Early Grade Reading Programme Evaluation Findings

Technical Report

August 2024

Acknowledgements

Suggested Citation: Zhou, T., Shilakoe, L., Polzer Ngwato, T., Ting (McClelland), L. & Prew, M. (2024) Early Grade Reading Programme I Evaluation Findings Technical Report. Department of Basic Education, Tshwane.
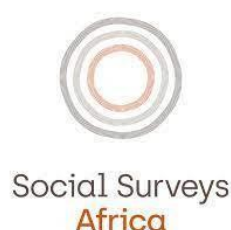
Table of Contents

Acronyms and Abbreviations

CAPS	Curriculum and Assessment Policy Statement
CPTD	Continuous Professional Training and Development
CWPM	Correct Words per Minute
DBE	Department of Basic Education
DH	Department Head
EA	Education Assistant
EFAL	English First Additional Language
EGMA	Early Grade Mathematics Assessment
EGRA	Early Grade Reading Assessment or Activity
EGRS	Early Grade Reading Study
EGRP	Early Grade Reading Programme
GGR	Group Guided Reading
GPLMS	Gauteng Primary Literacy and Mathematics Strategy
HL	(African) Home Language
IPA	Interpretative Phenomenological Analysis
LTSM	Learning and Teaching Support Material
M&E	Monitoring and Evaluation
Molteno	Molteno Language Institute
OLS	Ordinary Least Squares
PIRLS	Progress in International Reading Literacy Study
PLC	Professional Learning Community
RCT	Randomised Control Trial
RSP	Reading Support Project
SLP	Structured Learning Programme
SMT	School Management Team
SA	South Africa
TOC	Theory of Change
UNICEF	United Nations Children's Fund
ZAR	South African Rand

Executive Summary

Introduction

This report presents the findings of an independent evaluation of the Early Grade Reading Programme, implemented from 2021 to 2023 in the Dr Ruth Segomotsi Mompati District of the North West Province in South Africa. Commissioned by the Department of Basic Education (DBE) with support from UNICEF, the Hempel Foundation, and the Zenex Foundation, the intervention was delivered by the Molteno Language Institute (Molteno). Social Surveys Africa, commissioned by the Zenex Foundation, conducted the independent evaluation.

The EGRP is the latest in the Department of Basic Education's EGRS series of rigorous, large-scale studies aimed at testing interventions to strengthen foundational literacy outcomes within the public schooling system on a broad scale. Building on findings from previous studies on the efficacy of teacher coaching, the EGRP's central question is: Can coaching by Department Heads (DHs) produce similar improvements in learner reading outcomes as coaching by professional external coaches, but at a reduced cost?

The EGRP evaluation employed a mixed-methods approach. A randomised control trial design assigned 140 schools into two intervention groups (40 schools with external coaching and 40 schools with DH coaching, both accompanied by the 'base' intervention of structured learning programme materials and teacher training) and a control group (60 schools receiving only the 'base' intervention without coaching). Learners at these schools were assessed in November 2021 (Grades 1 and 2) and October 2023 (learners from the 2021 cohorts who were tracked into Grades 3 and 4, as well as a new cohort of Grade 2 learners) using standardised early grade reading assessments in Setswana Home Language and English as First Additional Language (EFAL). The methodology also included quantitative data collection on school-level evidence through interviews with teachers, DHs, and school managers, and qualitative data collection through six case studies (repeated in the same schools in 2021 and 2023), coach shadowing, classroom observations, and interviews with key informants involved in programme implementation, design, and funding. The key evaluation questions were whether the intervention led to the specified outcomes (changes in classroom teacher practices) and impacts (learner reading outcomes). The study also sought to identify the challenges of delivering coaching through DHs.

Key Findings

Implementation Fidelity

The base programme elements (production and distribution of Learning & Teaching Support Materials (LTSM) and teacher training) were implemented with sufficient fidelity and quality. The implementation of the pivotal coaching elements, however, was limited by late commencement of the DH coaching components and the uneven dosage of external coaching support to different schools. Across both the external and DH coaching streams, the coaching approach did not fulfil the key characteristics of 'instructional coaching', namely being intensive, individualised and developmental.

Teacher Outcomes

Teacher interviews and classroom observations in 2021 and 2023 confirmed an increase in teacher confidence and an improvement in classroom practices. Teachers attributed this to the structured lesson plans, teacher training and coaching (although the support offered by coaches was uneven). The more advanced reading methodologies central to the EGRP design, such as shared reading and group guided reading, continued to pose challenges for teachers in both coaching streams. This suggests that the coaching approach did not adequately address the barriers teachers faced in implementing these new pedagogical methods in the classroom—a concern that was also noted at the midline and reflected in earlier EGRS studies.

Learner Outcomes

The endline (2023) learner assessments did not find that learners in DH coaching schools had significantly better outcomes than learners in control schools. The effect sizes of DH coaching compared to control schools, while positive, were small across the board (from 0.011 in Grade 3 EFAL to 0.09 in Grade 2 HL).

We interpret the absence of additional effects from DH coaching, compared to the control group, based on the following findings regarding the implementation of the DH coaching intervention:

- **Learners received a very limited amount of DH coaching over a brief period.** After being exposed to external coaching in the first year (alongside the external coaching stream), the amount of coaching decreased in the second year (as external coaches continued to lead the coaching, but with a lower coach-to-school ratio). DH coaching only began at the start of the third year. Previous studies suggest that at least two years of coaching exposure are necessary to observe significant effects (Fixsen, 2005).
- **The quality of DH coaching varied greatly:** case studies and coach shadowing indicated that the effectiveness of DH coaching was heavily dependent on the commitment level of individual DHs. Additionally, in-person DH training was inconsistently scheduled, with most sessions occurring late in 2023. Although variations in motivation and skill are inherent in any large-scale educational intervention, this inconsistency may have reduced the overall effectiveness of the approach in this study, particularly when combined with the short implementation timeframe.
- **DH coaching was not 'instructional' in nature;** it lacked intensity, individualisation, and a developmental focus. Instead, it followed the lead of the external coaches (as directed by the implementing agency's overall coaching strategy), utilising a standardised, coach-driven (rather than teacher self-reflection-led), compliance-focused approach.
- The learner outcomes in DH schools do not suggest that these schools failed to see improvements in reading results during the intervention. Rather, the improvements

were not significantly greater than those achieved by schools that only received the base intervention. The evaluation's qualitative findings confirmed that the lesson plans and training provided through the base intervention, when combined with coaching, led to better classroom practices. This is likely to result in improvements in learner outcomes when compared to 'business as usual' schools that did not receive these inputs.

External coaching significantly improved both HL and EFAL learner outcomes in Grade 2, compared to learners in control schools, by adding more than 0.255 and 0.323 standard deviations respectively. In Grade 3, external coaching improved EFAL learning by about 0.175 standard deviations. No significant effect was found for external coaching in Grade 3 HL or for Grade 4 learning in either language. While this effect size would be considered large in relation to other international large-scale interventions (with a median effect size of 0.1 standard deviations in low and middle income countries) (Evans 2020). However, when these results are assessed against the reading benchmarks, they translate to a 3-6% reduction in the number of learners reading no words and a 4-5% increase in the proportion of learners meeting the benchmark. This underscores the complexities inherent in designing and implementing large-scale interventions aimed at addressing reading backlogs at a systemic level.

The modest impact of external coaching on Grades 3 and 4 (cohorts A and B of learners who were tracked from Grades 1 and 2 in 2021, respectively) is likely attributable to the learning backlogs caused by COVID-19.

● Cohort A and B were both subject to cumulative COVID-19 learning backlogs due to 2020 learning losses (in Grades R and 1, respectively) plus limited learning in 2021 (in Grades 1 and 2). Cohort B also experienced an intervention 'fade out' effect in 2023 since the intervention did not continue into Grade 4. In contrast, cohort C (Grade 2 in 2023) experienced entirely post-COVID-19 primary schooling and benefited from the more effective years of intervention implementation (2022 and 2023).
● While these cohort effects apply across all treatment groups, it is our view that the external coaching was not implemented with enough dosage or quality and did not take advantage of opportunities for reinforcing programme content, such as regular school-based workshops, which could have played a role in supporting teachers to address the cumulative COVID-19 backlog hole.

Given the COVID context challenge and the implementation fidelity concerns, the lack of measurable impacts of DH coaching at the learner level does not, however, directly imply that DH coaching cannot be effective. Based on qualitative insights, the evaluation concludes that DH coaching may be a viable means of integrating coaching into the public schooling system at scale if the following are in place:

1. Recognition of the coaching role of DHs by the provincial department, with concomitant adaptations to DH promotion/selection/succession policies & processes;
2. Careful selection of the DH coach which may, in the transition period from the current DH cohort to a future 'coaching-enabled' DH cohort, mean bypassing the existing DH and appointing a new DH who is committed to the role;

3. Education Assistants with sufficient training in the same structured learning programme utilised by the DH and teachers to support DHs and enable them to have time to observe and support teachers;
4. A curriculum coverage and lesson plan adherence data tool which tracks teacher use of the structured learning programme, provides DHs (and external support personnel) with real-time, easy-to-interpret evidence, and therefore enables targeted prioritisation of coaching time to teachers with the most severe backlogs;
5. Sufficient regular DH coach training, structured as continuous professional development;
6. Effective external coach support for the DH coach, with a higher initial support dosage that tapers off to a lower dosage once the DH has shown evidence of reaching a level of coaching competency; and
7. A longer period of support for the DH coach.

Introduction

This report summarises the findings of an independent evaluation of the Early Grade Reading Programme, implemented from 2021 to 2023, in the Ruth Mompati District of the North West Province in South Africa. Commissioned by the Department of Basic Education (DBE) with support from UNICEF, the Hempel Foundation and the Zenex Foundation, the intervention was implemented by the Molteno Language Institute (Molteno). The independent evaluation was conducted by Social Surveys Africa as commissioned by the Zenex Foundation.

The audience for this report includes the implementers, education officials (at national, provincial and district level), project funders, and other education stakeholders interested in improving Early Grade Reading.

Background to the Early Grade Reading Programme

South Africa has made a national policy commitment to ensure that learners are able to read for meaning by the age of ten. International and national studies have established that current Foundation Phase learning outcomes fall far below this target and need to compare better with international standards. The 2021 Progress in International Reading Literacy Study (PIRLS) showed that 81% of Grade 4 learners had not yet learned to read with a minimum acceptable level of comprehension. Improving reading outcomes in the Foundation Phase (Grades 1 to 3) is, therefore, a national priority.

As pointed out by Global School Leaders' 2024 Evidence Review, the South African situation is not unusual. "Nearly 7 in 10 students in [developing countries] are not mastering basic reading skills by age 10 (World Bank, 2022)…While only 14% of students in places like Europe and North America struggle with basic maths and reading, the numbers soar to 84-88% in Sub-Saharan Africa and 76-81% in parts of Asia (Clarke, 2022). Given this, there's a global push from governments and educational groups to find ways to boost learning and ensure every child gets a quality education." (Global School Leaders 2024)

Since 2015, the National Department of Basic Education has been leading a series of rigorous large-scale studies to test interventions to strengthen foundational literacy outcomes within the public schooling system at scale. These Early Grade Reading Studies (EGRS), designed as randomised control trials, have iteratively built up systematic evidence on the efficacy of different models of teacher support in teaching early-grade reading. This report summarises the findings of the Early Grade Reading Programme (EGRP) impact evaluation, which is the latest in the EGRS series.

The first in the series was the EGRS I, implemented in the North West province from 2015 to 2017, focused on reading in the home language of Setswana. EGRS II, implemented in Mpumalanga province from 2017-2019, focused mainly on English as a First Additional Language (EFAL). This was followed by the Reading Support Programme (RSP) in the North West (2019-2021), which included both Setswana home language and EFAL. EGRS I and II found that a package of teacher support consisting of Learning & Teaching Support Materials (LTSM) and teacher training, reinforced through literacy coaching of teachers, was effective in improving learner reading outcomes in Setswana and EFAL, respectively. This combination of training, coaching and provision of LTSMs for foundational learning, supported by rigorous qualitative and quantitative research methods and focused on rural

and other marginalised communities, is known as the 'triple cocktail' (Fleisch 2022). The 'triple cocktail' has been found to have a persistent impact. The original cohort of learners from the EGRS I study were reassessed in Grade 4 and Grade 7 and had retained improved reading skills in comparison with a control group.

Both EGRS I and II found external teacher coaching to significantly impact learners' reading performance (Taylor, Cilliers, Prinsloo, Fleisch, & Reddy, 2017). Learners in the schools that received external on-site coaching for two years were 40% of a years' worth of learning ahead of their peers in schools with no intervention ('business as usual' control schools) (Taylor et al., 2018).

However, external coaching is expensive and requires extensive capacity outside the public education system, neither of which fulfils the Department of Basic Education's requirements for a sustainable methodology for improving reading within the education system at scale.

EGRS II tested virtual coaching as a less expensive alternative to on-site coaching but found that it had a limited impact on teaching practice and learner outcomes (Cilliers et al., 2020; Kotze et al., 2019). The EGRP was, therefore, designed to test the efficacy of an alternative on-site coaching model embedded in the public schooling system using existing education system resources and personnel. In this model, Foundation Phase Department Heads (DH) assume the role of coach for teachers in their schools while, in turn, being supported by a smaller number of external coaches. The EGRP study compares the efficacy of DH coaching with the previously tested efficacy of external coaching by including both models in the same study and comparing them both to a set of 'control' schools receiving a 'base' Structured Learning Programme (SLP) including Learning & Teaching Support Materials (LTSM) and teacher training.

Therefore, the EGRP's guiding question is: Can teacher coaching by Department Heads have a similar impact on learner reading outcomes as coaching by professional external coaches, but at a lower cost?

In response to the EGRS findings and recommendations, the DBE has developed an improvement plan for reading outcomes endorsed by South Africa's National Cabinet. The plan aims to institutionalise the successful aspects of the interventions in the previous EGRS studies. Some of the critical items emerging from the improvement plan include:

- Adopting a structured learning programme using daily lesson plans which incorporate critical reading materials as a means of implementing the curriculum;
- developing guidelines for on-site coaching by specialised reading coaches and for institutional support for teachers and coaches;
- conducting further research to strengthen the programme, including implementing it on a larger scale;
- developing reading norms for African languages and
- evaluating the cost-effectiveness of using different support methods with teachers.

These efforts are all underway and are seen as necessary scaffolding for implementation at scale. The EGRP study is part of this larger plan to identify interventions that can be sustainably implemented at scale. The intention is to learn both at the design and implementation levels: what works in principle and what will work at scale in practice, given the current capacity and resources in the education system. Globally, evaluations are

increasingly considering implementation science to interpret whether measured intervention outcomes are due to design or contextual implementation factors. The current report draws on qualitative and quantitative methods to address these two levels of analysis.

Literature review

This brief literature review examines the key elements of the main hypothesis underpinning the EGRP study: that Department Heads can effectively coach their teachers to enhance the teaching of reading, with limited coaching provided to the DHs themselves, supported by training for both teachers and DHs, alongside the provision of appropriate LTSMs.

International literature highlights the critical role of teachers in student learning and school effectiveness, a factor often overlooked in school effectiveness research (White & Barber, 1997). Traditionally, school success in the developing world has been associated with the availability of resources and quality learning materials (Levin & Lockheed, 1993). However, over the past two decades, there has been growing recognition of teachers' pivotal role in both the success and failure of schools (Day, 2013). Increasing attention is paid to how teachers translate workshop-acquired knowledge and skills into classroom practice (Dadds, 1994; Gabriel, 2005; Germuth, 2018). Research suggests that teachers often struggle to implement what they learn in workshops, with limited uptake of new knowledge and skills. Germuth (2018: 78) notes that while teachers may acquire new knowledge and skills, they rarely put this knowledge into practice.

> "Research has shown that only ten per cent of teachers can transfer a new skill to actual practice when no additional support is provided. In comparison, embedded support for implementation can result in over 90% of teachers transferring the skills they developed to their practice. Additional research reveals that coaching is one way to change teacher practice successfully."

While researchers point out that researching the transfer of skills from professional development programmes to changes in teacher practice is difficult (Cochran-Smith et al. 2015), other research shows that if the teacher has support, someone who shows them in real-time how the innovation can work in their classroom or supports them with suggestions when they are trying out the innovation, gives the teacher confidence and has a significant impact on their ability to adopt the innovation (Germuth 2018, Fleisch et al. 2016).

Coaching teachers has been discussed for decades (see Fullan & Hargreaves 1992), but it is a relatively new concept that is a core part of teacher development. In the early years of the twenty-first century, various research studies have suggested that instructional leadership by school managers should involve coaching and mentoring their staff (Annenberg Foundation 2004; Knight 2011; Williamson 2012). It is pointed out that coaching is a critical element of continuous professional training and development in other professions.

The first systematic usage of the 'triple cocktail' in South Africa (LTSM, teacher training and teacher coaching), using one-on-one coaching of teachers by professional external coaches, was in the Gauteng Primary Language and Mathematics Strategy (GPLMS) (Fleisch et al. 2016). This study also introduced a second element: systematic and rigorous design planning, monitoring and evaluation (M&E) , which has been critical in professionalising such studies in South Africa, leading to the EGRS, RSP and EGRP randomised control trials (Bisgard et al. 2020).

For a real improvement in literacy levels of South African children, all teachers across the country need access to the 'triple cocktail', meaning that cost-effectiveness for scale is

crucial. While training and development of the LTSMs have been in place for some years, coaching has been problematic to implement at scale: external coaches are expensive, and there are simply too few educators with the requisite skills and experience to coach teachers across the country (Cilliers et al. 2020).

The literature contains some concerns about the effectiveness of scaling successful early-grade teacher coaching pilots (Kraft et al., 2018). Kraft et al's review of 60 studies that used causal research design on relevant teacher coaching projects showed large statistically significant effects of coaching on teacher practice and learner performance in pilot projects, but these effects were dramatically reduced when the pilot was taken to scale. Even in developed countries, a problem faced when scaling was accessing adequate numbers of competent coaches. Various studies have shown the impact of greater and lesser levels of coaching and concluded that reduced dosage from coaches in both the USA and Kenya are associated with much lower levels of impact on teacher practice and learner performance (Blazar & Kraft; Piper & Zuilkowski 2015). Reducing the dosage of coaching to control costs therefore does not seem to be a viable scaling option.

In addition to literature on coaching, there is a relevant body of research on the effects of school leadership on learning outcomes. Recent studies show that school leaders are second only to teachers in influencing student outcomes (de Barros et al., 2019; Bush et al., 2022; Cilliers et al., 2022; De Hoyos et al., 2021; Tavares, 2015; UNESCO, 2018; VVOB, 2018). Although school leadership interventions often focus on School Management Teams (SMTs), Department Heads also play those operative management roles found to lead to improved learning outcomes, namely: "championing reflective teaching practices, overseeing lesson planning, and fostering a culture of collaborative learning among educators (Bellibaş et al., 2021; Lemos et al., 2021)... [as well as] promoting teamwork among teachers and enhancing their job satisfaction (Wills & van der Berg, 2021; Shava & Heystek, 2021; Tavares, 2015)."

The EGRP posits that DHs can play the instructional leadership role, which essentially is the same as the role of the external coach, to teachers in their schools if provided with external training and support. As they are already working in schools, using the DH as the coach is cost-effective in the long run, and builds capacity in each primary school. But can the DHs play this role effectively? There is very little local or international literature on this proposition. This study, therefore, breaks new ground and will contribute to both the national and global debates on teacher coaching at scale.

The EGRP's approach to DH training and support is fairly hands-on, understanding that DHs, while already in leadership positions, also require training and coaching to implement new practices. While some management practices are introduced to the public schooling system through a new policy or circular, expecting schools to interpret and institutionalise the changes on their own, the EGRP provided concrete and personalised support to DHs over two full years.

The ability of DHs to play a role akin to external coaches depends on whether they can fulfil the characteristics of instructional coaching, defined as

> "site-based [professional development] designed to develop theory and use demonstration, observation and feedback to improve classroom practice" (Walpole et al, 2010, p118).

Instructional coaching is "probably the best-evidenced form of CPD currently known to humanity"[1]. This is because it is context-specific, personalised, practical, and happens in a teacher's classroom with her learners. This means that the teacher can see that the innovation can work with her learners, her level of resourcing, and her specific classroom. Overall, the aim of instructional coaching is for the coach and teacher to work closely together to improve the teacher's skills through a self-reflective process involving professional conversations that develop and implement evidence-based teaching practices[2] and so improve learner performance.

The EGRS II summary on virtual coaching (2019)[3] describes five characteristics that underpin instructional coaching relationships as distinct from other in-service training models:

- rather than a one-to-many mode of imparting knowledge, a coaching process is a **tailored and individualised interaction** where both parties form a close but boundaried professional relationship focused on the specific needs of the teacher (Majerowicz & Montero 2018);
- this relationship necessitates a **sustained interaction** over a long period (more than a year);
- given the intimacy and length it takes, coaching is considered an **intensive learning experience**;
- instructional coaching relationships are context-specific and focused on **specialised discrete skills** (Kraft, Blazar and Hogan 2018) and
- the role of a coach is fundamentally **supportive and developmental** before being evaluative (Mraz et al., 2016).

Given that DHs have existing roles and relationships with teachers, the EGRP asks whether DHs can prioritise sufficient time to build consistent and intense professional relationships and shift the nature of their often compliance—and evaluation-driven roles to supportive and developmental relationships.

Intervention Design

The implementing agency implemented the intervention in the three sub-districts of Kagisano Molopo, Greater Taung, and Naledi in Dr Ruth Segomotsi Mompati (RSM) District in the North West Province of South Africa.[4] 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 LTSM for the project. This district is poor and houses some of the most underperforming schools in South Africa. It is often overlooked by state and NGO interventions.

---

[1] https://blog.irisconnect.com/uk/coaching-for-teachers Posted 29 January 2020

[2] https://www.aitsl.edu.au/tools-resources/resource/instructional-coaching Accessed 12 October 2021

[3] https://www.education.gov.za/Portals/0/Documents/Publications/EGRS/EGRS%20II%20Website%20Upload/Reports/2019_Virtual%20Coaching%20Model%20Summary%20REVISED.pdf

[4] Both EGRS I and the RSP took place in Dr Kenneth Kaunda and Ngaka Modiri Molema districts of the North West Province. The intention in selecting a new district within the same province was to use the same language resources and assessments (Setswana) to enable comparability across studies, but to measure intervention impact without cross-fertilisation from previous interventions.

Figure 1: Distribution of sampled schools across three sub-districts shown by intervention arm
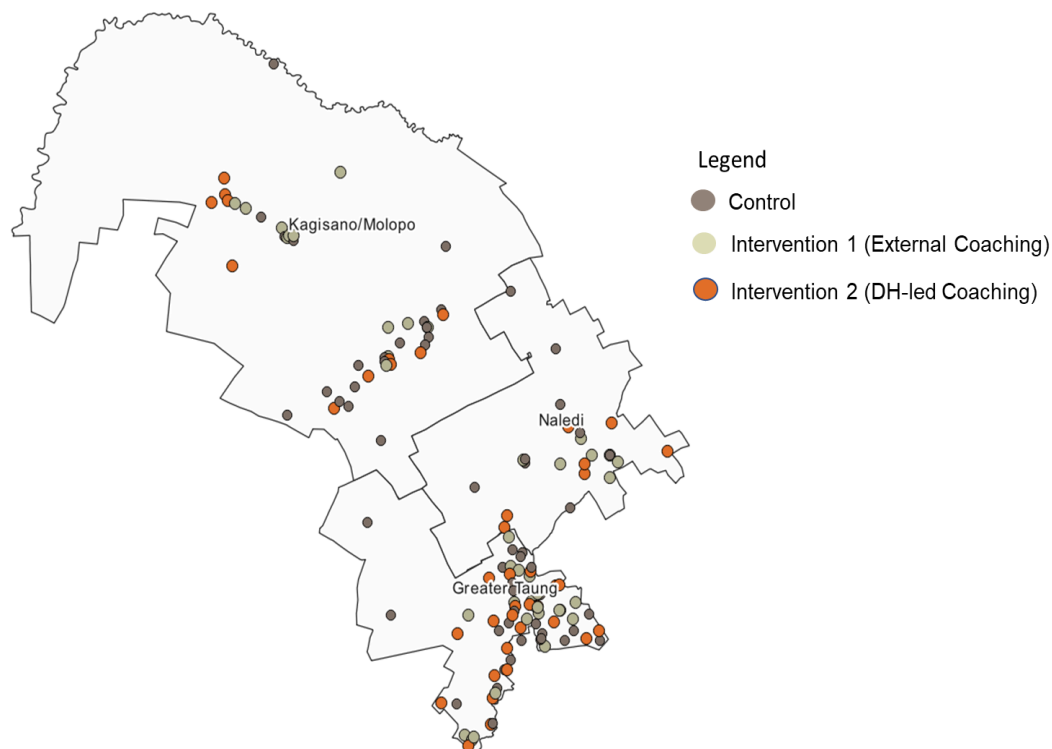
In late 2020, the Department of Basic Education, with its funding partners UNICEF and the Hempel Foundation, appointed Molteno as the external EGRP implementation agency. While most previous interventions in the EGRS series and RSP had been implemented by consortia of implementing agencies, a single agency was responsible for all aspects of the EGRP.

The EGRP was implemented simultaneously by all Grade 1, 2, and 3 teachers in all participating schools across all three implementation years.

As mentioned above and summarised in Figure 2, the EGRP study compared two coaching interventions: an external coaching model replicating the coaching approach found to be highly effective in the previous EGRS studies and a new Department Head (DH) coaching model. Each was implemented in 40 schools. A control group of 60 schools constituted the third 'stream' of the study. As per standard practice for randomised control trials, schools were randomly assigned to an intervention stream by the DBE, with extensive secondary data analysis on school size, performance and context to ensure balance between the groups. Small schools with multigrade classes were excluded from the sample.

Figure 2: High-level Intervention Design
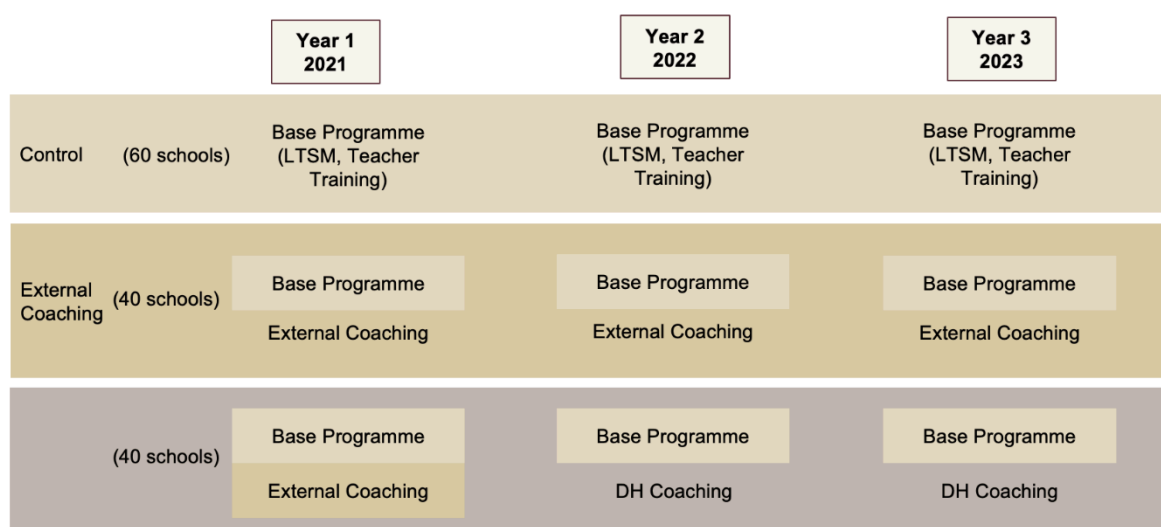
Table # provides more detail on the constituent elements of each intervention component.

Table 1: Detailed intervention components of each intervention element in more detail

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 two days of training at the beginning of each term throughout the intervention (1 day HL, 1 day EFAL)	Teachers receive two days of training at the beginning of each term throughout the intervention (1 day HL, 1 day EFAL)	Teachers receive two days of training at the beginning of each term throughout the intervention (1 day HL, 1 day EFAL)
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

All three streams received the same base programme consisting of Learning & Teaching Support Materials (LTSM) in Setswana and EFAL, and teacher training in using these LTSMs. In the first intervention year (2021), teachers in both coaching streams received regular on-site coaching visits by professional external coaches. The two coaching models diverged in the second intervention year, with the external stream continuing to receive external coaching visits. In contrast, the DH coaching stream shifted from external to DH coaching (Figure #).

Figure #: External and DH coaching Impact Chain
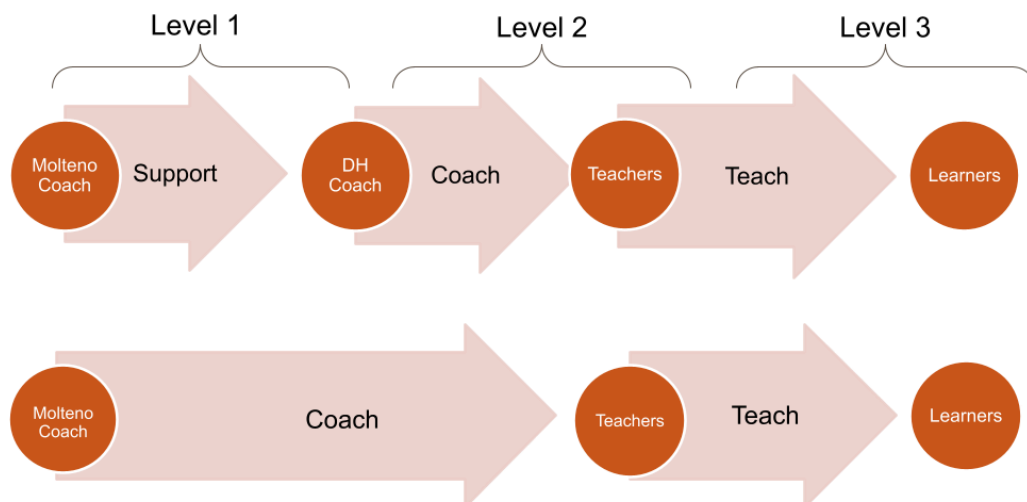
The implementing agency hired and trained eight external coaches as the primary contact between the intervention and schools. These coaches conduct teacher training and provide content-based coaching. Most coaches had experience with the Reading Support Programme and completed the University of Johannesburg's Foundation Phase Literacy Coaching qualification. In the first year, the coaching ratio was 1 coach to 10 schools. From the second year, the ratio in the external coaching stream was 1:7 while in the DH coaching stream it increased to 1:20.

External coaching model

In this model, the coach informs the school and teachers in advance to prepare for observation sessions, observing best-case scenarios. Each visit involves pre-classroom discussions, classroom observations, and post-observation discussions. The coach may model lesson approaches and conduct needs-based workshops after school to address common issues. An average of 12.5 coaching visits per teacher per year was expected.

This intervention arm receives weekly **virtual coaching** via WhatsApp, with a coach reminding teachers of lesson plans and providing resources. Monthly discussions focus on literacy teaching and EGRP programme content. Once a term, the coach engages in one-on-one sessions with each teacher. This coaching session will often include the following:

- 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.

DH coaching model

The two external coaches assigned to the DH stream trained the DHs in 2022 on how to coach and modelled coaching in schools while the DHs shadowed them. In 2022 and 2023, DHs were expected to coach teachers independently, while receiving in-person and virtual support from the external coaches who occasionally joined them in observing lessons conducted by their teachers and engaged with the DH on the way she coached her colleague. The external coaches were also expected to use their presence in the school to run a School Based Workshop (SBW) with the foundation phase staff and the DH. While the dosage expectation of DH coaches was not as clearly defined in the intervention design, the broad expectation was that each Foundation Phase teacher would be observed by the DH at least once a month during term time.

