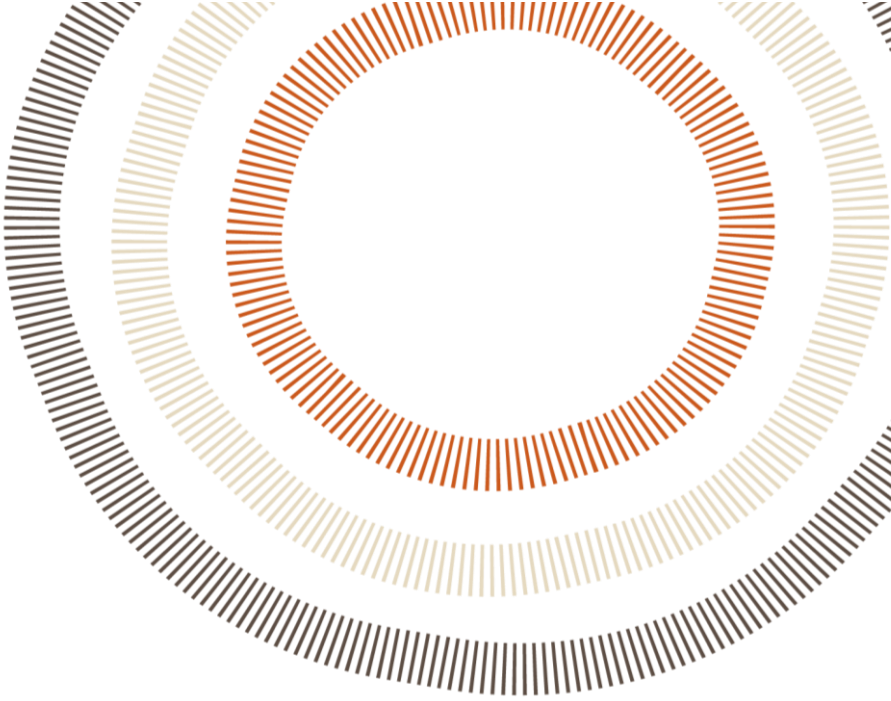

Early Grade Reading Programme Evaluation

Midline Report 2021

Acknowledgements

This intervention programme is Led by the National Department of Basic Education (DBE) and the North West provincial Education Department.

Implementation is done by Molteno Institute for Language and Literacy and funded by The Hempel Foundation in conjunction with UNICEF South Africa (SACO)

Evaluation is done by Social Surveys Africa (SSA) and funded by Zenex Foundation.

Table of Contents

List of Tables

List of Figures

Abbreviations and Acronyms

ANA	Annual National Assessments
CAPS	Curriculum and Assessment Policy Statement
CLPM	Correct Letters Per Minute
CWPM	Correct Words Per Minute
DBE	Department of Basic Education
EFAL	English as First Additional Language
EGRS	Early Grade Reading Study
EGRS II	Second Early Grade Reading Study
EGRA	Early Grade Reading Assessment
GGR	Group guided reading
HOD	Head of Department
ICC	Intra-class Correlation Coefficient
ICT	Information and Communications Technology
ISPFTED	Strategic Planning Framework for Teacher Education & Development
LTSM	Learning and Teaching Support Materials
OLS	Ordinary Least Squares
PIRLS	Progress in International Reading and Literacy Study
PDE	Provincial Department of Education
PLC	Professional Learning Communities
PMT	Project Management Team
RCT	Randomised Control Trial
RON	Rapid Object Naming
RSP	Reading Support Programme
SMT	School Management Team
SSA	Social Surveys Africa
SLP	Structured Learning Programme
ToC	Theory of Change

Executive Summary

This report presents the findings of the midline evaluation of the Early Grade Reading Programme (EGRP). The project implementation started in 2021 in the Dr Ruth Segomotsi Mompati District and will continue up to 2023. The project's evaluation is being implemented by Social Surveys Africa (SSA) on behalf of the Zenex Foundation, in partnership with the Department of Basic Education (DBE).

The importance of the EGRP cannot be overstated: reading ability among South Africa's youth, which is rooted in their learning trajectory in the early grades, is a real concern as it is holding back national development and the ability of many adults to rise out of poverty. The ability to read with understanding is key to all learning, formal employment and life chances.

The EGRP builds on lessons learned in EGRS I and EGRS II, which showed, respectively for Setswana home language reading in North West Province and EFAL reading in Mpumalanga, that a basic package of teacher support through training, materials provision (reading materials and lesson plans), and coaching is effective in improving learner reading outcomes. The most significant positive impact on learners' reading performance from both these studies was observed when support to teachers included coaching by external professional early-grade reading coaches. However, whilst impactful, the professional coaching component is also the most expensive part of these interventions.

This latest iteration of these studies uses an experimental design to test various hypotheses about cost-effective ways of supporting early-grade teachers implementing a range of ways of teaching reading. In particular, it aims to test whether Foundation Phase department heads (DHs) can successfully coach their colleagues if given low-dosage professional external coaching support themselves.

The EGRP's Theory of Change (ToC), again building on the findings and successes of the EGRS 1 and 2 studies, uses a combination of learning and teaching support materials (LTSMs), teacher training and two different modalities of coaching to improve the way early-grade reading is taught. The ToC assumes that if teaching improves, then learner reading skills will improve. The project is targeted at 140 schools which are divided randomly into three intervention groups: firstly, a control group of 60 schools that just receives a base program of teacher training and Learning and Teaching Support Materials (LTSMs), secondly, an intervention group of 40 schools that gets the full package of the base program plus external professional coaching throughout the three years, and finally, a third intervention group of 40 schools, which similarly receives the base program the same dosage of external professional coaching in the 1st year, but where the school DH becomes coach from the 2nd year of the project. The success of this third group will be largely determined by learner performance in early grade reading assessment (EGRA) tests, to be conducted in 2023 at the endline evaluation, compared to the performance of learners in the other two groups of schools.

The evaluation is designed as a randomised control trial with a learner cohort design over three years (2021 midline and 2023 endline). This report is written as a midline evaluation since no baseline assessment was conducted before the commencement of the intervention. The evaluation was commissioned in late 2020 and could not commence in practice until early 2021. Its conceptualisation as a baseline and endline impact study would have required a baseline measurement to be taken in early (term 1 or 2) 2021. This would, however, have been difficult to compare with a term 3 endline (2023), and a term 1 or 2 endline (2023) would have shortened the intervention period too much. It was therefore decided to conduct the initial round of learner assessments in term 4 of 2021, by which point the learners in the two intervention arms had already received almost a full year of coaching support plus the base program.

There were a number of implementation challenges in the initial year of implementation fidelity, specifically in terms of the operative coaching dimension, due to a combination of Covid 19 and other technical challenges. On this basis, we debated whether the teachers and learners in the two coaching intervention arms were likely to have received enough input to expect an effect. If not, one could have called the initial learner assessment a de facto baseline. However, there was some level of intervention, and we felt it would be methodologically inappropriate to simply discount this. For this reason, we have analysed the learner assessments in the current report as a midline assessment, using the appropriate statistical techniques detailed further in the report. However, to support a reflection on implementation issues and the COVID-19 pandemic, this report may be viewed as a baseline.

This report presents the findings from the quantitative components of the evaluation study design at midline, including the administration of adapted EGRA, tests to Grade 1 and 2 learners; surveys of teachers, Department Heads and Principals; a parent survey; a collection of school administrative data; and observation of basic school infrastructure conditions. The qualitative elements of the evaluation design (case studies) are presented in a separate report. Similarly, a separate report also covers the analysis of programme implementation fidelity data.

The midline data collection took place in November 2021. This report intends to test whether there are any significant differences between the schools in the control arm that received only the base program and the two intervention arms, which may bias the endline results. The evaluation also predicts the role of contextual factors as determinants of learner reading performance and explores the challenges of coaching delivery through DHs.

This report presents results of the midline learner performance in Setswana (HL) and English First Additional Language (EFAL). This is tested through an adapted one-on-one EGRA test version administered with each learner and a set of group-administered tests.

Literacy data in both HL and EFAL were collected from Grade 1 and Grade 2 learners, randomly sampled, in each of the 140 EGRP schools. In total, 5 812 learners were tested in November 2021 by 20 researchers. These Setswana-speaking researchers had successfully completed rigorous training to prepare them to accurately administer the adapted EGRA tests on tablets and administer the various questionnaires and observation tools which were to provide supplementary data on the school and home environment of the learners.

It should be noted from the outset that the first year of the programme, in 2021, was delivered during the COVID-19 pandemic. During this year, the country was experiencing a range of lockdown regulations which meant that schools were regularly disrupted and limited attendance to half the learners on a daily basis. This placed constraints on the delivery of key programme activities, with consequent fidelity challenges, which led to the possibility of differential programme implementation as the EGRP implementation partner and schools sought to navigate through this challenging period. The main impact was seen in the loss of teaching and learning time as schools implemented rotational learning, and some faced temporary closure from time to time to meet the COVID-19 regulatory requirements. Additionally, some of the teaching methodologies recommended by the EGRP, such as group-guided reading and various group activities, could not be implemented in the manner and frequency planned. The rotational system implemented mainly in poorer schools to avert overcrowding made it difficult for scaffolded learning to take place systematically and consistently. Further, as this evaluation took place while the restrictions were still in place, teachers had to observe COVID-19 protocols. So, even if teachers could implement

group activities as required by EGRP, they could not demonstrate this during the evaluation activities such as case study observations.

These losses in learning imply that some learners are experiencing an onset accumulation of learning backlogs that might hamper their acquisition of foundational reading skills and future learning potential. School programmes to manage these backlogs will be required, and teachers would need support to assist their learners to recover these losses, especially in poor schools that could not support their learners and parents more effectively during these COVID-19 schooling disruptions. Some elements of EGRP, such as school-based workshops, will be critical in helping teachers in this regard.

The main findings of the midline evaluation are that:

- There was no statistically significant difference in learner performance in the intervention schools receiving external coaching compared to the control schools, which received the base program only. This is a positive finding in terms of the validity of the RCT study design but also shows that there is not yet any measurable impact of the intervention on learner performance after a full year of implementation.
- There was a statistically significant difference in learner performance at Grade 1 between schools in the external coaching stream and those in the DH coaching stream, with learners in the external coaching stream performing better, even though by design, there should not have been any difference in the intervention received by these schools in the first year of the intervention. Since no statistically significant differences in school, teacher, home context, or learner characteristics could be found, the reasons for this difference were explored in the implementation fidelity evaluation. This was done through comparing the sample data on the dosage each of the intervention arms received. The results showed that the external coaching stream had 35% of teachers reporting receiving a full dosage of coaching compared to 41% for DH led coaching stream. The results for teacher training dosage indicated that 81% of teachers in the external coaching stream attended all training sessions compared to 70% in the DH led coaching.
- 70% of Grade 1s could not sound letters fluently, and 14% could not sound any single letter. More significantly, 83% of the Grade 1 learners could not read a word of the oral reading fluency test. At the other end of the scale, 17% of Grade 1s could read satisfactorily.
- 58% of Grade 2s could not read a word in HL, whilst 74% could not read a word in EFAL. These learners have failed to gain decoding skills, so the learners remain 'non-readers.' On the other end of the spectrum 29% of Grade 2s could read at a satisfactory level in HL, whilst 18% could read at a satisfactory level in EFAL
- Interestingly, and worryingly, very few Grade 2s (13%) could read some words, but with limited fluency, so within one classroom, there is likely to be a bimodal distribution of reading ability with a few students who can read at the expected level of fluency for their grade[1] and able to understand what they are reading and the majority who cannot read a single word – even in Grade 2 after two years in school. This means that differentiated teaching is essential, even though teachers tend to find this difficult to implement. Results from the group-administered tests were consistently better than those from the one-on-one tests. They provided greater

[1] Expected levels of fluency are based on national benchmarks. These have been developed by the DBE and are language specific. For Grade 2 Setswana (HL) the expected level of fluency is 40 CWPM (correct words per minute) and for EFAL 30 CWPM. For Grade 1 the benchmark is the ability to sound out 40 correct letters per minute (40 CLPM) in HL. There is no Grade 1 benchmark for expected level of fluency in reading HL or EFAL words.

differentiation between learners: the one-on-one tests exhibited many floor effects, particularly for the more predictive tests, such as oral reading fluency and reading comprehension.

- As expected, girls did significantly and consistently better than boys across all the tests and in each intervention group.
- No school-based factors or teacher attributes were found to have a strong statistically significant impact on learner reading scores. Home-environment and learner-specific factors which have a statistically significant impact on learner scores were:
    - parents/caregivers having a tertiary education compared to those with no schooling or partial schooling
    - homes having some books, and particularly more than 5 books
    - learner absenteeism from school.

The report makes some recommendations, although being a midline survey, this is not the main purpose of the report:

Intervention Implementation Recommendations

- The EGRP Implementing Partner should carefully consider why, after nearly a year of coaching engagement in intervention schools, there is no measurable difference in learner reading performance compared to control schools.
- The school observation results found that not all intervention classrooms had teaching aids on the walls, especially in the DH coaching stream, which should be addressed urgently by the EGRP Implementing Partners.
- The programme needs to identify schools in the DH coaching intervention group where the DH is intends to leave the DH position before the programme close-out and ensure the potential successor (where possible) is prepared for the role. Otherwise, a DH turnover or DH vacancies may endanger the comparative design of the intervention and evaluation.

Study Design Recommendations

- More research is needed into why learners do consistently better in the group sub-tasks compared to individual one-on-one tasks.
- For the endline data collection phase, we recommend using only the home context questions posed to learners as part of the learner assessment and not repeating the questionnaire sent home to parents, given the low response rate of the latter.

1 Introduction

The ability to read is a fundamental skill required to operate in modern society. It remains an essential building block for progressing academically and economically. The economic growth prospects of a country are also linked to early grade literacy. One study found that a 10 per cent increase in the share of children achieving basic literacy increases a country's annual economic growth rate by 0.3 per cent (Hanushek and Woessmann, 2012). The same study also found that a 10 per cent increase in the share of students with advanced literacy skills is associated with 1.3 percentage points higher annual economic growth. The combination of improvements in basic literacy and an increasing share of students with advanced literacy skills was found to have a much stronger effect on economic growth. This study suggests that a country needs a workforce with at least basic literacy skills to implement imitation and innovation strategies (Hanushek and Woessmann, 2012). The benefits of improved country-level literacy extend to other economic, political, and social benefits such as improved political participation, health, and gender equality (UNESCO, 2006)

At an individual level, children who learn to read effectively in the early grades are unlikely to struggle to develop more advanced skills, which are often acquired through reading. However, children who cannot follow written information and instructions and cannot communicate well in writing are, in later grades, at risk of falling further and further behind those who can read effectively (Gove & Watterberg, 2011). If effective interventions are not implemented to arrest this cumulative literacy lag, the literacy gap between good and poor readers will widen. This is known as the Matthew Effect (Stanovich 1986).

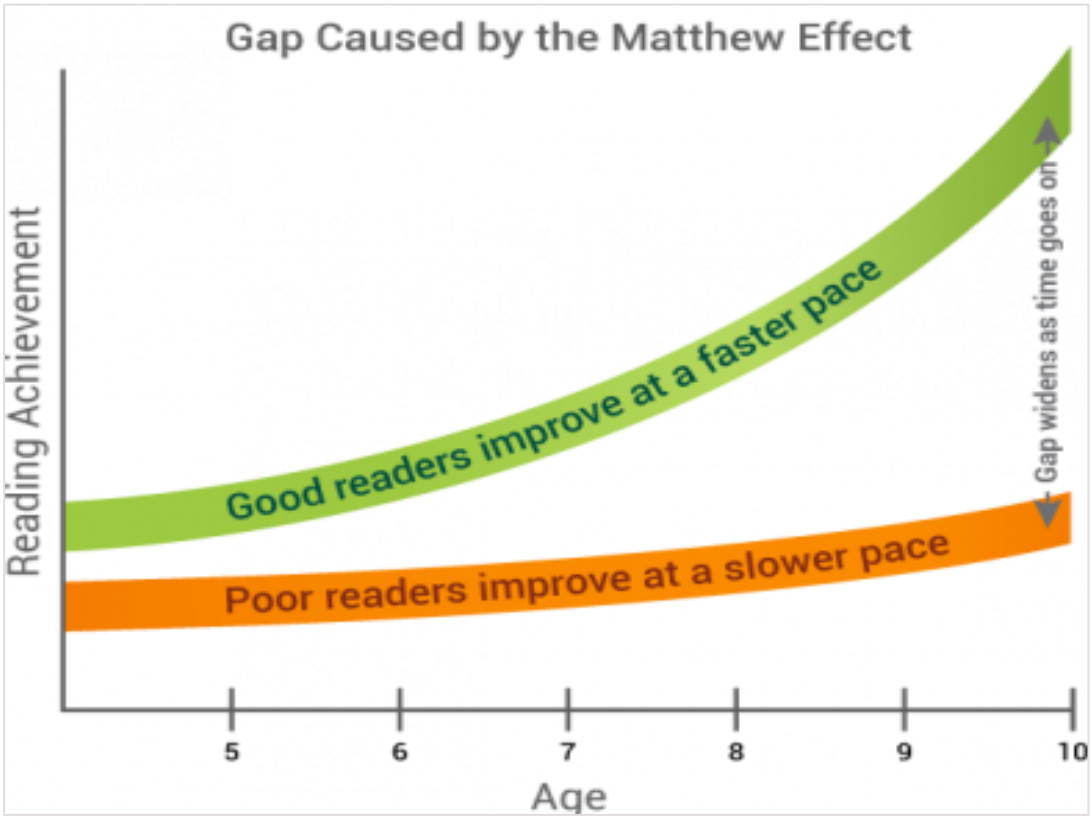

Figure 1: The Matthew Effect

Effective readers will often continue absorbing increasing amounts of written information, building their vocabularies, and strengthening their comprehension. In contrast, ineffective readers succumb to demotivation, reading a fraction of the quantity and failing to comprehend more complex information (Stanovich, 1986). Learning to read is taught in the early Grades 1 to 3 of primary schooling. Children who have not yet acquired this skill by the end of Grade 3 face constraints in further learning.

South Africa's literacy rates across all languages in primary schooling are below what is expected for the country's economic development level and its investment levels in public education. Various studies and large-scale reading tests have recorded low performance in reading at the foundation level (Grades 1 to 3). These studies confirm that most learners are not reading at age-appropriate levels, and 78% of learners cannot read for meaning in any language by the end of Grade 4 (PIRLS, 2016). Particularly concerning is that most of the children who have failed to learn to read for meaning by Grade 4 are learning to read in their African home language. These learners still have to face even steeper demands as further learning shifts from mainly Home Language (HL) in the foundation phase to English in Grade 4, which for these learners is their First Additional Language (EFAL). For example, in 45% of South African Grade 4 classrooms, no learner could read in their home language and make inferences (PIRLS Intermediate International Benchmark, 2016).

Though English is introduced from Grade 1 as an additional language to smooth out this transition in Grade 4, the effects of poorly taught foundation phase EFAL may be potent in limiting further learning for many children. Furthermore, most children have limited access to grade-appropriate reading materials, especially in indigenous languages, which should be the foundation for their development of reading skills.

Researchers and policymakers are increasingly recognising the importance of identifying and implementing effective interventions in the early grades. Van der Berg (2015) shows that the patterns of poor performance in Annual National Assessment (ANA) in later grades across school quintiles already exists by Grade 4. This suggests that learning and developmental deficits from early grades are being carried forward into later grades, significantly impacting on learners' ability to learn and so limiting their later performance.

So, the inability of South African learners to read for meaning in their HL by the end of Grade 3 has a serious impact on their ability to learn and access the curriculum at later grades in English (Van der Berg, Spaull, Wills, Gustafsson, Kotze, 2016). The benefits of effective interventions at lower grades are thus likely to have a knock-on effect throughout a person's life; hence research and policy must be focussed on improving early childhood development, including at the Foundation Phase of the schooling system.

Put simply, if a child fails to learn to read in their first three years in school, then their chances to ever read significantly diminishes, which in turn will usually condemn them to the margins of the economy, and so to a life of relative poverty.

There is considerable evidence in South Africa pointing to the reality that most South African teachers currently do not have the content knowledge and pedagogical skills to effectively convey the curriculum (Venkat & Spaull, 2015). It is thus not surprising that there is a widespread acceptance that instructional practices are a critical lever in improving learning outcomes despite the influence of many other factors (Coe, 2014). Impacting this change at the instructional core is essential to sustaining improvements in how early grade literacy is taught in schools. Alsofrom (n.d)

cites City et al. (2009) in asserting that changes in teaching practices, student engagement, and content are important in attaining measurable improvements in learners' reading performances. Such changes in teaching practices in schools, in turn, require changes in teaching tasks, materials used, and changes in teachers' knowledge.

Interventions targeting influencing instructional practices become more critical when one compares the learning outcomes of schools in wealthy and poorer communities. The differences in outcomes reflect different instructional practices in the schools across these communities (Hoadley, 2012). The patterns of inequality in South Africa continue to echo loudly in learning outcomes, with no-fee schools (quintile 1, 2, 3) which cater for lower socio-economic communities, bearing the brunt of poor performance. This is more evident when one considers that 80% of the students attaining distinctions in Mathematics of Physical science are from only 200 schools out of a total of 6 476 high schools, with almost 93% of these 200 schools charging significant fees (Schotte et al., 2018). Therefore, it is appropriate that policymakers and educational interventions focus on schools in these poorer communities in addition to targeting instructional practices.

The Department of Basic Education (DBE) and its partners have been working to find systematic ways of turning around the reading outcomes of schools from these lower quintiles through its Early Grade Reading Studies (EGRS) in the North West and Mpumalanga provinces. These studies have demonstrated that significant early grade reading outcomes can be achieved through packaged resource interventions that include: supporting teachers in classrooms with expert reading coaches; providing scripted lesson plans; training teachers; and providing learning and teaching support materials. Social Surveys Africa (SSA) was contracted by the Zenex Foundation to carry out implementation fidelity and impact evaluation of the EGRP on behalf of the DBE.

2 Background to the Early Grade Reading Programme

The DBE has been working since 2015 on a series of early grade reading studies designed as randomised control trials. These studies have been building up systematic evidence in an iterative manner on the efficacy of different models of teacher support in the teaching of early grade reading. The first in the studies series was the EGRS I, implemented in the North West province from 2015 to 2017, focusing on Setswana. This was followed by the Reading Support Program (RSP) in the North West - this programme which focussed on Setswana and English First Additional Language teaching practice was implemented from 2019 and is still in progress, and lastly EGRS II, implemented in Mpumalanga province from 2017-2019, focusing on English as a first additional language. The Early Grade Reading programme (EGRP) is the DBE's most recent iteration of these studies, whose design has retained the effective core programme from EGRS I and II (materials provision and teacher training) for all schools. It, however, adds a comparative design for the two coaching approaches (coaching by internal department heads (DH) as compared to external coaching).

EGRS I found that a basic package of teacher support through teacher training; materials provision (reading and visual materials, big books and lesson plans); and literacy coaching of teachers are effective in improving learner reading outcomes in Setswana. EGRS II had the same findings for English First Additional Language (EFAL) reading. The most significant positive impact on learners' reading performance from both these studies was observed when support to teachers included coaching by external professional early grade reading coaches. Learners in the schools that received this intervention had a 40% improvement in reading outcomes within two years of coaching intervention compared to the control schools whose teachers had not received on-site coaching (Taylor et al., 2018).

However, whilst impactful, the professional coaching component is also the most expensive part of these interventions. This presents sustainability and scaling concerns if the DBE implemented this across the South African public schooling system. The option to do this coaching virtually would have been preferable as a cost-saving measure, but this was tested in EGRS II and was found to have limited impact on the teachers and was less effective than on-site coaching (Fleisch, 2018). The analysis of the model concluded that virtual coaching's

> "*inability to see the teacher in the classroom potentially undermines the foundational pillars of observation, demonstration and feedback that form effective instructional coaching*" (EGRS Mpumalanga, 2019: 7).

Given this cost and scaling limitation of the external coaching model, the EGRP focuses on testing the efficacy of an alternative on-site coaching model which can be embedded in the public schooling system using existing education system resources and personnel. This option looks at the efficacy of the Foundation Phase Departmental Head (DH) assuming the role of the external coach in each school. It is expected that the DH coaching model would be more cost-effective and may offer more extensive coaching support to teachers as DHs will only be coaching a few teachers in their schools and are resident in the schools continually. Once the DHs are trained in the first year of the project, they will be given limited support by external professional coaches. The other schools in the experimental design will have direct support for their teachers by the professional coaches throughout the life of the project. The EGRP will thus examine the effectiveness of this in-

house DH led coaching model compared to that of the external professional coach on learner performance using early grade reading assessment (EGRA) in HL and EFAL.

The EGRP was co-developed in collaboration with several structures within the DBE, including Curriculum Policy, Support and Monitoring Directorate; Teachers, Human Resource and Institutional Development branches (Branches C and T); and the Research Coordination and Monitoring and Evaluation (RCME) Directorate.

The EGRP Implementation environment and COVID-19

EGRP was initiated in 2021 when the country was still undergoing varied levels of COVID-19 related lockdowns. This followed 2020, when schools were closed entirely for much of the year, resulting in learning gains for many Grade 1 and 2 learners being lost or compromised. The loss of learning in this period remains a significant concern in the schooling system in South Africa and globally. By 2021 many learners were still experiencing various levels of school closures due to COVID-19 (UNESCO, 2021). In South Africa, one study estimated that the majority of primary school learners lost between 70% to a full year of learning from March 2020, when the lockdowns began to June 2021. The same study estimates a decline in average school attendance reported by households with learners enrolled in Grade 1 – 12 from an estimated 97.9% in 2018 to 94.8% by April 2021 (Shephered & Mohlohlwane, 2021).

These losses in learning imply that some learners are experiencing an onset accumulation of learning backlogs that might hamper their acquisition of foundational reading skills and future learning potential. Addressing learning backlogs often requires more time than is available in the current schooling schedules (Christie, 2022). It also places pressure on teachers as they begin dealing with learners with increasingly differentiated capacities and reading abilities. School programmes to manage these backlogs will be required, and teachers would need support to assist their learners to recover these losses, especially in poor schools that could not support their learners and parents more effectively during these COVID-19 schooling disruptions. Some elements of EGRP, such as school-based workshops, will be critical in helping teachers in this regard.

At a programme delivery level, the COVID-19 regulations in place during 2021 and into early 2022 put significant constraints on the schooling system as they impacted learner and teacher attendance, teaching practices, and school management. It is expected that this might have affected the implementation fidelity of some aspects of the EGRP activities in 2021 and into 2022. External coach visits and learner and teacher attendance were negatively affected by illness, school lockdowns and rotational attendance. The teacher survey results indicate that teachers' and learners' attendance were negatively affected in 2020 and 2021. These challenges could have diminished the realisation of some outcomes related to coaching, learning, and teaching time. COVID-19 regulations negatively impacted the practicality of organising all group-based teaching methodologies, such as group-guided reading and shared reading, which are at the core of the EGRP model.

Further, as this evaluation took place while the restrictions were still in place, teachers had to observe COVID-19 protocols. So, even if teachers had the knowledge and ability to implement group activities as required by EGRP, they could not demonstrate this during the evaluation. The

details of the fidelity of implementation of the EGRP are contained in the EGRP implementation fidelity report.

This report is written as a midline evaluation since no baseline assessment was conducted in 2020 or early 2021 before the commencement of the intervention. Since the current evaluation was commissioned in late 2020 and could not commence in practice until early 2021, its conceptualisation as a baseline and endline impact study would have required a baseline measurement to be taken in early (term 1 or 2) 2021. This would, however, have been difficult to compare with a term 3 endline (2023), and a term 1 or 2 endline (2023) would have shortened the intervention period too much. It was therefore decided to conduct the initial round of learner assessments in Term 4 of 2021, by which point the learners in the two intervention arms had already received almost a full year of intervention support.

We note practical challenges experienced during the initial year of implementation, specifically in the operative coaching dimension, due to a combination of Covid 19 and other technical challenges. On this basis, we debated whether the teachers and, therefore, learners in the two coaching intervention arms were likely to have received enough inputs to expect an effect. If not, one could have called the initial learner assessment a de facto baseline. However, there was some level of intervention, and we felt it would be methodologically inappropriate to simply discount this. For this reason, we have analysed the learner assessments in the current report as a midline assessment, using the appropriate statistical techniques, as described further in the report.

2.1 Theory of Change

The EGRP Theory of Change (ToC) is adopted from the EGRS studies. At the core of this ToC is how the programme influences change at how early grade reading is taught by focusing on teaching practices, learner engagement, and content. These teaching practices require changes at the level of teaching tasks, materials used, and most notably at the level of teacher knowledge and behaviour.

Achieving this change requires understanding of how children learn to read. A vast body of knowledge demonstrates that a child who learns to read from a language that they speak at home finds it easier to read in a second language and establish a much deeper conceptual understanding (UNESCO, 1953; 2005; 2008). This is because a child learning to read in a foreign language is concurrently trying to execute two new tasks: learning to read and learning a foreign language. In comparison, a child learning to read a language frequently used at home only has to learn to read. This implies that the skills this child acquires are done at a much more conceptually profound level and can be used to further learn to read in another language.

Previous studies have confirmed that reading is a skill that any child can learn. Despite this, research, studies show significant numbers of children in schools struggling to read fluently and understanding what they are reading. The reason for this is that teachers have not received sufficient training to teach reading effectively (UNESCO, 2015). What is known about the acquisition of early grade reading ability has shifted from looking at reading acquisition as a natural process akin to when a child learns to talk. Neuroscience studies investigating live brain activity confirm that learning to read is a complex process that a child cannot achieve naturally but must

be taught systematically. While a child can learn to talk from consistently hearing conversations, the same does not apply to reading and exposure to text.

Learning to read is a technical process that is most effective when done through a systematic phonics approach rather than a whole word approach. The former requires children to learn the sound of letters and how they sound in a word and then use this to construct whole words. In contrast, the latter makes reading whole words the first step to learning how to read. Written text represents a code that a child learning to read should unscramble or decode through discovering and recalling that a particular arrangement of letters represents certain sounds. This combination of letter recognition and phonological awareness, gained through speaking and listening to others speak, results in the ability to decode text. As a child becomes proficient at decoding text through systematic teaching and practice and building up vocabulary, they progress to recognise words. As word recognition strengthens through practice and teaching, the child becomes fluent and can read a series of connected texts and make meaning out of it.

Though more complex than a whole word approach, the phonics approach to learning how to read allows a child to remember words better and gives the child the skill to decode new words more easily. This approach requires able teachers who can facilitate systematic teaching and sustain a consistent but progressive practice routine for the children to master decoding, word recognition, and reading fluency in pursuit of reading comprehension. At each stage of reading acquisition, teachers and learners should have access to appropriate teaching and learning materials.

This highly structured and systematic approach to phonics, grounded in research and practice, shaped the EGRP pathway to teaching reading. This approach requires a change in teaching practices, teaching tasks, learner engagement, content and materials to teach early grade reading. To manage this shift, the EGRP ToC has built into its design interventions support to teachers at both the cognitive and affective levels through training, coaching and using teaching materials. All these elements interact in ways that are expected to trigger a sustained change at the instructional core.

Lesson Plans and Learning Teaching and Support Materials (LTSM)

Lesson plans provide the core of a Structured Learning Programme (SLP) that teachers and coaches can follow. Integrated with LTSMs, they release the value of SLPs and provide details of the required core methodologies and teaching routines in early Grade language instruction (Fliesch, 2018). They form the foundation upon which teachers are trained, coaches give teachers feedback, demonstrate core methodologies, and conduct needs-based workshops at schools.

The SLP thus explicitly defines the expected level and quality of language instruction needed to change the instructional core. These lesson plans are broken down into daily plans providing a structured and systematic approach required to teach reading through phonetics as per Curriculum and Assessments Policy Statements (CAPS) for the Foundation Phase. Lesson plans and LTSMs also contribute to teachers feeling ready for change through feeling supported (Alsofrom, 2018). Lesson plans also afford teachers more time to spend on valuable instruction as they no longer have to do the often tedious task of drawing up lesson plans.

Teacher training

This component addresses the knowledge gap in teaching reading using phonetics, and it builds feelings of being supported through a change process that may be new and possibly intimidating to most teachers. The training covers a wide variety of early grade reading content, including effective reading comprehension strategies and methodologies, listening, speaking, phonics, writing, and guidance on using lesson plans, tablet resources, and LTSMs.

Coaching

The EGRP coaching model is premised on the understanding that coaching is more than just observation and feedback by a more experienced practitioner. Instead, it highlights the importance of demonstration, the opportunity for modelling that a teacher can learn from, and instructional coaching as a developmental process rather than an evaluative one.

This is important in ensuring that coaches are seen as supportive but critical friends to teachers as they go through a transformative but potentially difficult process of changing their teaching practice and attitudes whilst embracing new methods. As coaches train, encourage teachers to implement new strategies and lesson plans, and give feedback, the teachers are expected to feel supported and grow confident in applying more changes. They also deepen their understanding and skills (Alsofrom, 2018).

The coaching intervention will change in Year 2 of the implementation as Departmental Heads (DHs) will take over the coach's role in one cohort of 40 schools. In this cohort, on-site coaching dosage will be reduced whilst the DHs receive increased virtual support from two external professional coaches to assist them to transition into this role. This will coincide with the external coach creating and supporting a professional learning community (PLC) that the DH will facilitate within the school. In these PLCs, teachers will model their best practices to support each other and learn from each other. The role of the coach will mainly be a motivational one. It is expected that the instructional leaders and teachers in the school will grow enough interest to sustain the PLC as they begin to see its value. This approach will gradually phase out the external professional coach from the system as DHs take over their role.

All these elements (lesson plans, LTSMs, teacher training and coaching*)* are expected to improve the teachers and transform the way they teach early grade reading, leaning on a structured learning programme, leading to better reading outcomes for learners.

