

**Final Evaluation of the Early
Grade Reading
Programme II
(EGRP II)
South Africa
2024-2025**

FINAL REPORT

Evaluation Report

**Prepared by
Social Impact Insights Africa**

**Submitted to UNICEF
13 March 2026**

ACRONYMS AND ABBREVIATIONS	4
EXECUTIVE SUMMARY	4
1. INTRODUCTION	10
1.1 Purpose, objectives and scope of the final evaluation	10
1.2 Evaluation framework and questions	11
1.3 The intended use and users	12
1.4 Evaluability Assessment	13
1.5 Structure of the evaluation report	13
2. EVALUATION DESIGN AND METHODOLOGY	14
2.1 Evaluation design, methodology, and process	14
2.2 Instrument design, data collection and sample	16
2.3 Data Analysis	19
2.3.1 Quantitative: Learner assessment, implementation data and surveys	19
2.3.2 Qualitative: Case studies, observations, and KIIs	20
2.4 Limitations and challenges	20
2.5 Ethical Approach and Safeguarding	21
3. BACKGROUND AND CONTEXT	22
3.1 Policy and programming context	22
3.2 EGRP II Programme description and objectives	23
3.3 Human Rights, gender and equity issues in relation to early grade reading outcomes	24
3.4 Theory of Change	24
3.5 Key stakeholders	25
4. FINDINGS	27
4.2 Implementation fidelity	27
4.1.1 Teacher Training	27
4.1.2 Coach Training	30
4.1.3 LTSM	31
4.1.4 Coaching	40
4.1.5 Cluster workshops	43
4.1.6 Systemic reflections on training, LTSM and coaching dosage challenges	44
4.3 Teacher classroom practices	46
4.4 Learner outcomes	52
4.4.1 Description of learner assessment subtests	53
4.4.2 Learner Test Scores	53
4.4.3 Comparison of intervention streams	67
4.4.3.1. Grade 2 Home Language and EFAL Composite scores	68

4.4.3.2. Grade 3 Home Language and EFAL Composite scores	71
4.4.3.3 Distribution of effects by learner performance	73
4.4.3.4 Interpretation of learner outcomes	75
4.5 Sustainability Findings	76
EGRP II design factors	78
4.5.1 School-Based Professional Development that is Iterative and Participatory	78
4.5.2 Context-Responsive Resources and Local Adaptation	79
4.5.3 Policy Alignment and Government Leadership	79
4.5.4 Stakeholder Alliance	80
4.5.5 Financial Resources and Funding Mechanisms	81
4.5.6 Clarity on Which Design Elements Drive Embedding and Sustainability	81
4.5.7 Clear Messaging, Compelling Narrative	82
4.5.8 Clear Implementation Planning with Well-Defined Mandates	82
4.5.9 Professional Daily Practices (Implementation Capacity)	82
4.5.10 Management Capacity for Data-Driven Improvement	82
4.5.11 Alignment of Responsibility and Authority	83
5. CONCLUSIONS	86
6. LESSONS LEARNT AND RECOMMENDATIONS	88
6.1 Lessons learnt	88
6.2 Recommendations	89
6. REFERENCES	91
7. ANNEXURES	95
ANNEX 1: Terms of Reference	95
ANNEX 2: Evaluation Framework	95
ANNEX 3: Evaluability Assessment Findings	95
ANNEX 4: EGRP II Learner Assessment and Context Tools	95
ANNEX 5: EGRP II Case Study Manual	95
ANNEX 6: Training Observations tools	95
ANNEX 7: EGRP II Detailed Theory of Change	96
ANNEX 8: Teacher Training topics covered	98
ANNEX 9: Regression Models	100
ANNEX 10: Sustainability Methodology and Literature Review	105
ANNEX 11: Evaluation Team	114

TABLES OF FIGURES

Figure 1 Geographic Distribution of EGRP II Schools	12
Figure 2 EGRP II Theory of Change	24
Figure 3: 2024 Teacher attendance by Intervention and School Type	28

Figure 4: 2024 Teacher attendance by Intervention and School Type	28
Figure 5: Receipt of EGRP HL LTSM by year, school type and intervention	31
Figure 6: Usefulness of LTSMs	33
Figure 7: Time spent preparing lessons	34
Figure 8: Time spent preparing occupational tasks	34
Figure 9 Time spent preparing assessments	35
Figure 10 Tablet receipt functionality, and sharing by school type and treatment group	36
Figure 11 Distribution of coaching visits in Term 3 2025 by school type	37
Figure 12 Teacher reported quality of coaching	38
Figure 13 HL Methodologies ease of implementation	40
Figure 14 EFAL Methodologies ease of implementation	40
Figure 15 Adoption of electronic vs paper based lesson planning	42
Figure 16 Teacher experiences in accessing EGRP resources	43
Figure 17 Curriculum tracker usage patterns by teacher group	44
Figure 18 Grade 2 Subtask Correlation Matrix	48
Figure 19 Grade 3 Subtask Correlation Matrix	49
Figure 20 Grade 2 Benchmark Achievement	50
Figure 21: Grade 3 HL Benchmark Achievement	51
Figure 22: Grade 2 EFAL Benchmark achievement by treatment and school type	52
Figure 23: Grade 3 EFAL Benchmark achievement	53
Figure 24: EFAL Composite Score	54
Figure 25: Home Language Composite Score	54
Figure 26 HL Composite Density by treatment and school type	55
Figure 27: HL Composite Density (coaching only): Monograde vs Multigrade	56
Figure 28: HL Word Reading (coaching only): Monograde vs Multigrade	57
Figure 29: Grade 2 HL ORF (coaching only) Monograde vs Multigrade	57
Figure 30: Grade 2LSK cwpm (coaching only) Monograde vs Multigrade	58
Figure 31: Grade 2 EFAL Word Reading cwpm (coaching only): Monograde vs Multigrade	58
Figure 32: Grade 3 HL Subtask Densities (coaching only): Monograde vs Multigrade	59
Figure 33: Grade 3 EFAL cwpm Density (coaching only): Monograde vs Multigrade	59

TABLE OF TABLES

Table 1:Evaluation Criteria	13
Table 2: Intended users of the evaluation	13
Table 3: Evaluation Methods	16
Table 4 EGRP II Programme Elements	22
Table 5: Programme Stakeholders and their Roles	24
Table 6 EGRP II Assessment tasks administered at the endline	45
Table 7 Grade 2 Summary statistics per sub-task	46