The theory of change (ToC) in Figure # shows the intended linkages between the intervention elements and the intended pathways through which the coaching leads to sustained change in classroom practice.

The EGRP Design in International Comparison

When considering the EGRP's key design elements in international comparison, we see that it follows 'best practice' since it fulfils all ten of the RTI's Learning at Scale (2021) interim report recommendations for programmes to achieve impact at scale. This is despite the EGRP having been designed before this review report was published. The RTI recommendations are based on a review of "eight of the most effective large-scale education programs in LMICs" (RTI 2021), including interventions in India, Tanzania, Ghana, Kenya, Pakistan, Senegal and Nigeria. The ten recommended programme elements are:

1. Programme's teacher training focused on modelling and practising new skills,
2. Programme included structured teachers' guides,
3. Coaches were provided with structured tools to support teachers,
4. Programme used face-to-face training methods for their initial training,
5. Programme used direct-instruction pedagogical methods,
6. Student books were available at a 1:1 ratio for all students,
7. Programme utilised a phonics-based instructional methodology,
8. Programme increased the amount of instructional time in reading lessons,
9. Programme provided capacity building at a decentralised level,
10. Programme was designed to align with existing government education plans.

The EGRP Implementation Environment and COVID-19

Even the best-designed programme can face seemingly insurmountable contextual challenges. Previous studies in the EGRS series documented both learning outcomes prior to COVID-19 and the learning losses during the pandemic (2020-2021). South African schools lost 54% of contact time in 2020 and 22% in 2021 due to absenteeism and rotational attendance. This impact was even more pronounced in the foundation phase, where up to 65% of school contact time was lost because of rotational attendance policies. In the foundation phase, it is estimated that the learning losses in 2020 amounted to as much as 75% of a year's worth of learning at the Grade 3 level.

EGRP was initiated in 2021 when the country was still undergoing various COVID-19-related lockdowns. The COVID-19 regulations in place during 2021 and early 2022 impacted learner and teacher attendance, teaching practices, and school management, affecting the implementation fidelity of some EGRP activities in the first year. External coach visits and learner and teacher attendance were negatively affected by illness, school lockdowns and learner rotational attendance. COVID-19 'social distancing' regulations in the classroom also reduced the practicality of implementing group-based teaching methodologies, such as group guided reading (GGR) and shared reading (SR) which are at the core of the EGRP structured learning programme.

Figure #: Theory of Change
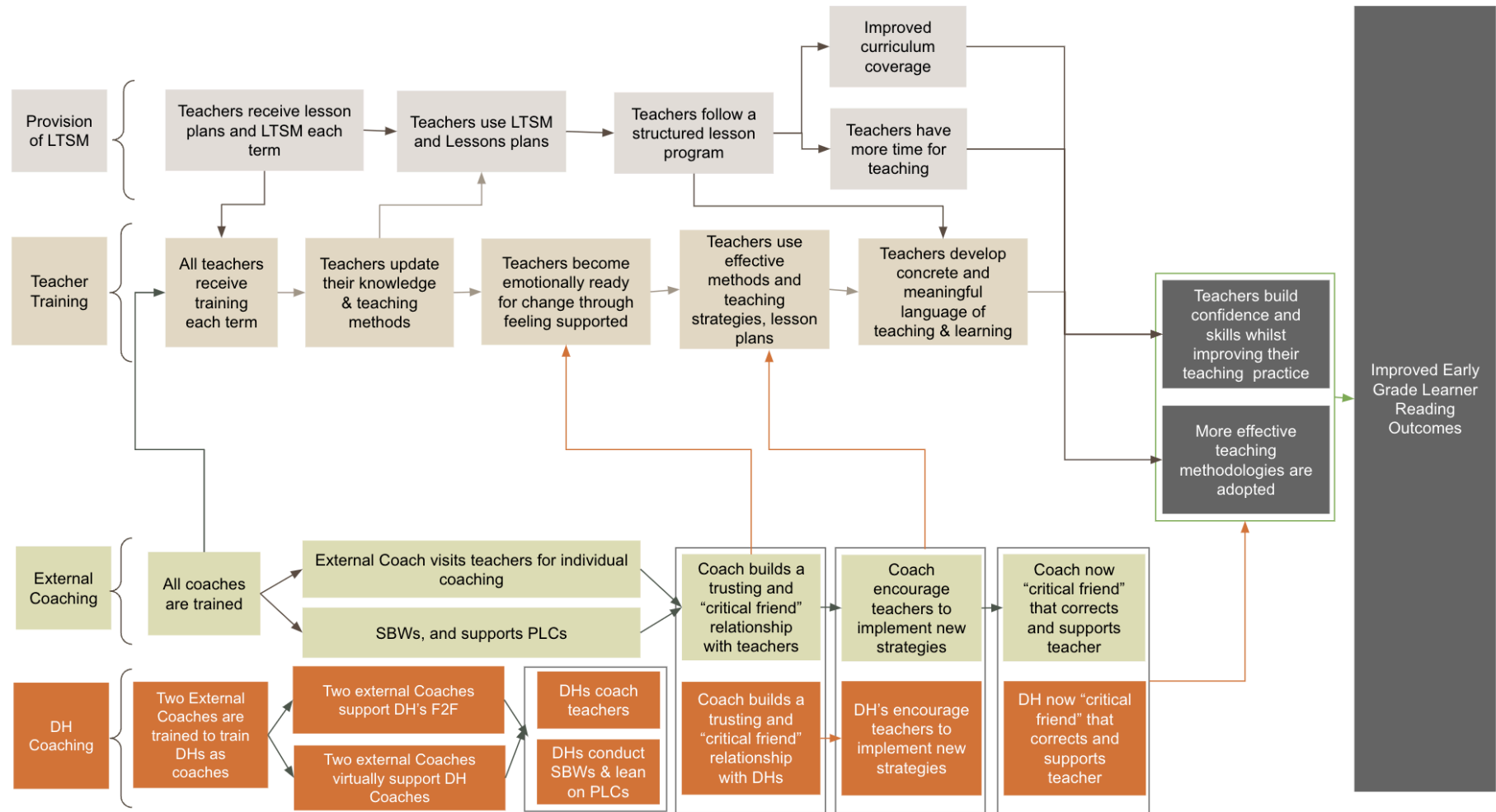
Evaluation Design

Evaluation Questions

The study sought to answer the following evaluation questions:

1. Was the programme implemented as intended (fidelity)?
2. Did the programme result in the specified outcomes and impacts?
3. What are the implementation challenges of delivering coaching through DHs?

Evaluation Scope and Mixed Method Design

Figure #: Evaluation Design
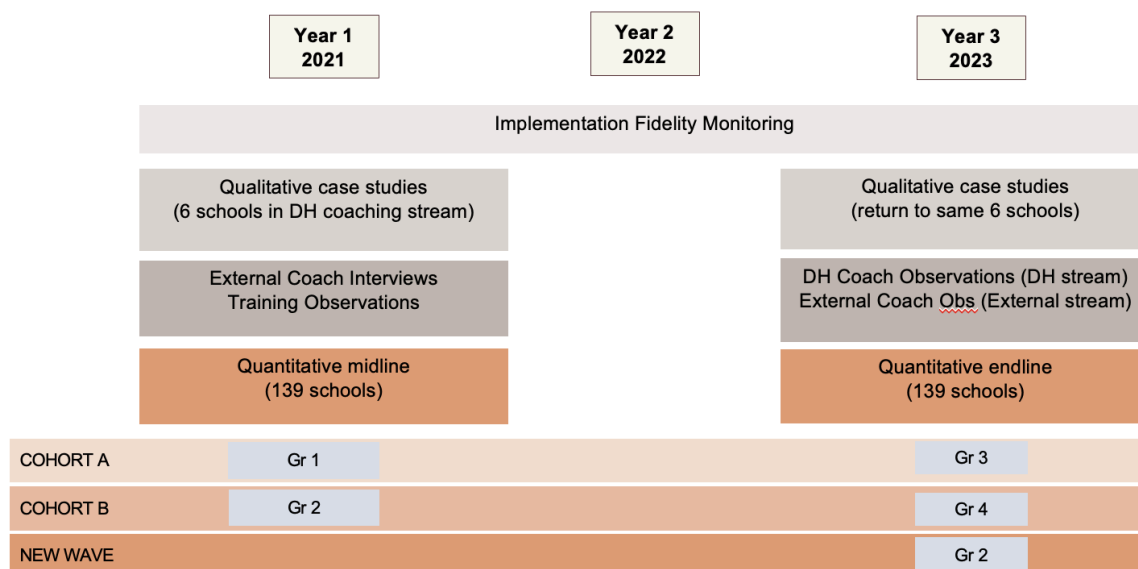
Given the EGRP's overall design as a randomised control trial, the primary aim of the evaluation was to establish the level of impact on learner outcomes for the DH coaching model compared to the external coaching model and the control schools. This was achieved by conducting standardised learner assessments in 2021 and 2023 across all 139 schools in the intervention.

No baseline assessment was conducted in 2020 or early 2021 before the commencement of the intervention. To enable comparability with the endline in 2023, the first round of quantitative data collection was conducted in the fourth term of 2021 after almost a year of intervention. Therefore, the first data collection round is treated as a midline rather than as a baseline. For this reason, the endline results are predominantly presented as a comparison between intervention groups rather than in terms of change between the midline and the endline.

As shown in Figure # above, the evaluation included three cohorts of learners. Two cohorts (A and B) followed learners who were in Grades 1 and 2 in 2021 and assessed them again in Grades 3 and 4 in 2023. Learners who repeated a grade were assessed together with their original cohort, e.g. a learner from cohort A would be reassessed in 2023 using Grade 3 assessments even if the learner remained in Grade 2. In 2023, a new group of Grade 2 learners was assessed. Their results were compared with the original 2021 Grade 2 group.

Each learner cohort experienced a different level of exposure to the intervention:

- **Cohort A** experienced three years of the intervention (Grade 1 in 2021, Grade 2 in 2022 and Grade 3 in 2023) but was affected significantly by COVID-19 restrictions on implementation and general learning in 2021, as well as being disadvantaged by a lack of Grade R learning in 2020.
- **Cohort B** experienced two years of the intervention (Grade 2 in 2021, Grade 3 in 2022). By the end of 2023 when they were reassessed, they had spent a year in Grade 4 with teachers who had not been included in the intervention, resulting in the possibility of a 'fade out' effect on learning outcomes.
- **Cohort C**, the 'new wave' of Grade 2 learners assessed in 2023, experienced two years of intervention (Grade 1 in 2022 and Grade 2 in 2023) without any COVID-19 effects or fade-out effects.

Considering this differential cohort exposure when interpreting the assessment outcomes is important.

In addition to learner assessments, extensive quantitative data was collected from teachers, DHs, principals and school administrators for the full sample of 139 schools to analyse fidelity, context and outcome-level factors. Table # shows the sample achieved for these instruments.

Table XXX: School-based Respondent Surveys Sample Achievement

Quantitative Tool Description	Midline (2021)			Endline (2023)		
	Target	Number achieved	Percentage Achieved	Target	Number achieved	Percentage Achieved
School Datasheet	139	135	**97%**	139	136	**98%**
Environment Scan	139	134	**96%**	139	132	**95%**
Principal	139	129	**93%**	139	119	**86%**
Teacher Interview	278	250	**90%**	278	325	**117%**
DH Interview	139	138	**99%**	139	112	**81%**
Learner Home Background Form	5560	1864	**34%**	6072	1957	**32%**

In addition, the imperative to consider the implementability of the DH model in the public education system as a whole led to the inclusion of a robust qualitative element in the form of six case studies (2021 and 2023), external coach interviews (2021 and 2022) and DH and external coach shadowing (2023). The case study and coach shadowing approach involved classroom and coaching observations, interviews with teachers, departmental heads, coaches, and principals, and collection of school-level context data. Interviews with implementing partners, programme designers, and funders were also carried out. The qualitative methods sought to understand:

- **Context**: the factors that could either drive or limit the adoption of the DH coaching model by schools;

- **Pathways and mechanisms**: how and why the DH coaching model works to change teacher practices in the classroom and therefore results in changes in learner outcomes;
- **Comparison**: the operative differences between external coaching and DH coaching pathways and mechanisms that may explain differences in impact between the two models.

Finally, the evaluators provided developmental support to the implementing agency in generating monitoring data. Despite limitations in the quality and completeness of the implementing partner's final monitoring data, it was also cautiously used to triangulate fidelity findings.

Instrument design

The evaluators and the Department of Basic Education (DBE) collaborated to develop the EGRP learner assessment instruments for both English First Additional Language (EFAL) and Home Language (HL) by adopting and adapting assessment tools used in previous studies conducted by the DBE, including the Early Grade Reading Studies (EGRS I, EGRS II) and the Reading Support Programme. Some elements of these tools were retained to maximise compatibility with other past and current DBE-led studies. In contrast, others were newly designed based on post-EGRS developments, such as the DBE's reading benchmarks project.

The study introduced group tasks in the EGRP assessments, which had not been used in previous EGRS studies. While this means there is no basis for comparison with earlier studies for these subtests, the addition of group tasks has been shown in other African countries to produce greater differentiation between learner scores and fewer floor effects. Group tasks also allow the introduction of simple writing tasks. All tools underwent a thorough pilot phase and received approval from the DBE before data collection.

The table below shows the sub-tasks assessed for each Grade. It also shows which subtests were included in the principal component analysis (PCA) generated to represent the aggregate learner achievement score used to compare outcomes across intervention streams. Subtests excluded were removed due to their independence from the intervention, as observed by the differences in their distributions.

Table #: Learner Assessment Sub-Tasks per Grade & Sub-Task Inclusion in PCA

Assessment	Sub-task	Grade 1 (2021)	Grade 2 (2021 & 2023)	Grade 3 (2023)	Grade 4 (2023)
HL one-on-one	Rapid Object Naming (RON)				
	Letter Sound Knowledge (LSK)			✔	
	Complex Consonants and Diacritics Knowledge (CCDK)			✔	
	Word Recognition/Reading (WR)		✔	✔	✔
	Oral Reading Fluency (ORF 1) 1 (1 minute)		✔	✔	✔
	Reading Comprehension 1 (ORF 1 Comp)		✔	✔	✔
	Oral Reading Fluency 2 (ORF 2) (1 minute)			✔	✔
	Reading Comprehension 2 (RC)			✔	✔
HL group test	Letter Sound Knowledge				

	Complex Consonant Sound Knowledge	✔		
	Word Recognition and Writing (WRW)	✔		
	Written comprehension (WC)		✔	✔
	Early Grade Mathematics test (EGMA)			
EFAL one-on-one	Oral Reading Fluency 1 (1 minute)	(2021 only)	X	X
	Reading Comprehension 1	(2021 only)	X	X
EFAL group test	Word Sound Knowledge	X		
	Written comprehension			X

Shaded   Subtest administered to learners in Grade
✔   Subtest included in HL PCA
X   Subtest included in EFAL PCA

Quantitative Sample Achievement and Learner Attrition

For the midline, data was collected over three weeks in November 2021. The endline data collection took place in late October 2023.

As shown in Table XXX, at midline, 2,912 Grade 1 learners and 2,897 Grade 2 learners were assessed. At the endline, 2,289 and 2,300 learners (respectively) were successfully reassessed from those cohorts. An additional 1,063 learners in Grade 2 in 2023 were also assessed. In total, across all cohorts and both midline and endline, 11,461 assessments were completed. Learner assessments are spread proportionately across the treatment groups.

Table XXX: Learner Sample Achievement

Cohort	Midline (2021)			Endline (2023)				
	Target (20 Gr1&2 learners per school)	Number achieved*	Percent Achieved	Target (all learners assessed at midline & 10 new cohort learners per school)	Revised Target (18 Gr3&4 learners & 8 new cohort learners per school)	Number achieved	Percent Achieved (Based on original target)	Percent Achieved (Based on revised target)
Cohort A: Grade 1/ Grade 3	2780	2912	**105%**	2912	2484	2289	79%	**92%**
Cohort B: Grade 2/ Grade 4	2780	2897	**104%**	2897	2484	2300	79%	**93%**
Cohort C: Grade 2 (2023)				1380	1104	1063	77%	**96%**

The midline sample was intentionally designed to be larger than necessary to ensure adequate statistical power at the endline, accounting for potential attrition. For the endline target sample, the evaluation partners agreed to include more extensive assessment tools and an additional Grade 2 cohort, leading to time constraints on the number of assessments that could be completed in each school within a day. Despite efforts to locate and reassess all learners who were evaluated at midline, some who were present could not be reassessed due to these time limitations. Approximately 21% of the learners in each cohort were not reassessed (see Figure XX for attrition and repetition rates). The primary reasons for learners not being assessed, in order of significance, were transfers to other schools, absenteeism on the assessment day, insufficient time for assessment, and learners either refusing to participate or not completing the assessment. Although this attrition rate is relatively high compared to other studies in the EGRS series, a regression analysis of attrition revealed no significant differences in attrition rates between the two coaching

treatment groups and the control group for both cohorts, indicating that attrition has not impacted the average cohort results.
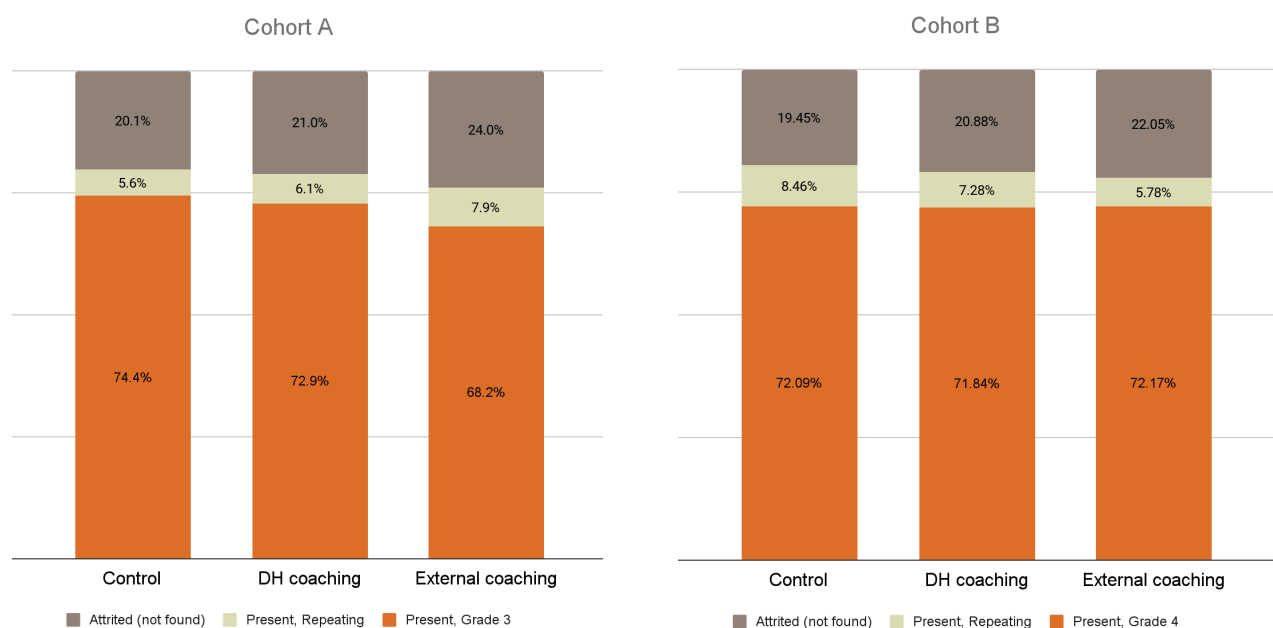
Cohort A

Cohort B

Figure XX: Attrition and Grade repetitions among learners from Cohort A (Grade 1 in 2021, Grade 3 in 2023) and Cohort B (Grade 2 in 2021, Grade 4 in 2023)

Data analysis

Case study and coach shadowing data were analysed using basic thematic analysis with Interpretative Phenomenological Analysis (IPA) coding to identify similarities and differences across teachers, coaches, and schools. The emphasis was on understanding the experiences of those involved in the programme. The EGRP case study report (2024) contains detailed qualitative findings, with key insights reflected in this report. Descriptive analysis of the school-based respondent surveys (with teachers, DHs, principals, etc.) was done in Tableau. Learner assessment data analysis processes are described in the chapter on learner assessment findings below.

Implementation Fidelity Findings

Implementation fidelity lies at the core of every intervention since the intended impact cannot be achieved when inputs and activities are not completed as planned. Insights into the strengths and weaknesses of EGRP implementation inform our understanding of the final learning outcomes. They also provide important insights on the conditions for scalable implementation of future interventions.

While in any intervention there may be necessary adjustments to implementation plans as learning occurs, and unintended positive impacts may arise from partial or adapted implementation, the intention of the EGRP was to test a specific intervention design for scaled application, meaning that fidelity and replicability is particularly important. We therefore designed the evaluation methodology to collect extensive fidelity data and describe the fidelity findings in detail. This report focuses on the fidelity of four key areas of programme delivery: Training, Coaching, LTSMs, and Teaching Practice, as outlined below.

It emphasises dosage, coverage, quality, participant responsiveness, and in certain instances, adherence.
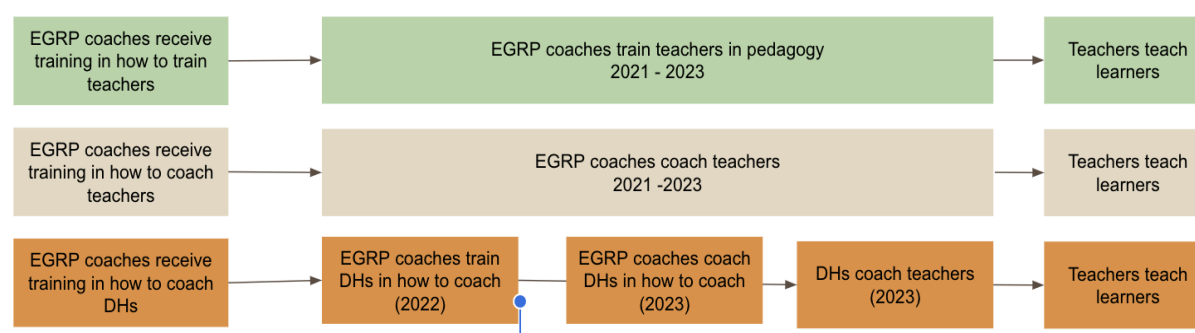
Figure XX: Implementation Fidelity Focus Areas

The fidelity insights described below are based on qualitative and quantitative data collected in different phases throughout the evaluation. Primary data was collected from over 800 respondents through midline and endline surveys, training observation, classroom observations and interviews during case studies and coach shadowing, and key informant interviews. Respondents included teachers, principals, Foundation Phase DHs, expert literacy coaches involved in the programme, implementing agency teams, the DBE, UNICEF and the ZENEX Foundation. Secondary data was sourced from the DBE's monitoring activities and the implementing partner's ongoing monitoring data, although some components of the latter were either unreliable or incomplete. This data supplemented the evaluation's primary findings and offered additional insights into specific indicators, particularly training quality, LTSM distribution and training attendance. A comprehensive understanding of the EGRPs implementation landscape was gained by triangulating insights from these diverse sources.

Table X: Number of teachers and departmental heads surveyed

Treatment	Teachers Surveyed (DHs surveyed)	
	Midline	Endline
External Coaching	110 (40)	87 (25)
DH Coaching	106 (38)	86 (29)
Control	163 (58)	145 (58)
Total	**379 (136)**	**318 (112)**

Figure XXX summarises the extent to which the different aspects of the EGRP theory of change were implemented with fidelity. Each component was assessed based on dosage, coverage and quality and rated to represent how well it was executed:

- Green Indicates that the element was executed successfully, with targets either met or exceeded, achieving a dosage of 81% or more.
- Amber Indicates that this aspect of the ToC was partially successful, with 50% to 80% of the dosage achieved
- Red Indicates that the component encountered significant challenges or was not implemented as intended, achieving 50% or less of the dosage.

Figure XXX: EGRP Theory of Change with Implementation Fidelity and Outcome RAG Rating[5]
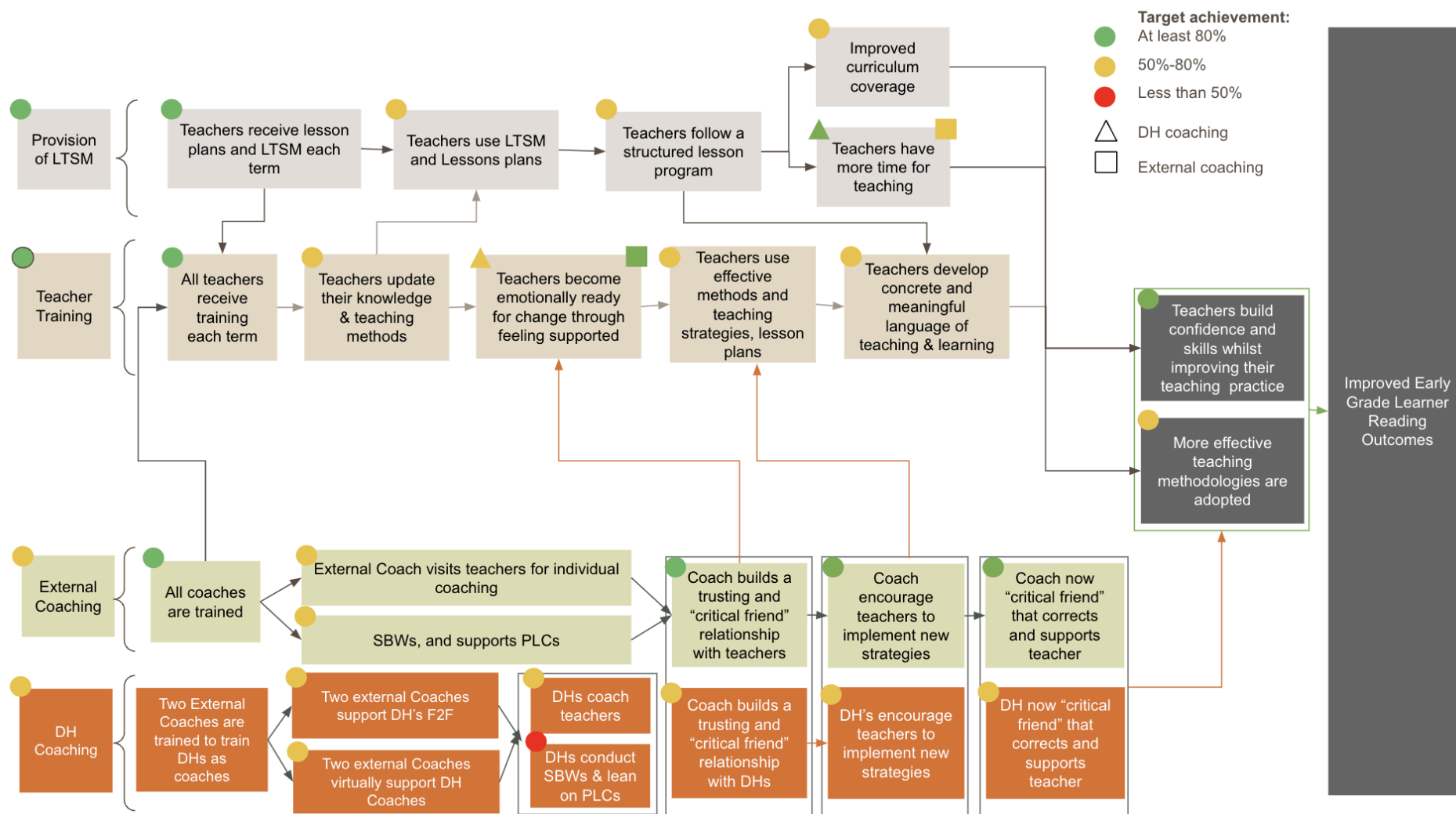
The subsequent sections of the report describe findings for each element of the Theory of Change in more detail. In this section we consider 'inputs' and 'outputs' (e.g. were planned tools distributed and activities completed at the required dosage, coverage and quality) and 'immediate outcomes' (e.g. were activities completed in a way that led to the expected behaviour change in participants and what evidence do we have that such behaviour change has occurred).

LTSM

The adaptation and distribution of LTSM was part of the base intervention included for all participating schools. There are two elements of this: lesson plans for the teachers to use in planning and structuring lessons, and printed reading materials for learners and teachers to use in the classroom (including Big Books, Sight Words Flash Cards, Discussion Posters, Phonic Friezes and graded readers in HL and EFAL). The lesson plans form the core of the structured reading programme and so their regular use by teachers is one of the most important impact chains in the intervention design.

An assessment of the quality of LTSM quality was not part of the current evaluation, but the materials were largely adapted from existing lesson plans and reading materials used in previous interventions and there was no indication during this study that there were concerns about their content.

Printed reading materials

Overall, the distribution of printed reading materials was good. Surveyed teachers across all programme schools reported receiving the relevant material in year 1, although it was late in Terms 1 and 2, and all top-up materials were received in Years 2 and 3. In the programme's inaugural year, a delay in the procurement process made it impossible for the prescribed EGRP EFAL reading materials to be produced and printed on time resulting in materials from a prior programme being used in Term 1 only. One large school reported receiving less than half of the required learner materials to implement some methodologies successfully (such as group-guided reading), but other large schools had sufficient materials. Many teachers expressed concerns about the durability of printed materials. By Term 3 of 2023, 13% of teachers indicated that their materials were still in excellent condition, 53% reported that their EFAL and HL materials were not in the best condition but still functional, and the rest said their materials were in poor condition. Wear and tear of printed materials suggests that teachers are using materials during their lessons but also confirms that printed materials must be regularly refreshed to enable ongoing availability for interactive reading practices such as paired and group-guided reading.

Scripted lesson plans and other digital materials on tablets

Similarly, while there were initial challenges with the distribution and use of tablets and the EGRP app in 2021, these were improved in 2022 and 2023. In mid-2021, teachers reported delays in receiving lesson plans on their tablets and case studies revealed that most teachers were not using their EGRP tablets in their classrooms as they were not confident in

using digital lesson plans. The process of updating and uploading app data was improved in subsequent years, resulting in increased app use in 2022 and 2023. 83% of teachers surveyed in 2023 reported that their tablets were still in working order.

The EGRP app's intention was to enable digital monitoring of lesson plan use and, therefore, curriculum coverage. The initial app design did not allow access to the app's back-end metadata, and when data became accessible for 2022 and 2023, it was incomplete. This represents a lost opportunity for linking teacher coaching (both external and DH) closely with digital evidence of lesson plan usage and is an area for improvement in future interventions.

Given the app data's incompleteness, it is impossible to reliably ascertain if lesson plan usage differed by intervention stream. The partial data suggests more regular usage by teachers receiving external coaching compared to the control group or to those who received DH-led coaching. This might reflect better curriculum coverage amongst those who received coaching, though it could merely reflect more compliance with accessing electronic lesson plans prompted by the external coach. The limited 2022 and 2023 EGRP app data as analysed by the DBE revealed a concerning overall trend across all treatment groups: that the final two weeks of the curriculum in each term were inadequately covered by teachers. This raises questions about the efficacy of the instructional support mechanisms put in place by DH coaches and external coaches to ensure comprehensive coverage of the curriculum to optimise learning outcomes.

Despite these data uncertainties about the consistency of lesson plan use, the mid-2023 classroom observations in the six case study schools showed a great improvement in actual classroom practice around lesson plans. Compared to 2021 lessons, which were often slow, the evaluation team observed most teachers using the tablets in class, showing confidence in their use, and implementing correctly-paced lessons with a wider range of activities as set out in the lesson plans.

Overall, as shown in Figure XXX above, we consider the LTSM component of the base intervention to have been implemented well at the level of delivery, but with significant caveats on the extent to which teachers consistently used the lesson plans and implemented the structured learning programme.

