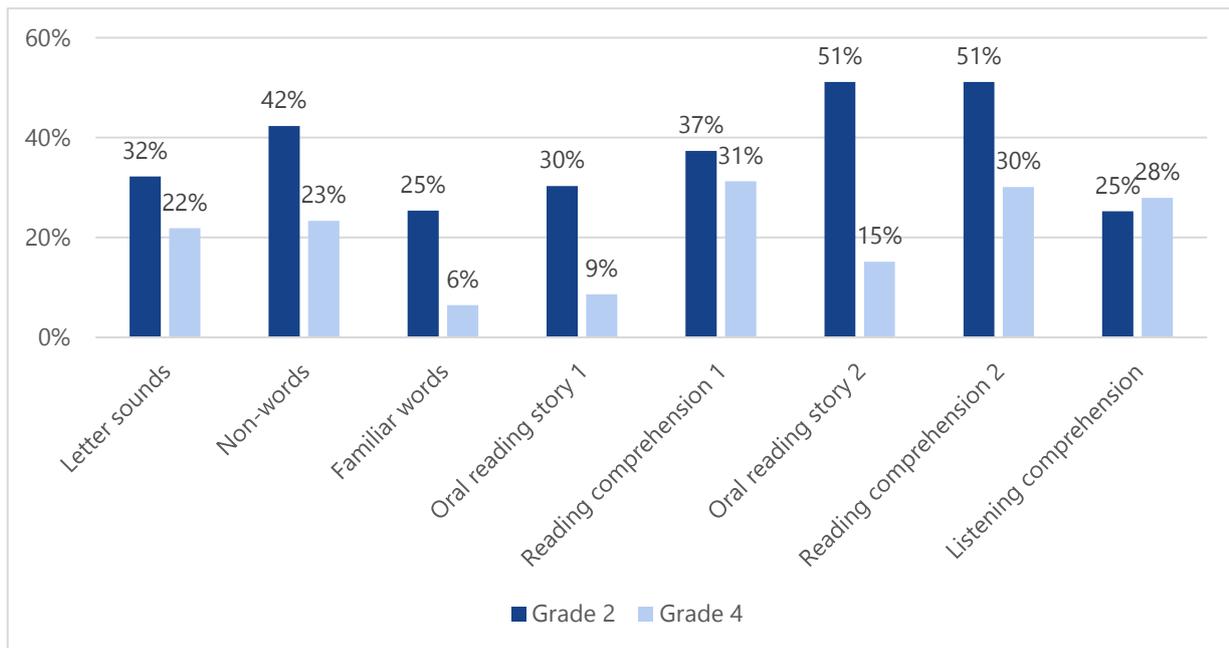


Figure 3: EGRA Subtasks Zero Scores by Grade



EGRA Finding 3: Learners performed well in the listening comprehension subtask, with around 75 per cent of learners in both grades responding to questions correctly 50 per cent of the time. Overall, learners did better in listening comprehension than in reading comprehension, indicating their oral language skills are developing more rapidly than their core reading skills, which is not uncommon.

EGRA Finding 4: The learners who took the second oral reading subtask read with nearly twice the accuracy rate as learners who took only the first oral reading subtask. On average, learners answered around half of the questions correctly in both passages. However, despite demonstrating reading fluency in these subtasks, the low reading comprehension scores indicate learners are taking limited

meaning from the passage they are reading and are unable to respond to – and possibly to understand – the questions. This could be related to limited vocabulary development (demonstrated by their poor performance on the familiar word subtask) and weak decoding skills (demonstrated by their poor performance on the unfamiliar word subtask).

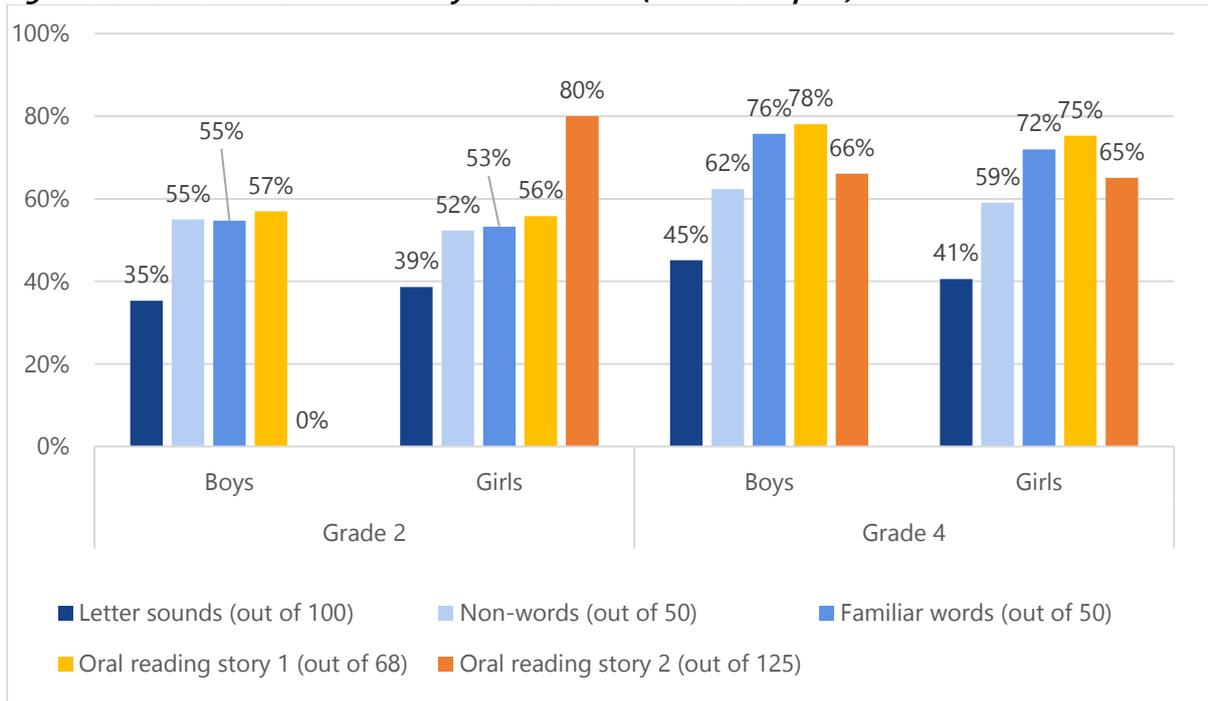
EGRA Finding 5: Results demonstrate clear progression between grades in every EGRA subtask except the second oral reading fluency test. The biggest gains are made between grade 2 and 4 learners in higher level reading skills, including familiar word reading and oral reading fluency. More moderate gains are made in bottom-up tasks like letter sound knowledge and unfamiliar word reading. This suggests that whole language, top-down reading skills grow at a more expected pace than bottom-up skills requiring phonics and decoding, as it appears teachers are not explicitly focusing on developing these foundational skills in their lessons in either grade. The lower mean result between grade 2 and grade 4 learners in second oral reading fluency subtask is likely a consequence of a small sample size, with few learners in either grade reaching the second-tier assessment level and those that did performing at different ability levels (meaning that one high or low performing learner in the group can widely vary the mean score). It is likely that, with a larger sample, progression of mean scores between grades would fall into a similar pattern in every subtask, including this one.

Table 10: EGRA Subtask Means - Excluding Zero Scores

Subtask	Mean (average number correct)	Std. Err.
Letter sounds (out of 100)	10.6	0.27
Non-words (out of 50)	11.8	0.23
Familiar words (out of 50)	14.5	0.21
Oral reading story 1 (out of 68)	19.7	0.32
Reading comprehension 1 (out of 5)	1.2	0.01
Oral reading story 2 (out of 125)	19.6	0.52
Reading comprehension 2 (out of 5)	1.1	0.01
Listening comprehension (out of 3)	1.6	0.02

Table 11: EGRA Subtask Mean of Items Correct vs Attempted by Grade

Subtask	No. of correct items		How many items attempted?		% correct (out of all attempted)	
	Grade 2	Grade 4	Grade 2	Grade 4	Grade 2	Grade 4
Letter sounds (out of 100)	9.2	11.7	25.0	27.2	37%	43%
Non-words (out of 50)	9.8	12.5	18.2	20.5	54%	61%
Familiar words (out of 50)	9.4	17.6	17.3	23.8	54%	74%
Oral reading story 1 (out of 68)	12.6	23.4	22.4	30.6	56%	77%
Oral reading story 2 (out of 125)	32.0	19.6	40.0	29.8	80%	66%

Figure 4: EGRA Subtask Items Correct by Sex and Grade (out of attempted)

3.3. Early Grade Math Assessment

EGMA Finding 1: To assess learners' emerging number sense (i.e., their understanding of the relationship between numbers), they identified numerals, compared quantities as expressed by two numbers and identified missing numbers in a sequence. Learners performed well in these subtasks, demonstrating competence in key procedural maths skills. There were few zero scores and notable average mean scores, with learners answering about 60 per cent of the items they attempted correctly in the number identification and quantity discrimination subtasks. Procedural fluency is critical to developing more advanced maths skills and must be mastered in the early grades; findings are generally positive in this direction.

EGMA Finding 2: Proficiency with simple operations is critical for developing learner's conceptual understanding and fluency with arithmetic operations. Tasks that include simple arithmetic formatted in word problems as well as computation-only problems assess more advanced number sense. Learners performed relatively well in the simple addition subtask, with only 21 per cent of learners scoring zero. Over twice as many scored zero on the simple subtraction subtask, indicating they are still developing their understanding of this more difficult conceptual skill. Mean scores demonstrate varied proficiency, with learners answering less than half the questions correctly on average in both subtasks. Notably, mastering fluency in single-digit addition is a core skill for later mathematical development and should be focused on as a predictor of future mathematics performance.

EGMA Finding 3: Majority of the learners who attempted the addition (67 per cent) and subtraction (77 per cent) level 2 subtasks scored zero. Positively, learners who responded to the items on both subtasks answered more than 60 per cent of the questions correctly, on average. As learners' progress, it is critical for them to develop and demonstrate more advanced – though still foundational – skills in conceptual mathematics, which is the basis for understanding all other higher order skills. Sharp differences in performance across the subtasks demonstrates that learners have notably higher skills in basic, procedural maths tasks, versus significantly lower skills in conceptual ones. These results suggest that learners' experience of mathematics instruction is more about memorisation of facts and rules than

development of strategies to find answers to problems. Conceptual maths must, therefore, be explicitly taught as mastering these foundational concepts is directly predictive of future performance.

Figure 5: Zero Scores by EGMA Subtask

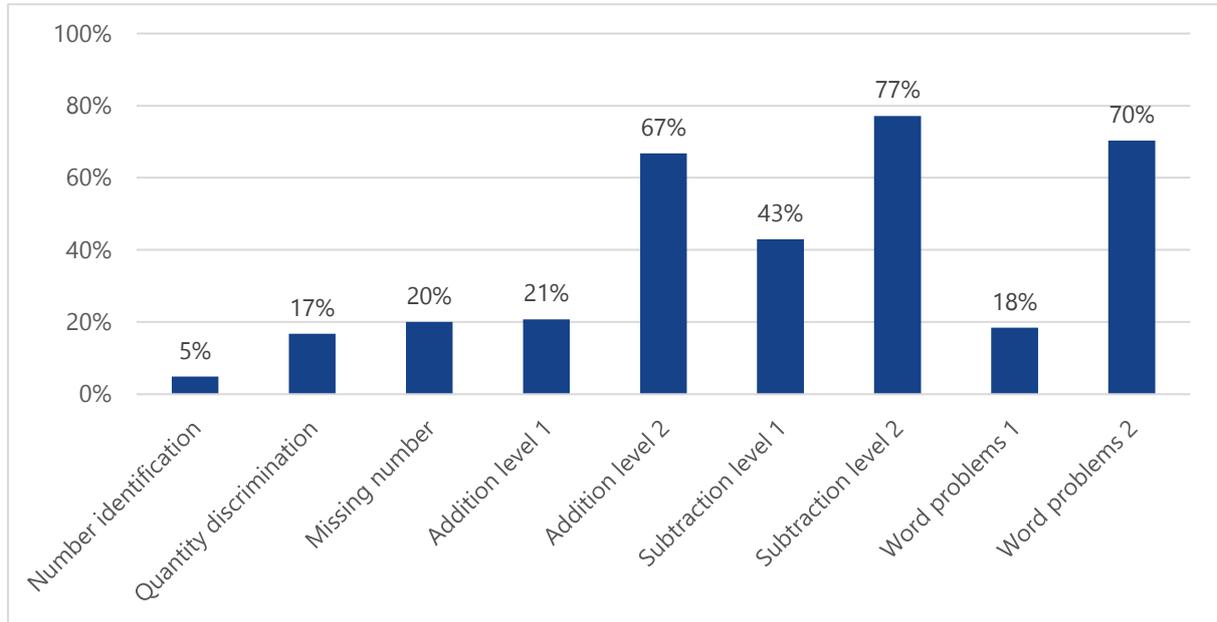


Figure 6: Zero scores by EGMA subtask and grade

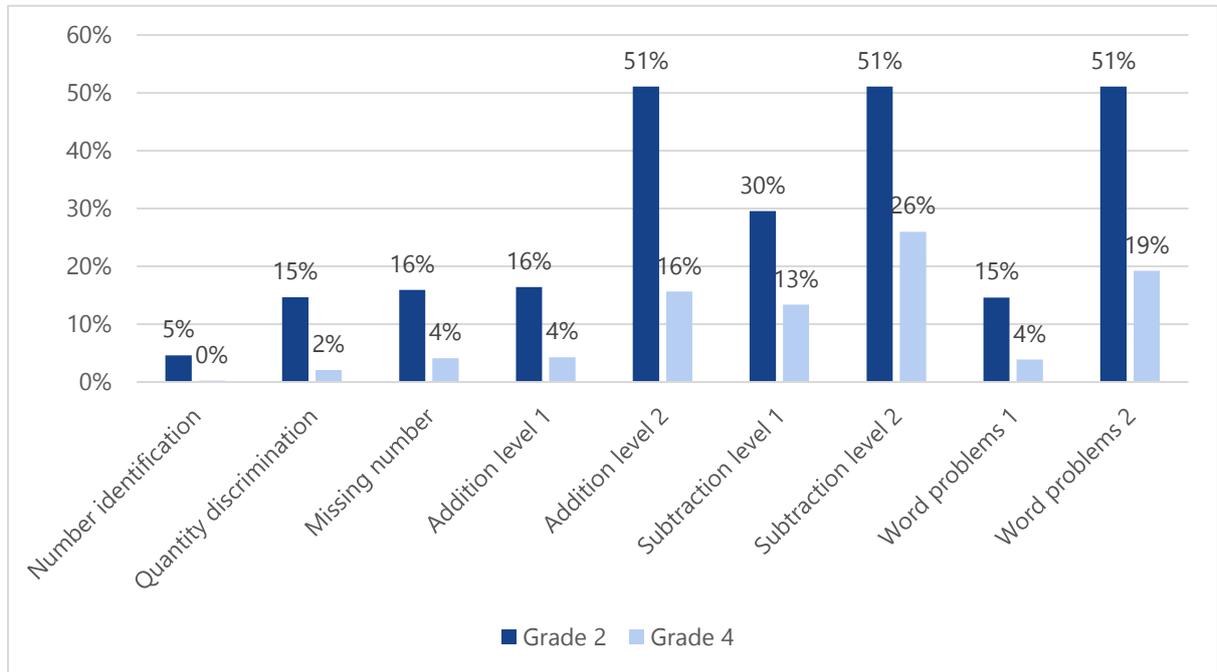
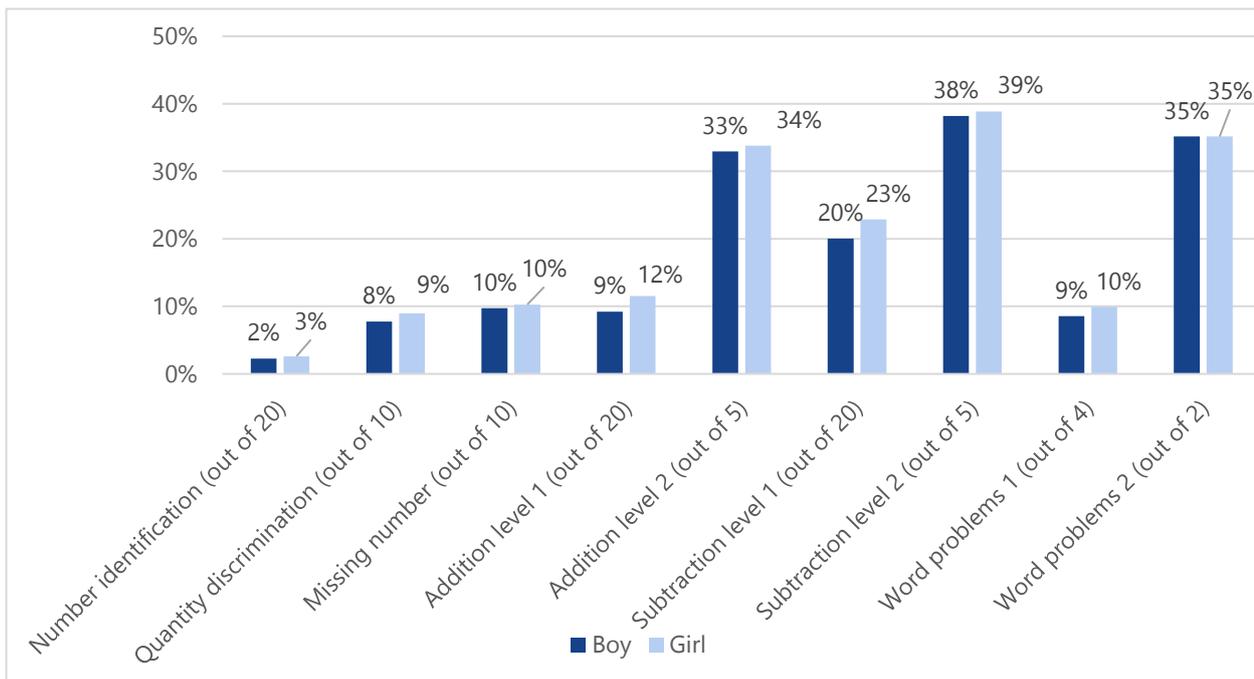


Figure 7: Zero scores by EGMA subtask and gender

EGMA Finding 4: Results demonstrate clear progression between grades in every EGMA subtask except the level two subtasks for addition, subtraction and word problems. The biggest gains are made between grade 2 and 4 learners in the number identification subtask. More moderate gains are made in all the other level one subtasks, where small mean gains are made between grades based on only 2-3 more items being attempted and answered correctly. Importantly, the attempted items and the accuracy of responses increases significantly between grades, with grade 4 learners answering between 20-30 per cent more of their attempted items correct than grade 2 learners.

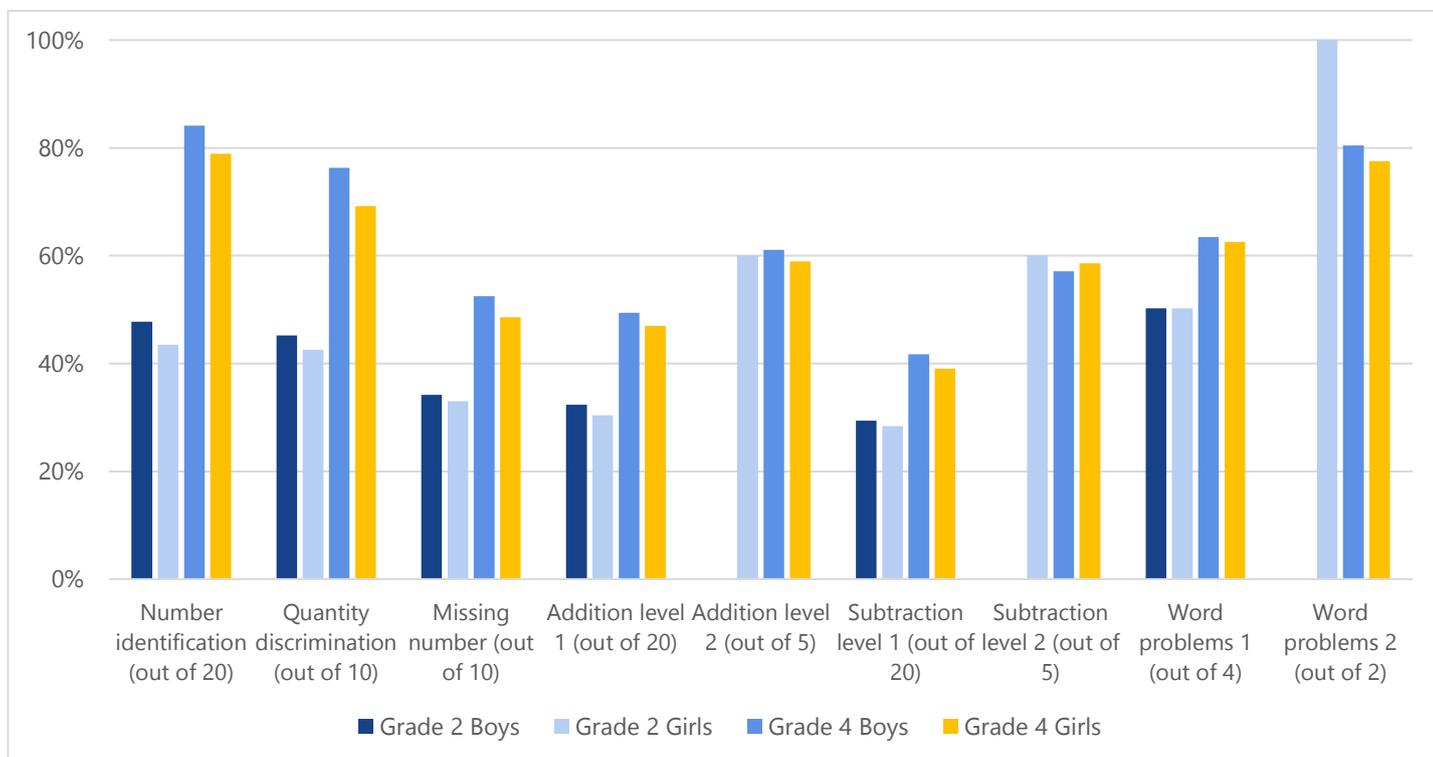
The lower mean result between grade 2 and grade 4 learners in the level 2 subtasks is likely a consequence of a small sample size, with few learners in either grade reaching the second-tier assessment level and those that did performing at different ability levels (meaning that one high or low performing learner in the group can widely vary the mean score). It is likely that, with a larger sample, progression between grades would fall into a similar pattern in every subtask, including this one.

Table 12: EGMA Subtask Means - Excluding Zero Scores

Subtask	Mean	Std. Err.
Number identification (out of 20)	12.8	0.10
Quantity discrimination (out of 10)	6.0	0.05
Missing number (out of 10)	4.3	0.04
Addition level 1 (out of 20)	8.2	0.07
Addition level 2 (out of 5)	3.0	0.03
Subtraction level 1 (out of 20)	7.2	0.08
Subtraction level 2 (out of 5)	2.9	0.04
Word problems 1 (out of 4)	2.3	0.02
Word problems 2 (out of 2)	1.6	0.01

Table 13: EGMA Subtask Mean of Items Correct vs Attempted by Grade

Subtask	No of correct items		How many items attempted? (N/A = all attempted)		% Correct (out of all attempted)	
	Grade 2	Grade 4	Grade 2	Grade 4	Grade 2	Grade 4
Number identification (out of 20)	9.1	16.3	17.9	19.3	51%	84%
Quantity discrimination (out of 10)	4.4	7.3	N/A	N/A	N/A	N/A
Missing number (out of 10)	3.4	5.1	N/A	N/A	N/A	N/A
Addition level 1 (out of 20)	6.3	9.6	10.3	11.9	61%	81%
Addition level 2 (out of 5)	3.0	3.0	N/A	N/A	N/A	N/A
Subtraction level 1 (out of 20)	5.8	8.1	10.0	10.5	58%	77%
Subtraction level 2 (out of 5)	3.0	2.9	N/A	N/A	N/A	N/A
Word problems 1 (out of 4)	2.0	2.5	N/A	N/A	N/A	N/A
Word problems 2 (out of 2)	2.0	1.6	N/A	N/A	N/A	N/A

Figure 8: EGMA subtask items correct by sex and grade out of total items

3.4. Additional findings

Additional Finding 1 – Statistically Significant Gender Differences in Performance: Boys outperformed girls on nearly every subtask in the EGRA and EGMA in both grades, though mean score differences were minimal, with boys answering at most 1-2 more items correctly than girls per subtask; other mean score differences were less than one more item answered correctly in favour of boys. Notably, however, these statistically significant differences in results span the range of foundational and conceptual skills assessed in both reading and maths.

What is statistical significance?

Statistical significance helps determine whether results in an evaluation are merely due to chance or can be attributed to a direct input from an intervention, or to a demographic difference in the sample population, like gender. When we say that results 'are statistically significant,' it means one group of learners is performing better than another group in key subtasks in a manner that is not due to chance. If this gap continues to grow and widen over time, the first group will likely noticeably outperform the second group in similar assessments, with the second group likely falling further behind due to barriers caused or furthered by that same demographic difference.

What are p-values?

A p-value less than 0.05 (typically ≤ 0.05) is statistically significant. A p-value higher than 0.05 (> 0.05) is not statistically significant and indicates strong evidence for the null hypothesis. A p-value greater than 0.05 means that no effect was observed. If a p-value is less than 0.05, it is flagged with one star (*). If a p-value is less than 0.01, it is flagged with 2 stars (**). If a p-value is less than 0.001, it is flagged with three stars (***). The smaller the p-value, the stronger the evidence that you should reject the null hypothesis.

The blue highlights in the **Table 14** below indicate subtasks with statistically significant results for boys over girls in more than half of the assessments. These findings indicate negative gains by boys over girls, meaning that they did not perform equally. Performance between grades, however, shows a strong positive significance in learning from grade 2 to grade 4. Notably, results for the highest level EGRA subtasks – where learners express their reading and processing fluency, as well as their comprehension skills – show no statistically significant differences, or mean score differences, between boys and girls. This is important, as it indicates that girls and boys are performing at similar levels in the higher order competencies involving cognition and comprehension.