ACRONYMS AND ABBREVIATIONS

BMGF	Bill and Melinda Gates Foundation
CAPS	Curriculum and Assessment Policy Statement
CPTD	Continuous Professional Training and Development
CWPM	Correct Words per Minute
CLPM	Correct Letters 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
EGRP II PMT	EGRP II Project Management Team
HL	(African) Home Language
IPA	Interpretative Phenomenological Analysis
KII	Key Informant Interview
LTSM	Learning and Teaching Support Material
M&E	Monitoring and Evaluation
OUP SA	Oxford University Press South Africa
PIRLS	Progress in International Reading Literacy Study
PILO	Program to Improve Learning Outcomes
PLC	Professional Learning Community
RCT	Randomised Control Trial
SIOC-CDT	Sishen Iron Ore Company's Community Development Trust
SLP	Structured Learning Programme
SBW	School Based Workshops
SMT	School Management Team
RSA	Republic of South Africa
TOC	Theory of Change
UNDP	United Nations Development Programme
UNEG	United Nations Evaluation Group
UNICEF	United Nations Children's Fund

EXECUTIVE SUMMARY

Background

South Africa faces a significant early grade reading crisis, with 81% of Grade 4 learners unable to read with comprehension, most acutely in lower quintile schools and compounded by COVID-19 learning losses. Since 2015, the Department of Basic Education (DBE) and its partners have run a series of Randomised Controlled Trials (RCTs) aimed at learning lessons to improve literacy outcomes in the Foundation Phase for the public schooling system. This series has consistently found that the combination of teacher training, learning and teaching support materials (LTSMs), and external instructional coaching (collectively known as the "triple cocktail") improves reading outcomes, with schools coached the EGRP I evaluation found that school-based coaches (Department Heads), could not replicate these results. Critically, all studies in this series prior to EGRP II were designed for monograde classrooms.

The EGRP II intervention was commissioned by the Department of Basic Education (DBE) as a continuation of the Early Grade Reading Studies series. The intervention was funded by UNICEF and the Zenex Foundation with additional contributions by other funders. The EGRP II implementing partner was Oxford University Press. The programme was implemented across 116 schools in poorer schools (quintiles 1-3) in the Frances Baard and John Taolo Gaetsewe districts in the Northern Cape, South Africa, in 2024 and 2025. EGRP II extends the proven "triple cocktail" model to multigrade schools for the first time, providing tailored teacher training, multigrade specific LTSMs, and differentiated external instructional coaching support alongside system level support and comparing the results to monograde schools also receiving materials, training and coaching.

Purpose, objectives, and methodology

This impact evaluation aims to generate evidence and lessons to inform the DBE's decision on scaling up the programme. Three key objectives guide the evaluation: (1) assessing the [impact](#) of external coaching, teacher training, and LTSM provision on reading outcomes in HL and EFAL; (2) comparing [intervention effectiveness](#) across multigrade and monograde settings while examining implementation fidelity, including contextual barriers to adoption; and (3) assessing programme [sustainability](#) including an assessment of system-level data.

Commissioned by UNICEF and the Zenex Foundation and carried out by Social Impact Insights Africa (SII Africa), the evaluation used a mixed-method approach. The [qualitative component](#) comprised two rounds of case studies across 10 schools, drawing on classroom observations, and teacher interviews, supplemented by monitoring data to understand implementation. Teacher and coach training observations were also carried out. The [quantitative phase](#) was a randomised controlled trial across 116 schools, comparing coached and non-coached schools (the latter receiving only learning and teaching support materials (LTSMs) and teacher training) in both monograde and multigrade settings. Quantitative data included Early Grade Reading Assessments (EGRA) and Early Grade Mathematics Assessments (EGMA) administered to 3 876 Grade 2 and 3 learners, alongside teacher and principal surveys, school observations, and administrative data. Group comparability was confirmed through sensitivity analyses, balance

tests, and composite outcome models. A key design limitation was the absence of baseline data, which precluded measuring learning growth over time or comparing the base programme against schools receiving no support. Design limitations included the absence of pre-intervention baseline data, which precluded measuring growth over time and comparing the effect of the base programme (non-coached group receiving LTSM and teacher training) with schools that received no support.

The evaluation serves stakeholders across national, provincial, and implementing partner levels, including the DBE, UNICEF, ZENEX, Gates Foundation, and Oxford University Press. The findings are also intended to inform broader policy dialogues and early grade reading research.

Key findings

EFFECTIVENESS FINDINGS

The assessment of implementation effectiveness presents an encouraging picture overall, with strong performance across teacher training, LTSM, coaching quality, and classroom practice. While dosage targets were not always fully met, the quality of implementation across core programme elements remained relatively good.

- **Implementation fidelity was generally high**, especially regarding LTSM provision and teacher training, despite the effects of initial contracting delays and contextual challenges (e.g., departmental austerity measures and teacher strikes impacting training, tablet donation timelines, etc.). Teachers gave strongly positive feedback on material quality and relevance, citing strong learner engagement (learners recalled favourite stories by name), print-rich classroom transformation, and significant time savings.
- **Coaching quality was generally high**, following core practices of instructional coaching. Coaching dosage was overall lower than planned, suggesting the need to reconsider practicable coaching dosages in future interventions. Dosage shortfalls have been a recurring challenge across similar programs in South Africa, possibly reflecting the structural and contextual barriers inherent in delivering intensive support at scale. Multigrade teachers received a significantly higher dosage of coaching than monograde teachers. Coach prioritisation of multigrade schools, which they acknowledged finding more complex to support, meant that their most intensive efforts were concentrated where the need for specialised guidance was greatest.
- **Coaching appears to have strengthened the uptake of programme components**. Coached teachers demonstrated more consistent attendance at training sessions than their non-coached counterparts and coaching deepened teachers' appreciation of and active engagement with LTSM. When asked if the multigrade-specific materials were more useful than monograde equivalents, 71% of coached multigrade teachers recognised their added usefulness, while only 32% of their non-coached peers did so.
- **Classroom practice experienced meaningful shifts particularly in multigrade settings**. Teachers towards ability-based grouping across grades, supported by structured occupational tasks that engaged different groups simultaneously. For many multigrade teachers, this represented a genuinely transformative change, as they had previously lacked

a clear methodology for managing simultaneous instruction. LTSMs played a central role in this shift, transforming classrooms into print-rich environments and meaningfully reducing the planning and administrative burden that disproportionately affects multigrade teachers.

- **Teachers showed improved adoption of core reading methodologies.** Stronger adherence to Shared Reading and Phonics routines and greater confidence in EFAL instruction were observed whilst Writing and Group Guided Reading remained the most challenging methodologies across both languages and school types, consistent with previous studies, though coaching appears to have partially mitigated these difficulties.