Teacher Training

Teacher training is the second element of the EGRP base intervention, complementing the provision of LTSM and scripted lesson plans. The implementation of this component was generally well executed in terms of coverage, dosage, and quality, although some logistical challenges were noted, particularly in the initial phases of the programme.

Teacher training achieved high coverage, with 98% of teachers at participating schools attending at least one training session. However, the dosage of training was slightly compromised in 2021, where only 72% of teachers attended all the training sessions (see Table X below). The main reasons cited for poor attendance were logistical challenges, such as transport difficulties and scheduling conflicts. Additionally, delays and changes to training dates due to administrative issues (e.g. venue availability and transport arrangements) were recurrent throughout the programme. Teachers frequently suggested that training dates be

communicated before they plan for the term ahead to avoid conflicts with their existing term plans.

Training attendance in 2022 mirrored the previous year's challenges, but there was a significant improvement in 2023, with over 85% of teachers attending all sessions. Of those who missed sessions, over half received catch-up training (see Table ==XX==). Teachers in the external and DH coaching arms received catch-up training during coach visits, which helped mitigate the impact of missed initial training.

==Table X:== Training coverage and dosage

Treatment	2021 (n=371)				2023* (n=309)			
	External Coaching	DH Coaching	Control	Overall	External Coaching	DH Coaching	Control	Overall
Coverage ≥ *100% of teachers attend at least 1 training session*	99%	98%	96%	**98%**	99%	100%	97%	**98%**
Dosage ≥ *90% of teachers attend all training sessions*	81 %	70%	67%	**72%**	89%	87%	86%	**87%**
Catch up trained ≥ *95% of teachers who miss training receive catch-up training*	63%	55%	41%	**51%**	70%	60%	47%	**58%**

*Only includes Term 1 to Term 3 training

==Table X:== Teacher training catch up training attendance

Treatment	Midline (2021)			Endline (2023)		
	External Coaching	**DH Coaching**	**Control**	**External Coaching**	**DH Coaching**	**Control**
None	1%	3%	3%	1%	0%	3%
1 session	4%	4%	3%	2%	8%	4%
2 sessions	3%	1%	8%	7%	5%	8%
3 sessions	11%	19%	19%	55%	54%	59%
4 sessions	84%	78%	71%	34%	33%	27%

Observations from training sessions and feedback from teacher surveys indicate that the quality of training was generally good. Training adhered to the planned content despite logistical setbacks, such as instances of late loading of app data like lesson plans. However, the delivery style and teacher interaction varied depending on the trainer's experience and preparation. Notably, in 2023, some teachers reported a decline in training quality due to the introduction of a new, less experienced trainer.

The training content covered essential literacy components such as phonics, phonemic awareness, vocabulary, fluency, and comprehension, with varying emphasis across different

grades. Specific methodologies, including group guided reading (GGR) and shared reading, were also part of the training, alongside classroom management strategies. In the evaluation's teacher survey, over 95% of teachers found the training sessions very useful.

Despite the high quality of content delivery, the programme lacked a reliable post-training assessment to gauge teachers' understanding and retention of the material, particularly in areas of difficulty. During coach observation and lesson observation, evaluators identified teachers struggling with shared reading and GGR in particular. Although teachers' self-reported survey data indicates high adherence to the use of these methodologies, 2023 classroom observation continued to show teachers struggling with them, with some observed improvement in GGR. Furthermore, a consistent challenge reported by teachers across both survey rounds was managing large classes with students of varying ability levels. This issue points to a need for more targeted support and training in differentiated instruction and classroom management techniques to equip teachers to handle diverse learning needs.

Foundation phase DHs are generally also foundation phase teachers and so were included in the teacher training component of the intervention and exposed to the structured learning programme alongside their colleagues. Since there was generally no process to assess trainees' comprehension of the training content, this was also not assessed for DHs. Given the importance of DHs in the EGRP design, it would be valuable to include such an assessment in future and to plan for additional targeted support in cases where DHs require it.

Coach Training

A prerequisite for effective coaching is that coaches are trained. There are three types of coach training in the EGRP model: training of external coaches, specific training for the two external coaches assigned to support the DH coaches, and training for the DHs.

External coach training

In 2021, the evaluation team observed coach training sessions in Term 2 and 3, while for 2022 and 2023 DBE training quality assurance monitoring reports were relied upon. These observations aimed to evaluate the nature, quality and content of the coach training and to assess the effectiveness of the EGRP coaching model. The training sessions were conducted by the implementing agency, with the coach coordinator, coaches, subject advisors, and district coordinators in attendance. The recruited coaches had all received specialised training prior to the EGRP since they had been involved in the RSP project, completed a coaching course, and had an education background either as teachers, trainers or subject advisors. The EGRP external coach training, therefore, mainly focused on the specific pedagogies included in the structured learning programme, assuming extensive prior knowledge of most aspects of literacy teaching and coaching practice.

Coach training sessions were conducted once a term over the three years of the programme. These trainings were intended to prepare coaches to:

- train teachers (all three years).
- coach teachers (all three years).

- train DH coaches (two coaches only as of 2022).

The expected training coverage was achieved, with all coaches attending at least one training session per year. However, the expected coach training dosage, where all coach training sessions are attended by all coaches, was not met. Given the importance of the coaches within the overall EGRP theory of change, differential training exposure and skill levels can adversely affect the rest of the impact chain.

In 2021, one coach was changed in August, and the new coach was only exposed to two training sessions that year. Another coach missed two training sessions. In 2022, coach training attendance data was unavailable for two coaches, but it is known that the expected training dosage was not met as not all coaches attended all four sessions. Two coaches were changed in 2022. For those coaches who missed training or joined the programme later, it is unclear how catch-up training was conducted to ensure that all coaches had the same knowledge base. This concern was raised by one of the coaches at the end of the first year, stating that they do not learn at the same pace or have the same knowledge base.

In terms of training quality, the evaluation team noted early on that the first coach training sessions mainly focused on preparing coaches to train teachers, with virtually no time spent on how to coach the teachers they were supporting. Considering the crucial methodological differences between training and instructional coaching (as defined in the literature review section above), the focus on training over coaching was a concern. In subsequent training sessions, however, coaching methodologies were visibly integrated. All the training sessions were well-organised with the requisite materials available. Although the engagement of training attendees was very high, and included activities like a mock lesson 'dry run' that gave coaches a chance to demonstrate what they have learnt and to display training and coaching expertise, the training was generally described as "facilitator-centric."

Training for the two External coaches supporting DHs

The two external coaches who provide support to DH coaches received their own training in August 2022, totalling 9.5 hours. It is unclear what kind of support or specialised training the two dedicated external coaches had received before this, especially in comparison to their external coaching stream counterparts. It is notable that the external coach training took place after the commencement of DH training in April 2022, suggesting that the initial DH training was generic coaching training that had not been adapted to the specific DH context. 2022 was intended to be the year when DHs would begin implementing DH coaching. However, due to delays in commencing their own training, they instead shadowed external coaches as part of their training. The training DHs underwent included modules on literacy coaching, integrating coaching with DH planning, classroom culture, a deeper understanding of listening and speaking, knowledge and pedagogy, planning for assessment, and Portfolio of Evidence (POE) review.

The dedicated external coaches also received a 5-hour virtual training session, but the implementing agency's monitoring data does not specify when this occurred in 2022. This session covered topics such as the analysis of pre/post test results, DHs role as a coach, instructional leadership, and building confidence in managing teachers.

The dependence of the EGRP DH coaching model on two individuals to train and support all participating DHs can be seen as a strength or a weakness for implementation at scale. It is a strength to the extent that it may be easier to recruit a small number of skilled individuals to sit at the apex of a cascade capacity building model than to depend on large numbers of variably skilled people. On the other hand, if those key individuals are not extremely skilled, or if there is turn-over in individuals, the impact chain for the entire intervention can be at risk. While there was no evidence to suggest that either of the two external DH-support coaches was unqualified, there was also no evidence that they had been selected out of all the other external coaches due to specific qualities or skills. The implementing agency did not seem to be intentional about whether direct external coaching and DH-support coaching required different skills or traits and what these might be, meaning an opportunity was lost to learn lessons about the external support coach selection and training components of the DH model. If a future scaled model is designed to depend on a similarly small number of individuals, which would be necessary to contain costs, it will be important to define clearly how to select, train and support those key individuals.

Training of DHs as coaches

According to the programme design, DHs were expected to begin coaching teachers in their schools at the start of the second year, indicating that their preparation should have commenced in 2021. However, DH training only began late in the fourth term of 2021, and the second year of the programme was used for DHs to shadow external coaches from April 2022. It was not until January 2023 (the final year of the programme) that DHs began coaching teachers independently, resulting in less than 10 months of independently DH coaching before the endline assessment in October 2023. This was a major deviation from the original DH coaching model, potentially impeding the ability to assess the effectiveness of this form of coaching within the evaluation time frame. Research indicates that full implementation of such models typically takes two years or more to yield measurable outcomes (Fixsen, 2005).

Despite the challenges, the training of DHs attended over and above the coach shadowing had shows that 96% of DHs surveyed reported that they attended at least one training session and most had received all of the 8 DH training modules by the end of 2023.

The implementation agency's monitoring data in 2022 shows that DH specific training was conducted virtually and 48 of 49 prospective DH coaches received some form of training (Virtual coaching coverage = 98%, dosage is unknown due to lack of reliable data)

Figure XX: DH Coach Training Module Coverage *(n = 27)*

DH Coach Training Module Coverage chart:
- Classrooom culture__: 89%, 24
- Instructional leadership__: 85%, 23
- Literacy coaching__: 96%, 26
- Pedagogic knowledge__: 85%, 23
- None of the modules above__: 4%, 1
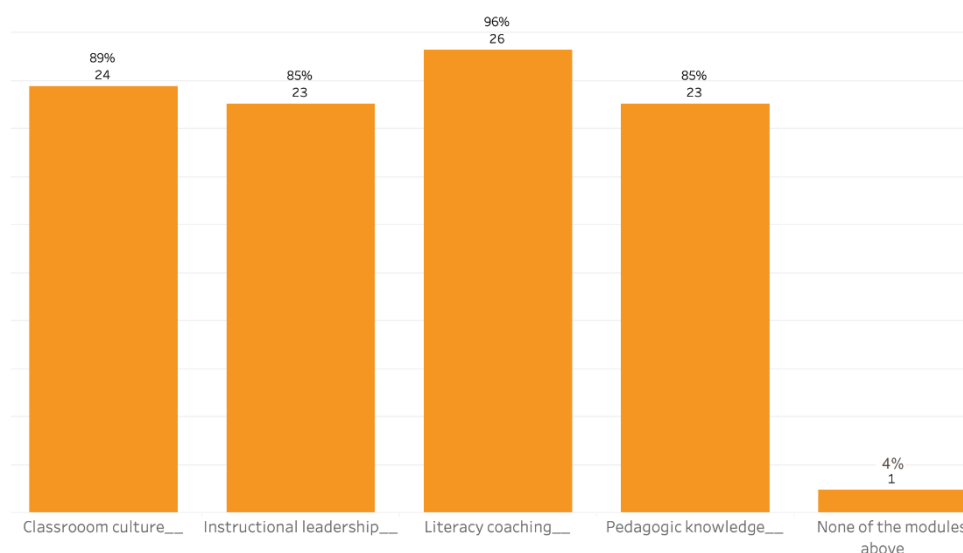
Data from the Department of Basic Education quality assurance reports in 2023 indicated a 97% engagement rate among DH trainees, with most questions and concerns from trainees being addressed appropriately. The high engagement reflects a strong interest of DHs to ensure they are equipped to take on the role. DHs were asked how useful they perceived each of these training topics to be (Figure XX). The majority found them to be useful. Adherence to training content was also found to be high.

Figure XX: Quality of DH training:

Quality of DH training chart:

Category	Very useful, I learned a lot	Not useful to me. I already knew the content	Refused to answer
Classroom Management	8 / 80%	1 / 10%	1 / 10%
Curriculum Coverage	8 / 80%	1 / 10%	1 / 10%
Instructional Leadership	8 / 80%	1 / 10%	1 / 10%
Observation Feedback	9 / 90%		1 / 10%
Soft Skills	8 / 80%	1 / 10%	1 / 10%
Subject Meetings	8 / 80%	1 / 10%	1 / 10%

Legend:
- Refused to answer
- Not useful to me. I already knew the content...
- Somewhat useful, learned some new things
- Very useful, I learned a lot

However, the DBE reported uneven timing of the DH training sessions: they did not occur in Term 4 of 2022 and Terms 2 and 3 of 2023 as planned. To compensate for the missed sessions, 2023 Term 1 training occurred over two days and Term 4 training was extended to three days of which only two took place due to district-level concerns about the feasibility of DHs travelling long distances for three consecutive days. This uneven distribution and overall reduction in training time likely impacted the depth and breadth of training content that could be covered, potentially affecting DHs' preparedness for their coaching roles.

In summary, both coaching streams were compromised from the outset by weaknesses in the coach training processes, especially with regard to the training of DHs.

External vs DH coaching coverage, dosage and quality

External coaching was initiated in all 80 coaching schools in 2021. As described above, from 2022 onward, six coaches continued coaching 40 schools directly while two coaches supported the 40 schools in the DH coaching stream.

While all teachers in both coaching streams received at least one coaching session each year (representing good coverage), the external coaching intervention consistently failed to meet dosage requirements over the three years (Table X). In 2021, only 13% of teachers received the full expected dosage of 12 to 13 coaching visits. According to the endline teacher survey, 69% of teachers in the external coaching stream and only 47% of teachers in the DH coaching stream received the prescribed coaching sessions in 2022. The lower coverage in the DH coaching stream can be attributed to the higher coach to school ratio (1:20) which the original design intended as appropriate for a 'light-touch' DH support role but in 2022 actually required external coaches to be present in schools directly coaching teachers while DHs observed.

Table X: External and DH coaching coverage and dosage

	2022		2023	
	External Coaching (n=58)	DH Coaching (n=55)	External Coaching (n=58)	DH Coaching (n=56)
Coverage ≥ 95% of teachers are coached	100%	95%	100%	96%
Dosage **External Coaching:** all teachers should received ≥ 13,5% visits **DH Coaching:** All teachers should receive ≥ 1 visit per month	62 %	47%	47%	42%

Note 1: 2023 DH coaching dosage target was adjusted to 10 visits because data collection was in November.
Note 2: 2022 DH coaching data based on shadowing EGRP DH coaches

In 2023, the coaching dosage declined to an average of 47% in the external coaching stream. Notably, across all three years, some teachers received significantly more visits than prescribed (Table XX and Table XX below), which compromised the coaching opportunities for others. In 2022, 62% of teachers received more than the 13 prescribed visits. Similarly, in 2023, almost all teachers who met the dosage requirement received more visits than necessary.

While some uneven distribution of coaching attention might reflect appropriate adaptation to teacher needs, with weaker teachers receiving more attention than stronger teachers, there was no evidence of systematic rating of teacher capacity and tracking of teacher improvements based on the additional coaching attention received. For example, had lesson plan use data from the EGRP app been available to coaches and systematically used to target coaching resources and time, differential dosage may have been justified. In the absence of such evidence, the uneven distribution of coaching sessions suggests an important breach in programme fidelity which is likely to have undermined the intended support structure for the use of prescribed programme materials and methodologies, potentially affecting the overall efficacy of the programme.

Figure XX: Number of teacher lesson observations conducted by DH coaches experienced by teachers *(n = XXX)*
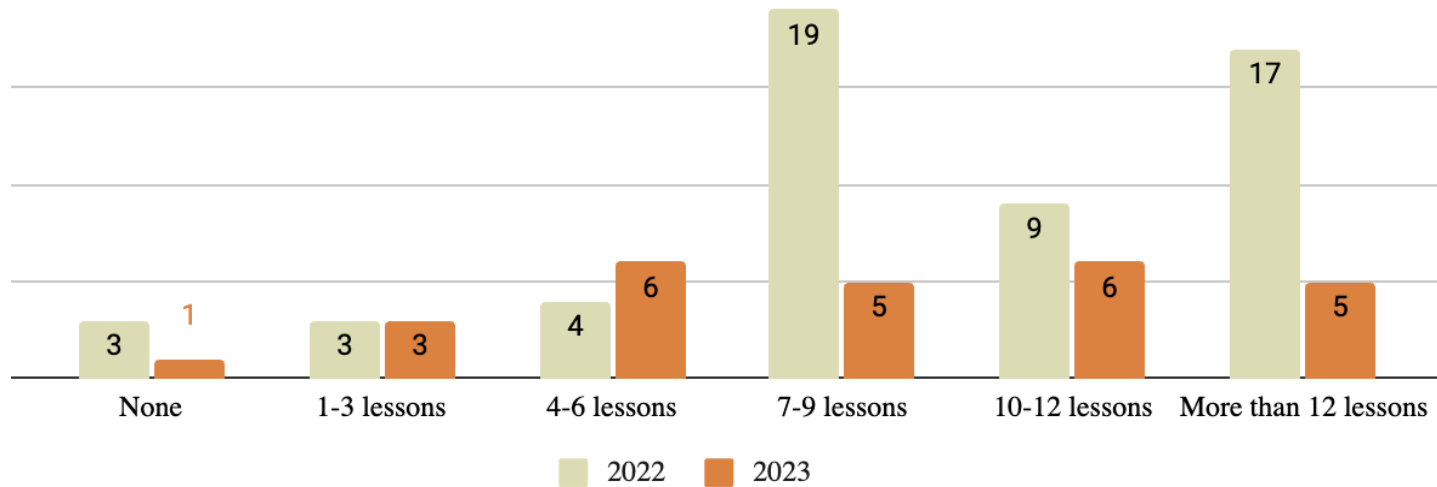
Figure XX: Number of teacher lesson observations conducted by External coaches experienced by teachers *(n=XX)*
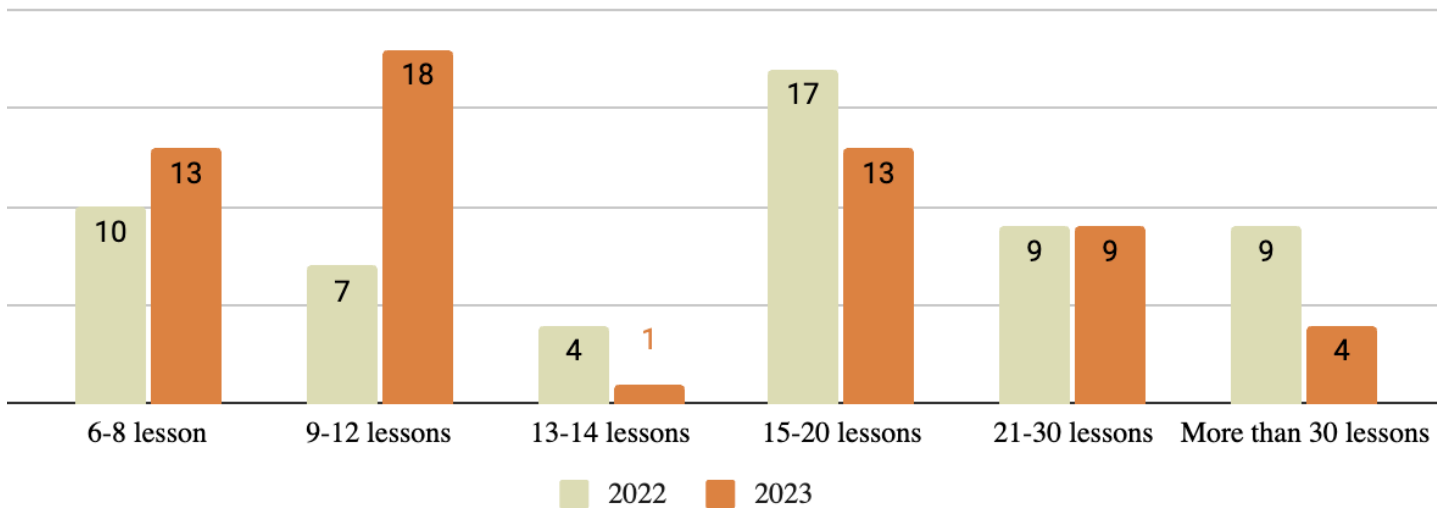
Virtual coaching was also implemented, showing higher levels of fidelity for DH coaches who were supported by their external coaches, but less so for external coaches directly supporting teachers. The dosage for DH coaches was set at a session at least once per month, with a self-reported fidelity of 90% (n= xxx) of teachers confirming they received support. In contrast, the target for direct coaching was weekly engagement, with a reported fidelity of 65% (n = xxx) of teachers confirming they were supported. Interviews with coaches, however, revealed a lack of responses during weekly engagements with teachers, noting that the conversations were often one-sided, with only sporadic engagements from the teachers.

Figure XX: Virtual coaching dosage *(n=xxx)*
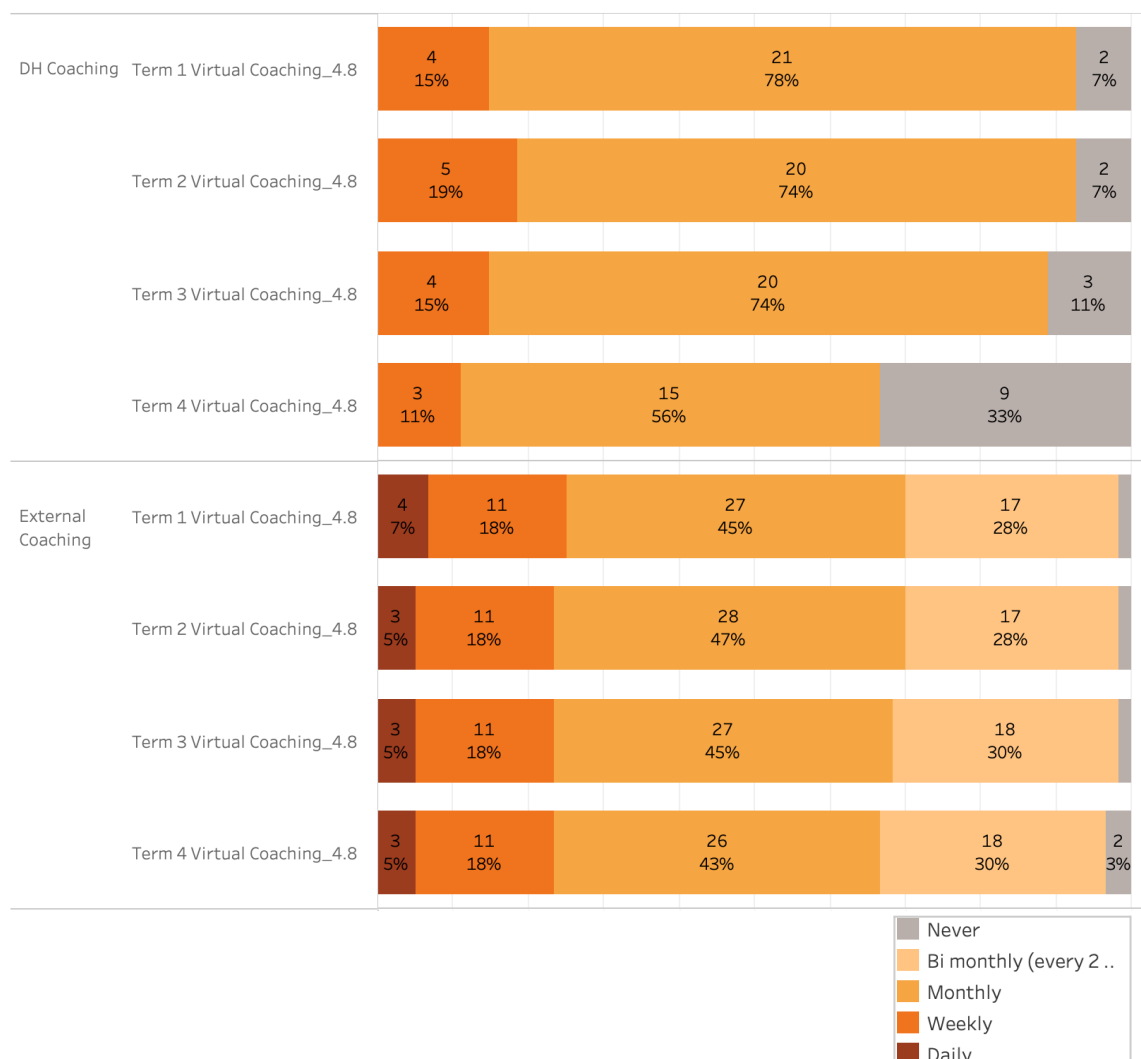
Regarding the quality of coaching in general, there were also variations in the quality of engagement between coaches and teachers depending on the coach's level of commitment to the programme and the nature of the relationship built with teachers. External coaches were very competent and teachers welcomed their feedback, except for one who was observed to give inputs that did not accord with teachers' actual practices. The level of commitment varied greatly, however, among DHs acting as internal coaches, as observed through the evaluation case studies. Two DHs were highly motivated - one mature with considerable experience, the other young and passionate about teaching and reading generally - and teachers in both their schools valued their coaching and respected their knowledge and skill. On the other end of the spectrum were two elderly DHs at the cusp of

retirement who did not accept the coaching role. In 2021, both these DHs were seen to be committed and effective in their conventional DH roles, but at endline it seems the EGRP expectations of the DHs to be coaches set the schools back and eroded the DHs' commitment to early grade reading. This may be as a result of the programme being introduced as one they had to adopt rather than one they could choose voluntarily.

One of the most important fidelity findings related to the quality of coaching by both external and DH coaches in terms of the expected components of instructional coaching. As discussed in the literature review, instructional coaching is intended to be an individualised relationship that is supportive and developmental rather than compliance and evaluation-driven. While external coaches were consistently observed to be technically competent in the EGRP literacy methodologies, only two of the eight coaches took a developmental approach to their coaching roles in that they assisted teachers to identify knowledge and skills they need to enhance and guide them in how to build their own individual skills base. The coaching was generally generic, following standard content and methodological steps, and was not differentiated for the needs of a particular teacher.

This standardised approach was passed on to DH coaches, although the DH coaches tended to be more pragmatic and more allowing of innovation than their external coach counterparts. Also important is the level of confidence demonstrated by the DHs, especially those individual DHs or external coaches who took a more developmental approach to their coaching practice by shifting away from a compliance driven approach. Overall, DHs' performance varied, with confident DHs excelling through proactive development and weaker DHs struggling with feedback and preparedness. Confidence emerged as a key differentiator in DH performance.

Despite some positive cases of good coaching practices from DHs, most observed DH coaching was not developmental and did not centre teacher self-reflection and problem-solving in its practice. Where pre-or post-observation conferences were held with teachers, both the external coaches supporting DHs and the DHs dominated these sessions rather than encouraging teachers to reflect on their own practice. DHs tended to focus on compliance with the application of EGRP literacy strategies. Perhaps this expectation for improved practices is premature, considering the delay in DH-led coaching which only effectively began in the final year of the intervention. Implementation evidence in education research suggests that full implementation typically takes 2-3 years or longer to reflect data that evaluation teams can use to assess the success of innovation (Fixsen, et al., 2005).

The combination of limited independent DH coaching time with teachers (late onset and low dosage) and a compliance-driven and standardised rather than individualised developmental approach to coaching suggests that the DH coaching stream did not in fact implement the intended methodology of instructional coaching which requires consistent, intensive, individualised and developmental relationship-building between coach and teacher.

School-based workshops
School-based workshops (SBWs) are a crucial element of the coaching design for both DH and external coaches, intended to be held with multiple teachers after lesson observations as a form of Professional Learning Community (PLC) within each school. However, in 2021, no monitoring data was reported for the SBWs and PLC activities, and triangulated primary

data from teacher surveys also indicated that PLC formation and operations were not prioritised.

The inconsistent application of school-based workshops, PLC development, and the generally low coaching dosage across both coaching streams in the first year of EGRP implementation represented a missed opportunity. In the post-COVID-19 school lockdown environment, these interventions were critical levers that could have supported schools and teachers in managing learning losses and classroom challenges related to the pandemic.

Figure XXX: SBW regularity by intervention arm and year (2022 and 2023) (Teacher Survey)
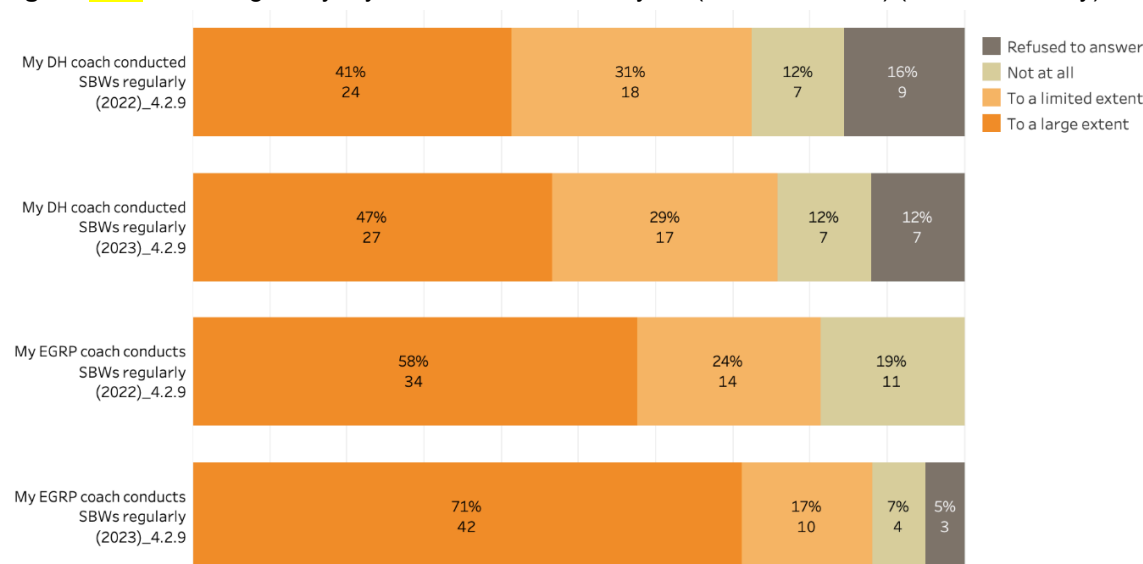
In 2022 and 2023, most external coaches carried out SBWs, with 58% of teachers in the external coaching stream reporting that their coach regularly held SBWs in 2022, increasing to 71% in 2023. Although DHs were present in schools and so could have conducted SBWs with their teachers more easily, they did so less regularly, with 41% of teachers reporting regular SBWs in 2022, which slightly improved to 47% in 2023.

Across both coaching streams, the implementation fell short of the required dosage (85% of visits including workshops) and coverage (all teachers experiencing SBWs). In terms of quality, DH coach SBWs were observed to be mostly dominated by the coach, with teachers remaining passive throughout the workshops. Teachers generally felt that these sessions just reiterated the content of the training sessions and did not add much value, rather than fulfilling their intended purpose of enabling peer support and practical problem solving among teachers as a group. The sessions ranged between 30 and 60 minutes.

The uneven and inadequate application of SBWs posed a significant gap in the programme's implementation, particularly considering the critical role these elements play in reinforcing training, fostering collaborative learning, and addressing ongoing challenges in the classroom.

Outcome-level Findings

The combined application of the LTSM, training and coaching components is intended to result in two key outcomes in the EGRP theory of change:

- Teachers build confidence and skills while improving teaching practice
- More effective teaching methodologies are adopted

These are necessary steps toward achieving the intended impact of improved learner reading performance.

To understand the programme's performance at an outcomes level, primary data sources were utilised, including surveys with teachers and DHs collected in Term 4 of 2021 and 2023, along with classroom observations and coach shadowing activities conducted midyear in both years. In addition, in-depth and semi-structured interviews with teachers, DHs, the eight coaches, and the coach coordinator were conducted in 2021 and 2023. These diverse data sources provided a comprehensive view of the programme's effectiveness. Surveys offered quantitative insights, while classroom observations and coach shadowing provided qualitative context.