Table 14: EGRA and EGMA mean scores by subtask and gender

Subtask	Boys		Girls		Difference
	Mean	Std. Err.	Mean	Std. Err.	p-value
Letter sounds (out of 100)	10.7	0.35	10.7	0.41	0.0
Non-words (out of 50)	12.3	0.32	11.1	0.32	-1.2***
Familiar words (out of 50)	15.1	0.30	13.8	0.29	-1.3***
Oral reading story 1 (out of 68)	20.0	0.43	19.4	0.46	-0.6
Reading comprehension 1 (out of 5)	1.2	0.02	1.2	0.02	0.0
Oral reading story 2 (out of 125)	18.7	0.64	20.6	0.83	2.0
Reading comprehension 2 (out of 5)	1.0	0.01	1.1	0.02	0.1***
Listening comprehension (out of 3)	1.6	0.02	1.6	0.02	0.0
Number identification (out of 20)	13.2	0.14	12.4	0.14	-0.9***
Quantity discrimination (out of 10)	6.2	0.07	5.8	0.07	-0.5***
Missing number (out of 10)	4.4	0.06	4.2	0.06	-0.2***
Addition level 1 (out of 20)	8.4	0.10	8.0	0.11	-0.4***
Addition level 2 (out of 5)	3.1	0.05	2.9	0.05	-0.1
Subtraction level 1 (out of 20)	7.4	0.11	7.0	0.12	-0.4***
Subtraction level 2 (out of 5)	2.9	0.05	2.9	0.06	0.1
Word problems 1 (out of 4)	2.3	0.02	2.3	0.02	0.0
Word problems 2 (out of 2)	1.6	0.02	1.6	0.02	-0.1**

Table 15: EGRA and EGMA mean scores by subtask and grade

Subtask	Grade 2		Grade 4		p-value
	Mean	Std. Err.	Mean	Std. Err.	
Letter sounds (out of 100)	9.2	0.41	11.7	0.35	0.0000**

Non-words (out of 50)	9.8	0.43	12.5	0.27	0.0000**
Familiar words (out of 50)	9.4	0.25	17.6	0.28	0.0000**
Oral reading story 1 (out of 68)	12.6	0.44	23.4	0.40	0.0000**
Reading comprehension 1 (out of 5)	1.1	0.02	1.2	0.02	0.0000**
Listening comprehension (out of 3)	1.7	0.02	1.5	0.02	0.0000**
Number identification (out of 20)	9.1	0.14	16.3	0.10	0.0000**
Quantity discrimination (out of 10)	4.4	0.07	7.3	0.06	0.0000**
Missing number (out of 10)	3.4	0.06	5.1	0.05	0.0000**
Addition level 1 (out of 20)	6.3	0.10	9.6	0.09	0.0000**
Subtraction level 1 (out of 20)	5.8	0.13	8.1	0.09	0.0000**
Word problems 1 (out of 4)	2.0	0.02	2.5	0.02	0.0000**

4. SURVEY COHORT EQUIVALENCY AT BASELINE

As previously noted, data collection for the baseline survey was conducted in two rounds in each district, with 150 schools visited in round 1 and 110 schools visited in round 2. While schools were selected using the same metrics, the timing of the sample's data collection differed, with cohort 2 fieldwork running later in the school term (e.g., after additional weeks of learning, studying and final exams).

At baseline, we hoped to prove equivalency between the two cohorts in rounds 1 and 2 and demonstrate the data is comparable, despite being collected at different times. However, analysis suggests there are differences at the learner level between the two cohorts, with cohort 2 performing better than cohort 1. The findings are statistically significant on several key EGRA and EGMA subtasks.

This section explores the differences in results between the two cohorts by comparing the timing of data collection week-on-week against learner performance. Importantly, there are no observable demographic differences in the learner or school sample between the two cohorts; schools and learners were selected using the same criteria, and data was collected by the same group of enumerators following the same protocols. As such, the last finding summarises possible reasons for the difference in results based upon other factors, including more time spent on task teaching and learning during the term.

Cohort Equivalency Finding 1 – Statistically Significant Differences by Cohort: Learners in cohort 2 performed better than learners in cohort 1 in every subtask: the mean scores for cohort 2 were higher than those in cohort 1, and this performance was statistically significant in nearly every category. Interestingly, both cohorts attempted to answer virtually the same number of items in each subtask. But, learners in cohort 2 answered more of those items correctly than those in cohort 1.

Literacy results show that oral reading fluency scores are essentially the same across the cohorts for both story reading subtasks. However, learners in cohort 2 performed noticeably better in subtasks that measured foundational literacy skills, including letter sound knowledge, decoding and word recognition. Although learners in cohort 2 are not reading faster or more accurately than those in cohort 1, they do appear better at foundational skills, noticeably outperforming learners in cohort 1 in mean scores in these subtasks. This is important, as grasping these core skills is key to reading success.

Cohort 2 learners are performing better statistically across all foundational maths skills, though their actual mean performance is only marginally better than learners in cohort 1. There is little difference in performance between the cohorts in level 2 maths subtasks, indicating little change in performance regarding more conceptual maths skills. Cohort 2 learners performed slightly better performance

statistically in both simple and complex word problems, which may be an early indication of more rapidly developing conceptual maths skills that require higher order thinking and processing, though this must be monitored over time to prove true.

Table 16: Subtask mean scores compared by cohort

Subtask	Period	Mean	Attempted	% correct
Letter sounds	Round 1	9.7	26.12	37%
	Round 2	12.3	26.58	46%
Non-words	Round 1	10.9	20.00	55%
	Round 2	13.0	19.78	66%
Familiar words	Round 1	13.9	21.29	65%
	Round 2	15.1	21.38	71%
Oral reading story 1	Round 1	19.4	28.25	69%
	Round 2	20.2	27.10	75%
Oral reading story 2	Round 1	19.4	30.43	64%
	Round 2	19.8	29.03	68%
Number identification	Round 1	12.5	18.58	67%
	Round 2	13.2	18.73	70%
Addition level 1	Round 1	7.9	11.18	71%
	Round 2	8.6	11.14	77%
Subtraction level 1	Round 1	7.1	10.38	68%
	Round 2	7.4	10.27	72%

Table 17: Statistical significance of scores by cohort

Subtask	Round 1		Round 2		Difference
	Mean	Std. Err.	Mean	Std. Err.	
Letter sounds (out of 100)	9.7	0.34	12.3	0.43	2.6***
Non-words (out of 50)	10.9	0.31	13.0	0.33	2.1***
Familiar words (out of 50)	13.9	0.28	15.1	0.31	1.2***
Oral reading story 1 (out of 68)	19.4	0.42	20.2	0.47	0.8
Reading comprehension 1 (out of 5)	1.2	0.02	1.1	0.02	0.1***
Oral reading story 2 (out of 125)	19.4	0.71	19.8	0.74	0.4
Reading comprehension 2 (out of 5)	1.1	0.02	1.0	0.01	0.1*
Listening comprehension (out of 3)	1.6	0.02	1.6	0.02	0
Number identification (out of 20)	12.5	0.13	13.2	0.15	0.7***
Quantity discrimination (out of 10)	5.8	0.07	6.2	0.08	0.4***
Missing number (out of 10)	4.0	0.05	4.7	0.07	0.7***
Addition level 1 (out of 20)	7.9	0.10	8.6	0.11	0.7***
Addition level 2 (out of 5)	3.0	0.04	3.1	0.05	0.1
Subtraction level 1 (out of 20)	7.1	0.11	7.4	0.12	0.3**
Subtraction level 2 (out of 5)	2.9	0.05	2.9	0.06	0
Word problems 1 (out of 4)	2.2	0.02	2.4	0.03	0.2***
Word problems 2 (out of 2)	1.5	0.02	1.6	0.02	0.1***

Figure 9: Percent of correct EGRA subtasks by data collection round (out of attempted)

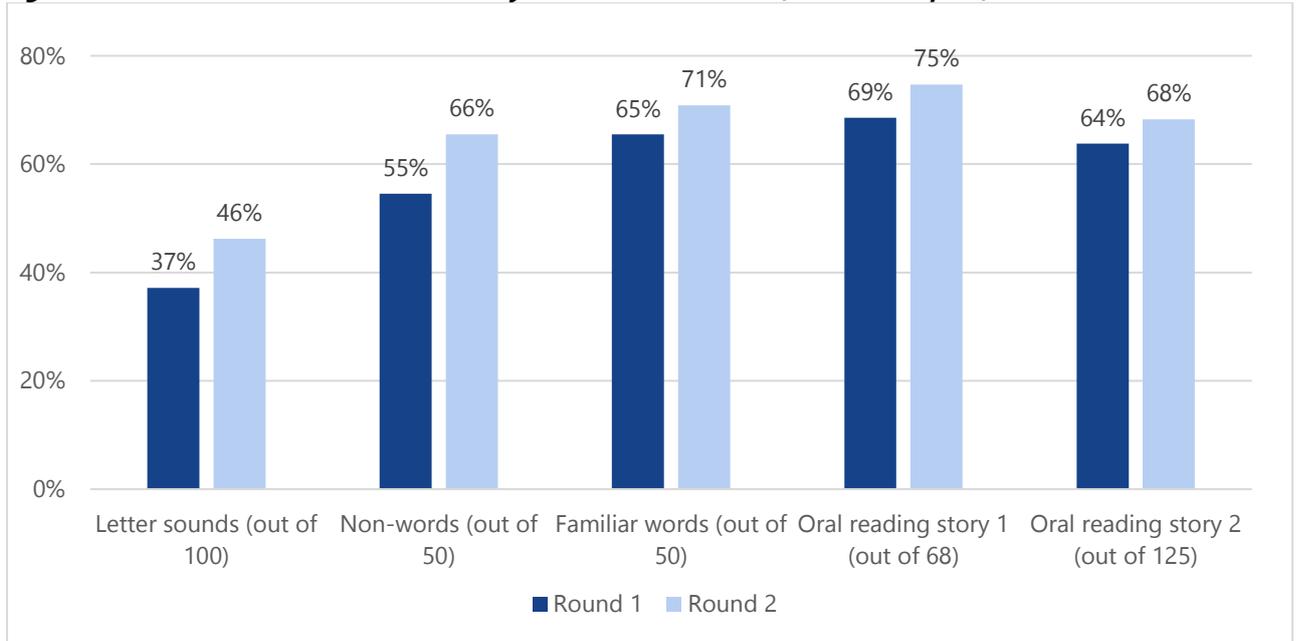
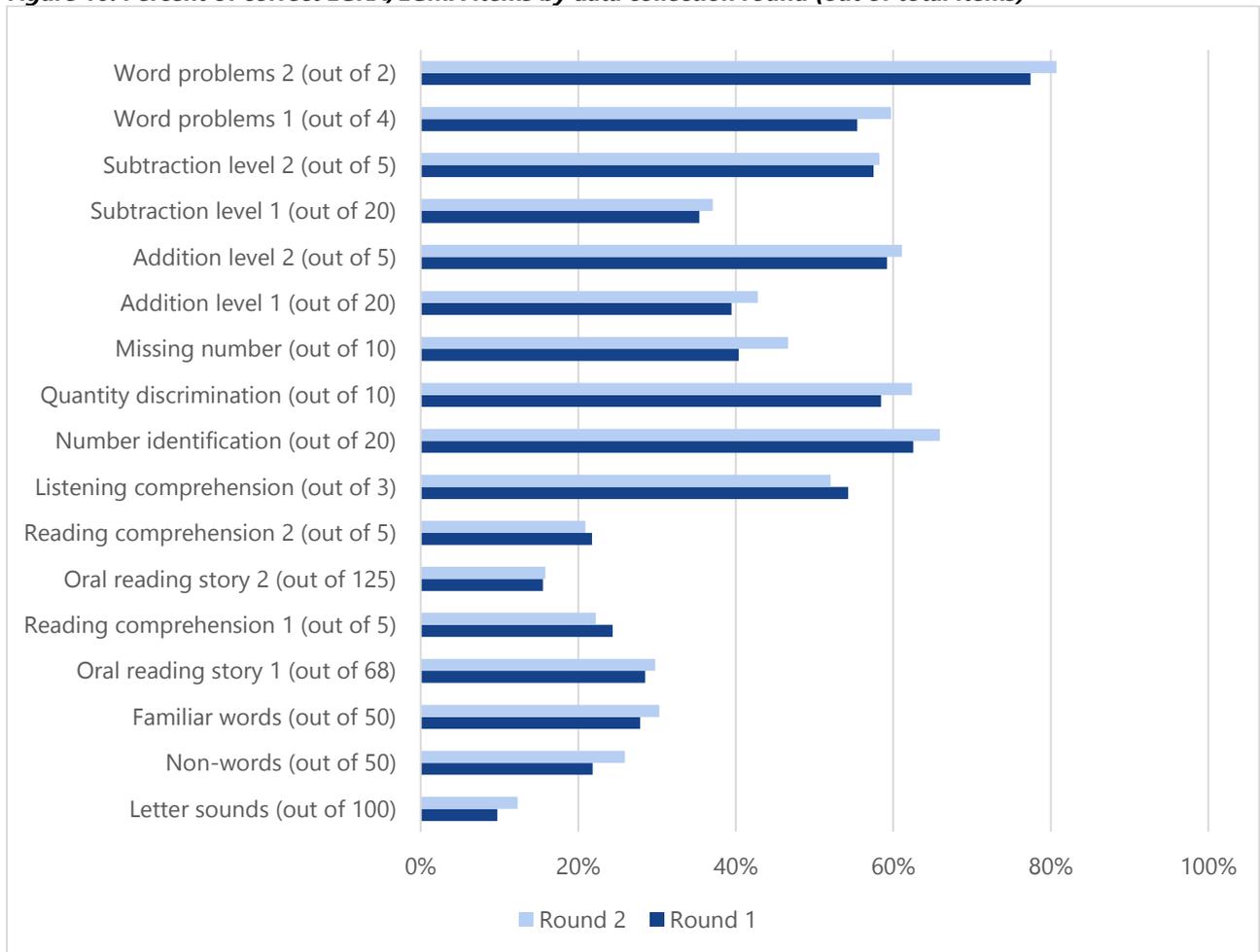


Figure 10: Percent of correct EGRA/EGMA items by data collection round (out of total items)



Cohort Equivalency Finding 2 – Differences Reduce when Cohorts are Compared Week-on-Week Across Survey Rounds: Both mean score differences and the statistical significance of findings across

multiple subtasks reduces between cohorts when learner data is compared week-on-week from the last week of round 1 of the survey to the first week of round 2. Findings are still statistically significant for foundational reading skill subtasks, but not others.

Table 18: Subtask mean scores compared by week and cohort

Subtask	Period	Mean	Attempted	% correct
Letter sounds	Last full week of round 1	10.1	25.92	39%
	First full week of round 2	12.5	26.61	47%
Non-words	Last full week of round 1	10.7	18.81	57%
	First full week of round 2	13.1	19.94	65%
Familiar words	Last full week of round 1	13.6	20.42	67%
	First full week of round 2	15.2	21.38	71%
Oral reading story 1	Last full week of round 1	19.5	27.36	71%
	First full week of round 2	20.2	27.20	74%
Oral reading story 2	Last full week of round 1	18.7	28.89	65%
	First full week of round 2	19.8	28.87	69%
Number identification	Last full week of round 1	13.3	18.77	71%
	First full week of round 2	13.1	18.87	69%
Addition level 1	Last full week of round 1	8.4	11.37	74%
	First full week of round 2	8.7	11.32	77%
Subtraction level 1	Last full week of round 1	7.5	10.50	71%
	First full week of round 2	7.7	10.55	73%

Table 19: Statistical significance of scores by week and cohort

Subtask	Last full week of round 1		First full week of round 2		Difference
	Mean	Std. Err.	Mean	Std. Err.	
Letter sounds (out of 100)	10.1	0.74	12.5	0.60	2.3***
Non-words (out of 50)	10.7	0.52	13.1	0.55	2.4***
Familiar words (out of 50)	13.6	0.45	15.2	0.51	1.5**
Oral reading story 1 (out of 68)	19.5	0.71	20.2	0.75	0.8
Reading comprehension 1 (out of 5)	1.2	0.04	1.1	0.02	-0.1***
Oral reading story 2 (out of 125)	18.7	1.19	19.8	1.17	1.1
Reading comprehension 2 (out of 5)	1.1	0.03	1.0	0.02	0.0
Listening comprehension (out of 3)	1.6	0.04	1.6	0.04	0.1
Number identification (out of 20)	13.3	0.23	13.1	0.24	-0.1
Quantity discrimination (out of 10)	6.1	0.12	6.2	0.13	0.1
Missing number (out of 10)	4.2	0.10	4.7	0.11	0.5***
Addition level 1 (out of 20)	8.4	0.18	8.7	0.17	0.3
Addition level 2 (out of 5)	3.0	0.07	3.1	0.08	0.1
Subtraction level 1 (out of 20)	7.5	0.19	7.7	0.18	0.2
Subtraction level 2 (out of 5)	3.1	0.09	3.0	0.10	-0.1
Word problems 1 (out of 4)	2.4	0.04	2.4	0.04	0.1
Word problems 2 (out of 2)	1.6	0.03	1.6	0.03	0.0

Cohort Equivalency Finding 3 – Limited Variation in Scores for Cohort 2 Week-on-Week: Almost no variation is seen in performance among cohort 2 week-on-week in the survey, which was expected given that the data was collected back-to-back with only a weekend break. Learners performed nearly identically in mean scores, except for subtasks related to comprehension, where learners in week 2 slightly outperformed learners from week 1. Importantly, tracking learner progress developmentally over time may provide a better understanding of how learners acquire and apply new knowledge and skills to their learning, with the ultimate goal of achieving a high level of comprehension and conceptual thinking.

Table 20: Subtask mean scores compared by week for cohort 2

Subtask	Period	Mean	Attempted	% correct
Letter sounds	Round 2, Week 1	12.5	26.61	47%
	Round 2, Week 2	12.9	26.49	49%
Non-words	Round 2, Week 1	13.1	19.94	65%
	Round 2, Week 2	13.2	19.57	68%
Familiar words	Round 2, Week 1	15.2	21.38	71%
	Round 2, Week 2	15.7	21.71	72%
Oral reading story 1	Round 2, Week 1	20.2	27.20	74%
	Round 2, Week 2	20.8	27.40	76%
Oral reading story 2	Round 2, Week 1	19.8	28.87	69%
	Round 2, Week 2	21.0	30.14	70%
Number identification	Round 2, Week 1	13.1	18.87	69%
	Round 2, Week 2	13.4	18.69	72%
Addition level 1	Round 2, Week 1	8.7	11.32	77%
	Round 2, Week 2	8.6	11.12	78%
Subtraction level 1	Round 2, Week 1	7.7	10.55	73%
	Round 2, Week 2	7.4	10.11	73%

Table 21: Statistical significance of scores by week for cohort 2

Subtask	Round 2, Week 1		Round 2, Week 2		Difference
	Mean	Std. Err.	Mean	Std. Err.	
Letter sounds (out of 100)	12.5	0.60	12.9	0.75	2.7***
Non-words (out of 50)	13.1	0.55	13.2	0.49	2.6***
Familiar words (out of 50)	15.2	0.51	15.7	0.48	2.1***
Oral reading story 1 (out of 68)	20.2	0.75	20.8	0.70	1.3
Reading comprehension 1 (out of 5)	1.1	0.02	1.2	0.03	0.0
Oral reading story 2 (out of 125)	19.8	1.17	21.0	1.21	2.2
Reading comprehension 2 (out of 5)	1.0	0.02	1.1	0.03	0.0
Listening comprehension (out of 3)	1.6	0.04	1.5	0.03	-0.1
Number identification (out of 20)	13.1	0.24	13.4	0.23	0.1
Quantity discrimination (out of 10)	6.2	0.13	6.3	0.12	0.2
Missing number (out of 10)	4.7	0.11	4.8	0.10	0.5***
Addition level 1 (out of 20)	8.7	0.17	8.6	0.16	0.2
Addition level 2 (out of 5)	3.1	0.08	3.0	0.07	0.0
Subtraction level 1 (out of 20)	7.7	0.18	7.4	0.18	-0.1
Subtraction level 2 (out of 5)	3.0	0.10	2.9	0.09	-0.2

Word problems 1 (out of 4)	2.4	0.04	2.4	0.04	0.0
Word problems 2 (out of 2)	1.6	0.03	1.6	0.03	0.0

Cohort Equivalency Finding 4 – Potential Reasons for Performance Variation: Given that the demographics of the sample were the same across cohorts 1 and 2, reasons for variation in performance could be down to a specific factor, or mix of factors, that affected results. While it is not possible to deduce the exact reason for the performance improvement based upon the data, the following factors may have influenced the overall outcome. They include:

1. A longer time spent in school, and on task for learners in cohort 2, allowing their capabilities to improve over time.
2. Enumerator error in assessment administration (in the favour of cohort 2) due to fatigue, carelessness or lack of focus. This can affect many group members at the same time and can be common towards the end of the assessment phase.
3. The teachers and/or learners in cohort 1 shared information about the assessment with teachers and/or learners in cohort 2.
4. Greater focus and ability in cohort 2 learners to adapt to and take an assessment, given the overlap of phase 2 data collection and end of year exams. Children might have been more accustomed to studying and testing, giving them a greater attention span and attention to detail that could have been applied to the assessment.

5. LEARNER PROFILES AND PROGRESS TRACKING 2014 TO 2021

A recognised challenge of large-scale education research projects is determining the most appropriate way to effectively report findings for a wide variety of audiences (e.g., researchers, ministry officials, donors, and other relevant stakeholders). It is important to consider technical rigor and accessibility, while ultimately providing results that can be used to inform policy and instruction.

By categorising learners into learning profiles based on their performance, we can better track progress over time, pinpoint their particular instructional needs and inform curriculum reform. In the report, we divide learners into learning profiles based on their reading ability (non-reader, pre-reader, emerging, progressing, proficient and exceeds expectations), and then examine the relationship among these profiles and their reading skills on a variety of EGRA subtasks to determine the instructional requirements to promote learners from one profile to the next. A similar process is followed for the EGMA.

This approach allows us to map measurable progress in learners' reading and maths skill development across the assessments, developmentally and over time. By allowing us to compare the distributions of learner profiles across proficiency levels, rather than looking only at the milestone achievement level in terms of overall or mean scores, the results will give us far more information about the interventions needed to ensure learners master key foundational skills in the early grades as they progress to fluent readers who comprehend text.

The following tables and graphs track learner achievement against core reading and maths profiles, which are measured across a progressive developmental scale as learners develop their skills over time. Results from the 2021 baseline are compared to data from the 2014 national EGRA and EGMA assessment using learner profile bands. They are provided here as a means of reflecting on changes to date in early grade reading and maths performance among learners.

Though they do not offer a direct statistical comparison across subtasks, it is possible to explore changes in the number of non-readers in the population and the degree to which learners were able to attempt items on specific subtasks. These reflections may provide some evidence of whether there have been any improvements in learning outcomes since the introduction of initiatives post-2014 aimed at improving the teaching of reading.

5.1. Learner profiles for reading

Learner Profiles Finding 1 – Reading Fluency and Comprehension: Learner profiles for oral reading fluency indicate that learners are progressing as readers from grade 2 to grade 4, with the share of non-readers in grade 4 reducing significantly from that in grade 2. Significantly more grade 4 learners are at an emerging or progressing level in reading fluency than those in grade 2, indicating an upward, positive trend in their literacy development. However, trends in reading comprehension do not follow suit, with limited progress between grades.

The majority of learners in grades 2 and 4 are not able to comprehend the text they read (64-73 per cent), while only 25-30 per cent are at an emerging skill level. These findings suggest that, while learners are increasingly able to identify words and read text fluently, their understanding of those words and content is limited, as is their conceptual ability to apply concepts learned to other tasks. Increasing comprehension skills must be a central goal on instruction.

Getting more learners reading more of the time, reducing non-readers and growing the proportion of capable readers who can comprehend text – both within and between grades – must be a clear goal of any national literacy program. Using learning profiles to monitor and track this achievement in individuals and groups of learners provides a clear mechanism for diagnosing progress, communicating results, and getting teachers and parents on board with discussing and monitoring change.

Table 22: Learner profiles for oral reading fluency by grade

Pupil profile oral reading fluency	Grade 2	Grade 4
Non-reader (0 CWPM)	57.2%	17.0%
Pre-reader (1 - 5 CWPM)	11.1%	18.5%
Emerging reader (6 - 15 CWPM)	19.0%	34.4%
Progressing reader (16 - 28 CWPM)	6.8%	18.2%
Proficient reader (29 - 40 CWPM)	1.4%	4.2%
Exceed expectations reader (41+ CWPM)	4.6%	7.6%

Figure 11: Distribution of grade level oral reading fluency scores by learner profile

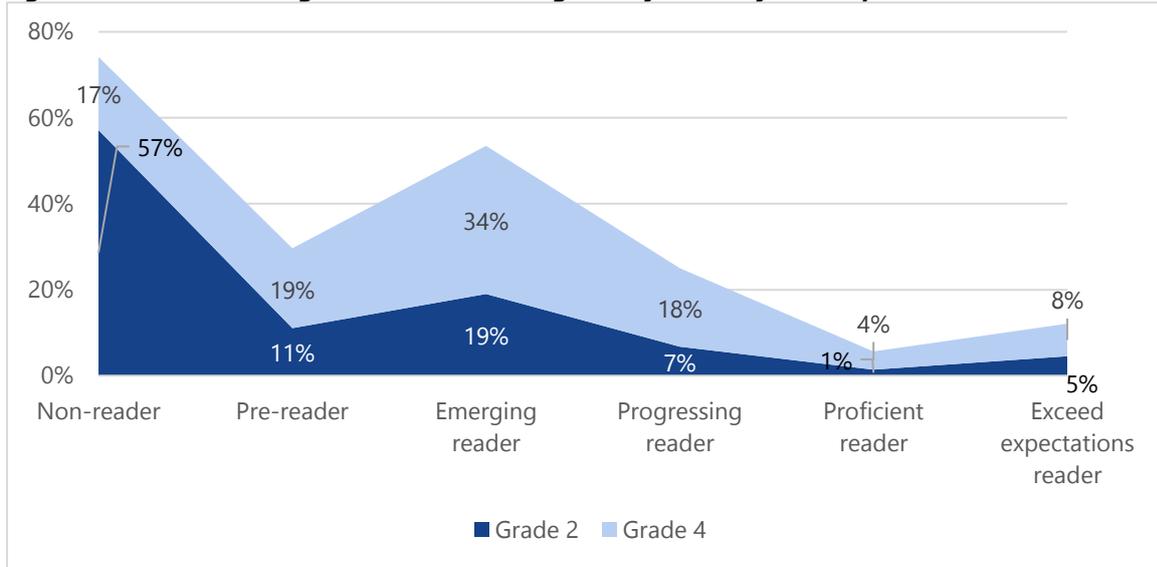


Table 23: Learner profiles for reading comprehension by grade

Pupil profile	Grade 2	Grade 4
Non-reader/Pre-reader (0 questions correct)	73.0%	64.0%
Emerging reader (1 question correct)	25.3%	30.1%
Progressing reader (2-3 questions correct)	0.0%	5.4%
Proficient reader (4 questions correct)	1.5%	0.3%
Exceed expectations reader (5 questions correct)	0.3%	0.3%

Figure 12: Distribution of grade level reading comprehension scores by learner profile