IMPACT FINDINGS

The assessment of the impact of the intervention, constrained by the absence of a baseline, cautiously concludes that the programme specifically benefited multigrade schools.

- **Coaching reduced multigrade disadvantage (Grade 2):** The evaluation confirmed the structural disadvantage inherent in Grade 2 multigrade classrooms, with uncoached multigrade learners performing at lower literacy levels than their monograde counterparts. However, this gap disappeared by Grade 3, where performance levels were comparable across both school types.
- The most significant evaluation finding was that *the addition of coaching resulted in Grade 2 multigrade learners overcoming the multigrade disadvantage*, achieving literacy outcomes comparable to, and in several subtasks exceeding, those of their (uncoached and coached) monograde peers. In Grade 2 multigrade schools, coaching produced statistically significant gains of between 0.35 and 0.47 standard deviations across Setswana Home Language word reading, oral reading fluency, and English First Language word reading. These effect sizes are considered large by international benchmarks.
- **Coaching did not generate significant learning advantages for monograde learners (Grade 2 or in Grade 3 (mono- or multigrade):** No comparable difference between coached and base programme learners was measured in monograde schools at Grade 2, and coaching effects were not statistically significant at Grade 3 in either school type.
- **Potential explanations for the multigrade Grade 2 coaching effect include:**
 - **Growth from a lower base:** multigrade teachers had previously received no dedicated support with appropriate LTSM and pedagogies, so the provision of tailored attention and context-specific resources is likely to have generated heightened motivation and stronger implementation among the teachers.
 - **Larger coaching dosage:** multigrade teachers received more coaching support than monograde teachers.
 - **Grade 1 advantage:** since the EGRP II was a two-year programme, the Grade 2 learners at the endline had been exposed to quality teaching support since Grade 1, while the Grade 3 learners did not have the benefit of supported teaching in this core foundational learning year. This may have impacted on the cumulative learning journey.
- **Despite relative gains, absolute performance levels remain critically low, reflecting the depth of South Africa's literacy crisis.** Across all Grade 2 and Grade 3 Setswana contexts, DBE benchmark achievement ranged from 18% to 35%, indicating that the large majority of learners, whether coached schools or not, are still failing to read at grade appropriate levels.

Cross-study comparisons, treated with caution given differences in design and context, suggest the EGRP II gains were similar to comparable contexts without intervention. The most notable gains were in multigrade coached schools, where Grade 2 home language oral reading fluency nearly doubled, pointing to targeted multigrade coaching as EGRP II's greatest area of added value.

SUSTAINABILITY FINDINGS

The sustainability assessment presents a mixed picture, with genuine progress in government ownership alongside significant structural challenges and binding constraints.

- **Multigrade Setswana LTSMs are the most durable programme legacy because of their context responsiveness and local adaptation.** The Setswana LTSM represents the programme's most enduring material legacy. As the first structured reading resources of their kind tailored for multigrade classrooms in the Northern Cape, they fill a gap no other South African programme currently addresses, and have generated strong teacher and official endorsement and ownership that is likely to sustain their active use beyond the programme.
- **Genuine progress was made in achieving policy alignment and building government ownership, leadership, and capacity.** However, embedding remains incomplete and structurally fragile, compromised by fiscal austerity, unresolved SA shortages, and the absence of formally assigned accountability, authority and resources for continuation of activities.
- **Coaching is the most effective but the least sustainable element.** The proposed abridged model for sustaining an equivalent of coaching using SAs and champion teachers in cluster workshops is a pragmatic adaptation given the system constraints, but it is qualitatively different from the in-classroom, individualised coaching that drove results, and its effectiveness remains unproven.
- **The introduction of the National Home Language Reading Programme across Northern Cape districts outside the EGRP II footprint adds a further risk if not integrated with the reading strategy.** While the two programmes are considered complementary, a government-owned initiative carries institutional gravity that an externally funded programme cannot match, and in a fiscally stressed system, planning energy and budget allocations will likely gravitate toward it. If EGRP II is not explicitly integrated into the Provincial Reading Strategy, its long-term embedding beyond the current funding period remains vulnerable, and austerity is likely to sharpen rather than ease this tension.
-

Evaluation key conclusions

- **Targeted coaching as an equaliser for multigrade schools:** Grade 2 learners in coached multigrade schools achieved literacy outcomes comparable to or exceeding monograde peers despite historical structural disadvantages, suggesting that coaching, training, and customised materials offer a viable pathway to address inequalities in rural multigrade schools.

- **Coaching played a critical role in strengthening the uptake and appreciation of programme components.** This compounding effect suggests that coaching is not merely an add-on to the programme but an enabling or mediating condition that amplifies the value of its other elements. Coached teachers demonstrated more consistent training attendance and deeper engagement with learning and teaching support materials than their non-coached counterparts.
- **Pedagogical flexibility over rigid standardisation:** Success in multigrade classrooms stemmed from fundamental shifts toward ability-based grouping rather than rigid grade-level teaching. Given that most South African classrooms have wide variation in learner ability, the differentiated instructional approaches developed for EGRP II may offer valuable support to teachers in monograde settings facing similar challenges.
- **Implementation quality shapes programme outcomes:** Training attendance was high, materials well received and coaching quality relatively good, a higher dosage for multigrade schools may have produced differential outcomes between treatment groups.

Key recommendations

A. Recommendations for the DBE and PED

1. Prioritise targeted support for multigrade early grade teachers. The DBE should assist PEDs in developing and delivering specialised training for multigrade teachers, with particular emphasis on mastering Group Guided Reading, Shared Reading methodologies, use of occupational tasks and learner grouping. These approaches enabled multigrade teachers to fundamentally rethink classroom management and simultaneously instruct learners across multiple grades and ability levels.
2. Standardise and improve the validity of SBAs: The DBE, in collaboration with provincial education departments, should develop a repository of CAPS aligned standardised assessment tasks for the Foundation Phase, updated annually to prevent teaching to the test. Regular reliability testing against externally facilitated assessments would provide a systemic resource for future evaluations, reducing the need for separate baseline assessments
3. Strengthen Subject Advisor capacity to support sustainability. Prioritise addressing Subject Advisor shortages and investing in their ongoing capacity development. SAs are central to sustaining school-level support and play a critical role in embedding programme gains within the system.

B. Recommendations for future research and programme design

1. Establish reliable baseline data collection protocols. All future interventions should prioritise collecting baseline learner performance data before programme commencement. The absence of such data in this evaluation limited causal claims and the measurement of learning growth over time. Improved system-generated data, including regular systemic assessments and standardised school-based assessments, may reduce this need in future, though their applicability to each evaluation context should be established in advance.