Teacher confidence

The coaching theory of change posits that teacher confidence and self-efficacy are enhanced through coaching, enabling teachers to adapt literacy strategies to their contexts. However, the evaluation showed that most DH and external coaches focused more on compliance with EGRP literacy strategies than fostering teacher self-reflection and problem-solving. This limited teacher innovation, though it might not be crucial for improving learner outcomes. Effective coaching is likely to require a longer period to transition from compliance to innovation and self-efficacy, suggesting extended studies are needed. Despite these challenges, some case studies indicated increased teacher confidence and innovation, particularly with those DH and external coaches who exhibited more developmental coaching practices.

Teachers reported significant changes in teaching practices and classroom environments, attributing these improvements to the EGRP. Figure XXX shows self-reported 2021 and 2023 data on how easy teachers find the implementation of some methodologies and activities in the classroom. Across all the treatment groups, teachers generally reported that it was easier to implement activities in 2023 than in 2021, with the exception of shared reading which teachers in control schools found harder in 2023 than in 2021. The teacher survey data showed no consistent differences across treatment groups, although teachers in control schools started off struggling the most with GGR and phonics and then mostly catching up with the other treatment groups by the endline. This suggests that increased confidence may be due to the lesson plans and training, rather than to reinforcement through either coaching model.

Not surprisingly, teachers across treatment groups were the least confident in their delivery of GGR in both years (see circle in Figure XXX). The evaluation team noted that although a handful of teachers had mastered the more challenging methodologies, such as GGR and shared reading, some had not fully grasped how to deliver them. Those who had grasped these also began to innovate and adapt the methodology to their classroom context or personal preferences, such as designing easels, having learners hold up the big book, putting up copies of big book pages on the classroom walls and experimenting with learner seating arrangements. The range of observed innovations in GGR approaches was broader

in the schools with DH coaches as the DHs tended to be more pragmatic and less prescriptive in relation to their colleagues than the external coaches.

Figure XXX: Teacher-reported ease of implementing activities at midline and endline[6]
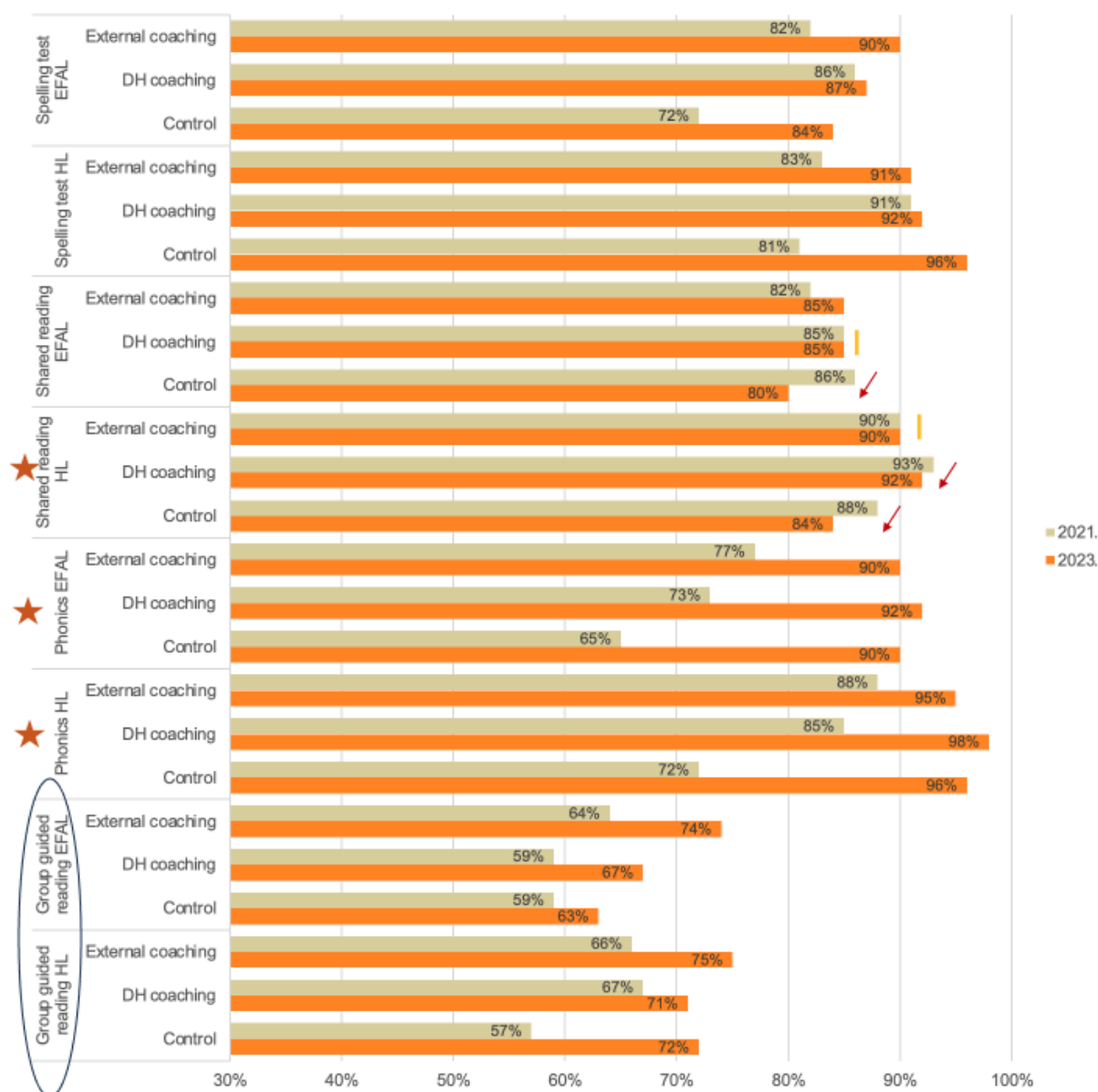
Teachers were also asked questions about feeling supported by their school management team (SMT). Despite SMT support not being a central design feature of the EGRP, there were improvements across all the domains (recognition, mentoring and curriculum support, fostering reading culture, modelling of lessons, and observations done by SMT) in all the treatment groups, including the control group. External coaching consistently led to the most substantial changes across domains. DH coaching and the base programme also had a positive impact but were more modest. Overall, it seems that the base programme itself had a positive effect on the work of both teachers and the SMT.

---

[6] The x-axis is truncated for ease of interpretation

Classroom practice

Data on changed classroom practice is based on classroom observations during the 2021 and 2023 case studies in six DH coaching schools and the 2023 coach shadowing in external coaching schools. Evidence from teacher, DH and coach interviews is also considered.

Based on the evaluation team's past experience in assessing classroom teaching quality, it is generally noted that the quality of teaching reading in almost every classroom observed for the EGRP study (DH and external coaching) compares favourably with the best examples in rural and peri-urban schools a decade ago. This is partly due to the abundance of learning and teaching support materials (LTSMs) provided by the project and previous projects. This has resulted in print-rich environments and reading corners. However, an important gap is the absence of learner work displayed on classroom walls. This is a common feature of effective early-grade classes worldwide. There were no observable differences between the two coaching streams on this indicator.

A more essential long-term positive trend is that teachers observed in 2023 showed greater understanding of how learners progress through the various stages of learning to read. This has had a profound impact on the way they teach, especially in Grade 3, where in the past teachers too often focused on teaching letter and word sounds as they did not know how to teach learners to turn these words into sentences and paragraphs. Lessons are now fast-paced with a variety of activities, which teachers ascribe to their use of the scripted lesson plans. Teachers also expressed much less frustration with fitting the different components into a lesson compared with their feedback in the midline research. Two caveats to the benefits of these changed classroom practices are that a) some teachers feel that the fast timing further disadvantages slower learners, and b) classroom observations still showed a general lack of teacher-learner interactions in classrooms.

EGRP protocols suggest seating learners in their ability groups to facilitate the differentiation of tasks. This was increasingly observed at the endline, although Covid-19 social distancing protocols may have been the reason for limited application at midline. The advantage of learners being seated in this way was evident when GGR was conducted because teachers could provide different occupational tasks to learners based on their skill levels. This is a relatively high-level skill for teachers to master. One observed full-service school kept its standard seating protocol, grouping weaker and stronger learners together to allow stronger learners to assist weaker ones. It was positive to see that the external coach did not attempt to change this deferring to the school's specific context.

Overall in schools where classroom observations were conducted, the teachers were on track with curriculum coverage at the endline, and only one school was a week behind on the curriculum tracker. This was an improvement on midline observations and indicates that most schools have adapted to the expected pace of curriculum coverage. However, as highlighted earlier, the limited EGRP app data still indicates that many teachers did not complete the last two weeks of the curriculum each term in 2023.

There were no observed differences in classroom practice between the two coaching arms.

It was expected that if teachers used the scripted lesson plans, the time they spent preparing for lessons and other administrative tasks would be lower. Self-reported survey data indicates teachers spent less time preparing lessons across all three treatment groups. It was assumed that this time would be reallocated to teaching; however, only the DH coaching teachers reported an increase in teaching time, which may suggest that teachers did their lesson preparation outside of their school times.

Enabling and Constraining Factors in DH Coaching Practice

Observations from baseline and endline case studies, coach interviews, and DH surveys reveal interesting insights into DHs' constraints and opportunities in relation to the coaching role. Three specific system-level enabling factors for DH coaching, which could become barriers if ignored, emerged from the evaluation: time constraints, DH succession planning, and DH recruitment and promotion.

Time Constraints: At the beginning of the EGRP, time constraints for DHs were anticipated. It was assumed that their administrative and supervisory duties could be adjusted to include coaching. However, DHs have full teaching loads in addition to their administrative duties, leaving little time for coaching tasks. During classroom observations and DH interviews, it was found that DHs with Education Assistants (EAs), often present in schools as part of the separate Presidential Youth Employment Initiative (PYEI), used them to supervise classes while they conducted coaching, thus freeing up the necessary time. Although the EAs were not part of the EGRP design, their presence in many schools fortuitously identified a mechanism through which the DH time constraint could be addressed. While using largely untrained EAs to 'babysit' DH classes while the DH coaches other teachers may have unintended negative consequences for the learners in DH-taught classes, the training of EAs and their incorporation into structured learning programmes, especially for foundational literacy, is steadily improving in South Africa (see Moyo & Polzer Ngwato 2024) and so incorporating EAs into the system may be an essential ingredient to freeing up DHs' time for coaching.

Succession Planning: Managing DHs approaching retirement is a critical constraint. If the DH coaching model is scaled, an effective succession plan must be implemented. This involves strategies for engaging near-retirement DHs to either commit to the coaching role or pass it on to colleagues. Advance planning for transferring the coaching role when an experienced DH coach retires is also necessary.

Recruitment and Promotion: Scaling the DH model requires integrating coaching skills into recruitment and promotion procedures, unlike the current system of promotion based on seniority. This might be challenging given existing practices in managing DH succession and promotion by schools, education districts, and teacher unions, but discussions should explore ways to incorporate coaching skills into these processes.

Impact-Level Findings (Learner Assessments)

Learner Results against Reading Benchmarks[7]

Before comparing learner reading outcomes across the three treatment groups, we describe the basic reading fluency skills measured in the overall midline and endline samples against the DBE oral reading fluency benchmarks for each grade.[8]

In the 2021 midline assessment, learner performance in both Grade 1 and Grade 2 was very poor in HL and EFAL:

- **Grade 1** (HL benchmark:40 letters per minute): 17% of learners met or exceeded the benchmark and 14% could not sound out a single letter. 82% could not read a single word in their HL.
- **Grade 2** (HL benchmark: 40 oral reading fluency (ORF) correct words per minute(cwpm): 15% met or exceeded the benchmark and 59% could not read a single word.
- **Grade 2** (EFAL benchmark: 30 ORF cwpm): 9% met or exceeded the benchmark and 74% failed to read a single word.

At the endline, we see an overall improvement in home language reading in Grade 2 HL, with 35% of learners achieving the word reading fluency benchmark and only 32% unable to read a single word. EFAL ORF was not repeated in the endline Grade 2 assessment. This Grade 2 HL improvement from 2021 to 2023 probably reflects a general post-COVID-19 reading recovery rather than being attributable to the EGRP intervention.

Grade 3 and 4 endline assessments included two ORF HL subtasks and one ORF in EFAL. Results show some improvement in benchmark achievement compared to the midline results for the same cohort, but it is unclear how much of this is due to natural maturation effects.

- **Grade 3** (HL benchmark: 60 cwpm)**:** for the first ORF, 32% met the benchmark (18% zero correct), and for the second, 40% met the benchmark (17% zero correct).
- **Grade 3** (EFAL benchmark: 50 cwpm): 36% met the benchmark (22% zero correct).
- **Grade 4** (EFAL benchmark: 70 cwpm): 23% of learners met the benchmark (33% zero correct).

Grade 4 HL benchmarks are not yet available.

The large percentage of Grade 4 learners who could not read one word (33%), and the fact that this is a larger percentage than the Grade 3 group, may be due to COVID-19 learning backlogs that were never caught up, since the Grade 4 (2023) cohort missed all of Grade 1 and most of Grade 2 during COVID when the foundations and reading should have been laid.

These findings are summarised in Figure <mark>#</mark> which shows the general trend of increasing benchmark achievement and the reduction in non-reading from 2021 to 2023 within each

---

[7] Setswana HL and EFAL Benchmarking EGR skills in SA Technical report
https://pdf.usaid.gov/pdf_docs/PA00ZFC5.pdf
[8] This analysis is based on the consolidated learner results across intervention streams for matched learners in each cohort. Cohort A is 2289 learners, Cohort B is 2300 learners and Cohort C is 2300 Grade 2 learners in 2021 and 1063 Grade 2 learners in 2023.

cohort. The increased zero-score between Grade 1 in 2021 and Grade 3 in 2023 for Cohort A is because Grade 1 measures letter recognition while Grade 3 measures word reading. Overall, however, achievement of the benchmarks in 2023 remains low (36% and below).

Figure ##: Overall learner performance in Oral Reading Fluency against reading benchmarks by Cohort, Grade and Language
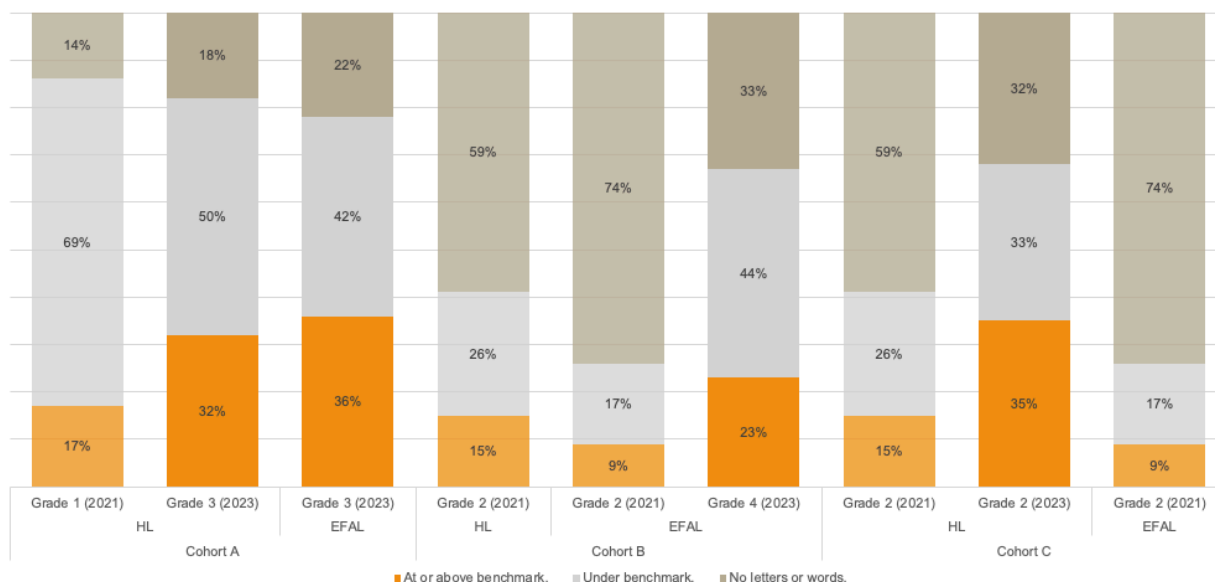
A density plot provides a descriptive analysis of the distribution of composite learner scores[9] for HL and EFAL and shows the overall pattern of learner reading skills within a Grade (combining all results across intervention streams). As shown in Figure ##, the distribution for HL is bimodal to varying degrees for all Grades, meaning that a portion of the class has very low reading skills (below -1 in the standardised composite score distribution), while another group is reading at acceptable levels (around 1 and above in the standardised score), with few learners in between. The Grade 2 (2021) pattern is more skewed to lower reading ability than the 2023 grades 2, 3 and 4, reflecting a general improvement in reading ability between 2021 and 2023. When comparing the 2021 and 2023 Grade 2 results we see how the above finding on improved benchmarks achievement is reflected as a positive shift in the overall distribution of learning. Despite this improvement, the 2023 Grade 2, 3 and 4 HL distributions still show around 40% of learners achieving at the low end of the score range.

---

[9] The construction of the composite learner score through principal component analysis is described below.

Figure ##: Density curves showing the distribution of HL and EFAL composite scores for HL and EFAL
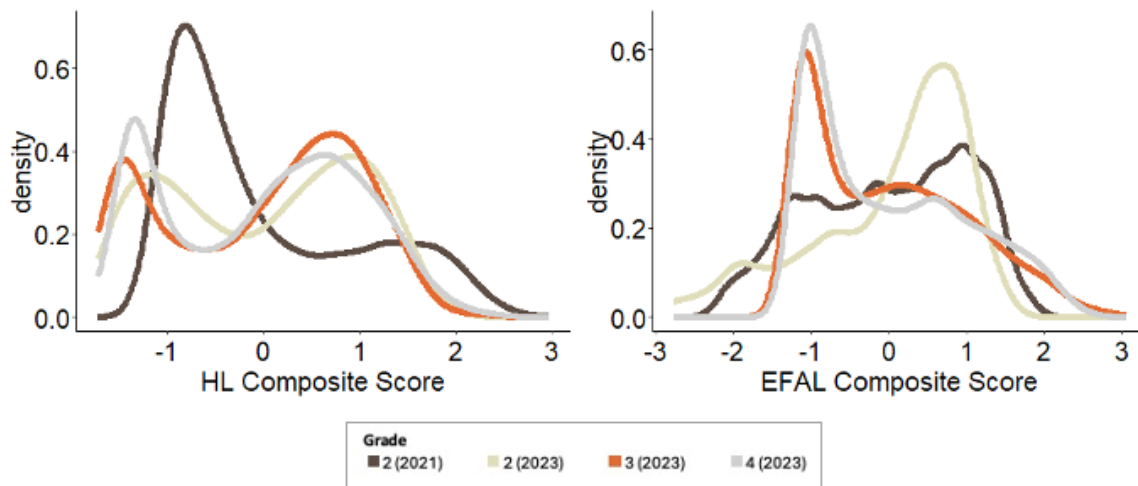
Studies have documented this bimodal distribution in early grade reading performance in several Sub-Saharan African countries, reflecting the stark differences in reading proficiency levels within the same classrooms and schools. While some learners are making significant progress, a large portion are 'behind the curve' for grade-level skill expectations, meaning that they either require concentrated remedial attention or remain at risk of falling ever further behind due to a lack of foundational skills. This bimodal distribution therefore places significant pressure on teachers, who are essentially teaching one class, using a curriculum designed for one ability continuum, while there are actually two distinct ability levels (or 'grades') within the same class.

For EFAL, the results are more normally distributed rather than bimodal, with a spread across performance levels in 2021 Grade 2, a skew towards positive results for Grade 2s in 2023 and a skew towards low performance for Grades 3 and 4. This suggests a continued need for intensive catch-up support to Grade 3 and 4 learners (2025's Grades 4 and 5), but also offers hope that the 2023 Grade 2s have substantially recovered from (or did not initially acquire) COVID backlogs.

Descriptive Analysis of Learner Assessment

The descriptive analysis indicated that there were minimal differences between the treatment groups across all cohorts. Slight variations were observed (with 1 to 2 more correctly answered questions) in Wave 2, Grade 2 subtask LSK (Table XX) and Grade 3 subtasks HL ORF ALT, EFAL ORF, and EFAL WR (Table XX). However, no differences were found for Grade 4 (Table XX).

For the subset of subtasks related to ORF across all cohorts, it was noted that there were very high ceiling effects. Most if not all learners were reaching the maximum number of correct answers during the 3 minute task, and even before the 3 minute duration. Upon closer inspection, from the distribution analysis of the ORF subtasks (Figure XX), and sanity checks with other subtasks (Figure XX), that the correct interpretation for ORF was to analyse it at the 1 minute mark, so words correct per minute. Additionally, the following outliers were also excluded:

For subtask ORF (Oral Reading Fluency) for endline grade 3 and 4, words correct per minute for HL was capped at 300 and for EFAL was capped at 250. Similarly, wave 2 grade 2, words correct per minute was also capped for HL and EFAL but at 150. Anything above this number did not make sense, this was cross-checked with potential errors due to technical system errors from Tangerine. This resulted in a total of 9 observations excluded:

- Grade 2 Wave 2 – excluded 1 observation
- Grade 3 – excluded 2 observations
- Grade 4 – excluded 6 observations

All other subtasks distributions were as expected: high floor effects, and bimodal distributions, excluding RON which was normally distributed (Figures XX in Appendix).

Table xx: Cohort C - Grade 2, 2023 Summary Statistics per subtask

Language	Subtest	Treatment Arm	n	zero	min	max	p10	p90	mean	sd	se
Setswana Home Language	Rapid Object Naming	Control	421	0.48%	0.00	29.00	11.00	22.00	16.40	4.35	0.21
		DH Coaching	329	0.00%	1.00	30.00	10.00	21.00	15.81	4.43	0.24
		External Coaching	313	0.32%	0.00	33.00	11.00	22.00	16.84	4.46	0.25
	Letter Sound Knowledge	Control	421	2.61%	0.00	70.00	9.00	70.00	44.78	21.63	1.05
		DH Coaching	329	3.65%	0.00	70.00	9.80	70.00	44.29	22.77	1.26
		External Coaching	313	0.96%	0.00	70.00	12.00	70.00	46.48	20.38	1.15
	Word Reading	Control	421	9.98%	0.00	55.00	1.00	38.00	16.63	14.44	0.70
		DH Coaching	329	13.98%	0.00	59.00	0.00	40.00	17.53	15.66	0.86
		External Coaching	313	8.95%	0.00	59.00	1.00	40.00	18.99	15.12	0.85
	Oral Reading Fluency	Control	421	33.02%	0.00	49.00	0.00	49.00	27.85	21.78	1.06
		DH Coaching	329	32.83%	0.00	49.00	0.00	49.00	27.83	21.71	1.20
		External Coaching	313	28.12%	0.00	49.00	0.00	49.00	30.71	21.23	1.20
	Oral Reading Fluency Comprehension	Control	421	42.28%	0.00	5.00	0.00	4.00	1.87	1.87	0.09
		DH Coaching	329	43.16%	0.00	5.00	0.00	4.00	1.81	1.84	0.10
		External Coaching	313	36.10%	0.00	5.00	0.00	4.00	2.12	1.86	0.11
	Complex Consonant Sound Knowledge	Control	421	3.33%	0.00	5.00	1.00	5.00	3.48	1.55	0.08
		DH Coaching	329	6.69%	0.00	5.00	1.00	5.00	3.43	1.62	0.09
		External Coaching	313	3.83%	0.00	5.00	2.00	5.00	3.78	1.44	0.08
	Word Recognition and Writing	Control	421	18.29%	0.00	5.00	0.00	5.00	3.05	1.90	0.09
		DH Coaching	329	17.93%	0.00	5.00	0.00	5.00	3.05	1.94	0.11
		External Coaching	313	16.29%	0.00	5.00	0.00	5.00	3.33	1.88	0.11
English First Additional Language	Word Sound Recognition	Control	421	2.38%	0.00	10.00	3.00	9.00	6.71	2.49	0.12
		DH Coaching	329	2.13%	0.00	10.00	2.00	9.00	6.53	2.62	0.14
		External Coaching	313	0.00%	1.00	10.00	3.20	9.00	7.26	2.22	0.13

Table xx: Grade 3, 2023 Summary Statistics per subtask

Language	Subtest	Treatment Arm	n	zero	min	max	p10	p90	mean	sd	se
Setswana Home Language	Letter Sound Knowledge	Control	978	3.48%	0.00	110.00	12.00	75.00	46.69	23.23	0.74
		DH Coaching	662	4.98%	0.00	110.00	13.10	77.00	48.09	23.34	0.91
		External Coaching	649	3.39%	0.00	110.00	12.00	76.00	47.37	22.65	0.89
	Word Reading	Control	978	9.71%	0.00	66.00	1.00	47.00	25.44	17.29	0.55
		DH Coaching	662	8.46%	0.00	72.00	1.00	46.00	25.95	16.76	0.65
		External Coaching	649	9.24%	0.00	71.00	1.00	48.00	26.28	17.05	0.67
	Complex Consonant & Diacritic Knowledge	Control	978	16.67%	0.00	45.00	0.00	41.00	21.14	14.92	0.48
		DH Coaching	662	16.62%	0.00	45.00	0.00	39.00	20.84	14.34	0.56
		External Coaching	649	15.56%	0.00	45.00	0.00	41.00	22.29	14.70	0.58
	Oral Reading Fluency 1	Control	978	20.04%	0.00	58.00	0.00	58.00	41.91	23.57	0.75
		DH Coaching	662	16.16%	0.00	58.00	0.00	58.00	43.33	22.31	0.87
		External Coaching	649	17.72%	0.00	58.00	0.00	58.00	43.11	22.69	0.89
	Oral Reading Fluency 1 Comprehension	Control	978	23.62%	0.00	7.00	0.00	7.00	3.68	2.55	0.08
		DH Coaching	662	19.94%	0.00	7.00	0.00	7.00	3.75	2.45	0.10
		External Coaching	649	21.73%	0.00	7.00	0.00	7.00	3.82	2.55	0.10
	Oral Reading Fluency 2 Comprehension	Control	978	26.18%	0.00	6.00	0.00	6.00	3.21	2.35	0.08
		DH Coaching	662	24.17%	0.00	6.00	0.00	6.00	3.25	2.29	0.09
		External Coaching	649	24.19%	0.00	6.00	0.00	6.00	3.27	2.29	0.09
	Oral Reading Fluency 2	Control	978	17.69%	0.00	64.00	0.00	64.00	47.43	25.01	0.80
		DH Coaching	662	16.31%	0.00	64.00	0.00	64.00	48.43	24.16	0.94
		External Coaching	649	14.95%	0.00	64.00	0.00	64.00	49.00	23.79	0.93
	Group Written Comprehension	Control	978	30.27%	0.00	7.00	0.00	6.00	2.92	2.47	0.08
		DH Coaching	662	29.46%	0.00	7.00	0.00	6.00	2.90	2.46	0.10
		External Coaching	649	29.12%	0.00	7.00	0.00	6.00	2.92	2.47	0.10
English First Additional Language	Word Reading	Control	978	24.64%	0.00	104.00	0.00	49.00	21.69	19.82	0.63
		DH Coaching	662	21.45%	0.00	100.00	0.00	47.00	21.55	19.11	0.74
		External Coaching	649	22.34%	0.00	82.00	0.00	51.00	23.20	19.93	0.78
	Oral Reading Fluency 1	Control	978	23.82%	0.00	71.00	0.00	70.00	41.43	28.76	0.92
		DH Coaching	662	21.00%	0.00	71.00	0.00	70.00	42.03	28.12	1.09
		External Coaching	649	20.34%	0.00	71.00	0.00	70.00	43.35	28.53	1.12
	Oral Reading Fluency 1 Comprehension	Control	978	27.81%	0.00	11.00	0.00	9.00	3.35	3.25	0.10
		DH Coaching	662	25.83%	0.00	11.00	0.00	8.00	3.11	3.01	0.12
		External Coaching	649	23.73%	0.00	11.00	0.00	9.00	3.61	3.23	0.13

Table xx: Grade 4, 2023 Summary Statistics per subtask

Language	Subtest	Treatment Arm	n	zero	min	max	p10	p90	mean	sd	se
Setswana Home Language	Letter Sound Knowledge	Control	978	3.48%	0.00	110.00	12.00	75.00	46.69	23.23	0.74
		DH Coaching	662	4.98%	0.00	110.00	13.10	77.00	48.09	23.34	0.91
		External Coaching	649	3.39%	0.00	110.00	12.00	76.00	47.37	22.65	0.89
	Word Reading	Control	978	9.71%	0.00	66.00	1.00	47.00	25.44	17.29	0.55
		DH Coaching	662	8.46%	0.00	72.00	1.00	46.00	25.95	16.76	0.65
		External Coaching	649	9.24%	0.00	71.00	1.00	48.00	26.28	17.05	0.67
	Complex Consonant & Diacritic Knowledge	Control	978	16.67%	0.00	45.00	0.00	41.00	21.14	14.92	0.48
		DH Coaching	662	16.62%	0.00	45.00	0.00	39.00	20.84	14.34	0.56
		External Coaching	649	15.56%	0.00	45.00	0.00	41.00	22.29	14.70	0.58
	Oral Reading Fluency 1	Control	978	20.04%	0.00	58.00	0.00	58.00	41.91	23.57	0.75
		DH Coaching	662	16.16%	0.00	58.00	0.00	58.00	43.33	22.31	0.87
		External Coaching	649	17.72%	0.00	58.00	0.00	58.00	43.11	22.69	0.89
	Oral Reading Fluency 1 Comprehension	Control	978	23.62%	0.00	7.00	0.00	7.00	3.68	2.55	0.08
		DH Coaching	662	19.94%	0.00	7.00	0.00	7.00	3.75	2.45	0.10
		External Coaching	649	21.73%	0.00	7.00	0.00	7.00	3.82	2.55	0.10
	Oral Reading Fluency 2	Control	978	17.69%	0.00	64.00	0.00	64.00	47.43	25.01	0.80
		DH Coaching	662	16.31%	0.00	64.00	0.00	64.00	48.43	24.16	0.94
		External Coaching	649	14.95%	0.00	64.00	0.00	64.00	49.00	23.79	0.93
	Oral Reading Fluency 2 Comprehension	Control	978	26.18%	0.00	6.00	0.00	6.00	3.21	2.35	0.08
		DH Coaching	662	24.17%	0.00	6.00	0.00	6.00	3.25	2.29	0.09
		External Coaching	649	24.19%	0.00	6.00	0.00	6.00	3.27	2.29	0.09
	Group Written Comprehension	Control	978	30.27%	0.00	7.00	0.00	6.00	2.92	2.47	0.08
		DH Coaching	662	29.46%	0.00	7.00	0.00	6.00	2.90	2.46	0.10
		External Coaching	649	29.12%	0.00	7.00	0.00	6.00	2.92	2.47	0.10
English First Additional Language	Word Reading	Control	978	24.64%	0.00	104.00	0.00	49.00	21.69	19.82	0.63
		DH Coaching	662	21.45%	0.00	100.00	0.00	47.00	21.55	19.11	0.74
		External Coaching	649	22.34%	0.00	82.00	0.00	51.00	23.20	19.93	0.78
	Oral Reading Fluency 1	Control	978	23.82%	0.00	71.00	0.00	70.00	41.43	28.76	0.92
		DH Coaching	662	21.00%	0.00	71.00	0.00	70.00	42.03	28.12	1.09
		External Coaching	649	20.34%	0.00	71.00	0.00	70.00	43.35	28.53	1.12
	Oral Reading Fluency 1 Comprehension	Control	978	27.81%	0.00	11.00	0.00	9.00	3.35	3.25	0.10
		DH Coaching	662	25.83%	0.00	11.00	0.00	8.00	3.11	3.01	0.12
		External Coaching	649	23.73%	0.00	11.00	0.00	9.00	3.61	3.23	0.13