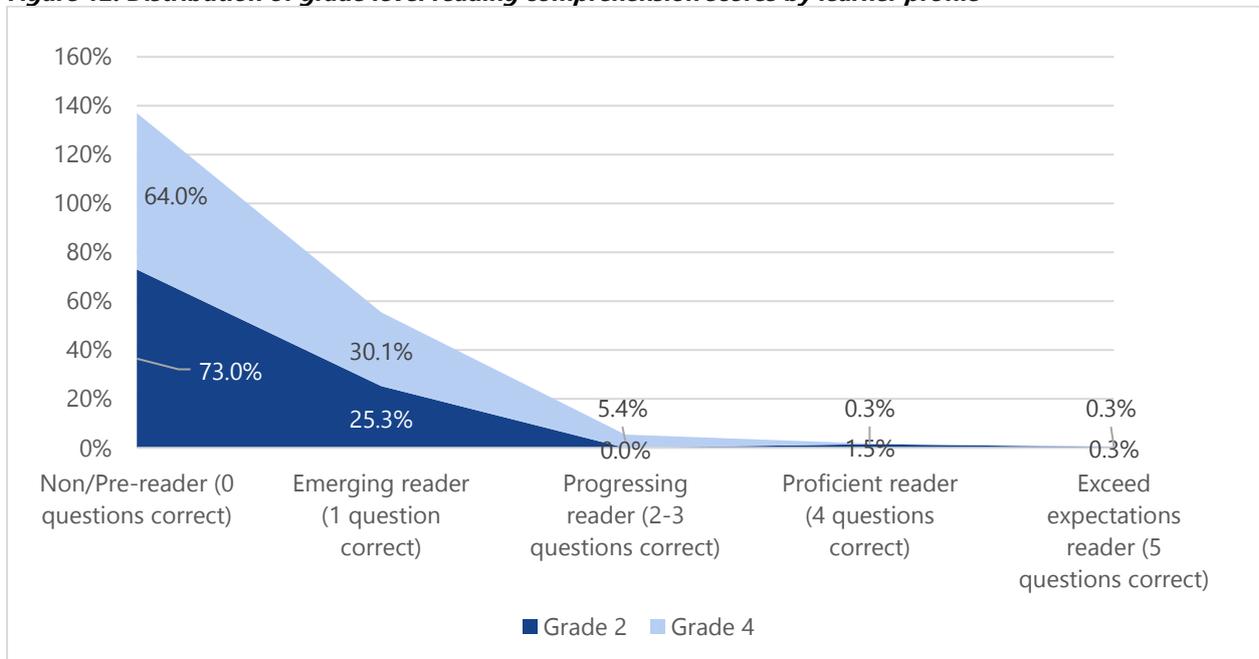


Figure 13: Distribution of grade level letter sound scores by learner profile

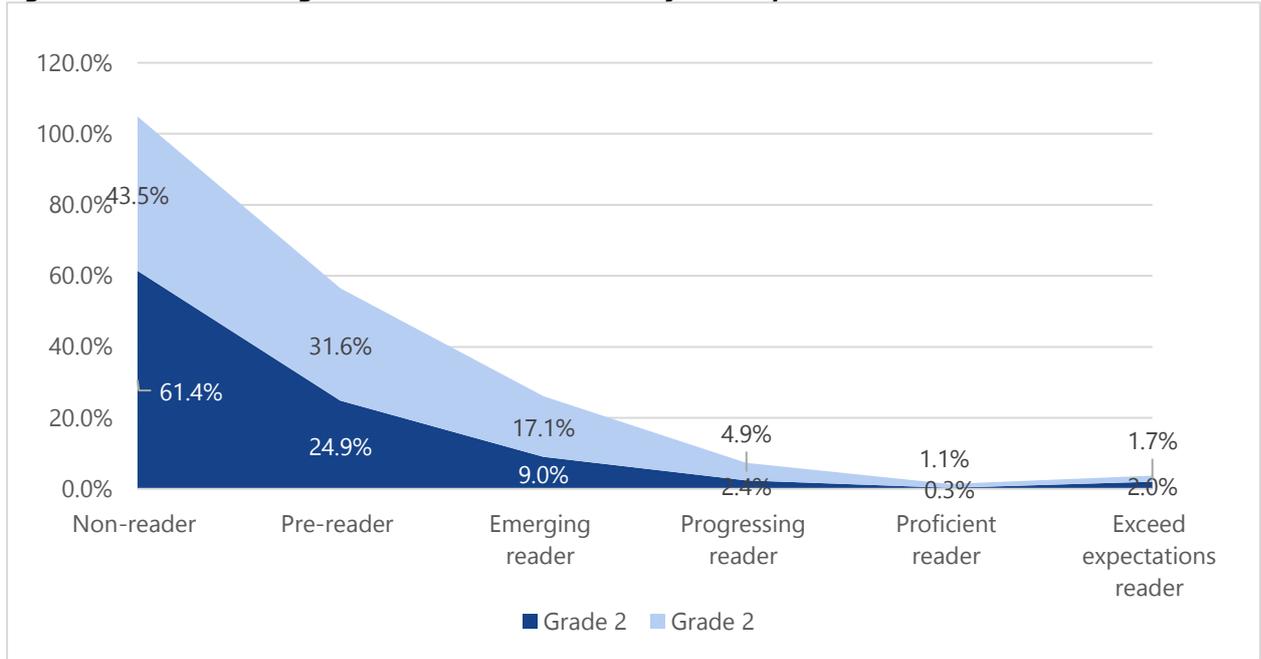


Figure 14: Distribution of grade level familiar word scores by learner profile

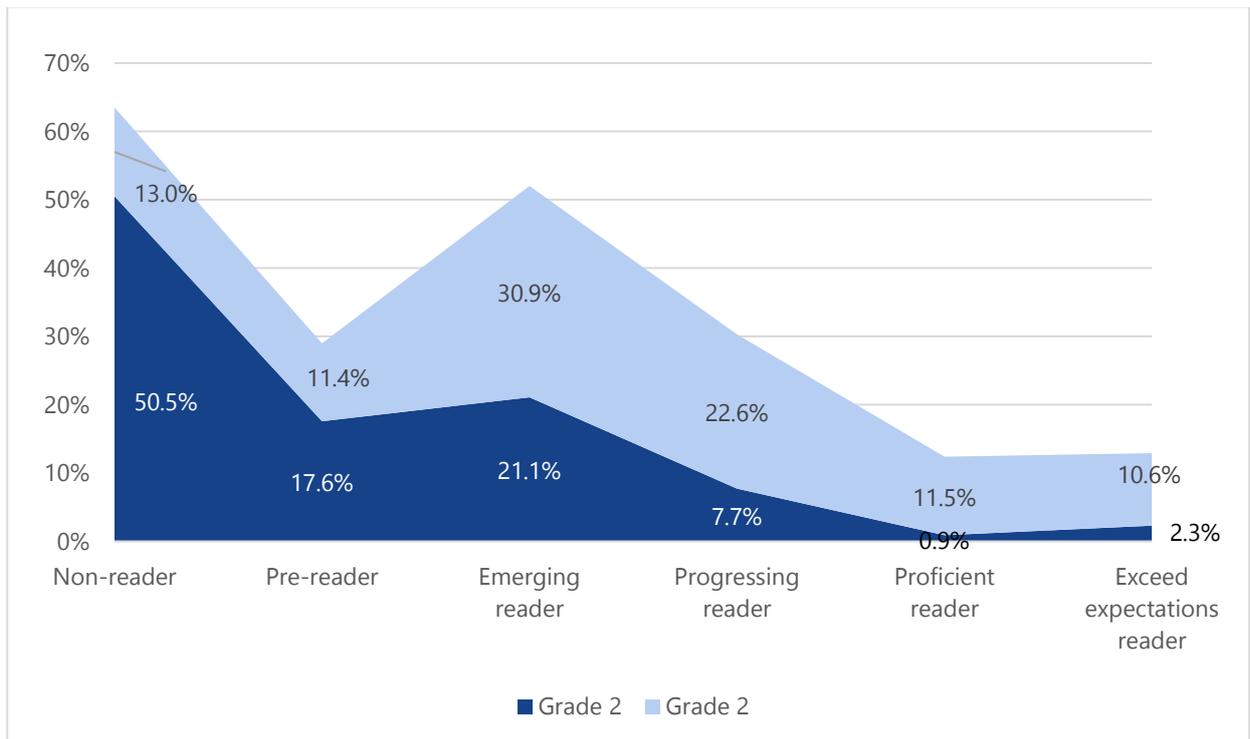


Figure 15: Distribution of grade level unfamiliar word scores by learner profile

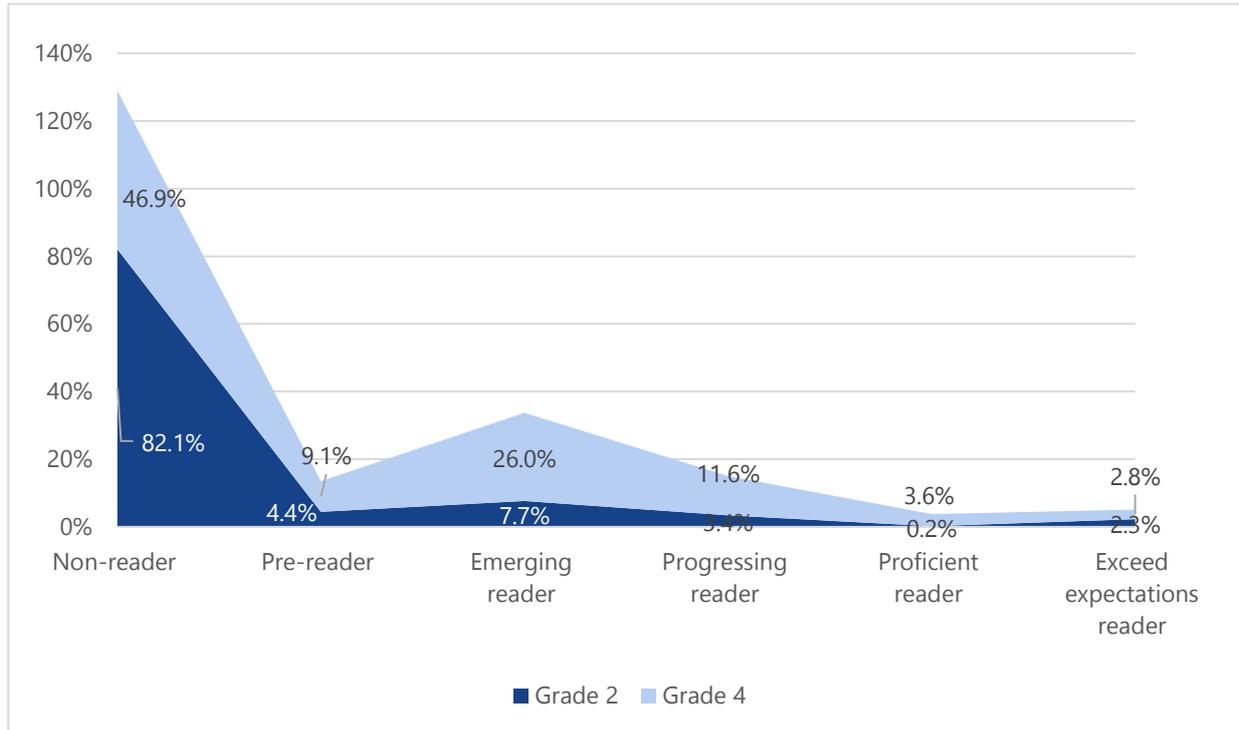
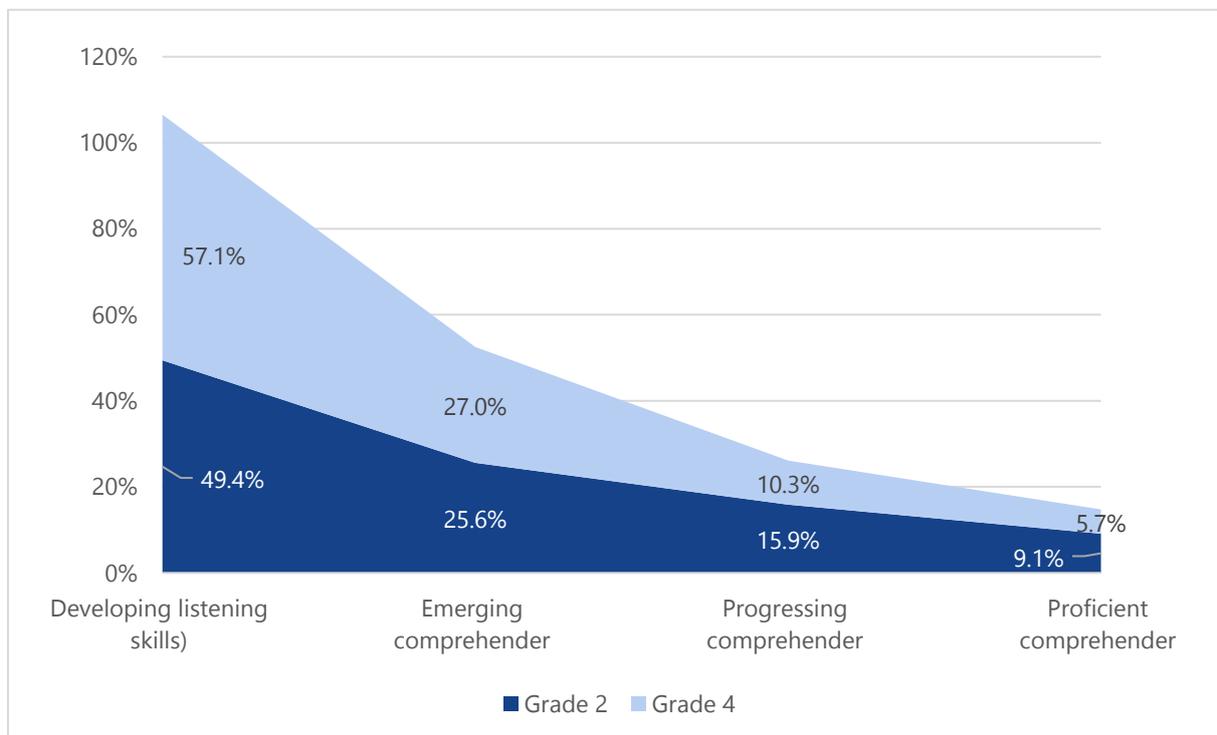


Figure 16: Distribution of grade level listening comprehension scores by learner profile



Learner Profiles Finding 2 – School Language and Literacy: The language spoken at school does impact learners’ scores. As previously described in the learning outcomes section on the EGRA, when English is the school language learners perform better, especially in literacy tasks, with an increase in scores of up to 8 points. Conversely, when the school language is Krio or another local language, learner performance is negatively affected.

Learner Profiles Finding 3 – Home Language and Literacy: The language spoken at home plays an even greater role: speaking English at home relates to an increase in literacy scores of up to 24 points. Again, effects are stronger for literacy tasks, while scores for numeracy tasks are only improved by 1-2 points. Scores are negatively impacted when the home language is a local language different from Krio, while speaking Krio at home has a slightly positive effect. The links between speaking Krio at home and children performing better in English literacy tasks are not clear; a possible explanation is that speaking Krio (as opposed to another local languages) is correlated with other household characteristics such as the education level of the parents, their location in an urban centres and access to electricity, which in turn can help learner performance.

5.2. Learner profiles for numeracy

Learner Profiles Finding 4 – Numeracy Development: Learner profiles for missing number indicate that learners are gaining developmental maths skills from grade 2 to grade 4, with the majority performing at the progressing level in grade 4, just below proficiency level where they can work independently and apply maths skills to conceptual tasks. Significantly more grade 4 learners are at a progressing level in key numeracy tasks than those in grade 2, indicating an upward, positive trend in their maths skills development and overall fluency. However, trends in subtraction do not follow suit, with limited progress between grades.

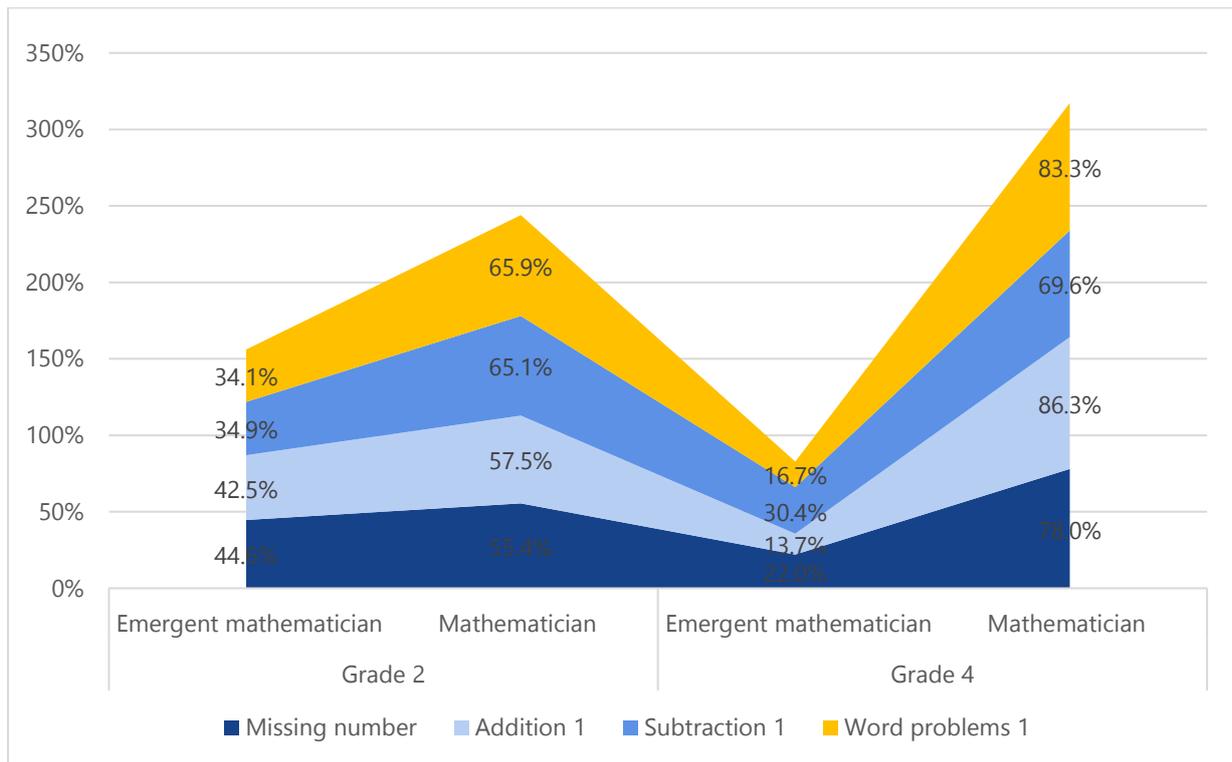
The majority of learners in grade 2 are at an emerging skill level, meaning they are still trying to master basic number facts and computational skills, while their conceptual capacities are still developing. Grade 4 learners demonstrate significant improvement in basic maths skills, as well as application of concepts to higher level tasks.

These findings suggest that, while learners are increasingly able to complete basic computations, their understanding of mathematical facts and strategies applied in new ways is limited, as is their conceptual ability to apply processes learned to other tasks. Increasing learners' conceptual skills must be a central goal of instruction; this can be achieved, in part, by ensuring learners are confident in their computational skills and knowledge and application of number facts.

Getting more learners confident doing maths and capable of applying concepts readily to new equations must be a clear goal of any national maths and science programme. Using learning profiles to monitor and track this achievement in individuals and groups of learners provides a clear mechanism for diagnosing progress, especially in key science, technology, engineering and maths (STEM) subject areas.

Table 24: Learner profiles for numeracy by grade

Subtask	Grade 2		Grade 4	
	Emergent	Progressing	Emergent	Progressing
Missing number	44.6%	55.4%	22.0%	78.0%
Addition 1	42.5%	57.5%	13.7%	86.3%
Subtraction 1	34.9%	65.1%	30.4%	69.6%
Word problems 1	34.1%	65.9%	16.7%	83.3%

Figure 17: Distribution of grade level numeracy scores by learner profile

5.3. Progress tracking 2014 to 2021

As accurately noted in the 2014 national EGRA and EGMA report, it is generally not recommended to directly compare EGRA statistics like mean scores or the proportion of readers achieving a numerical fluency target across multiple countries or languages. This is because the structure of languages – such as the length and complexity of words, grammatical structures, writing systems and symbols – varies tremendously, as do contextual factors such as instructional strategies, delivery systems and language of instruction policies that affect learning. As such, direct comparisons of reading fluency scores across languages and locations holds limited meaning, and while possibly interesting to explore and examine, must be taken in context. Moreover, measuring competency or fluency by a singular metric or numerical value – such as a correct word per minute target – is often restrictive and uninformative, frequently serving only to measure deficits in children’s abilities rather than assessing their progressive mastery of key competencies and comprehension skills over time. In keeping, utilising a singular score like correct words per minute to compare learning outcomes across diverse groups of learners in widely spaced academic years using different tests presents limitations, especially if the findings are used in a strictly evaluative capacity as they lack direct comparability from a statistical perspective.

Yet, comparing zero scores across learner cohorts, even in different geographies and languages, is possible because ‘zero’ has the same weight in every system. It also provides a fairly universal indicator of low or illiteracy in a language, as non-readers in a specific category are unable to read any letters or words, or in other words, demonstrate that reading skills.

Learner Profiles Finding 5 – Comparing Zero Scores in 2014 and 2021: Zero scores in letter sound, non-word, familiar word, oral reading and reading comprehension were identified and compared from 2014 and 2021 assessment data. Noting the comparability of these results, described above, clear trends in performance emerge from the data. In grades 2 and 4 in 2014, the overwhelming majority of learners were non-readers of letters, words and texts. There were few differences in grade-level performance on

zero scores for many subtasks in 2014 and 2021, though there was significant improvement in performance from the first assessment in 2014 to the one administered in 2021, with a strong reduction in non-readers on this subtask between years. Other subtasks in the EGMA follow a similar pattern in zero scores, with higher numbers of zero scores in the 2014 study than in the 2021 survey, and grade level reductions in zero scores both times.

Figure 18: Proportion of learners scoring zero on EGRA subtasks in 2014 and 2021

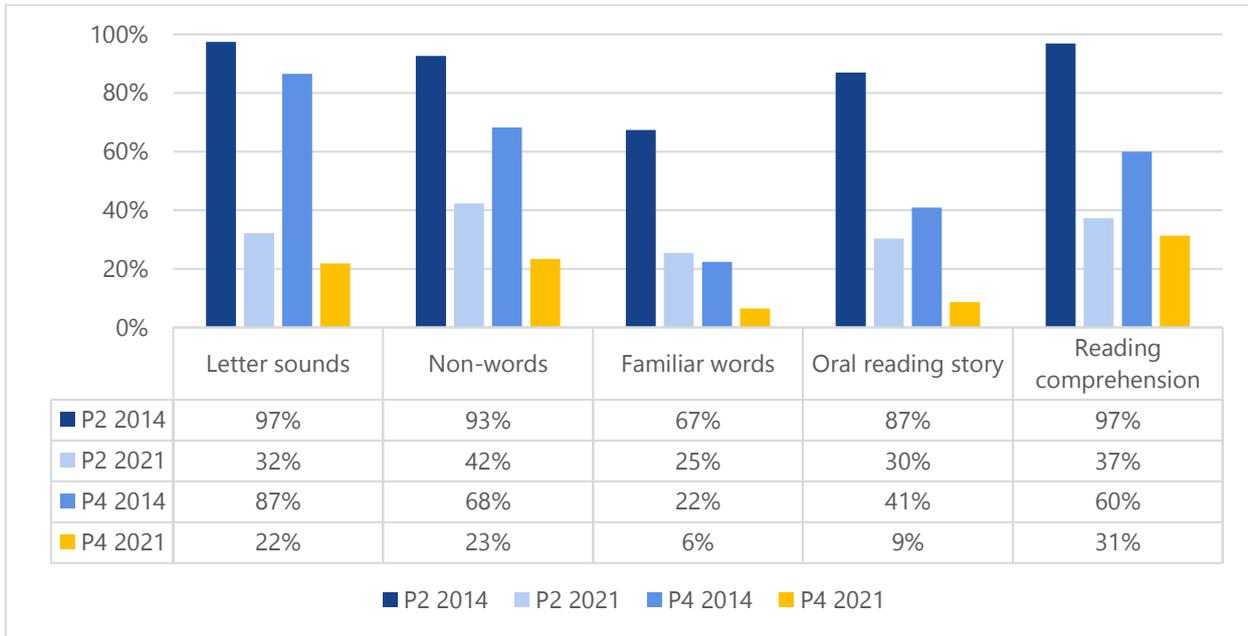
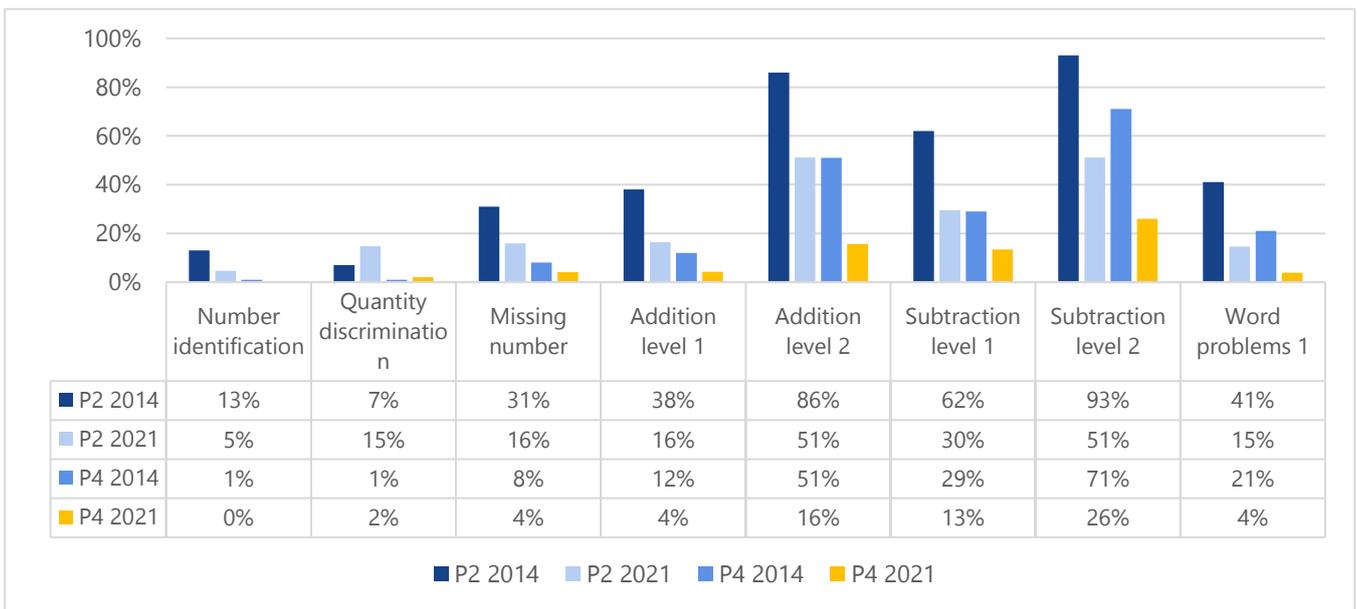


Figure 19: Proportion of learners scoring zero on EGMA subtasks in 2014 and 2021



Learner Profiles Finding 6 – Comparing Learner Profiles in 2014 and 2021: As mentioned, comparison of assessment data between two diverse cohorts of learners across two widely spaced years like 2014 and 2021 – and using distinctive tests with potentially divergent levels of complexity, untested item reliability, differing subtask design, item selection and length – should not be used to prove any statistically significant results. Rather, such data can help to identify trends in learning that may emerge between those two years when examining the evidence for patterns.

To support this process, the 2021 learner profiles for EGRA and EGMA were slightly restructured to align with the performance bands created for the 2014 EGRA and EGMA assessment. Learners were assigned into each category based on their performance on a subtask so grade-level results across both evaluation years could be reviewed and compared. Overall, there are differences in performance by grade within the 2021 assessment year, though they do not appear in the 2014 data. There are also greater variations in performance in 2021, with learners demonstrating a wider distribution of key competencies. 2014 results are largely clustered between grades, demonstrating little grade-level progression in key skills.