These reading outcomes are to be measured using a contextualised early grade reading assessment (EGRA) in HL and EFAL. This is in line with the earlier iterations of the EGRS. However, one innovation is the introduction of group administered tests alongside the more conventional one-on-one testing modality. This is potentially another cost cutting element which could lead to good practice, as group administered tests are easier to implement and can be replicated easily by teachers and integrated into their normal assessment regimen. Group administered tests have been successfully piloted by RTI (2015) and at scale in the Soma Umenye Project in Rwanda

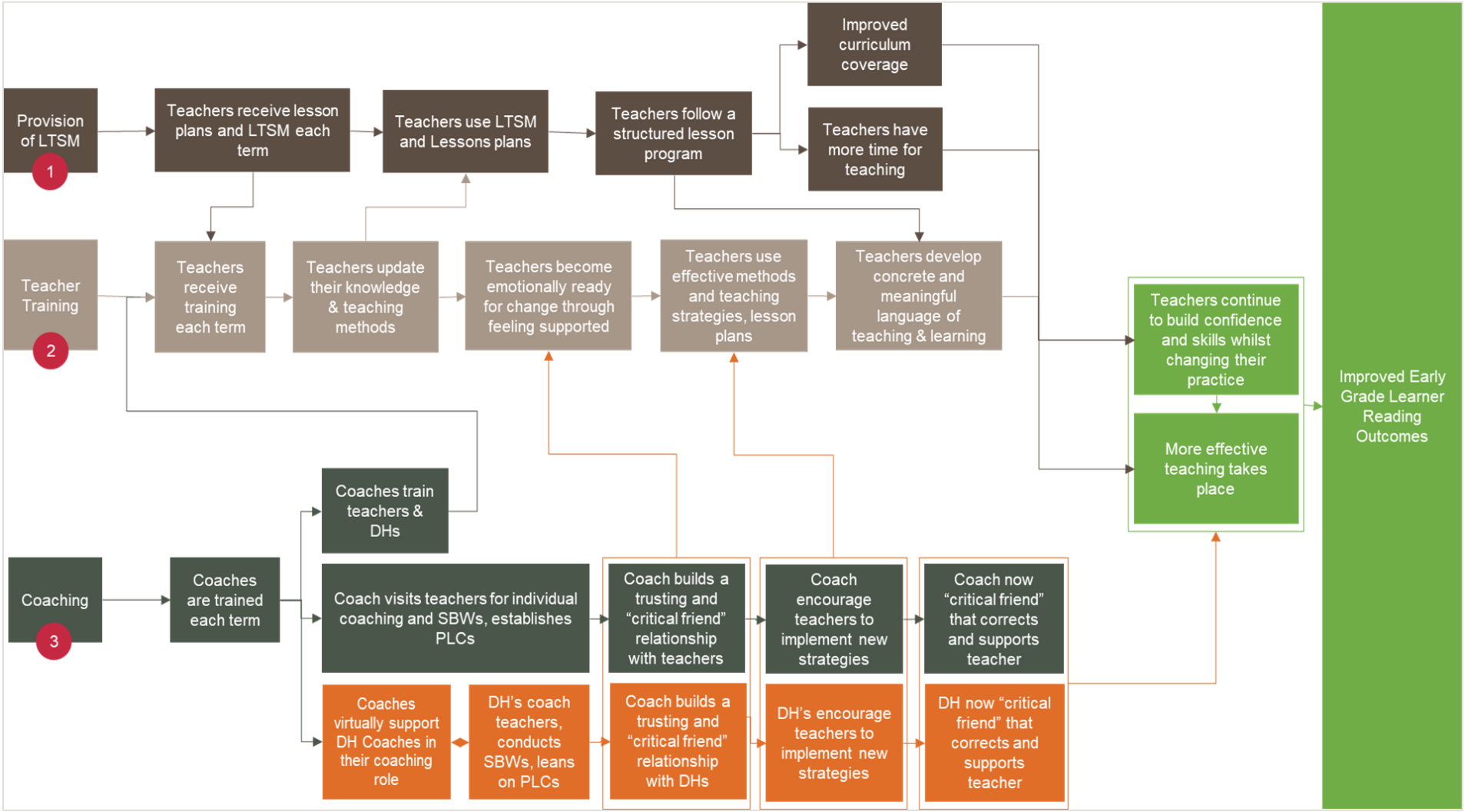

Figure 2: Programme Theory of Change

2.2 Description of Interventions

The EGRP study has three intervention arms at improving early grade reading outcomes in Setswana and EFAL. The three arms provide different combinations of three components: Learning and Teaching Support Materials (LTSM), Teacher Training, and two different modalities of Coaching support to teachers.

Control Group[2]
Teacher Training and the provision of lesson plans and LTSM (Base programme) (60 schools)

The foundation phase teachers in these schools receive Setswana (the HL of all the schools in the EGRP) and EFAL lesson plans before the beginning of each term as well as printed LTSMs. Each teacher also receives a tablet. The tablets contain lesson plans that are updated every term, videos, and other LSTMs such as assessment plans, weekly routines, trackers, and reflection trackers. The material on tablets is supported by various reading and big books – for learners – in Setswana and EFAL.

The teachers also receive teacher training before the commencement of each term over the three-year EGRP implementation period. The teacher training sessions are conducted by eight external professional coaches using a cascaded model immediately following their own training by Molteno. Teachers who miss this training are expected to receive catch-up training when their coaches visit their schools. The package of LTSM and Training provided to the teachers constitutes the base programme given to all the intervention arms. Previous EGRS interventions indicated that this package improves the teaching of reading.

Intervention 1
Base programme ***+*** *external professional coaching (40 schools).*

In addition to the base programme, the foundation phase teachers in these schools receive coaching support from external coaches for the entire duration of the EGRP intervention. The coaching component consists of on-site coaching and continuous virtual coaching between coach site visits.

On-site coaching

Teachers receive coaching support visits by an external literacy expert once a month. With each visit, the coach needs to inform the school and the teacher in advance so that the teacher prepares for the session. This allows the coach to observe what would be the best-case scenario as it is assumed the teacher would seek to give their best delivery of the planned lesson activities under observation. The coaching routine each school visit consists of the coach carrying out pre-classroom observation discussions with each teacher based on lesson plan activities for the day, followed by at least three classroom observations and post-observation discussions across the different teachers. Depending on the coach's assessment

[2] The term control group is used in relation to that this cohort of schools did not and will not receive any of the two different modalities of coaching which are being tested by the study. These schools however receive the base intervention just like other two coaching cohorts, making the intervention common across all cohorts. The learner results of the two coaching cohorts are thus referred to as intervention groups in this study and their results will be compared against the control group and against themselves

of how the coaching observation session is progressing, the coach may sometimes model and demonstrate the expected approaches and methodologies within the lesson in the classroom.

The coach is also expected to carry out needs-based workshops after school with all the foundation phase teachers to address any common issues that might have been observed. The teachers will further receive support from the PLC that the external coaches will establish in the second year of the programme.

Previous EGRS studies found the coaching intervention to be particularly effective in improving learner reading outcomes in both Setswana and EFAL.

Virtual coaching

This intervention arm also receives weekly virtual coaching, mainly through the WhatsApp social media platform. Through this platform, the coach sends a message to the teachers' WhatsApp group at the start of every academic week, reminding them of the key learnings in the lesson plan for that week and sending video clips and other tools to assist the teachers with the week ahead.

Once a month, the coach also initiates a brief discussion forum on the group, focusing on general matters about literacy teaching, core methodologies, and the delivery of the EGRP programme content. Once a term, the coach further engages with each teacher in a short one-on-one virtual coaching session. This coaching session will often include:

- Establishing the teacher's level of proficiency and engagement through data analysis and questioning
- Praising the teacher for evident strengths and efforts
- Identifying challenges and addressing them
- Sending short video or voice clips to address challenges
- Documenting the content of the coaching session.

The schools in this intervention arm are comparator schools as any improvements in their learners' reading g performance would be compared to learners' reading performance in the third group of schools.

Intervention 2
Base programme **+** *external coaching in Year 1* **+** *DH coaching in Year 2 and 3 (40 schools).*

Teachers in this intervention arm schools receive the same intervention as intervention 1 teachers in Year 1. However, in years 2 and 3 (2022 and 2023), teachers at these schools will receive coaching from their Departmental Head (DH), who will take on the role of in-person coach to their colleagues. To support the DHs in their role as coaches in Year 2, two external professional coaches will each support 20 schools in this cohort virtually and in person. In Year 1 of implementation, the DHs would have also been exposed to the training that all the teachers attend. The learner reading outcomes from this intervention arm will be compared against the outcomes from schools in the Intervention 1 group of schools to determine the efficacy and value for money of using the DHs as coaches.

Table 1: Description of intervention arms and support offered

Activities	Base programme (control)	Intervention arm 1: External coaching	Intervention arm 2: DH Coaching
Provision of lesson plans	Tablet-based HL and EFAL scripted plans	Tablet-based HL and EFAL scripted plans	Tablet-based HL and EFAL scripted plans
Provision of LTSM	Paper-based · HL and EFAL Big Books · HL and EFAL Sight Words Flash Cards · HL and EFAL Discussion Posters · HL and EFAL Phonic Friezes	Paper-based · HL and EFAL Big Books · HL and EFAL Sight Words Flash Cards · HL and EFAL Discussion Posters · HL and EFAL Phonic Friezes	Paper-based · HL and EFAL Big Books · HL and EFAL Sight Words Flash Cards · HL and EFAL Discussion Posters · HL and EFAL Phonic Friezes
Training of teachers	Teachers receive training at the beginning of each term throughout the intervention	Teachers receive training at the beginning of each term throughout the intervention	Teachers receive training at the beginning of each term throughout the intervention
Coaching of teachers	None	External Coach visits each teacher in the classroom from Year 1 to Year 3	External Coach visits teacher in the classroom in year 1
Coaching of DHs	None	None	2 External coaches coach DHs remotely at a ratio of 1 coach to 10 schools
Coaching of Teachers by DHs	None	None	Teachers coached by DHs in Years 2 and 3
Virtual coaching support	None	Ongoing remote support by external coach from year 1 to Year 3	None
Establishing PLCs	None	Established in year 2 and implemented throughout	Established in year 2 and implemented throughout
School-based workshops	None	Carried out with each coach visit	Carried out with each coach visit

2.3 Research site

The EGRP is being implemented in the three sub-districts of Kagisano Molopo, Greater Taung, and Naledi in Dr Ruth Segomotsi Mompati (DR RSM) District in the North West Province of South Africa. The rationale for selecting this remote district in the North West Province is that the region has a relatively uniform home language and HL instruction in Setswana, making it affordable to produce learning and teaching support materials for the project. This district is also poor and houses some of the most underperforming schools in South Africa, while often overlooked by state and NGO interventions. The provincial department of education (PDE) officials have been supportive of this intervention and of past studies and interventions that have been carried out. This buy-in from the PDE is vital as the PDE has committed time, access and resources to facilitate various programme interventions such as training and access to schools, teachers and learners.

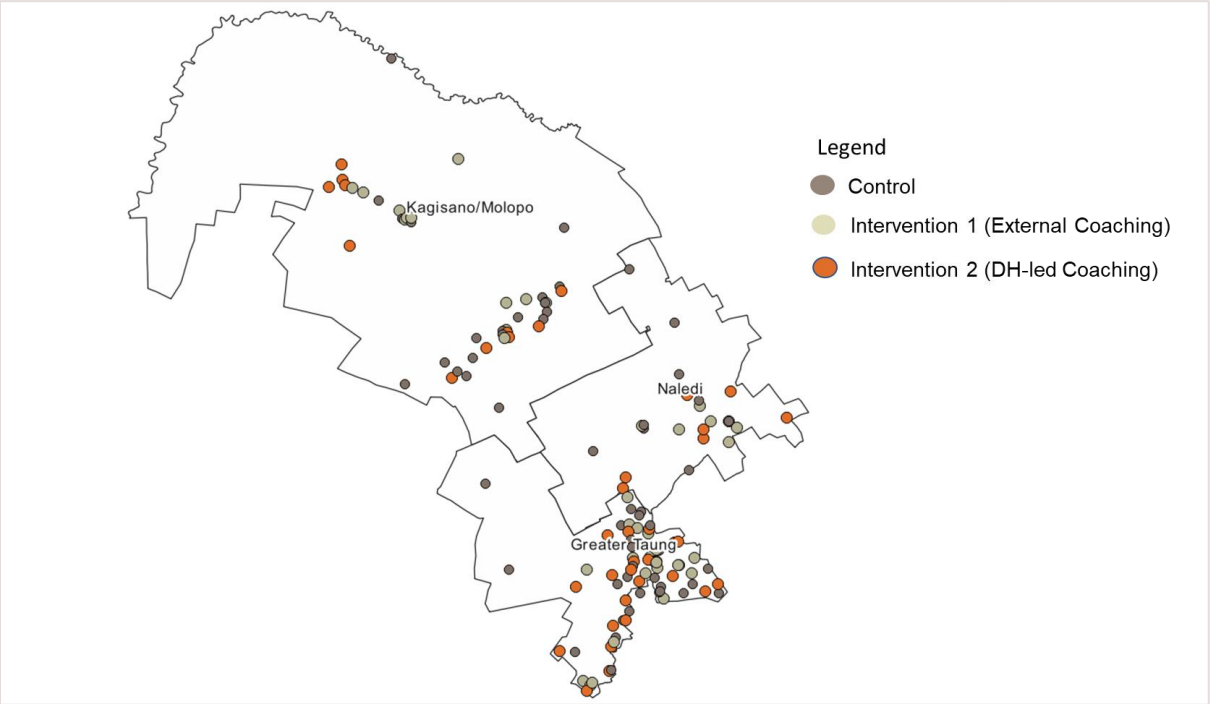

Figure 3: Distribution of sampled schools across three sub-districts shown by intervention arm

Table 2: Number of sampled schools

Subdistrict	Number of schools	Base programme (control)	Intervention arm 1: External coaching	Intervention arm 2: DH coaching
Greater Taung	73	30	21	22
Kagisano	43	20	12	11
Naledi	24	10	7	7
Grand Total	**140**	**60**	**40**	**40**

3 Evaluation Design

3.1 Scope of the Evaluation

The evaluation measures both implementation fidelity and the impact of the EGRP interventions on learner reading outcomes. This report covers impact measurement in the form of the initial midline observations, with the endline report coming out in early 2024. Implementation fidelity is captured in a standalone report. It covers dosage, adherence, quality of delivery, participant responsiveness, and programme differentiation fidelity aspects after the first year of implementation.

Impact measurement focuses on the effect of the intervention on learner performance assessments in Setswana (HL) and English First Additional Language (EFAL). The evaluation also uses a model to predict the role of contextual factors as determinants of learner reading performance. This will shed light on their possible role in contributing towards differential intervention impacts, if any are observed. The challenges of coaching delivery through Foundation Phase Departmental Heads are also explored.

Impact measurement will thus consist of:

- Midline measurement of learner performance in HL and EFAL across all three intervention arms after one year of intervention
- Endline measurement of learner performance in HL and EFAL across all three intervention arms after three years of intervention implementation.

The comparison of these will establish the impact of:

- DH coaching on EFAL & HL learning outcomes
- External coaching on EFAL & HL learning outcomes
- Base programme on EFAL & HL learning outcomes.

Evaluation Questions

The evaluation addresses the following questions and their subsequent sub-questions as indicated in Table 3. which includes reports where they are addressed.

Table 3: Evaluation Questions

Evaluation Question	***Midline Report 2021***	***Fidelity Report 2021***	***Case Study 2021/ 2023***	***Endline Report 2024***
1. Did the programme result in the specified outcomes and impacts?				
How successful are DHs as coaches compared to the external professional coaches?				✓
How successful are DHs as coaches compared to the external professional coaches?				✓
What has been the impact of each of the two coaching interventions on learning outcomes for Setswana and EFAL at the end of the third year?				✓

Evaluation Question	*Midline Report 2021*	*Fidelity Report 2021*	*Case Study 2021/ 2023*	*Endline Report 2024*
Which intervention-external factors influence learner reading performance and lead to differential intervention impacts?	✓			✓
To what extent do teaching practices in the classroom reflect the practices conveyed by teacher training and coaching?		✓	✓	✓
2. Is the programme implemented as intended?				
Are lesson plans and other LTSMs being used by teachers?		✓	✓	✓
What dosage of termly training have teachers received?		✓		✓
What is the dosage, nature and quality of the coaching provided to teachers by the DH?		✓		✓
What is the dosage, nature and quality of coaching provided to teachers by the external professional coaches?		✓		✓
3. What are the implementation challenges of delivering coaching through DHs?				
What are the DHs' time and capacity constraints		✓		✓
How well have PLCs been functioning and why (enabling conditions)?		✓		✓

3.2 Intervention assignment and sample selection

Similar to previous EGRS studies, this study is also designed as a randomised controlled trial (RCT). It is made up of three program arms receiving differentiated components.

The DBE selected the sample from an eligible pool of schools through a series of exclusions and elimination. Based on information received from the district officials, the elimination process excluded schools too small to choose 20 learners in each of Grades 1 and 2. These schools would have had to do multi-grade teaching, making it impractical to implement a structured learning programme through grade-specific scripted lesson plans. Similarly, to safeguard against escalating project implementation costs, large schools with large enrolments were excluded from the sample. All schools falling within the quintiles 4 and 5 were also excluded, leaving the sample with only non-fee-paying schools. Schools that did not have Setswana as a language of instruction were also excluded. Schools that were used for pilot purposes or that were randomly selected as replacement schools were also excluded. This exclusion led to a reduced sampling frame of 140 schools, out of 291 state primary schools in the district.

The next stage stratified these schools into 10 strata, each composed of 14 schools of similar characteristics. This process arranged the 140 schools by sub-district, by whether the schools participated in the School Reading Improvement Programme (PSRIP), and by the level of DH commitment (proxied by whether the school's Foundation Phase DH participated in the first training workshops that were offered to all 140 schools). This stratification allows for valid and

balanced estimation of sub-group effects. It is thus possible, for example, to extract a reliable estimate of the programme effect on schools that participated in the PSRIP or those that had committed DHs. Table 4 shows the assignment per strata of the percentage of schools that were part of the PSRIP, and Table 5 shows the percentage of schools with committed DHs.

Table 4: Percentage of schools in PSRIP shown by strata

Stratum	Percentage of schools in PSRIP	Number of schools in Stratum
1	0%	14
2	64%	14
3	100%	14
4	100%	14
5	100%	14
6	7%	14
7	7%	14
8	0%	14
9	64%	14
10	100%	14

Table 5: Percent of schools whose DH attended training shown by strata

Stratum	Percentage of schools whose DH attended training	Number of schools in Stratum
1	43%	14
2	36%	14
3	0%	14
4	86%	14
5	100%	14
6	7%	14
7	0%	14
8	93%	14
9	86%	14
10	100%	14

After the stratification was completed, schools within each stratum were randomly assigned using a random number generator to each program arm, such that each stratum had 4 schools assigned to intervention 2, 4 schools assigned to intervention 1, and 6 schools assigned to the base program (control group).

3.3 Instrument development and piloting

SSA and the DBE co-created the EGRA learner assessment instruments for both EFAL and HL by adopting and adapting assessment tools used in previous studies conducted by the DBE. These include the Early Grade Reading Studies (EGRS I, EGRS II) and the RSP. Some aspects of these tools were new and had to be designed afresh. The same approach was also taken for the context tools. This was favoured because it allowed for a high degree of continuity with previous studies. The instruments design process was carried out from September to October 2021, and all instruments were signed off by the DBE before piloting and primary data collection.

For Grade 1 learners, five one-on-one HL individual tests were designed: Rapid Object Naming, Letter Sound Knowledge, Oral Reading Fluency, Word Recognition, and Reading Comprehension, including one group-administered subtest EFAL Word Sound Knowledge.

Likewise, for Grade 2 learners, ten sub-tasks consisting of seven one-on-one administered sub-tasks and three group tests were designed and piloted for EFAL and HL. For HL, five one-on-one sub-tasks (Rapid Object Naming, Letter Sound Knowledge, Oral Reading Fluency, Reading Comprehension, Word Recognition) and two group-administered sub-tasks (Complex Consonant Sound Knowledge and Word Recognition Writing) were designed and piloted. For EFAL, two one-on-one administered sub-tasks consisting of Oral Reading Fluency and Reading Comprehension, and one group-administered task, Word Sound Knowledge, were also designed and piloted.

The context tools co-developed with DBE include the Teacher Survey, Departmental Head Survey, Principal Survey, School datasheet, and the School Environment/ Functionality Tool.

A team of 10 field researchers were trained over five days to collect pilot data for learners in Grades 1 and 2 and context data in five schools in the Dr RSM District in the North West Province. Data collection in the sixth school was cancelled due to an outbreak of an upper respiratory tract infection amongst learners in the school, resulting in over 90% of all learners not attending school that week. After this pilot, an analysis of the results indicated that some sub-tests had floor effects. These subtests were redesigned and re-piloted in another school in Rustenburg.

3.4 Data collection

3.4.1 Recruitment and training

SSA ran an extensive recruitment process that involved analysing its internal database of trained assessors who had administered EGRA on previous studies and launching an aggressive recruitment drive for suitable candidates. This process yielded over 180 eligible applications, with 65 shortlisted based on the following criteria:

- Language - fluency in both Setswana and English
- Gender – with a preference for females as most foundation phase school environments are predominantly female
- Tertiary qualification – a tertiary qualification (preferably an undergraduate degree in social sciences/arts at a minimum)

- Data collection experience – some data collection or capturing experience (preferably working with Social Surveys Africa)
- Experience in working in schools – preference for candidates who have worked in a foundation phase school environment
- Experience working with children – preference for candidates with experience working with children under the age of 10

This process resulted in 55 field researchers being recruited for training, with 10 selected for additional training as field researcher supervisors and for piloting instruments.

The training was co-designed and co-presented by DBE and SSA's core team of early learning and literacy specialists, experienced field leads, researchers and data managers. It was broken down into two five-day phases. The first phase, carried out in October 2021, focused on ten field researchers who were to be used in the pilot and engaged as field researcher supervisors in the main data collection. After analysing pilot results and finalising instruments, the second training phase was carried out in November 2021. This retrained the field researchers' supervisors and a larger pool of the remaining recruits.

The central focus of the training was on how to administer and accurately capture learner EGRA for the different subtests using Tangerine. It also included field protocols and sampling processes. Field team supervisors received additional training on tracking sample completion, communicating with school Principals, managing the field teams, and reporting any concerns to the SSA office.

All field researchers who attended the training were rigorously tested on their comprehension of the study objectives, assessment and Tangerine rules, ability to administer tests to individual learners and groups of learners, and their comportment and engagement throughout training. The critical testing, undertaken at regular intervals through both training processes, was Inter-rater Reliability (IRR) tests. IRR uses a range of indicators that are captured on Tangerine to assess how well each assessor performs in relation to the group during practice rounds of learner assessment. Outliers in speed, accuracy and starting and ending the subtest are easy to identify. Participants who were consistent outliers were removed from the team to collect EGRA data. Only those who passed all the tests were selected to be part of the final team, while those who qualified but performed slightly lower than the others were placed as backup team members.

3.4.2 Sample Achievement Limitations

The sampling of learners in schools included a random selection of all learners in Grade 1 and Grade 2 in each school. In principle, this meant that in schools with more than two classes per grade, more than two teachers would have had to be interviewed so that all learners would be matched to their teachers. Though ideal, this was not practical given the time limit field researchers had to carry out learner assessments, Principal surveys and collect school-level environmental data. It also would have escalated costs significantly as data collection teams would have been expected to spend more than one day in some of the larger schools. The result of this is that 71% of learners were matched to interviewed teachers. Absent teachers on the days of the survey also contributed to this gap in matching.

School rotation related to the COVD-19 pandemic affected the number of learners and, in some cases, teachers who were present at the school at the time data collection teams visited. Though SSA teams made prior arrangements and gave reminders of the teams' scheduled visits, some schools still could not have at least 20 learners per grade available. This affected 24 schools. Table 6 shows the number of learners successfully tested per school.

Table 6: Number of learners successfully tested per school

Number of learners	No. of schools with the number of Grade 1 learners	No. of schools with the number of Grade 2 learners
25	3	2
24	3	4
23	13	4
22	84	72
21	11	24
20	10	13
19	3	4
18	0	1
17	2	2
16	4	2
15	0	1
14	2	3
13	0	2
12	0	0
11	1	1
8	0	1
7	0	1
6	1	0
5	1	0
4	0	1
0	1	1
Total	**139**	**139**

Midline data was collected from a total of 139 schools instead of 140 schools. The school that was not included was a misallocated multi-grade school that had not received any interventions. No replacement school was provided to SSA during data collection. Of the 139 schools where midline data was collected, one school had no foundation phase learners present due to a learner rotation system which alternates foundation phase and intersen attendance on a two-week interval. The data collection period fell within the two weeks when foundation phase learners were not at school. Surveys with the Principal and teachers as well as the school data sheet and school environment checklist were completed.

In a few cases where the Principal was not available, the deputy, acting Principal, or senior teacher was surveyed, and where alternates were not available, contact details of the Principal were collected, and the survey was administered telephonically. Since most DHs are also foundation phase teachers, they were often surveyed for both roles. There were cases of teachers who were absent and thus could not be surveyed. In smaller schools with only one Grade 1 and one Grade 2 class, if a teacher was absent on the day of the survey, the full sample of teachers could not be interviewed.

3.4.3 Fieldwork

SSA teams of field researchers collected midline data over three weeks in November 2021. A total of 50 assessors, grouped into teams of 5 field researchers (10 teams in total), was deployed to randomly select and test 40 learners per school (20 in Grade 1 and 20 in Grade 2), interview the Principal or school management team (SMT) member, the DH, and at least two foundation phase teachers per school. Each team had a lead field supervisor who facilitated all communications with schools and the SSA office, reported daily completion rates on real-time digital tools, managed ongoing learner assessment quality control and administered context tools. Learner home background questionnaires were sent to schools to send to the parents or caregivers several weeks in advance and were collected from the school on the assessment day. These were given to all learners in Grades 1 and 2 to fill in.

The midline data were collected digitally by trained field researchers. Learner assessments were administered using Tangerine, whilst all the other context tools were collected through KoboCollect.

Table 7: Number of assessments and surveys conducted

			Grade 1 learners assessed	Grade 2 learners assessed	Home background questionnaire	Teacher survey	DH survey	Principal survey	School environment tool	School datasheet
Program Arm	Control	Intended number	1180	1180	2360	118	59	59	59	59
		Number achieved	1222	1208	1153	162	59	55	57	54
		%	**100%**	**100%**	**49%**	**100%**	**100%**	**93%**	**97%**	**92%**
	External Coaching	Intended number	800	800	1600	80	40	40	40	40
		Number achieved	851	829	760	111	40	37	40	38
		%	**100%**	**100%**	**48%**	**100%**	**100%**	**93%**	**100%**	**95%**
	DH Coaching	Intended number	800	800	1600	80	40	40	40	40
		Number achieved	842	860	802	99	37	37	40	37
		%	**100%**	**100%**	**50%**	**100**	**93%**	**93%**	**100%**	**93%**
	Total Intended number		2780	2780	5560	278	139	139	139	139
	Total achieved		2915	2897	2716	372	136	129	137	129
	Percentage achieved		100%	100%	49%	100%	98%	93%	99%	93%

Table 7 shows the number of Grade 1 and Grade 2 learners who participated in both the group and individual assessments. For the learner midline assessments, where possible more learners were assessed than the intended number to ensure full sample coverage and increase the possibility of linking learners' individual and group assessments.

In aggregate, per program arm, the full intended number of learner assessments were reached across all the program arms. This was, however, due to oversampling of learners in some schools. There were 14 schools where less than 20 Grade 1 learners were assessed, 18 schools where less than 20 Grade 2 learners were assessed, and 1 school where no learner assessments were conducted. See the Sample Coverage Limitations section for a detailed breakdown of schools where learners were either under or oversampled. Under sampling was mainly due to small enrolment sizes, Covid-19 related attendance rotations, and early dismissal of foundation phase learners.

Overall, 49% of the expected Home Background questionnaires were returned by the day of the school visit. This low sample coverage was overcome by including a short set of home background questions in the individual learner assessment tool to ensure that this did not compromise having basic background information about all learners. A methodology lesson from the midline may be that the background questions asked of learners are sufficient and that the parent questionnaire may not be necessary in the endline iteration.

Many Foundation Phase teachers are also Foundation Phase Departmental Heads in their schools; therefore, they were asked the same teacher-specific questions on teaching practices, LTSM use, EGRP coaching, training, and overall experience in the programme in addition to DH- specific questions. The Foundation Phase DHs that are not foundation phase teachers were only represented in the DH survey.

Full sample coverage was not reached for the DH survey because some DHs or senior staff members were not present on the day of the visit, and telephonic appointments were not honoured. Similar challenges were experienced with the Principal Survey. Two schools had recently lost their Principals and had no Deputy Principal as a result school leadership interviews could not be conducted in those schools. The School Environment scan could not be conducted for two schools because of time constraints caused by negotiating access to learners with class teachers and members of the SMT. In these two schools, the learner assessments and teacher interviews were prioritised. For the school datasheet, some schools did not have the necessary information on record and could not complete the datasheet. In other schools, the administrator was unavailable on the day of the visit, these forms were emailed to them for completion.

All field researchers were given a hard copy linking form, which required that they write down the Tangerine assigned unique ID, along with the SSA unique ID (which was constructed using the assigned school ID, grade of the learner and a number from 1 to 25). The linking form also required that field researchers write the learner, teacher and school name for tracking purposes should the incorrect ID be captured. The linking form was used to clean the data received in cases where the same SSA or Tangerine ID was assigned more than once on the individual assessment, group assessment or home background survey datasets. The linking form also enabled the linking of learners to surveyed teachers.

A total of 2,915 (100%) Grade 1 individual assessments were matched to their respective group assessments and were accepted after the quality control process was implemented. Similarly, 2,897 (99%) matched Grade 2 learner assessments were accepted. Additionally, teacher questionnaires were linked to learners through the linking forms.

All matched learner assessments collected from schools were used in the midline data analysis to create an attrition buffer for the 2023 endline data collection.

A summary of the individual assessments matched to group assessments and the home background survey and teacher survey are shown in Tables 8, 9, and 10 below.

Table 8: Number of learner assessments conducted

		No. of learners	No. of individual assessments matched to group assessments	Percentage
Grade 1	Control	1222	1222	100%
	External Coaching	851	851	100%
	DH coaching	842	842	100%
	Total	**2915**	**2915**	**100%**
Grade 2	Control	1208	1208	100%
	External Coaching	829	829	100%
	DH coaching	863	860	99%
	Total	**2900**	**2897**	**100%**

Table 9: Number of learner consent and background forms received

		No. of learners	No. of home background forms matched to learner assessments	Percentage
Grade 1	Control	604	551	91%
	External Coaching	401	366	91%
	DH coaching	387	372	96%
	Total	**1392**	**1289**	**93%**
Grade 2	Control	549	501	91%
	External Coaching	359	318	89%
	DH coaching	415	369	89%
	Total	**1323**	**1188**	**90%**

Table 10: Number of teachers interviews conducted

	No. of teacher surveys conducted (all grades)	No. of Grade 1 and 2 teacher surveys conducted	No. of teacher surveys matched to learner assessments	Percentage
Control	168	155	129	77%
External Coaching	112	107	95	85%
DH coaching	103	93	76	74%
Total	**383**	**355**	**300**	**79%**

3.4.4 Data cleaning

Data cleaning was performed during the second week of fieldwork at the end of data collection. Quality control checks were performed in the R software package on a rolling basis to identify data points that needed cleaning. Cleaned data points are limited to the following variables:

- School name (schools swapped between two teams and the wrong school had been selected on the drop-down list)
- Team number
- Teacher name (spelling)
- Learner name (spelling)
- Learner unique ID
- Last attempted for auto-stopped questions (despite this being highlighted in the assessor training, there were cases where the assessor marked the last attempted item as the last item instead of the fifth item in instances where the first five attempts were incorrect)

No other significant data quality issues were identified.

4 Midline Findings

This section highlights all the findings from the midline data collection process. The first subsection describes all the characteristics of the sample from school-level characteristics through to learner-level characteristics. The section homes in on learner performance in the midline assessment composite index analyses as well as explains balance tests conducted.

4.1 Sample characteristics

4.1.1 School-level characteristics

The sample of schools selected to receive different programme variations was explicitly targeted at lower quintile or non-fee-paying schools, which are the country's most poorly serviced and under-resourced schools (White and Dyk, 2019). As seen in Table 11, nearly half of the schools sampled are Quintile 1 schools, whilst approximately a quarter is in each of Quintiles 2 and 3. The majority (76%) of these schools are in rural areas, and only a quarter of the schools are in urban/ peri-urban areas.

The Principal and at least one Departmental Head (DH) were surveyed at each school. Most schools had someone appointed as Principal or acting Principal on the day of the school visit; only two schools did not have someone in that capacity. The majority of schools (92%) also had a Departmental Head.

Both teacher and DH surveys enquired about the existence of a Professional Learning Community (PLC) in the school and how often the PLC meets. An adapted definition of PLCs outlined by the Integrated Strategic Planning Framework for Teacher Education and Development (ISPFTED) was used when engaging teachers to establish a common understanding of what a PLC is since they are known by a range of names. The educators' responses show that 58% (n = 82) of schools have a PLC, and of those schools, 61% (n = 50) meet at least once a term. The remaining 39% (n = 32) of school PLCs meet as and when needed. The educators reported that 28% of the PLCs for the control and 24% of DH coaching schools were specialised literacy PLCs, whilst only 13% of external coaching schools have literacy PLCs.

We posit that the existing PLCs continue meeting because teachers found them helpful in improving their classroom practices and carrying out their work, as indicated from the teacher's survey results. The most prominent type of PLCs found are school-level grade and phase meetings where teachers share new strategies they have implemented and analyse and solve problems experienced in the classroom. The responses given by teachers regarding the sustainability of these structures indicated that these structures operated independently of the coaches and existed before the on-site coaching intervention. Though cluster-level PLCs involving several proximate schools were not common, data from the EGRP Case Study carried out in 2021 indicates that where these existed, they were teacher-initiated and very effective for sharing information and resources between schools – especially in rural areas. In one case, the teachers cited that due to the villages' distance away from the main towns where training and workshops are conducted, the schools organised so that one or two schools attend and then share information with the rest. Interviews with EGRP external coaches in 2021 and early 2022 revealed that some EGRP coaches were yet to set up PLCs at a cluster level.

A school environmental checklist was used to establish the extent to which schools were conducive to teaching and learning. This tool was administered by a field researcher who was guided around the school by someone who knows the school environment and available facilities well. During this school observation/ walkabout information about the presence and functionality of school facilities (such as toilets, libraries, technology), availability of LTSM, and observed school management practices and protocols, amongst other themes, was collected. It was observed that 31% (n = 42) of schools have a functional school library. Most schools either did not have a designated library or had libraries which were merely textbook storage rooms which were inaccessible to learners. In some schools, the library doubled as a staff room or administrators' office and is inaccessible to learners. The 41 schools with libraries characterised as functional are schools with libraries with novels, readers and other types of books not typically used in the classroom where learners can read or borrow books to take with them. These libraries were also accessible to learners. On the other hand, reading corners were observed in most foundation phase classrooms in 73% of the schools (n = 99). Interestingly, a slightly greater proportion of schools receiving the external coaching intervention arm had reading corners in foundation phase classrooms than those in the control and DH coaching group.

The majority of classrooms (over 90%) have reading materials and teaching aids visible and available. Inconsistencies were, however, noted in that whilst EGRP materials were displayed on most classroom walls, this was not the case for all of them. This observation was most prevalent in DH coaching intervention schools, where 13% did not have the teaching aids displayed on classroom walls as expected.