Subtest Correlation Matrices

Correlation matrices indicated that the subtask RON showed a different distribution to the other subtasks (see Appendix: Distribution Analysis). Therefore it was excluded in the PCA construction. The correlation matrix is shown for HL and EFAL separately:

Table xx: Wave 2, Grade 2, Subtest Correlation Matrix

	HL letter sound knowledge	HL word recognition	HL oral reading fluency	HL comprehension oral reading fluency	HL complex consonant sound knowledge (Gr)	HL word recognition writing
HL letter sound knowledge	1					
HL word recognition	0,76	1				
HL oral reading fluency	0,73	0,93	1			
HL comprehension oral reading fluency	0,72	0,83	0,84	1		
HL complex consonant sound knowledge (Gr)	0,55	0,54	0,54	0,56	1	
HL word recognition writing	0,77	0,71	0,68	0,7	0,62	1

Table xx: Wave 1, Grade 2, Subtest Correlation Matrix

	HL letter sound knowledge	HL word recognition	HL oral reading fluency	HL comprehension oral reading fluency	HL complex consonant sound knowledge (Gr)	HL word recognition writing
HL letter sound knowledge	1					
HL word recognition	0,77	1				
HL oral reading fluency	0,73	0,94	1			
HL comprehension oral reading fluency	0,69	0,84	0,87	1		
HL complex consonant sound knowledge (Gr)	0,56	0,56	0,52	0,51	1	
HL word recognition writing	0,77	0,78	0,74	0,73	0,61	1

Table xx: Grade 3, Subtest Correlation Matrix

	HL Letter Sound Knowledge	HL Complex Consonant Diacritic Knowledge	HL Word Reading	HL Oral Reading Fluency ALT	HL ALT Comprehension	HL Oral Reading Fluency ALT2	HL Oral Reading Fluency Text	HL Writen Comprehension Grp	EFAL Oral Reading Fluency	EFAL ALT Comprehension	EFAL Word Reading
HL Letter Sound Knowledge	1										
HL Complex Consonant Diacritic Knowledge	0,7	1									
HL Word Reading	0,69	0,87	1								
HL Oral Reading Fluency ALT	0,64	0,83	0,92	1							
HL ALT Comprehension	0,59	0,77	0,81	0,79	1						
HL Oral Reading Fluency ALT2	0,65	0,84	0,93	0,95	0,77	1					
HL Oral Reading Fluency Text	0,58	0,78	0,82	0,79	0,84	0,79	1				
HL Writen Comprehension Grp	0,52	0,72	0,78	0,78	0,73	0,78	0,75	1			
EFAL Oral Reading Fluency	0,56	0,79	0,87	0,9	0,71	0,9	0,71	0,75	1		
EFAL ALT Comprehension	0,49	0,67	0,71	0,7	0,69	0,69	0,68	0,65	0,77	1	
EFAL Word Reading	0,58	0,79	0,86	0,86	0,73	0,86	0,73	0,73	0,92	0,77	1

Table xx: Grade 4, Subtest Correlation Matrix

	HL Word Reading	HL ORF Alt	HL Alt COMP	HL ORF Alt 2	HL ALT COMP 2	HL WC Grp	EFAL ORF	EFAL Comprehension	EFAL WC
HL Word Reading	1								
HL ORF alt	0,93	1							
HL Alt COMP	0,9	0,95	1						
HL ORF Alt 2	0,8	0,8	0,78	1					
HL ALT COMP 2	0,71	0,71	0,72	0,79	1				
HL WC Grp	0,78	0,78	0,76	0,76	0,67	1			
EFAL ORF	0,86	0,91	0,76	0,88	0,68	0,75	1		
EFAL Comprehension	0,69	0,71	0,75	0,69	0,72	0,65	0,8	1	
EFAL Writing Comprehension	0,66	0,68	0,66	0,65	0,58	0,68	0,73	0,71	1

Comparison of Intervention Streams

Analysis of learner assessment results was conducted in R and included the following processes[10]:

- While the comparability of the treatment groups had been confirmed at the midline, further balance tests (Table XX in Appendix XX) were conducted to ensure comparability at the endline. Based on these, learner gender and age were included in the intervention comparison regressions as controls.
- As the basis for the intervention comparison regressions, two aggregate performance scores were generated for each learner through principal component analysis (PCA)[11]: one for HL and one for EFAL. Table ## above shows the subtests included in each PCA. The first principal component explains over 70% of the variation (see Appendix XX).
- Before inclusion in the PCA, all sub-tests were tested for floor and ceiling effects. Only the Oral Reading Fluency (ORF) subtests were found to have ceiling effects at the three-minute cut-off-point but none at the one-minute point and so were included in the PCA at one minute, which accords with the correct words per minute (cwpm) ORF subtests used in EGRS I and II.
- All cohort averages include learners who have repeated a grade, therefore showing the full spectrum of learner abilities that started with that cohort in 2021.

Home Language and EFAL Composite Scores

Analyses of composite Home Language and EFAL scores show that external coaching had a positive and significant effect on Grade 2 HL and EFAL scores and on Grade 3 EFAL scores compared to the control group. The effect size for Grade 2 is substantial, indicating a notable improvement. Furthermore, the effect for EFAL was larger than for HL.

Table X: Home Language and English as a First Additional Language Regression Results[12]

Dependent variable: Each Language Composite Score, OLS Clustered Standard Errors						
	Grade 2, Wave 2 COHORT C (NEW) Grade 2 in 2023		Grade 3 COHORT A Grade 3 in 2023		Grade 4 COHORT B Grade 4 in 2023	
	HL	EFAL	HL	EFAL	HL	EFAL
External coaching	0.255***	0.323***	0.128	0.175**	0.068	0.064
	(0.088)	(0.090)	(0.081)	(0.081)	(0.077)	(0.082)
DH coaching	0.090	0.038	0.046	0.011	0.029	0.014
	(0.105)	(0.096)	(0.074)	(0.074)	(0.068)	(0.068)
Control mean	-0.09	-0.08	-0.05	-0.05	-0.02	-0.02
Observations	1,057	1,057	2,249	2,249	2,260	2,260
R^2	0.106	0.089	0.125	0.093	0.101	0.094
Adjusted R^2	0.094	0.076	0.119	0.087	0.095	0.088

Note: All regressions include individual, strata and district controls.
Statistical significance is indicated as follows *p<0.1 **p<0.05 ***p<0.01

---

[10] Further detail will be included in the full technical report

[11] Principal component analysis (PCA) was constructed using correlation matrices with standardised means of zero and standard deviations of 1. See Appendix: PCA construction

[12] In this table, numbers represent standard deviations (SD) and numbers in brackets are standard errors (SE). OLS Regressions are reported with clustered standard errors at the school level for the HL composite score results, which were run separately from the EFAL results. This table combines both outputs for readability. The average means for the control group in each cohort is included for reference. All the composite scores were normed to a standard deviation of 1 and a mean of 0 and vary between -2 and +2.

Additional analysis comparing the sample excluding repeaters was also conducted, as well as tests for any individual heterogeneity impacts on the intervention. Similarly, regressions for each subtask were also performed. However, these did not change the overall outcomes, but confirmed the regression results seen above. (see Appendix: Regressions excluding repeaters and for all subtasks and Heterogeneity regressions).
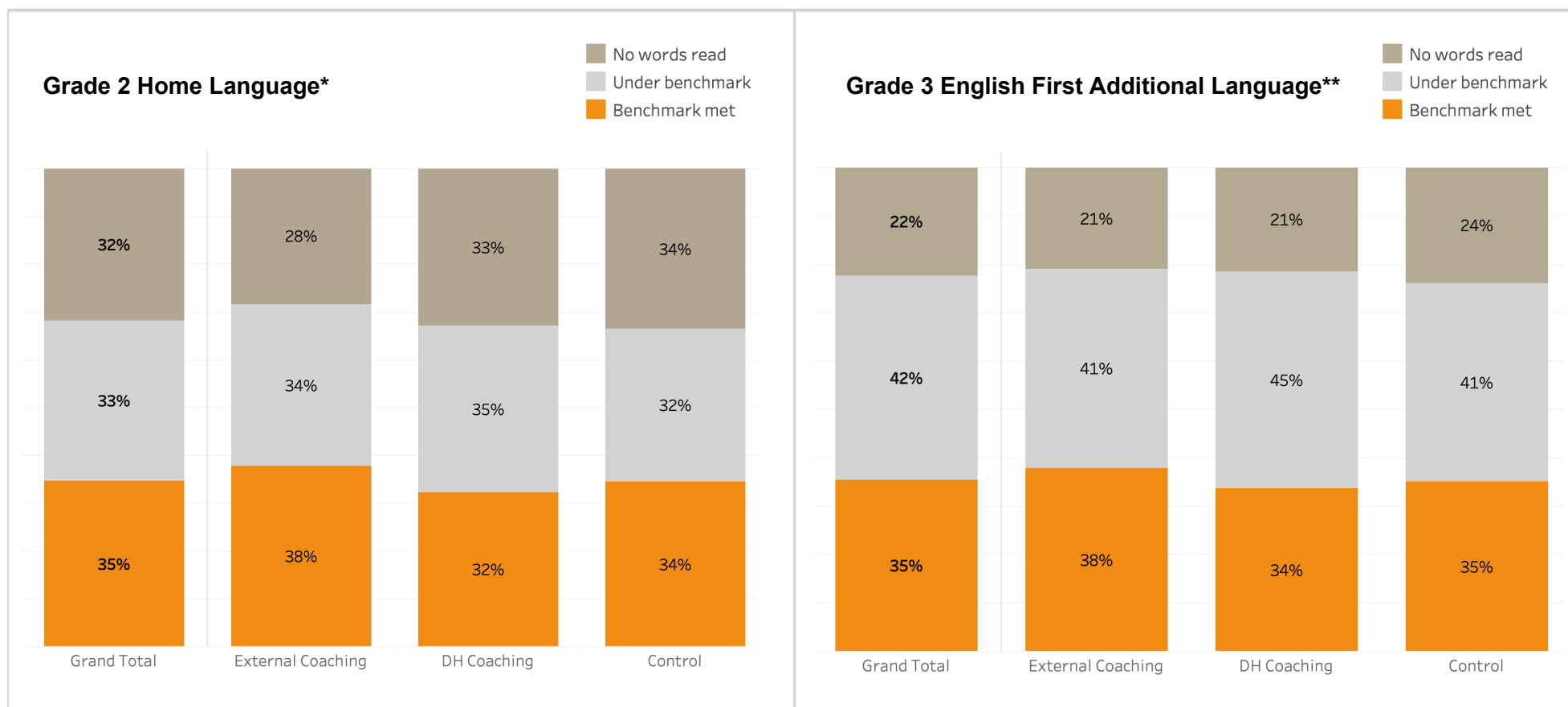
Based on an analysis of 130 randomised controlled trials (RCTs) aimed at improving educational outcomes in low- and middle-income countries, Evans and Yuan (2020) found that the median effect size for successful interventions was 0.10 standard deviations (SD), with 0.38 SDs being the 90th percentile. For large-scale studies involving over 5,000 learners, such as the Early Grade Reading Programme (EGRP), the average effect size tends to be smaller, around 0.05 SDs. In this context, the EGRP's effect sizes of 0.255 SDs for external coaching in Grade 2 HL and 0.323 SDs for EFAL are relatively high by international standards.

However, when measured against reading benchmarks (see Figure below)—which assess a learner's ability to read a specified number of correct words per minute (CWPM) on an oral reading fluency task but exclude other subtasks—these large effect sizes have not resulted in significant progress in overall reading ability. In Grade 2 HL, where the benchmark is 40 CWPM, the external coaching group saw 6% fewer learners who could not read a single word correctly compared to the control group, while 4% more learners reached the benchmark. In Grade 3 EFAL, with a benchmark of 50 CWPM, the trend was similar: 3% fewer learners in the external coaching group were unable to read a single word, and 3% more learners met the benchmark compared to the control group.

This suggests that while the programme is effectively building lower-order skills such as letter-sound recognition and word reading, it may not be adequately addressing higher-order skills necessary to overcome existing reading backlogs in the system.

An additional analysis of the Grade 2 assessments was conducted to compare Cohort B (learners exposed to the programme for only one year and severely affected by COVID-19 rotations) with Cohort C (learners exposed to the programme since Grade 1 without COVID-19-related disruptions).

This analysis revealed significant and positive findings for external coaching and the 2023 cohort (year dummy), while DH coaching was insignificant for both HL and EFAL scores. At the sub-task level, external coaching generally had a positive and significant effect, except on lower-order skills such as rapid object naming and letter-sound knowledge. DH coaching remained insignificant across all subtasks. The 2023 cohort, which had uninterrupted schooling and greater programme exposure, showed highly significant effects across all subtasks, highlighting the major negative impact that COVID-19 had on the 2021 cohort ==(see Table XX).==

Grade 2 Home Language*

	No words read	Under benchmark	Benchmark met

	Grand Total	External Coaching	DH Coaching	Control
No words read	32%	28%	33%	34%
Under benchmark	33%	34%	35%	32%
Benchmark met	35%	38%	32%	34%

Grade 3 English First Additional Language**

	Grand Total	External Coaching	DH Coaching	Control
No words read	22%	21%	21%	24%
Under benchmark	42%	41%	45%	41%
Benchmark met	35%	38%	34%	35%

Grade 2 HL Benchmark is 40 CWPM
**Grade 3 EFAL Benchmark is 50 CWPM*

Figure X: Benchmark achievement by treatment group for Grade 2 HL and Grade 3 EFAL

	Dependent variable: Grade 2 Main Regression	
	HL Composite Scores OLS, Clustered se Total Sample	EFAL (EFAL Group WSK)- Standardised OLS, Clustered se Total Sample
	HL	EFAL
External coaching	0.219*** (0.098)	0.245** (0.108)
DH coaching	0.042 (0.081)	0.071 (0.093)
2023	0.603*** (0.084)	0.506*** (0.094)
External coaching	0.039 (0.118)	0.034 (0.128)
DH coaching	0.037 (0.133)	-0.064 (0.124)
Observations	3,910	3,910
R^2	0.161	0.120
Adjusted R^2	0.157	0.116

Note: All regressions have individual, district and stratum controls
*Statistical significance is indicated as follows *p<0.1 **p<0.05 ***p<0.01*

	Dependent variable: Grade 2 Cohort Analysis							
	Rapid Object Naming (HL)	Letter Sound (HL)	Word Recognition (HL)	Oral Reading Fluency ALT (HL)	ORF Comprehension (HL)	HL G Complex Consonant Sound Knowledge	HL G Word Recognition Writing	EFAL Word Sound Knowledge
	Out of 36	Out of 70	Out of 60	Out of 49	Out of 5	Out of 5	Out of 5	Out of 10
External Coaching	0.230 (0.299)	3.511 (2.226)	2.795** (1.151)	3.888* (2.015)	0.347** (0.156)	0.436*** (0.167)	0.378* (0.201)	0.688** (0.293)
DH Coaching	-0.088 (0.284)	0.803 (1.889)	-0.258 (0.904)	-0.370 (1.653)	0.019 (0.134)	0.324** (0.163)	0.104 (0.164)	0.192 (0.252)
2023	0.497 (0.326)	15.494*** (1.854)	7.590*** (1.099)	12.614*** (2.028)	0.795*** (0.157)	0.602*** (0.145)	1.164*** (0.148)	1.369*** (0.254)
External Coaching	0.371 (0.463)	-0.372 (2.662)	0.523 (1.535)	2.517 (2.720)	0.056 (0.222)	0.008 (0.205)	0.065 (0.244)	0.092 (0.347)
DHCoaching	-0.293 (0.519)	-0.081 (3.078)	1.626 (1.715)	3.071 (3.131)	0.070 (0.232)	-0.234 (0.240)	0.022 (0.237)	-0.174 (0.335)
Midline Mean	15.83	29.62	9.17	13.37	1.07	3	1.93	5.45
Observations	3,910	3,910	3,910	3,910	3,910	3,910	3,910	3,910
R	0.029	0.164	0.144	0.140	0.109	0.089	0.135	0.120
Adjusted R	0.024	0.160	0.140	0.136	0.105	0.084	0.131	0.116

Note: All regressions have individual, district and stratum controls
*Statistical significance is indicated as follows *p<0.1 **p<0.05 ***p<0.01*

The regressions also show that the effect of external coaching was positive for Grade 3 HL and both Grade 4 languages, but it was not significantly different from control schools. The Grade 4 effect size for the external coaching stream is also smaller than for Grades 2 and 3, which may reflect a tapering-off effect given that the intervention did not include Grade 4. A separate analysis for Grade 3 HL results excluding repeaters[13] shows a positive and significant effect for the external coaching stream, but this overestimates the intervention impact by excluding weaker learners.

DH coaching outcomes were not found to be significantly different from control group outcomes at any Grade level or in any language, and the effect sizes, while positive, were small across the board (from 0.011 in Grade 3 EFAL to 0.09 in Grade 2 HL).

Regressions were conducted for each HL sub-test that makes up the composite score and the results supported the composite score results for each cohort. All Grade 2 HL subtests were significant for external coaching and none for DH coaching. For Grade 3 HL, CCDK, ORF 1, ORF 1 comprehension and ORF 2 were significant but only at the $p < 0.1$ level, with no effect for DH coaching. For Grade 4 HL, neither external nor DH coaching showed any significant effects for any of the subtests.

Distribution of Impact by Learner Performance

Quantile regressions were conducted for Grades 2, 3, and 4 at the endline to determine if the programme had differential effects on learners with varying performance levels. The differences across the performance distribution were negligible for DH coaching across all the grades and languages. This was also the case for external coaching in Grade 3 EFAL and Grade 4s in both languages.

Among Grade 2 learners in HL, the effect of external coaching was positive (above zero standard deviations) across the entire distribution of learners, but it was especially positive for lower-mid-range learners. As shown in Figure XX, learners between the 20th and 60th quantile benefited more than the average for Grade 2 learners (represented by the solid red line in the graph). Compared to the average 10.2 months of learning gains compared to controls (0.255 standard deviations), this group of lower-mid-range performing learners gained up to 18 months of HL learning (0.45 standard deviations).

---

[13] Repeaters are learners who were in Grade 1 in 2021 and were found in Grade 2 in 2023 but were assessed using the Grade 3 (cohort A) assessment tool in 2023

Figure X: External Coaching Grade 2 HL quantile plot[14]
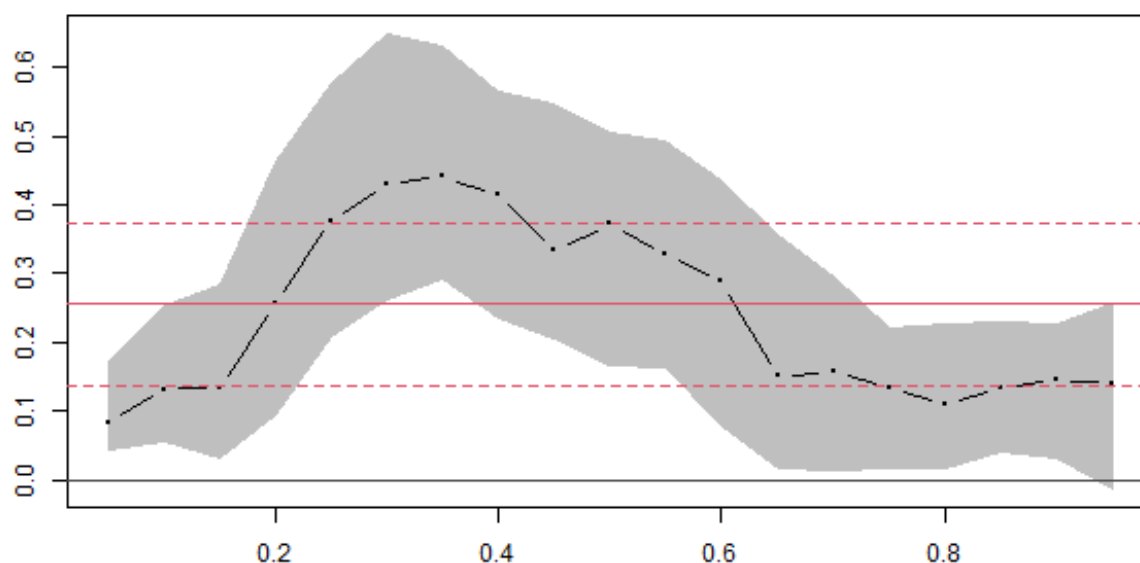
Interpretation of Impact Results

Two key questions emerge from the impact results:

- Why was DH coaching not found to be effective, compared to the control group?
- What explains the positive effects of external coaching on Grade 2 but not on Grades 3 (HL) and 4 (both languages), as expected based on previous studies?

In interpreting the comparative results for external and DH coaching, two important factors must be taken into account: the nature of the 'control' schools with which the coaching schools are compared, and the post-COVID-19 context.

The EGRP 'control' schools received an extensive base programme (LTSM and teacher training). Our qualitative results confirmed that the lesson plans and training led to improved classroom practice, which is likely to translate into improvements in learner outcomes compared to 'business as usual' schools without such inputs, as shown in past studies. The learners in both coaching streams were therefore compared with learners who had probably already achieved some level of improved reading outcomes, although, since no true baseline data was collected and no 'business as usual' schools were included, the evaluation design did not allow the scale of improvement to be measured. The results therefore do not mean that DH schools did not see improved learner reading results over the course of the intervention, but simply that these improvements were not greater than those achieved by the schools exposed only to the base intervention.

We interpret the lack of additional DH coaching effects, compared to controls, based on the following findings on the implementation of the DH coaching intervention:

---

[14] The x-axis represents the distribution of learner performance (using the composite HL PCA score) and the y-axis is standard deviations from the control group. The red solid line represents the overall average SD while the broken red lines represent the upper and lower bounds of the 95% confidence interval around the average. The black dotted line represents the standard deviations from the control group for each segment of the Grade 2 performance distribution. The black dotted line above the red dotted line shows the learners who gained significantly more than the average for the Grade.

- **Learners were exposed to very limited dosage of DH coaching and for a short period of time**: after exposure to external coaching in year 1 (along with the external coaching stream), coaching exposure reduced in year 2 (since external coaches were still largely doing the coaching but at a lower coach to school ratio) and DH coaching only commenced at the beginning of year 3. Previous studies have posited that at least two years of exposure to a coaching intervention is required to observe effects (Fixsen, 2005).
- **DH coaching quality was highly variable**: case studies and coach shadowing showed that DH coaching quality depended on the level of commitment by individual DHs. Furthermore, DHs training was not consistently timed, with most training sessions taking place late in 2023. While variation in motivation and skill will always be part of the schooling system, and so is built into any similar intervention at scale, it may have reduced the average efficacy of the approach for this study, especially when coupled with the short implementation period.
- **DH coaching was not 'instructional'** in that it was not intensive, individualised and developmental but rather followed the lead of the external coaches (as guided by the implementing agency's overall approach to coaching) in using a standardised, coach-led (not teacher self-reflection led), compliance-driven approach.

This interpretation implies that DH coaching might be effective if implemented
- for a longer period of time,
- with processes in place to increase consistent DH acceptance and adoption of the coaching role (see recommendations below), and
- with an adaptation to the coaching approach to include greater emphasis on the instructional (individualised, developmental) aspects over compliance with the use of specific pedagogies and standardised classroom methodologies.

When considering reasons for the observed positive effect of external coaching for Grade 2 (cohort C) but less for Grades 3 (cohort A) and 4 (cohort B), we must consider the combination of the post-COVID-19 context with the differential intervention exposure of each of the intervention cohorts (as described in the section on Evaluation Design above). To recap, cohort A experienced three years of the intervention (Grade 1 in 2021, Grade 2 in 2022 and Grade 3 in 2023) and were affected significantly by COVID-19 restrictions on implementation and general learning in 2021, as well as being disadvantaged by a lack of Grade R learning in 2020. Cohort B experienced two years of the intervention (Grade 2 in 2021, Grade 3 in 2022) while also being affected by cumulative backlogs from a Covid-era Grade 1 (2020) and Grade 2 (2021). By the end of 2023 when they were reassessed, they had spent a year in Grade 4 with teachers who had not been included in the intervention, resulting in the possibility of an additional 'fade out' effect on learning outcomes. Cohort C, the 'new wave' of Grade 2 learners assessed in 2023, experienced two years of intervention (Grade 1 in 2022 and Grade 2 in 2023) without COVID-19 effects or fade-out effects.

Given this context, cohorts A and B both had strong mitigating factors against measurable overall learning improvements: for cohort A, the cumulative COVID-19 effect of 2020 learning backlogs plus limited learning in 2021, and for cohort B, the fade-out effect of 2023. In contrast, cohort C experienced entirely post-COVID-19 primary schooling and benefited from the more effective years of intervention implementation (2022 and 2023). These cohort effects apply across all treatment groups, and also made it more challenging to effectively implement the coaching intervention which might have reversed these COVID-19 learning

losses. It is our view that the external coaching was not implemented with enough dosage or quality, and did not include specific reinforcing methodologies like regular school-based workshops, and therefore could not pull the external coaching schools in cohorts A and B out of the cumulative Covid backlogs hole.

Furthermore, there were implementation challenges with the external coaching stream that mirror some of the challenges experienced with DH coaching, including insufficient dosage and uneven distribution of coaching dosage across teachers, and a lack of instructional coaching practice (or the presence of compliance-driven coaching practice). These may have contributed to lower than expected efficacy of the external coaching intervention in Grade 3 (HL) and Grade 4.

A further reason for stronger Grade 2 effects of external coaching may lie in the way the pedagogies core to the Structured Learning Programme were implemented. The CAPS approach to reading is based on cumulative application of lower order skills (such as phonics and phonemic awareness, letter sound recognition and word reading) and higher order skills (such as oral reading fluency and comprehension). The Grade 2 curriculum focuses more strongly on the lower order skills (letter sounds and word reading), while Grades 3 and 4 require the higher order skills to be in place. This skills progression in the curriculum is also reflected in the subtasks of each grade's EGRA assessment tools. While classroom observations and teacher surveys consistently found that teacher skills in the lower order skills had been improved (probably through a combination of the lesson plans and teacher training but possibly reinforced through external coaching), the pedagogies intended to build higher order skills were consistently weaker. Shared reading and especially group guided reading were generally weakly implemented in the classrooms, and neither of the coaching modalities, although intended to improve these methods, seemed to be effective in doing so. If the intervention was therefore more effective in improving lower order reading skills than in improving higher order skills, it is logical that a greater effect would be seen at Grade 2 than in the higher grades.

In summary, and taking the quantile regression results into account, the EGRP external coaching intervention as implemented substantially benefited younger (Grade 2) learners who had not experienced substantial COVID-19 learning losses. Among these, lower-to-mid-range performing learners benefited even more. This is likely to have a lasting effect on those Grade 2 learners as they pass through the rest of the education system with improved reading fundamentals in both HL and EFAL. The larger improvement among lower-to-mid-range performers is progressive and contributes to addressing the challenge of early learning backlogs. However, whether due to contextual factors such as COVID or implementation factors such as coaching approach, the EGRP could not confirm the overall (across Grade 3 languages) and lasting (to Grade 4) efficacy of external coaching. It did not quantitatively show any top-up benefits of DH coaching in comparison with a basic Structured Learning Programme intervention.

The viability of the DH model for application at scale

Despite the lack of measurable impact, the evaluation did find qualitative reasons to believe that DH coaching may, in principle, be a viable option for internalising coaching within the public education system. This is because some of the barriers to viability that were originally

predicted were not found to be as strong as expected. These expected barriers related to time and the DH-teacher relationship.

- **Time Constraints**: as noted in the section on enabling & constraining factors for DH coaching, the evaluation confirmed that DHs have full teaching and administrative loads, leaving little time for coaching tasks and specifically for observing teachers in the classroom. However, the study also found that DHs with PYEI Education Assistants (EAs) used them to supervise classes while they conducted coaching. Incorporating EAs into a DH coaching intervention design may therefore effectively address this constraint. This suggests that instead of a 'triple cocktail' of LTSM, training and coaching, a 'quadruple cocktail' of LTSM, training, coaching and Education Assistants may be required.

  While the system-wide introduction and maintenance of EAs introduces an additional cost factor to the scaled-up response, this may be lower than (or cumulatively equivalent to) the cost of professional external coaches, but easier to staff (given that EAs require much lower previous experience and qualification levels) and having higher positive 'externalities' beyond the education system (e.g. large-scale youth employment and work experience), making it easier to fund within a combination of government and (potentially) public-private partnership systems.

- **DH-teacher relationship**: an initial concern for the design of the DH coaching model was whether DHs would find it structurally difficult to fulfil both quality control/evaluative roles (as set out in their primary DH duties to monitor curriculum coverage and similar functions) and the more developmental coaching role. While this evaluation found that DHs did indeed largely apply a compliance approach to coaching, this was not because they were 'stuck in their ways' or faced pressure and incentives from other parts of the education system (i.e. SMT or district demands), but rather because this was the model taught to them by the external coaches and the implementing agency. In fact, some highly motivated DHs took on the coaching role with enthusiasm and skill, including being flexible and innovative in supporting their teachers. This suggests that, in principle, were a more developmental coaching model to be presented to DHs, with strong initial external modelling and support, it may well be adopted by a sufficiently large proportion of DHs to be effective. The coaching approach therefore becomes a question of implementation quality rather than a question of design barriers.

Overall, the mixed methods evaluation findings suggest that both external and DH coaching models seem to have equivalent pathways and classroom practice outcomes. Both models depend to a large extent on similar enabling factors: supportive school managers who prioritise foundation phase literacy teaching, committed teachers who accept the teaching strategies and utilise the materials provided, and individual commitment by coaches (whether DHs or external) to their roles.

In addition to the EA discussed above, a key structural consideration that is specific to DH coaching relates to the appropriate and continuous selection of DHs for the coaching role. This consideration affects both DH recruitment/promotion and succession planning.

- **DH Recruitment/Promotion:** in contrast to external coaches, who are specifically hired based on their coaching skills, DHs are currently promoted into their positions for a range of factors, mostly not related to coaching ability. However, should coaching become a core element of the DH job description, the intentional consideration of coaching-related skills and personality traits would need to be included in the recruitment and promotion process. It is understood that changing the 'rules' of both succession and promotion may be structurally challenging, given how

schools, education districts and teacher unions currently manage DH succession and promotion, but discussions should be held about options for integrating coaching skills and role descriptions into these processes.

- **Succession planning:** similarly, once DHs are trained as coaches, the pro-active management of DH succession becomes an important element of system-wide adoption, with clear ways of engaging near-retirement DHs to either commit to the coaching role or pass it on to other colleagues as well as planning in advance for how the coaching role is passed on when an experienced DH coach retires.

Finally, while Subject Advisors and District Officials were included in the EGRP design through consultation on the lesson plans and some inclusion in training sessions, their overall roles in the DH coaching processes are not clear.

We therefore conclude that DH coaching may be a viable means of integrating coaching into the public schooling system at scale where the following are in place:
1. Recognition of the coaching role of DHs by the provincial department, with concomitant adaptations to promotion/selection/succession policies & processes;
2. Careful selection of the DH coach which may, in the transition period from the current DH cohort to a future 'coaching-enabled' DH cohort, mean bypassing the existing DH and appointing a new DH who is committed to the role;
3. Education Assistants with sufficient training in the same structured learning programme utilised by the DH and teachers to support DHs and enable them to have time to observe and support teachers;
4. A curriculum coverage and lesson plan adherence data tool which tracks teacher use of the structured learning programme, provides DHs (and external support personnel) with real-time, easy-to-interpret evidence, and therefore enables targeted prioritisation of coaching time to teachers with the most severe backlogs;
5. Sufficient regular DH coach training, structured as continuous professional training and development;
6. Effective external coach support for the DH coach, with higher initial support dosage that tapers off to a lower dosage once the DH has shown evidence of reaching a level of coaching competency;
7. A longer period of support for the DH coach, including DH training and external coaching support.