Figure 20: Letter sound fluency 2014 and 2021

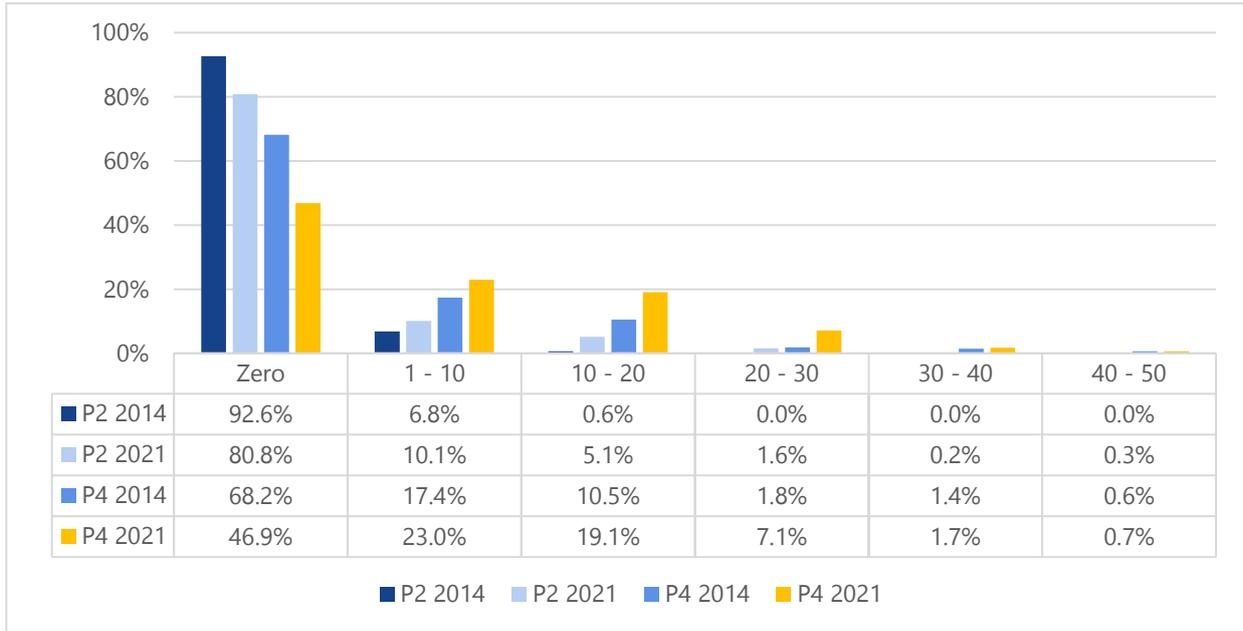


Figure 21: Invented word fluency 2014 and 2021

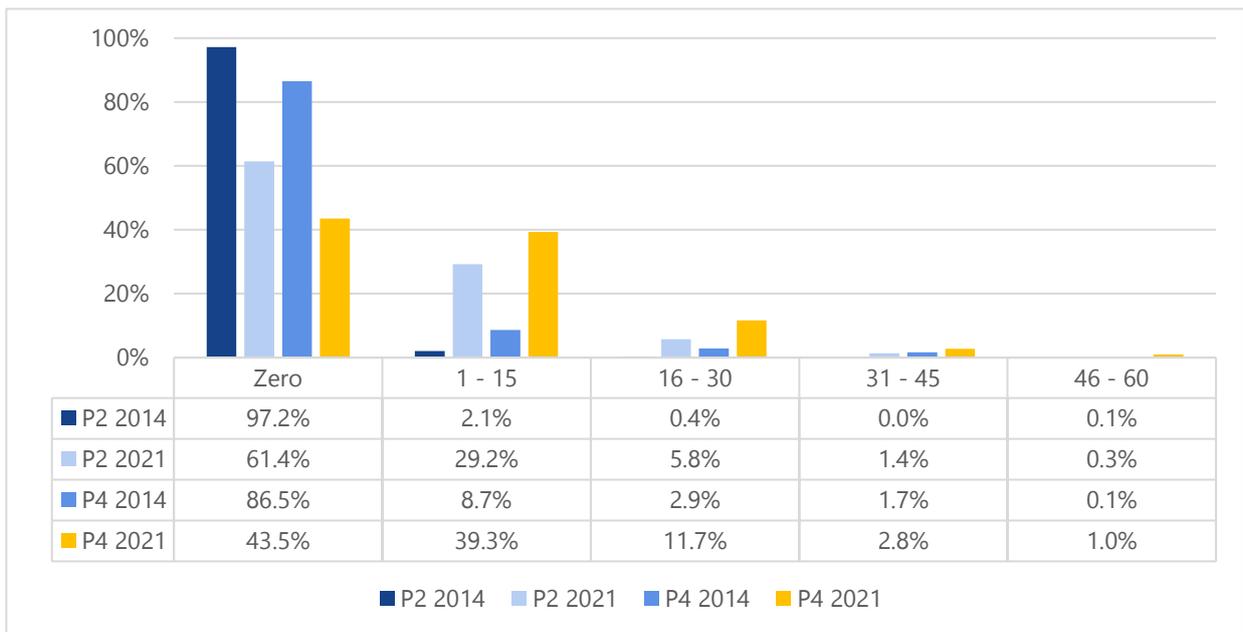


Figure 22: Familiar word fluency 2014 and 2021

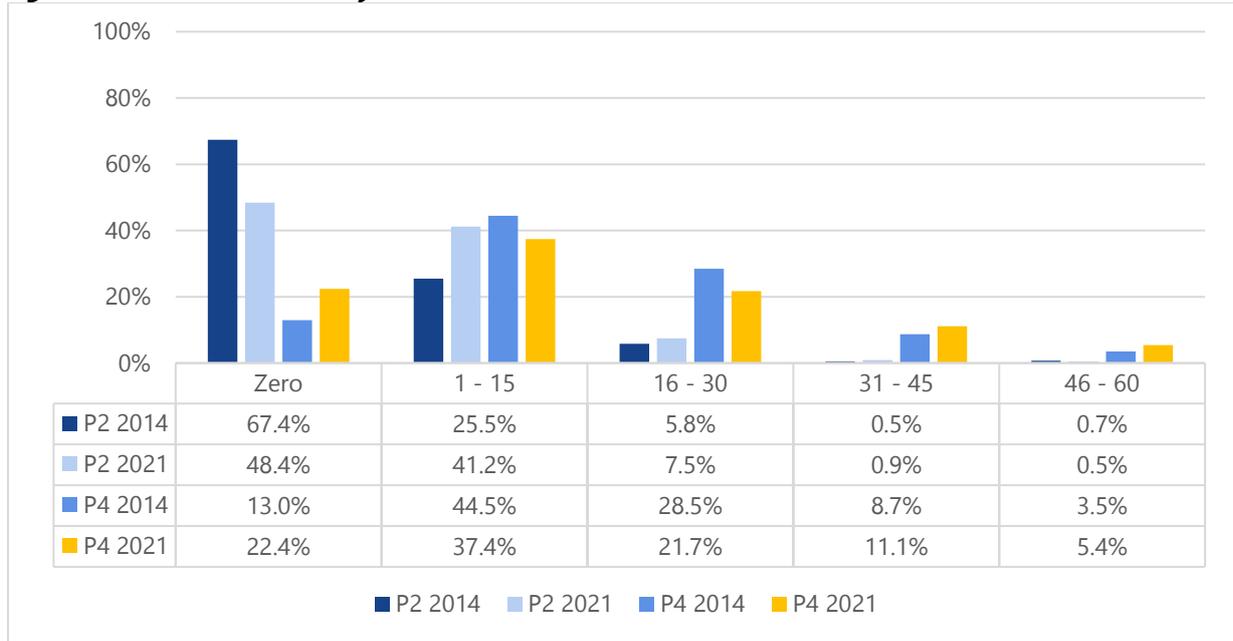


Figure 23: Oral reading fluency 2014 and 2021

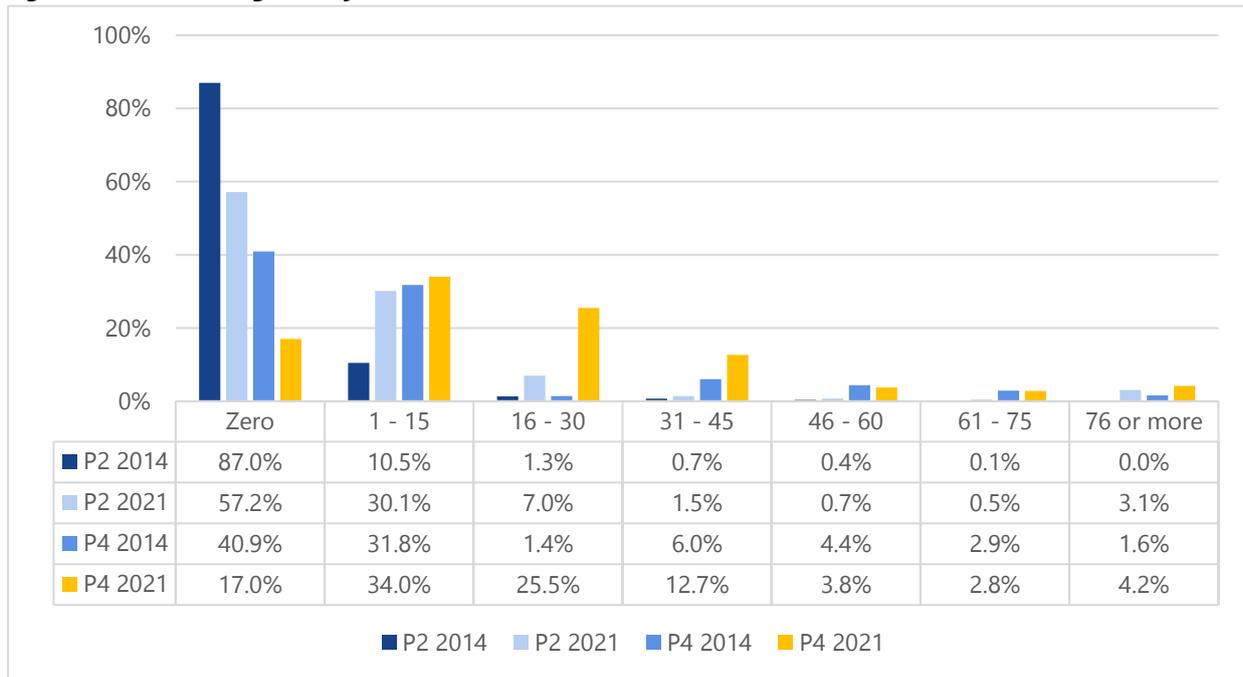
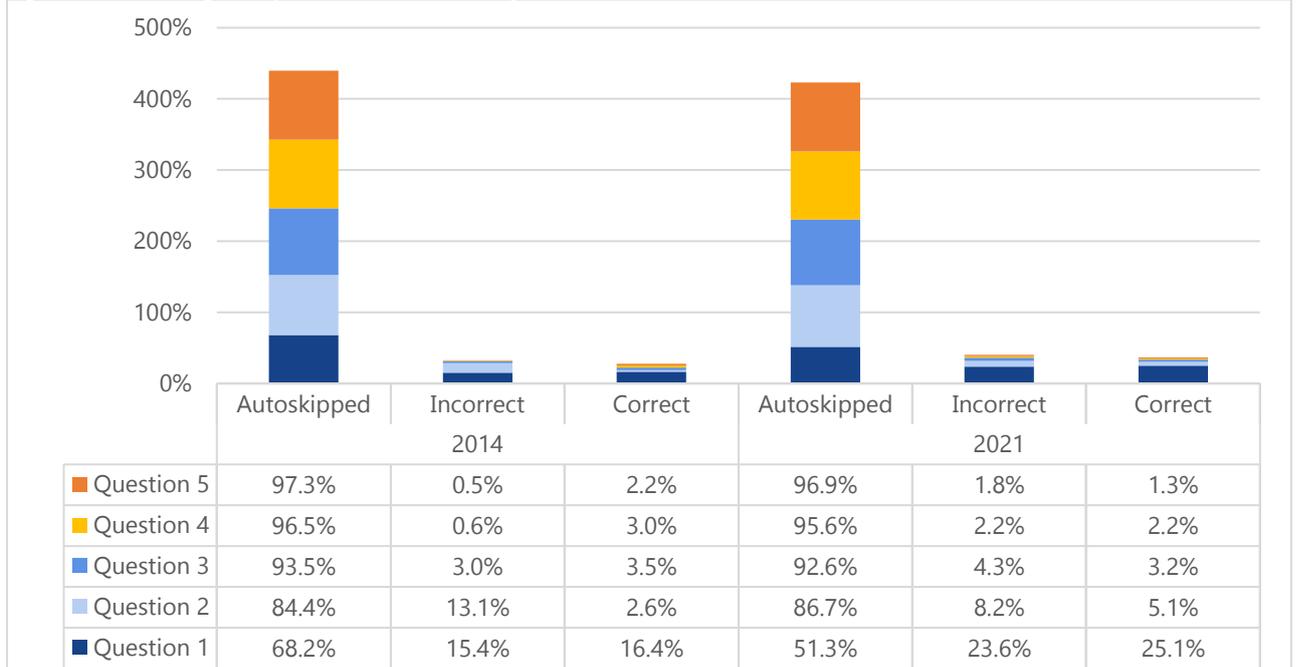
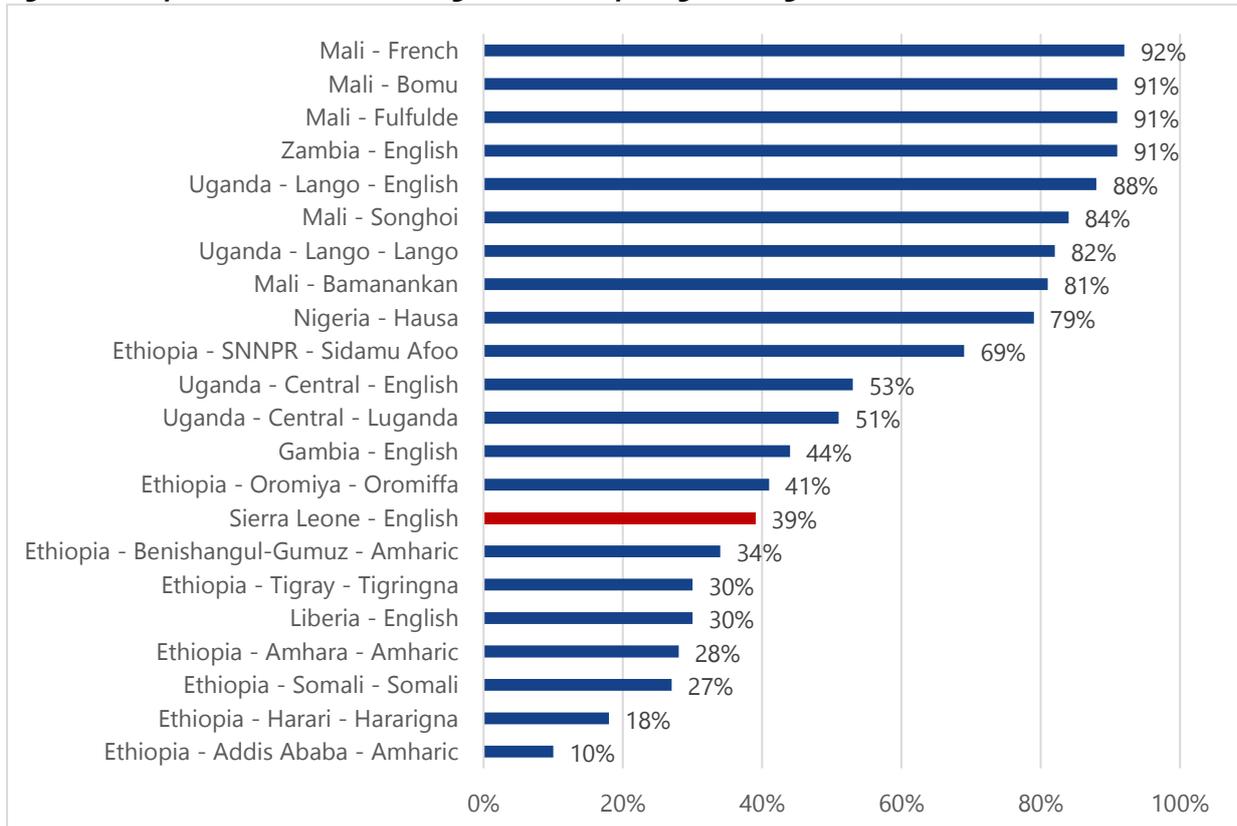


Figure 24: Reading comprehension item analysis 2014 and 2021



Learner Profiles Finding 6 – Comparing Zero Scores Internationally: Below is a graph from a brief by RTI International on Assessing Early Grade Reading Skills in Africa. Most EGRA applications in Africa reveal that a high percentage of learners are unable to read at all. The data from that chart is compared to findings from the 2021 study in Sierra Leone for learners with zero scores on the oral reading fluency test. Sierra Leone falls in the upper half of countries in this analysis, demonstrating fewer zero scores than other countries, especially those administering English language assessments.

Figure 25: Proportion of learners scoring zero on oral passage reading across countries



6. FINDINGS ON TEACHERS

The following section details key findings on grade 2 and 4 teachers in the study. Data is presented on their certification and professional development experiences, their exposure to literacy instruction and training, and opinions and expectations about learner performance and competencies in the early grades. This information is important to inform findings and recommendations on instructional and pedagogical interventions that will help teachers be more effective and successfully raise literacy and numeracy outcomes for developing learners in line with grade level expectations.

6.1. Teacher workforce profile and management

Teacher Finding 1 – Certification: Nearly 25 per cent of the teachers surveyed have only a basic education or West African senior school certificate, meaning they have not received certified training beyond primary or secondary school to become a teacher. Other teachers surveyed have either no certification whatsoever (about 1 per cent) or have identified a (possibly) non-recognised certification that they hold outside of teaching (about 11 per cent). Positively, about 50 per cent of teachers surveyed have a teacher’s certificate, while another 12 per cent have a higher teacher certificate for either primary or secondary. Just over 2 per cent have a university level certification in teaching or education at the diploma, degree or master’s level.

Table 25: Teacher Certification Levels

Highest education level	Percent
Teacher Certificate	49.2%
Higher Teacher Certificate Primary	10.7%
Higher Teacher Certificate Secondary	1.7%
Basic Education Certificate	3.6%
West African Senior School Certificate	20.6%
Diploma in Teaching	1.5%
Bachelor’s in education or other	0.6%
Master’s in education or other	0.2%
None	0.6%
Other	11.3%

Teacher Finding 2 – Experience: On average a teacher has worked 8.6 years at the current school and 11.6 years in total. The average age of teachers is 39 years. The overall picture is that of an experienced work force. And one with long-term ties, investment and time spent in their current school and working with its communities.

Teacher Finding 3 – Years teaching verses years certified: There is an average difference of 6 years between total years a teacher has worked and the years since they have acquired a certificate. This indicates that it has been common for teachers to start working without a certificate, though many acquire one later. Indeed, 70 per cent of uncertified teachers have been working for 6 years or less. A substantial number of uncertified teachers – 22 per cent (89 teachers) – have more than 10 years of experience.

These findings can help to inform reflections on the country’s teacher workforce management strategy, especially for the recruitment and placement of new hires and for the possible reallocation and distribution of certified teachers already in the system. Moreover, these findings indicate that the upgrading and certifying of in-service teachers through targeted training will be critical to ensure there

continues to be a consistent, immediate supply of skilled teachers in the system to meet current demand. Furthermore, it should inform government’s plans for the review of the roles of community and volunteer teachers.

6.2. Professional development and literacy instruction

Teacher Finding 4 – Professional Development: Positively, nearly 40 per cent of the teachers surveyed received professional development and training within the last two years, regardless of their certification status; however, the majority – 60 per cent – did not (regardless of certification status). It is not clear whether these teachers were not offered training or whether if offered, they declined. On average, each teacher who benefitted participated in two trainings. Teachers were mostly trained on teaching literacy and numeracy, classroom instructional practices and lesson planning. Other trainings included approaches to guidance and counselling and inclusive education.

Teacher Finding 5 – Performance on Letter Sounds: Just 35 per cent of teachers have received in-service training in how to teach letter sounds. They also struggle to demonstrate this competency, given that it is a skill that must be practiced and mastered, regardless of a teacher’s first language: when asked to say the sounds of five English letters, teachers only pronounced three, on average, correctly. These findings highlight gaps in teacher’s instructional knowledge regarding a set of key foundational skills critically needed for reading success – letter-sound knowledge, phonological awareness and alphabetic principle. Teacher findings corroborate EGRA results, where over 50 per cent of learners were unable to pronounce even one letter sound correctly, indicating a lack of knowledge and skill across both groups regarding phonics and phonemic awareness. Explicit training in literacy instruction is urgently needed for teachers to address these disparities for themselves and their learners.

Figure 26: Teachers Receiving Recent Professional Development

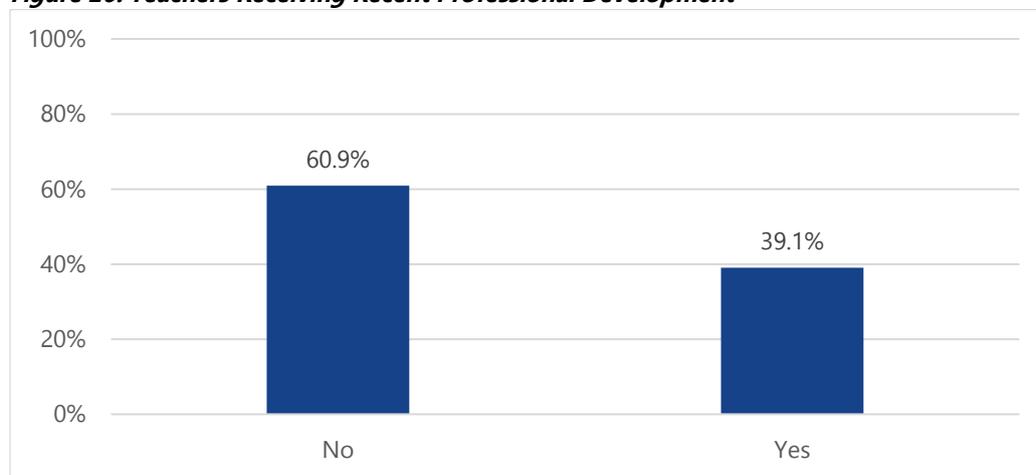


Table 26: Mean number of professional development trainings received

Subtask	Mean	Std. Err.
Number of professional development trainings received	2.1	0.10

Figure 27: Professional Development Training Content

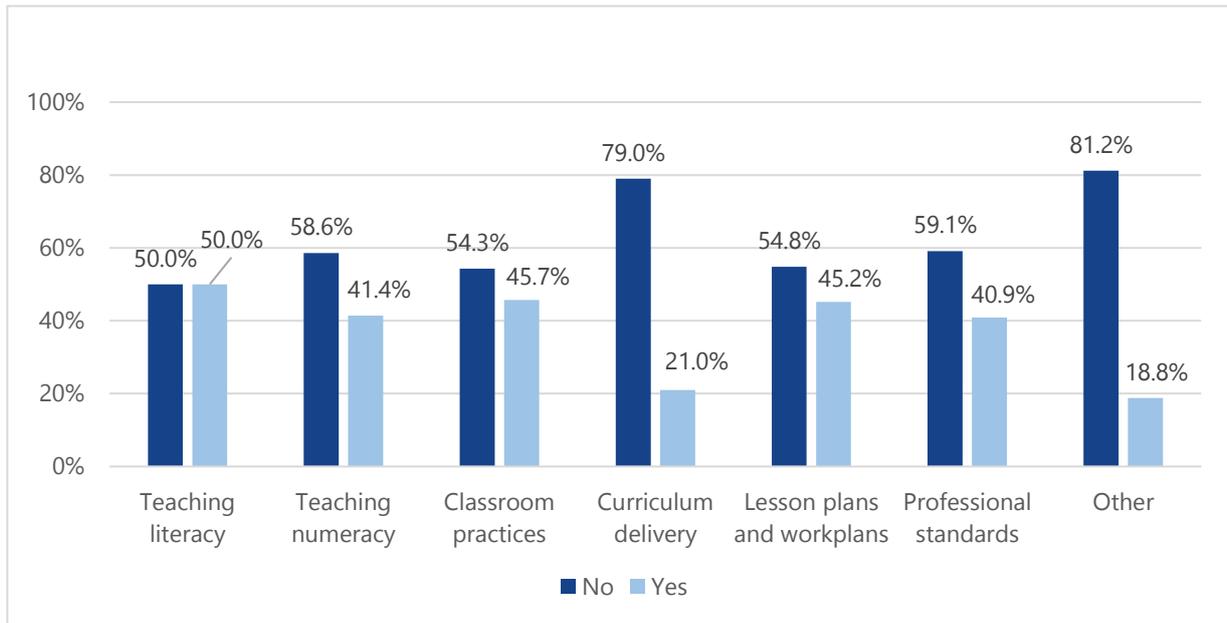


Table 27: Other Training Content

Other training content	Frequency	Percent
Guidance and counselling	4	20%
Child-centred teaching techniques	2	10%
COVID-19 SoPs	2	10%
Code of conduct	1	5%
Sexual reproductive health counselling	1	5%
Inclusive education	3	15%
Classroom management	2	10%
Pedagogical training	4	20%
Role of PTA and SMC	1	5%

Figure 28: Teacher's Trained on Letter Sounds

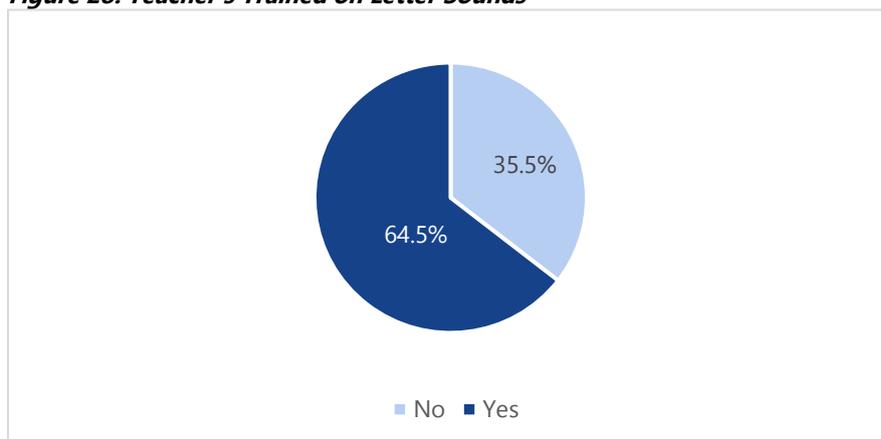


Table 28: Teacher Performance on Letter Sounds

Subtask	Mean	Std. Err.
Letter sounds (out of 5)	2.9	0.09

6.3. Teacher expectations and learner performance

Teacher Finding 6 – Learner Expectations: Teachers generally professed low expectations of learner performance in literacy and numeracy across the early grades. Over 80 per cent felt that children should not be able to fluently read a short passage until grade 2 or grade 3, while nearly 70 per cent believed children should not be able to comprehend what they read until grade 3 or grade 4. Over 50 per cent believed children should be able to count to 100 by grade 2 – the same grade they said learners should be able to add, subtract and multiply 1-digit numbers.