2. Investigate coaching effects in multigrade contexts through rigorous experimental design. The promising results in coached multigrade classrooms call for further exploration through studies employing baseline measures and experimental designs to validate or challenge these findings. Research should examine mechanisms driving efficiency gains, pedagogical shifts, and learner outcomes, as well as explore whether multigrade-adapted approaches could benefit monograde classrooms with diverse learner ability groups.
3. Refine coaching dosage and quality standards. Given that dosage targets have gone unmet across successive similar education programmes, future research should investigate the contextual and systemic factors constraining delivery, and examine what minimum effective dosage is required to produce meaningful outcomes.

1. INTRODUCTION

1.1 Purpose, objectives and scope of the final evaluation

The purpose of this impact evaluation is to assess the second Early Grade Reading Program (EGRP II) across three critical dimensions: learner reading outcomes, fidelity of implementation, and pathways to long-term sustainability. The evaluation provides evidence-based insights to inform the Department of Basic Education's (DBE) decisions on scaling up the programme to other provinces. By examining what works, how it works, and under what conditions, the evaluation contributes to the body of knowledge on improving learner outcomes and supports strategic planning for expanding effective reading interventions across South Africa's education system.

The **objectives** of the evaluations were to:

- Assess the effectiveness of external coaching, teacher training, and Learning and Teaching Support Materials (LTSM) provision in improving learners' reading outcomes in Setswana Home Language (HL) and English First Additional Language (EFAL).
- Compare results between monograde and multigrade classes in the Northern Cape and identify contextual factors that support or hinder EGRP II effectiveness and adoption in multigrade compared to monograde classrooms.
- Assess implementation fidelity to determine whether the intervention was delivered as designed and identify contextual factors contributing to any variations.
- Explore the utility of using existing national system-level data to assess EGRP II impact and identify gaps and areas where national data systems should be strengthened.
- Assess programme sustainability beyond the intervention period.

The **scope of the evaluation** covers the full implementation period of the EGRP II from 2024 to December 2025. It encompasses five main components:

- **Evaluability Assessment:** Since baseline data was not collected prior to the intervention's start in 2024, this assessment evaluated the usability of existing system-level data for establishing a proxy baseline and for continuous monitoring of EGRP outcomes. It also identified gaps and areas requiring improvement in national data systems.

- **Implementation Fidelity:** This measured whether EGRP II was delivered as intended, focusing mainly on core EGRP II Theory of Change (ToC) elements, training, coaching and LTSM indicators.
- **Intermediate Outcomes and Impact:** Assessed the programme's effects on teaching practices and learner achievement.
- **Sustainability Assessment:** Examined the programme's long-term viability and potential for embedding within the education system.

Methodologically, the evaluation employed both quantitative and qualitative approaches. The quantitative component included primary data collection in 118 schools to assess Grade 2 learners using standardised Early Grade Reading Assessment (EGRA) and Early Grade Mathematics Assessment (EGMA) tools in addition to principal, teacher and Departmental Head (DH) surveys. The qualitative component comprised 10 in-depth case studies with two rounds of data collected in the same schools in 2024 and 2025. All data collection instruments were designed or adapted for the evaluation and quantitative tools were piloted prior to fieldworker training and data collection.

The evaluation's **geographical scope** was limited to the Northern Cape, South Africa's largest province by land area. Characterised by arid landscapes and sparse population, the province has a high prevalence of multi-grade schools. All data collection activities were conducted within two of the province's five districts, Frances Baard (FB) and John Taolo Gaetsewe (JTG), where EGRP II was being implemented.

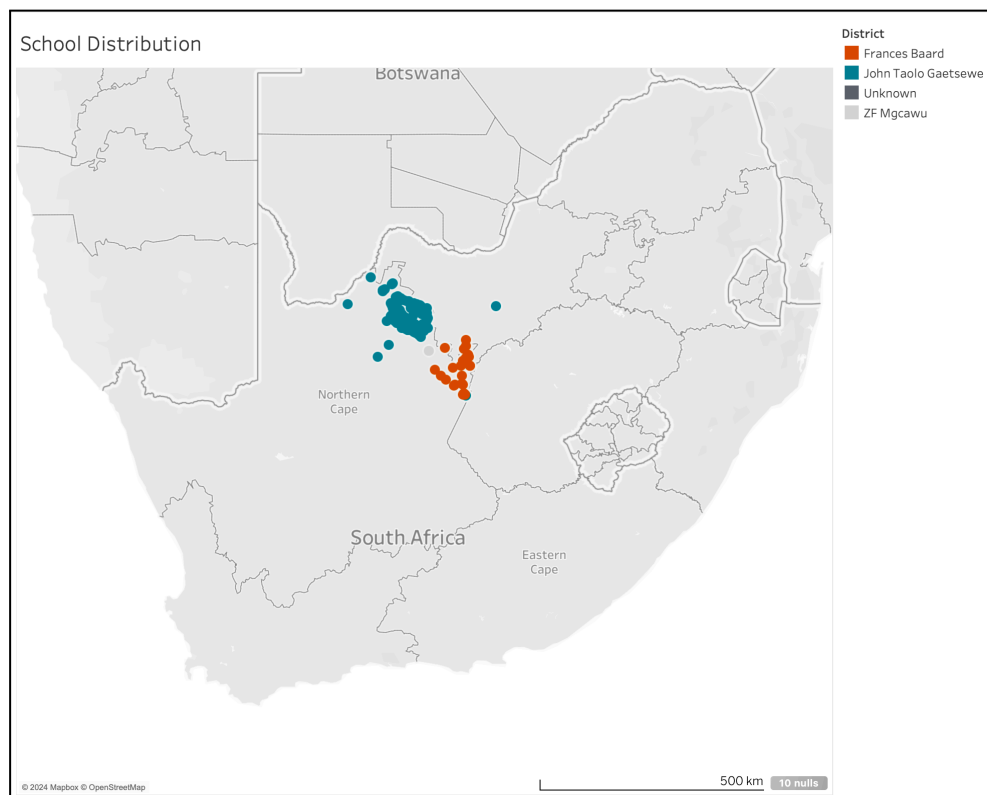


Figure 1 Geographic Distribution of EGRP II Schools

1.2 Evaluation framework and questions

The table below outlines the evaluation criteria for EGRP II, based on OECD DAC criteria and UNICEF's definition of impact. In accordance with the Terms of Reference, the evaluation focuses on effectiveness, impact, and sustainability, excluding relevance and coherence.