Any future assessment of the effectiveness of DH coaching on learner outcomes would need to have the following characteristics to allow for reliable impact measurement:
1. Separate the process for testing 'system internal' DH coaching implementation viability (i.e. can the conditions under which DHs work in the public schooling system be adapted to include effective coaching responsibilities) from testing 'system external' support requirements (i.e. the costs and levels of implementation capacity required from external actors to provide the coaching training and coaching support). Further, the external support capacities for coaching support should be institutionally distinct from the external capacities for LTSM and teacher training.
2. Include a real baseline assessment of both learner outcomes and teaching practices so that change over time can be measured at a school-by-school level;
3. Set out clearly defined measures based on which effective DH coaching practice will be measured so that the DH cohort can be categorised into high and low fidelity and quality and relative learner impacts can be judged accordingly;
4. Ensure a consistent support package to be in place (i.e. Education Assistants, external training and external coaching support for the DHs);

5. Allow at least two full years of DH coaching to be in place before the endline assessment.

Conclusions

The EGRP was designed to test whether DH coaching could achieve similar improvements in learner outcomes as external coaching had achieved in previous studies. This could not be shown at an impact level, but rather than being a design fault this may be due to contextual factors such as Covid-19, and implementation constraints such as placing external implementation responsibility on a single agency, and the amount of time spent on the coaching model.

By tracking implementation fidelity and outcomes through a variety of methodologies, the study suggests that DH coaching may be a viable model for integrating coaching into the public education system at scale, if specific design factors are considered.

The study also shows that external coaching, as implemented in the current intervention, was confirmed to be highly effective in conducive learning contexts such as 'normal' post-Covid learning and teaching. However, it could not reverse the adverse effects of Covid-19 learning backlogs for learner cohorts who had been exposed to two years of Covid-era learning losses.

Recommendations

Recommendations on the base intervention (LTSM & teacher training):

- Continue development of the EGRP app to enable DHs and external coaches to continuously monitor teacher use of lesson plans through dashboards and automated feedback. This can strengthen the linkage between coaching and curriculum coverage and enable the targeting of teachers who are falling behind.

- For teacher training, introduce post-training assessments to gauge teachers' understanding of the material, particularly in relation to challenging methodologies such as shared reading and GGR. Training can also focus more on differentiated instruction and classroom management techniques to equip teachers to handle diverse learning needs.

Recommendations on Coaching (all models):

- When planning an intervention at scale, intervention protocols need a certain level of standardisation. However, for a coaching intervention, this standard process should focus on following steps that allow for adaptation to each school's and teacher's needs, which may include variation in dosage as well as content, so that weaker teachers receive more support than teachers who are already strong.

- Train coaches to engage with teachers about learners' understanding of the lesson content and whether the lesson is effectively building on earlier lessons and skills learned in those lessons, in addition to offering practice and feedback on the application of specific literacy strategies in the classroom.

- Encourage a more tailored and developmental approach rather than a one-size-fits-all coaching model.

- Clarify the design of school-based workshops so that they fulfil the role of peer learning and do not repeat content and interactions covered in other parts of the programme. They should emphasise the importance of interactive sessions and teacher engagement.

Recommendations on School- and System-level Enabling Factors for DH Coaching:

- Consider the processes through which DHs are informed about/recruited into a new coaching role to manage possible rejection of the role based on feelings of not having been consulted or not receiving the required support. This may be especially important for older DHs or those close to retirement

- Consider adaptations to standard DH promotion and succession processes to take coaching potential and skills into account

- Integrate Education Assistants into the design of DH coaching models

Recommendations on Large-scale Intervention Implementation Design:

- The EGRP implementation design tested two questions simultaneously: firstly, whether DHs could be trained and supported to coach teachers, and secondly, whether this could be done at relatively low cost with limited external support (i.e. two external DH-support coaches). The current evaluation suggests that limitations in the second factor (implementation of the intervention with limited resources) have made it impossible to ascertain the answer to the first question. While application at scale fundamentally includes questions of cost and implementation capacity, in addition to questions of system adoption (e.g. by DHs), these questions can be addressed through different learning processes.

- The 'triple cocktail' (or, as recommended, the 'quadruple cocktail') requires a wide range of distinct implementation skills, from materials production to app design and data management, and from training to coaching. These skills are rarely found to be equally strong within one implementing agency. In any case, a stable system-wide implementation capacity would mean that impact within any part of the system does not depend on a few individuals or one agency. We therefore recommend that future studies spread implementation capacity across a consortium of implementing agencies, as was the case in previous EGRS, rather than tasking a single agency with the full spectrum of implementation tasks.

References

Alsofrom, K. (n.d.). The Second Early Grade Reading Study: How and Why does Coaching Work to Improve Teaching Practices in the EGRS II? An examination of causal mechanisms. Unpublished paper prepared for DBE.

Annenberg Institute (2004), http://annenberginstitute.org/pdf/InstructionalCoaching.pdf

Barber, M. (1996) The Learning Game. London: Victor Gollancz.

Baskwill, J. & Whitman, P. (1999). Every Child Can Read. Scholastic: Canada.

Biscard, J. et al. (2020) Reading Support Project: Evaluation Report. Pretoria: USAID.

Blazar, D. & Kraft, M.A. (2015) Exploring Mechanisms of Effective Teacher Coaching: A Tale of Two Cohorts from a Randomised Experiment https://doi.org/10.3102/0162373715579487

Bradley, H., Conner, C. and Southworth, G. (eds) (1994) Developing Teachers Developing Schools. London: David Fulton

Cilliers, Jacobus & Fleisch, Brahm & Kotze, Janeli & Mohohlwane, Nompumelelo & Taylor, Stephen & Thulare, Tsegofatso. (2022). Can virtual replace in-person coaching? Experimental evidence on teacher professional development and student learning. Journal of Development Economics. 155. 102815. 10.1016/j.jdeveco.2021.102815. https://www.researchgate.net/publication/357939473_Can_virtual_replace_in-person_coaching_Experimental_evidence_on_teacher_professional_development_and_student_learning

City, A., Elmore, R., Fiarman, S.E., and Teitel, L. (2009). Instructional Rounds in Education: A Network Approach to Improving Teaching and Learning. Harvard Education Press: Cambridge.
Cochran-Smith et al. (2015) Critiquing Teacher Preparation Research: An Overview of the Field, in Journal of Teacher Education vol 22 no. 2.

Dadds, M. Bridging the Gap: Using the School-based Project to Link Award-Bearing Inset to School Development, in Bradley, H., Conner, C. and Southworth, G. (eds) Developing Teachers Developing Schools. London: David Fulton.

Day, C. (2013) Teacher Quality in the 21st Century: New Lives, Old Truths, in Zhu, X. & Zeichner, K. (eds) Preparing Teachers for the 21st Century. Heidelberg: Springer.

Department of Basic Education. (2018). Second Early Grade Reading Study: Year 1 Report.

Department of Basic Education. (2019). Available at https://www.education.gov.za/Programmes/EarlyGradeReadingStudy.aspx

EGRS II (2019). Virtual Coaching Model Summary June 2019. Available at https://www.education.gov.za/Portals/0/Documents/Publications/EGRS/EGRS%20II%20Website%20Upload/Reports/2019_Virtual%20Coaching%20Model%20Summary%20REVISED.pdf

Evans, D. and Yuan, F., 2020. "How Big Are Effect Sizes in International Education Studies?" CGD Working Paper 545. Washington, DC: Center for Global Development. https://www.cgdev.org/publication/how-big-are-effect-sizes-internationaleducation-studies

Fixsen, D. L., Naoom, S. F., Blase, K. A., & Friedman, R. M. (2005). Implementation research: a synthesis of the literature.

Fleisch, B. (2022) The Education Triple Cocktail: System-wide Instructional Reform in South Africa. Cape Town: UCT Press.

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

Fleisch, B., Schöer, V., Roberts, G., & Thornton, A. (2016). System-wide improvement of early-grade mathematics: New evidence from the Gauteng Primary Language and Mathematics Strategy. International Journal of Educational Development, 49, 157–174. https://doi.org/10.1016/j.ijedudev.2016.02.006

Freire, P. (1972) Pedagogy of the Oppressed. London: Penguin

Fullan, M. & Hargreaves, A. (eds) (1992) Teacher Development and Educational Change. London: Falmer Press.

Gabriel, J.G. (2005) How to Thrive as a Teacher Leader. Alexandria VA: ASCD.

Germuth, A.A. (2018) Professional Development that Changes Teaching and Improves Learning, in Journal of Interdisciplinary Teacher Leadership vol 2, no. 1 pp77-90.

Global School Leaders (2024) Evidence Review 2024 - A Review of Research on the Role of School Leaders in Improving Student Learning Outcomes. https://www.globalschoolleaders.org/_files/ugd/c5f2d7_04c556b2aa9d4a4ea1e089c4551ad88f.pdf

Greenhalgh, T., Robert, G., Macfarlane, F., Bate, P., & Kyriakidou, O. (2004). Diffusion of innovations in service organisations: systematic review and recommendations. The Milbank Quarterly, 82(4), 581-629.

Hargreaves, A. (1998). The Emotions of Teaching and Educational Change in Hargreaves, A., Lieberman, A., Fullan, M., and Hopkins, D. (eds) International Handbook of Educational Change. Kluwer International Handbooks of Education vol 5 pp 558 – 575. Springer: Dordrecht

Howie, S.J., Combrinck, C., Roux, K., Tshele, M., Mokoena, G.M., & McLeod Palane, N. (2017). Progress in International Reading Literacy Study 2016: South African Children's Reading Literacy Achievement. Pretoria: Centre for Evaluation and Assessment, University of Pretoria.

Johnson Sirleaf, E. (2011). Foreword. In A. Gove & A. Wetterberg (Eds.), The Early Grade Reading Assessment: Applications and Interventions to Improve Basic Literacy (1st ed.). RTI Press. https://doi.org/10.3768/rtipress.2011.bk.0007.1109, USA.

Knight, J. (2011) What Good Coaches Do. In Education Leadership vol. 69 no. 2 pp 18 – 22.

Kotze, J., Fleisch, B., and Taylor, S. (2019) Alternative forms of early grade instructional coaching: Emerging evidence from field experiments in South Africa in International Journal of Educational Development, vol 66, April 2019, pp 203-213.

Kraft, M., Blazar, D., & Hogan, D. (2016). The effect of teacher coaching on instruction and achievement: A meta-analysis of the causal evidence. Review of Education Research.

Levin, H. and Lockheed, M. (eds) (1993) Effective Schools in Developing Countries. London: Falmer Press.

Lyon ###

Majerowicz, S., & Montero, R. (2018). Teaching be taught? Experimental evidence from a teacher coaching program in Peru. [Job market paper]. Retrieved https://scholar.harvard.edu/smajerowicz/publications/job-market-paper-can-teaching-be-taught-experimental-evidence-teacher

Mraz, M., Salas, S., Mercado, L., & Dikotla, M. (2016). Teaching Better, Together: Literacy Coaching as Collaborative Professional Development. In English Teaching Forum (Vol. 54, No. 4, pp. 24-31). US Department of State. Bureau of Educational and Cultural Affairs, Office of English Language Programs, SA-5, 2200 C Street NW 4th Floor, Washington, DC 20037.

Nadupalli, S. (n.d.). Ten-year study shows South African school reading literacy is slow to improve. Research Matters, University of Pretoria. Retrieved 27 October 2021, from https://www.up.ac.za/research-matters/article/2801832/ten-year-study-shows-south-african-school-reading-literacy-is-slow-to-improve

Neuman, S. B., & Cunningham, L. (2009). The impact of professional development and coaching on early language and literacy instructional practices. American educational research journal, 46(2), 532-566.

Patzer, R. (2020) https://blog.irisconnect.com/uk/coaching-for-teachers. Posted 29th January 2020.

Piper, B., & Zuilkowski, S. S. (2015). Teacher coaching in Kenya: Examining instructional support in public and non formal schools. Teaching and Teacher Education, 47, 173-183.

Sailors, M., Hoffman, J. V., David Pearson, P., McClung, N., Shin, J., Phiri, L. M., & Saka, T. (2014). Supporting change in literacy instruction in Malawi. Reading Research Quarterly, 49(2), 209-231.

Shanklin, N. L. (2006). What Are the Characteristics of Effective Literacy Coaching? Denver, CO: Literacy Coaching Clearinghouse.

Taylor, S., Cilliers, J., Prinsloo, J., Fleisch, B., & Reddy, V. (2019). Improving early grade reading in South Africa, 3ie Grantee Final Report. New Delhi: International Initiative for Impact Evaluation (3ie).

UNESCO (1953). The use of 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: Paris

Van Deventer, I. & Kruger, A.G. (2003) An Educator's Guide to School Management Skills. Pretoria: Van Schaik

Walpole, S., McKenna, M. C., Uribe-Zarain, X., & Lamitina, D. (2010). The relationships between coaching and instruction in the primary grades: Evidence from high-poverty schools. The Elementary School Journal, 111(1), 115-140.

Wayne, A. J., Yoon, K. S., Zhu, P., Cronen, S., & Garet, M. S. (2008). Experimenting with teacher professional development: Motives and methods. Educational Researcher, 37(8), 469-479.

White, J. and Barber, M. (1997) Perspectives on School Effectiveness and School Improvement. London: Institute of Education.

Williamson, R. (2012) Coaching Teachers: An Important Principal Role. Education Partnerships Inc.

Yeatman et al. (2012) https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0049790. Accessed March 2021.

Yoshikawa, H., Leyva, D., Snow, C. E., Treviño, E., Barata, M., Weiland, C., ... & Arbour, M. C. (2015). Experimental impacts of a teacher professional development program in Chile on preschool classroom quality and child outcomes. Developmental psychology, 51(3), 309.

Appendices

Distribution Analysis of Sub-Tasks

Figure XX: Grade 2 (Wave 1 vs Wave 2) ORF ALT
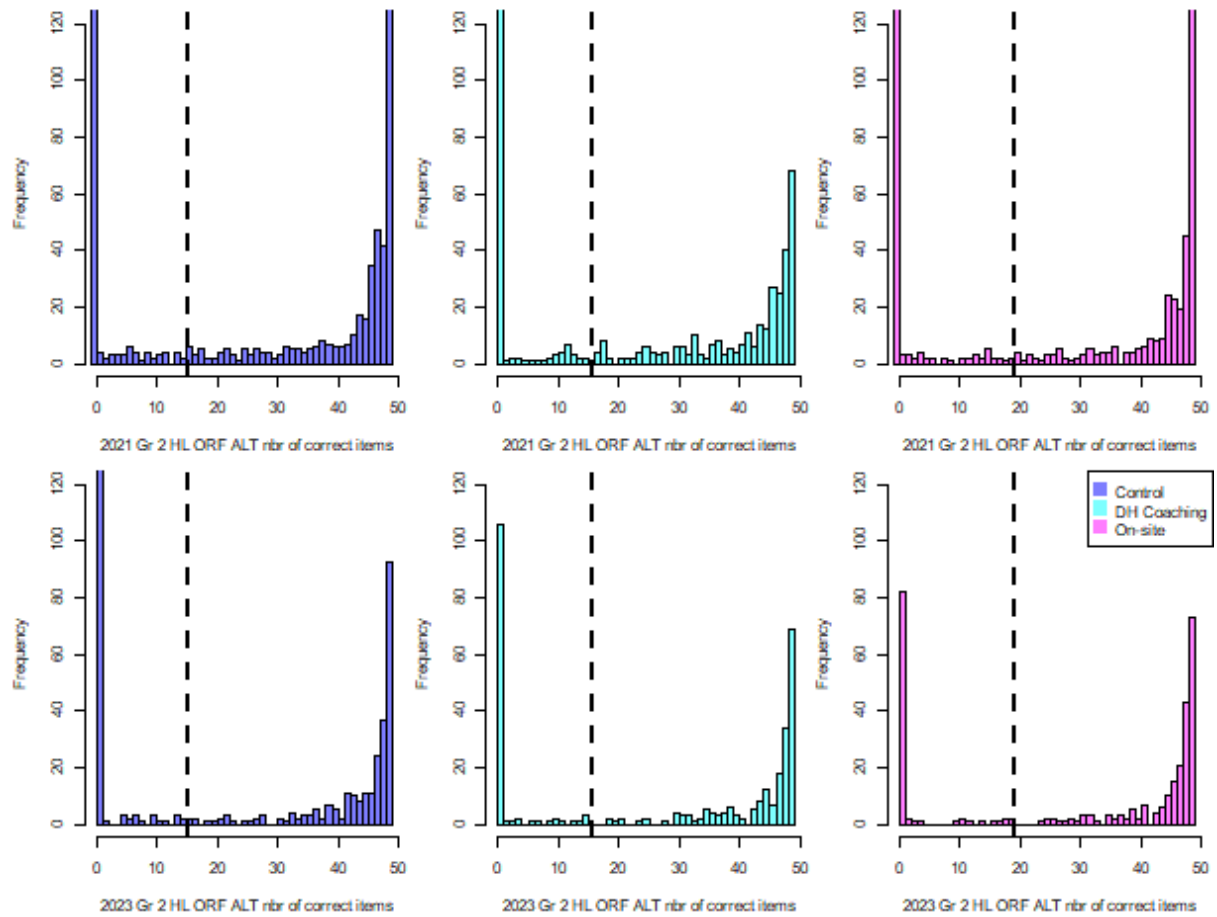
Figure XX: Grade 2 (Wave 1 vs Wave 2) ORF ALT @ 1 minute
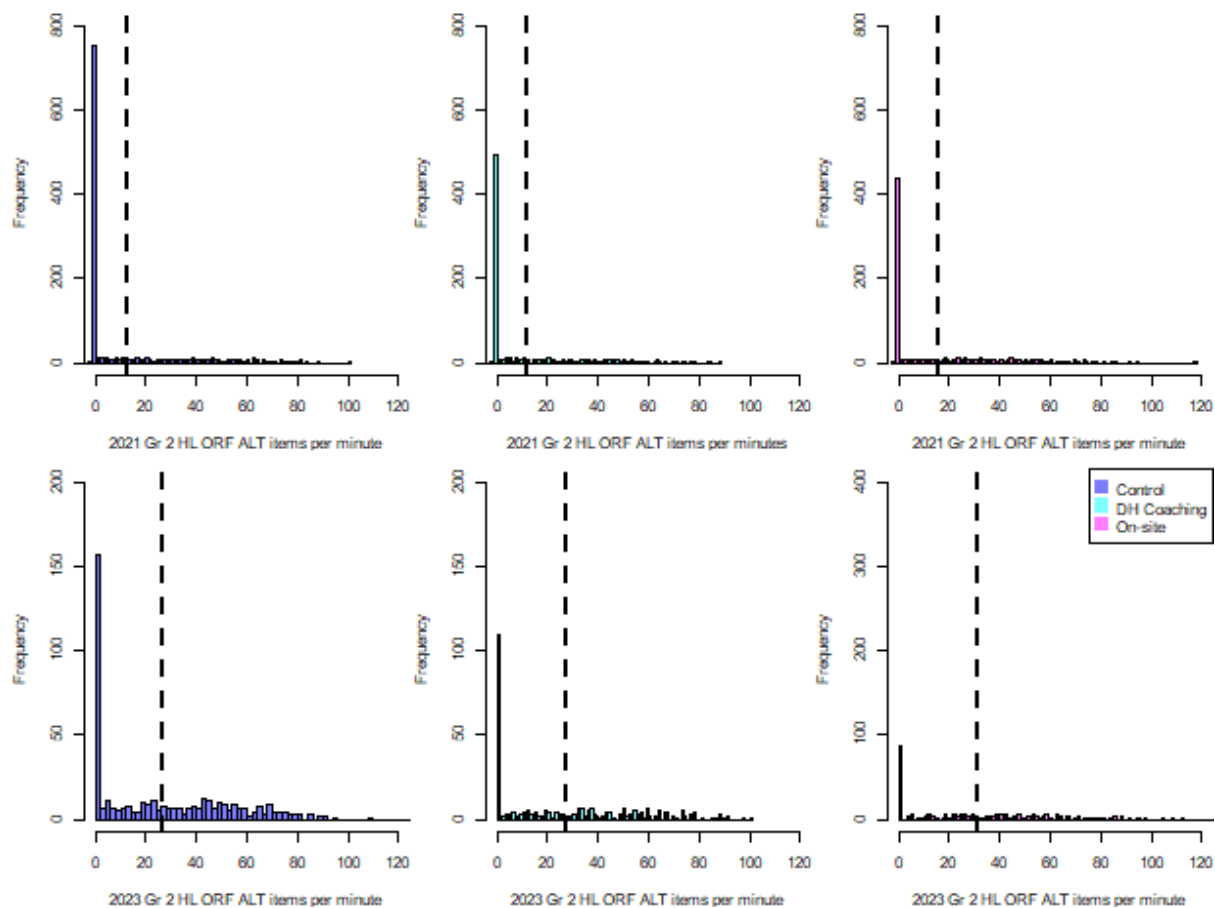
Figure XX: Grade 3 HL ORF ALT
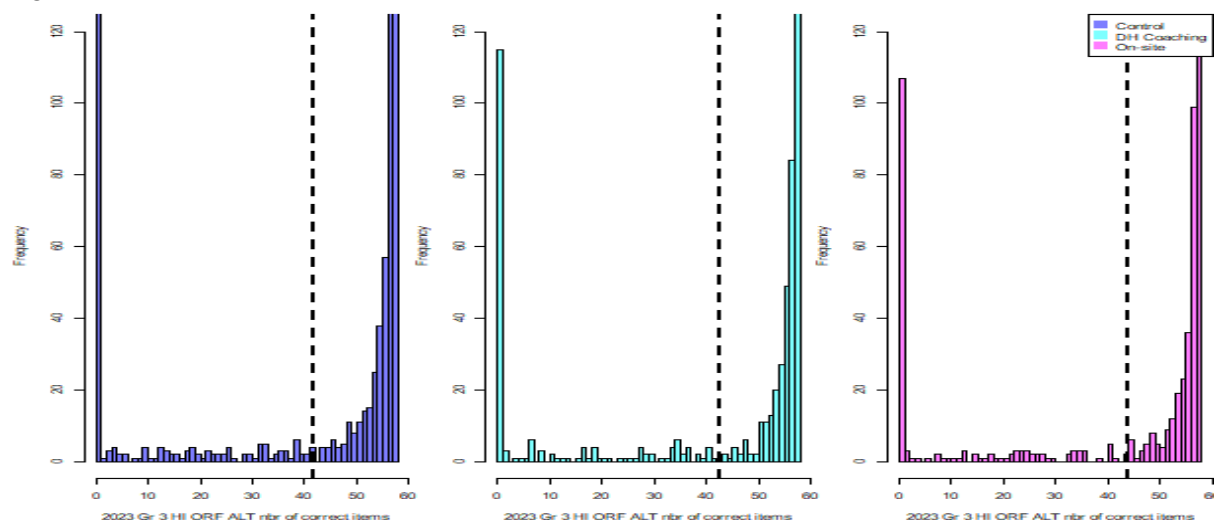
Figure XX: Grade 3 ORF ALT @ 1 minute
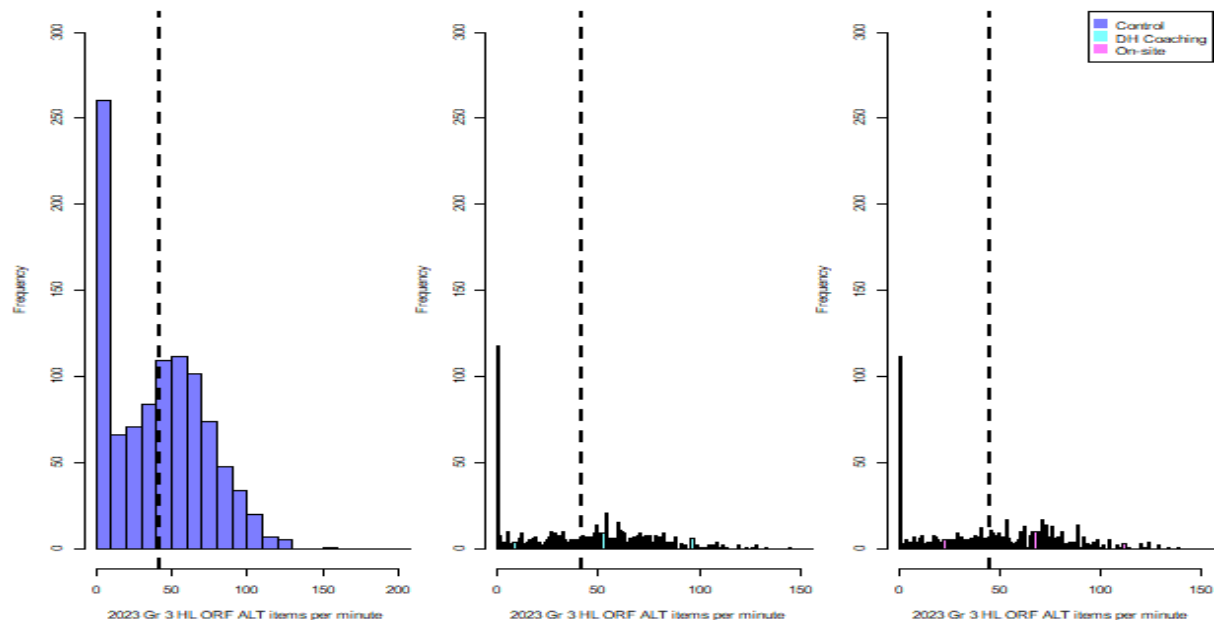
Figure XX: Grade 3 ORF ALT 2
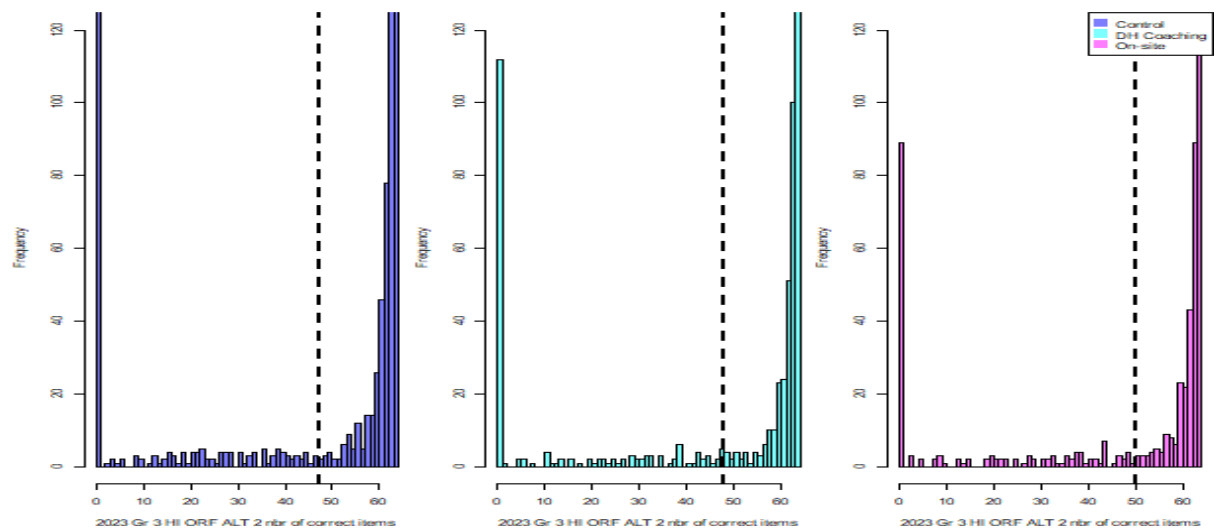
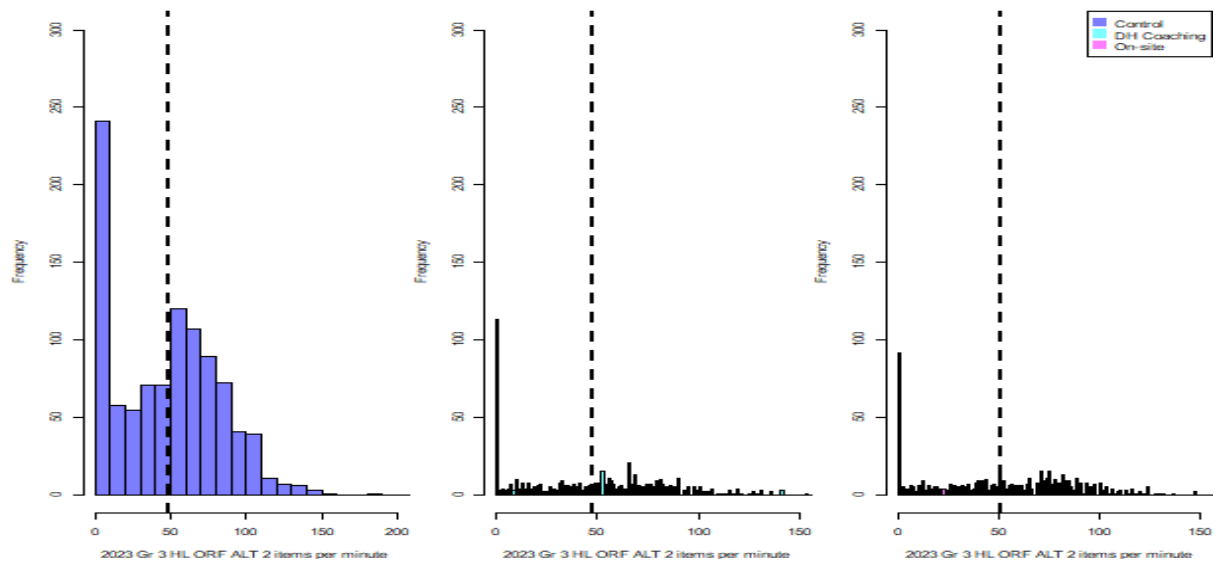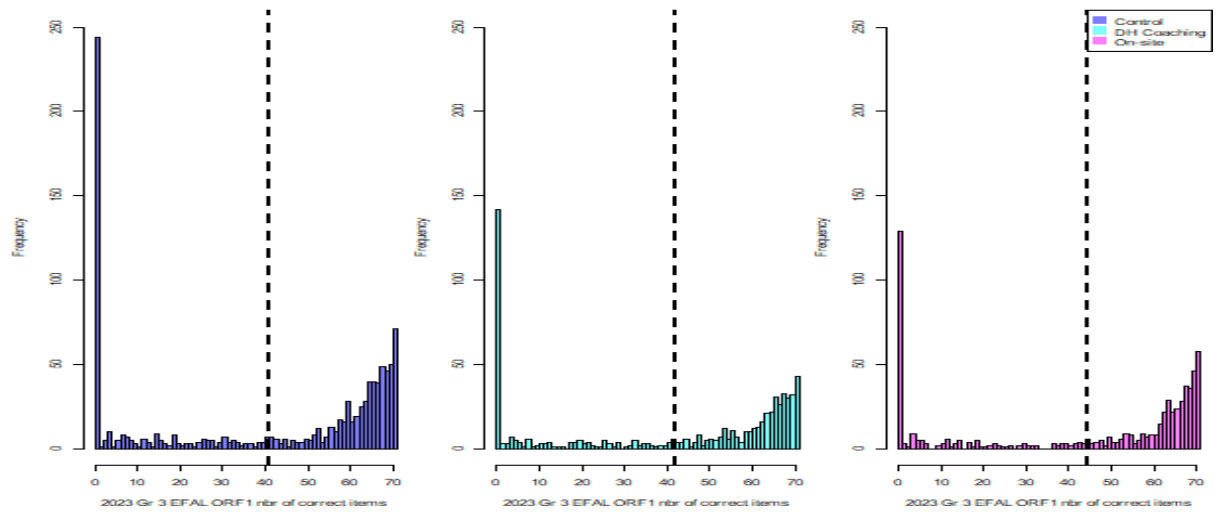## Figure XX: Grade 3 ORF ALT 2 @ 1 minute