Most schools (97%) followed the timetable, with learners observed to be in class when required. However, this was not always the case for a small number of DH coaching schools (8%, n = 3) which had most learners still roaming around or playing when teaching was meant to be in progress.

The school-level sample characteristic was further disaggregated into the three intervention arms. The sample looked fairly balanced across the different school characteristics such as urbanity, school support & governance, school environmental factors and to some extent school quintile. With regards to school quintile, although the control group seems to disproportionately have more quintile 1 schools, however socio-economic conditions in quintile 1-3 are usually found to be similar.

Table 11: School-level sample characteristics by intervention arm

	Total number of schools	% of schools	Intervention arm		
			Control	External coaching	DH coaching
School Quintile					
Quintile 1	67	48%	58%	40%	43%
Quintile 2	37	27%	24%	25%	33%
Quintile 3	35	25%	19%	35%	25%
Urbanity					
Rural	105	76%	75%	72%	80%
Urban	34	24%	25%	28%	20%
School support and governance					
PLC presence	92	67%	66%	68%	68%
DH presence	128**	92%	95%	90%	90%
Principal presence	129**	100%	100%	100%	100%
School environment					
School library availability	42	31%	29%	26%	38%
Early grade reading corner availability in classrooms	99	73%	68%	77%	75%
EGRP material availability in classrooms	132	97%	95%	97%	100%
Teaching aids availability in classrooms	125	92%	93%	95%	87%
Class times are structured	132	97%	98%	100%	92%
Grounds and buildings well cared for	90	66%	64%	67%	69%
Learner toilets clean and well cared for	92	68%	76%	69%	54%
	126	93%	95%	95%	87%

Note: Asterisk (**) the reported figures not from all 139 schools, but rather on the available observations

4.1.2 Learner characteristics

Learner characteristics were obtained in two ways. Firstly, every learner who participated in the individual assessment was asked context questions such as their attendance of early

childhood development facilities (ECD), if they receive homework support at home, if someone reads to them at home, and other specific details. Secondly, a paper-based self-completed questionnaire was shared with all caregivers, allowing for the home background of each learner to be understood in greater detail. This dual process sought to manage the anticipated risk of low return rates of filled out home background questionnaires.

Table 12: Learner self-reported sample characteristics by intervention arm

	Total number of learners		Percentage of learners		Intervention arm					
					Control		External coaching		DH coaching	
	Gr 1	Gr 2	Gr 1	Gr 2	Gr 1	Gr 2	Gr 1	Gr 2	Gr 1	Gr 2
Learner age										
Younger than 5	3	1	0%	0%	0%	0%	0%	0%	0%	0%
5 years old	13	1	0%	0%	1%	0%	0%	0%	0%	0%
6 years old	1330	21	46%	1%	43%	0%	49%	1%	46%	1%
7 years old	1365	1116	47%	39%	48%	38%	45%	42%	46%	35%
8 years old	166	1413	6%	49%	6%	50%	6%	45%	6%	50%
Older than 8	38	345	1%	12%	2%	11%	12%	12%	1%	13%
Gender										
Boys	1414	1388	49%	48%	49%	47%	48%	49%	48%	49%
Girls	1501	1509	51%	52%	51%	53%	52%	51%	52%	51%
Attendance of Grade R/ ECD										
ECD/Creche	1539	1619	53%	56%	52%	51%	49%	59%	57%	60%
Grade R	1195	1137	41%	39%	42%	44%	43%	36%	37%	36%
Neither	181	141	6%	5%	5%	5%	7%	5%	6%	4%
Support with homework at home										
Yes	2750	2789	94%	96%	94%	96%	95%	97%	94%	97%
No	165	108	6%	4%	6%	4%	5%	3%	6%	3%

Table 12 (above) reflects the learner-reported background information because only 49% of paper-based questionnaires were returned to schools. This table shows that most Grade 1 learners are between six and seven years old (93%), whilst the other 7% of learners were over the age of seven. Most Grade 2 learners are between seven and eight years (88%) by the end of the school year, while the rest are older than eight years. This is the expected age distribution across the two grades in accordance with the South African Schools Act which requires children to be in Grade 1 by the time they turn seven years old.

The majority of learners reported having someone in the household who assists them with their schoolwork (over 94%) and who reads to them at home. These details mirror data

collected from caregivers, albeit from only 49% of questionnaires received back. A slightly greater proportion of Grade 2 learners reported receiving support at home than their Grade 1 counterparts.

Most learners were exposed to early childhood development education before their enrolment into the school, with the most common form being creche or an ECD programme that does not necessarily follow the Grade R curriculum.

The gender distribution of the learners assessed is relatively evenly split between boys and girls, with slightly more girls than boys in both grades.

4.1.3 Caregiver characteristics

Caregiver characteristics were collected from home background questionnaires that were returned to schools. As expected, this revealed that the primary caregivers of children tend to be female (mothers 64%; and grandmothers 20%), and this was similar across all intervention arms as reported in Table 13.

The caregivers' age profile shows that only 9% of learners are from households where the primary caregiver is younger than 25 years old, while the most prominent age profile is 25 - 35 years of age (33%), followed closely by the middle age group of 35 - 44 years (27%). Generally, the age distribution of primary caregivers is similar across the three intervention arms.

Table 13: Caregiver sample characteristics by intervention arm

	Total number of Learners	Percentage of Learners	Intervention arm		
			Control	External coaching	DH coaching
Main caregiver					
Mother	1698	61%	61%	61%	62%
Father	58	2%	2%	2%	2%
Aunt/uncle	280	10%	14%	9%	11%
Grandmother	567	21%	20%	22%	20%
Grandfather	40	1%	1%	3%	1%
Other	81	3%	3%	2%	3%
Caregiver age					
Younger than 25	222	9%	11%	6%	8%
25 - 34 years old	832	33%	33%	33%	35%
35 - 44 years old	688	28%	27%	28%	29%
45 - 54 years old	379	15%	15%	17%	15%
55 years +	369	15%	15%	16%	13%
Caregiver education					
No formal education	180	7%	7%	9%	6%
Didn't finish matric	1518	61%	62%	55%	66%
Matric	675	27%	27%	30%	25%

	Total number of Learners	Percentage of Learners	Intervention arm		
			Control	External coaching	DH coaching
Undergraduate degree	51	2%	2%	3%	2%
Post - graduate qualification	55	2%	2%	4%	1%
Other	5	0%	0%	0%	0%
Caregiver employment status					
Working for pay	473	18%	18%	16%	21%
Not working for pay	2122	82%	82%	84%	79%
English language use are home					
I hardly understand English	629	25%	27%	25%	24%
I can speak a little	1048	42%	42%	38%	47%
I read and write in English fluently	805	32%	32%	37%	30%
Support learners with homework					
Yes	2561	94%	95%	94%	94%
No	154	6%	5%	6%	6%
No. of books at home					
No books	463	19%	17%	19%	20%
1-5 books	1447	58%	58%	60%	57%
6-10 books	384	15%	17%	13%	15%
11-25 books	105	4%	4%	5%	4%
More than 25 books	83	3%	3%	3%	3%
LSM Score					
1 - 2	910	37%	36%	33%	41%
3 - 4	772	31%	29%	36%	30%
5 - 6	639	26%	28%	23%	25%
7 - 8	96	4%	5%	4%	2%
9 - 10	64	3%	3%	3%	1%

Nearly two-thirds of the primary caregivers (68%) either did not complete matric and in 7% of cases had no formal education. This might have a bearing on the support these caregivers can give learners at home. A greater proportion of caregivers in the DH coaching intervention arm had not completed matric than any other intervention arm. Overall, only 2% of caregivers have pursued tertiary education. More parents in the DH coaching and control groups have not completed matric in relation to their counterparts in the external coaching group.

Caregivers were also asked about the presence of books within the home as a proxy for existing reading culture within the household[3]. Although a large proportion (77%) of the sample reported having at most 5 books in their home, 91'% of caregivers reported that they read to

[3] Note that this does not tell of the presence of early grade friendly books within the home

their children more than once a week. This, coupled with the fact that only 32% of the learners' caregivers report being able to read and write in English, may indicate that the homework support the children receive is in HL and so while possibly building an active reading culture, it will not build learner knowledge of the English language.

The caregivers were also asked about their employment status, and multiple questions sought to establish the socio-economic conditions of the learners in their homes. The caregivers reported that only 18% of learners come from homes where the primary caregiver is employed and working for pay through formal, informal, or self-employment. The remaining 82% of learners come from a home where their primary caregiver is unemployed.

An LSM index was allocated to each household based on the ownership of a basket of goods (such as a television, smartphones, fridge, washing machine and other to establish a proxy for their living conditions. As with other measures, this measure has its limitations: mainly because it assumes that by having certain goods at a specific time, one's standard of living is better, without considering that some may have intentionally chosen not to have those goods. However, it does provide some indication of the socio-economic standing of households in a way that is easy to collect in a short self-completed form. As expected, due to the employment status of caregivers and the Quintiles of the schools sampled, most households that learners came from were assigned LSM scores lower than 7. Over 60% of learners' households were assigned an LSM score lower than 5; these are the LSM groups assumed to have the least affluence, where some high school education was completed, and are primarily rural. The characteristics of these LSM groups corroborate the other information provided by caregivers through the home background questionnaire.

4.1.4 Teacher characteristics

These characteristics reported in Table 14 were collected from a teacher survey completed face-to-face by a total of 361 early grade teachers interviewed across the 138 schools, averaging approximately three foundation phase teachers per school.

Table 14: Teacher sample characteristics by intervention arm

	Total number of Teachers	Percentage of Teachers	Intervention arm		
			Control	External coaching	DH coaching
Age					

	Total number of Teachers	Percentage of Teachers	Intervention arm		
			Control	External coaching	DH coaching
18 - 34 years	35	10%	11%	8%	9%
35 - 44 years	64	18%	20%	14%	18%
45 - 54 years	149	41%	38%	42%	45%
55 -64 years	112	31%	30%	36%	27%
65 years +	1	0%	1%	0%	0%
Education level					
Matric/ National Senior Certificate	1	0%	0%	1%	0%
Higher Certificate	12	4%	3%	5%	4%
Diploma	168	55%	53%	55%	59%
Bachelor Degree	70	23%	26%	19%	23%
Post Graduate Certificate	27	9%	12%	5%	7%
Post Graduate Degree	27	9%	6%	14%	7%
Years of experience					
1 -5 years	73	20%	20%	20%	21%
6 - 10 years	70	19%	24%	12%	20%
11 - 20 years	62	17%	15%	17%	20%
21 - 30 years	102	28%	30%	32%	19%
31 - 40 years	53	15%	10%	18%	19%
40 years +	1	0%	1%	0%	0%
Formal training to teach foundation phase					
Yes	243	80%	78%	77%	85%
No	62	20%	22%	23%	15%
Attendance of training					
None	4	1%	2%	0%	1%
1 session	10	3%	3%	3%	3%
2 sessions	17	5%	8%	3%	1%
3 sessions	60	17%	17%	11%	21%
All sessions	270	74%	70%	83%	74%
Use of methodologies in lessons					
Group guided reading HL *[0h30−1h15] ≥ 2 days/week**	271	90%	82%	96%	95%

	Total number of Teachers	Percentage of Teachers	Intervention arm		
			Control	External coaching	DH coaching
Group guided reading EFAL *[0h30−1h15] ≥ 1 day/week**	250	84%	84%	84%	85%
Spelling test HL *> 1 day/week**	301	99%	100%	100%	99%
Spelling test EFAL *> 1 day/week**	272	92%	93%	93%	89%
Phonics HL *[4h30−5h] ≥ 2 days/week**	289	96%	97%	94%	95%
Phonics EFAL *[15 min/session] ≥ 1 day/week**	291	99%	98%	100%	98%
Shared reading HL *[15min/session] ≥ 2-3 days/week**	289	96%	92%	97%	100%
Shared reading EFAL *[0h20-1h15] ≥ 1 day/week**	285	96%	96%	98%	95%
Differentiated activities	288	95%	95%	92%	96%
Experience with SMT					
Teachers feel supported	215	89%	88%	89%	91%
Teachers feel recognized	207	86%	89%	81%	87%

** Note that the italicised text represents the amount of time recommended for each methodology according to the Curriculum and Assessment Policy Statement, which the EGRP lesson plans align with. The numbers in brackets represent the minimum and maximum time recommended weekly.*

Similar to the EGRP Case Study Report findings, nearly three-fourths of the teachers surveyed are over 45 years old, and a third are over the age of 55 (close to retirement age). Nearly half of the teachers (43%) reported having over 20 years of teaching experience. Teachers in the external coaching intervention arm reported having slightly more experience than the other intervention arms, with 68% reporting more than ten years of teaching experience. 80% of Grade 1 and Grade 2 teachers reported having formal training to teach the foundation phase. Regarding Foundation Phase teaching training, DH coaching intervention schools had the greatest proportion of teachers who had received formal teaching in early grades.

More than half of the teachers surveyed (55%, n = 168) hold a teaching diploma as their highest qualification, whilst 41% (n = 142) of teachers reported having either a bachelor's degree or a postgraduate qualification. The external coaching intervention arm has the greatest proportion of teachers with qualifications with an NQF rating lower than 6.

When Foundation Phase teachers who are not DHs were asked whether they felt supported by their SMT in their work, 89% (n = 215) reported that they felt supported. A greater proportion of teachers teaching in DH coaching schools felt supported by the SMT. In contrast, a slightly higher proportion of teachers teaching in control schools felt recognised and appreciated by their SMT.

Teachers were further asked about their attendance of EGRP training, with most reporting that they attended all the training sessions (75%, n = 270). Attendance by teachers was better in the external coaching intervention arm, where 83% of teachers surveyed reported having attended all sessions compared to 70% and 74% in the control and DH coaching intervention arms, respectively. The teachers were then asked how frequently they integrate teaching methodologies such as group-guided reading (GGR), spelling tests, phonics and shared reading (SR) in their lessons. Most teachers interviewed reported using these methodologies at the frequency advised by Curriculum and Assessment Policy Statement, which the EGRP lesson plans align with, as indicated in Table 14. It is observed that more teachers report using GGR in their Home Language (HL) lessons than in the English First Language (EFAL) lessons. This was more pronounced in intervention schools than in control schools. Classroom observations conducted as part of the 2021 EGRP Case Study found that teachers were more comfortable teaching HL over EFAL, enabling easier adoption of new teaching methodologies in HL lessons.

Overall, there is a lower proportion of teachers using GGR compared to other methodologies. Data from interviews conducted with eight external EGRP coaches indicates that teachers generally struggle with applying GGR and shared reading correctly, often confusing the two in the process. This is not surprising since 38% of teachers interviewed (n = 289) for HL and 40% (n = 253) for EFAL reported GGR being the most challenging methodology. This finding is consistent with the 2021 EGRP Case Study classroom observations, which confirmed that teachers struggled to apply the methodology correctly in classrooms. EGRP coaches interviewed in late 2021 and early 2022 were almost unanimous in their view that teachers were not using these methodologies as a regular practice because even learner responses and reactions to the methodology suggested that it was something new to them or something they were not accustomed to. Some coaches pointed out that it was clear that most of the teachers did not prepare well enough to apply these methodologies in their classes despite knowing that they would be observed but instead preferred to lean on methods they were more comfortable with. Below are some pull quotes from the coaches:

> *They [teachers] are not yet comfortable with group guided reading as a methodology. They still need to get better at it. With my findings, I can say that they did not understand it. Sometimes you think that they hear and understand everything but when you go to the schools you realize that they are doing the opposite (EGRP Coach 1)*

> *When I observe teachers doing group guided reading, it becomes so tense in the classroom and it's like learners are doing something new that they are not used to do and they are scared to do. (EGRP Coach 2)*

These findings may imply that whilst 90% and 84% of the teachers self-reported using GGR as per CAPS recommended frequency in HL and EFAL, respectively, there is a possibility that given the challenges around the correct adoption of GGR in the classrooms, most teachers did not use it correctly or regularly enough to gain confidence to extract significant value from it for their learners. The higher numbers self-reported by the teachers might be reflecting a desirability bias or a compliance approach. The DBE and the implementing partner could benefit from a review of the use of GGR or devise strategies to strengthen its use given the realities of its mixed adaptation.

4.1.5 Departmental head characteristics

Face-to-face structured interviews were conducted with at least one Departmental Head (DH) from each school. Table 15 presents some of the findings and sample characteristics of these interviews.

Most DHs are mature and well-qualified teachers: over the age of 45 years (76%) with over 10 years of foundation phase teaching experience (88%), and 28% have pursued postgraduate qualifications related to teaching. Interestingly, only 17% of DHs have more than 10 years of experience in the DH position, implying that most were only appointed to these positions in the later part of their careers.

Table 15: DH sample characteristics by intervention arm

	Total number of Learners	Percentage of Learners	Intervention arm		
			Control	External coaching	DH coaching
Age					
18 - 34 years	8	6%	5%	3%	11%
35 - 44 years	11	8%	14%	0%	8%
45 - 54 years	72	53%	53%	58%	49%
55 -64 years	45	33%	29%	40%	32%
Education level					
Higher Certificate	1	1%	2%	0%	0%
Diploma	24	18%	22%	20%	8%
Bachelor's degree	24	18%	12%	23%	2%
Post Graduate Certificate	7	5%	7%	5%	3%
Post Graduate Degree	15	11%	7%	15%	14%
Years of teaching experience					
1 -5 years	1	1%	0%	0%	3%
6 - 10 years	15	11%	15%	3%	14%
11 - 20 years	34	25%	31%	15%	27%
21 - 30 years	60	44%	44%	53%	35%
31 - 40 years	26	19%	10%	30%	22%
40 years +	0	0%	0%	0%	0%
Years in DH position					
1 -5 years	85	63%	66%	62%	58%
6 - 10 years	28	21%	20%	19%	23%
11 - 20 years	17	13%	10%	14%	15%

	Total number of Learners	Percentage of Learners	Intervention arm		
			Control	External coaching	DH coaching
20 years +	6	4%	3%	5%	5%
Is the DH a foundation phase teacher					
Yes	126	93%	92%	98%	89%
No	10	7%	8%	3%	11%
Formal training to teach the foundation phase					
Yes	58	43%	37%	53%	41%
No	13	10%	12%	10%	5%
Previous coaching experience					
Yes	14	39%	-	-	39%
No	22	61%	-	-	61%

Most (93%) Foundation Phase DHs are also teachers in the phase. The DHs were asked whether they were planning to leave the DH position or retire in the next two years, and 29% said they intended to resign or retire from the position in the near future. A follow-up question was asked as to whether there is a succession plan for the next person to take up the position to which 77% DHs reported that someone is being groomed to take over their position. This aspect is of particular interest for the DH coaching intervention arm, as withdrawal of the DH directly affects programme fidelity. Homing in on the DH coaches arm shows that 31% of DHs report planning to leave their post whilst only 64% have identified and begun grooming a successor. It is critical that the programme implementation singles out schools with DHs intending to leave the DH position before programme close-out and take the potential successor(s) alongside the DHs through the support and coaching activities – this will ensure implementation integrity and programme continuity.

DHs in the coaching intervention arm (both DH coaching and external coaching cohorts) were further asked a series of questions related to the coaching role some will occupy. They were asked if they had previous coaching experience, and 61% (n = 22) reported no coaching experience. They were further asked if they were interested in assuming the coaching responsibilities, and 86% (31) indicated that they were interested as it already relates to their KPIs, 8% (n = 3) were somewhat curious, and 6% (n = 2) were not interested at all. These questions sought to provide insights into the likelihood of DH-led coaching innovations being sustainable in practice.

4.1.6 Principal sample characteristics

Structured interviews were also conducted with the Principal of each school, and a total of 129 Principals' responses were captured; the other Principals were unavailable on the day of the visit and did not honour telephonic interview appointments. Some of the details of the fielded information are reported in Table 16.

Table 16: Principal sample characteristics by intervention arm

	Total number of Principals	Percentage of Principals	Intervention arm		
			Control	External coaching	DH coaching
Age					
18 - 34 years	1	1%	2%	-	-
35 - 44 years	1	1%	-	3%	-
45 - 54 years	86	67%	76%	65%	54%
55 - 64 years	41	32%	22%	32%	46%
Years of experience					
1 - 5 years	72	56%	61%	46%	59%
6 - 10 years	29	23%	20%	30%	19%
11 - 20 years	18	14%	13%	19%	11%
21 - 30 years	9	7%	6%	5%	11%
Principal teaches in the Foundation Phase					
Yes	5	4%	2%	1%	2%
No	123	96%	98%	99%	98%
Frequency of meetings with FP teachers					
Weekly	16	13%	9%	17%	14%
Monthly	46	36%	40%	31%	35%
Once a term	57	45%	44%	47%	43%
Never	9	7%	7%	6%	8%

Similar to DHs, Principals were also mainly middle-aged (98%), while over half (56%) reported having less than five years of experience as a Principal. All Principals were also subject teachers, with only 4% teaching foundation phase learners. This preference is likely due to the admiration attached to teaching older learners and the flexibility that teaching in the Intersen Phase offers them to also fulfil managerial duties. When asked how frequently they meet with Foundation Phase teachers, 45% of the Principals reported meeting only once a term, while 49% reported meeting at least once a month.

4.2 Midline learner assessment results

This section presents the results of learner reading outcomes in both HL and EFAL for Grade 1 and Grade 2 after almost one year of programme implementation (data collection was carried out in November 2021). It also details the distribution of learner scores per sub-task and highlights any differences in learner performance per intervention arm. Ceiling or floor effects are also assessed per sub-task.

Using Principal component analysis (PCA), a composite index was constructed to enable a more compact analysis and presentation of results. This combined several sub-task results per grade into a single score reflecting overall performance.

Previous reading studies (EGRS I and EGRS II) assessed learners at the beginning of the academic year to establish pre-literacy levels (i.e., especially for Grade 1 learners). In contrast, the EGRP programme assessed learners after being exposed to the core intervention (LTSM and training) and external coaching for most of a year. This implies that the 2021 round of data collection is in fact a midline whose results will be compared with learner performance after the complete immersion of the DH coaching intervention. The midline will thus be used as the basis for a follow up end-line data collection and analysis to be conducted in late 2023, where the same learners will be assessed.

4.2.1 Description of Learner Assessment subtests

The main goal of learning to read is comprehension – the ability to extract and construct meaning through interacting with written language. Gaining the skills to read fluently and comprehend what you have read is a developmental process, a complex skill that is only built when a range of sub-skills or lower-order skills have been built and mastered (Chall, 1996). The range of sub-tasks on the early grade reading assessment subtests included in this midline, and this study in general, measure a range of skills from lower-order skills (such as identifying letter sounds), which are predictive of future reading achievement, to higher-order skills (reading and comprehension).

For both grades, the assessments were administered for HL (Setswana) and EFAL. Individual sub-tasks were administered to a learner one-on-one by trained SSA field researchers, whilst group sub-tasks were administered to the same learners grouped as 20 per class by a field researcher. The table below shows the sub-tasks assessed at midline for Grade 1 (Table 17) and Grade 2 (Table 18) learners and the time spent on each sub-task.

The Grade 1 learners faced only five tests: four HL tests conducted with each learner individually, and one, the only EFAL test, as a group assessment.

Table 17: Grade 1 learner assessment sub-tasks

Grade 1 Assessment Type	Test Item	Indicative Administration Time (min)	Total Administration Time (min)
HL one-on-one test	Rapid Object Naming	5	20
	Letter Sound Knowledge	5	
	Oral Reading Fluency	5	
	Reading Comprehension	5	
EFAL group test	Word Sound Knowledge	10	10
			30

For Grade 2 learners, five HL individual sub-tests and two HL group tests were designed, whilst EFAL had two individual tests and one group test.

Table 18: Grade 2 learner assessments sub-tasks

Grade 2 Assessment Type	Test Item	Indicative Administration Time (min)	Total Administration Time (min)
HL one-on-one test	Rapid Object Naming	5	25
	Letter Sound Knowledge	5	
	Oral Reading Fluency (ORF)	5	
	Reading Comprehension	5	
	Word Recognition	5	
HL group test	Complex Consonant Sound Knowledge	10	20
	Word Recognition and Writing	10	
EFAL one-on-one test	Oral Reading Fluency	5	10
	Reading Comprehension	5	
EFAL group test	Word Sound Knowledge	10	10
			65

The decision as to which sub-tasks should be administered to Grade 1s and which to Grade 2s were determined by a number of factors based on the grade-specific curriculum. Grade 1s were predominantly tested on HL as EFAL is not the main focus in Grade 1. Grade 2 learners faced a broader range of sub-tasks as they are expected to be able to read words and so text in their HL and EFAL after nearly two years of schooling. The majority of sub-tasks for Grade 2 learners were still in HL. This is the language in which the learners are assumed to have gained most reading competence. As indicated earlier, learners learn to read first in their familiar HL and then transfer these decoding skills to EFAL. In Grade 2, they are still developing basic reading skills and have not become fluent readers even in HL.

The object naming sub-task, the letter sound grids, and the structure of the reading texts and comprehension questions were replications of tests used in the earlier iterations of the EGRS or were adapted using standard EGRA rules. This allows for a comparison of learner performance over time and between tests.

The decision as to which sub-tasks should be administered using a group-administered process – which is much faster as it tests 20 learners at a time – was based on the experience of which tests can be administered to a group and which can only be administered individually to each learner. For instance, reading fluency along with comprehension cannot be group administered tasks, while letter sounds, word recognition and dictation can be structured as group-administered tasks. For letter sounds and word recognition, the learners are given five options, and they choose the one that reflects the letter sound or word that has been read to them.

The group tasks were added in the EGRP and had not been used in previous EGRS studies. While this means that there is no basis for comparison from earlier studies, the addition of group tasks brings another dimension to the testing and has been shown in other African countries to produce greater differentiation between learner scores and less floor effects.

Having group tasks also allows the introduction of simple writing tasks - the Grade 2 'word recognition and writing' subtask involves learners writing the words that they hear spoken by the researcher singly and in a short sentence. This test can be made more complex in Grade 3 and so measure learner abilities and development in turning the read word into a meaningful sentence.

4.2.2 Learner Test Scores

This section gives a detailed breakdown on learner performance per grade by sub-task. The tests administered were standard sub-tasks expected of learners to know at each grade. Table 19 presents the average performance and summary statistics per grade.

Grade 1 learners performed reasonably well in EGRA HL visual tasks (Rapid Object Naming), given they had 20 seconds to identify 36 objects.

This test simply measures recall and speed in responding to instructions and visual prompts. It is a pre-reading task which tests understanding of instructions and uses visual prompts to measure basic pre-reading skills and speed.

The other sub-tasks appeared difficult for the Grade 1 learners and present floor effects, particularly in the HL oral reading fluency and reading comprehension sub-tasks. These floor effects might be due to a lack of exposure to higher order reading skills during Grade 1 and so the inability to identify and blend letter sounds into words and then make meaning of them. More worrying is that most learners also struggled with HL letter-sound knowledge, with only signs of better performance in the 90th percentile, as shown in Table 19 and 20. For Grade 1 learners, the DBE benchmark for HL ORF is "sounding at least 40 correct letters per minute". A similar trend is noticed in the EFAL word sound knowledge, administered as a group task, where good performance only emerged in the 75th percentile. Although these tasks proved difficult for learners, with extreme floor effects, the sub-task isolates already strong learners based on their performance. It also provides a basis for comparison in later years of the intervention.

Table 19: Grade 1 Average Performance per sub-task

Assessment Type	Subtask	Observations	Mean	SD	se	Cronbach
HL one-one-one test	Rapid Object Naming (/36)	2915	14.15	4.39	0.08	0.89
	Letter Sound Knowledge (/70)	2915	18.71	18.31	0.34	0.98
	Oral Reading Fluency (/49)	2915	6.16	14.71	0.27	0.99
	Reading Comprehension (/5)	2915	0.30	0.98	0.02	0.89
HL group test	Letter Sound Knowledge (/5)	2915	3.4	1.61	0.03	0.74
EFAL group test	Word Sound Knowledge (/5)	2915	2.29	1.53	0.03	0.66

Table 20: Grade 1 Summary Statistics per sub-task

Assessment Type	Subtask	Observations	Score attained at each percentile					Min	Max	Target
			p10	p25	p50	p75	p90			
HL one-one-one test	Rapid Object Naming (/36)	2915	9	11	14	17	19	0	33	36
	Letter Sound Knowledge (/70)	2915	0	3	13	30	47	0	70	70
	Oral Reading Fluency (/49)	2915	0	0	0	0	38	0	49	49
	Reading Comprehension (/5)	2915	0	0	0	0	1	0	5	5
HL group test	Letter Sound Knowledge (/5)	2915	1	2	4	5	5	0	5	5
EFAL group test	Word Sound Knowledge (/5)	2915	0	1	2	4	4	0	5	5

From Table 21, it is evident that there were no sub-tasks with ceiling effects for Grade 2; however, there were some with floor effects where the average performance was below what is expected of learners in the grade. For instance, sub-tasks such as oral reading fluency, word recognition, and especially reading comprehension in both languages seemed difficult for most learners. Table 22 shows that the 50th percentile of learners scored a zero in oral reading fluency and reading comprehension, HL and EFAL. Although the DBE benchmark for HL oral reading fluency at Grade 2 is “40 words correctly read per minute” and for EFAL oral reading fluency is “30 words correctly read per minute”.

However, there were some sub-tasks where they performed reasonably well, such as on rapid object naming (where the 50th percentile was able to identify 16 objects correctly out of 36 objects in 20 seconds), HL letter-sound knowledge (where the 50th percentile was able to identify 27 letters out of 70 letters correctly in 60 seconds), complex consonant sound knowledge (HL) and word sound knowledge (EFAL). Overall, Grade 2 learners performed reasonably well on phonetic sub-tasks but struggled with reading fluency and comprehension, especially in EFAL. This indicates that many learners are able to associate the letter sound and its written form. However, after two years of schooling, learners should be able to use this basic building block to read words and sentences and so read sentences and make sense of text. It seems that most learners are unable to do this. This indicates that the majority of learners are struggling to read – although as with Grade 1 there is at least one learner who is able to read fluently and answer all the questions on the text in HL and in EFAL.

Table 21: Grade 2 Average performance per sub-task

Assessment Type	Subtask	Observations	Mean	SD	se	Cronbach
HL one-on-one test	Rapid Object Naming (/36)	2897	15.83	4.44	0.08	0.90
	Letter Sound Knowledge (/70)	2897	29.57	22.36	0.42	0.99
	Word Recognition (60)	2897	9.15	12.02	0.22	0.97
	Oral Reading Fluency (49)	2897	16.42	21.22	0.39	0.99
	Reading Comprehension (/5)	2897	1.06	1.69	0.03	0.99
EFAL one-on-one test	Oral Reading Fluency (/62)	2897	10.08	19.63	0.36	0.99
	Reading Comprehension (/5)	2897	0.34	0.96	0.02	0.83
HL group test	Complex Consonant Sound Knowledge (/5)	2897	2.97	1.63	0.03	0.70
	Word Recognition and Writing (/10)	2897	3.83	3.99	0.07	0.95
EFAL group test	Word Sound Knowledge (/10)	2897	5.45	2.7	0.05	0.78

Table 22: Grade 2 Summary Statistics per sub-task

Assessment Type	Subtask	Observations	Score attained at each percentile					Min	Max
			p10	p25	p50	p75	p90		
HL one-on-one test	Rapid Object Naming (/36)	2897	10	13	16	18	21	0	35
	Letter Sound Knowledge (/70)	2897	1	9	27	48	64	0	70
	Word Recognition (/60)	2897	0	0	3	16	29	0	60
	Oral Reading Fluency (/49)	2897	0	0	0	44	49	0	49
	Reading Comprehension (/5)	2897	0	0	0	2	4	0	5
EFAL one-on-one test	Oral Reading Fluency (/62)	2897	0	0	0	3	51	0	62
	Reading Comprehension (/5)	2897	0	0	0	0	1	0	5
HL group test	Complex Consonant Sound Knowledge (/5)	2897	1	2	3	4	5	0	5
	Word Recognition and Writing (/10)	2897	0	0	2	8	10	0	10
EFAL group test	Word Sound Knowledge (/10)	2897	2	3	6	8	9	0	10

Before getting into greater detail on learner performance per sub-task test, a presentation of correlation results between sub-tasks is made for both grades. While correlation does not imply causation, it is possible to reasonably but carefully infer dependencies between these tests based on early grade reading literature and the progressive sequence in which these concepts are taught to learners in practice. Although floor effects for sub-tasks such as oral reading fluency and reading comprehension for Grade 1 may affect correlation analysis, for example, the analysis still provides some useful insights.

Table 23, for instance, presents correlation results between sub-task performances for Grade 1 that reveals some interesting relationships between the sub-task results. It shows that learners' performance in HL letter-sound knowledge is positively correlated with oral reading fluency, reading comprehension, and word sound knowledge (EFAL and HL). The positive relationship observed between HL oral reading fluency and HL letter-sound knowledge is unsurprising since learners who do well in reading need more robust letter recognition and phonological awareness, which builds their ability to decode text and later decode whole words and subsequently read fluently.

Equally interesting is the relatively higher positive relationship between HL letter-sound knowledge and EFAL word sound knowledge. Though not ascribing causation, it gives credence to studies that found that a child who learns to read from a language they regularly speak at home finds it easier to read in a second language (UNESCO, 1953). Put another way, it implies that once a child understands how decoding works in their HL, they can use the same knowledge and skills in a second language, EFAL. Also noteworthy is the highest positive correlation which is between HL oral reading fluency and HL reading comprehension, which confirms what is already known about reading fluency being a prerequisite to being able to understand what has been read.

In comparing individual sub-tasks to group assessment, the correlation between HL letter sound knowledge subtask (individual) is better correlated with EFAL word sound knowledge (group) than HL letter sound knowledge task (group task). The expectation was that the same task performed in a group will be highly correlated with the individual subtask, however there may be some reasons why this is not the case. That is the group task are usually performed in a setting which is more natural to the learner and the individual subtask are more demanding and detailed (than the group task), for instance for the HL letter sound knowledge subtask learners are required to identify 70 objects per minute whereas for the group task it's just 5 letters that need to be identified.

Table 23: Grade 1 Subtest Correlation Matrix

	HL rapid object naming	HL letter-sound knowledge	HL oral reading fluency	HL reading comprehension	HL letter sound knowledge - group	EFAL word sound knowledge - group
HL rapid object naming	1.00					
HL letter-sound knowledge	0.22	1.00				
HL oral reading fluency	0.10	0.66	1.00			
HL reading comprehension	0.08	0.49	0.79	1.00		
HL letter-sound knowledge - group	0.15	0.51	0.29	0.19	1.00	
EFAL word sound knowledge - group	0.14	0.61	0.44	0.30	0.61	1.00

The Grade 2 correlation matrix is presented in Table 24. This shows similar findings from Grade 1 sub-task correlation results, albeit showing a much stronger relationship on selected sub-tasks. For example, letter-sound knowledge in HL was fairly positively correlated with most sub-tests across both languages. The HL word recognition sub-task is especially strongly correlated with sub-tasks such as oral reading fluency in HL and EFAL, HL reading comprehension (with a reasonably positive result for EFAL comprehension), complex consonant sound knowledge and word sound knowledge. As letter and word recognition are the basis for all other higher-order reading skills, this relationship is to be expected.