Teachers do not expect incoming grade 1 learners to demonstrate any emergent literacy or numeracy skills, an indication they do not expect learners to be school-ready when they begin primary. There is little change in their expectations for learners once they are in grade 1: the only skill most teachers think grade 1 learners should master is the ability to recite the alphabet. These low expectations are likely driven by a range of factors, including teacher’s personal experiences in the classroom with learners, gaps in their training and knowledge of the curriculum, and their beliefs about learning. Learners who enter primary school without a nursery education may struggle to gain foundational skills in the early years, leading to their poor performance and slow overall growth. To address this, teachers must be taught how children acquire literacy and numeracy and understand the key competencies they must progressively demonstrate in the early grades to become fluent readers. Refer to **annex 9.11** for additional regression analysis demonstrating a statistically significant relationship between a learner having attended nursery school and their overall performance in literacy and numeracy.

Table 29: Teacher Responses on Child Competency Expectations by Grade

Skill to be demonstrated	Before P1	P1	P2	P3	P4 or later
Pupils should first be able to read aloud a short passage with few mistakes in?	0.0%	4.5%	40.0%	41.1%	14.5%
Pupils should first be able to write their name in?	0.0%	11.1%	56.0%	28.4%	4.5%
Pupils should first be able to understand stories they read in?	0.0%	5.0%	27.1%	37.3%	30.7%
Pupils should first be able to recognise letters and say letter names in?	0.5%	28.2%	44.7%	18.7%	7.9%
Pupils should first be able to sound out all letters in the alphabet in?	0.2%	17.4%	41.1%	30.5%	10.8%
Pupils should first be able to sound unfamiliar words in?	0.0%	5.9%	22.8%	41.8%	29.6%
Pupils should first be able to understand stories they hear in?	0.0%	4.3%	27.5%	34.5%	33.6%
Pupils should first be able to recite the alphabet in?	1.1%	42.4%	39.5%	14.0%	2.9%
Pupils should first be able to count numbers up to 100 in?	0.7%	15.2%	55.4%	25.3%	3.5%
Pupils should first be able to add and subtract one-digit numbers in?	0.0%	16.5%	59.1%	21.0%	3.5%
Pupils should first be able to add and subtract two-digit numbers in?	0.0%	3.5%	53.9%	34.0%	8.7%
Pupils should first be able to multiply one-digit numbers in?	0.0%	6.3%	44.8%	39.8%	9.1%
Pupils should first be able to recognise and identify basic shapes in?	0.7%	8.9%	29.7%	37.2%	23.6%

6.4. Lesson observations

This section explores the instructional methods and classroom practices teachers use through lesson observations, including analysis of videos taken of them teaching. This is intended to understand which

key methods they employ, what resources they use, and how the classroom environment affects teaching and learning.

Teacher Finding 7 – Productive Lesson Time: Observation data, collected every 10 minutes of the 40-minute class, consistently shows that focus gets lost over time in the subject being taught, especially during the last 10 minutes of the class. This means that almost a quarter of the lesson time goes unused. Teachers must be supported to create lesson plans that reflect the needs of their learners and their attention span (e.g., the younger the learner the shorter the concentration time) and apply measures to make sure that all the time is used productively in a range of activities and tasks. If classes across all grades are not well synchronised, learners might get distracted towards the end of the lesson seeing that the class has ended for other grades. One possible measure would be to make sure all classes within the school start and end at the same time, perhaps keeping time with a bell.

Table 30: Minutes spent teaching

Activity	1 st 10 mins	2 nd 10 mins	3 rd 10 mins	4 th 10 mins
In class teaching	9.8	9.5	9.2	4.2

Teacher Finding 8 – Language of Instruction: Literacy classes are taught in English, which is the language of instruction in schools. It is often the second or third language of teachers and learners alike. The education sector lacks a clear language of instruction policy, as well as guidelines on mother tongue instruction. During the observations, English was spoken about 70 per cent of the time in both P2 and P4 classes, though its use tended to decrease in both grades as the class progressed. Krio was spoken for about 1 minute every 10 minutes and another minute was spent speaking other local languages. The use of different languages suggests that most of the lesson is conducted in English, however, in certain cases the teacher or the learners switch to a local language to explain something, utilising code-switching strategies to check for understanding and comprehension.

Table 31: Language spoken during the lesson

Language	1 st 10 mins	2 nd 10 mins	3 rd 10 mins	4 th 10 mins
Speaking English	7.1	6.9	6.7	4.5
Speaking Kiro	0.9	1	1	0.9
Speaking other local language	0.9	1	1.1	0.9

Teacher Finding 9 – Instructional Methods: Across both lesson types, instructional time is split between various, largely teacher-led, activities. Teachers spend nearly half of their time (about 45 per cent) lecturing, and the remaining time writing on the board and have learners repeat or recite new content. The least amount of time in the lesson is spent asking learners questions. These methods and time spent on them remains largely the same across the entire lesson, indicating that little progression is made in teacher-learner actions that helps learners progress from performing simple to more complex tasks.

Table 32: Time allocation during literacy observations

Activity	1 st 10 mins	2 nd 10 mins	3 rd 10 mins	4 th 10 mins
Repeating/recitation	2.7	2.8	2.9	2.2
Writing on board	2.9	2.6	2.6	2
Lecturing	4.7	4.5	4.4	3.3
Listening to learners	2.5	2.7	2.9	2.1
Asking questions	2.2	2.2	2.4	2

Table 33: Time allocation during numeracy observations

Activity	1 st 10 mins	2 nd 10 mins	3 rd 10 mins	4 th 10 mins
----------	-------------------------	-------------------------	-------------------------	-------------------------

Repeating/recitation	2.9	2.9	2.9	2.3
Writing on board	3.5	3	3.1	2.1
Lecturing	4.7	4.6	4.5	3.3
Listening to learners	2.5	2.9	2.9	2.2
Asking questions	2.3	2.6	2.5	1.8

Teacher Finding 10 – Positive Actions: Teachers overall show that they understand and implement positive practices and behaviours during their lessons related to classroom management. They freely engage their learners and encourage their participation during interactions. However, as observed previously, positive actions diminish in the last 10 minutes of the lesson. There is some room for improvement in teacher actions, especially when it comes to continuous assessment and lesson planning. Please refer to **figures 29 to 32** on **page** Error! Bookmark not defined. below for the visual representation of the positive actions performed by teachers in the classroom.

Teacher Finding 11 – Negative Actions: Teachers do not appear to have prepared the lesson in about 40 per cent of cases for grade 2 classes and about 30 per cent for grade 4 classes which might indicate that teachers underestimate the need to prepare for lower classes that are easier for them. For the majority of the time, teachers do not keep track of learners' performance; head teachers could tackle this by requesting periodic reports. Please refer to **figures 33 to 36** on **page 51** below for the visual representation of the negative actions performed by teachers in the classroom.

Teacher Finding 12 – Instructional Content for Literacy: During literacy classes, reading tasks for grade 2 focus on sounds, letters and words while for grade 4 more time is dedicated to sentences. Time spent covering stories is however very low, which is interesting considering learners' performance on the oral reading fluency subtask. A similar pattern of increased complexity of topics covered from grade 2 to 4 was observed during writing tasks. Please refer to **figures 37 to 40** on **page 52** below for the visual representation of the literacy instructional content used by teachers in the classroom.

Teacher Finding 13 – Instructional Content for Numeracy: During numeracy classes, tasks for grade 2 focus on single- and double-digit addition while for grade 4 more time is dedicated to subtraction, multiplication and fractions. Both grade 2 and 4 spend nearly equal amounts of instructional time counting, identifying number facts and reciting them orally. However, grade 4 learners spend more time practicing with missing number tasks and comparing sets. Please refer to **figures 41 to 44** on **page 53** below for the visual representation of the numeracy instructional content used by teachers in the classroom.