Figure XX: Grade 3 EFAL ORF

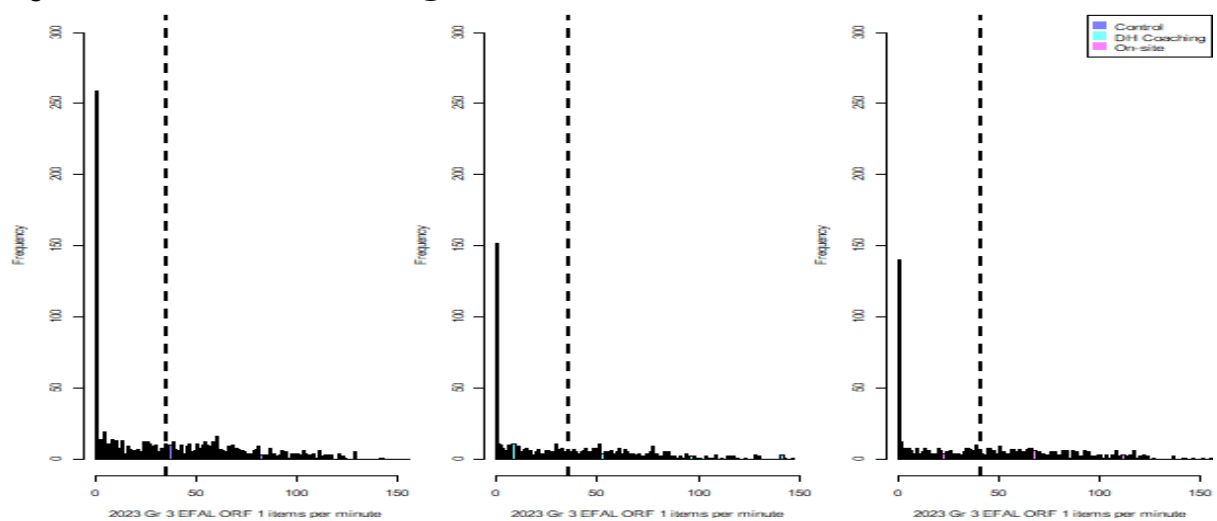## Figure XX: Grade 3 EFAL ORF @ 1 minute
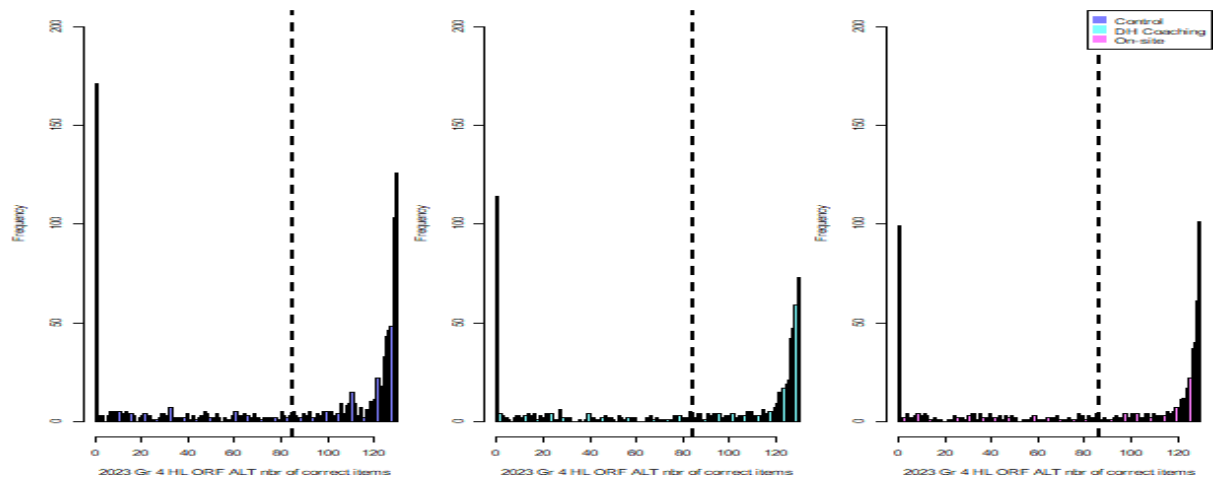
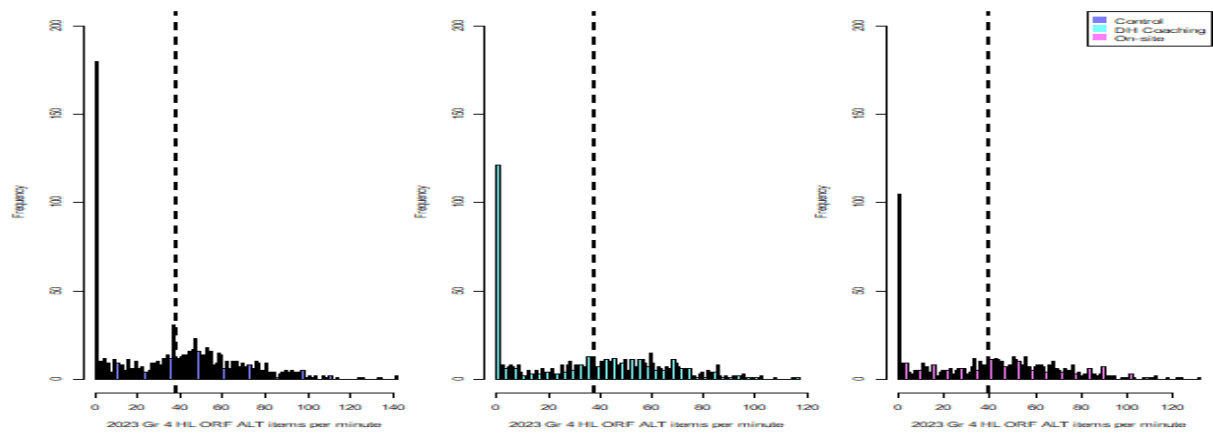Figure XX: Grade 4 ORF ALT 1
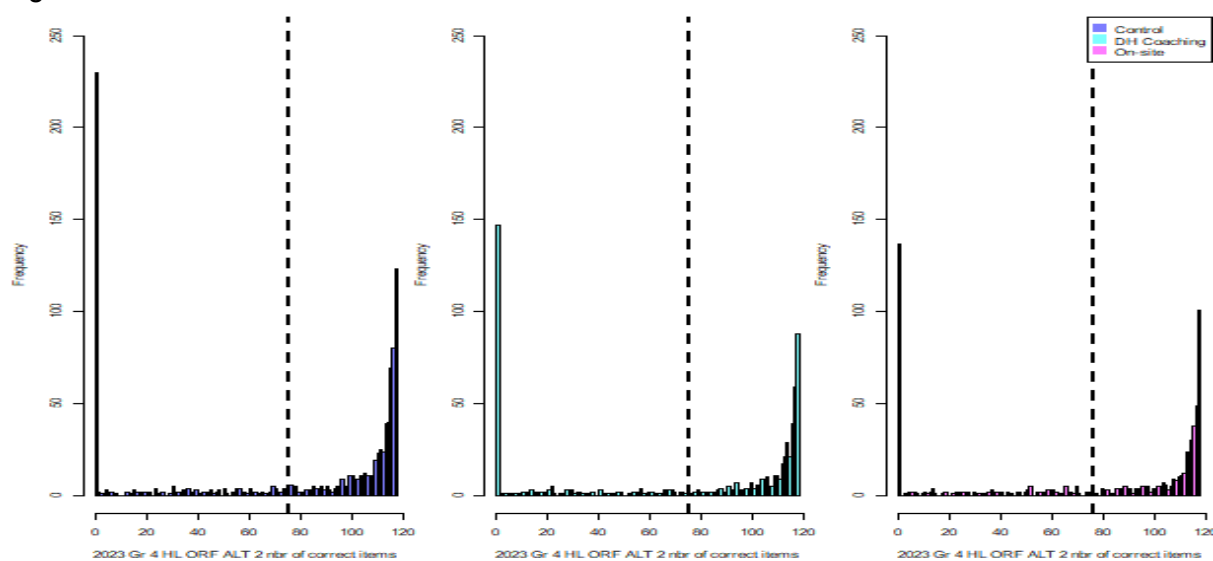

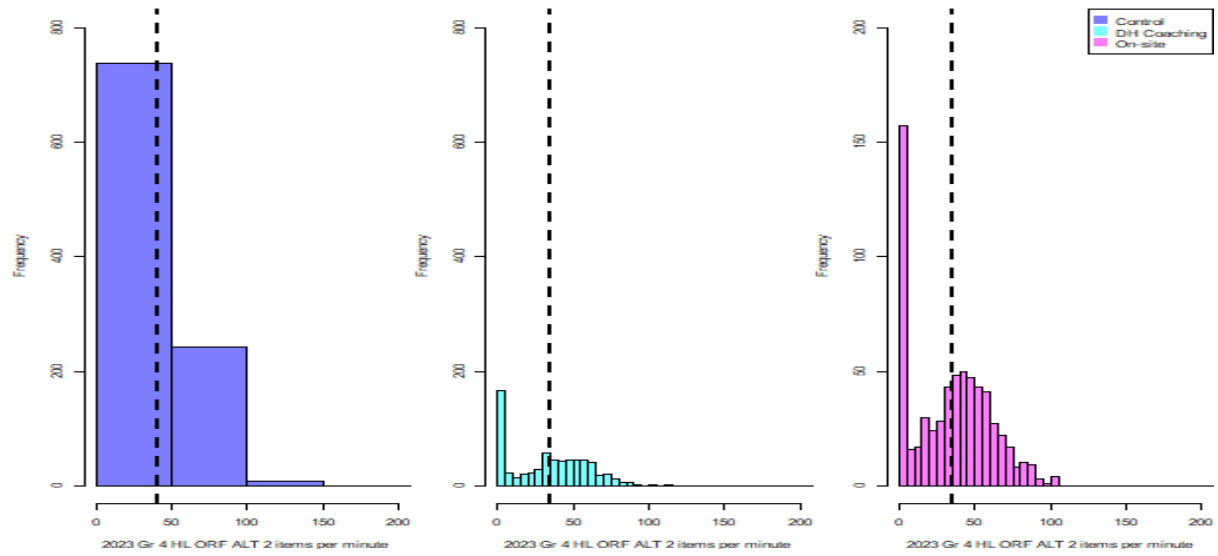Figure XX: Grade 4 ORF ALT 1 @ 1 minute
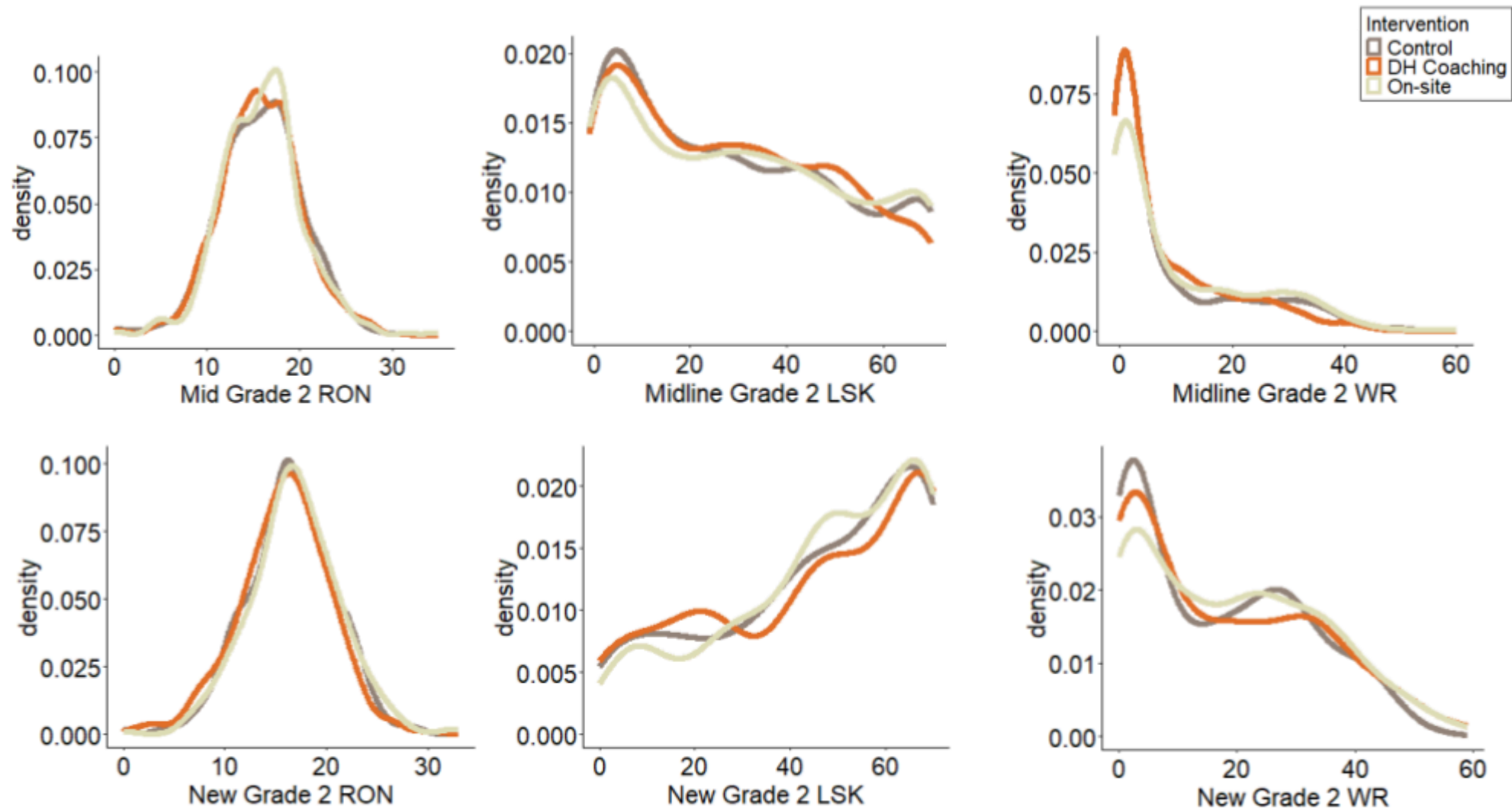

Figure XX: Grade 4 ORF ALT 2

Figure XX: Grade 4 ORF ALT 2 @ 1 minute

Appendix XX: Kernel Densities: Grade 2 Midline and Endline Rapid Object Naming, Letter Sound Knowledge and Word Recognition/ Reading

Appendix XX: Kernel Densities: Grade 2 Midline and Endline Oral Reading Fluency @3minutes, Oral Reading Fluency Correct Words per Minute, Oral Reading Fluency Comprehension
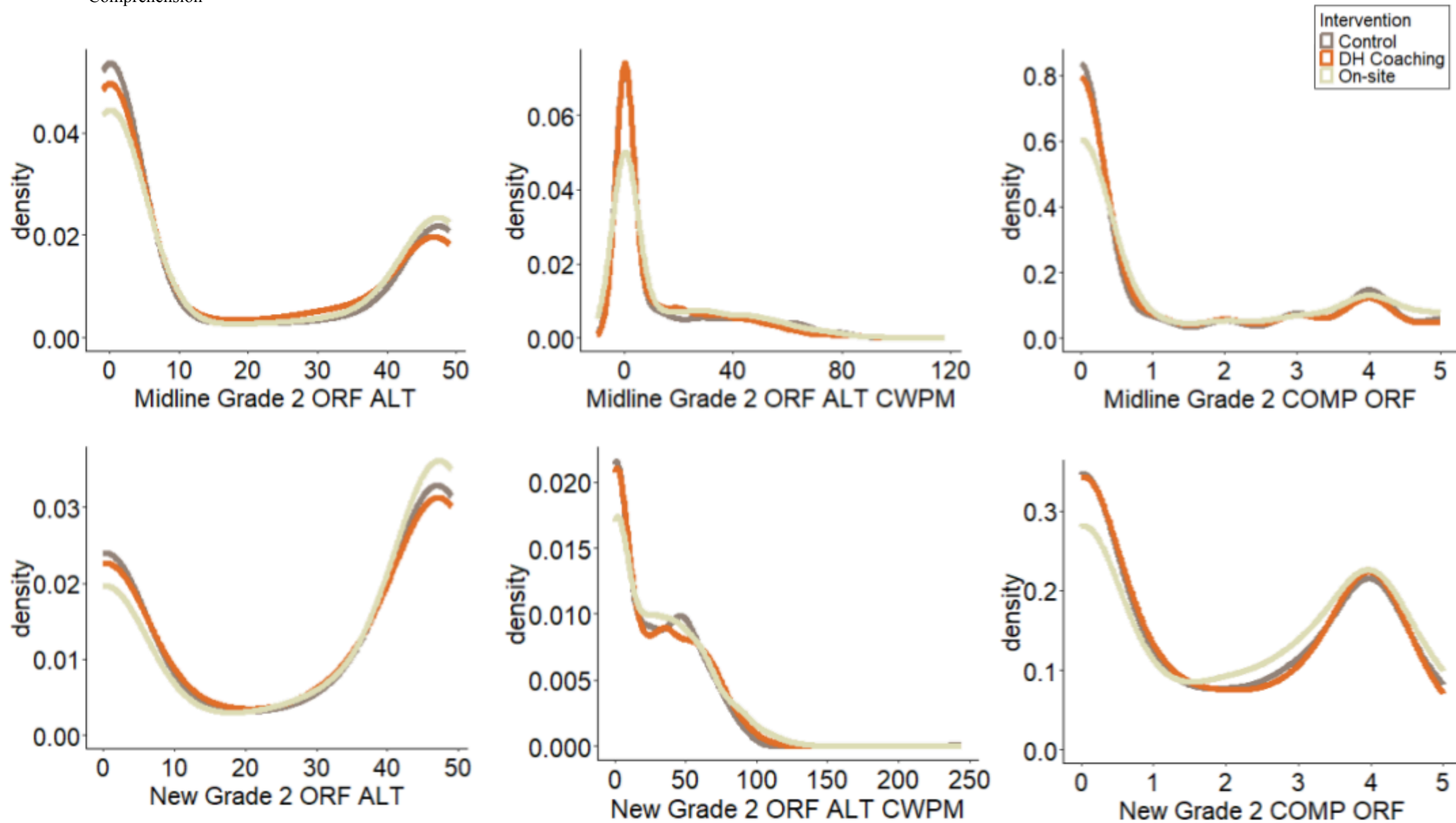
Appendix XX: Kernel Densities: Grade 2 Midline and Endline Group Test HL Complex Consonant Sound Knowledge, HL Word Recognition and Writing and EFAL Word Sound Knowledge
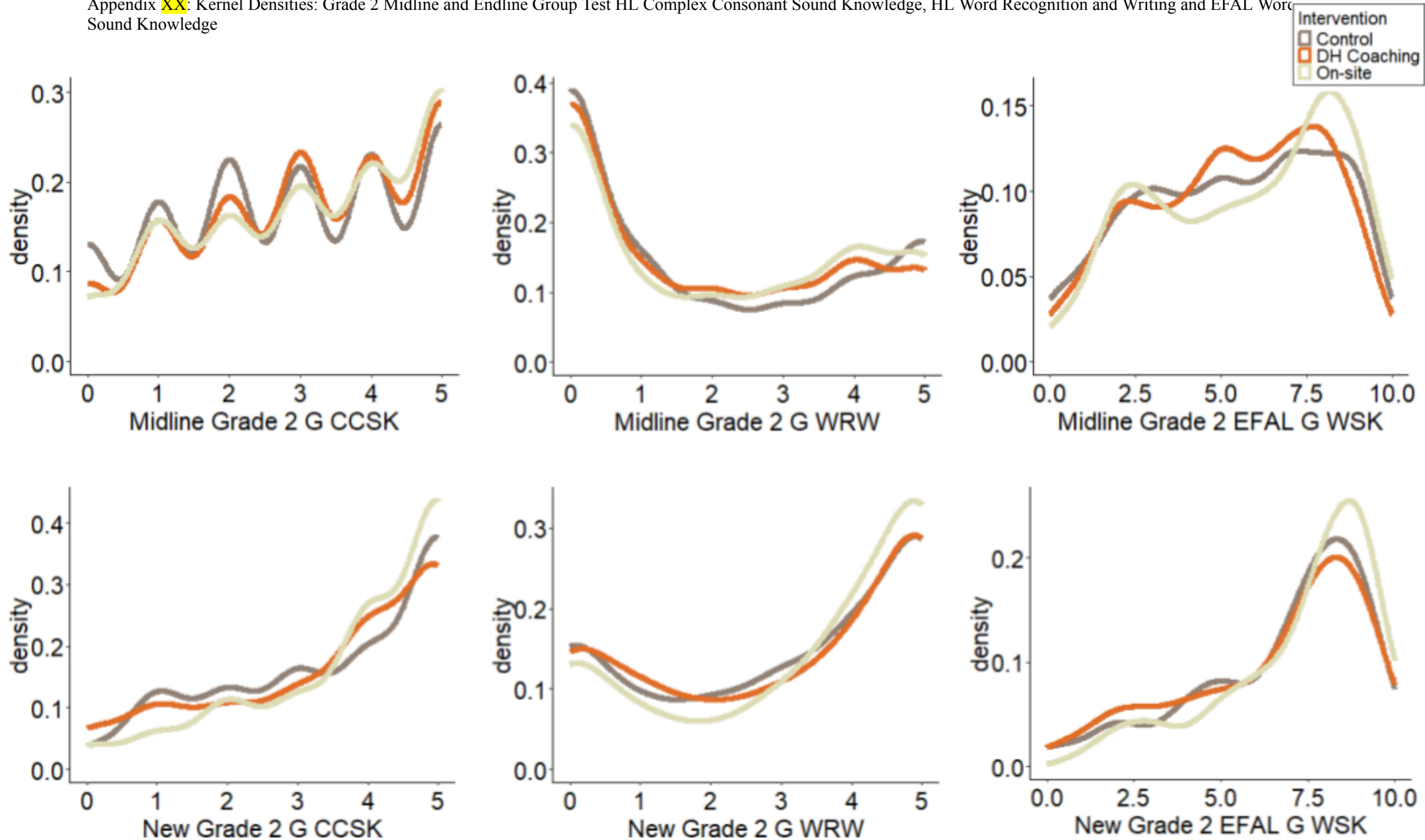
Appendix XX: Kernel Densities: Grade 3 HL Reading Fluency 1& 2 @3min, HL Reading Fluency 1& 2 Correct Words Per Minute and Comprehension
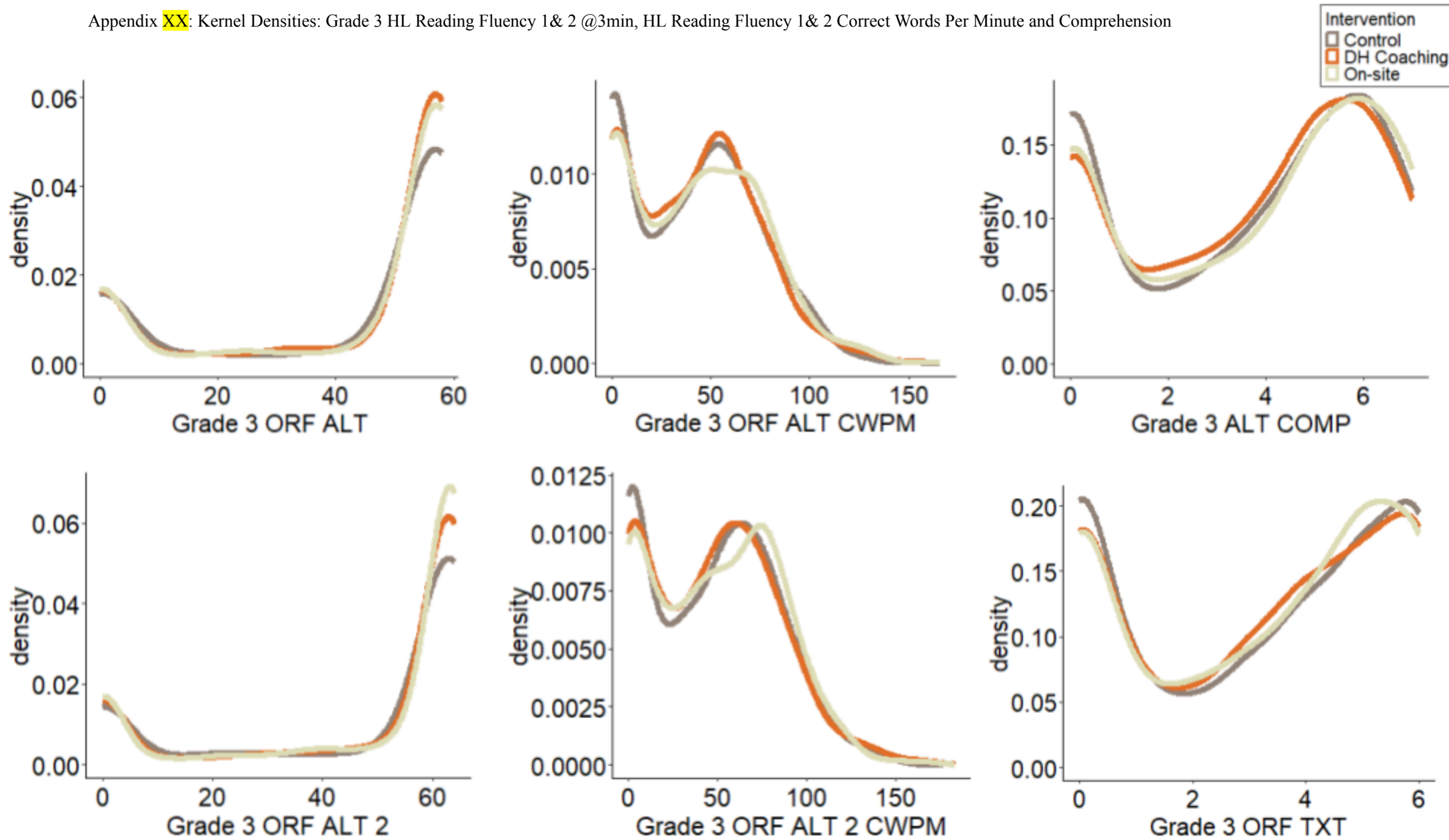
Appendix XX: Kernel Densities: Grade 3 EFAL Reading Fluency @3min and CWPM, HL Letter Sound Knowledge, Word Reading, Complex Consonant and Diacritic Knowledge
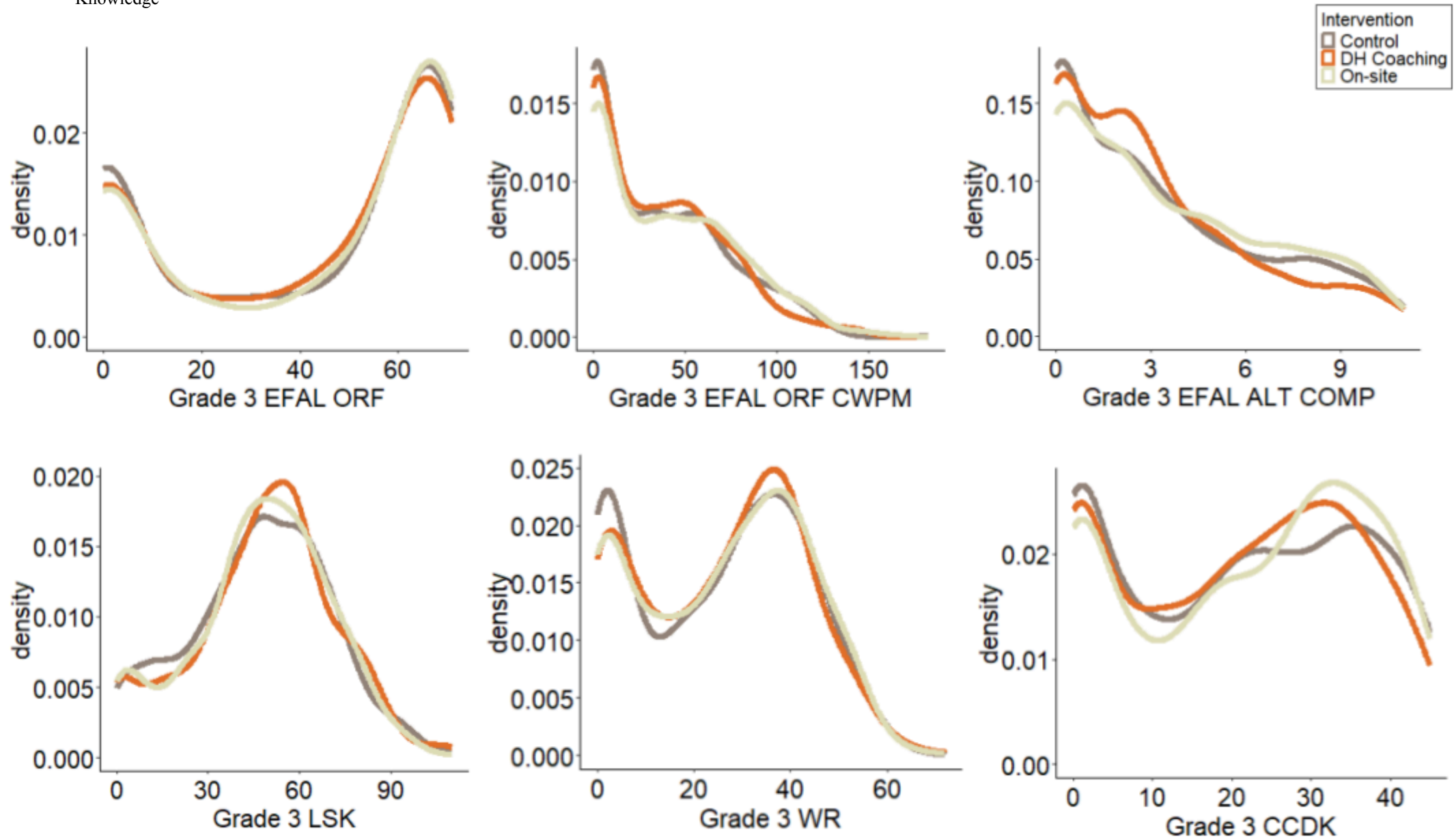
Appendix XX: Kernel Densities: Grade 3 EFAL Word Reading, Group HL Written Comprehension, and Grade 4 Oral Reading Fluency 1 & 2 at 3 minutes and CWPM
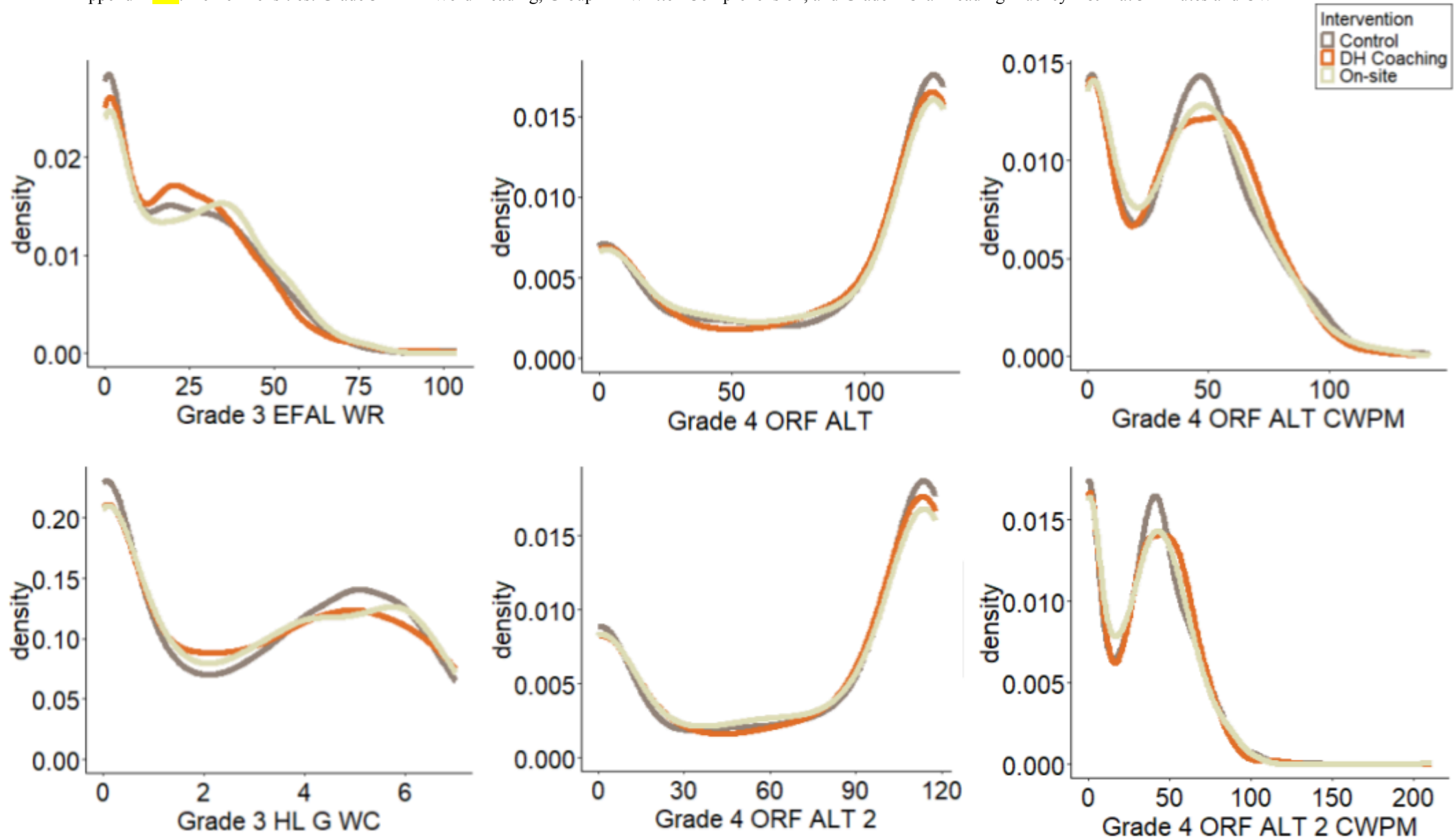
Appendix XX: Kernel Densities: Grade 4 HL Letter Sound Knowledge, Oral Reading Fluency 1 and Comprehension, and EFAL Oral Reading Fluency at 3 min and CWPM and Comprehension
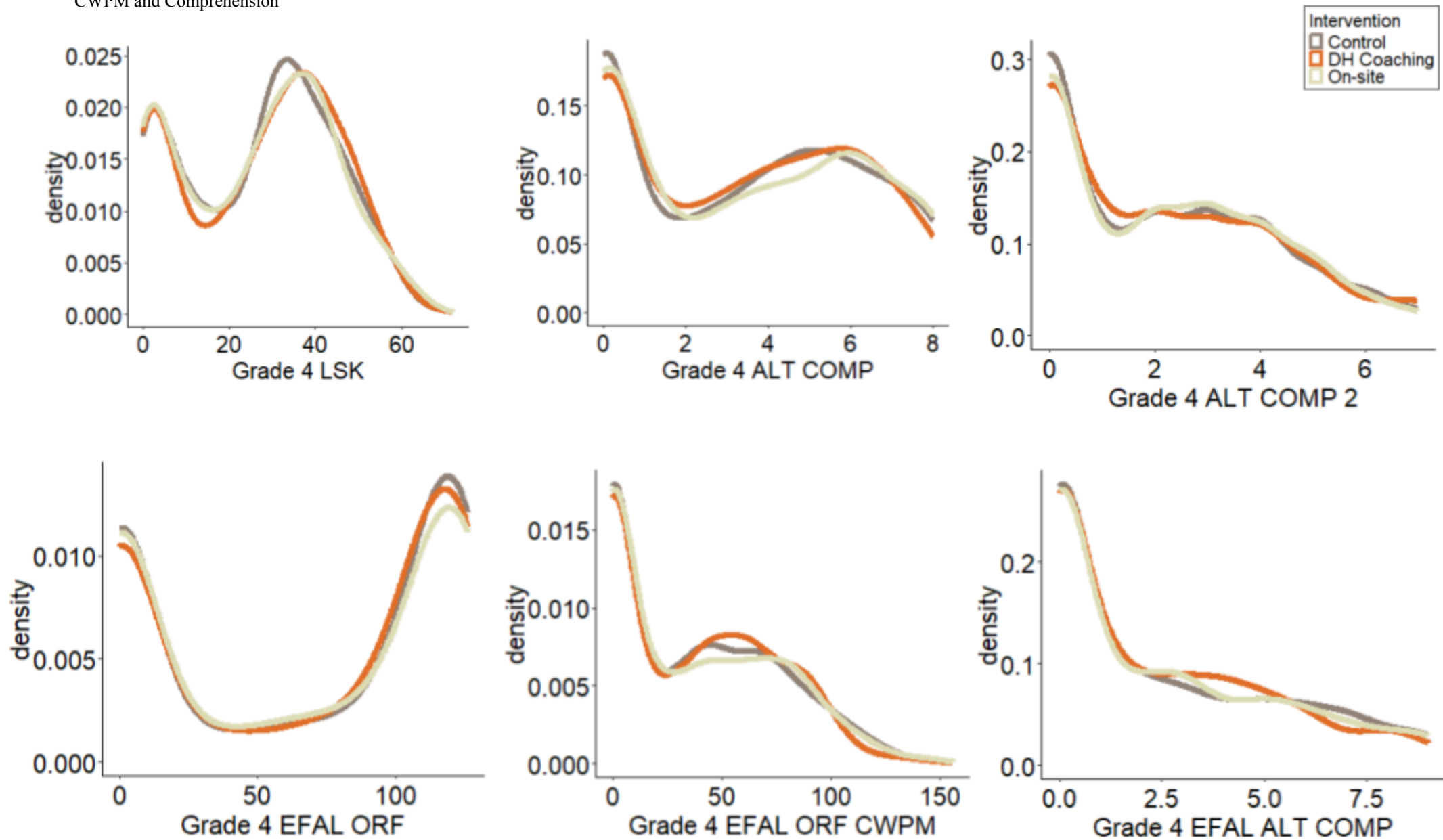
Appendix XX: Kernel Densities: Grade 4 HL and EFAL Group Written Comprehension