Also notable is the relatively strong relationship between HL oral reading fluency and reading comprehension, which is consistent with early grade reading acquisition literature. Similarly, a fairly strong relationship between some HL sub-tasks and EFAL sub-tasks is observed. This suggests that being competent in reading in HL (Setswana) may be a good predictor for knowledge in EFAL sub-tasks. The correlation matrix also suggests that being able to recognise words may improve the ability to read fluently and read for comprehension. This again is consistent with the vast body of knowledge in early grade reading literature.

Table 24: Grade 2 Sub-test Correlation Matrix

	HL rapid object naming	HL letter sound knowledge	HL word recognition	HL oral reading fluency	HL reading comprehension	EFAL oral reading fluency	EFAL reading comprehension	HL complex consonant sound knowledge - group	HL word recognition and writing- group	EFAL word sound knowledge - group
HL rapid object naming	1.00									
HL letter sound knowledge	0.33	1.00								
HL word recognition	0.26	0.77	1.00							
HL oral reading fluency	0.23	0.77	0.86	1.00						
HL reading comprehension	0.21	0.68	0.84	0.87	1.00					
EFAL oral reading fluency	0.18	0.60	0.84	0.73	0.77	1.00				
EFAL reading comprehension	0.13	0.41	0.59	0.50	0.61	0.71	1.00			
HL complex consonant sound knowledge - group	0.16	0.56	0.55	0.57	0.50	0.44	0.30	1.00		
HL word recognition and writing- group	0.22	0.78	0.80	0.83	0.75	0.65	0.45	0.63	1.00	
EFAL word sound knowledge - group	0.22	0.68	0.63	0.63	0.57	0.52	0.38	0.63	0.69	1.00

Evaluation of Individual Assessments versus Group-Based Assessment

This section seeks to understand whether there is a difference in performance between individually administered subtask compared to group administered sub-task. Therefore, the intra-class correlation coefficient (ICC) or Rho value is used to calculate the extent to which writing in a group leads to similar scores in comparison to individually administered assessments. The ICC score is between 0 and 1 and can be interpreted as the total amount of variation that is attributable to between unit rather than within unit difference

Both the HL and EFAL subtask scores were used in calculating the ICC score were standardized before the ICC coefficient was calculated. The rationale for the standardization was because the various subtasks have different number of items and standardization avoids having any of the test scores unfairly influence the ICC score.

In order to calculate the ICC values for both Grade 1 and 2, there was a need to choose a similar subtask.

For Grade 1 learners, since HL letter sound knowledge was assessed at both individual and as a group subtask, it was preferred to use in calculating the ICC scores. The calculated ICC scores for Grade 1's are in Table 25 and this can be interpreted as the between unit variation accounting for 25.5% of the total variation, thus most of the variation was within the given subtask. This makes sense since performance in Grade 1 were below par for many of the subtask with the exception of rapid object naming subtask.

Table 25: Intra-class Correlation (Grade 1)

Grade	Subtask	One-on-One vs Group subtask
Grade 1	ICC score	0.255
	95% confidence interval	0.09<ICC<0.532

Likewise for Grade 2 learners, word recognition subtask was used in estimating the ICC score (i.e. due to it being assessed at both individual and group level subtask). From Table 26, the calculated ICC score for Grade 2 learners is 0.51, implying that the between unit variance accounts for 51% of the total variation.

Table 26: Intra-class correlation coefficient (Grade 2)

Grade	Subtask	One-on-One vs Group subtask
Grade 2	ICC score	0.51
	95% confidence interval	0.10<ICC<0.71

The following section presents each sub-task in greater detail, focusing on the distribution of learner performance per sub-task across intervention arms per grade.

Sub-task 1: Rapid Object Naming (RON)

This visual prompt task required a learner to name objects they recognise and after agreement is reached with the enumerator about identifying each object, quickly and automatically name them in their home language (Setswana) from pictorial stimuli presented by a field researcher. Even though the RON subtest is not a test of reading skills and is not taught in school, it is considered to be predictive of later reading outcomes. It is, therefore, a good candidate for a baseline control, in this case a midline control to be included in the endline regression equation.

Exactly the same test was administered to both Grade 1 and 2 learners, and learners in both grades had 20 seconds to name as many objects as possible from a maximum of 36 objects presented to them. The test consists of 6 pictured items (pig, book, table, chair, dog, sun) which are presented in various different orders.

An auto-stop rule that automatically discontinued the test was implemented if a learner could not identify the first 5 consecutive image prompts on the grid. Where a learner kept quiet for 5 seconds on a stimulus, the nudge rule was applied where the field researcher prompted the learner to attempt naming the next item on the pictorial stimuli.

Grade 1 learners performed well, given that they were supposed to identify nearly 2 objects in less than a second (i.e., 36 pictures of objects in 20 seconds). The 10th percentile of the learners correctly identified 9 images out of 36 in total whilst the 50th percentile correctly identified 14 pictures within 20 seconds. None of the Grade 1 learners was able to identify all 36 pictures correctly within the 20 seconds allocated (i.e., the highest was 33 objects correctly identified).

Figure 4 shows a good spread of achievement with this subtask with neither a floor nor ceiling effect of the scores but instead resembles a fair approximation of a normal distribution. This also confirms a good balance across intervention groups. Figure 5 indicates a close link between attempted learner responses and correctness, implying that learners who tried were more likely to correctly identify the object in the picture. Given the low skill required for this task, this was to be expected.

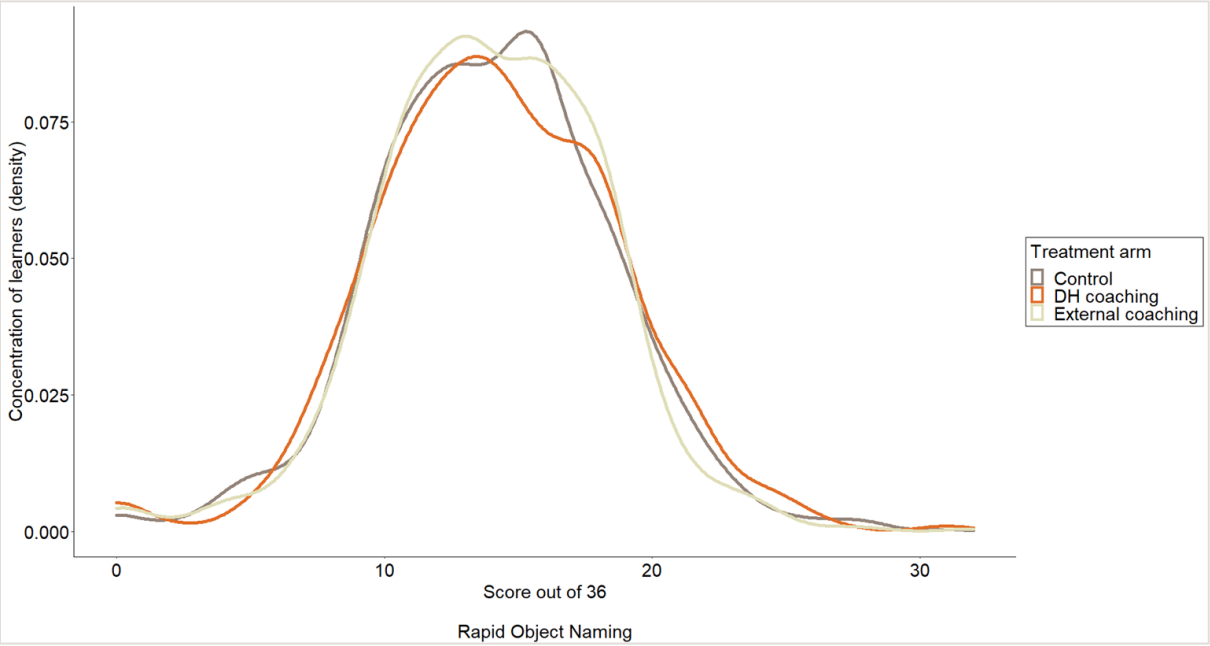

Figure 4: Grade 1 Kernel Density Curve for Rapid Object Naming (RON) by intervention arm

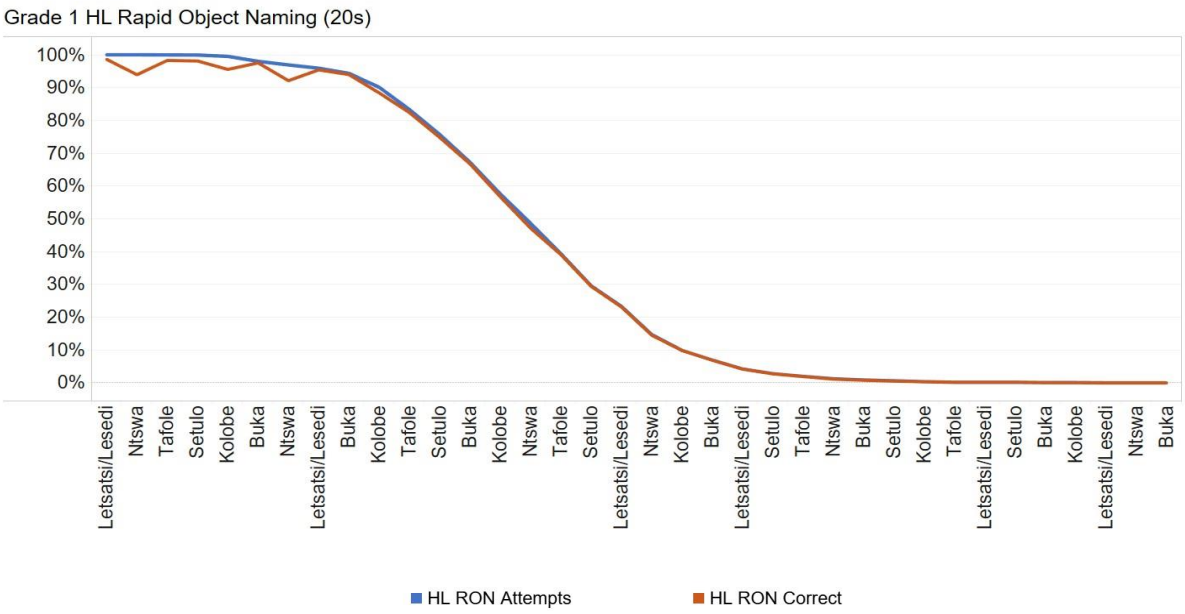

Figure 5: Grade 1 Number of Objects correctly identified in HL Rapid Object Naming sub-task

Grade 2 learners performed better than the Grade 1 learners as expected, with the 10th percentile able to correctly identify 10 out of 36 objects and the 50th percentile of learners able to identify 16 objects correctly within 20 seconds. The maximum number of objects correctly identified was 35 out of 36 within the 20 seconds. Figure 6 reflects a good spread of achievement with this sub-task, presenting neither a floor nor ceiling effect of the scores but approximating a normal distribution. It also confirms a good balance across intervention groups.

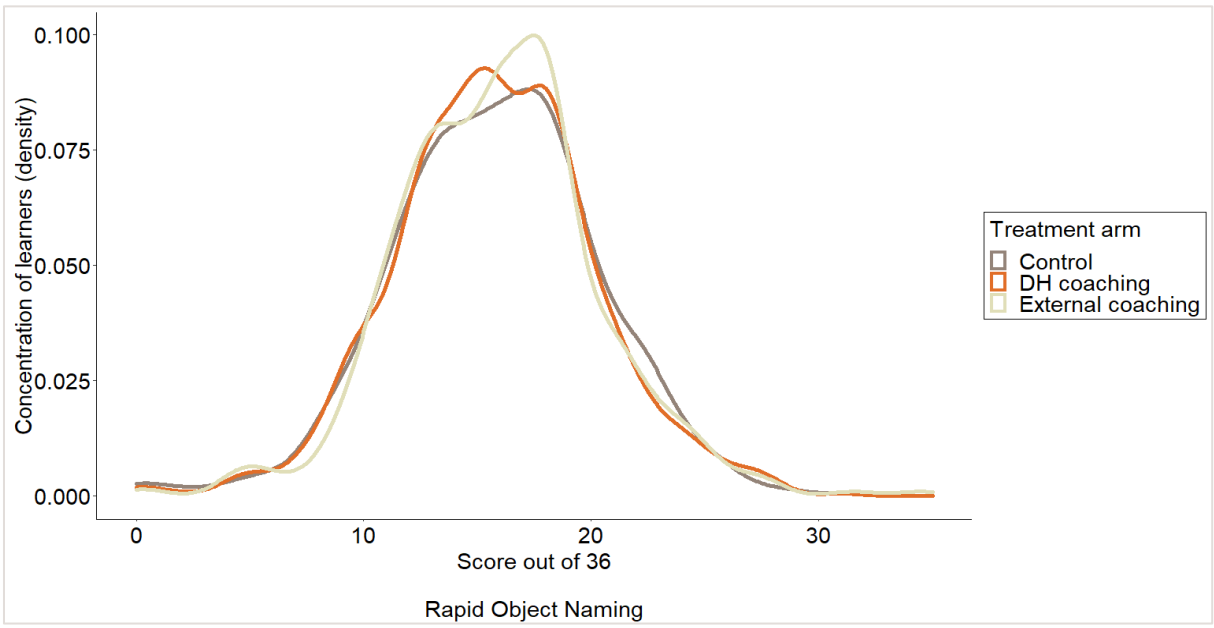

Figure 6: Grade 2 Kernel Density Curve for Object Naming by intervention arm

Similar to Grade 1 learners, Figure 7 indicates a close link between attempts and correctness on the rapid object naming sub-task for Grade 2 learners. Learners who attempted were more likely to identify the object in the picture. As expected, Grade 2 learners performed slightly

better on this task, with the 50th percentile of Grade 2 learners identifying 16 objects compared to 14 objects correctly identified by Grade 1 learners.

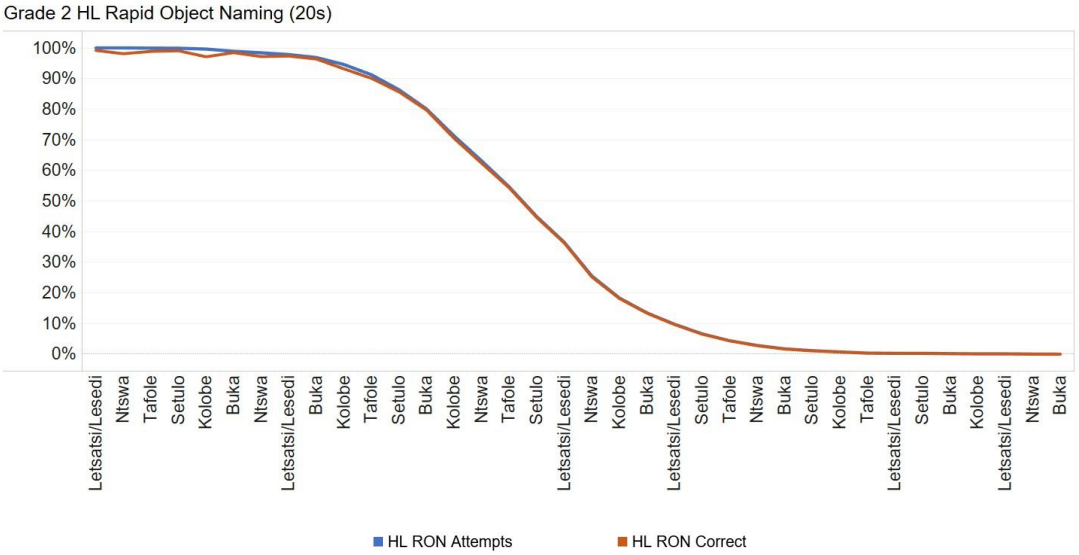

Figure 7: Grade 2 Number of Objects correctly identified in HL Rapid Object Naming sub-task

Sub-task 2: HL Letter-Sound Knowledge

This sub-task is deemed a good predictor of later reading success. Learners were given 60 seconds to read as many letter sounds as possible from a grid with a total of 70 letters. The test was designed so that the letter sequence on the grid was based on the frequency on which the letters appear in the Setswana language and in early grade textbooks, with letters presented in both upper and lower case. Grade 1 and 2 learners are presented with the same grid of letters. Field researchers captured the number of letters attempted, and the number of letters correct within the 60 seconds given for this task. As expected, the number of letters attempted (averaging 28) was higher than the number of letters correctly read (19 on average). This test was administered to both grades.

The results in Figure 8 for Grade 1 learners exhibit a substantial floor effect on this sub-task. The 50th percentile could only correctly pronounce 13 letters within 60 seconds out of the expected maximum of 70 letters. The kernel density curves were closely identical across the intervention arms. However, the distribution tended to score slightly better on this sub-task for the external coaching intervention arm. This trend is investigated further in the balance test to understand if these differences are due to random chance or are statistically significant.

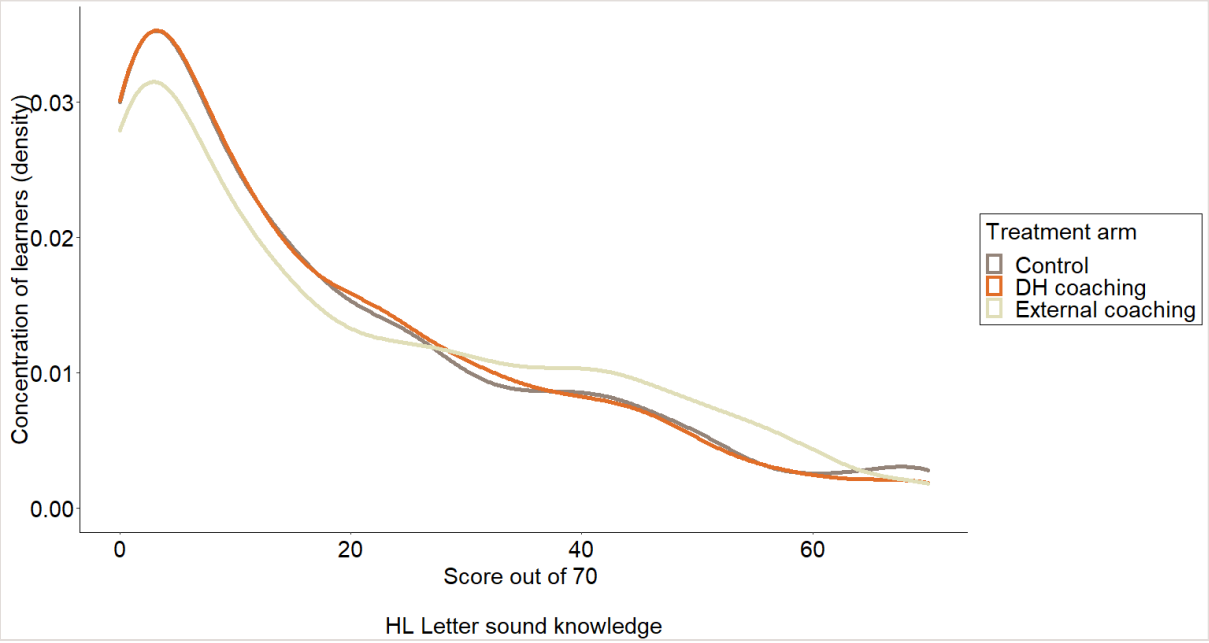

Figure 8: Grade 1 Kernel Density Curve for Letter sound knowledge by intervention arm

Figure 9 below shows a drop in attempts of approximately 14% of the learners because they could not correctly read any of the first 5 letters presented to them (auto-stop rule). It is also clear that learners struggled with some letter sounds such as “h” and “d”, which show sharp declines in performance.

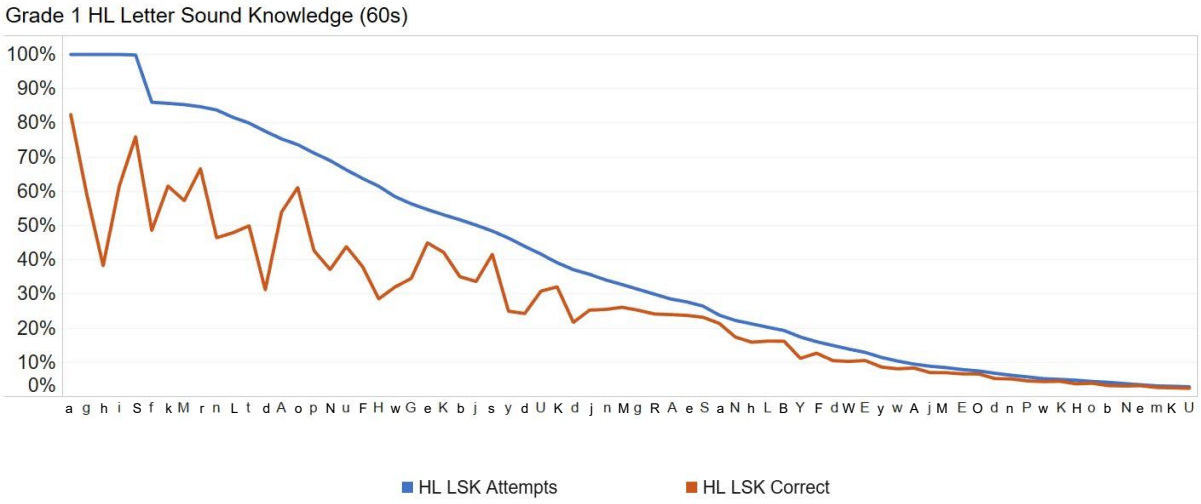

Figure 9: Grade 1 Number of letters correctly read in HL Letter Sound knowledge subtask

Grade 2 learners performed better in letter-sound knowledge in HL than Grade 1 learners as expected, although Figure 10 shows their results are still skewed, implying most learners did not perform well. Within sixty seconds, the 50th percentile of Grade 2 learners was able to identify 27 letter sounds correctly (compared to the 50th percentile of 19 for Grade 1 learners). This implies that, on average, a Grade 2 learner is able to read correctly only 8 more letter sounds in 60 seconds than a Grade 1 learner. The 75th and 90th percentile of Grade 2 learners correctly identified 48 and 64 letter sounds, respectively, within 60 seconds.

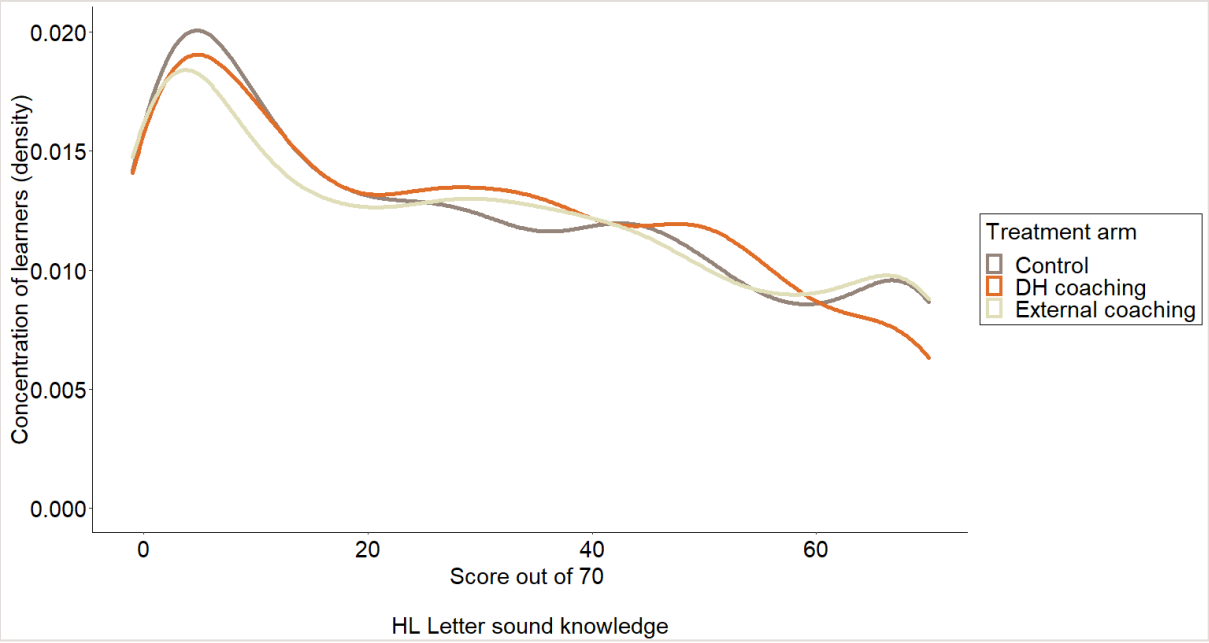

Figure 10: Grade 2 Kernel Density Curve for HL Letter sound knowledge by intervention arm

Similar to Grade 1 learners, Figure 11 shows a drop of approximately 10% of the learners in the number of attempts due to the auto-stop rule indicating that these learners could not correctly read any of the first 5 letters presented to them. This is particularly worrying as these learners have been in school at least two years and have yet to understand that the written letter relates to a specific sound – this is the most basic knowledge needed before a learner can start reading. Grade 2 learners generally struggled with the same letter sounds as the Grade 1s, including “h”, “d”, and “y”, which from the figure show sharp declines in performance

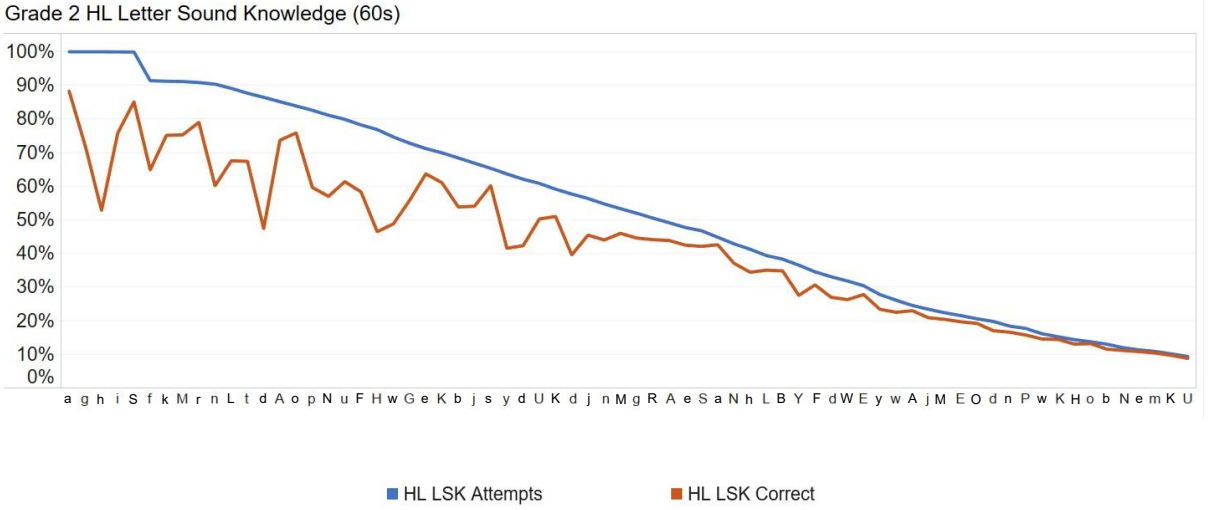

Figure 11: Grade 2 Number of Letters correctly read in Letter Sound knowledge sub-task

Sub-task 3: HL Word Recognition

This sub-task, administered to Grade 2 learners only, assessed learners’ ability to correctly read out as many words as possible presented to them in a grid by a field researcher within 60 seconds. An auto-stop rule was applied.

The results in Figure 12 shows that, on average, the Grade 2 learners did not perform well on this sub-task, with the 50th percentile of learners only able to read out three words correctly within 60 seconds. The 75th percentile could only read 15 words correctly, whilst the 90th percentile could only read 29 words. Since word recognition is a prerequisite for reading fluency and later reading comprehension, these results are expected to hinder performance on these sub-tasks given the relatively strong correlation observed between them. Figure 13 indicates learners struggled with reading words such as "*ntlo*" and "*poo*".

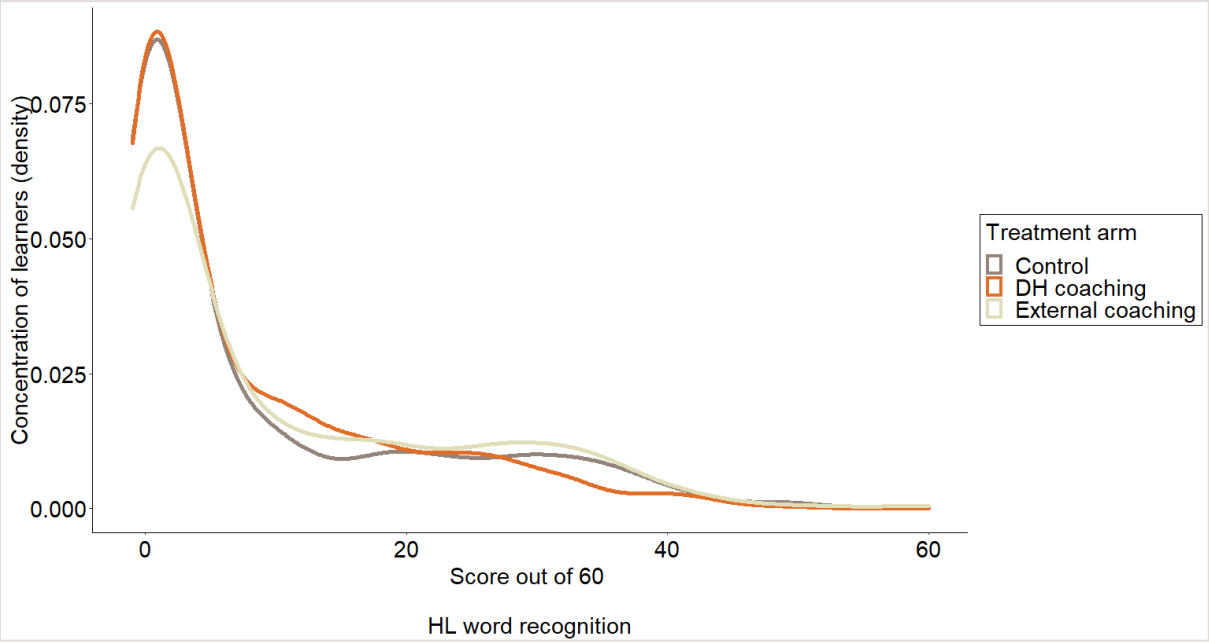

Figure 12: Grade 2 Kernel Density Curve for Word recognition by intervention arm

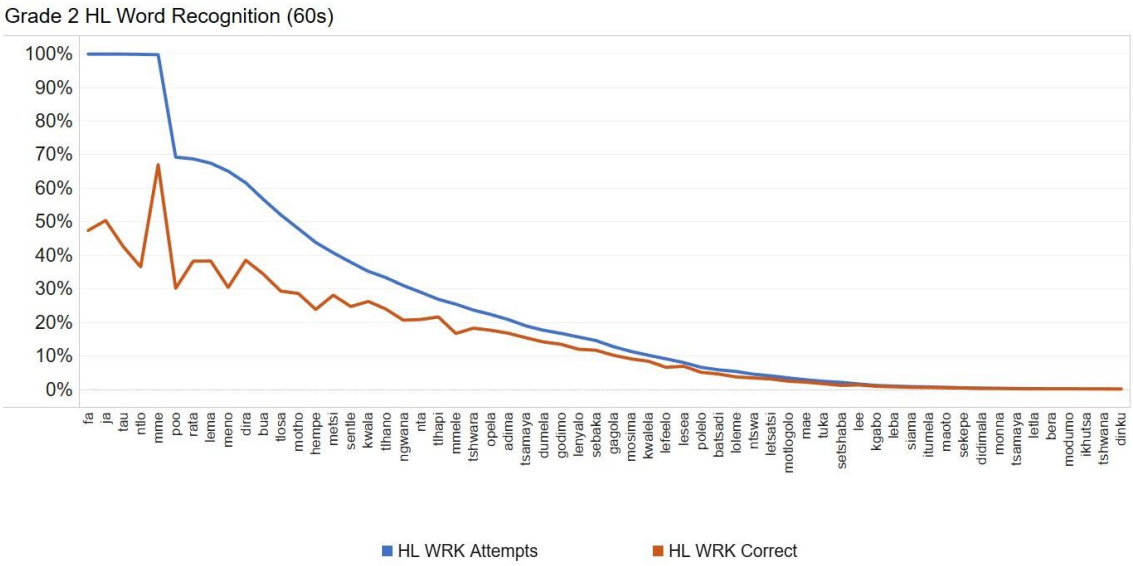

Figure 13: Grade 2 Number of HL words correctly recognised

Sub-task 4: HL Word Recognition and Writing – Group Task

A word recognition and writing task in HL was administered as a group task for Grade 2 learners only. Learners had to identify and write down the familiar words a field researcher

read out. They were all taken from Grade 1 and 2 texts. Figure 14 shows that learners did not perform well, with the 50th percentile of learners only able to identify and write down 2 out of 10 words in their home language. However, the 75th percentile of Grade 2 learners correctly identified and wrote down eight words correctly, and the 90th percentile attained a perfect score (i.e., scored 10 out of 10 words correctly). This performance shows two extremes where learners who could recognise and write words read out to them did so well whilst 37% of the learners failed to write even one word correctly. This reflects an inability to translate a spoken word into the written form.