Figure 32: P2 Literacy - Teacher positive actions

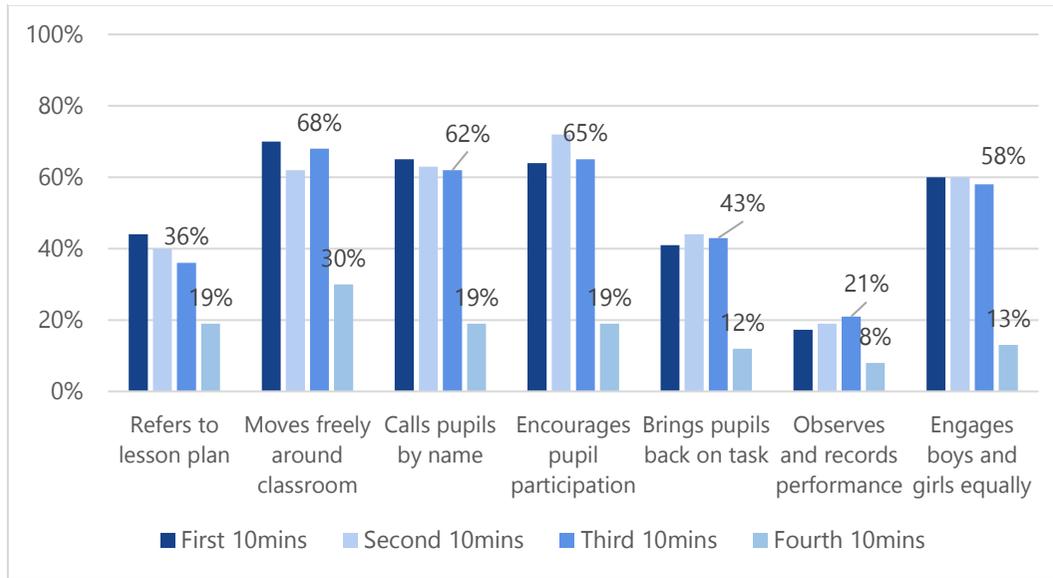


Figure 31: P4 Literacy - Teacher positive actions

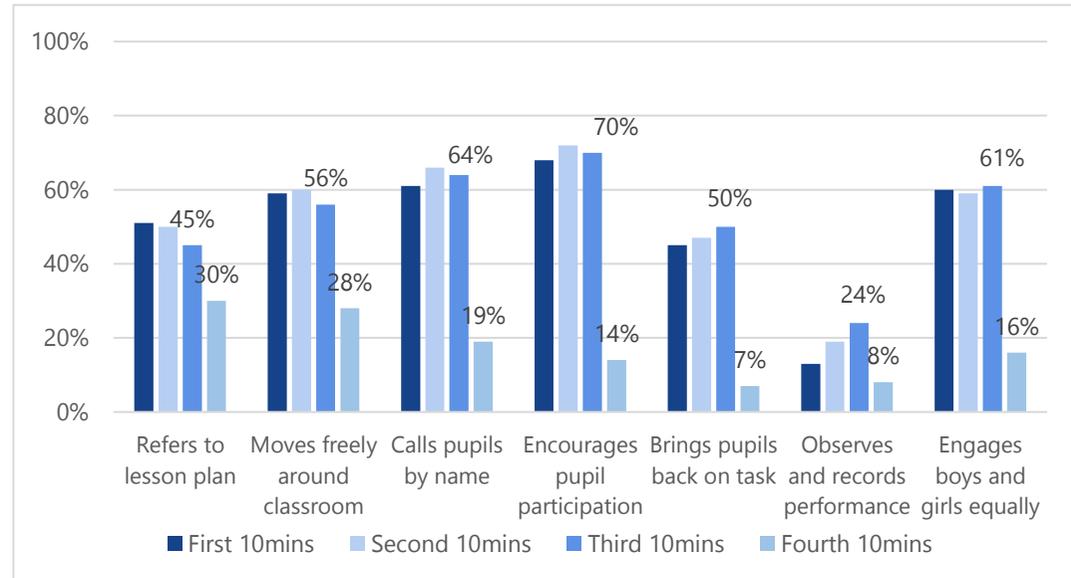


Figure 30: P2 Numeracy - Teacher positive actions

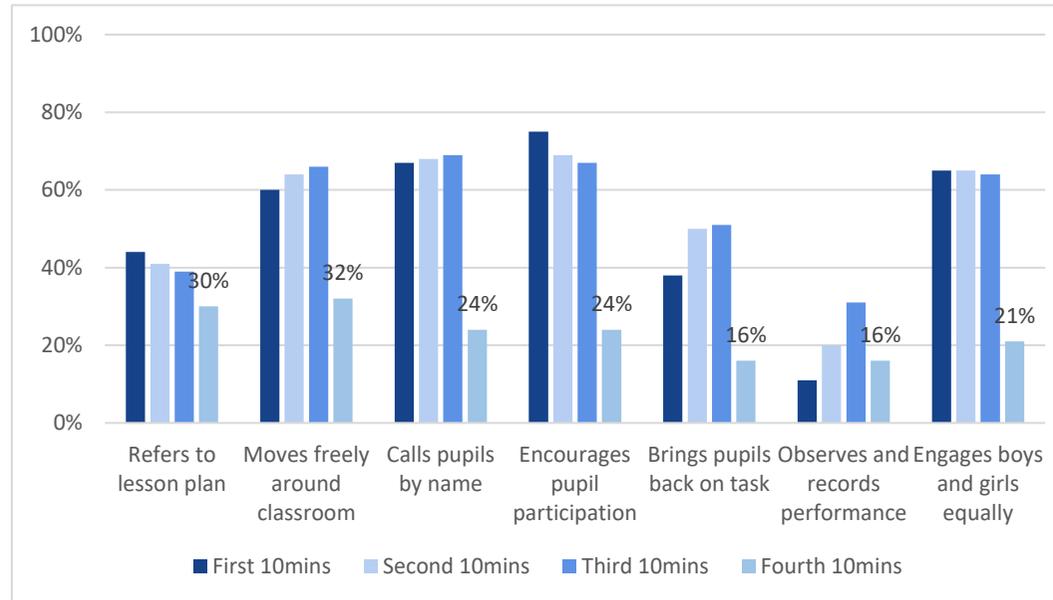


Figure 29: P4 Numeracy - Teacher positive actions

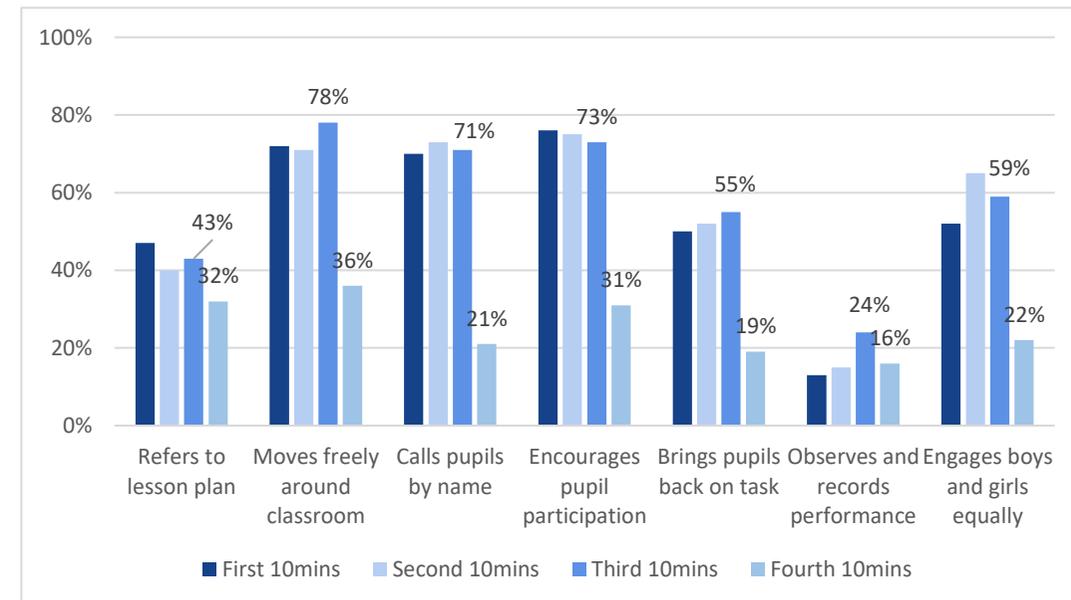


Figure 36: P2 Literacy - Teacher negative actions

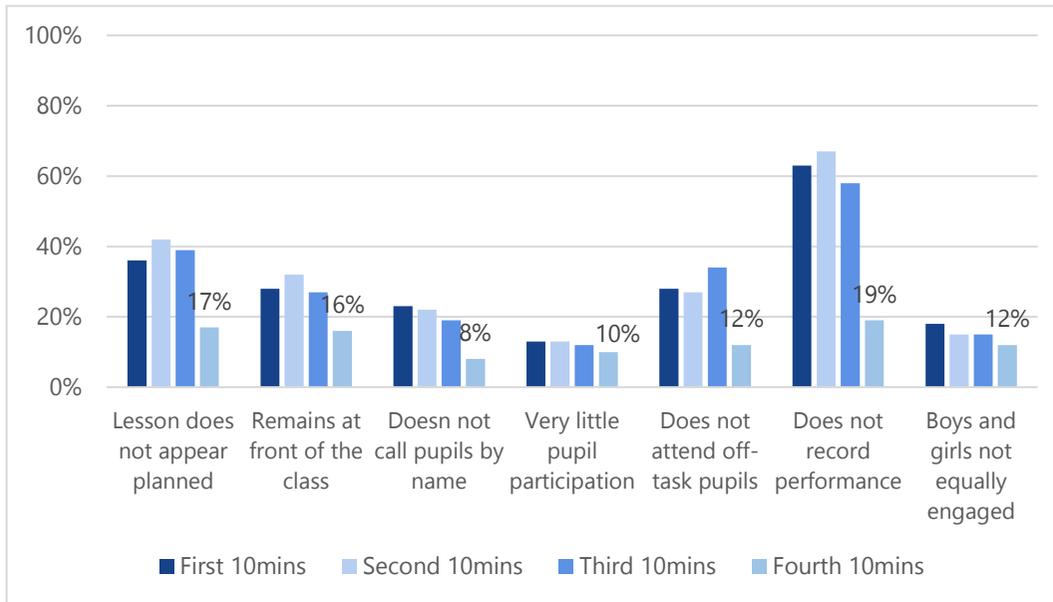


Figure 35: P4 Literacy - Teacher negative actions

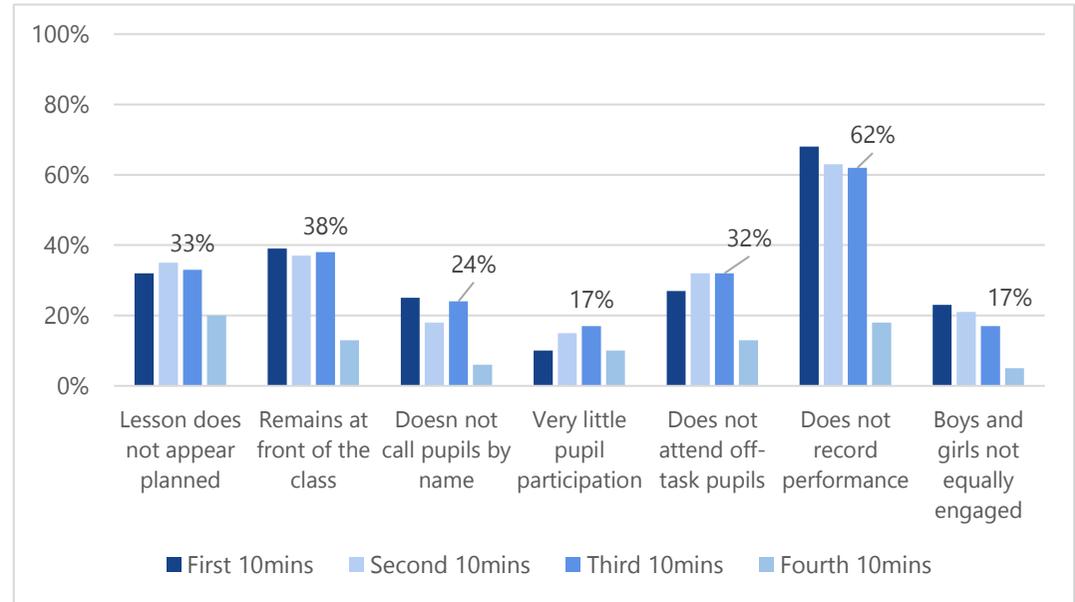


Figure 34: P2 Numeracy - Teacher negative actions

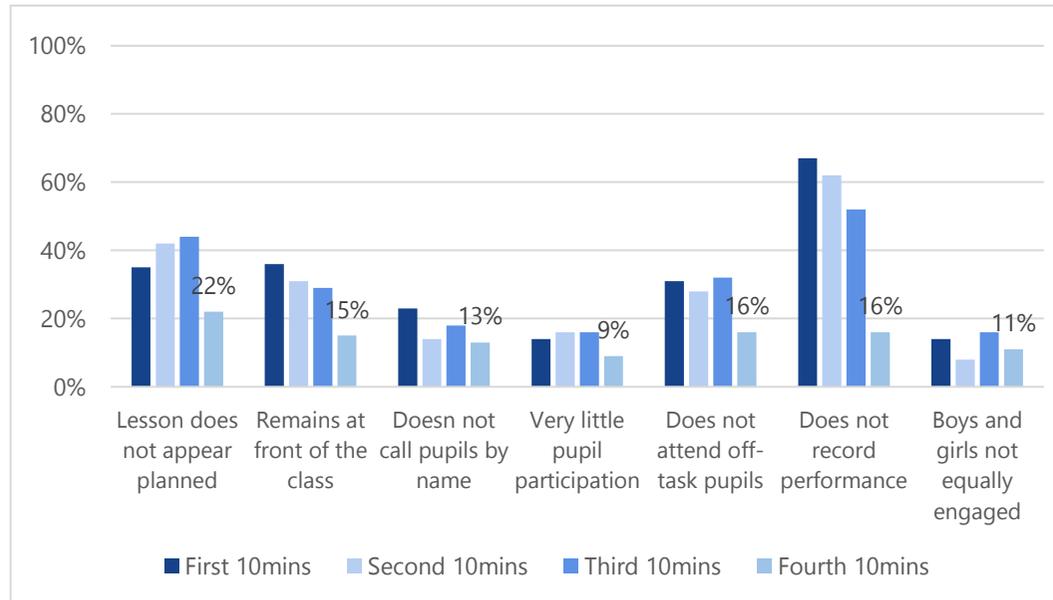


Figure 33: P4 Numeracy - Teacher negative actions

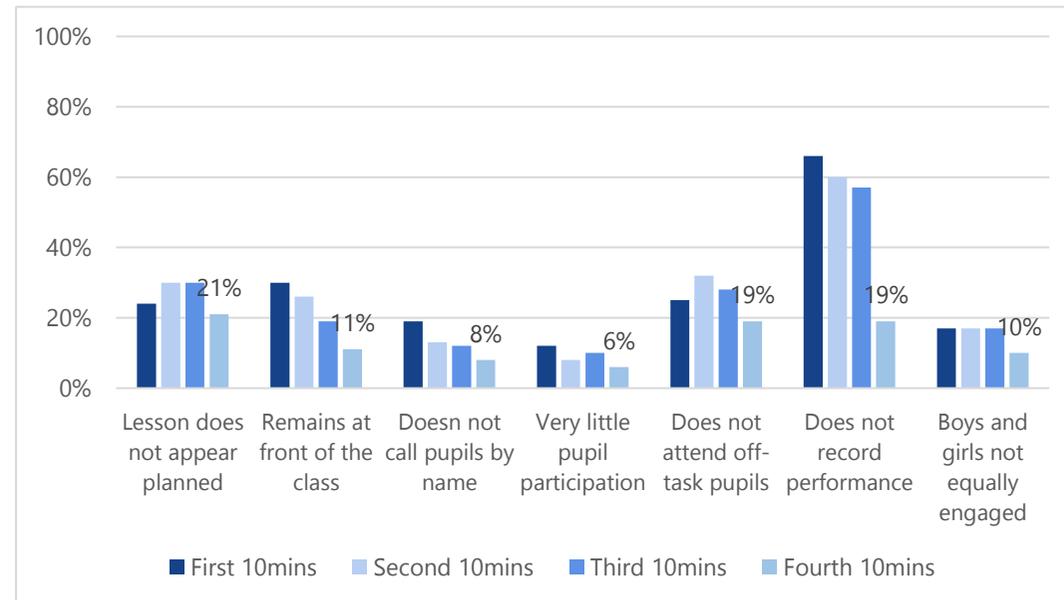


Figure 40: P2 Reading - Instructional content

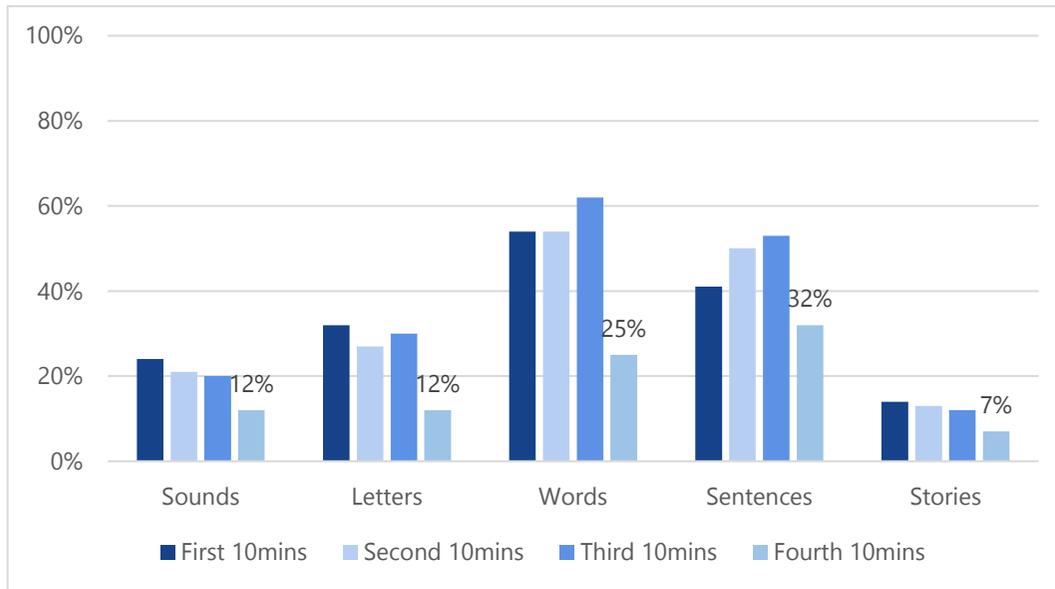


Figure 39: P4 Reading - Instructional content

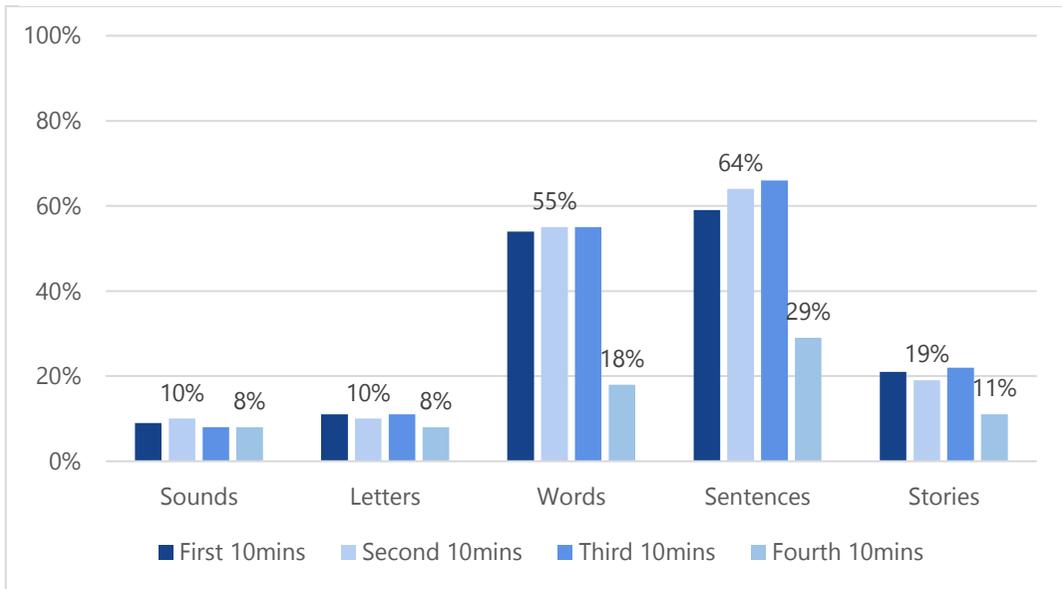


Figure 38: P2 Writing - Instructional content

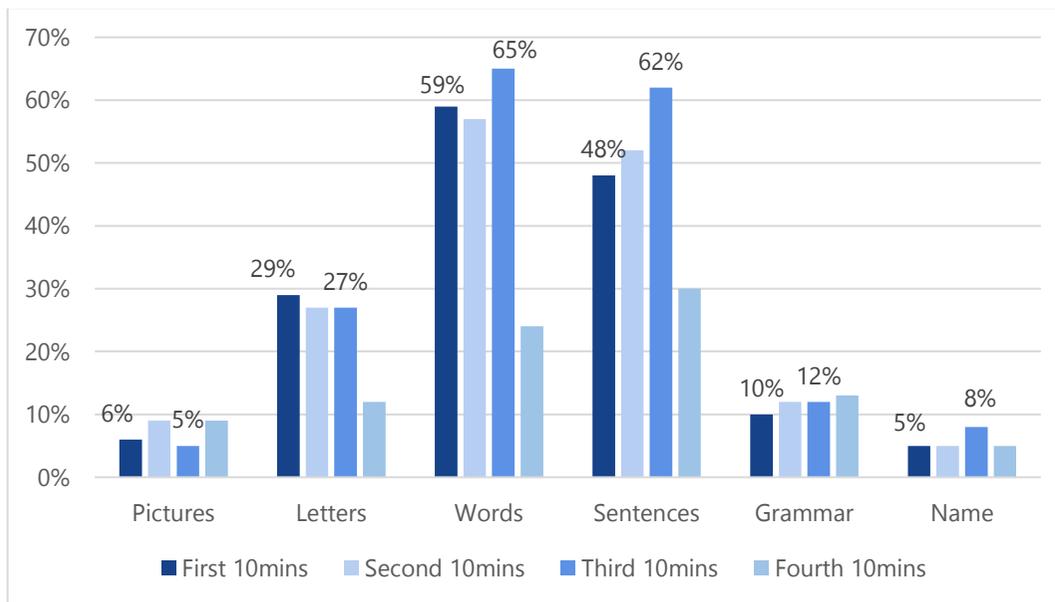


Figure 37: P4 Writing - Instructional content

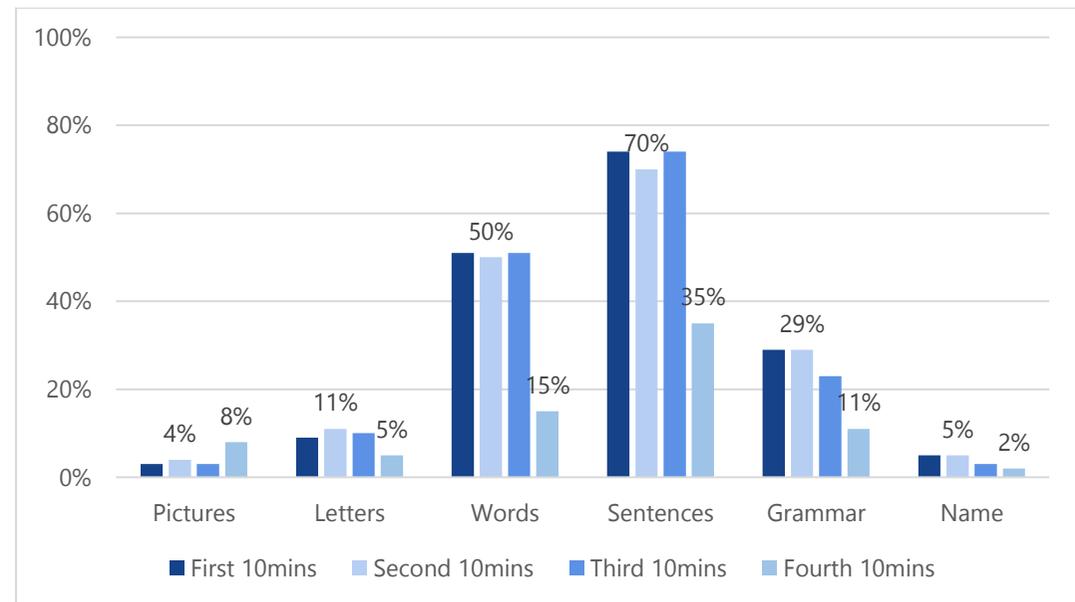


Figure 44: P2 Calculation and Measurement - Instructional content

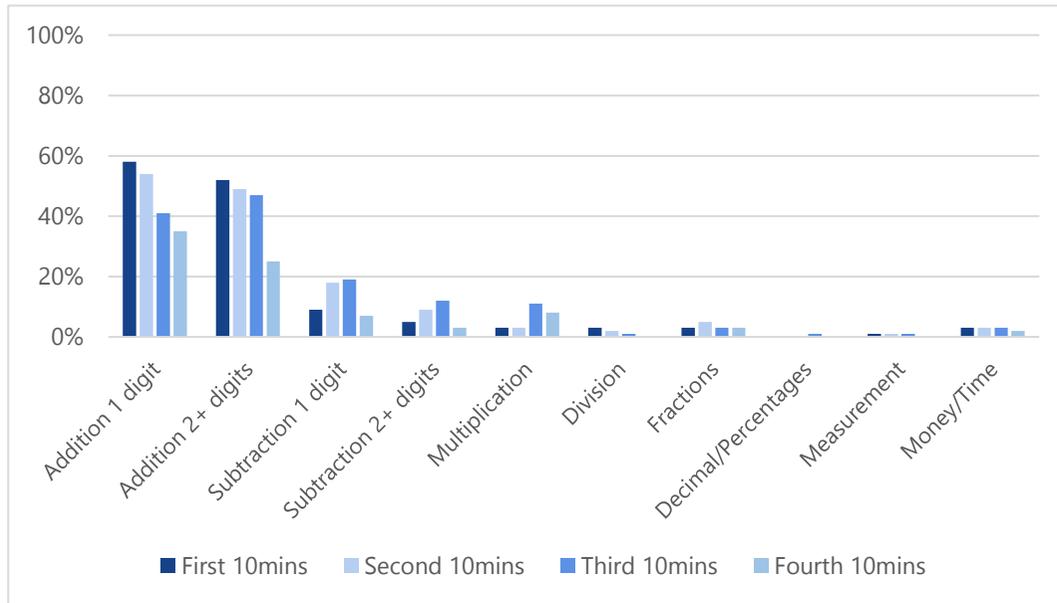


Figure 42: P4 Calculation and Measurement - Instructional content

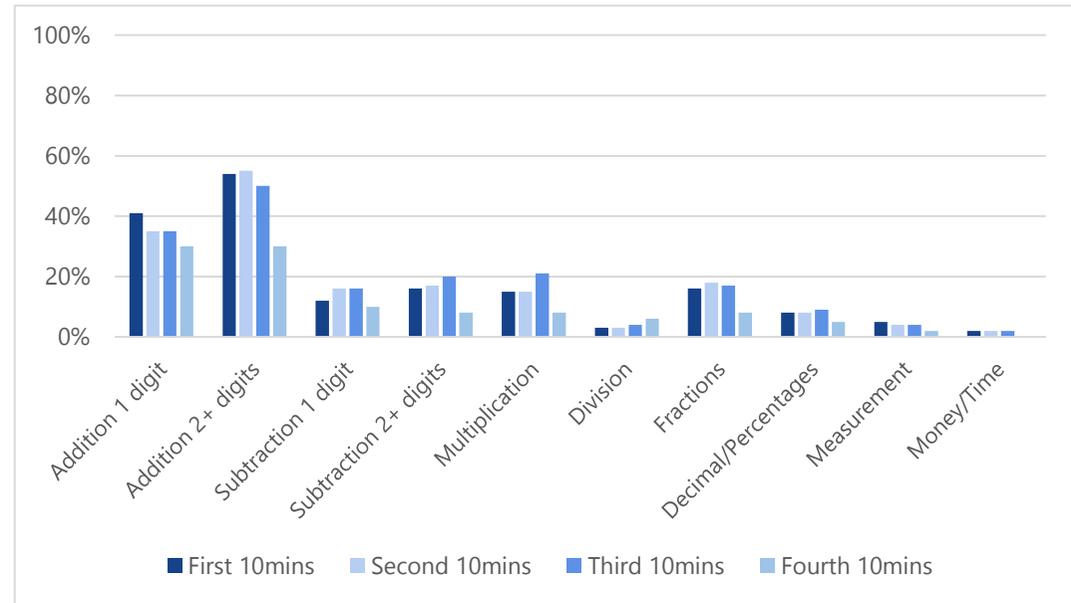


Figure 43: P2 Numbers - Instructional content

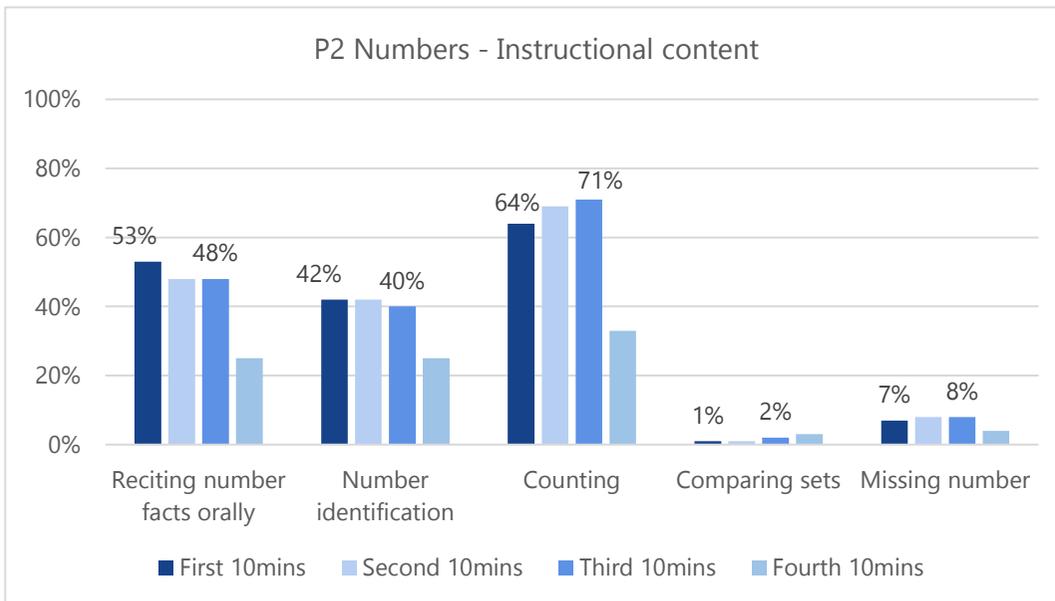
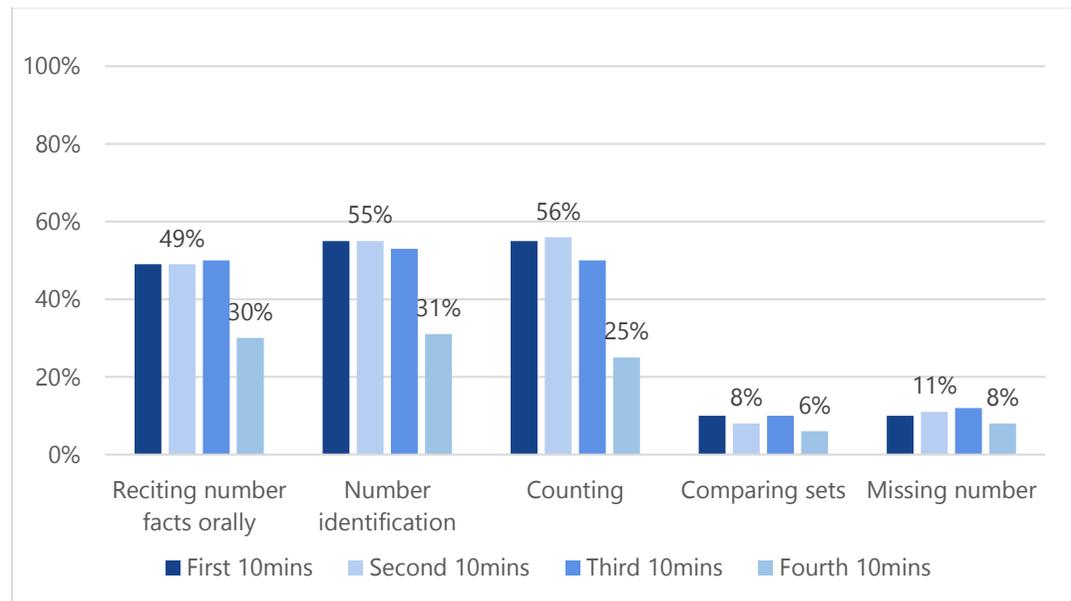


Figure 41: P4 Numbers - Instructional content



Teacher Finding 14 – Performance and Experience: Teaching experience does not have a significant effect on positive actions during class. Several reasons might contribute to this result. First, the sample size is not very large (about 128 observations). Second, it is not clear what effect should be expected: on one hand, with experience, teachers should know how to best manage learners; but on the other hand, with more years of service, a teacher might lose motivation or have gaps in their knowledge about current teaching practices. These two effects, going in opposite directions, can partly explain the low correlation between experience and performance.

Positive Actions	Has been teaching for more than 5 years	
	Coefficient	p-value
Positive actions during literacy	3.7	0.25
Positive actions during numeracy	2.5	0.48

Teacher Finding 15 – Performance and Certification: Having a certificate has a positive effect on teaching performance (measured by positive actions) during literacy classes, but not during numeracy classes. The magnitude of the positive effect is, however, not particularly large, as having a certificate increases positive actions by only 7 percentage points.

Positive actions	Certificate			Certificate OR more than 5 years of experience		
	No	Yes	Difference	No	Yes	Difference
Positive actions during literacy	39.8	47	7.2**	39.8	45.1	5.3
Positive actions during numeracy	45	49.5	4.5	44.9	48.9	5

TEACH Finding 1 – Time on Task: When using the TEACH tool to assess the videotaped classroom observations, teachers do not score particularly well on the time on task category: the average score was 2.6 out of 5, indicating that there are at least between 2 and 5 children who are off task at any given point in the lesson time interval snapshots.

Snapshot of time on task	Score (out of 5)
Time on Learning - 1st Snapshot	2.61
Time on Learning - 2nd Snapshot	2.70
Time on Learning - 3rd Snapshot	2.60

TEACH Finding 2 – Quality of Teaching Practices: Teachers scored on average 3.4 in the “supporting learning environment” category, which indicates that they do use positive language, call learners by name and say thanks when needed but they do so irregularly; similarly, a score of 3.2 in “positive behavioural expectations” indicates that teachers do not fully manage to control misbehaviour in class and when they use phrases like “keep quiet” or “sit down” they are not fully effective. A score of 3 in lesson facilitation means that the lesson objective could be inferred in some cases and the explanation of content was clear but, in most cases, did not link to other content.

Teacher scored poorly in all other categories which were graded between 1 and 2.3 out of 5. A low score in socio-emotional skills is associated to teachers not assigning work to do in pairs and a low score on feedback shows that teachers only let learners know if their answer is correct or incorrect without explanation. Teachers were not asking questions or asking few learners, not giving individual or group work and not adjusting teaching for learners in most cases (scoring only 2.3 in “checks for understanding”) and most teachers only asked closed ended questions, without giving the learners any thinking tasks, they spoke and learners listened (scoring 2 in “critical thinking”).

Quality of teaching practice	Score (out of 5)
Supportive learning environment	3.41
Positive behavioural expectations	3.17
Lesson Facilitation	3.02
Checks for understanding	2.29
Feedback	1.39
Critical Thinking	2.00
Autonomy	2.33
Perseverance	2.00
Socioemotional skills	1.10

7. FINDINGS ON SCHOOLS

The following section details high level findings on the schools that participated in the study. Data is presented on the presence and availability of key teaching and learning materials, as well as the main facilities at the school and their condition. This information is important to inform findings and recommendations on strengthening school and classroom infrastructure and resources, and to identify school-level inputs that will help teachers be more effective and improve their instructional methods.

7.1. Learner enrolment and attendance

Additional data on learner enrolment and attendance was captured for each school. Understanding the distribution of learners within and across local geographies and sites is important to inform recommendations to improve teacher workforce management, including the supply (hiring), posting and transfer of teachers based on evidence and need of the demand. Findings can also inform future acquisition and allocation of teaching and learning resources to schools, as well as other hardware inputs.

School Finding 1 - Learner Enrolment: Enrolment diminishes over time by grade: grade 6 has, on average, half of the total number of learners enrolled as grade 1 (for both boys and girls). This could be due to a high drop-out rate from lower to upper primary, and to learners repeating grades throughout the primary cycle – especially in grade 1. High rates of repetition in primary schools is also evidenced by the high level of mismatch between learner’s ages and the grade they are in, as there are a significant number of overage children in all grades in every school. It is well documented that attending classes at the wrong age level can have a demotivating effect on children, who often decide eventually to drop-out.

School Finding 2 - Learner Enrolment: We observe a large discrepancy between learners enrolled and learners present at school according to data from school rosters; boys across all grades have an average attendance rate of 53 per cent and girls 54 per cent. This primarily reflects the need to get children to attend school regularly but can also be a signal that class registers need to be updated and continuously monitored by school management and other stakeholders, as they might contain children that are effectively out of school.

Table 34: Boys’ and girls’ enrolment and attendance from school rosters

Grade	Boys present (average per class)	Boys enrolled (average per class)	Girls present (average per class)	Girls enrolled (average per class)
P1	19	34	21	36

P2	14	27	15	29
P3	14	25	16	27
P4	13	23	13	25
P5	12	21	13	22
P6	7	17	8	18

7.2. Teacher assignment and attendance

School Finding 3 – Teacher Assignment: Schools have on average slightly more than one teacher assigned per grade according to the school registers examined during the baseline. While this may seem adequate, data shows their qualifications do not always match the grade level and subject they are teaching. In other schools, there are multiple streams for a given grade, requiring more teachers per class; not all schools have them. Some of these positions are filled by local recruitment strategies managed by schools, who sometimes hire uncertified teachers available locally. Recruitment and assignment of teachers to posts should reflect the needs and size of the school, and staff allocations should be structured accordingly using government-approved teachers who are on the payroll. These findings require further examination as part of a broader national teacher workforce management strategy.

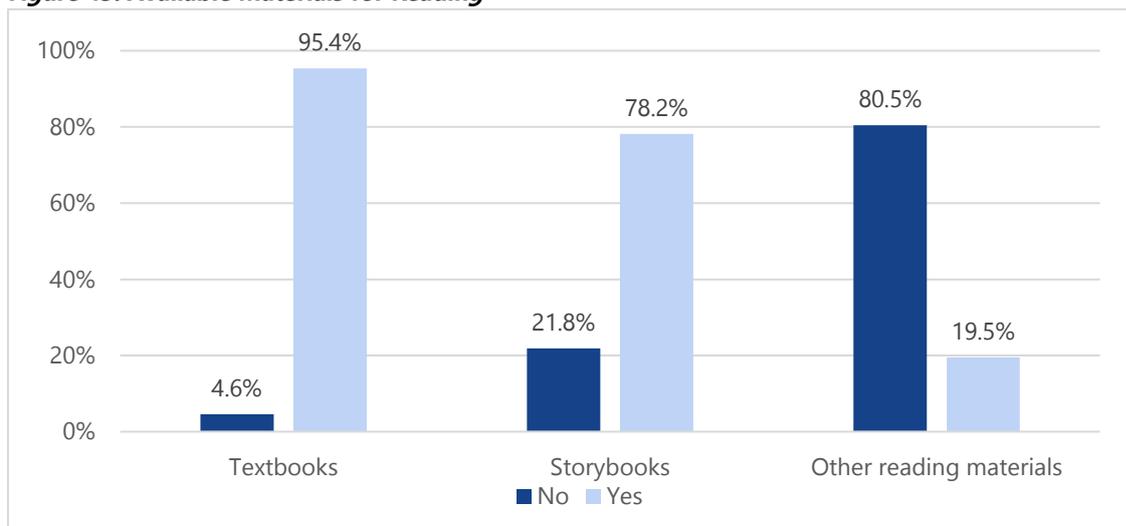
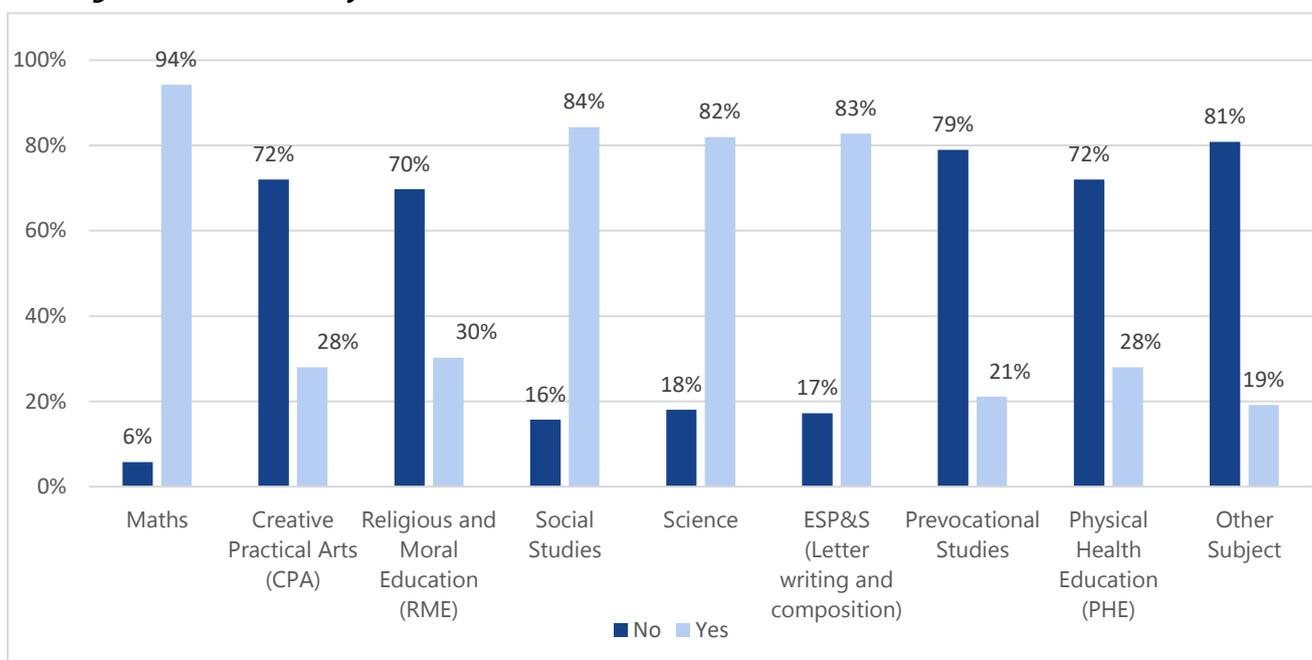
School Finding 4 – Teacher Attendance: The attendance rate is high for teachers at 89.9 per cent on average across all schools. Teachers worked with or were assisted by a co-teacher in 15.7 per cent of the numeracy classes observed and 14.1 per cent of literacy classes. These are positive findings indicative of a committed workforce who cares about showing up. This is corroborated in lesson observation findings, which showed that teachers stay in class teaching their lesson the great majority of the time.

Table 35: Teacher allocation by grade

Grade	Teachers assigned (average per class)
P1	1.3
P2	1.3
P3	1.2
P4	1.2
P5	1.1
P6	1.2

7.3. Teaching and learning materials

School Finding 5 – Teaching and Learning Materials: Most schools had basic materials available for teaching reading, including textbooks and storybooks; however, access to additional reading materials and learner primers was limited. In general, schools appear to have very few supplementary resources such as teacher’s guides, audio-visual materials and lesson planning documents. Schools generally lacked textbooks for four out of eight subjects, though they are mostly non-academic. About 95 per cent of schools had maths textbooks, though only 84 per cent had ESP&S (letter writing and comprehension) texts. Schools must have a range of core and supplementary materials to support learning. Their acquisition must be prioritised alongside other system inputs to improve teacher effectiveness.

Figure 45: Available Materials for Reading**Figure 46: Textbook Subject****Table 36: Other Available Teaching and Learning Materials**

Other Available Teaching and Learning Materials	Frequency	Percentage
Audio visual materials	1	2%
Duster	8	16%
Flash cards	3	6%
Teaching aids including	3	6%
Lesson note	4	8%
Lesson Plan Manual	10	20%
Pencil, pens, eraser, markers, sharpeners, crayons, etc.	10	20%
Pointers	2	4%
Mathematical set	5	10%
Prescribed books by the government, English book three and four	1	2%
Teaching guide	1	2%

Other Available Teaching and Learning Materials	Frequency	Percentage
Class register	1	2%
Scheme of work	2	4%

School Finding 6 - Supporting Materials During Different Tasks: The chalkboard is the most commonly used support by teachers for both literacy and numeracy classes. The second and third supporting materials for reading classes are textbooks and reading books; but, they are not widely used with 20 per cent and 10 per cent of classes, respectively, regularly using them. Paper is, on the other hand, commonly used (in over 60 per cent of cases) as a support for writing tasks. Numeracy classes are supported by textbooks and counters, but this is not a widespread practice as that happened in no more than 20 per cent of observations.