Table 1: Evaluation Criteria

Evaluation Criteria	Application to the EGRP II evaluation
Effectiveness	The extent to which EGRP II achieved its intended results and contributed to expected outcomes. This includes analysing results and identifying differences between monograde and multigrade classrooms, as well as understanding the factors that enabled or hindered achievement.
Impact	What was the impact of the coaching intervention on learning outcomes in Setswana HL and EFAL in both multigrade and monograde classrooms after two years of implementation?
Sustainability	To what extent did EGRP II foster national ownership of its methods and results, and how effectively did partnerships with stakeholders strengthen national capacities to ensure sustainability of efforts and outcomes?

Annexure 1 of this report contains a detailed evaluation matrix with indicators, data sources, and data collection methods.

1.3 The intended use and users

The evaluation serves multiple stakeholders across national, provincial, and implementing partner levels. The table below identifies the primary users, outlining their interests and their involvement in the evaluation process.

Table 2: Intended users of the evaluation

Users	Interest in the evaluation	Involvement in the evaluation
District and Subject Advisors	- Adopting/adapting strengthened EGRA and EGMA assessment tools and training approaches - Good practices and lessons learned for improving reading outcomes, teacher and school support	- Participate as Interview and KII respondents - Facilitate access to schools for all evaluation data collections activities - Provide administrative data for planning and evaluation
Oxford University Press South Africa	- Project assessment and organisational learning. - Recommendations for further activities and participation.	- Provide implementation data and work plan - Facilitate access to coaches, sample LTSM, and the EGRP Platform - Participate as interview respondents
Northern Cape Provincial Education Department of	Project assessment and organisational learning for future programming	- Participate as Interview and Key Informant Interviews (KII) respondents - Facilitate access to schools for all evaluation data collection activities

Users	Interest in the evaluation	Involvement in the evaluation
Education National DBE	• Good practices, and lessons learned • Recommendations for improving reading outcomes, programme sustainability, teacher and school support. • Use of system-level data to identify gaps and inform solutions	• Provide administrative data for planning and evaluation • Member of the EGRP II Project Management Team
National DBE		• Convene the EGRP II Evaluation Project Management Team. • Coordinate and provide access to system-level data for use as a proxy baseline • Technical evaluation management - signing off on tools, methodologies, analysis
UNICEF		• Hiring agency, evaluation management, contact point for evaluation team • Provide feedback on deliverables and signing off • Participate in interviews and KIIs • Member of the EGRP II Project Management Team (PMT)
ZENEX Foundation		• Participate in interviews and KIIs • Member of the EGRP II Project Management Team
Bill and Melinda Gates Foundation		• Funding agency • Provide feedback on deliverables

1.4 Evaluability Assessment

An evaluability assessment was conducted early in the programme implementation (mid-2024). This timing was consistent with the UNICEF Evaluability Assessment guidance. The summary results of the evaluability assessment are reflected below, showing overall positive evaluability, except for the lack of a quantitative baseline. This limitation is discussed further in the methodology section below.

	Key Question	Status	Comment
Relevance	Is the results framework of the programme coherently articulated and aligned with the country's context and national priorities?		The results framework is mainly coherent and aligned with national priorities. The programme has continuity and alignment with previous research in early-grade reading.
	Does the programme have a clear theory of change/logic model? Does it address the problems identified?		The program has a clear ToC but the sustainability element has not been fully reflected in the ToC.
Adequacy and validity of the indicators, tools and systems for monitoring, measuring and verifying results	Do the outputs, outcomes, and overall goal follow the result chain logic?		Planned activities, intended effects, and ultimate objectives are logically linked and form a coherent chain of cause and effect (though not always linearly).
	Is the results chain coherent, logical, and articulated?		More clarity needed for (i) specific activities for M&E support to provinces and district - activity plan, specifying activities, timeline and roles (ii) sustainability plans need more articulation.
	Are the results clear and realistic? Are they measurable (quantitatively or qualitatively)?		Valid methods exist for assessing inputs, outputs, and outcomes, both qualitatively and quantitatively. However, Programme duration is too short to realise impact. Aspects with SA roles are likely to not be feasible given their shortage.
	Does the sector programme have a monitoring system to gather and systematise the information with defined responsibilities, sources and periodicity?		There is a structured system for tracking the programme's progress and performance, including data capturing tools, defined roles for monitoring and reporting, and sources. Gaps were observed with the structure of the data which makes it difficult, though possible, to track some metrics consistently.
	Are baselines in place for the indicators part of the results framework?		The initial analysis of 2023 SBA data by the DBE has, in our view, shown that this data is not appropriate for reliably measuring intervention impact because (1) lack of correlation with standardised test results on previous studies and, (2) non-comparability of intervention-school SBAs over time due to school level decisions, and teacher training and coaching aimed at improving school-based assessment practices.
KEY YES TO SOME EXTENT NO EASY TO REMEDY HARD TO REMEDY VERY HARD TO REMEDY			

1.5 Structure of the evaluation report

The report is organised as follows:

- **Section 2** presents a brief overview of the methodology, with detailed information provided in the annexes.
- **Section 3** details the background, including the policy and country context, and describes the programme.
- **Section 4** presents the main findings across six key areas: programme effectiveness and teacher classroom practices, learner outcomes, sustainability outcomes, evaluability assessment outcomes, and school-based assessment (SBA) analysis outcomes.
- **Section 5** summarises the main conclusions.
- **Section 6** outlines lessons learnt and recommendations.

2. EVALUATION DESIGN AND METHODOLOGY

2.1 Evaluation design, methodology, and process

The evaluation is **utilisation-focused**, ensuring that findings inform decision-making and enhance current and future programming. It adopts a forward-looking approach with strong emphasis on learning and adaptation. It is also **theory-based**, using the ToC and programme objectives as the foundation for assessing key programme components.

- **Co-creation and co-ownership:** The evaluation team worked closely with the EGRP II Project Management Team (PMT) throughout the evaluation, adopting a collaborative approach to stakeholder engagement, evaluation design, and implementation. This included clarification workshops, data interpretation workshops, ongoing dialogue on evaluation methods and activities, and participative problem solving to ensure meaningful participation of key stakeholders and quality assurance of evaluation design and outputs.
- **Developmental evaluation:** The evaluation team shared insights throughout the programme, enabling implementers to adapt and improve implementation whilst maintaining focus on agreed outcome variables and core intervention activities.