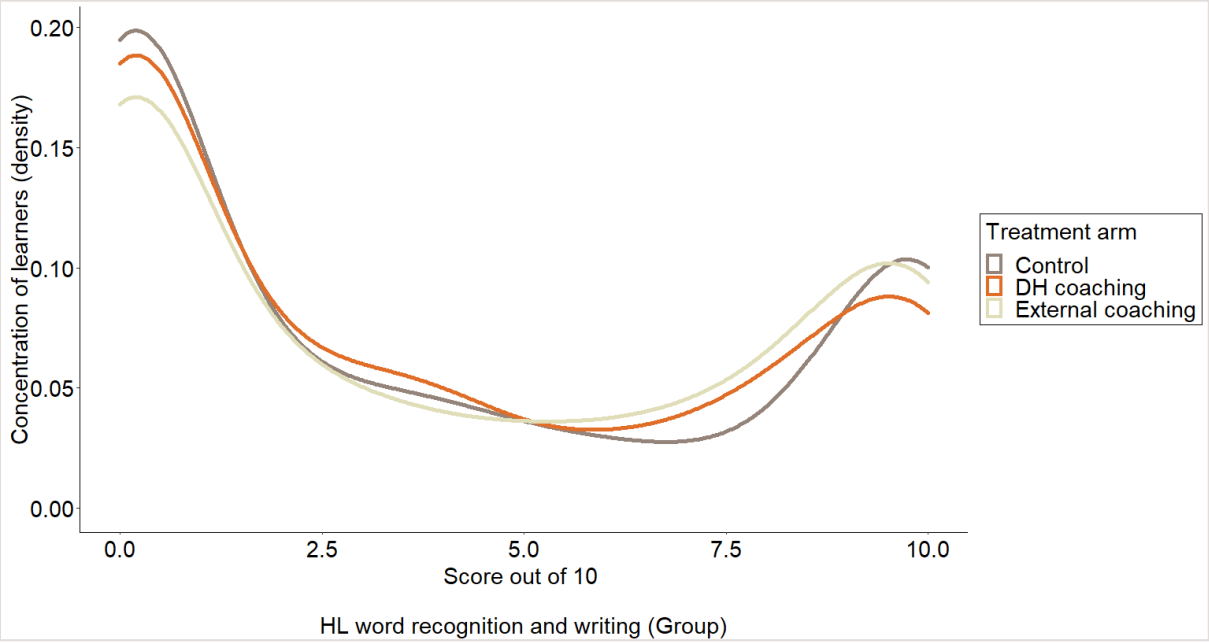

Figure 14: Grade 2 Kernel Density Curve for word recognition and writing by intervention arm.

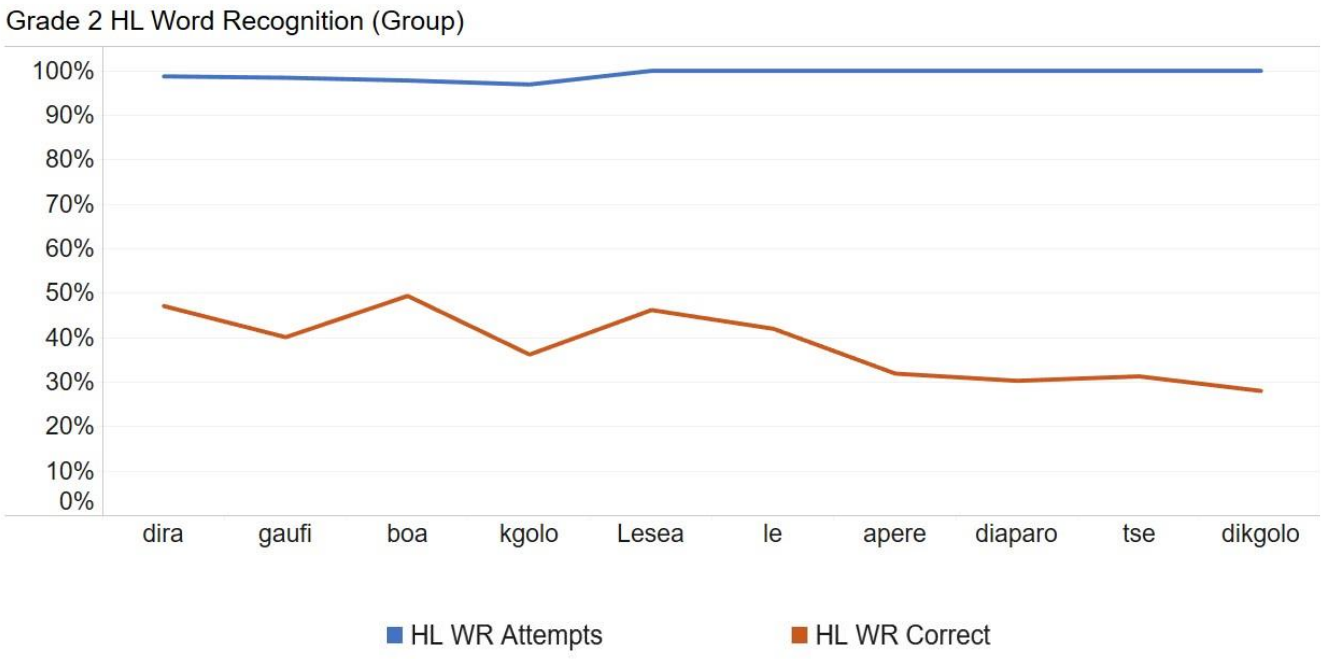

Figure 15: Grade 2 Number of words recognised and written down correctly

Sub-task 5: HL Oral Reading Fluency

Grade 1 and Grade 2 learners were tasked with reading the same story with five sentences consisting of 49 words in the home language. A learner in Grade 2 would be expected to read 40 of these words correctly. Sixty seconds were allocated for the timed story reading, and another 120 seconds (2 minutes) were given for the learner to read further to understand the full text. An auto-stop rule was applied if the learner could not read the first 5 words. After 60 seconds, the researcher stopped the timer and the tablet registered how many words the learner had read correctly, while the learner continued for another 2 minutes to finish reading the text.

As expected, Grade 1 learners scored very poorly in this sub-task (Figure 16), with even learners in the 75th percentile not managing to read a single word in these sentences. However, learners in the 90th percentile were able on average to read 37 words out of the 49 words.

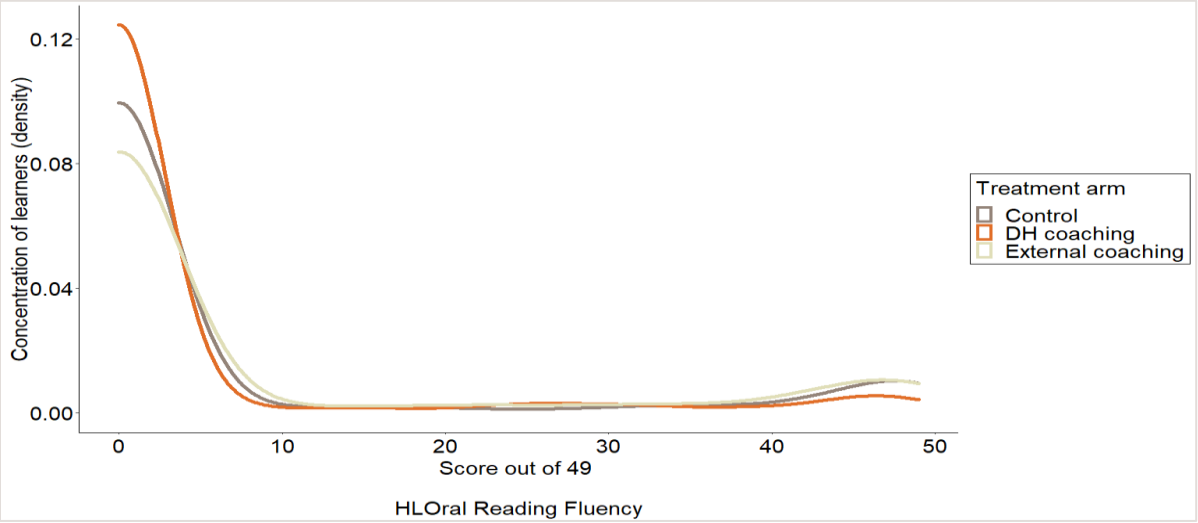

Figure 16: Grade 1 Kernel Density Curve for sentence reading by intervention arm

Oral reading fluency portrays a floor effect as reflected in Figure 17, with about 80% of the learners unable to correctly read any of the first five words resulting in the massive drop in participation due to the auto-stop rule. On average, the number of words correctly read in the sentence is below 20%. The learners also struggled with words longer than five letters, such as "*tshwanetse"* and "*sengwe*". This is problematic as Setswana has many compound words.

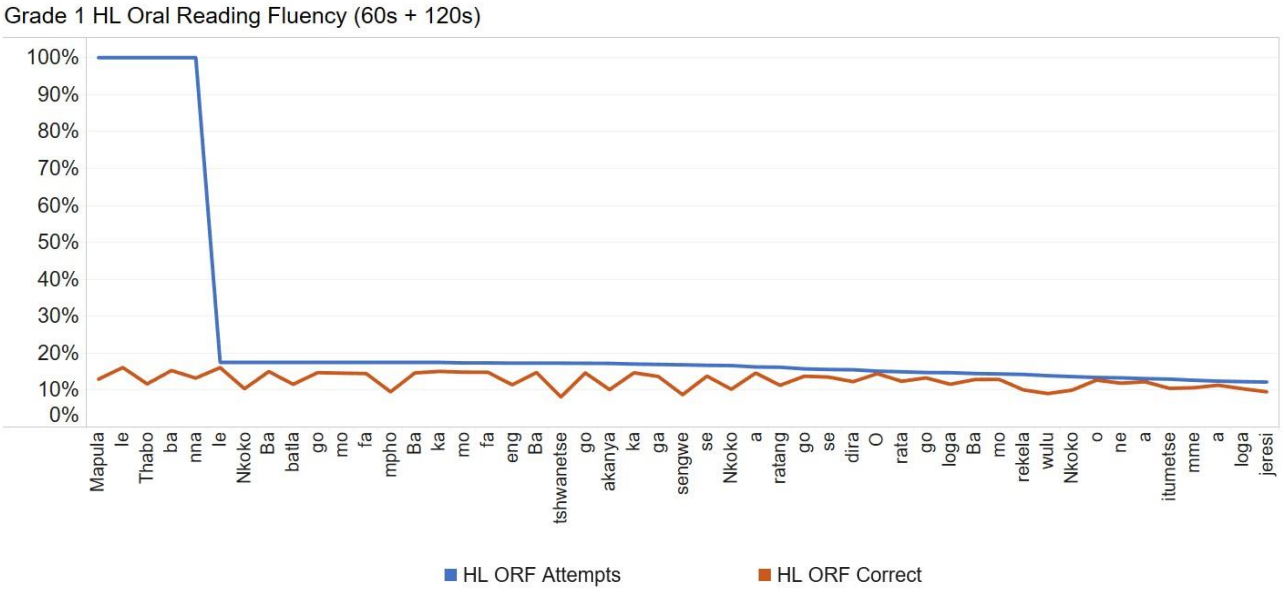

Figure 17: Grade 1 Number of words correctly read in the oral reading fluency sub-task

On average, Grade 2 learners also scored poorly in the oral reading fluency sub-task, although as expected they performed better than Grade 1 learners. Figure 17 shows that the 50th percentile of learners could not correctly read a single word in any sentence. However, learners in the 75th percentile could correctly read 44 words (90%) within 60 seconds out of the maximum of 49 words, and the 90th percentile could read all the words correctly. This indicates that about a quarter of the Grade 2 learners are reading at the speed and accuracy in their home language which is at or above the benchmark for expected level of reading fluency.

This means learners performed poorly on this task on average, but those who can read did so fast and accurately. These results present an apparent schism in learner performance: a few learners can read fluently, and the vast majority cannot read at all with relatively few learners exhibiting average competence in reading. Figure 18 depicts this by two peaks in performance with one on the far left, and the other on the far right, which shows reading performance is skewed on both tails albeit with a lower density on the higher-performing side. The trend is similar across all the intervention arms, but with the DH coaching schools continuing to under-perform relative to the external coaching schools.

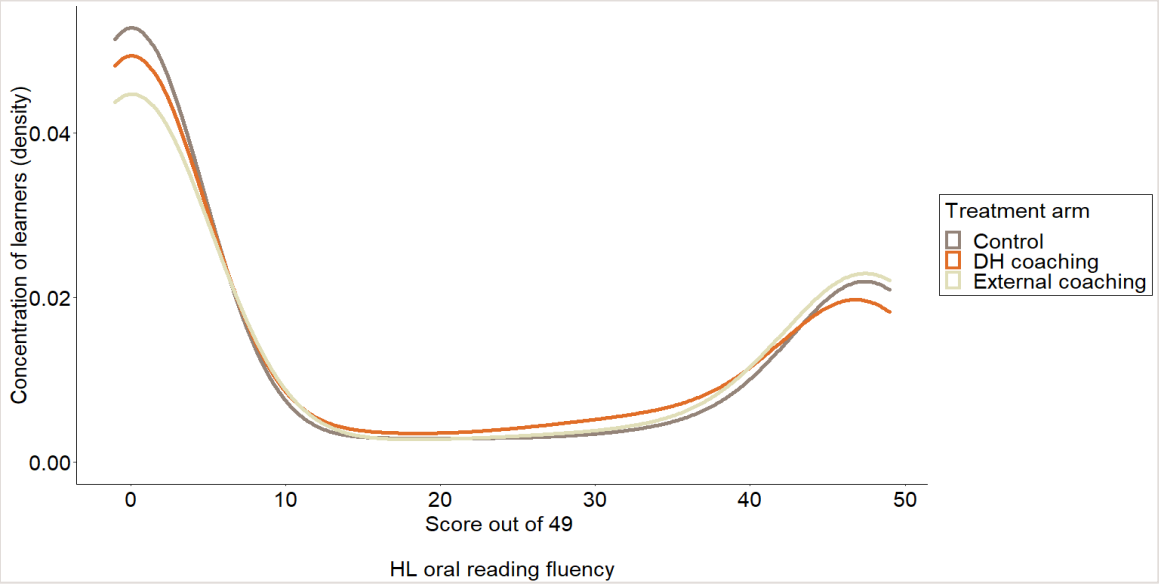

Figure 18: Grade 2 Kernel Density Curve for sentence reading by intervention arm

Figure 19 further illustrates that Grade 2 learners had a consistently poor performance in oral reading fluency, with the number of words correctly read being slightly below 40%. About 60% of learners could not read any word correctly. Similar to Grade 1 learners, Grade 2 learners also struggled with words that were longer than five letters, such as “tshwanetse’, “sengwe”, and “itumetse”.

This poor performance in oral reading fluency for both Grade 1 and 2 is not surprising considering that learners performed poorly in letter sound and word recognition, which positively correlate with reading fluency. This may also indicate the lack of exposure of learners to reading tasks in the classroom, especially for Grade 1 learners. This was observed in the case study conducted in 2021 in six schools participating in the EGRP intervention, which also indicated that in some of the case study schools, teaching was still teacher-centred with very little learner practice and participation. The loss of classroom time related to COVID-19 and rotational learning may have compounded the poor performance in reading and compromised the delivery of the curriculum as originally designed. This was confirmed by 48% of the teachers surveyed who reported that they could not cover the planned curriculum due to the COVID-19 situation and related disruptions. The effect of this is that scaffolded learning and systematic teaching of reading as envisaged by the EGRP ToC has been compromised in many schools.

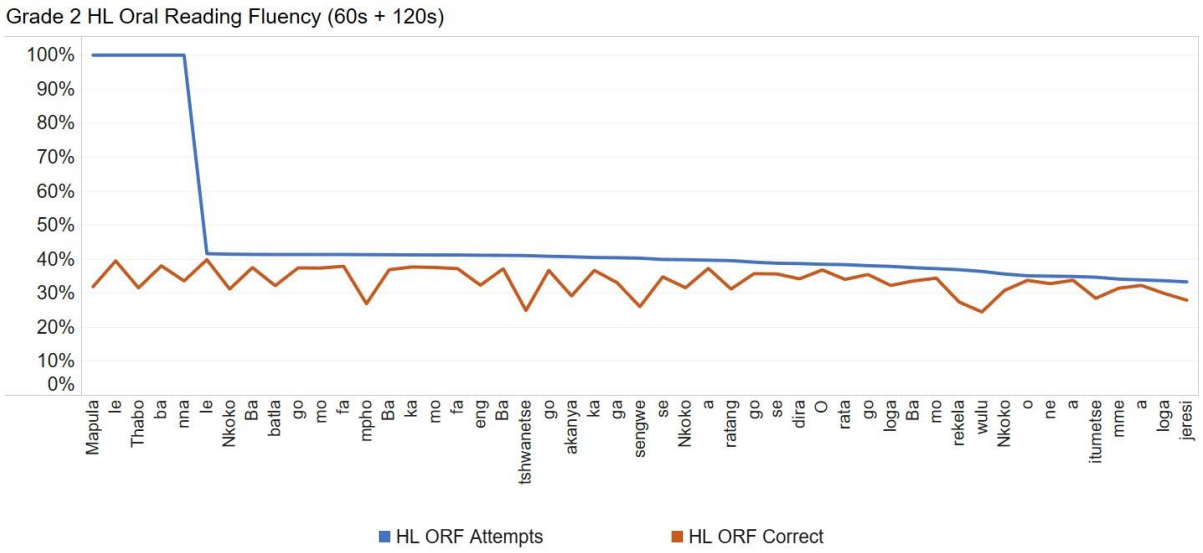

Figure 19: Grade 2 Number of words correctly read in the oral reading fluency sub-task

Sub-task 6: HL Comprehension

By design, the comprehension sub-task was dependent on the oral reading fluency sub-task. Learners were thus only asked questions for the parts of the story they could read in the oral reading fluency sub-task. So, if a learner failed to read the first line, they were not asked any comprehension questions at all. Field researchers were trained only to ask the first question if they read the first line only. Similarly, if they read the first two lines, they were asked questions 1 and 2, but not questions 3,4 and 5.

This performance correlated with the reading task, which was also poor. Figure 20 shows that the 75th percentile of Grade 1 learners did not answer any comprehension question correctly, whilst the 90th percentile of learners only answered one question correctly.

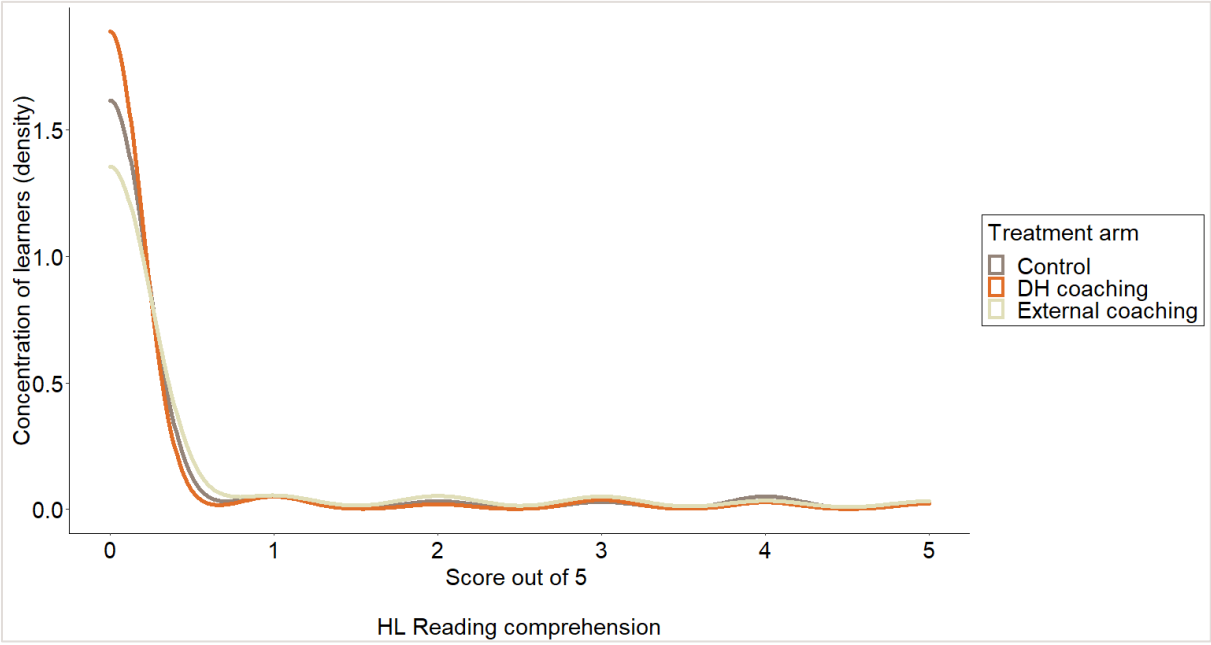

Figure 20: Grade 1 Kernel Density Curve for HL reading comprehension by intervention arm

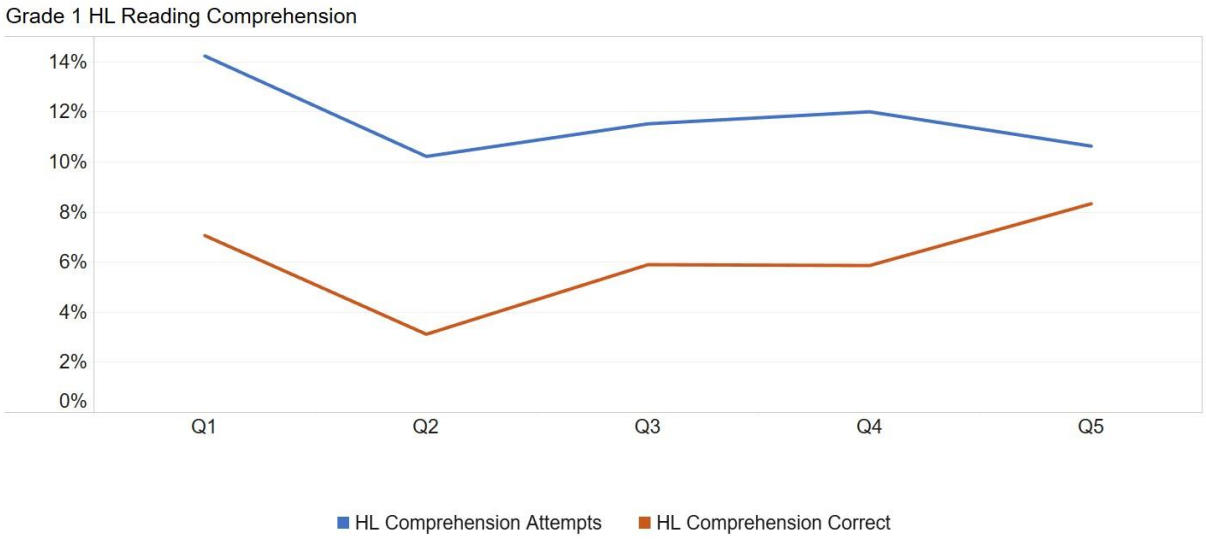

Figure 21: Grade 1 Number of HL comprehension questions correctly answered

Likewise, the comprehension sub-task for Grade 2 learners depended on the HL oral reading fluency sub-task. Similarly, the Grade 2 performance on this sub-task was also poor as shown in Figure 22 with the 50th percentile of learners not asked any comprehension questions because they could not read the text presented to them in the prior oral reading fluency sub-task. However, the 75th and 90th percentile of learners correctly answered 2 and 4 comprehension questions respectively out of a maximum of 5.

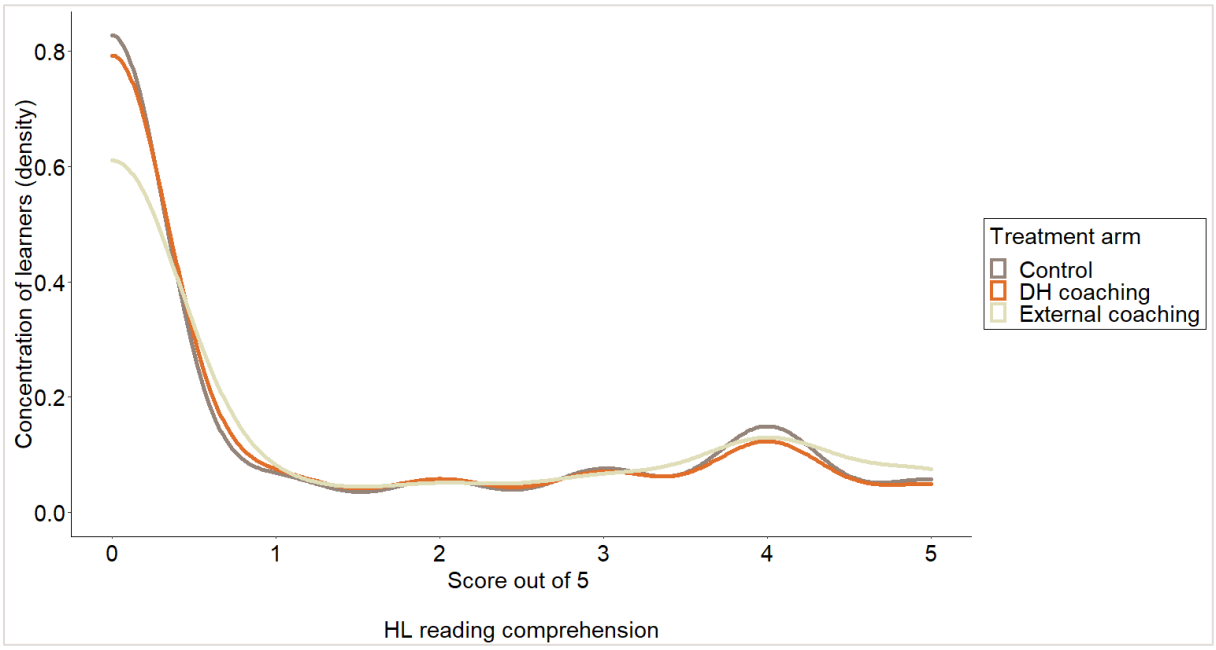

Figure 22: Grade 2 Kernel Density Curve for HL reading comprehension by intervention arm.

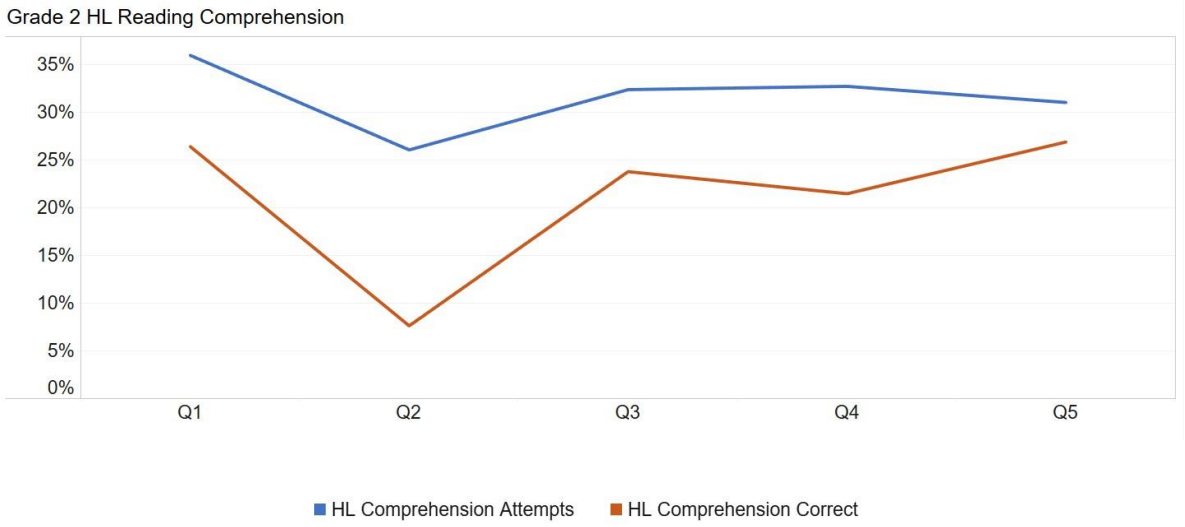

Figure 23: Grade 2 Number of comprehension questions correctly answered

Sub-task 7: EFAL Oral Reading Fluency

Grade 2 learners were tested with EFAL oral reading fluency tasks similar to those for HL. The same rules and time allocation in HL were applied to these sub-tasks in EFAL. The only difference was that the story they read was different in EFAL, consisting of 62 words. On average, the learners performed poorly on this task. For instance, Figure 24 shows that the 50th percentile of learners could not read a single word, whilst the 75th percentile of learners could only read 3 words out of a maximum of 62 words in the text. However, learners in the 90th percentile were able to correctly read 51 (82%) words out of the maximum of 62 words. This is well over the 30 CWPM which is the benchmark for EFAL fluency in Grade 2. Learners in the control group performed worse than learners in the other intervention arms.

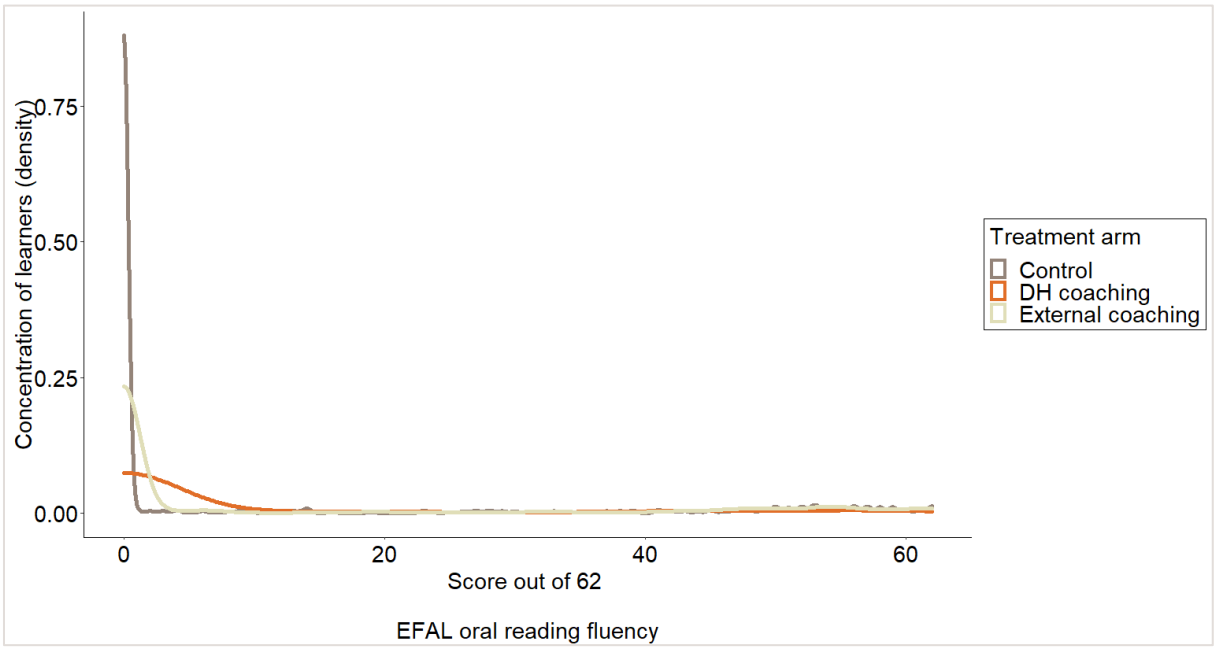

Figure 24: Grade 2 Kernel Density Curve for oral reading fluency by intervention arm

The EFAL Grade 2 oral reading fluency assessment (Figure 24) indicated a poorer performance than the same students performed in HL's oral reading fluency assessment as 20% or fewer learners could correctly answer the questions on the sub-task.

Figure 25 illustrates that only just over 20% of the Grade 2 learners could read any English words from the oral reading text. As with the HL oral reading text, learners struggled to read longer words, such as “walked”, “chased” and “called”.

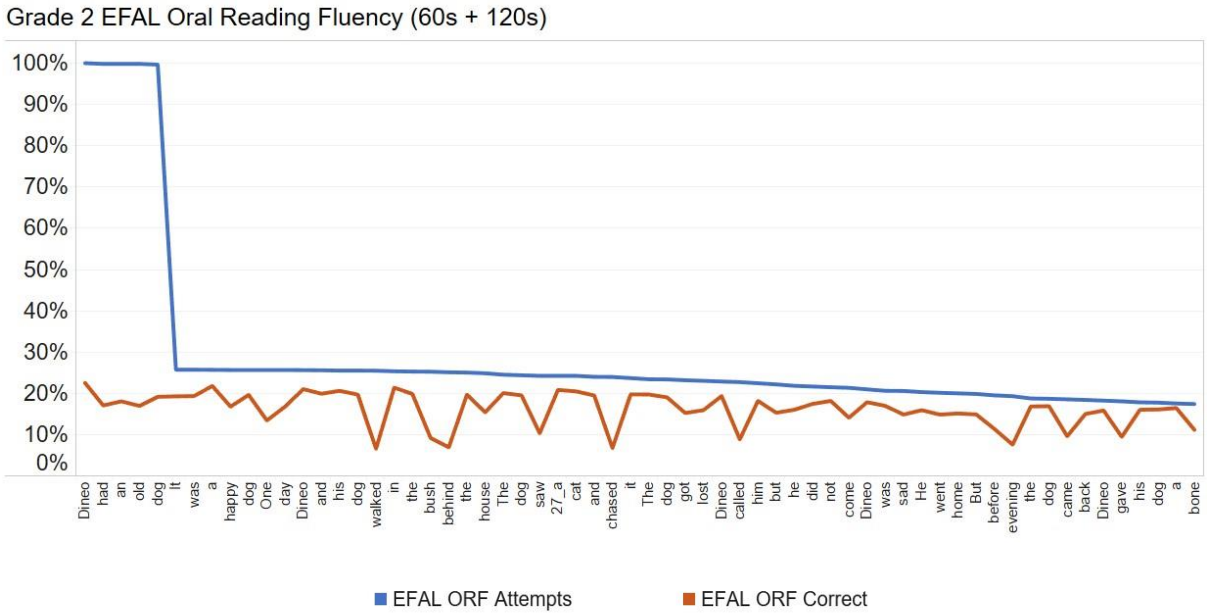

Figure 25: **Grade 2 Number of words correctly read in the EFAL oral reading fluency sub-task**

Sub-task 8: EFAL Comprehension

Like the comprehension sub-task in HL, the EFAL reading comprehension sub-task was dependent on the learner's performance in the EFAL oral reading fluency sub-task. This means that learners were only asked questions for the parts of the EFAL story they had managed to read. Since most learners could not read a single word, they were not asked any comprehension questions. This is illustrated in Figure 26. Only about 30% of learners were asked even one question, and Figure 27 shows that even if they were asked the question, only about a third (on questions 1 and 5) to a half (on the other questions) of learners got the answer correct. This is in contrast to the HL comprehension exercise, which showed that if the learner could read the text, they could understand the story and so answer the questions correctly. In EFAL, more than half of the learners who could read the words did not seem to understand the story so they could not answer the questions correctly.