Figure 47: Supporting materials during reading

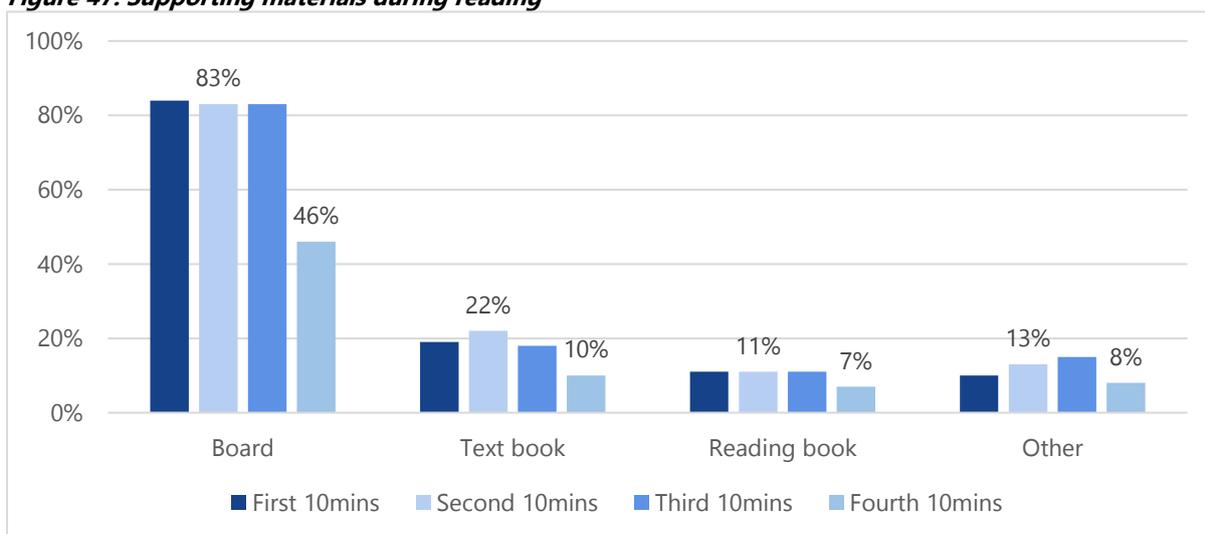


Figure 48: Supporting materials during writing

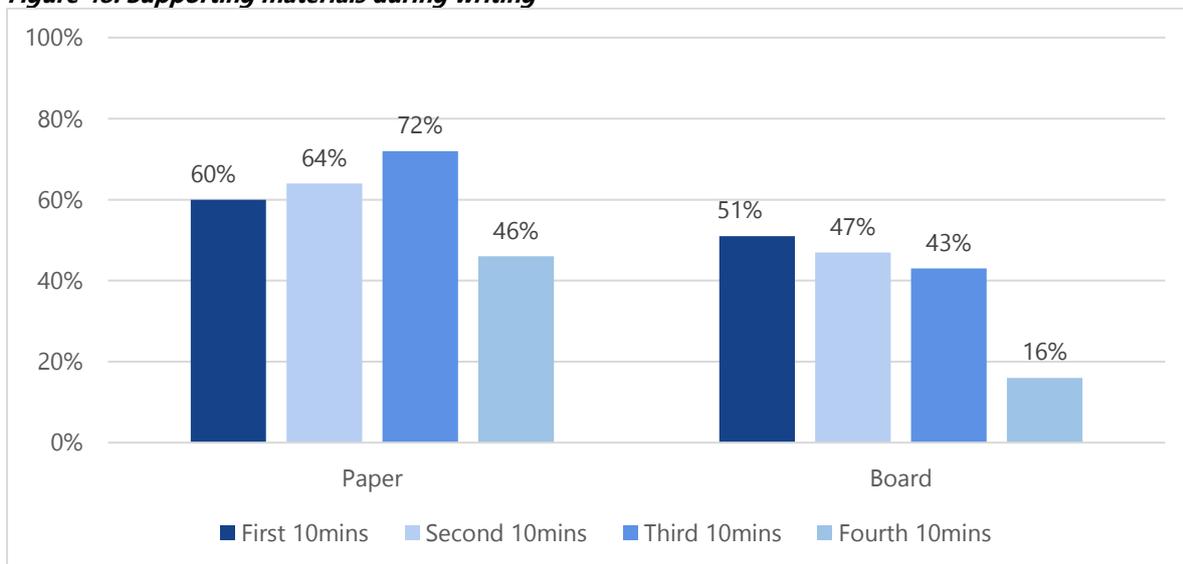
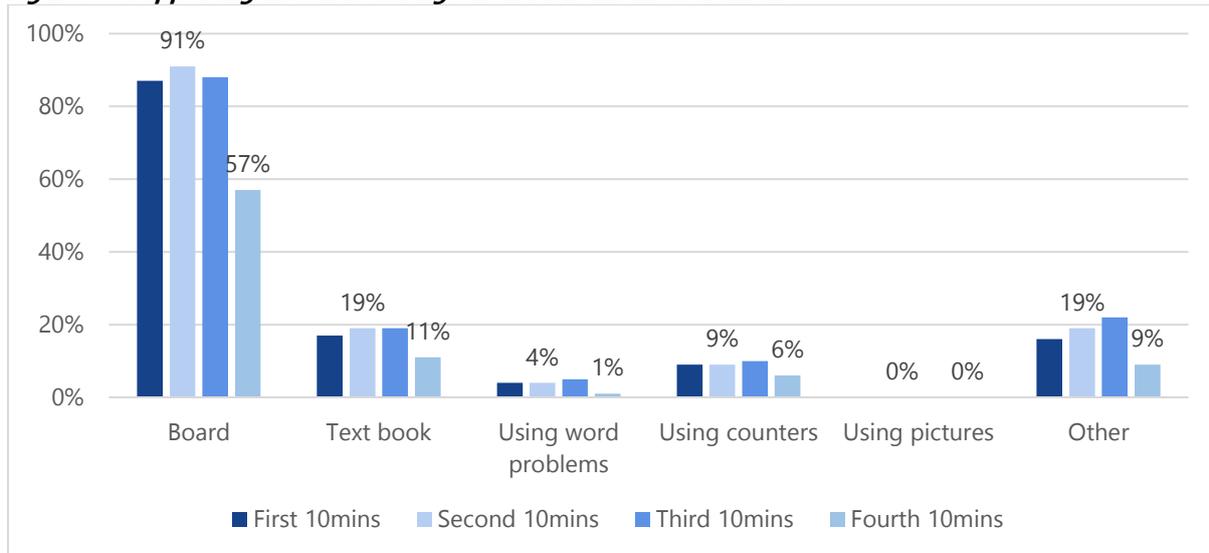
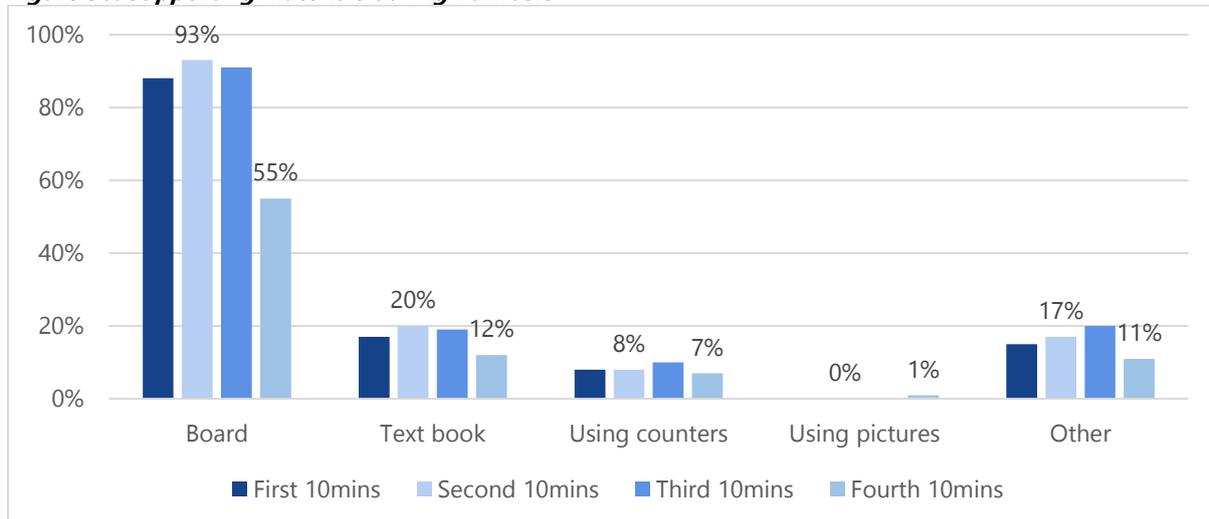


Figure 49: Supporting materials during calculation and measurement**Figure 50: Supporting materials during numbers**

School Finding 7 – Other Materials: Only 3.5 per cent of literacy classrooms and 0.9 per cent of numeracy classrooms had a clock. This can make it hard for teachers to keep track of time during the lesson and to distribute tasks appropriately across the 40-minutes of instructional time. Most classes do not have materials on the walls that can help learners in different subjects, especially with tips and reminders about key reading and maths strategies or tips. It would be relatively simple and cheap for school staff to create and display such stimuli to improve the learning environment and classroom conditions across a range of schools.

Table 37: Instructional material on the classroom walls to support learning

	Literacy			Numeracy		
	No	Somewhat	Yes	No	Somewhat	Yes
Does the classroom have materials hanging on the walls and made available throughout the classroom to help with learner instruction in the subject area?	56.4	5.3	38.3	73.5	4	22.4

	Literacy			Numeracy		
	No	Somewhat	Yes	No	Somewhat	Yes
Does the classroom have materials hanging on the walls and made available throughout the classroom to help with learner instruction in other subjects ?	59	32.6	8.4	56.5	8.3	41.7

Table 38: Other findings that affect learners' ability to learn how to read

Other findings that affect learners' ability to learn how to read	No	Yes
School has a room(s) where textbooks and storybooks are kept?	25.3%	74.7%
Do children use the books/reading materials during class?	11.1%	88.9%
Do children take the books home?	63.6%	36.4%
Does the school provide meals to learners?	67.4%	32.6%
Are the meals provided through a feeding programme?	3.5%	96.5%
Do parents pay or contribute to the school meals?	14.1%	85.9%
Does the school provide meals to teachers?	81.2%	18.8%
Does the school have a cooking area?	8.2%	91.8%

7.4. Water, hygiene and sanitation facilities

School Finding 8 – WASH Facilities: Schools have a range of WASH facilities available, though very few schools have all the WASH facilities they should have – and need – as per national guidelines. Over half of the facilities inspected were in good working order; the remainder of the facilities had minor issues but were in use. Learners and teachers only have access to safe drinking water in around 50 per cent of the schools.

In 22 per cent of schools there are no separate latrines for boys and girls; in 40 per cent of schools there are no latrines for teachers. In schools that do have latrines, around 10 per cent of them were reportedly hazardous and unsafe, though they were still in use. The huge majority of schools lack adequate facilities for menstrual hygiene management for girls – a large gap in infrastructure that must be addressed to make schools safe and accessible for all learners.

Table 39: Inquiry on WASH Facilities within the school

Inquiry on WASH Facilities within the school	No	Yes
Are there adequate facilities for menstrual hygiene management for girls?	87.7%	12.3%
Are there menstrual hygiene disposal facilities for girls?	90.0%	10.0%
Is there soap to wash hands for menstrual hygiene management for girls?	87.4%	12.6%
Are the menstrual hygiene management facilities for girls accessible?	88.9%	11.1%

Table 40: WASH Facilities and Condition

WASH Facility Type	Found		Working condition				Hygiene condition			Accessibility		
	No	Yes	Good, in use	Minor issues, in use	Hazardous, unsafe, in use	Not in use	Very clean	Minor issues	Inadequate	Fully accessible	Partially accessible	Not accessible
Water tank	78.9%	21.1%	41.8%	21.8%	0.0%	36.4%	38.2%	32.7%	29.1%	58.2%	16.4%	25.5%
Borehole	54.8%	45.2%	40.7%	27.1%	3.4%	28.8%	44.9%	26.3%	28.8%	71.2%	9.3%	19.5%
Well	83.1%	16.9%	43.2%	38.6%	6.8%	11.4%	36.4%	52.3%	11.4%	86.4%	9.1%	4.6%
Safe drinking water, learners	49.0%	51.0%	70.7%	29.3%	0.0%	0.0%	70.7%	28.6%	0.8%	88.0%	12.0%	0.0%
Safe drinking water, teachers	51.7%	48.3%	71.4%	27.8%	0.0%	0.8%	69.8%	29.4%	0.8%	84.9%	14.3%	0.8%
Latrines, boys	22.2%	77.8%	37.9%	48.8%	11.8%	1.5%	25.6%	62.6%	11.8%	77.8%	19.2%	3.0%
Latrines, girls	21.8%	78.2%	38.2%	49.5%	10.3%	2.0%	28.4%	60.8%	10.8%	76.0%	21.6%	2.5%
Latrines, teachers	38.3%	61.7%	54.7%	39.1%	3.7%	2.5%	48.5%	45.3%	6.2%	83.2%	14.9%	1.9%
Handwashing, learners	36.4%	63.6%	53.0%	41.0%	0.6%	5.4%	52.4%	42.8%	4.8%	84.9%	10.8%	4.2%
Hand-washing teachers	44.4%	55.6%	59.3%	37.2%	1.4%	2.1%	59.3%	40.0%	0.7%	89.0%	9.0%	2.1%

8. CONCLUSIONS AND RECOMMENDATIONS

The following section summarises key conclusions and corresponding recommendations from the baseline survey's findings.

8.1. Learners

Conclusion 1 – Literacy: Learner profiles for oral reading fluency indicate that learners are progressing as readers from grade 2 to grade 4, with the share of non-readers in grade 4 reducing significantly from that in grade 2. Significantly more grade 4 learners are at an emerging or progressing level in reading fluency than those in grade 2, indicating an upward, positive trend in their literacy development.

Comparison of zero scores between 2014 and 2021 national assessment data indicates a substantial, positive reduction in the proportion of non-readers in both grades 2 and 4 and across all subtasks from 2014 to 2021, with a clear decrease in the number of learners unable to read a single word or letter. Emergent literacy skills also develop in a similar way, with more learners demonstrating basic competencies in lower-level reading skills from 2014 to 2021 and within grades each time.

However, trends in reading comprehension do not follow suit, with limited progress between years and grades. The majority of learners in 2021 in grades 2 and 4 are not able to comprehend the text they read (64-73 per cent), while only 25-30 per cent are at an emerging skill level, performing below expected proficiency. These findings suggest that, while learners are increasingly able to identify words and read text fluently, their understanding of those words and content is limited, as is their conceptual ability to apply skills they have learned to other tasks.

Recommendation 1 – Literacy: Getting more learners reading more of the time, reducing non-readers and growing the proportion of capable readers who can comprehend text – both within and between grades – must be a clear goal of any national literacy programme. Such a programme should be a priority for the government. Using learning profiles to monitor and track this achievement in individuals and groups of learners provides a clear mechanism for diagnosing progress, communicating results, and getting teachers and parents on board with discussing and monitoring change. Moreover, teachers need explicit instruction, regular support and ongoing professional development to teach well and effectively; this should be provided in earnest at both the pre- and in-service levels. Coordinated efforts must be made to train and certify them in the key skills they need to teach the children of today.

Conclusion 2 – Numeracy: Results demonstrate clear progression between grades in every EGMA subtask except the level two subtasks for addition, subtraction and word problems. The biggest gains are made between grade 2 and 4 learners in the number identification subtask. More moderate gains are made in all the other level one subtasks, where small mean gains are made between grades based on only 2-3 more items being attempted and answered correctly. Importantly, the attempted items and the accuracy of responses increases significantly between grades, with grade 4 learners answering between 20-30 per cent more of their attempted items correct than grade 2 learners.

Recommendation 2 – Numeracy: As learners' progress, it is critical for them to develop and demonstrate more advanced – though still foundational – skills in conceptual mathematics, which is the basis for understanding all other higher order skills. Sharp differences in performance across the subtasks demonstrates that learners have notably higher skills in basic, procedural maths tasks, versus significantly lower skills in conceptual ones. These results suggest that learners' experience of mathematics instruction is more about memorisation of facts and rules than development of strategies to find answers to problems. Conceptual maths must, therefore, be explicitly taught as mastering these foundational concepts is directly predictive of future performance.

8.2. Teachers

Conclusion 3 – Teachers: Nearly 25 per cent of the teachers surveyed have only a basic education or West African senior school certificate, meaning they have not received certified training beyond primary or secondary school to become a teacher, though some have participated in non-certificate in-service training programmes. Another 12 per cent of the teachers surveyed have either no certification whatsoever or have identified a non-recognised certification that they hold outside of teaching.

Positively, about 50 per cent of teachers surveyed have a teacher’s certificate; while others have higher level degrees, that do not necessarily qualify them for primary teaching. Teachers with a certificate demonstrated more positive actions when teaching literacy than those without, though there was no similar impact in numeracy instruction. Years of experience did not correlate with teacher’s positive actions in the classroom in either literacy or numeracy lessons, indicating that the length of time a teacher spends working in the profession may have a limited impact on their overall effectiveness.

Recommendation 3 – Teachers: These findings can help to inform reflections on the country’s teacher workforce management strategy, especially for the recruitment and placement of new hires and for the possible reallocation and distribution of certified teachers already in the system. Moreover, these findings indicate that the upgrading and certifying of in-service teachers through targeted training will be critical to ensure there continues to be a consistent, immediate supply of skilled teachers in the system to meet current demand.

Conclusion 4 – Teachers: Teachers generally professed low expectations of learners’ performance in literacy and numeracy across the early grades, as well as limited development of basic skills and competencies over the same time. Over 80 per cent felt that children should not be able to fluently read a short passage until grade 2 or grade 3, while nearly 70 per cent believed children should not be able to comprehend what they read until grade 3 or grade 4. Over 50 per cent believed children should be able to count to 100 by grade 2 – the same grade they said learners should be able to add, subtract and multiply 1-digit numbers.

Recommendation 4 – Teachers: These low expectations are likely driven by teacher’s personal experiences in the classroom, as well as gaps in their pre-service training. Additionally, most learners who enter primary school without a nursery education may struggle to gain foundational skills in the early years, leading to their poor performance and slow overall growth. To counteract this, teachers must learn how children acquire literacy and numeracy and the key competencies they must progressively demonstrate developmentally in the early grades to become fluent readers. They must also understand how to align their instruction to the level of their learners, close these gaps and monitor performance and improvement.

8.3. Schools

Conclusion 5 – Schools: Enrolment diminishes over time by grade; grade 6 was found to have, on average, half of the total number of learners enrolled as in grade 1 (for both boys and girls). This is likely due to a high drop-out rate from lower to upper primary, and to learners repeating grades throughout the primary cycle; steep increases in enrolment in recent years may also be a major factor. High rates of repetition in primary schools is also evidenced by the high level of mismatch between learner’s ages and the grade they are in, as there are a significant number of overage children in all grades in every school. When learners did go to school, they had an average attendance rate of 53 per cent for boys and 54 per cent for girls.

Recommendation 5 – Schools: It is well documented that attending classes at the wrong age level can have a demotivating effect on children, who often decide eventually to drop-out. Likewise, high

repetition rates have the same negative effect. A more detailed cohort analysis of learners in the system at various grade levels will provide a more accurate picture on which to base decisions and inform programming. An additional way to counteract this challenge lies with investing in alternative education programmes for overage youth who move in and out of the formal basic education. Flexible pathways in and out of the formal system that allow multiple entry and exit points and pathways to acquiring recognised certifications for non-formal or technical education must be made available and put on offer to children who fall out of the formal education system but still need skills to work and earn.

Conclusion 6 – Schools: Schools have on average slightly more than one teacher assigned per grade. While this may seem adequate, data shows their qualifications do not always match the grade level and subject they are teaching. In other schools, there are multiple streams for a given grade, requiring more teachers per class which not all schools have available to them.

Recommendation 6 – Schools: Recruitment and assignment of teachers to posts should reflect the needs and size of the school, and staff allocations should be structured accordingly. These findings require further examination as part of a broader national teacher workforce management strategy.

9. ANNEXES

9.1. Population and sample distribution per district

Table 41: Population and sample distribution per district

District	School Owner	Population		Sample	
		Freq	Percent	Freq	Percent
Bo	Community	42	1%	2	1%
Bo	Government	58	1%	3	1%
Bo	Mission/religious group	345	8%	17	7%
Bo	Private	16	0%	2	1%
Bombali	Community	27	1%	2	1%
Bombali	Government	35	1%	2	1%
Bombali	Mission/religious group	153	3%	7	3%
Bombali	Private	8	0%	2	1%
Bonthe	Community	11	0%	2	1%
Bonthe	Government	41	1%	2	1%
Bonthe	Mission/religious group	99	2%	5	2%
Bonthe	Private	1	0%	0	0%
Falaba	Community	13	0%	2	1%
Falaba	Government	18	0%	2	1%
Falaba	Mission/religious group	89	2%	5	2%
Kailahun	Community	17	0%	2	1%
Kailahun	Government	49	1%	3	1%
Kailahun	Mission/religious group	210	5%	11	4%
Kailahun	Private	1	0%	0	0%
Kambia	Community	24	1%	2	1%
Kambia	Government	74	2%	4	2%
Kambia	Mission/religious group	185	4%	11	4%
Kambia	Private	2	0%	0	0%
Karene	Community	44	1%	2	1%
Karene	Government	65	1%	3	1%
Karene	Mission/religious group	116	3%	6	2%
Kenema	Community	28	1%	2	1%
Kenema	Government	52	1%	3	1%
Kenema	Mission/religious group	349	8%	18	7%
Kenema	Private	12	0%	2	1%
Koinadugu	Community	25	1%	2	1%
Koinadugu	Government	33	1%	3	1%
Koinadugu	Mission/religious group	91	2%	6	2%
Koinadugu	Private	2	0%	0	0%
Kono	Community	90	2%	5	2%
Kono	Government	64	1%	3	1%
Kono	Mission/religious group	238	5%	13	5%
Kono	Private	5	0%	1	0%

District	School Owner	Population		Sample	
Moyamba	Community	23	1%	2	1%
Moyamba	Government	41	1%	2	1%
Moyamba	Mission/religious group	260	6%	15	6%
Moyamba	Private	3	0%	0	0%
Port Loko	Community	36	1%	2	1%
Port Loko	Government	109	2%	6	2%
Port Loko	Mission/religious group	258	6%	14	5%
Port Loko	Other	3	0%	1	0%
Port Loko	Private	8	0%	4	2%
Pujehun	Community	5	0%	0	0%
Pujehun	Government	34	1%	2	1%
Pujehun	Mission/religious group	177	4%	9	3%
Tonkolili	Community	27	1%	2	1%
Tonkolili	Government	73	2%	4	2%
Tonkolili	Mission/religious group	340	7%	16	6%
Tonkolili	Private	1	0%	0	0%
Western Area Rural	Community	36	1%	2	1%
Western Area Rural	Government	47	1%	4	2%
Western Area Rural	Mission/religious group	99	2%	4	2%
Western Area Rural	Private	43	1%	3	1%
Western Area Urban	Community	32	1%	1	0%
Western Area Urban	Government	44	1%	2	1%
Western Area Urban	Mission/religious group	100	2%	6	2%
Western Area Urban	Other	1	0%	0	0%
Western Area Urban	Private	65	1%	2	1%

9.2. Detailed study methodology

9.2.1. Quantitative sampling – schools and learners

School Selection

The sampling frame began with the 2019 Annual School Census (ASC) list of primary schools as provided by UNICEF/MBSSE where the sample of 260 schools for this study were obtained from an initial list of 7,154 primary schools. Only schools that meet a pre-defined selection criteria were eligible for sampling.

To achieve the recommended sample size of 10 learners per grade, schools that had an enrolment of at least 30 learners in Grade 2 in 2019 were considered. To achieve a high level of confidence in the findings and generate enough data for analysis, the selection criteria only considered schools that:

- had an enrolment of at least 30 learners in grade 1; and
- had an active grade 4 in 2019 (enrolment not zero)

The sample was taken from a population of 4,597 primary schools that met the eligibility criteria above, representing 64.3 per cent of all the 7,154 primary schools in Sierra Leone (as per the 2019 school census). Schools with higher numbers of learners were purposefully selected to ensure the sample size could be met in each site.

As a result, a sample of 260 schools were drawn using proportional to size allocation with simple random sampling without replacement in each stratum. In the population, there were 16 districts and five school ownership categories (community, government, mission/religious, private and others). A total of 63 strata were made by forming combinations of the 16 districts and school ownership categories. In each stratum, a sample size was computed proportional to the total population and samples were drawn randomly without replacement. Drawing from other EGRA/EGMA studies conducted by Montrose in the past, a backup sample of up to 78 schools (30 per cent of the sample population) with which enumerator teams can replace sample schools was also drawn.

In the distribution of sampled schools by ownership, majority of the sampled schools are owned by mission/religious group (62.7 per cent, $n=163$) followed by the government owned schools at 18.5 per cent ($n=48$). Additionally, in school distribution by district, majority of the sampled schools (54 per cent) were found in Bo, Kambia, Kenema, Kono, Port Loko and Kailahun districts. Refer to **annex 9.1** for details on the population and sample distribution by district.

Learner Selection

Systematic random sampling was undertaken to select the required number of learners in each grade. The enumerator team leader was required to randomly select ten learners of proportional gender representation. Pupils were drawn from 2 grades per school, grade 2 and grade 4.

Selection Procedure

1. For each class where sampling was being done, learners present were asked to form two (2) queues: one for girls and another one for boys.
2. The number of learners in each queue was counted and noted down (#girls, #boys)
3. The sampling interval "i" for each queue was obtained. "i" for girls = $\#girls/5$ and "i" for boys = $\#boys/5$. Then the result round-off to the nearest whole number was calculated⁸.
4. Where there were multiple streams, we included both (or all three) streams in the sampling frame, i.e., lining up boys and girls from all streams in a single line, to create the interval. Once the sample was drawn, enumerators began conducting the assessment with learners.

A total of 10 learners, 5 girls and 5 boys, were randomly selected from each grade, 2 and 4. This brought the total sample of learners to 20 learners per school, or around 5,200 learners in 260 schools. Even if the total sample of learners was not met due to low enrolment and attendance in grades 2 and 4 on the day of the visit, the findings would remain reliable and representative given that sampling calculations were based on the school as the unit of selection, rather than the learner.

9.3. EGRA/EGMA

Objectives of tool changes

1. Use one passage (or set of letters or words) across all assessments in the study to avoid ceiling effects that would occur if the items were scored only on accuracy across two different tests.
2. Avoid creating multiple passages or word lists when resources and time for both assessment development and enumerator training are limited – as in this study.
3. Spending more time learning to administer one test will greatly improve enumerator accuracy and reliability, which will support higher quality data collection, greater consistency, and better fieldwork outputs.

⁸ The number of learners is divided by 5 because the sample population drawn from each class is expected to have equal representation of girls and boys.

Benefits of tool changes

1. Key subtasks are timed at 60 seconds throughout the assessment. Learners are scored for both accuracy (correct answers) as well as time (the faster you finish the subtask correctly, the more competent you are) as they answer each subtask.
2. Therefore, it is possible to evaluate the individual skills of learners across a range of ability levels using the same harmonised assessment of core foundational skills.
3. Learners who perform better on the subtask will complete tasks faster and with more accuracy; their correct word per minute scores will reflect this. Success can also possibly be measured against national targets for grade 2 and 4 reading and maths.

Table 42: Summary of EGRA Subtask Changes

Subtask	Change Made	Justification for Change
Letter name knowledge	Removed subtask from assessment	The letter sound knowledge subtask remains in the assessment to test alphabetic principle, while letter name knowledge is removed. Given the length of the assessment it is useful to remove any partially overlapping subtasks (where possible) to reduce test time.
Oral reading fluency and comprehension	Renamed original subtask as level 1 and added another passage as level 2	A second, longer, Primary 4 level oral reading fluency and comprehension subtask was added after the initial oral reading task. Only Primary 4 learners who read at least one word of the level 1 story correctly will take it (i.e., they did not discontinue their oral reading fluency test).