Summary of the evaluation methodology

The evaluation was designed as a randomised control trial (RCT) using a participatory, mixed-methods approach. Both primary and secondary data were employed to address the evaluation questions. The evaluation methodology consisted of core methods as reflected in Table 3 below. These include

- Secondary data reviews/analysis
- Case studies

- Observations
- Learner assessments
- Surveys
- Key informant Interviews
- Focused Group Discussions

The methodology was implemented in three stages.

1. First, the evaluability assessment was carried out to determine if the project was evaluable. This also reviewed SBA data.
2. Second, case studies were conducted in two rounds: an initial round in 2024, followed by a second round in 2025 in the same schools. Two rounds of teacher and coach training observations were also conducted.
3. Finally, the endline stage comprised learner assessments and teacher surveys conducted in October 2025.

Although the ToR intended to use SBA system data as a proxy baseline, the evaluability assessment found this data to be unreliable for comparisons between schools. The assessment tested the extent to which SBA data in the North West province was correlated with the independent learner assessment results of the EGRP I evaluation in 2024 in the same schools. While SBA data captured some expected performance differences between schools, such as Quintile 4-5 schools outperforming Quintile 1-3 and lower scores in multigrade classrooms, it did not provide a reliable picture of actual learner performance ranges within or between schools. Inflated average scores (64-77%) and the absence of correlation between SBA and EGRA scores for the same schools indicate that SBAs lack the objectivity of independent standardised assessments. A detailed methodology, analysis and findings of the assessment of the school-based data is presented in Annexure 2.

Consequently, the EGRP II evaluation lacks a baseline and cannot quantitatively determine the impact of the overall EGRP II intervention (base programme and coaching programme) in relation to no support. It could only assess the differential impact of coaching compared to the base programme. This means that any positive effects of the base programme on its own (by providing appropriate LTSM and teacher training without coaching) could not be measured or reported on.

Table 3 summarises the methods used to assess learner effects, classroom and coaching practice outcomes, evaluability, sustainability, and implementation fidelity.

Table 3: Evaluation Methods

Evaluation Criteria	Evaluation Sub-Question	Case Study	Observation	Doc Review	Surveys	Learner Assessment	KII
Intermediate outcomes and impact – has the programme made a difference							
Impact	What is the overall impact of the coaching interventions on learning outcomes for Setswana and EFAL at the end of the 2 years?						

Impact	What is the impact of the coaching intervention on learning outcomes for both Setswana HL and EFAL in multigrade and monograde classrooms respectively at the end of 2 years of implementation.						
Impact	What are the mechanisms that made the second Early Grade Reading Programme successful?						
Impact	To what extent are classroom practices reflecting core methodologies conveyed by training and coaching?						
Impact	To what extent are coaching practices reflecting core methodologies for monograde and multigrade classrooms?						
Implementation Fidelity/Effectiveness - was the programme implemented as intended							
Effectiveness	How were the core methodologies trained, coached & implemented similarly and differently in monograde and multigrade schools by the implementing partner.						
Effectiveness	Are lesson plans and other LTSM delivered and being used by teachers?						
Effectiveness	What dosage of termly training have teachers & coaches received?						
Effectiveness	To what extent are classroom practices reflecting core methodologies conveyed by training and coaching?						
Effectiveness	To what extent are coaching practices reflecting core methodologies for monograde and multigrade classrooms?						
Sustainability - to what extent are EGRP II capacity development activities sustainable							
Sustainability	How can EGRP II be institutionalised at school, district and provincially?						
Sustainability	How successful has the capacitation of Subject Advisors been to support embedding of the intervention at district and provincial level?						
Sustainability	How successful and sustainable has the M&E capacity building for Subject Advisors been?						
Sustainability	How well have Professional Learning Communities PLCs been functioning and why?						
Evaluability - to what extent is school-based data useable to provide a baseline for EGRP II							
Evaluability	How reliably does SBA data reflect learner outcomes that can be compared across schools						

2.2 Instrument design, data collection and sample

School and Learner Sample: School sampling was carried out by the DBE using randomisation from qualifying quintile 1-3 schools. A sample of 20 randomly selected learners was identified in each school.

Table 4 DBE School sampling frame

Intervention Group	Number of schools		
	Monograde	Multigrade	Total
Base Programme (LTSM and training only)	42	20	62
Coaching (LTSM, Training and Coaching)	37	17	54
All	79	37	116

EGRA learner Assessment tools: The EGRP II endline assessment battery measured reading proficiency in Home Language (Setswana) and English First Additional Language (EFAL) across Grades 2 and 3. The instruments were adapted from proven measures used in EGRS, Funda Uphumelele National Reading Study, and EGRP I, and piloted to ensure suitability for multigrade contexts. SII Africa's evaluation team collaborated with the DBE, ZENEX, and UNICEF in two workshops to finalise the learner assessment and context tools. A Setswana language expert reviewed the instrument to ensure contextual relevance to the Northern Cape variant. The expert's input on language nuances and variations was incorporated into the tools and further refined during pilot training. Detailed Learner Assessment tools are contained in Annexure 3.

Case Study: This assessed coaching and teaching practices through classroom observations, teacher and principal reflections, and focus group discussions with teachers and learners. Data collection tools were either designed by the evaluation team or adapted from previous studies (EGRP I, Reading Support Program, EGRS) and signed off by the PMT during a case study design workshop. Ten schools were purposively selected from 60 intervention schools (4 monograde, 6 multigrade) based on demonstrated programme adoption capacity. The evaluation team engaged with the implementing partner, coaches, subject advisors, and district officials to collaboratively identify schools showing strong uptake to document successful implementation and observe teaching methodologies consistently across two consecutive years, which would not be feasible in schools struggling with basic functionality or programme adoption. Two rounds of data collection in the same schools were completed in 2024 and 2025. Detailed Case Study Methodology is in Annexure 4

Table 5: Case Study sample

Intervention Group	Number of schools		
	Monograde	Multigrade	Total
Base Programme (LTSM and training only)	2	3	5
Coaching (LTSM, Training and Coaching)	2	3	5
All	4	6	10

Surveys: These surveys were administered to teachers, departmental heads, and school principals during the 2025 endline data collection. Survey data established programme and contextual information, triangulated evaluation findings on implementation fidelity and teacher experiences, and verified implementation data and emerging case study themes. In each school, at least one available programme teacher was selected for interview, with 18 schools where two teachers were interviewed, totalling 204 teachers. With approximately 600 programme teachers, this sample achieved a 92.1% confidence level with a 5% margin of error, allowing some degree of generalisation of findings.