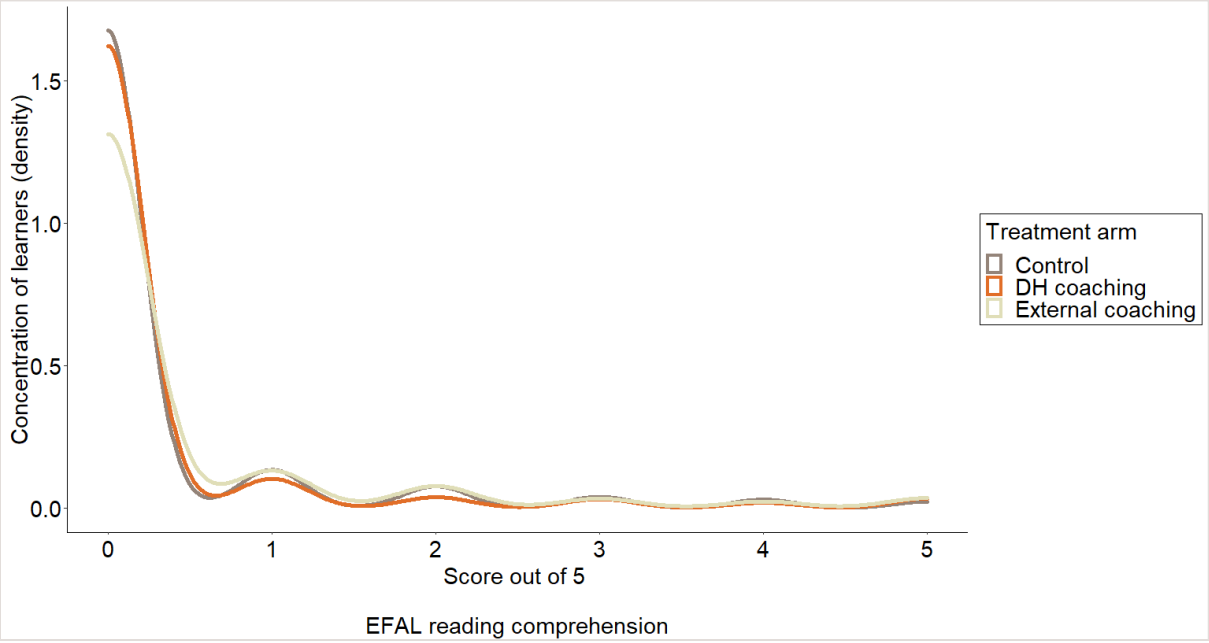

Figure 26: Grade 2 Kernel Density Curve for EFAL reading comprehension by intervention arm

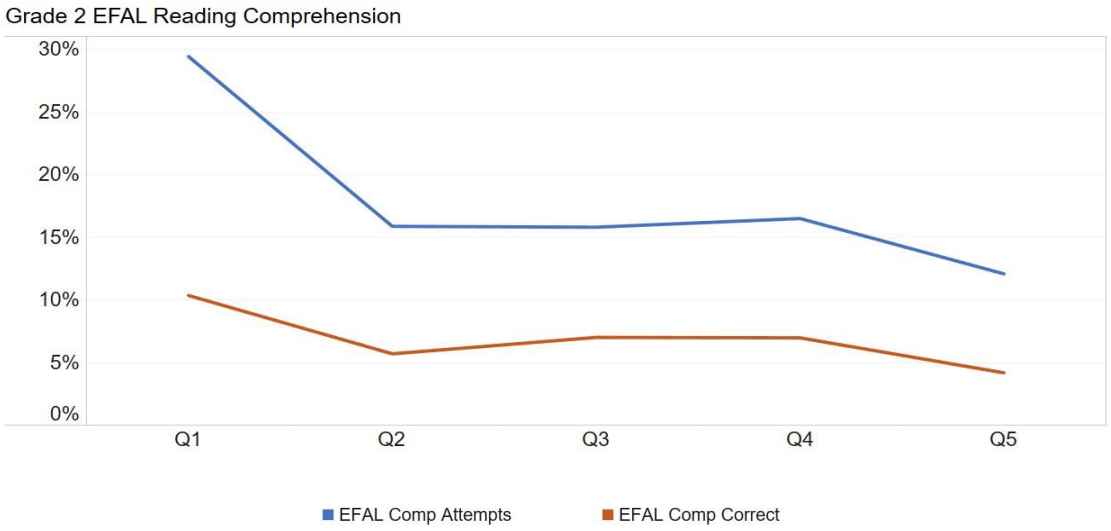

Figure 27: Grade 2 Number of EFAL comprehension questions correctly answered

Sub-task 9: EFAL Word sound knowledge

This was a group-administered sub-task where the Grade 1 learners were required to individually identify and circle on their scripts EFAL words read out to them by a field researcher. In this task, the learners were read five words in turn. For each word, the learners had five options to choose from on their scripts. The learners were allowed 20 seconds to circle the correct item before the field researcher read out the next word.

Figure 28 indicates that learners performed relatively well on this sub-task, with the 50th percentile of learners scoring 2 out of 5 words correctly, and the 75th percentile of learners scored 4 out of 5. It is worth noting that learners in the DH coaching intervention group on average did worse than learners in the external coach intervention and even the control groups. This is the same as with a number of the sub-tasks.

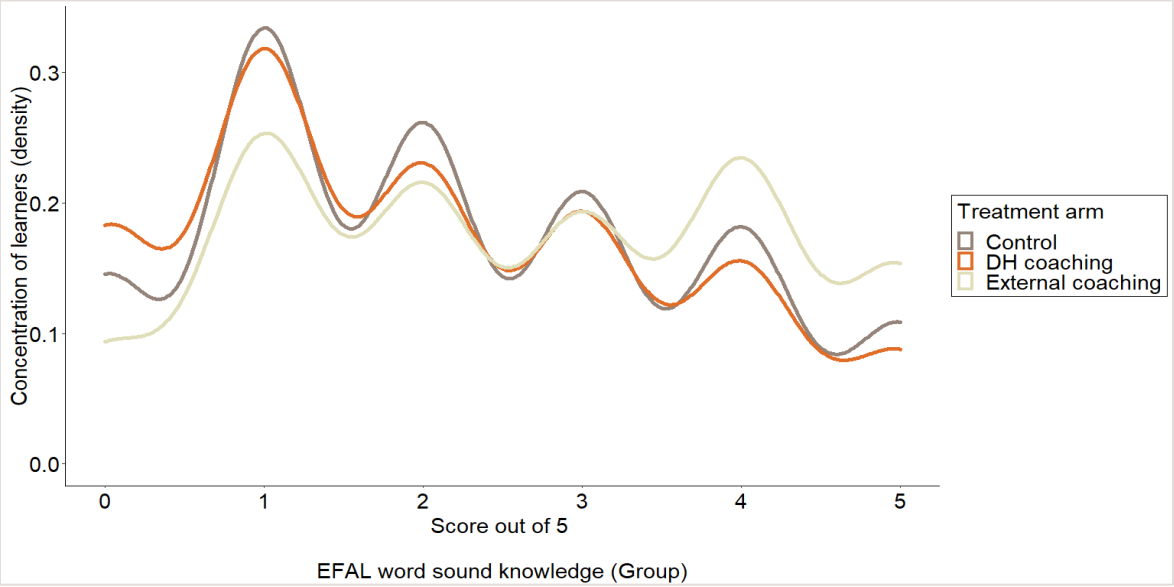

Figure 28: Grade 1 Kernel Density Curve for EFAL word sound knowledge by intervention arm

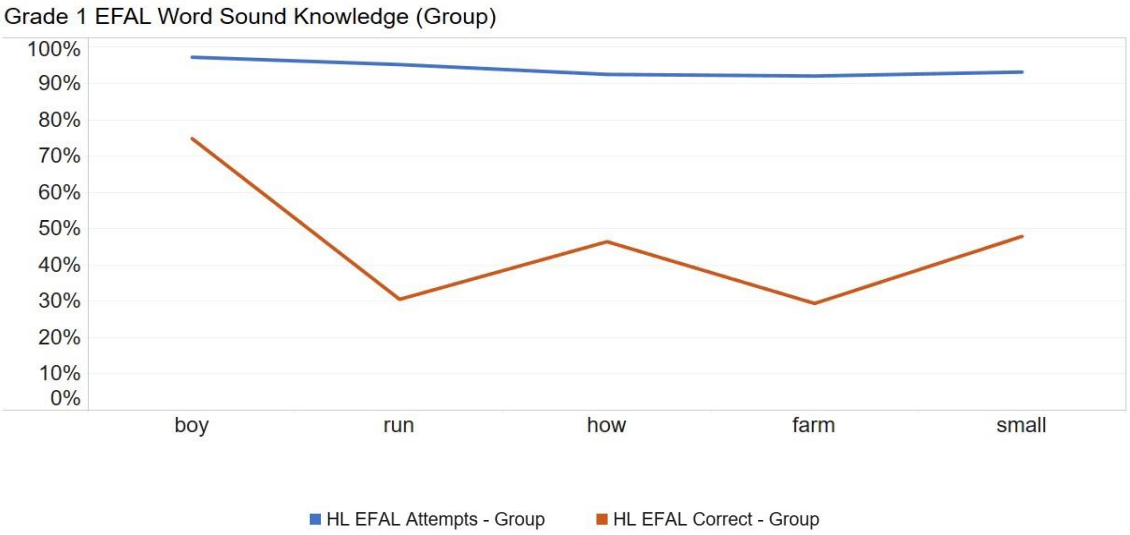

Figure 29: Grade 1 Number of EFAL words sounds correctly identified

A similar assessment was administered to Grade 2 learners as a group sub-task, but this time the field researcher read out 10 words in EFAL, and the learners had 20 seconds to circle the correct items in turn.

From Figure 30 it can be seen that on average, Grade 2 learners performed comparatively well on this sub-task, with the 50th percentile of learners scoring 6 out of 10 words correctly. The 75th percentile of learners scored 8 out of 10. The density curve is right-skewed, implying that most learners did well on this sub-task.

There may be a number of reasons why learners generally performed better in the group sub-tasks compared to the individual sub-tasks. The obvious answer is that learners in a group setting were able to cheat. However, given that many learners were shown in other tests to be

poor readers copying from each other is as likely to produce wrong answers as correct ones. So, the most likely explanation is (i) that pure guessing will score an average of 2 out of 10; (ii) learners feel more comfortable with this sort of test (multiple choice in a classroom with their friends) than being sat with a stranger doing one-on-one tests. The latter point has been noted in other countries – where learners are tested by their own teacher or in groups they tend to attempt more answers and do better than being tested alone by strangers.

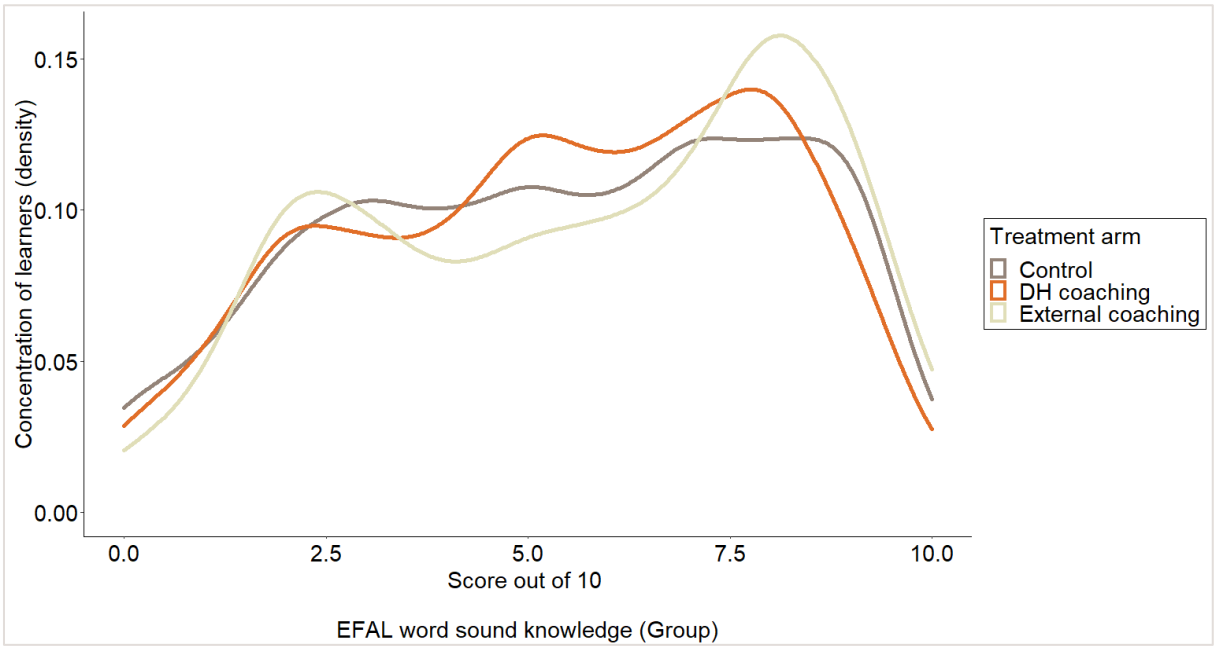

Figure 30: Grade 2 Kernel Density Curve for EFAL word sound knowledge by intervention arm

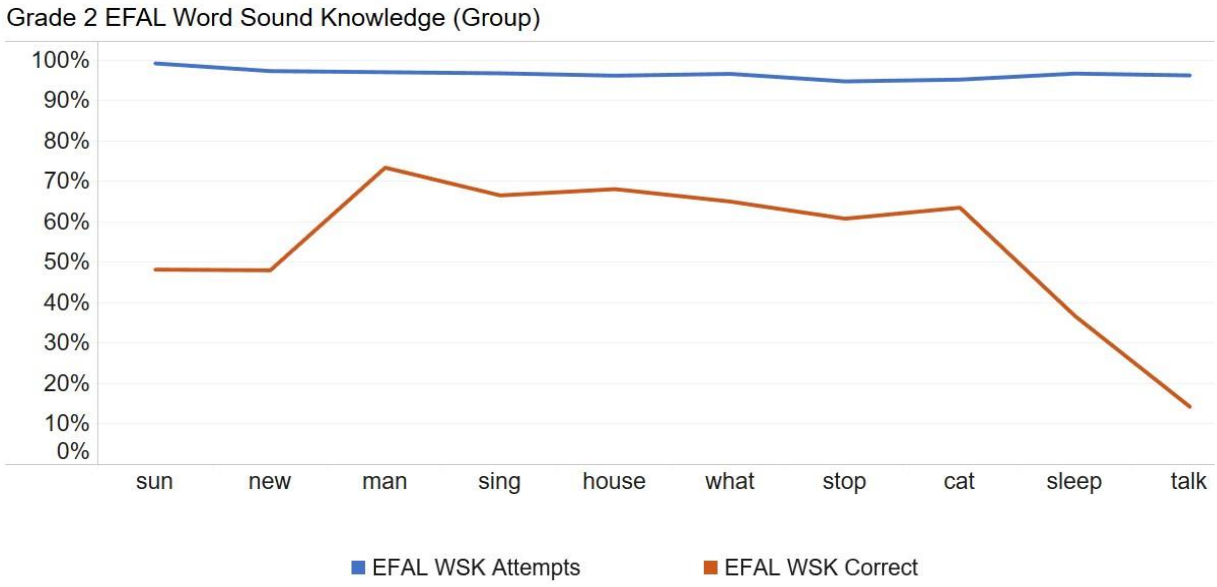

Figure 31: Grade 2 Number of EFAL words sounds correctly identified

Sub-task 10: HL Complex Consonant Sound Knowledge

This was a group-administered sub-task where Grade 2 learners were required to individually identify and circle HL complex consonants on the scripts that they were given, as the field researcher read them out. In this task, learners read the sound of five complex consonants in

HL. On their script, they had five options to choose from and 20 seconds to make their choice by ringing the answer.

On average, learners performed well on this sub-task (Figure 32), with the 50th percentile scoring 3 out of 5 complex consonants correctly and the 75th percentile of learners getting 4 out of 5 correct. Figure 33 shows that the level of attempts was above 90% for each consonant read, and on average, the number of learners that correctly identified the consonant sounds was 60%. Not surprisingly, the learners found the two letter compounds easier to identify correctly than the three letter compounds.

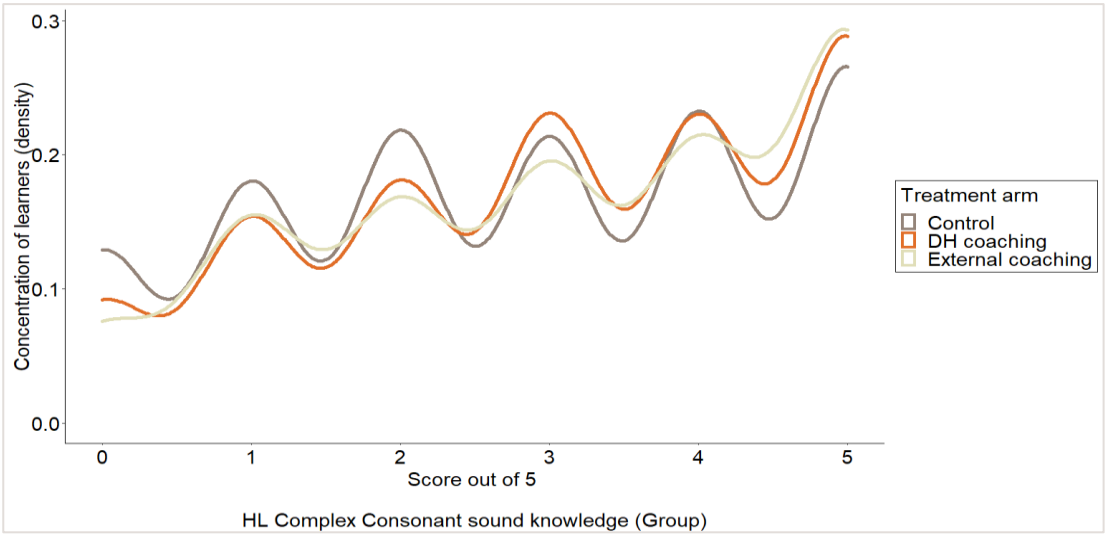

Figure 32: Grade 2 Kernel Density Curve for HL Complex Consonants sound knowledge by intervention arm

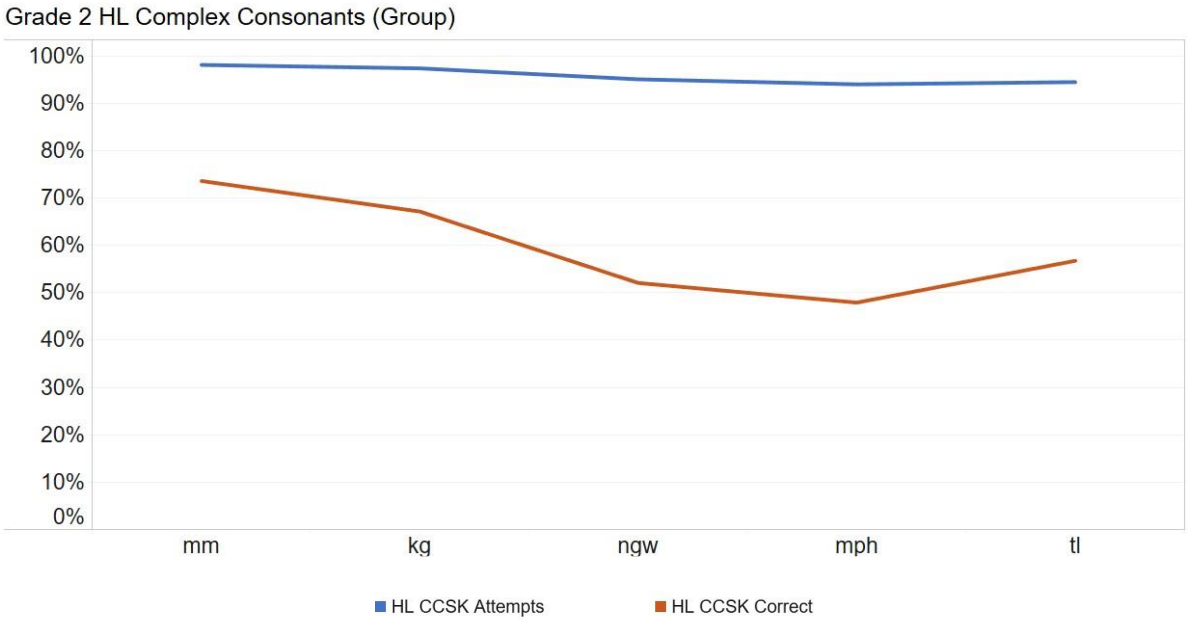

Figure 33: Grade 2 Number of HL Complex Consonants correctly identified

4.2.3 Composite Index Analysis

Using principal component analysis (PCA), the different sub-task performance was combined to give a single composite index. For Grade 1 learners, the first principal component accounts for 50.3% of the variance out of all components. Although the explanatory power of the first principal component is not as high, the overall composite score is still useful as it represents a single picture of the Grade 1 learner's performance.

The first principal component is used as an overall composite score since it tends to account for the majority of the explanatory power (50.3% in this case) in the individual subtask scores, the second principal component usually captures even less of the explanatory power in the subtest scores (i.e. about 16% of explanatory power in this case), and the subsequent principal components account for even lesser of the variation. The first principal component thus serves as a single concise index to measure overall learner performance. More importantly, the performance of Grade 1 represented by this composite score across intervention arms, depicted in Figure 34, is that it is a standardized score (with a zero mean and standard deviation of 1) with a long left tail (left-skewed). The standardized composite score enables ease of interpretation and comparison of performance across the grade.

A left-skewed distribution, implies that the average composite score was less than the median composite score of Grade 1 learners (Note: mean = median for normally distributed scores), which means that some of the learners attained worse scores than expected. The distribution of the standardized composite score was similar irrespective of the intervention arm.

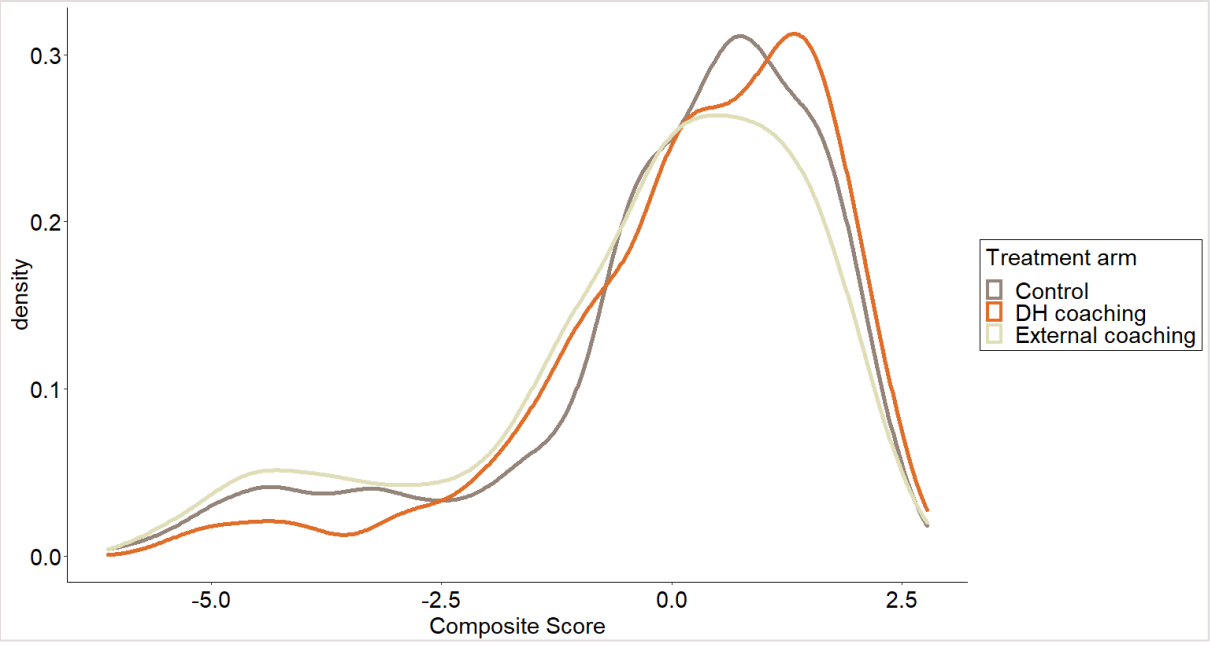

Figure 34: Grade 1 Overall learner performance distribution by intervention arm

For Grade 2 learners, the same PCA analysis was used to combine the sub-tasks; the first Principal component accounts for 63% of the variance out of all the principal components. This first Principal component thus captured many variations, implying that the different sub-tasks

for Grade 2 learners were coherent and tended to measure the same underlying task, such as reading/comprehension skills.

A higher variance for the first principal component also implies that test scores per learner were correlated. For example, if a Grade 2 learner performed well in HL letter-sound knowledge, they were more likely to perform well in reading in their home language. Likewise, the composite scores were standardized. Figure 35 shows the performance distribution of all Grade 2 learners across intervention arms. The distribution is right-skewed (i.e., with a long right tail), which implies some Grade 2 learners attained better scores than expected.

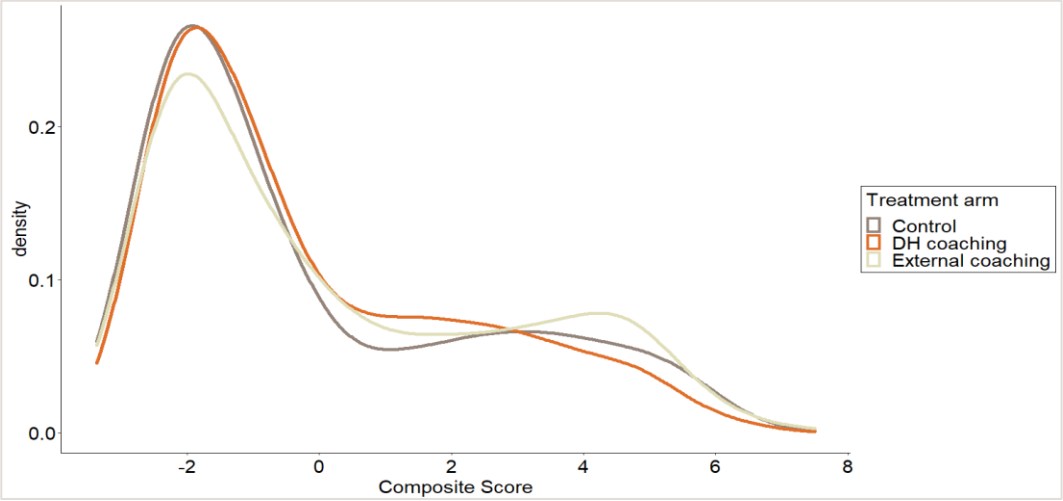

Figure 35: Grade 2 Overall learner performance distribution by Intervention

4.2.4 Evaluation of Intervention arms on performance (after one year of implementation) and Balance Test

Schools enrolled in this EGRP programme may differ in observable and unobservable ways, which might impact or bias the study outcomes. Random allocation of schools to intervention groups ensures that selection bias is reduced and the allocation of schools to either intervention or control groups is balanced. The fundamental assumption of RCT sample design is to ensure that no intervention group or control schools are inherently biased.

SSA ran a battery of balance tests to understand the effect of sampling variables on learner performance. The variables tested were (i) the impact of the Primary School Reading Improvement Programme (PSRIP), which is operational in some of the sample schools; (ii) DH commitment; and (iii) controlling for strata. The test results indicated a fairly balanced distribution of performance by these sampling variables. Additionally, since the rapid object naming (RON) subtask is a cognitive measure and not necessarily an outcome of the programme, the RON subtask performance was used as a proxy to measure if the sample was fairly balanced across the intervention arms. From both Table 27 and 28, we can see that for the RON subtask, there is no statistically significant result across intervention arms for both Grade 1 and Grade 2 learners; therefore, we conclude that the sample is fairly balanced across the intervention arms.

As articulated before, this assessment has been done after some learners have been exposed to the core programme (teacher training and LTSM provision) and external coaching for a year. The following checks are thus done to test if learners' exposure to the intervention arms at midline has resulted in any observable differences in learner performance for each sub-task. If there are any differences, one needs to test for statistical significance per sub-task. This will be achieved by comparing the interventions to the control group by learner performance.

For Grade 1 learners, Figure 36 presents learner performance on the different sub-tasks by intervention group versus the control group. Some slight differences are observed in the HL letter-sound knowledge and HL oral reading fluency; however, the other test results look like there is no difference in performance across intervention groups.

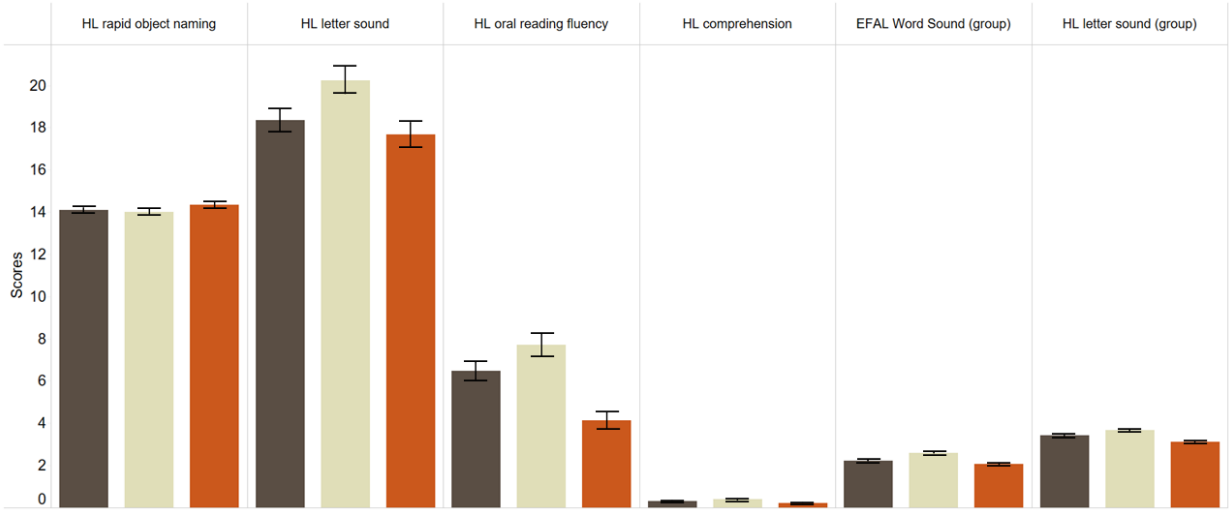

Figure 36: Grade 1 sub-task performance by intervention

Table 27 reports results that test if these differences are statistically significant or due to random chance. Each column in the table represents a separate regression run for each sub-task on intervention indicators, controlling for strata and district fixed effects. In the regression analysis, standard errors were also clustered at the school level. The numbers in parentheses for both interventions represent the standard errors, whereas the numbers above the parenthesis represent the mean differences (regression coefficients) obtained from regression analysis by comparing the control to the intervention groups. Stars represent the significant differences in the table. One star will indicate that the difference is significant at 10%, two stars indicate a significance level of 5%, and the three stars indicate that the difference is significant at a 1% level.

The results show statistically significant differences in learners' performance in both the group-administered ‘HL Letter sound knowledge’ and the individual administered ‘HL reading fluency’ task (only significant at 10%) for the DH coaching cohort relative to the external coaching group. Similar significant differences are observed for the group administered ‘EFAL word sound knowledge’. This result is unexpected since the DH coaching cohort received the same intervention as the external coaching arm, using the same coaches.

The row before the final row displays the p-value for the pair-wise comparison test comparing the means between the two intervention arms (i.e., external versus DH coaching). A p-value of less than 0.05 will indicate that there is a statistically significant difference between the two interventions for the specific learning subtask. With the exception of Rapid Object Naming

(which is a pre-reading task which is not taught) and HL letter sound knowledge subtask, the remaining subtask showed a consistent significant difference between the intervention groups, with the external coaching group having performed consistently better than the DH coaching group. This result is unexpected since both intervention groups received the same intervention at midline. This may point to EGRS implementation challenges that need to be investigated further to establish why this is the case.

Table 27: Grade 1 Intervention Assessment per sub-task

	Rapid Object Naming (HL)	Letter Sounds (HL)	Reading Fluency (HL)	Comprehension (HL)	Group Letter Sounds (HL)	Group Word Sounds (EFAL)
External Coaching	-0.08	0.63	0.72	0.07	0.18	0.31*
	(0.3)	(2.17)	(1.81)	(0.1)	(0.17)	(0.17)
DH Coaching	0.13	-1.57	-2.52*	-0.09	-0.44**	-0.28*
	(0.26)	(1.98)	(1.5)	(0.09)	(0.18)	(0.15)
Control mean	14.09	18.71	6.75	0.32	3.44	2.26
External = DH p-value	0.517	0.259	0.014**	0.067*	0.00***	0.00***
N	2631	2631	2631	2631	2631	2631

*Note: Each column represents a separate regression on intervention dummies and stratification dummies. Standard errors are clustered at the school level. Significance level: *for p<.1;**p<.0.5; ***p<.0.01. (standard error) in parentheses*

For Grade 2 learners, Figure 37 presents learner performance on the different sub-tasks by comparing the control group to intervention groups. This shows differences in the HL letter-sound knowledge and oral reading fluency (i.e., both HL and EFAL). However, further significant tests show that none of these results is statistically significant (Table 28); after controlling for strata and the intervention arms and clustering standard errors at the school level. This result implies that after one year of exposure, there have not been any significant differences in Grade 2 learner performance in the intervention schools compared to those in the Control group.

Likewise, for Grade 2 learners, the row before the final row in Table 28 displays the p-value for the pairwise comparison test comparing the means between the two intervention arms (i.e. external versus DH coaching). A p-value of less than 0.05 will indicate that there is a statistically significant difference between the two interventions for the specific learning subtask. With the exception of HL word recognition & writing and EFAL reading fluency which presents statistical significance, the remaining subtask show that after 1 year of implementation, there has been no difference in performance with regards to the intervention arms versus the control group.

In summary, after one year of intervention, one would expect to see some differences between the control and intervention groups. Similar to Grade 1 results, this might be related to implementation challenges experienced in 2021. The implementation fidelity analysis will monitor this closely.