Table 43: Summary of EGMA Subtask Changes

Subtask	Change Made	Justification for Change
Addition – level 1 and level 2	Renamed original subtask as level 1 and added another set of addition problems as level 2	A second, more difficult, addition subtask was added after the initial addition task. Only Primary 4 learners who answer at least one level 1 addition question correctly will take it (i.e., the learner did not discontinue the level 1 subtraction task).
Subtraction – level 1 and level 2	Renamed original subtask as level 1 and added another set of addition problems as level 2	A second, more difficult, subtraction subtask was added after the initial subtraction task. Only Primary 4 learners who answer at least one level 1 subtraction question correctly will take it (i.e., the learner did not discontinue the level 1 subtraction task).
Word problems – level 1 and level 2	Renamed original subtask as level 1 and added another set of word problems as level 2	A second Primary 4 level set of word problems was added after the initial word problems task. Only Primary 4 learners who answered at least one level 1 word problem correctly will take it (e.g., they did not discontinue the level 1-word problems).

Below is a table showing the subtasks that are included in the current version of both learning assessment tools.

Table 44: Tasks that shall be included in the learning assessment prototypes

EGRA subtasks	EGMA Subtasks
	Number identification
Letter recognition – letter sounds	Quantity discrimination
Familiar word reading	Number patterns (missing numbers)
Non-word reading	Addition Level 1 & 2
Oral reading fluency & Reading comprehension 1 & 2	Subtraction Level 1 & 2
Listening comprehension	Word problems 1 & 2

9.4. The World Bank Teach tool

Teach is a free classroom observation tool that provides a window into one of the less explored and more important aspects of a learner's education: what goes on in the classroom. The tool is intended to be used in primary classrooms (grades 1-6) and was designed to help low- and middle-income countries track and improve teaching quality.

Teach can be used as a tool as a tool for system diagnostic and for professional development. As a system diagnostic, Teach allows governments to monitor the effectiveness of their policies to improve teacher practices. As a professional development tool, Teach is used to identify individual teachers' strengths and weaknesses. Teach differs from other classroom observation tools in that it captures (i) the time teachers spend on learning and the extent to which learners are on task, and (ii) the quality of teaching practices that help develop learners' socioemotional and cognitive skills.

As part of the Time on Task component, 3 snapshots of 1–10 seconds are used to record both the teacher's actions and the number of learners who are on task throughout the observation. The Quality of Teaching Practices component, on the other hand, is organised into 3 primary areas as shown below: Classroom Culture, Instruction, and Socioemotional Skills. These areas have 9 corresponding elements that point to 28 behaviours. The behaviours are characterised as low, medium, or high, based on the evidence collected during the observation. These behaviour scores are translated into a 5-point scale that quantifies teaching practices as captured in a series of two, 15-minute lesson observations.

		
<p>CLASSROOM CULTURE:</p> <p>The teacher creates a culture that is conducive to learning. The focus here is not on the teacher correcting students' negative behaviors but rather the extent to which the teacher creates a supportive learning environment and sets positive behavioral expectations.</p>	<p>INSTRUCTION:</p> <p>The teacher instructs in a way that deepens student understanding and encourages critical thought and analysis. The focus here is not on content-specific methods of instruction, but rather, the extent to which the teacher facilitates the lesson, checks for understanding, provides feedback, and encourages students to think critically.</p>	<p>SOCIOEMOTIONAL SKILLS:</p> <p>The teacher fosters socioemotional skills that encourage students to succeed both inside and outside the classroom. To develop students' social and emotional skills, the teacher instills autonomy, promotes perseverance, and fosters social and collaborative skills.</p>

9.5. Data collection tools

Attached as a separate document.

9.7. Research ethics

Montrose followed the UNEG Ethical Guidelines for Evaluations' and the 'UNICEF Procedure for Ethical Standards in Research, Evaluation, Data Collection and Analysis'. Additionally, Montrose paid heed to the ethical guidelines on conducting research with children.⁹

Montrose expected no physical, psychological, social, or legal risks to respondents. The main risk is of a breach of confidentiality. This risk was mitigated by storing all identifiable data securely using encrypted, password-protected files, and by anonymising data (removing participant names) prior to analysis. If at any point monitoring showed any potential harm to participants because of participation in the study, Montrose was to consult immediately with UNICEF on further measures, including potentially halting the study. Montrose had no reason to believe that there are any risks to study participants.

Additionally, one of the main priorities for the enumerators, prior to beginning the survey, was to ensure that they conduct the survey in a safe and secure environment during daytime where the respondent feels comfortable answering questions. Learner assessments were conducted in an area designated by the Headteacher, preferably within the school compound where learners and enumerators had the full visibility of the staff room. Respondents who wanted to skip questions or withdraw from further participation were informed of the freedom to do so at any time.

The study was designed to protect confidentiality of subjects. However, it was necessary to follow individuals over time, so Montrose assigned each participant a unique data identifier that was used throughout the study to establish a link between the name of the individual and their data. Information from the coversheets include names and identifying information. That information was saved separately and securely and was not linked during data entry. All computers that contained this project data were password protected and stored in secure locations.

Montrose implemented appropriate steps to adequately protect human subjects against potential breaches in confidentiality. The protection of respondents in this study follow established procedures such as using a respondent's name only during data collection and separating the identifying information from the response immediately after the interview so that individuals could not be matched with their responses. Montrose protected subject privacy by storing all identifiable data in encrypted form with password controls. Other than names and contact information, no sensitive information was gathered. Soft copy data was stored on an online server that was also encrypted and password protected.

A designated member of the senior Montrose team was responsible for data security and only the enumerators and Montrose staff had access to respondent information. As planned, paper forms containing data will be destroyed three years after the completion of the study, while the electronic records will remain anonymised and stored in a research data repository. Data entry took place inside a locked room under the oversight of Montrose in-county partner and additional data entry specialists. Data capturers were contracted and signed confidentiality agreements in addition to working on Montrose-owned computers, with data files collected and stored securely in encrypted format at the end of each day of data entry.

9.8. Ethical considerations due to COVID-19

In addition to following UNEG Ethical Guidelines for Evaluations' and the 'UNICEF Procedure for Ethical Standards in Research, Evaluation, Data Collection and Analysis', Montrose employed the following

⁹ Berman, Gabrielle (2020). Ethical Considerations for Evidence Generation Involving Children on the COVID-19 Pandemic, *Innocenti Discussion Papers* no. 2020-01, UNICEF Office of Research - Innocenti, Florence

guidelines while conducting fieldwork due to the current COVID-19 pandemic. Throughout this study, Montrose followed the guidelines and standard operating procedures from Ministry of Health and Sanitation (MoHS).

Pre-data collection

- To safeguard enumerators, the research team and host communities from increased risk of contracting COVID-19, and where possible, Montrose did its best to ensure enumerators could work/collect data in the districts/communities where they already lived.
- In preparation for data collection, enumerators underwent a six-day in-person training. The training was held in facilities where there was sufficient space to allow social distancing, and to minimise contact and reduce the number of people in the training at one time.
- The training venue was arranged per the WHO prescribed social distancing measures by hiring a room with twice the seating capacity. A well-ventilated room was sought out to ensure appropriate air circulation.
- During the face-to-face training, enumerators were expected to wash their hands before entering the venue, always wear their face masks while in the training room and have their temperature checked whenever they entered the premises.
- To ensure that appropriate hygiene was maintained within the training venue, table surfaces, and electronic tablets were wiped at least three times a day before use. Enumerators were also equipped with hand sanitisers at the beginning of the training for their own personal use.
- In addition to child protection policy training, enumerators received a copy of the COVID-19 SOPs provided by MoHS which were reviewed at the start of each day of the training.

During data collection

- For the duration of the fieldwork, enumerators were equipped with non-reusable facemasks to protect them and participants from infection. Each enumerator used one facemask per day.
- Each enumerator team was furnished with facemasks for each of the participants. Anyone that was to be interviewed or participate in the focus group discussions was required to wear a protective facemask before they could be engaged.
- Each team was also provided with a thermometer that was to be used to measure their temperature. The temperature reading for each enumerator was to be recorded by the team leader to help with contact tracing purposes in case of an infection.
- Depending on the location of the data collection site, enumerators were to walk or be driven in well-ventilated vehicles. Vehicles were to seat no more than four people (including the driver) and where possible all windows were to be left open. Travel using a vehicle was not permitted if windows could not be left open e.g., when it is raining or if the window was faulty. As much as possible, Montrose endeavoured to hire vehicles that were in excellent condition, and/or use other safer locally available transport options.
- Enumerators were encouraged to desist from eating while in the field.
- To reduce the risk of contracting COVID-19, enumerators were to maintain the prescribed social distancing measures between themselves and participants during data collection.

9.9. Additional tables and graphs

9.9.1. Data collection summary and respondent sample

Round	No. of schools	P2 Assessments		P4 Assessments		Total Pupils	Head Teacher Questionnaire
		Boys	Girls	Boys	Girls		
Round 1 (Planned)	150	750	750	750	750	3,000	150
Round 1 (Actual)	147 (98%)	750 (100%)	733 (97.7%)	702 (93.6%)	698 (93%)	2,883 (96%)	147 (98%)

Round	No. of schools	P2 Assessments		P4 Assessments		Total Pupils	Head Teacher Questionnaire
		Boys	Girls	Boys	Girls		
Round 2 (Planned)	110	550	550	550	550	2,200	110
Round 2 (Actual)	113 (101%)	513 (93.2%)	497 (90%)	488 (88.7%)	492 (89.4%)	1,990 (90.4%)	113 (101%)
Total (Planned)	260	1,300	1,300	1,300	1,300	5,200	260
Total (Actual)	260	1,263	1,230	1,190	1,190	4,873	260
Gap	0	37	70	110	110	327	0
% Reached	100%	97.1%	94.6%	91.5%	91.5%	93.7%	100%

Round	Teacher Questionnaire		Lesson Observation Tool		School Observation	Grade Roster	Employees Roster
	P2	P4	Literacy	Numeracy			
Round 1 (Planned)	150	150	150	150	150	150	150
Round 1 (Actual)	139	129	124	126	147	134	0
Round 2 (Planned)	110	110	110	110	110	110	110
Round 2 (Actual)	101	103	97	88	113	130	128
Total (Planned)	260	260	260	260	260	260	260
Total (Actual)	240	232	221	214	260	264	242
Gap	20	28	39	46	0	0	18
% Reached	92.3%	89.2%	85%	82.3%	100%	101%	93%

Round	P2 Lesson Observation Video		P4 Lesson Observation Video		FGD with PTA	Interview with DEO
	Literacy	Numeracy	Literacy	Numeracy		
Round 1 (Planned)	15	15	15	15	32	16
Round 1 (Actual)	11	16	7	15	19	2
Round 2 (Planned)	11	11	11	11	32	-
Round 2 (Actual)	9	8	11	7	22	3
Total (Planned)	26	26	26	26	64	16
Total (Actual)	20	24	18	22	41	5
Gap	6	2	8	4	23	11
% Reached	77%	92%	69%	85%	64%	31%

9.9.2. Challenges and mitigation measures during fieldwork

Table 45: Challenges experienced in the fieldwork scheduling and school sample

Topic	Details	Mitigation Measure Taken
School Sample	Poor road conditions caused by flooding that resulted in some schools in Tonkolili and Falaba being inaccessible.	<ul style="list-style-type: none"> The existing communication protocol and presence of a TIMAP staff in the field allowed for the timely replacement of any schools we couldn't reach. The replacement school was notified in time for data collection within the same week. Inaccessible schools were replaced.
	The available national list of schools provided for sampling was outdated, resulting in a substantial number of schools being replaced because the schools were either not found or not accessible by road/vehicle.	<ul style="list-style-type: none"> A back up list was generated at the time of initial sampling to allow for one-to-one replacement of schools in the sample by similar demographic characteristics. Details were recorded in team activity reports for each school.
Timing of Assessment – Round 1 and Round 2	Changes in school calendar that meant that schools were found closed on the day of the assessment due to mid-term breaks during Round 1 of data collection.	<ul style="list-style-type: none"> The enumerator deployment plan was revised to conclude later in June and schools were revisited where necessary or replaced in round 2. Muslim schools are closed on Fridays, so data collection continued in those locations from Sunday-Thursday only.
	Some enumerators couldn't continue to round 2 of the data collections due to sickness and bereavements in their families.	<ul style="list-style-type: none"> Assistant team leaders from high functioning teams were selected to replace any team leaders that had dropped out of the study. These new team leaders received close monitoring and extra attention from the Research Specialist as they took on their new duties. Teams were reconstituted to 4 enumerators each, which was sufficient to collect data in the schools given lower attendance rates for teachers and learners.
	Delays in finalising contract discussions for the additional 110 schools resulted in round 2 of data collection starting later than planned, with a 3-day gap between each round. As a result, some schools had finished conducting lessons and taken their end of term exams by the time they were visited during round 2.	<ul style="list-style-type: none"> The goal is to ensure equivalency between the two cohorts in rounds 1 and 2 to prove the data is comparable despite being collected at different times. While schools were selected using the same metrics, the timing of the sample's data collection differed, with cohort 2 fieldwork running later in the school term (e.g., after additional weeks of learning, studying and final exams). We must prove there are no inherent differences between the two samples of teachers and learners due to these differences. We'll run analysis to this effect and present it in the report.

Table 46: Challenges experienced in collecting quantitative data

Theme	Details	Mitigation Measures
Learner Data	Failure to get the intended 5 girls and 5 boys per grade due to the existing class composition or attendance on the day of the assessment.	<ul style="list-style-type: none"> The gender distribution of the sample was intended to be $\sim\pm 5$ in each grade, so findings are acceptable.
	Even with mobilisation efforts and confirming the number of learners in the school prior to the assessment with school administrators, some schools did not have the required number of learners in attendance on the day of the assessment. This affected the ability of the teams to collect complete learner data in all schools in round 2.	<ul style="list-style-type: none"> The learner and teacher sample, while reduced in some cases, was largely met, as was the school sample. The data collected is enough to provide a statistically sufficient, representative sample for analysis.
Teacher Data	Even with the delivery of the letter from the Ministry and a follow-up phone call from the enumerator team leader, some teachers and head teachers and teachers were not found in the school on the day of the assessment.	<ul style="list-style-type: none"> This was mitigated where possible with repeated contacts to the head teachers and rescheduling where necessary. Where possible the teams mobilised teachers to the school once they arrived if some of the teachers were staying nearby.
	<p>Teacher samples were affected by:</p> <ul style="list-style-type: none"> Schools not having enough teachers in the target grades. Teachers not coming to school as they are not being paid; the head teachers in some cases cannot hold them accountable since they are not on the payroll. Some teachers preferred to work in their farms than come to school to teach even if on the government payroll, as they had not been paid for a long time. In some schools in round 2, the head teachers were the only teachers in the school, and they were running six classes alone. 	<ul style="list-style-type: none"> Data was collected from teachers found in the school on the day of the assessment.

Table 47: Challenges experienced in collecting qualitative data

Theme	Challenges	Mitigation Measures
Lesson Observation Videos	Some classrooms were too dark to be filmed well for the lesson observation.	<ul style="list-style-type: none"> Videos were attempted in these locations and will be reviewed for inclusion in data analysis if clear enough.
PTA Data	Some PTA focus group discussions were not held because participants were absent/not mobilised by the head teacher.	<ul style="list-style-type: none"> These were skipped and not replaced, as there were enough PTA observations from other schools.
District Official Data	District education officers were not available to be interviewed.	<ul style="list-style-type: none"> Where possible, DEO interviews were re-scheduled while others were not held completely.

Table 48: Challenges in data quality assurance

Topic	Details	Mitigation Measure
Connectivity and Daily Reporting	Internet connectivity issues hindered daily upload of data while in the field, as a result the monitoring team had to wait until the end of the week to review the data that was collected.	<ul style="list-style-type: none"> The team leaders' WhatsApp group was used by the technical team to maintain frequent contact with all teams and to provide daily summaries of the teams' performance and troubleshoot technical and logistical challenges. It also allowed for the timely identification of inaccessible schools by the TIMAP staff and technical team. Data was saved to the tablet and uploaded when an internet connection was available and strong enough to support the uploading exercise.
In-Person Technical Support	In-country technical support was needed for the enumerators after Montrose concluded its visit.	<ul style="list-style-type: none"> The lack of a technical oversight in the field was mitigated by Montrose hiring a national Research Specialist (with extensive EGRA/EGMA experience) who visited all teams and supervised the intake process at the end of each data collection week.

9.10. Reliability analysis

In this study, we carried out reliability analyses on the English EGRA and EGMA subtasks. To do this, Pearson correlations was first used to compute the reliability coefficients among each tool's subtasks. Strong correlations among subtasks are desired because they typically indicate consistency in the performance of sampled pupils across subtasks. Additionally, the Cronbach's alpha for all the subtasks in each tool was computed to measure the internal consistency of the tool. When combined with item-level analyses, internal consistency offers insights on subtask functioning.

Cronbach's alpha estimates items' consistency in measuring the concepts of interest within each subtask (Aron, Coups, & Aron, 2013; Cronbach, 1951). We first present the Pearson correlations for each subtask of the English EGRA tool, then the Cronbach's alpha scores for the English EGRA tool subtasks, and lastly, the Pearson correlations and Cronbach's alpha scores for the EGMA tool subtasks.

9.10.1. English EGRA tool analysis

Table 49 shows the correlations for the English EGRA subtasks. All the coefficients were statistically significant ($p < .001$), except letter sound fluency and reading comprehension 1, and listening comprehension 2. The pairs of oral reading fluency and non-word decoding (0.69), familiar word fluency and non-word decoding (0.73), oral reading fluency 2 and oral reading fluency 1 (0.67), and oral reading fluency and familiar word fluency (0.88) were positive and highly correlated. Positive and moderate correlations were observed between the pairs of oral reading fluency 2 and non-word decoding (0.52), and oral reading fluency 2 and familiar word fluency (0.62). Low positive correlations were observed among the remaining subtasks except listening comprehension that had negative correlations with all other subtasks, except with reading comprehension 1 (0.12) and reading comprehension 2 (0.00).

Table 49: Pearson correlations for EGRA subtasks in English

EGRA Subtasks	Letter-sound fluency	Non-word fluency	Familiar word fluency	Oral reading passage 1 fluency	Reading comprehension 1	Oral reading passage 2 fluency	Reading comprehension 2	Listening comprehension
Letter-sound fluency	1							
Non-word decoding	0.48***	1						
Familiar word fluency	0.39***	0.73***	1					
Oral reading passage 1 fluency	0.36***	0.69***	0.88***	1				
Reading comprehension 1	0.02	0.15***	0.24***	0.30***	1			
Oral reading passage 2 fluency	0.28***	0.52***	0.62***	0.67***	0.16***	1		
Reading comprehension 2	0.08***	0.20***	0.31***	0.31***	0.28***	0.26***	1	

EGRA Subtasks	Letter-sound fluency	Non-word fluency	Familiar word fluency	Oral reading passage 1 fluency	Reading comprehension 1	Oral reading passage 2 fluency	Reading comprehension 2	Listening comprehension
Listening comprehension	-0.15***	-0.19***	-0.24***	-0.21***	0.12***	-0.18***	0.00	1

* $p < .05$, ** $p < 0.01$, *** $p < 0.001$.

Table 50 shows the Cronbach's alpha coefficient for the English EGRA subtasks. Similar to **Table 52** below, this table, shows the correlations between each subtask and the entire scale (i.e., item-test correlations); the correlations between each subtask and the scale including all other subtasks (i.e., item-rest correlations); and the overall Cronbach's alpha. We followed the accepted rules (Bland & Altman, 1997; Tavakol & Dennick, 2011), listed below, to make decisions on the alpha statistics:

- $\alpha \geq 0.9$: Excellent
- $0.9 > \alpha \geq 0.8$: Good
- $0.8 > \alpha \geq 0.7$: Acceptable
- $0.7 > \alpha \geq 0.6$: Questionable
- $0.6 > \alpha \geq 0.5$: Poor
- $\alpha \geq 0.5$: Unacceptable

The highest coefficient of 0.80 was found in reading comprehension 1 and listening comprehension. The lowest coefficient was 0.69 for familiar word fluency and oral reading fluency 1. These results show that the English EGRA tool was reliable for assessing the baseline study group of pupils, with an overall reliability coefficient of 0.78.

Table 50: Cronbach's alpha for EGRA subtasks in English

Subtask	Item-test correlation	Item-rest correlation	Alpha
Letter sounds	0.56	0.38	0.77
Non-word fluency	0.78	0.68	0.71
Familiar word fluency	0.87	0.80	0.69
Oral reading passage 1 fluency	0.86	0.79	0.69
Reading comprehension 1	0.41	0.21	0.80
Oral reading passage 2 fluency	0.74	0.64	0.74
Reading comprehension 2	0.48	0.30	0.78
Listening comprehension	0.38	0.17	0.80
Totals			0.78

9.10.2. EGMA tool analysis

We also carried out reliability analyses for the EGMA tool used in this baseline study. **Table 51** shows the bivariate correlations among the EGMA subtasks. All the correlations were statistically significant ($p < .001$). Strong positive correlations were observed between the pairs of quantity discrimination and number identification (0.87); missing number and number identification (0.71); addition level 1 and number identification (0.70); missing number and quantity discrimination (0.76); addition level 1 and quantity discrimination (0.71); subtraction level 1 and addition level 1 (0.72); and subtraction level 2 and addition level 2 (0.76). Moderate positive correlations were observed between the remaining pairs of subtasks.

Table 51: Pearson correlations for EGMA subtasks

EGMA Subtasks	Number identification	Quantity discrimination	Missing number	Addition level 1	Addition level 2	Subtraction level 1	Subtraction level 2	Word problems level 1	Word problems level 2
Number identification	1								
Quantity discrimination	0.87***	1							
Missing number	0.71***	0.76***	1						
Addition level 1	0.70***	0.71***	0.68***	1					
Addition level 2	0.54***	0.57***	0.54***	0.60***	1				
Subtraction level 1	0.57***	0.61***	0.63***	0.72***	0.56***	1			
Subtraction level 2	0.42***	0.46***	0.48***	0.48***	0.76***	0.56***	1		
Word problems level 1	0.58***	0.62***	0.68***	0.63***	0.46***	0.59***	0.43***	1	
Word problems level 2	0.47***	0.50***	0.52***	0.51***	0.68***	0.49***	0.62***	0.52***	1

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 52 shows the Cronbach's alpha coefficients for the EGMA tool. The highest coefficient of 0.93 was found in subtraction level 2, while the lowest coefficient was 0.92 for all the remaining EGMA subtasks. These results show that the EGMA tool was highly reliable for assessing the Sierra Leone baseline study group of pupils, with an overall reliability coefficient of 0.93.

Table 52: Cronbach's alpha for EGMA subtasks

Subtask	Item-test correlation	Item-rest correlation	Alpha
Number identification	0.82	0.76	0.92
Quantity discrimination	0.85	0.81	0.92
Missing number	0.84	0.78	0.92
Addition level 1	0.84	0.79	0.92
Addition level 2	0.80	0.73	0.92
Subtraction level 1	0.80	0.74	0.92

Subtask	Item-test correlation	Item-rest correlation	Alpha
Subtraction level 2	0.72	0.65	0.93
Word problems level 1	0.77	0.70	0.92
Word problems level 2	0.74	0.67	0.92
Totals			0.93

References

- Aron, A., Coups, E. J., & Aron, E. (2013). *Statistics for psychology* (6th ed.). Boston, Massachusetts, USA: Pearson.
- Bland, M. J., & Altman, D. G. (1997, February 22). Statistics notes: Cronbach's alpha. *British Medical Journal*, 314:572. doi:<https://doi.org/10.1136/bmj.314.7080.572>
- Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, 53–55. Retrieved from <https://doi.org/10.5116/ijme.4dfb.8dfd>

9.11. Additional regression analysis

Subtask	Had breakfast at home			Has electricity at home			Has a radio at home			Went to pre-school			Repeated grade			Read/study at home		
	Coef.	Std. Err.	p-value	Coef.	Std. Err.	p-value	Coef.	Std. Err.	p-value	Coef.	Std. Err.	p-value	Coef.	Std. Err.	p-value	Coef.	Std. Err.	p-value
Letter sounds	1.8	0.70	0.011*	0.9	0.56	0.11	-0.2	0.54	0.77	0.1	0.61	0.88	-1.6	1.68	0.33	0.8	0.67	0.24
Non-words	-0.2	0.61	0.77	1.9	0.47	0.001**	-0.5	0.46	0.25	3.3	0.51	0.001**	-5.5	1.49	0.001**	3.0	0.68	0.001**
Familiar words	1.2	0.53	0.021*	3.3	0.44	0.001**	1.2	0.42	0.001**	4.9	0.47	0.001**	-5.3	1.34	0.001**	6.0	0.52	0.001**
Oral reading story 1	0.8	0.79	0.32	4.7	0.66	0.001**	1.4	0.63	0.021*	7.1	0.70	0.001**	-4.0	2.10	0.06	8.0	0.80	0.001**
Reading comprehension 1	0.0	0.03	0.51	0.1	0.03	0.001**	0.0	0.03	0.33	0.1	0.03	0.06	-0.1	0.10	0.15	0.0	0.03	0.23
Oral reading story 2	0.7	1.32	0.59	7.2	1.09	0.001**	2.7	1.03	0.011**	8.0	1.16	0.001**	-8.6	3.58	0.021*	6.7	1.54	0.001**
Reading comprehension 2	0.0	0.03	0.67	0.0	0.03	0.23	0.0	0.02	0.56	0.0	0.03	0.77	-0.1	0.09	0.44	0.0	0.04	0.96
Listening comprehension	0.0	0.04	0.24	-0.1	0.04	0.05	-0.1	0.03	0.011**	-0.2	0.04	0.001**	0.0	0.09	0.59	-0.1	0.04	0.001**
Number identification	1.3	0.25	0.001**	1.4	0.22	0.001**	1.6	0.20	0.001**	1.6	0.23	0.001**	-1.5	0.59	0.011**	3.9	0.23	0.001**
Quantity discrimination	0.6	0.13	0.001**	0.7	0.11	0.001**	0.6	0.10	0.001**	0.5	0.12	0.001**	-0.4	0.31	0.17	1.7	0.12	0.001**
Missing number	0.7	0.10	0.001**	0.3	0.09	0.001**	0.3	0.08	0.001**	0.4	0.10	0.001**	-0.5	0.26	0.051*	0.7	0.10	0.001**
Addition level 1	0.6	0.18	0.001**	0.5	0.16	0.001**	0.5	0.15	0.001**	0.7	0.17	0.001**	-0.6	0.43	0.19	1.7	0.17	0.001**
Addition level 2	0.2	0.08	0.06	0.4	0.07	0.001**	-0.1	0.07	0.17	0.4	0.08	0.001**	-0.3	0.21	0.14	0.4	0.10	0.001**
Subtraction level 1	-0.1	0.20	0.59	0.2	0.17	0.14	0.2	0.16	0.34	0.5	0.18	0.001**	-0.4	0.52	0.39	1.1	0.20	0.001**
Subtraction level 2	0.1	0.10	0.16	0.3	0.08	0.001**	0.0	0.08	0.89	0.3	0.09	0.001**	-0.5	0.27	0.09	0.2	0.12	0.08
Word problems 1	0.1	0.04	0.011**	0.1	0.03	0.001**	0.1	0.03	0.011**	0.1	0.04	0.05	-0.3	0.10	0.011**	0.2	0.04	0.001**
Word problems 2	0.1	0.03	0.041*	0.1	0.03	0.07	0.0	0.03	0.07	0.0	0.03	0.59	-0.2	0.08	0.011*	0.0	0.04	0.21