Appendix <mark>XXX</mark>:

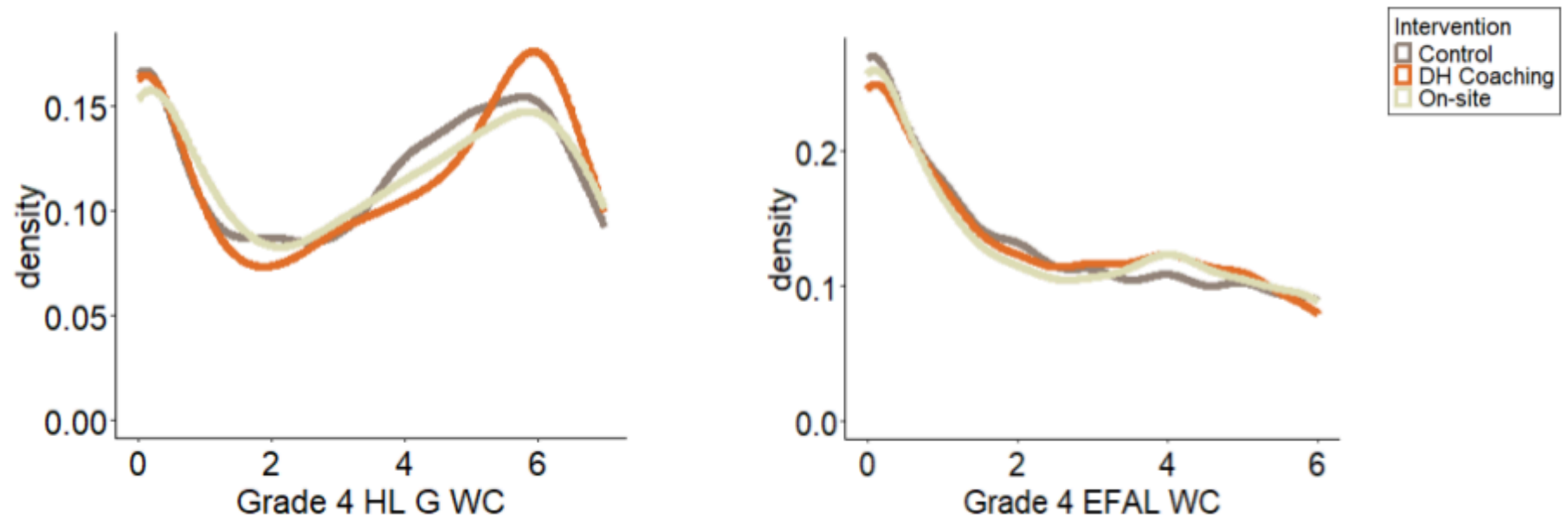|	Grade 2, Wave 2 COHORT C (NEW) Grade 2 in 2023					Grade 3 COHORT A Grade 3 in 2023					Grade 4 COHORT B Grade 4 in 2023				
	Rapid Object Naming (Items Correct)	Age	Gender	Helped with Homework at Home	Attended Grade R	Age	Gender	Read books at home	Helped with Homework at Home	Attended Grade R	Age	Gender	Read books at home	Helped with Homework at Home	Attended Grade R
	OLS	OLS	logistic	logistic	logistic	OLS	logistic	logistic	logistic	logistic	OLS	logistic	logistic	logistic	logistic
External coaching	0.527	0.035	-0.157*	0.022	-0.098	-0.095**	0.003	-0.078	0.470	0.172	-0.173	-0.103	0.018	0.518*	0.237
	(0.327)	(0.064)	(0.086)	(0.425)	(0.398)	(0.042)	(0.079)	(0.201)	(0.393)	(0.317)	(0.110)	(0.076)	(0.139)	(0.315)	(0.450)
DH coaching	-0.434	-0.015	-0.157*	0.123	0.493	-0.067	-0.023	-0.077	0.044	0.591*	-0.006	-0.047	0.252*	0.114	0.059
	(0.352)	(0.055)	(0.084)	(0.414)	(0.420)	(0.042)	(0.072)	(0.220)	(0.306)	(0.336)	(0.110)	(0.081)	(0.143)	(0.270)	(0.391)
Observations	1,057	1,057	1,057	1,057	1,057	2,249	2,249	2,249	2,249	2,249	2,260	2,260	2,260	2,260	2,260
R^2	0.042	0.045				0.027					0.009				
Adjusted R^2	0.030	0.033				0.021					0.003				
Log Likelihood			-725.760	-157.935	-163.945		-1,555.103	-1,149.103	-292.663	-245.991		-1,557.604	-1,140.172	-333.355	-288.409
Akaike Inf. Crit.			1,479.520	343.870	355.890		3,138.207	2,326.206	613.326	519.983		3,143.209	2,308.344	694.711	604.819

Note: All regressions have controls for stratum and district
*Statistical significance is indicated as follows *p<0.1 **p<0.05 ***p<0.01*

PCA Construction

The scree plots below show that the first principal component was able to explain just over 75% of all variations across all cohorts. The composite scores show a bimodal distribution for HL composite score and for EFAL composite scores show more of a floor effect.

Figure XX: A.) Grade 2, Wave 2 (Cohort C) Scree plot, B.) Grade 2, Wave 2 (Cohort C) Composite Score by treatment arm

Figure XX: A.) Grade 3 (Cohort A) HL Scree plot, B.) Grade 3 (Cohort A) HL Composite Score by treatment arm

Figure XX: A.) Grade 3 (Cohort A) EFAL Scree plot, B.) Grade 3 (Cohort A) EFAL Composite Score by treatment arm

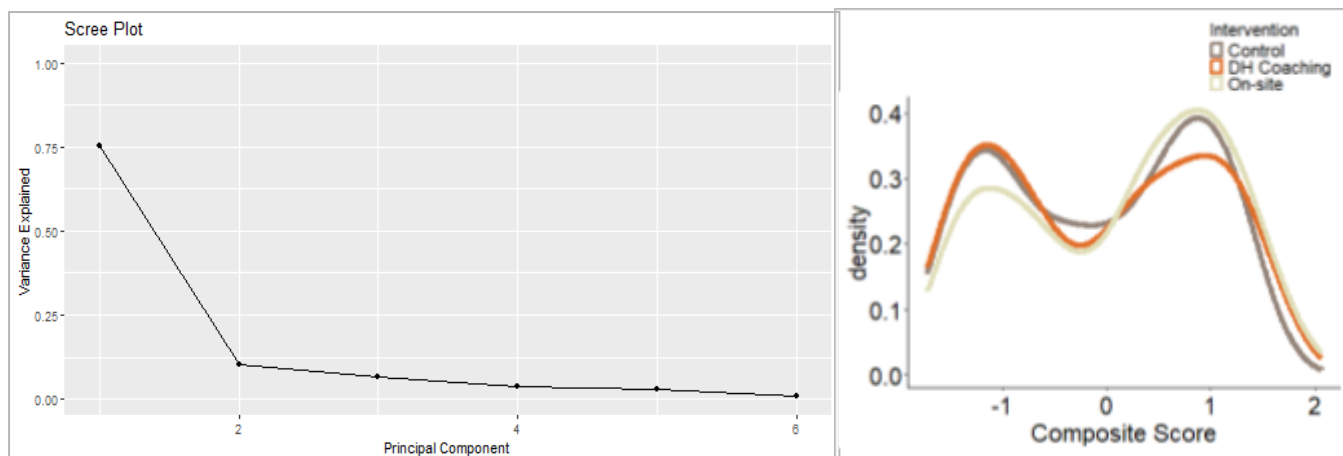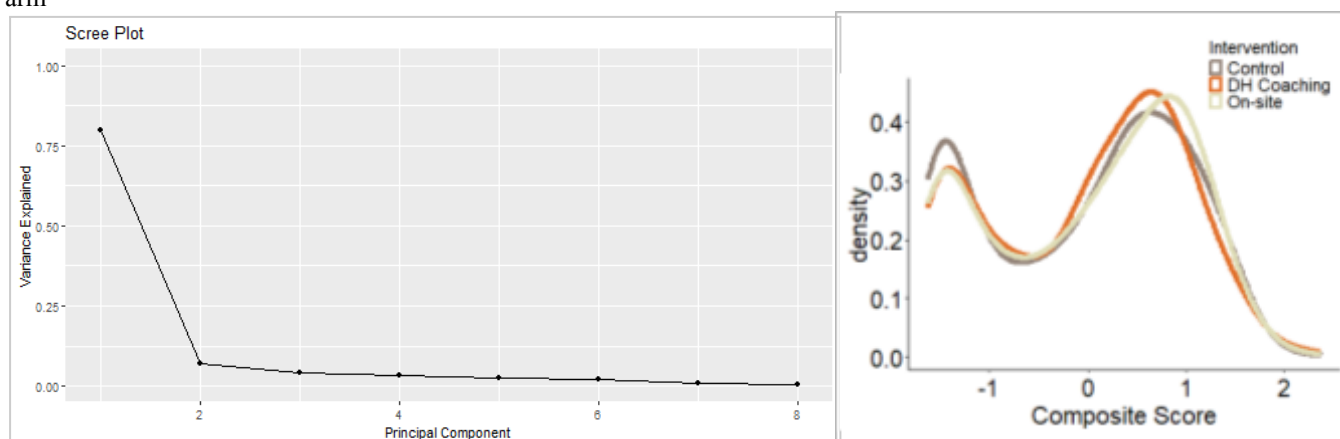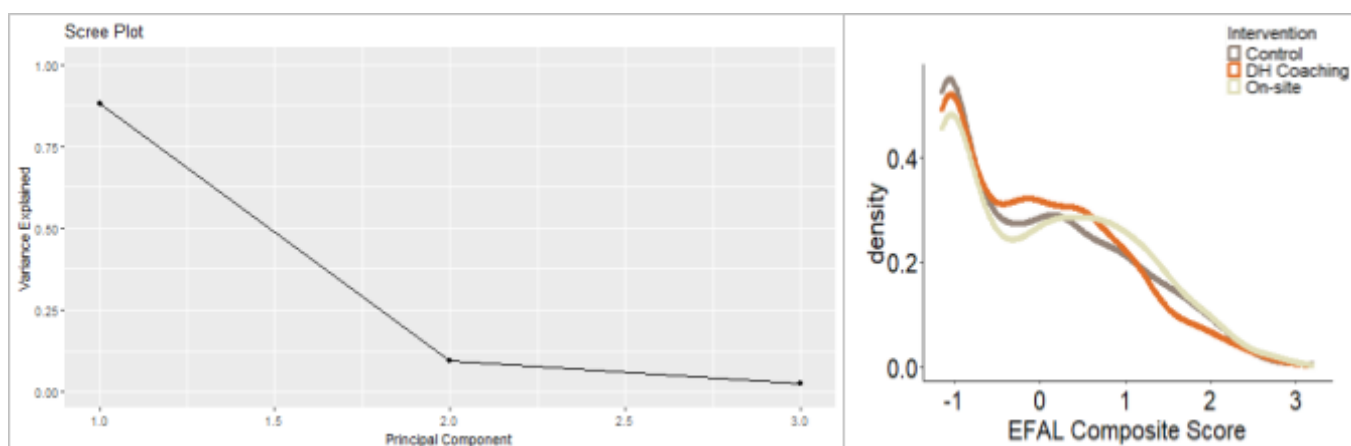Figure XX: A.) Grade 4 (Cohort B) HL Scree plot, B.) Grade 4 (Cohort B) HL Composite Score by treatment arm
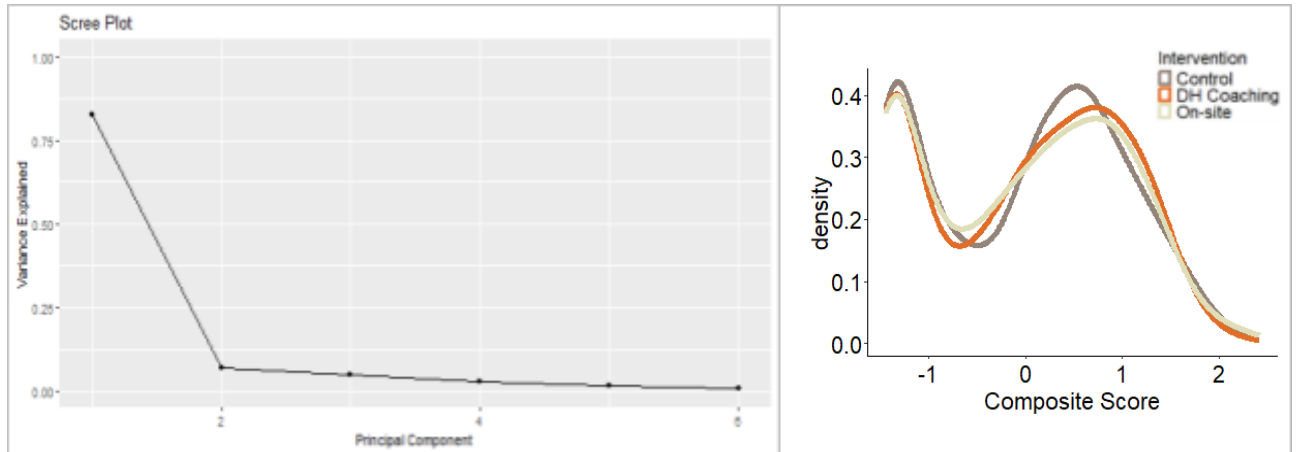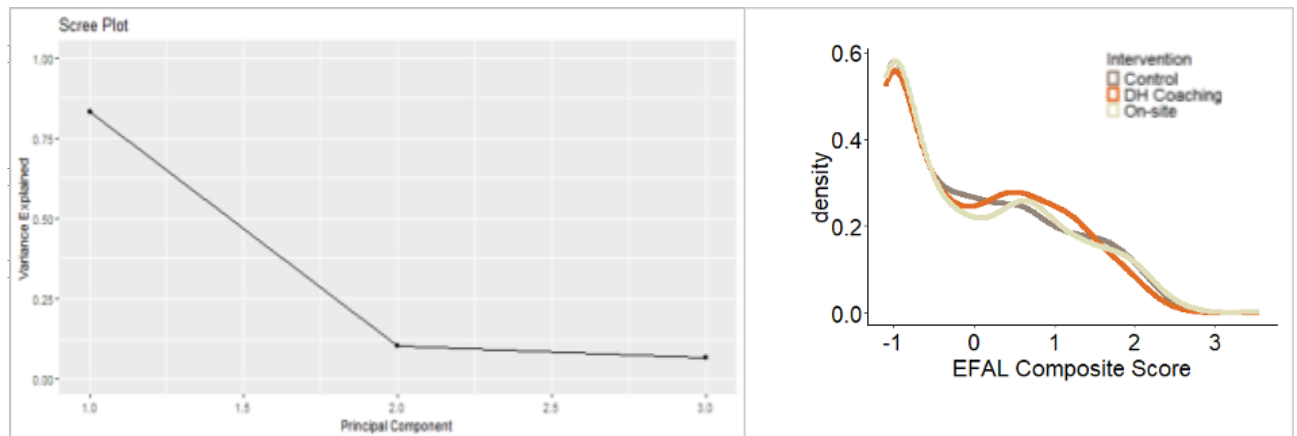

Figure XX: A.) Grade 4 (Cohort B) HL Scree plot, B.) Grade 4 (Cohort B) HL Composite Score by treatment arm

Regression Analysis

Appendix XX: Main regressions, excluding repeaters.

	Total Sample			Exclude Repeaters		
	Grade 2, Wave 2 COHORT C (NEW) Grade 2 in 2023	Grade 3 COHORT A Grade 3 in 2023	Grade 4 COHORT B Grade 4 in 2023	Grade 2, Wave 2 COHORT C (NEW) Grade 2 in 2023	Grade 3 COHORT A Grade 3 in 2023	Grade 4 COHORT B Grade 4 in 2023
Dependent variable: HL Composite Score, OLS Clustered Standard Error						
External coaching	0.255***	0.128	0.068		0.149*	0.025
	(0.088)	(0.081)	(0.077)		(0.087)	(0.080)
DH coaching	0.090	0.046	0.029		0.064	0.043
	(0.105)	(0.074)	(0.068)		(0.079)	(0.071)
Control Mean	-0.09	-0.05	-0.02		-0.05	-0.02
Observations	1,057	2,249	2,26		2,066	2,051
R2	0.106	0.125	0.101		0.108	0.104
Adjusted R^2	0.094	0.119	0.095		0.102	0.097
Dependent variable: EFAL Composite Score, OLS Clustered Standard Error						
External coaching	0.323***	0.175**	0.064		0.194**	0.018
	(0.090)	(0.081)	(0.082)		(0.089)	(0.085)
DH coaching	0.038	0.011	0.014		0.013	0.017
	(0.096)	(0.074)	(0.068)		(0.080)	(0.071)
Control Mean	-0.08	-0.05	-0.02		-0.05	-0.02
Observations	1,057	2,249	2,260		2,066	2,051
R2	0.089	0.093	0.094		0.082	0.099
Adjusted R^2	0.076	0.087	0.088		0.075	0.092

Note: All regressions have individual, district and stratum controls
Statistical significance is indicated as follows *p<0.1 **p<0.05 ***p<0.01

Sub Tasks

Grade 2 Wave 2

All sub tasks are positive and significance for on-site

Dependent variable: Grade 2, Wave 2 COHORT C (NEW)							
Home Language (Setswana)							EFAL
	Rapid Object Naming	Letter Sound Knowledge	Word Reading/ Recognition	Oral Reading Fluency CWPM	Oral Reading Fluency Comprehension	Group - Complex Consonant Sound Knowledge	Group - Word Recognition and Writing	Group - Word Sound Knowledge
	Out of 36	*Out of 70*	*Out of 60*	*Out of 49*	*Out of 5*	*Out of 5*	*Out of 10*	*Out of 10*
External coaching	0.540*	3.152*	3.368***	6.591***	0.402**	0.444***	0.454**	0.798***
	(0.324)	(1.898)	(1.215)	(2.351)	(0.162)	(0.144)	(0.183)	(0.223)
DH coaching	-0.403	0.957	1.594	3.181	0.095	0.110	0.150	0.093
	(0.353)	(2.356)	(1.538)	(2.780)	(0.189)	(0.164)	(0.195)	(0.238)
Control Mean	16.30	44.10	16.27	25.41	1.79	3.42	2.99	6.60
Observations	1,057	1,057	1,057	1,057	1,057	1,057	1,057	1,057
R2	0.050	0.103	0.083	0.088	0.083	0.066	0.100	0.089
Adjusted R^2	0.037	0.090	0.070	0.075	0.070	0.053	0.087	0.076

Note: All regressions have individual, district and stratum controls
Statistical significance is indicated as follows *p<0.1 **p<0.05 ***p<0.01

Grade 3

· HL CCDK, ORF ALT, ORF ALT2, ALT COMP is positive and significant for on-site

· All EFAL Sub tasks are positive and significant for on-site

	Dependent variable: Grade 3 COHORT A									
	Home Language (Setswana)							English First Additional Language		
	Letter Sound Knowledge	Complex Consonant Sound Knowledge	Word Reading	Oral Reading Fluency 1 CWPM	Oral Reading Fluency 1 Comprehension	Oral Reading Fluency 2 CWPM	Oral Reading Fluency 2 Comprehension	Word Reading	Oral Reading Fluency 1 CWPM	Oral Reading Fluency 1 Comprehension
	Out of 110	Out of 45	Out of 72	Out of 58	Out of 7	Out of 64	Out of 6	Out of 104	Out of 71	Out of 11
External coaching	1.134	1.989*	2.060	4.228*	0.354*	4.849*	0.231	3.189**	6.415**	0.469*
	(2.072)	(1.188)	(1.321)	(2.564)	(0.193)	(2.661)	(0.166)	(1.541)	(2.834)	(0.248)
DH coaching	0.490	0.023	0.862	1.365	0.199	1.468	0.113	0.822	0.702	-0.103
	(2.067)	(1.106)	(1.256)	(2.282)	(0.177)	(2.520)	(0.162)	(1.415)	(2.546)	(0.235)
Control Mean	46.81	20.97	24.96	40.53	3.60	46.59	3.14	21.68	36.51	3.32
Observations	2,249	2,249	2,249	2,249	2,249	2,249	2,249	2,249	2,249	2,249
R2	0.095	0.084	0.118	0.136	0.099	0.126	0.077	0.082	0.097	0.075
Adjusted R^2	0.089	0.077	0.112	0.131	0.093	0.120	0.071	0.076	0.091	0.069

Note: All regressions have individual, district and stratum controls
*Statistical significance is indicated as follows *p<0.1 **p<0.05 ***p<0.01*

Grade 4

· All subtask are positive but not significant for HL and EFAL

	Dependent variable: Grade 4 COHORT B								
	Home Language (Setswana)						*English First Additional Language*		
	Word Reading	Oral Reading Fluency 1 CWPM	Oral Reading Fluency 1 Comprehension	Oral Reading Fluency 2 CWPM	Oral Reading Fluency 2 Comprehension	Group - Written Comprehension	Oral Reading Fluency 1 CWPM	Oral Reading Fluency 1 Comprehension	Group - Written Comprehension
	Out of 72	Out of 130	Out of 8	Out of 118	Out of 7	Out of 7	Out of 126	Out of 9	Out of 6
External coaching	0.720	1.614	0.237	1.215	0.124	0.216	2.000	0.107	0.170
	(1.349)	(2.061)	(0.218)	(1.799)	(0.164)	(0.186)	(2.717)	(0.237)	(0.182)
DH coaching	0.612	0.434	0.046	0.401	0.049	0.142	0.290	-0.076	0.124
	(1.276)	(2.054)	(0.182)	(1.733)	(0.124)	(0.178)	(2.434)	(0.186)	(0.159)
Control Mean	26.69	38.16	3.35	33.63	2.13	3.34	37.19	2.41	2.19
Observations	2,260	2,260	2,260	2,260	2,260	2,260	2,260	2,260	2,260
R2	0.106	0.107	0.068	0.097	0.054	0.088	0.094	0.070	0.082
Adjusted R^2	0.100	0.101	0.062	0.091	0.048	0.082	0.088	0.063	0.076

Note: All regressions have individual, district and stratum controls
*Statistical significance is indicated as follows *p<0.1 **p<0.05 ***p<0.01*

Heterogeneity regressions

	Home Language (Setswana)								English First Additional Language			
	Grade 2, Pooled Wave 1 vs Wave 2		Grade 2 Wave 2		Grade 3		Grade 4		Grade 3		Grade 4	
	OLS		OLS		OLS		OLS		*OLS*		OLS	
	Gender	Age	Gender	Age	Gender	Age	Gender	Age	Gender	Age	Gender	Age
	-1	-2	-3	-4	-5	-6	-7	-8	-1	-2	-3	-4
External coaching	0.211**	-0.150	0.377***	0.128	0.145**	0.523	0.087	0.675	0.166**	0.279	0.088	0.290
	(0.101)	(0.439)	(0.100)	(0.717)	(0.070)	(0.656)	(0.071)	(0.492)	(0.071)	(0.656)	(0.071)	(0.490)
DH coaching	0.039	-0.023	0.162	0.503	0.119*	-0.525	-0.031	1.220***	0.119*	-0.377	0.008	1.170***
	(0.084)	(0.376)	(0.125)	(0.711)	(0.069)	(0.646)	(0.071)	(0.444)	(0.070)	(0.646)	(0.071)	(0.443)
sex	0.385***		-0.386***		0.717***		0.504***		0.609***		0.489***	
	(0.054)		(0.083)		(0.061)		(0.063)		(0.062)		(0.063)	
age		-0.022		0.017		0.033		0.003		0.018		-0.010
		(0.032)		(0.068)		(0.047)		(0.011)		(0.047)		(0.011)
External coaching: sex	0.014		-0.268**		-0.045		-0.036		0.008		-0.042	
	(0.078)		(0.124)		(0.096)		(0.098)		(0.098)		(0.098)	
DH coaching: sex	0.008		-0.164		-0.147		0.112		-0.214**		0.011	
	(0.076)		(0.148)		(0.095)		(0.096)		(0.097)		(0.097)	
External coaching: age		0.047		0.014		-0.046		-0.064		-0.012		-0.025
		(0.053)		(0.093)		(0.076)		(0.051)		(0.076)		(0.051)
DH coaching: age		0.007		-0.057		0.066		-0.122***		0.045		-0.118***
		(0.045)		(0.092)		(0.075)		(0.045)		(0.075)		(0.045)
External coaching: Endline	0.040	0.037										
	(0.118)	(0.119)										
DH coaching: Endline	0.037	0.034										
	(0.133)	(0.133)										
Observations	3,91	3,91	1,057	1,057	2,249	2,249	2,26	2,26	2,249	2,249	2,26	2,26
R2	0.161	0.123	0.109	0.041	0.124	0.016	0.102	0.036	0.094	0.018	0.093	0.040
Adjusted R²	0.157	0.119	0.096	0.026	0.118	0.009	0.095	0.029	0.088	0.011	0.087	0.033

Note: All regressions have individual, district and stratum controls
Statistical significance is indicated as follows *p<0.1 **p<0.05 ***p<0.01