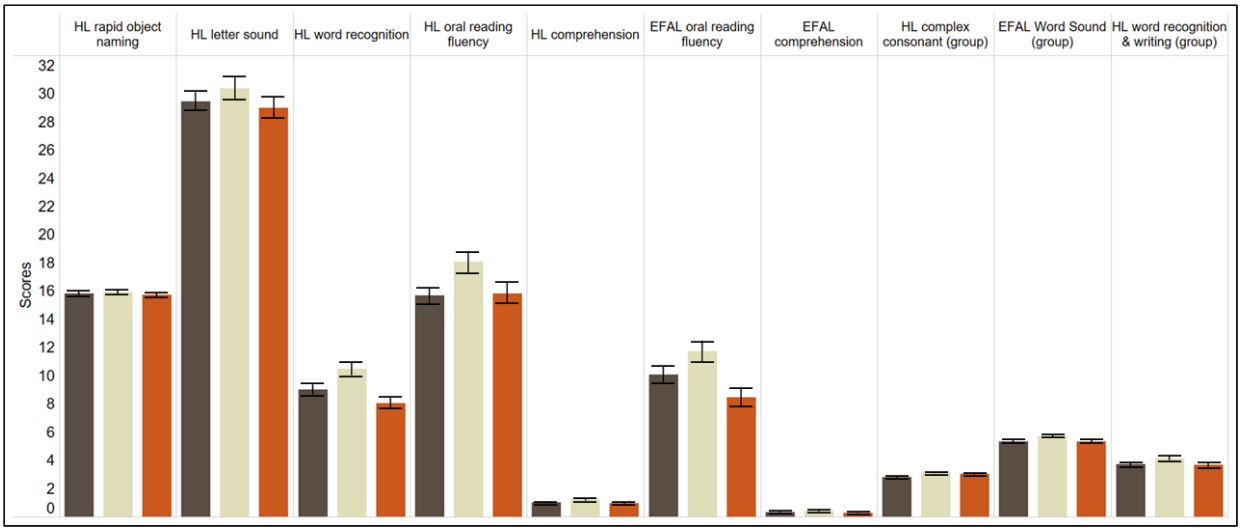

Figure 37: Grade 2 sub-task performance by intervention

Table 28: Grade 2 Intervention Assessment per sub-task

	Rapid Object Naming (HL)	Letter Sounds (HL)	Word Recognition (HL)	Reading Fluency (HL)	Comprehension (HL)	Reading Fluency (EFAL)	Comprehension (EFAL)	Group Complex Consonant (HL)	Group Word Reading /writing	Group Word Sound (EFAL)
External Coaching	0.09	0.94	1.51	2.46	0.22	1.99	0.12	0.31	0.44	0.40
	(0.31)	(2.26)	(1.14)	(2.03)	(0.16)	(1.64)	(0.08)	(0.18)	(0.43)	(0.30)
DH Coaching	-0.18	-0.45	-0.77	0.66	-0.02	-1.28	-0.007	0.23	0.01	0.05
	(0.3)	(2.11)	(1.05)	(1.94)	(0.15)	(1.51)	(0.08)	(0.18)	(0.39)	(0.28)
Control mean	15.85	29.18	8.83	15.39	1.00	9.84	0.32	2.82	3.67	5.31
External= DH p-value	0.355	0.45	0.029**	0.294	0.069	0.028**	0.061	0.544	0.252	0.187
N	2636	2636	2636	2636	2636	2636	2636	2636	2636	2636

*Note: Each column represents a separate regression on intervention dummies and stratification dummies. Standard errors are clustered at the school level. *for p<.1;**p<.0.5; ***p<.0.01*

4.3 Overall Learner Performance

Having noted the details of the sub-tasks and how the tests were administered and their validity as tests of learner performance and predictive capacity related to learner reading skills, this section looks at the overall performance of learners in the tests.

The learners fall into three categories: fluent readers who read at or above the benchmark for fluency, those who are beginning readers who cannot read with fluency, and non-readers who

cannot read a single word. The following graphics illustrate the proportion of learners who are attaining the expected level of fluency in HL and EFAL - that is they are considered fluent and have reached the benchmark. For Grade 1 the benchmark is the ability to sound out 40 correct letters per minute (40 CLPM) in HL. There is no standard for Grade 1s for reading words in HL or letters in EFAL. For Grade 2 the benchmark is 40 CWPM for HL and 30 CWPM for EFAL.

Figures 38, 39 and 40 show the results of this analysis.

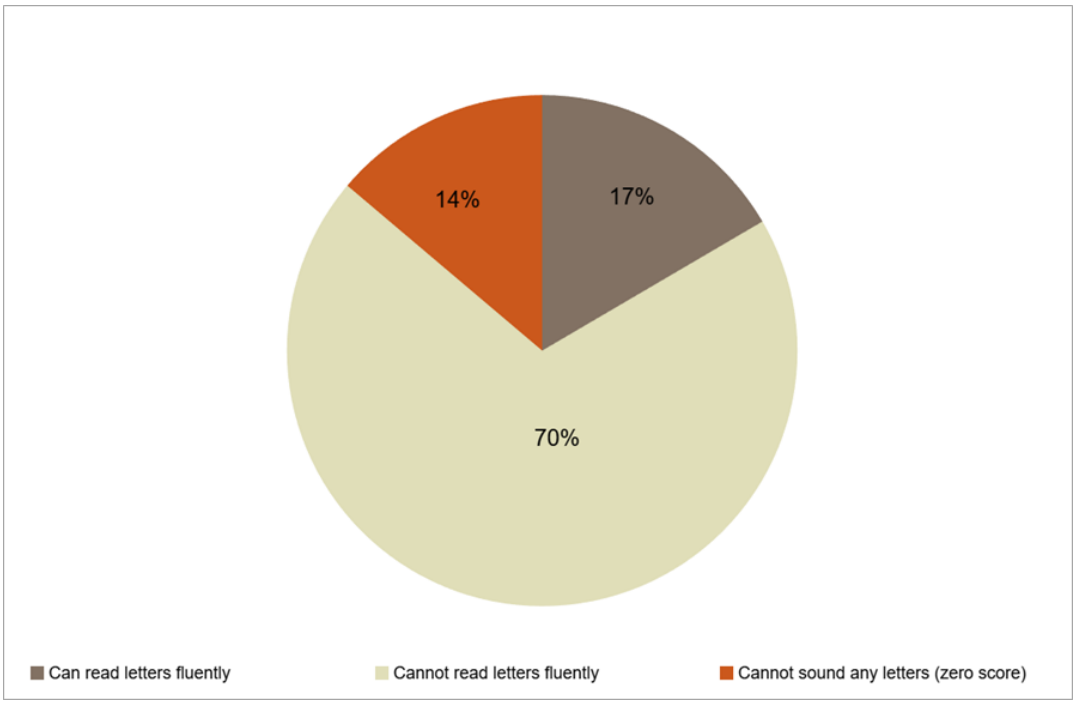

Figure 38: Grade 1 HL Letter Sound Knowledge by Gender (all intervention arms)

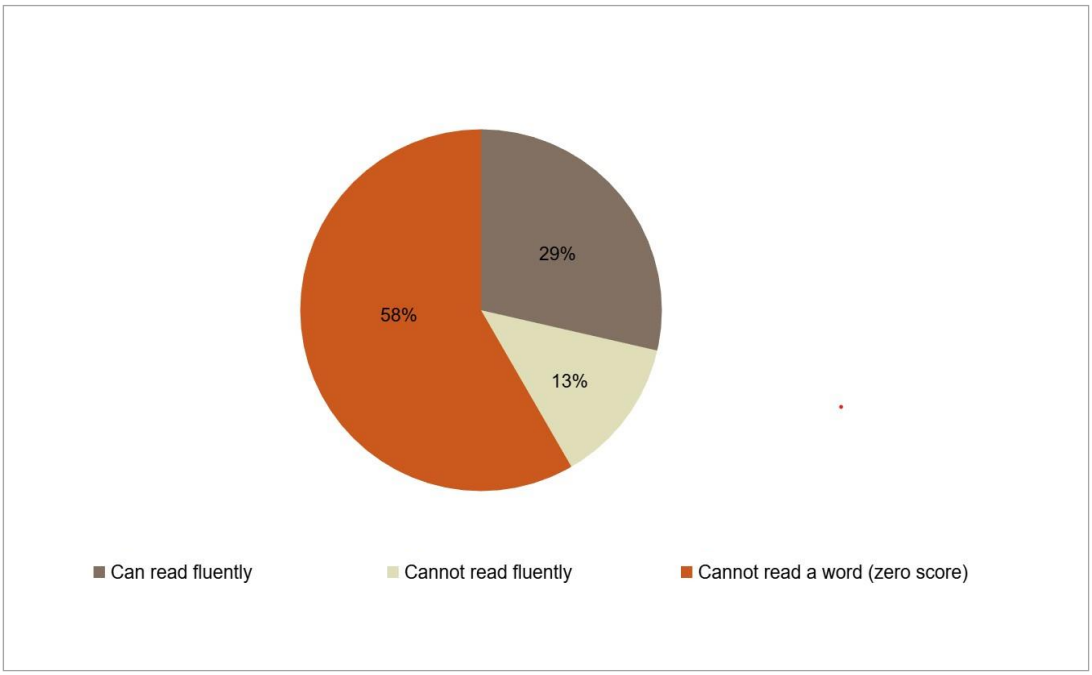

Figure 39: Grade 2 Oral Reading Fluency in Home Language (all intervention arms)

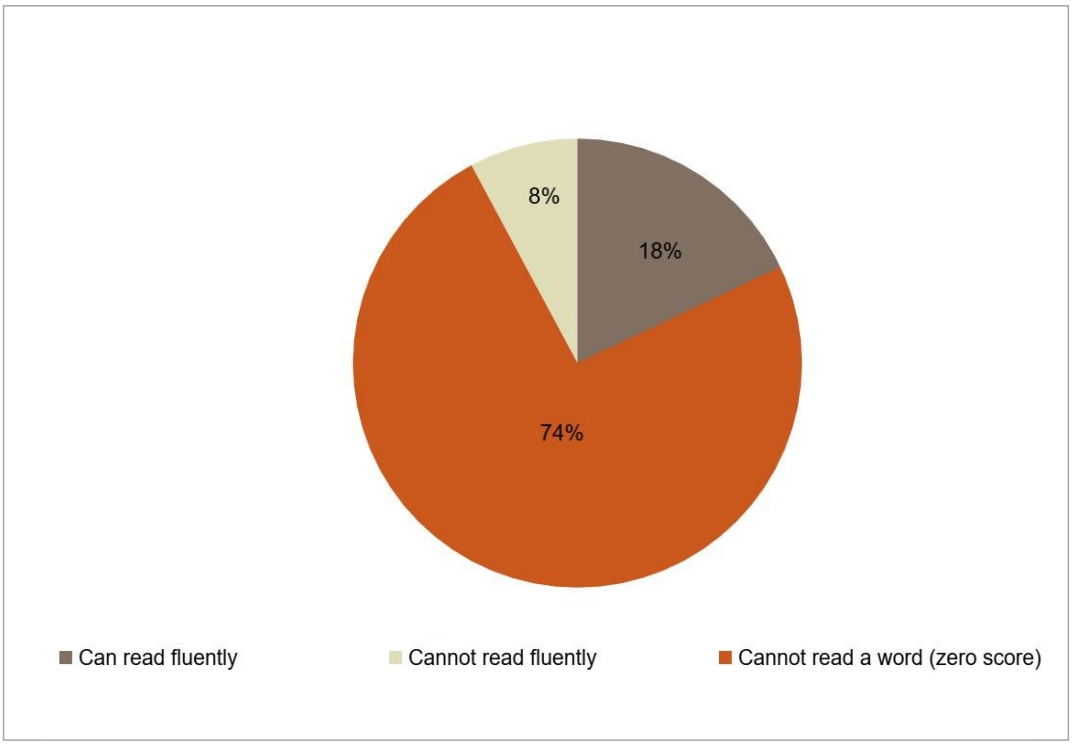

Figure 40: Grade 2 Oral Reading Fluency in EFAL (all intervention arms)

Not surprisingly, learner performance in their HL was considerably better than in EFAL. A quarter of the learners in Grade 2 across both the intervention groups and the control group could decode at least a few words in English and of these 18% had some competence in reading English.

Of particular concern is the number of learners who could not read a single letter or word – particularly in their home language. These are non-readers who after some time in school have not mastered the basics of decoding and so are unable to relate a sound to a written letter.

4.4 Influence of Gender on the Learner Performance

When the learner results are disaggregated by gender a clear pattern emerges girls in both Grades 1 and 2 outperform boys on every reading task. Interestingly the one task where the performance of boys is almost equal to that of girls is the Rapid Object Naming, which is a pre-reading recall task (which has not been taught) and not a reading task

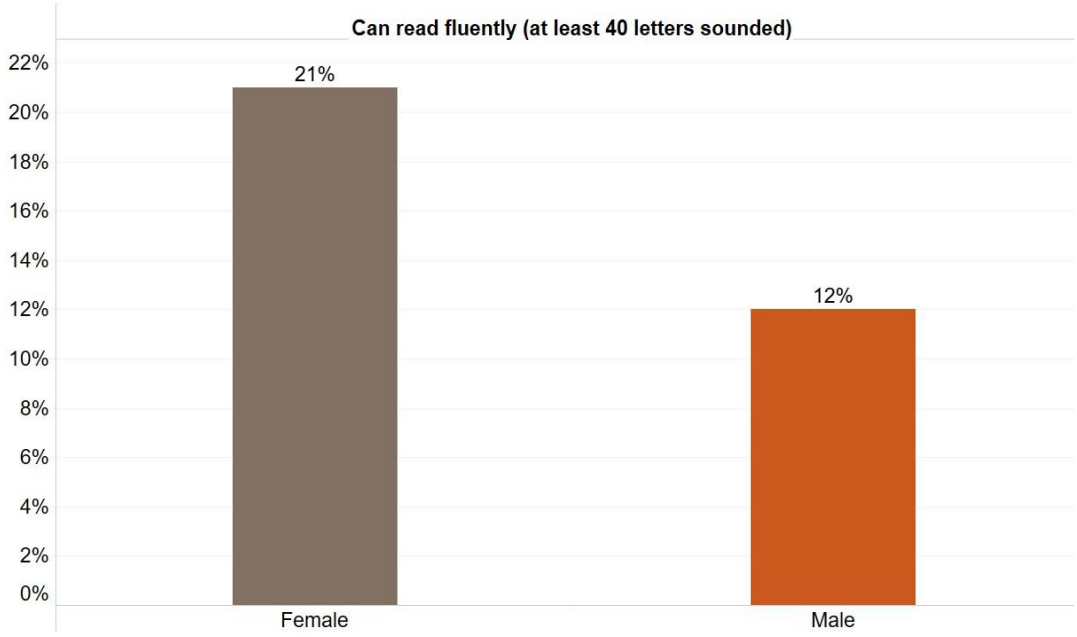

Figure 41: Grade 1 HL letter sound knowledge (all intervention arms)

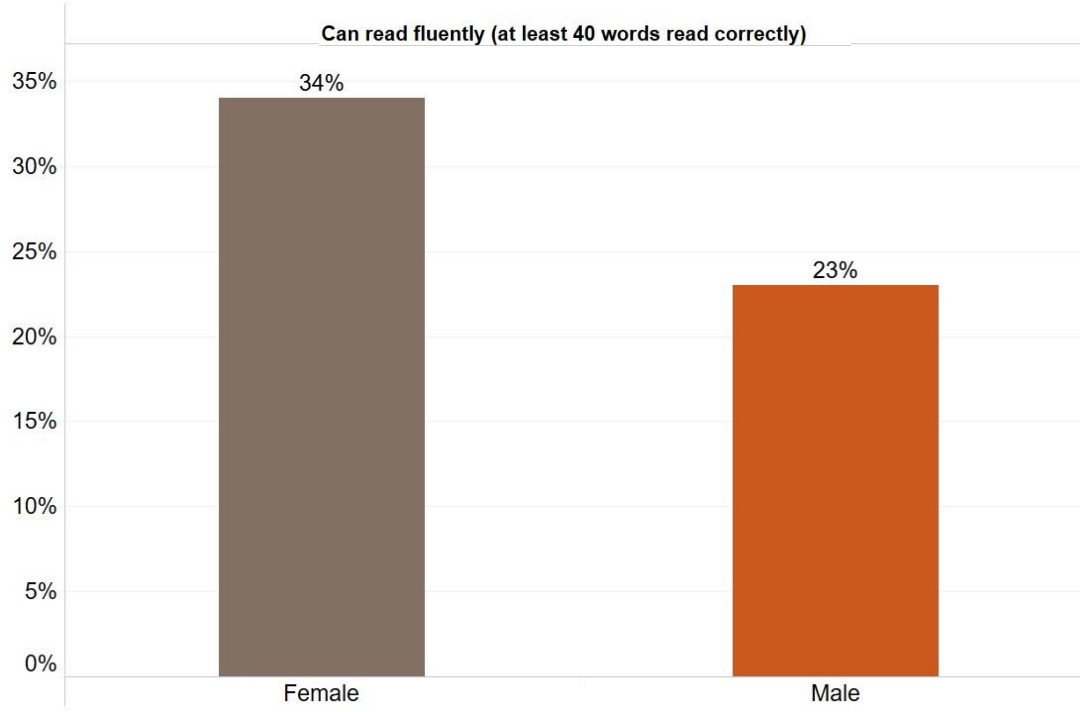

Figure 42: Grade 2 Oral Reading Fluency in HL by Gender (all intervention arms)

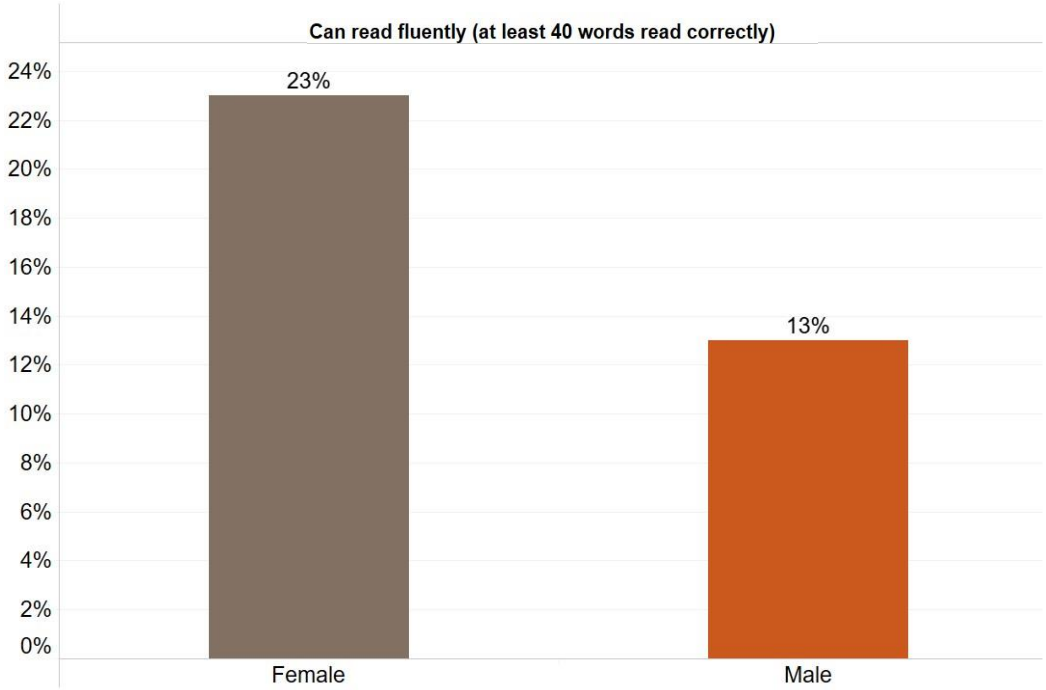

Figure 43: Grade 2 Oral Reading Fluency in EFAL by Gender (all intervention arms)

If we then look at the results from the perspective of the various intervention groups, the pattern remains consistent, as indicated in Figures 42, 43 and 44 across the grades and languages.

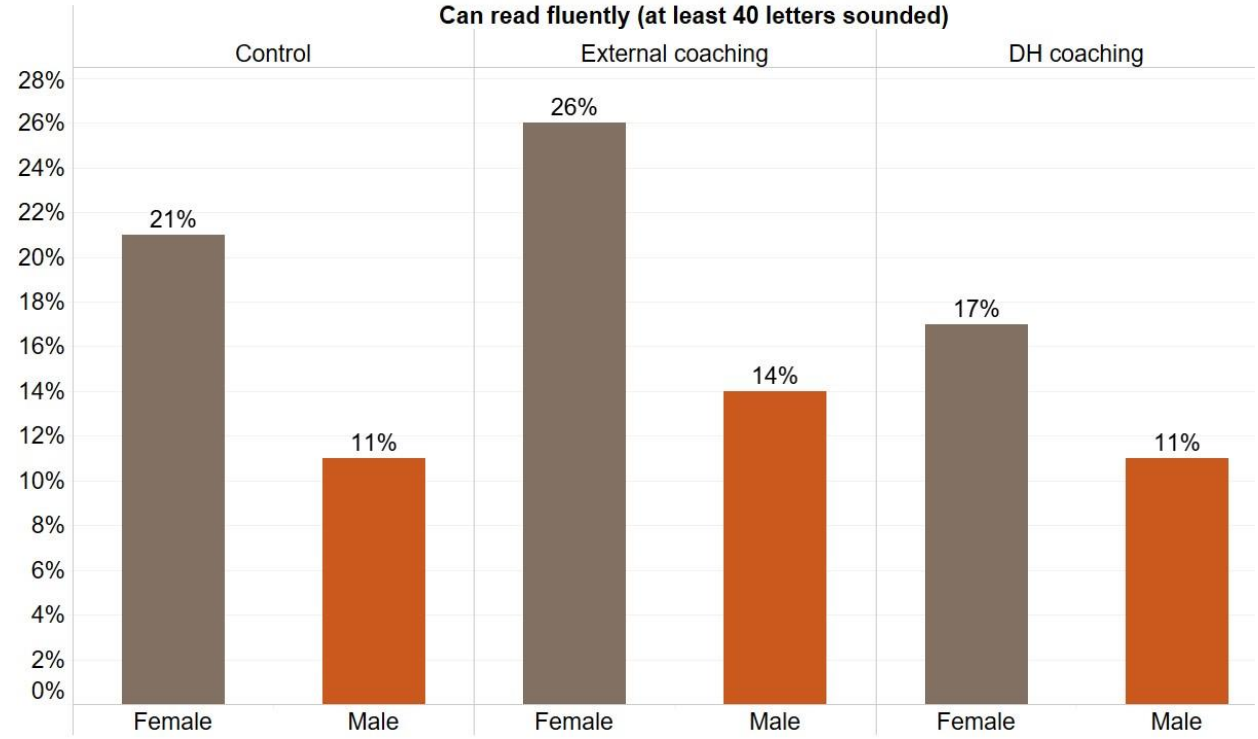

Figure 44: Grade 1 HL Letter Sound Knowledge by Gender and Intervention Group

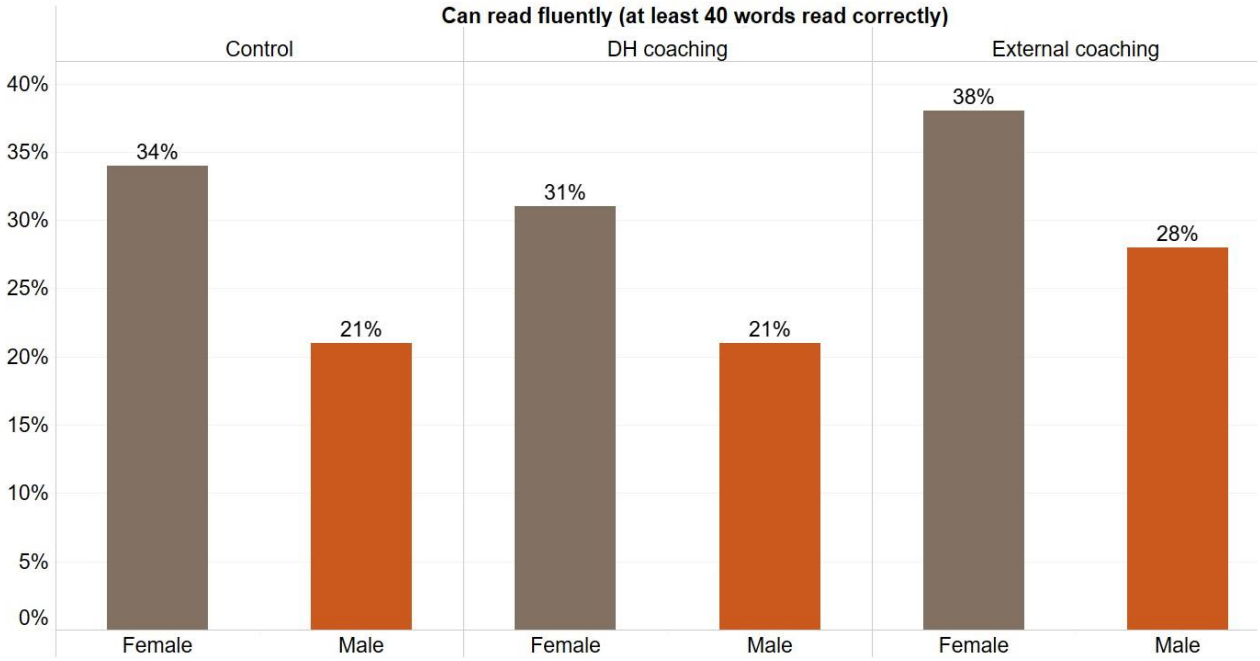

Figure 45: Grade 2 Oral Reading Fluency in HL by Gender and Intervention Group

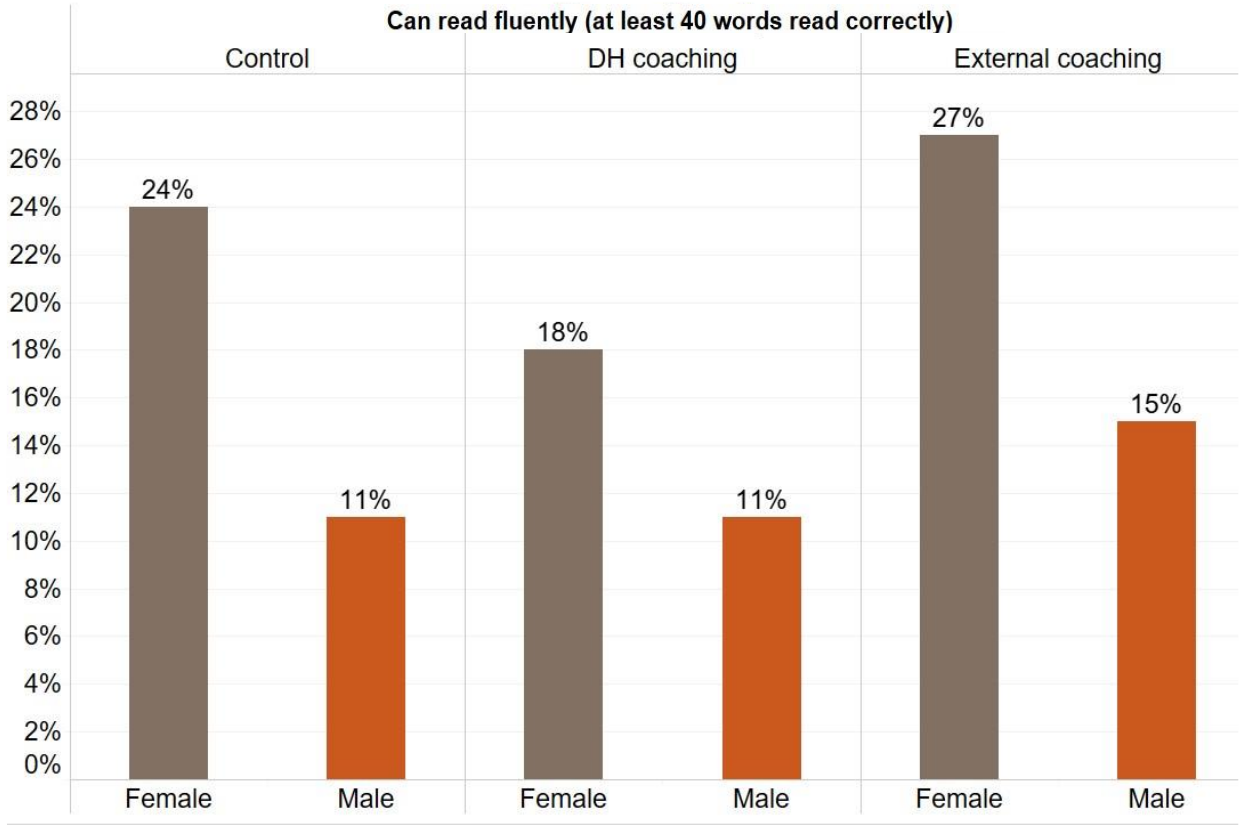

Figure 46: Grade 2 Oral Reading Fluency in EFAL by Gender and Intervention Group

These graphs illustrate that across both the intervention groups and the control group more males than females cannot read fluently or even read a word and more females than males can read at and beyond the benchmark that indicates fluency.

The last two graphs, Figures 47 and 48, show the relative performance by girls and boys in each grade (it is important to note that the various tests have different numbers of items as indicated in the graph so they cannot be directly compared). A glance at Figure 44 shows that the performance of girls even in Grade 1 far outstrips that of boys, particularly on the crucial reading sub-tasks of letter sounds and ORF. Figure 45 indicates that the differential between male and female performance in these key tasks has widened as we would expect based on the Mathew Effect. Interestingly, while girls still did better than boys in the group tasks the difference between the sexes was much less pronounced.

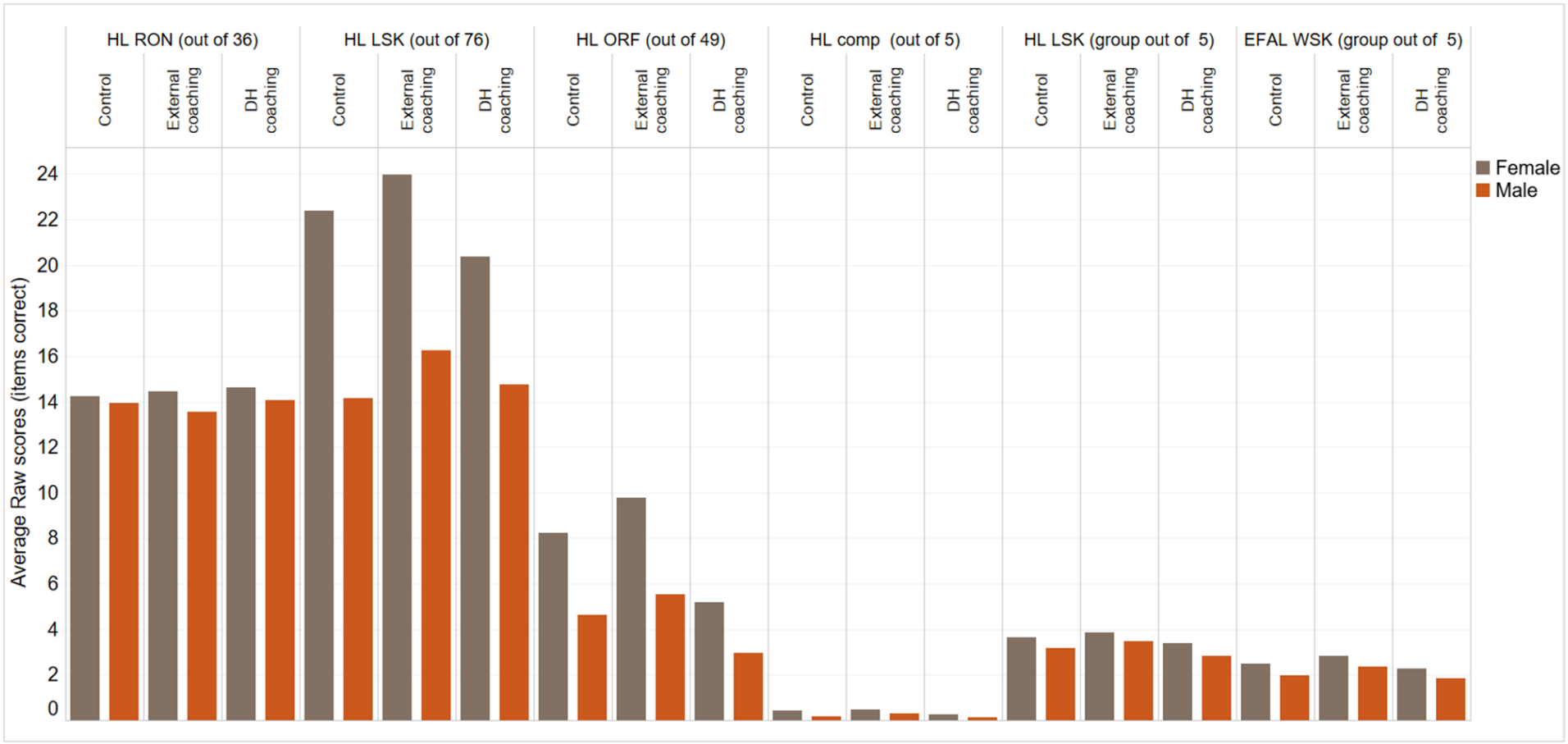

Figure 47: Grade 1 subtask performance by intervention group and gender

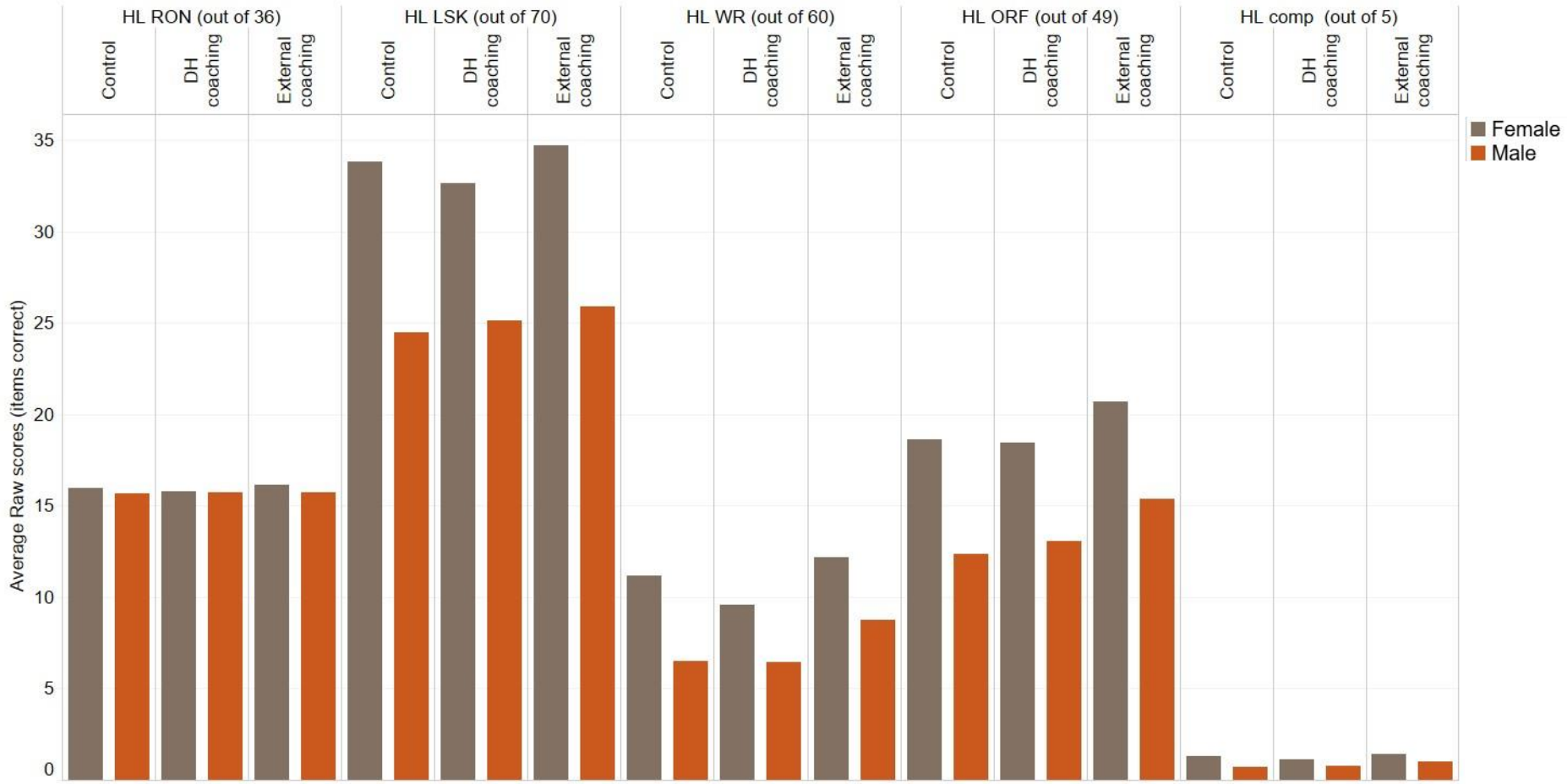

Figure 48: Grade 2 subtask performance by intervention group and gender (subtask 1-5)