Training observation and site visits: One coach training and one teacher training observation were conducted in each implementation year (2024 and 2025), with two trained observers attending entire sessions to enable collaborative reflection and validation. Teacher training observations covered both EFAL and HL instruction to capture the full scope of literacy training. A site visit shortly after programme inception included funders and government officials. The evaluation team presented the methodology, responded to questions, and established relationships with key stakeholders (coaches, teachers, implementing partner staff, district officials) to facilitate subsequent observations, interviews, and other evaluation activities. Tools for observing training are in Annexure 5.

Key Informant interviews: Key informant interviews gathered nuanced information on various evaluation dimensions, particularly sustainability, from government officials, project management team members, and implementing partner staff. Interview data served as a triangulation source for programme delivery aspects. Respondents were purposively sampled based on their specialised knowledge and organisational roles. Most interviews were conducted remotely.

Document reviews: Multiple secondary data sources were consulted across different phases of the evaluation to understand the programme context, implementation environment, and implementation plans as well as to verify findings from other data sources. In some cases, primary data sources were used to verify secondary data. These sources informed assessments of sustainability, evaluability, and implementation fidelity, among other evaluation components. The following secondary data sources were consulted:

Table 6 Secondary data sources consulted

Phase	Documents & Datasets to be reviewed
Instrument Development Data Collection Phases Evaluability Assessment Sustainability Assessment	• Past EGRS and EGRP I Evaluation Reports • EGRP II Concept Note • EGRP II Sustainability Plans • UNICEF Guidance Note for Conducting Evaluability Assessments • EGRP ToC, OUP Measurement Framework, EGRP II Implementation Plans • OUP Tools and Data - Training Attendance Registers, Teacher Training Evaluation Forms, LTSM Registers, Teacher pre and post test scores, Coach Training Evaluation Forms, Coach Monitoring Data, Coach Visits Plans, OUP Coaching Teachers Surveys, OUP Annual Teacher Surveys, Training Agendas, OUP Coaching Plans. • EGRP I Case Study Guide and Tools, • EGRP I, EGRS EGRA Tools, • CAPS, APPs, • Lesson Plans for EFAL & HL
Data Analysis & Reporting	• DBE School Based Data, • EGRP I and EGRS technical reports, • Tablet Data • OUP Monitoring Data • UNICEF Evaluation Policy • OUP Monitoring and Evaluation Report: June 2025-December 2025

2.3 Data Analysis

2.3.1 Quantitative: Learner assessment, implementation data and surveys

The learner data was analysed as follows:

- **Distribution analysis:** this was used to examine patterns, identify extreme outliers, and understand the distribution of outcomes for each subtask across treatment groups and school types.
- **Means testing between treatment and school type:** means testing using t-tests was also conducted to check which treatment and school type was statistically significantly different. Standard errors were clustered at the school level.
- **Sensitivity analysis:** this assessed result robustness by comparing full and trimmed datasets. The top two and bottom two performing schools were excluded separately, and means testing compared results from the full sample against each trimmed specification. Schools were ranked using Grade 2 Home Language Rapid Object Naming scores (selected for its similarity across treatment and school types) and Grade 3 EGMA mathematics scores.
- **Linear regression, multiple regression with covariates:** regression analysis was performed with separate models for each subtask. Five regression specifications with progressively added covariates were estimated to test the robustness of results and ensure coaching effects remained statistically significant across model specifications.

- **Principal Component Analysis:** composite scores were calculated for both home language and EFAL using Principal Component Analysis (PCA). The same five regression specifications were applied to these composite scores, yielding consistent results.
- **Quantile regression:** finally quantile regressions were run to identify which learner percentile group/s was/were most helped by coaching i.e. low performers, middle performers or top performers.

Survey and implementation data were analysed using:

- **Descriptive statistics:** including frequency distributions, percentages, and cross-tabulations by school type and treatment group to examine patterns in teacher practices, resource use, coaching quality, and implementation fidelity.

2.3.2 Qualitative: Case studies, observations, and KIIs

- **Content analysis:** this was employed to systematically identify, code, and categorise themes emerging from qualitative data sources, including interview transcripts, classroom observation notes, case study monographs, and open-ended survey responses. For case studies, content analysis was used with limited coding to maintain and foreground the richness of experiences as represented by the text and observations and not take a reductionist approach
- **Interpretative Phenomenological Analysis (IPA):** this was used to explore the lived experiences of teachers and coaches, focusing on how they made sense of the EGRP II intervention within their specific monograde and multigrade contexts. This approach provided deeper insight into individual perspectives on coaching quality, resource utilisation, professional development, and the meaning teachers attributed to changes in their practice.
- **Thematic analysis:** this was used to identify, analyse, and report patterns across the qualitative dataset. Themes were generated inductively from the data and deductively through alignment with the evaluation framework, allowing for both emergent insights and structured assessment of implementation components. This approach facilitated comparison of experiences across monograde and multigrade contexts and between treatment groups.

2.4 Limitations and challenges

Lack of a baseline: The evaluation does not include baseline data, as school-based assessments, that were expected to serve as a proxy baseline, were found to be unreliable in comparing learner performance across schools. The evaluation was therefore not able to assess impact quantitatively with confidence regarding change over time. The only quantitative statements that the evaluation can make use cross-sectional data to describe the relative learner outcomes of the four sub-groups in the sample (monograde base, monograde coaching, multigrade base, multigrade coaching). It is not possible to assess the impact of the base

programme on its own compared to schools that received no intervention at all, since intervention schools had no 'pre-intervention' baseline and a 'no intervention' comparison group was not included in the study at the endline.

While there is no quantitative 'impact over time' data, the evaluation team made every effort to link the endline learner assessment findings to insights from the qualitative methodologies, and to the teacher and SMT surveys during the endline data collection phase, which enables some plausible causal interpretation of changes created by the intervention over time.

Implementation data gaps: Implementation data quality improved significantly under EGRP II compared to EGRP I. The implementing partner was responsive to requests for improvement, and initial data structure limitations were progressively resolved after a collaborative workshop in 2024. However, some implementation data gaps remained. These gaps limited the extent to which some aspects of the fidelity dimensions could be fully assessed. The evaluation team addressed this by triangulating implementation records with teacher surveys and case studies to verify and supplement the available data. However, given the small sample sizes for multigrade teachers, survey findings should be interpreted with caution, since small shifts in responses can have disproportionately large effects on percentages.

Limited intervention timeframe: The programme duration was compressed, with some multigrade teachers receiving less than two years of full intervention due to delayed multigrade-specific materials design and production, though they received monograde specific materials and training in the first six months of the intervention period. This timeframe posed challenges for observing meaningful effects, as research indicates instructional coaching interventions typically require sustained engagement over multiple years to produce measurable shifts in practice and outcomes. The abbreviated timeline may have constrained the intervention's ability to demonstrate its full potential impact.