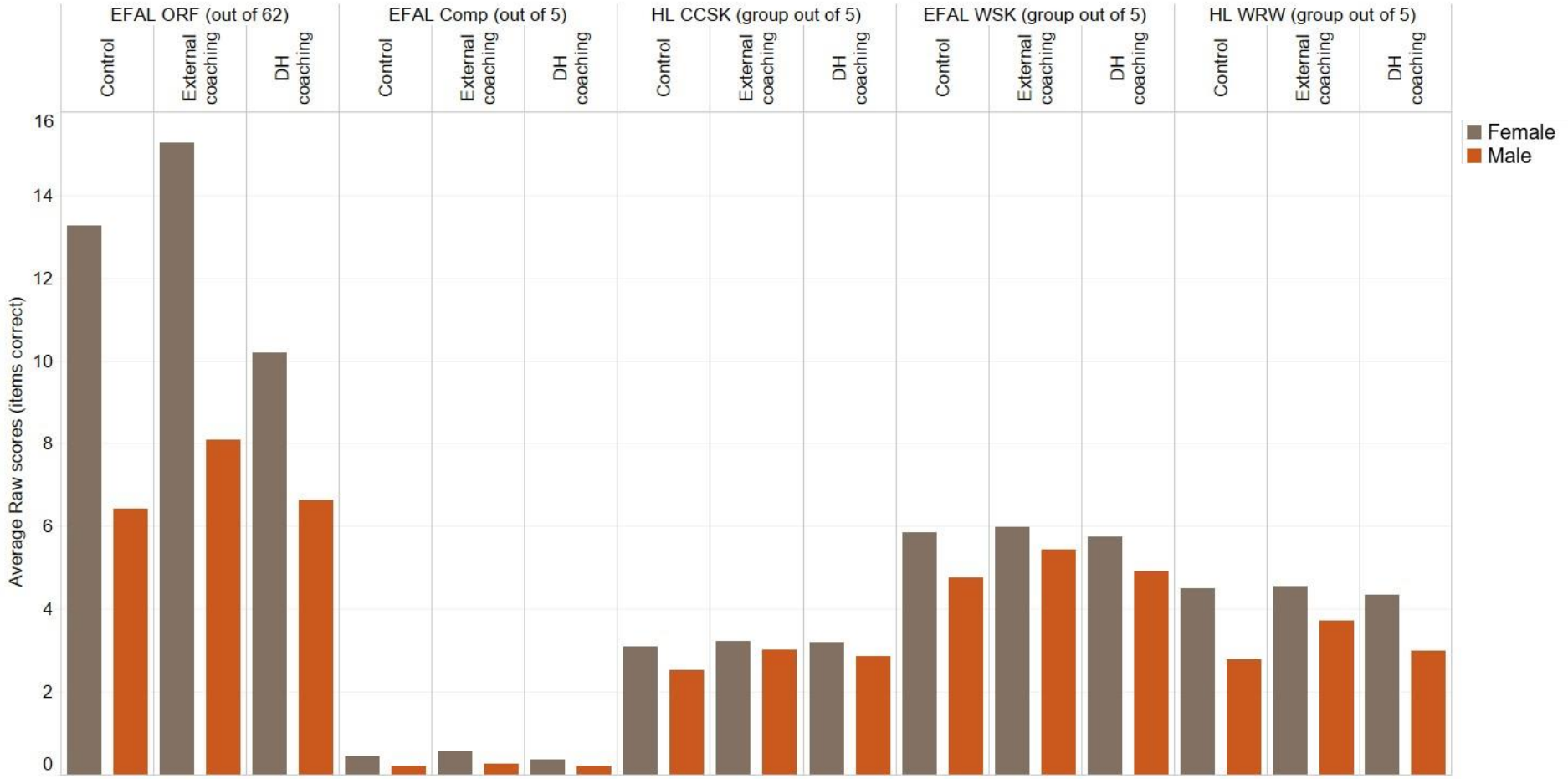

Figure 49: Grade 2 subtask performance by intervention group and gender (subtask 5-10)

4.5 Influence of context variables on learner test scores

This section presents results from the Ordinary Least Squares (OLS) regression runs on the extent to which learner, teacher, parent, and school environmental characteristics predict learner test scores. The analysis bases performance on the overall composite test score and the number of letters correctly read in the home language. The overall composite score is reported in terms of standard deviation, whereas the letter sounds are reported in letters read correctly within 60 seconds. These regression results do not establish causality but provide important insights for triangulating the findings of the evaluation study.

4.5.1 The influence of learner characteristics on learner test scores

Table 29 describes the learner characteristics (context variables) used in the linear regression models to predict their influence on learner performance (i.e., home language letter sounds and composite test score). For instance, there were three districts in the study, but Greater Taung was chosen as a reference variable for the analysis implying the model will only produce results for Kagisano and Naledi. The results will, however, be interpreted relative to Greater Taung. Likewise, the gender column was a dummy variable, and male learners were used as a reference variable, so the model will only produce results for female learners. Similarly, the model findings will be interpreted relative to male learners.

Table 29: Learner Characteristics used in Learner OLS model

Variable	Description	Variable Options	Reference
Subdistrict		Greater Taung	Yes
		Kagisano	
		Naledi	
Age	Learner Age	Age in years	
Gender		Female	
		Male	Yes
Homework help	Do they help you with homework at home? (learner response)	No	Yes
		Yes	
Pre-school Attendance	Did you attend grade R/creche?	None	Yes
		ECD/Creche	
		Grade R	
Read to child	Do they read books to you at home? (learner response)	No	Yes
		Yes	
Missed school	Has your child missed school this year for reasons other than rotations caused by Covid? (parent response)	None	Yes
		1-5 days	
		5+ days	

Table 31 shows the results of different learner characteristics correlated with their test scores after one year of implementation. Each column represents a separate regression run with the final column including all variables (predictors). Learners in Kagisano Molopo and Naledi seemed to have performed at a lower level than their counterparts in Greater Taung; however, only the results for Kagisano Molopo were statistically significant. This difference was unexpected given that Kagisano MoIopo has performed better than Greater Taung based on data from the last term of 2021 collected from schools through the Data-Driven Districts Database for Grade 3. The data shows the Kagisano Molopo region had a higher pass rate and average (in Grade 3) than Greater Taung and Naledi in both Home Language and EFAL (see Table 30). This difference will be monitored in the endline assessment in 2023.

Table 30: Fourth term Grade 3 results from DDD Database

District	Mathematics		Setswana Home Language		English First Additional Language	
	Pass Rate	Average	Pass Rate	Average	Pass Rate	Average
Greater Taung	94,59%	71,02%	89,71%	71,27%	89,71%	
Kagisano Molopo	98,79%	77,66%	90,17%	75,46%	90,17%	
Naledi	92,64%	70,91%	82,96%	67,51%	82,96%	

Learners' attendance of ECD or pre-school does not seem to impact as a significant majority of learners attended ECD or pre-school. As expected, the influence of learners missing school negatively impacted learner performance for the HL letter-sound knowledge sub-task, especially in the context of disruptions related to COVID-19 and rotational learning.

Similar to the previous EGRS report findings, learner age does not predict learner performance. However, as we have already seen gender is a predictor of learner performance, with female learners performing better than their male counterparts, with a significant gap of 8.82 points over their male counterparts in home language letter sounds. This difference, albeit in different magnitudes, is consistent with previous EGRS studies. It is also consistent with the analysis of fourth grade reading results from the PIRLS 2011 covering 45 countries, which found that, on average, girls outperformed boys in reading achievement in 40 of these countries. South Africa had the third-largest gender gap (in favour of girls), equivalent to one year of learning by Grade 4 (Mullis et al., 2012 as cited in van Broekhuizen & Spaull, 2017). Mohohlwane (2016) also notes that girls in North West Province enter school in Grade 1 with statistically significant advantages over boys in Setswana as indicated by their baseline home language scores in Setswana.

Several reasons are proffered in literature for these differences, but Broekhuizen & Spaull (2017:p2), citing Becker et al., (2010) reflect that the "*current leading explanation is that females have more traits and behaviours that are favourable for schooling in its current form"*..

Table 31: Results of learner characteristics correlated with learner test scores

	Average score			Letter sounds (letters correct)		
	(1)	(2)	(3)	(1)	(2)	(3)
Kagisano	-	0.14 (0.87)	0.12 (0.74)	-	-3.77 (-1.78)	-4.26* (-2.05)
Naledi	-	0.11 (0.43)	0.23 (0.91)	-	-3.47 (-1.67)	-2.65 (-1.36)
Age	0.27*** (3.32)	-	-0.08 (-1.10)	4.55*** (7.94)	-	1.05 (1.61)
Female	0.16 (1.79)	-	0.14 (-1.57)	9.22*** (12.42)	-	8.82*** (12.03)
Grade 2	-	-	1.00*** (3.77)	-	-	10.1*** (6.11)
Homework help	-	0.29 (1.34)	0.18 (0.85)	-	1.57 (0.96)	1.03 (0.67)
ECD/Creche	-	0.21 (0.8)	0.12 (0.49)	-	1.45 (0.67)	-0.02 (-0.01)
Grade R	-	-0.04 (-0.15)	-0.11 (-0.45)	-	2.09 (0.92)	0.69 (0.32)
Missed school (1-5 days)	-	-0.06 (-0.43)	-0.06 (-0.620)	-	-2.33* (-2.50)	-1.80* (-2.07)
Missed school (5+ days)	-	-0.03 (-0.13)	0.02 (0.09)	-	-3.57* (-1.90)	-3.31* (-1.94)
Read books - yes	-	-0.35 (-1.11)	-0.39 (-1.47)	-	2.00 (1.25)	1.04 (0.72)
constant	-1.91	0.02	0.22	-9.81	21.13	8.12
N	2431	2421	2418	2431	2421	2418
Adjusted R2	0.011	0.004	0.042	0.08	0.02	0.122

*Note: Each column represents a separate regression on learner characteristics and parent background information. Standard errors are clustered at the school level. * for p<.05; ** for p<.01; *** for p<.001*
t statistics in parentheses

4.5.2 The influence of parent characteristics on learner test scores

Table 32 describes the parent/caregiver characteristics that were used in the linear regression models to predict their influence on learner performance. The parent/caregiver level of education was categorized into 4 options where the "*No formal education*" option was used as the reference target and the model results for the three other options will be interpreted relative to "*No formal education*".

Other parent/caregiver context variables used in the OLS regression model include whether parents/caregivers assist learners with homework, do they "*read to their child"*, the "*number of books at home"*, *"employment status",* and "*socio-economic status"*.

Table 32: Parent/Caregiver Characteristics used in Learner OLS model

Variable	Description	Variable Options	Reference
Education	What is your highest completed education level?	No formal education	Yes
		Less than Matric	
		Matric	
		Post-Matric	
Homework help	Do you assist the learner with schoolwork after school?	No	Yes
		Yes	
Read to child	Do you read to your child? (parent response)	No	Yes
		Yes	
Books	How many books are there in your home?	None	Yes
		1-5 books	
		5+ books	
Employed	In the past 2 weeks, have you done any form of work for which you earned money?	No	Yes
		Yes	
SES (Socio-Economic Status)	Do you have any of the following things in your home?	13 items used to measure SES score	

Table 33 reports the results of parent/caregiver context factors on learner outcomes. It shows that learners with parents/caregivers who have attained a matric and especially a post-matric qualification performed significantly better at letter-sound knowledge than learners with parents/caregivers with *"no formal education"* (i.e. reference target). However, for overall performance (composite score) only, learners with parents/caregivers who possess a post-matric qualification performed better than those with no formal education. This might point to the differences in the quality of home support the two groups of parents/caregivers give.

Support with homework did not seem to influence learner performance since a significant number of parents/caregivers claimed to support their child/children with homework. This result was unexpected given the previous EGRS findings demonstrated a positive relationship between learner test scores and support with homework. Equally surprising was that reading to children did not strongly correlate with learner scores as expected. A plausible explanation might be that some variables such as "*homework help"* and "*read to child"* used in the model may have been susceptible to desirability bias, as parents/caregivers may be prone to give socially acceptable responses instead of responding truthfully. This bias was anticipated, prompting the fielding of the same questions to learners; however, the regression result from the learner responses remained the same.

The number of books in the home elicited strong results, with the home environment having more than five books reflecting a much stronger effect. This is expected and reflects the literature, which indicates that the presence of a range of books in the home encourages early engagement with the written word and an interest in reading as grown-ups tend to model reading where there are books available.

Parents/ caregivers being employed negatively affects learners' HL letter-sound knowledge performance, possibly because they face more time constraints than their unemployed counterparts. The socio-economic status of parents (proxied by the Living Standards Measure) does not predict learner performance. The lack of a strong predictive value might be associated with the general lack of variability in the socioeconomic status of parents whose children attend quintile 1 to 3 schools.

Table 33: Performance by Parent/Caregiver characteristics

	Average score			**Letter sounds (letters correct)**		
	(1)	(2)	(3)	(1)	(2)	(3)
Kagisano			0.08 (0.49)			-3.89* (-1.95)
Naledi			0.20 (0.79)			-2.43 (-1.15)
Less than Matric	-0.008 (-0.05)		0.02 (0.11)	0.64 (0.30)		-0.31 (-0.15)
Matric	0.20 (1.03)		0.22 (1.16)	6.10** (2.71)		4.67* (2.11)
Post-Matric	0.84* (2.38)		0.86* (2.43)	9.58*** (3.22)		7.98** (2.69)
Homework help		-0.18 (-0.90)	-0.25 (-1.25)		1.94 (0.99)	1.01 (0.58)
ECD/Creche			0.26 (1.53)			1.00 (0.59)
Grade R			0.18 (0.93)			-0.11 (-0.06)
Read to child – yes		-0.004 (-0.02)	-0.03 (-0.19)		-1.74 (-1.05)	-2.73 (-1.76)
Books (1-5)		0.27* (1.98)	0.24 (1.74)		5.27*** (4.38)	4.75*** (3.84)
Books (5+)		0.30 (1.92)	0.24 (1.53)		8.12*** (5.80)	7.26*** (4.90)
Employed - Yes	-0.09 (-0.65)		-0.11 (-0.80)	-2.93* (-2.36)		-3.00* (-2.46)
SES	0.00 (0.36)		0.005 (0.24)	0.40 (1.40)		0.39 (1.44)
constant	-0.01	0.03	-0.21	23.88	22.22	23.3
N	2427	2426	2416	2427	2426	2416
Adjusted R2	0.005	0.0003	0.005	0.02	0.01	0.04

*Note: Each column represents a separate regression on learner characteristics and parent background information. Standard errors are clustered at the school level. * for p<.05; ** for p<.01; *** for p<.001*
t statistics in parentheses

4.5.3 The influence of school Principal, socio-economic and school environmental factors on learner test scores

The socio-economic and school environment attributes used in the linear regression model are described in Table 34. Its results are reported in Table 35, which indicates that only a few results were statistically significant in this regression. It should be noted that the standard errors were clustered at the school level.

Table 34: School Principal, socio-economic and school environmental characteristics used in Learner OLS model

Variable	Description	Variable Options	Reference
Quintile	The schools in the lower quintiles (1 to 3) are declared no-fee schools, and do not charge school fees	Quintile 1	Yes
		Quintile 2	
		Quintile 3	
Age	Principal Age	Age in years	
Principal Tenure	How long have you been the Principal at this school?	tenure in years	
Phase	What phase were you teaching before becoming the Principal?	Foundation phase	Yes
		Intermediate phase	
		Senior phase	
Daily Learner Absenteeism	This term, on average what percentage of learners in the school were absent on any given day for reasons other than rotational schedules?	<10%	Yes
		11-20%	
		21-30%	
		31-40%	
Impact of COVID	For this year (2021), for how many days was the school closed due to COVID19 related factors?	Number of days	

Learner performance in both districts (i.e. Kagisano and Naledi) seems inferior to Greater Taung, since the regression coefficient was negative and sizable. Quintile status did not seem to predict learner performance well since all schools are no-fee paying schools, and their socio-economic conditions might be similar.

Principal age and Principal experience (i.e. how long the Principal has served at the school) does not predict learner performance. The phase the Principal taught as a teacher before becoming a Principal is significant. Surprisingly, Principals who were previously intermediate and senior phase teachers have learners who performed significantly better than Principals who were once foundation phase teachers. Absenteeism, as expected, significantly negatively affects learner performance.

Table 35: Learner performance by school Principal, location and socio-economic background characteristics

	Average score			Letter sounds (letters correct)		
	(1)	(2)	(3)	(1)	(2)	(3)
Kagisano	0.087 (0.85)	-	0.08 (0.74)	-2.45 (-1.25)	-	-3.60 (-1.82)
Naledi	-0.36* (-2.34)	-	-0.27 (-1.82)	-5.64** (-3.26)	-	-5.24** (-2.82)
Quintile 2	0.09 (0.79)	-	0.06 (0.50)	1.45 (0.71)	-	0.33 (-0.16)
Quintile 3	0.21 (1.78)	-	0.18 (1.52)	2.79 (1.47)	-	1.02 (0.55)

	Average score			Letter sounds (letters correct)		
	(1)	(2)	(3)	(1)	(2)	(3)
Age	-	0.016 (1.73)	0.01 (1.35)	-	0.20 (1.21)	0.14 (0.70)
Principal Tenure	-	-0.008 (-1.04)	-0.008 (-1.24)	-	-0.02 (-0.16)	-0.022 (-0.19)
Intermediate phase	-	0.23* (1.98)	0.20 (1.81)	-	4.43* (2.14)	3.96* (1.97)
Senior phase	-	0.13 (0.98)	0.13 (1.03)	-	6.03** (2.61)	6.08** (2.79)
Daily learner absenteeism (<10%)	-	-0.33* (-2.03)	-0.25 (1.51)	-	-5.11 (-1.54)	-3.50 (-1.01)
Daily learner absenteeism (11-20%)	-	-0.29 (-1.33)	-0.15 (-0.68)	-	-9.84** (-2.82)	-8.13* (-2.19)
Daily learner Absenteeism (21-30%)	-	-0.55* (-2.34)	-0.38 (-1.56)	-	-15.73** (-3.00)	-16.04** (-2.84)
Daily learner Absenteeism (31-40%)	-	-0.50* (-2.30)	-0.38 (-1.78)	-	-2.53 (-0.70)	-2.10 (-0.60)
Impact of COVID	-	-	0.01 (1.94)	-	-	-0.03 (-0.27)
constant	-0.019	-0.61	-0.68	24.83	16.03	19.8
N	5294	5289	5284	5294	5289	5284
Adjusted R2	0.004	0.002	0.006	0.011	0.024	0.033

*Note: Each column represents a separate regression on learner characteristics and parent background information. Standard errors are clustered at the school level. * for p<.05; ** for p<.01; *** for p<.001*
t statistics in parentheses

4.5.4 The influence of teacher attributes on learner test scores

Table 36 below describes the teacher attributes used in the linear regression model. As seen before, the table includes the variables of interest, the variable descriptions, the options per variable and the reference target.

Table 36: Teacher characteristics used in Learner OLS model

Variable	Description	Variable Options	Reference
Age	Teacher Age	Less than 45 years old	Yes
		45-54 years old	
		55+ years old	
Teaching experience	How many years have you been teaching in total?	number of years	

Learner Absenteeism	For the previous full week, how many learners in your class were absent on an average day?	number of days

Only a few results were statistically significant in the OLS model using teacher characteristics, as seen in Table 37. Teacher educational level and age do not predict learner performance. Teacher experience (i.e. how many years have you been teaching in total?) also surprisingly does not predict learner performance. This could be related to EGRP introducing new methods which older teachers often struggle with as they are more likely to be set in their ways than younger teachers. This is particularly the case as the EGRP requires teachers to access lesson plans, LTSM etc on tablets. Triangulated evidence from coach interviews, case studies, and teachers' self-reported survey data show that most teachers struggled with some core methodologies put forward by the EGRP. Therefore we can assume that change would only be noticeable in as far as the teachers (both experienced and less experienced) implement the new methodologies effectively.

As expected, learner absenteeism had a negative impact on HL letter-sound sub-task. Caution should, however, be taken when interpreting these results as the explanatory power of these models was very low, as indicated by the adjusted R-square.

Table 37: Learner Performance by teacher characteristics

	Average score			**Letter sounds (letters correct)**		
	(1)	(2)	(3)	(1)	(2)	(3)
Kagisano	-	-	0.03 (0.40)	-	-	-3.51 (-1.72)
Naledi	-	-	-0.25* (-2.04)	-	-	-3.71 (-1.63)
Quintile 2	-	-	0.07 (0.81)	-	-	-1.97 (-0.84)
Quintile 3	-	-	0.09 (1.53)	-	-	-1.12 (-0.54)
45-54 years old	-0.10 (-0.53)	-	-0.09 (-0.96)	0.77 (0.35)	-	1.55 (0.64)
55+ years old	-0.21 (-1.06)	-	-0.21 (-1.70)	-1.43 (-0.71)	-	0.32 (0.11)
Teaching experience	-	-0.006 (-0.90)	-0.0004 (-0.11)	-	-0.06 (-1.00)	-0.05 (-0.62)
Learner absenteeism	-	0.005 (0.31)	0.01 (1.02)	-	-0.73*** (-3.66)	-0.73*** (-3.88)
constant	0.12	0.078	0.05	24.84	28.81	30.5
N	4067	4067	4061	4067	4067	4061
Adjusted R2	0.001	0.0006	0.001	0.02	0.02	0.02

*Note: Each column represents a separate regression on learner characteristics and parent background information. Standard errors are clustered at the school level. * for p<.05; ** for p<.01; *** for p<.001*
t statistics in parentheses

5 Conclusions

The early grade reading assessment process unfolded as planned across all the 140 schools. In all schools the SSA team was welcomed and supported by the staff and school management team. In total 2 915 learners were assessed in Grade 1 and 2 897 in Grade 2, making a total of 5 812 learners who were assessed using contextualised EGRA sub-tasks based on those used in earlier EGRS studies.

The learner performances in both Grade 1 and Grade 2 were poor in HL and EFAL: 82% of Grade 1s and 58% of Grade 2s could not read a single word in their HL, and 74% of Grade 2s could not read a word of an EFAL text, even though all the words are grade appropriate and lifted from their textbooks and readers. Using the DBE benchmark for Grade 1 reading fluency,17% could sound out 40 correct HL letters per minute with fluency, with a larger proportion (70%) who can sound letters but with limited fluency that is below the acceptable DBE benchmark in HL. Meanwhile for Grade 2, using the DBE benchmarks, 29% of learners in HL and 18% in EFAL were reading at or above the benchmark, signifying expected performance, which is age-appropriate fluency. This is a pattern which has been noted in other learner assessments in Sub-Sahara Africa, particularly in 2nd grade, with learners largely split between those who cannot read at all and those who can read fluently, with relatively few learners in the middle – able to read a bit but slowly and with limited understanding. This is significant as teaching a class with these extremes in it is difficult.

Overall, the profiles of the two intervention groups and the control are statistically similar. This is a positive finding in terms of the evaluation study design since it suggests that there are no significant biases built into the different intervention groups, which may undermine the ability to reliably attribute endline results to intervention effects. There were, however, significant differences in Grade 1 learner performance between schools in the external coaching stream and the DH coaching stream, with learners in the External coaching stream performing better. This is in spite of both sets of schools supposedly receiving the same intervention in year 1. Since no statistically significant differences in school, teacher, home context or learner characteristics could be found, the reasons for this difference will be explored in the forthcoming implementation fidelity report.

In terms of the broader comparison between control and intervention schools, it is notable that after a year of receiving external professional coaching, the learners in the intervention schools (both external coaching and DH coaching streams) were not performing better than those in the control schools. This could relate to the quality of the coaching, the level of teacher uptake of the teaching approaches being coached or, more generally, through lack of fidelity exhibited in the delivery and take up of the innovations. The effect of COVID-19 disruptions and the rotational learning associated with it is also a possible contributor to this observation. As above, this will be explored further in the intervention fidelity report.

Many of the more predictive one-on-one administered sub-tasks produced strong floor effects. The same cohort of learners assessed in 2021 will be tracked through the next few years of their reading trajectory, this is not a serious problem. It should allow for real progress to be recorded as the majority of non-readers in Grade 1 and to some extent Grade 2 become readers in Grades 3 and 4.

When learner characteristics were compared with learner scores it was found that there is strong correlation between a learner's gender and their scores across all subtests. Girls did consistently better than boys in the Grade 1 and Grade 2 sub-tasks. This supports the literature (Mullis et al., 2012 as cited in van Broekhuizen & Spaull, 2017), which finds that this is consistent across countries in the early grades.

The learners did consistently better in the group sub-tasks compared to the individual tests, and so there are fewer floor effects and more differentiation between learners. The improved performance in the group sub-tasks may be due to learners feeling more confident and so performing better when sitting in their classroom surrounded by peers compared to sitting alone with a stranger. Possibly they also prefer multiple-choice type responses which do not require one to speak. However, it is also possible improved performance may be due to guessing the answer or even cheating from other learners. While the group sub-tasks have the potential to be seen as cheap and fast proxies for the standard EGRA one-on-one tests, the performance of learners in the group tests need to be compared with those from individual tasks to see that they do equate.

Not many external factors (either in the home environment or in the school environment) were found to have a statistically significant impact on learner reading scores. The home environment factors that did, include parents/caregivers having a tertiary education compared to those with no schooling or partial schooling and there being books in the home, particularly more than 5 books. These both could act as proxies for children growing up in an environment where reading for pleasure or for work reasons are a normal activity. In such households, children from a young age are likely to be exposed to books and have a good idea of how to engage with them. No school-based factors had a statistically significant impact on learner reading scores. Equally, none of the teacher attributes were found to have a statistically significant impact on their learners' reading performance. At the learner level the one factor which had a strong statistically significant relationship with learner performance was learner absenteeism. This is not surprising as time out of school will obviously relate to loss of learning time and so to less chance to learn to read, which as a skill requires sustained attention and work.

6 Recommendations

Intervention Implementation Recommendations

- The EGRP Implementing Partners should carefully consider why, after nearly a year of coaching engagement in intervention schools, there is no measurable difference in learner's reading performance compared to control schools. This suggests that the coaching intervention may either need a more robust or focussed design, or that implementation fidelity of the original design was low in the first year. The latter explanation will be explored in the forthcoming intervention fidelity report.

- The school observation results found that not all intervention classrooms had teaching aids on the walls. 12% of the classrooms in DH coaching schools were found to lack teaching aids on the walls. This should be addressed urgently by the EGRP Implementing Partners.

- It is critical that the programme singles out schools in the DH coaching intervention group where the DH is intending to leave the DH position before programme close-out and take

the potential successor (where possible) alongside the DH through support and coaching activities to prepare them for the DH coaching role. This will help ensure implementation integrity and programme continuity. Otherwise, a DH turnover or DH vacancies may endanger the comparative design of the intervention and evaluation.

Study Design Recommendations

- More research is needed into why learners do consistently better in the group sub-tasks compared to individual one-on-one tasks. As group sub-tasks create less floor effect than the one-on-one sub-tasks, they are arguably better at differentiating weak learners' performance. However, it is important to be sure that the improved performance is due to learners feeling more comfortable in their classroom doing a regular test rather than sitting with a stranger individually – which must be very intimidating for young children – and not due to cheating or the effect of guessing. Also, these sub-tasks need to be rigorously tested to equate them with the individual sub-tasks to prove that they have the same or similar predictive capacity.

- For the endline data collection phase, we recommend using only the home context questions posed to learners as part of the learner assessment and not repeating the questionnaire sent home to parents, given the low response rate of the latter.

7 References

Alsofrom, K. (2018). How and why coaching works to improve teaching practices in the EGRS II: An examination of causal mechanisms. Report to the Early Grade Reading Study (EGRS). Johannesburg: Government Printers.

Annie E. Casey Foundation. 2010. Early Warning: Why Reading By the End of Third Grade Matters. Baltimore, MD, USA: Annie E. Casey Foundation.

Chall, J.S. (1996) Learning to Read, the Great Debate. Harcourt Brace.

Christie, P (Ed.). (2022). Perspectives on Learning Backlogs in South Africa Schooling. Zenex Foundation

Coe, R. Aloisi, C. Higgins, S. & Major, L., (2014). What makes great teaching? Review of the underpinning research, London: Sutton Trust.

EGRS Mpumalanga (2019) Virtual coaching model summary. Department of Basic Education/USAID/ University of Witwatersrand.

Fleisch, B. (2018). The education triple cocktail: System-wide instructional reform in South Africa. Cape Town: University of Cape Town Press.

Hendrik van Broekhuizen & Nic Spaull, 2017. "The 'Martha Effect': The compounding female advantage in South African higher education," Working Papers 14/2017, Stellenbosch University, Department of Economics.

Hanushek, E. and L. Woessmann. 2012. Do Better Schools Lead to More Growth Cognitive Skills, Economic Outcomes, and Causation. Journal of Economic Growth, Vol. 17, Issue 4. 6

Gove, A. and A. Wetterberg. 2011. The Early Grade Reading Assessment: Applications and Interventions to Improve Basic Literacy. Research Triangle Park, NC, USA: Research Triangle Institute.

Hoadley, U., 2012. What do we know about teaching and learning in South African primary schools?. Education as Change, 16(2), p. 187–202.

Koo, T. and Li, M., 2016. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research

Research Triangle Institute (2015) Measurement and Research Support to Education Strategy Goal 1: Lot Quality Assurance Sampling (LQAS) Pilot in Tanzania: Final Report. Dar-es-Salaam: USAID.

Schotte, S., Zizzamia, R., & Leibbrandt, M. (2018). A poverty dynamics approach to social stratification: The South African case. World Development 110 (2018) 88-103.

Stanovich, K. 1986. Matthew Effects in Reading: Some Consequences of Individual Differences in the Acquisition of Literacy. Reading Research Quarterly. 11 (4).

Spaull, N., Pretorius, E. & Mohohlwane, N., 2020, 'Investigating the comprehension iceberg: Developing empirical benchmarks for early-grade reading in agglutinating African languages', South African Journal of Childhood Education 10(1), a773. https://doi.org/10.4102/sajce.v10i1.773

Spaull, N. & Makaluza, N., 2019. Girls Do Better, Agenda, 33:4, 11-28, DOI: 10.1080/10130950.2019.1672568

Matjila, D. & Pretorius, E., 2004. Bilingual and Biliterate? An exploratory study of Grade 8 reading skills in Setswana and English. Per Linguam, 20(1), p. 1–21

Mohohlwane, N.L., 2016. The contribution of Randomised Control Trials (RCTs) to improving education evaluations for policy: evidence from developing countries and South African case studies. Available at: http://wiredspace.wits.ac.za/handle/10539/22682 [Accessed February 7, 2022].

Mullis, I., O. Martin, M., Foy, P & Drucker, K. 2012a. PIRLS 2011 International Results in Reading. International Association for Evaluation of Educational Achievement (IEA). TIMSS & PIRLS International Study Center, Boston.

Van Broekhuizen, H., Spaull, N. 2017. The 'Martha Effect': The compounding female advantage in South African higher education. Stellenbosch Economic Working Papers: WP14/2017. Available at: http://https://ideas.repec.org/p/sza/wpaper/wpapers290.html[Accessed February 7, 2022].

UNESCO. 2006. Education for All Global Monitoring Report 2006: Literacy for Life. Paris, France: UNESCO.

Shepherd, D & Mohohlwane, N., 2021. The impact of COVID-19 in education - more than a year of disruption. National Income Dynamics (NIDS)-Coronavirus Rapid Mobile Survey (CRAM) Wave 5

Taylor, S., Cilliers, J., Prinsloo, C., Fleisch, B., & Reddy, V. (2017). The Early Grade Reading Study: Impact evaluation after two years of interventions. Department of Basic Education. Pretoria (Online). Available:https://www.jet.org.za/clearinghouse/primted/resources/language-and-literacy-resources-repository/egrstechnical-report-13-oct-2017.pdf

UNESCO (2021). One year into COVID-19 education disruption: Where do we stand? Article accessed 20 May 2021:https://en.unesco.org/news/one-year-covid-19-education-disruption-where-do-we-stand.

UNESCO (2015). EFA Global Monitoring Report. UNESCO: Paris.

UNESCO (1953). The use of the vernacular languages in education. Monographs on Foundations of Education, No. 8. UNESCO: Paris.

UNESCO (2005). Advocacy brief on mother tongue-based teaching and education for girls. UNESCO: Bangkok

UNESCO (2008). Mother Tongue Matters: Local Language as a Key to Effective Learning. UNESCO: Pari

Van der Berg, S. & Gustafsson, M. (2019). Educational Outcomes in Post-Apartheid South Africa: Signs of Progress Despite Great Inequality. In "South African Schooling: The Enigma of inequality (Eds: Spaull, N & Jansen, J). Springer.

Van der Berg, S., Spaull, N., Wills, G., Gustafsson, M., & Kotzé, J. (2016) Identifying the Binding Constraints in Education. A report commissioned by the South African Presidency

and funded by the European Union's Programme to Support Pro-Poor Policy Development (PSPPD) initiative.

Venkat, H. and Spaull, N. 2015. What do we know about primary teachers' mathematical content knowledge in South Africa? An analysis of SACMEQ 2007. International Journal of Educational Development. Vol. 41 Mar. p.121-130.

White, C. J., & Van Dyk, H. (2019). Theory and practice of the quintile ranking of schools in South Africa: A financial management perspective. South African Journal of Education, 39 (Supplement 1), s1-19

8 Annexures