2.5 Ethical Approach and Safeguarding

The evaluation followed UNEG Norms and Standards and UNEG Ethical Guidelines for Evaluation, UNICEF's Procedure on Ethical Standards in Research and Evaluation, UNICEF's Ethical Research Involving Children, and UNICEF's Guidance on Gender Integration in Evaluation. It was guided by ethical principles aimed at minimising harm whilst maximising benefits through confidentiality, informed consent, data protection, participant safety, and meaningful, inclusive participation. The evaluation adhered to "Do No Harm" and "Leave No One Behind" principles and integrated human rights, gender equality, and equity considerations throughout. All evaluators and data collectors confirmed their commitment to these policies and received induction training.

The DBE provided clearance for this study as data collection was approved by national and provincial government authorities and conducted in public schools.

3. BACKGROUND AND CONTEXT

3.1 Policy and programming context

South Africa aims to ensure that all 10-year-olds can read with comprehension by 2030, but current outcomes fall far short of that goal. The 2022 PIRLS study found that 81% of Grade 4 learners cannot read with comprehension, with the problem most severe in poorer, non-fee-paying schools (quintiles 1 to 3), where most of South Africa's 13 million public school children are enrolled. The COVID-19 pandemic worsened these challenges, with Foundation Phase learners losing up to 65% of learning time due to rotational attendance and absenteeism. Under-resourced schools were hardest hit, and these learning backlogs will continue to affect learners as they progress through the schooling system.

Since 2015, the DBE and its partners have conducted RCTs to test teacher support models for early-grade reading. EGRS I (2015 to 2017) focused on Setswana in North West Province, followed by the Reading Support Programme covering Setswana and English. EGRS II (2017 to 2019) focused on English in Mpumalanga. The Early Grade Reading Programme (2021-2023) in the North West compared internal department head coaching with external coaches, while providing training and materials to all schools.

These studies demonstrated that training, provision of LTSMs, and coaching significantly improved reading outcomes, with external professional coaches producing the strongest results. Learners in coached schools showed 40% better outcomes within two years (Taylor et al., 2018). However, external coaching is expensive, raising sustainability concerns for national rollout. Virtual coaching was tested as a cheaper alternative but proved less effective due to limited classroom observation and feedback (Fleisch, 2018). EGRP I then tested whether department heads could coach effectively, but could not demonstrate comparable results to external coaching.

EGRP I confirmed the efficacy of external coaching, with Grade 2 Setswana recording gains equivalent to 10.2 months of learning progress. Even more striking were the gains among mid-range performers between the 20th and 60th percentile, who achieved up to 18 months of learning progress. External coaching effects on Grade 2 EFAL were even stronger than those recorded for HL. These gains were however not replicated in Grade 3, where results were mostly insignificant with a small gain in EFAL, a pattern attributed to older cohorts having experienced cumulative learning losses during their foundational years due to COVID-19 disruptions in 2020 and 2021.

South Africa's monograde context

All the preceding interventions were designed for monograde classrooms, reflecting international practice in which "all teacher education, textbooks, sources, policies, and curricula were established according to monograde principles" (Joubert 2010). Administrative data is also structured based on the assumed conditions of a monograde setting. However, multigrade teaching remains significant in South Africa. Despite school consolidation policies since a 2015

Multi-grade Strategy (when 13% of learners were in multigrade schools), approximately 16% of learners (2 million) remain in officially designated multigrade classrooms, with higher proportions in rural provinces.

Adapting literacy interventions to multigrade settings matters not only for officially multigrade schools but for all public schools, especially those in quintiles 1 to 3. Previous EGRS studies indicate that Foundation Phase classrooms often contain learners with significantly diverse literacy levels. This effectively transforms them into disguised multigrade classes, but without the advantage of structured, differentiated teaching methods or specific training and resourcing for teachers and learners. Previous studies have also revealed that learning outcomes are typically bimodal rather than normally distributed, with skills falling either significantly below or at/above expected grade levels. Most South African teachers, therefore, face functionally multigrade classes while using materials designed for monograde settings. Therefore, insights from EGRP II are crucial for developing differentiated teaching approaches for both monograde and multigrade classrooms.

3.2 EGRP II Programme description and objectives

The EGRP II programme builds on EGRP I and other studies such as the EGRS I and II by applying the same 'triple cocktail' approach of external coaches, teacher training, and LTSM provision, but adapts it for multigrade classrooms. The key difference between EGRP II and EGRP I lies in the coaching model: while EGRP I tested whether departmental heads (existing school staff) could effectively coach teachers, EGRP II returns to using external literacy coaches only (as previously used in EGRS iterations). EGRP II also introduces structured lesson plans specifically designed for multigrade settings and provides system-level support to district and provincial education departments. The programme components of EGRP II are outlined in Table 7. A

Table 7 EGRP II Programme Components

Activities	Base Programme Monograde & Multigrade	External coaching + Base Programme Monograde & Multigrade
Provision of lesson plans	2025: Tablet-based HL and EFAL scripted plans	
Provision of LTSM	Paper-based Lesson Plans and Trackers HL and EFAL Big Books HL and EFAL Sight Words Flash Cards HL and EFAL Conversation and Handwriting Posters HL and EFAL Phonic Friezes Ulwazi Lwethu Anthologies Occupational Tasks/ Worksheets Methodology book Teacher's Manual	
Training of teachers	Teachers receive training at the beginning of each term throughout the intervention	
Coaching of teachers	None	External Coach visits each teacher in the classroom from Year 1 to Year 2
Virtual coaching support	None	Ongoing remote support by external coach from year 1 to Year 2
Establishing PLCs	None	Yes
System level support and capacity development	School, district and provincial levels	

3.3 Human Rights, gender and equity issues in relation to early grade reading outcomes

Gender distribution in South African reading studies is generally balanced, with a slight majority of girls, a balance maintained in this evaluation's sampling. Previous studies, including EGRP I, have consistently shown girls outperforming boys on reading subtests, such as oral reading fluency. The 2011 PIRLS study further confirmed this trend, finding girls outperformed boys in 40 of 45 participating countries. Notably, South Africa had the third-largest gender gap in favour of girls, equivalent to one year of learning by Grade 4 (Mullis et al., 2012, as cited in van Broekhuizen & Spaull, 2017). Furthermore, Mohohlwane (2016) reported that girls in North West Province enter Grade 1 with statistically significant advantages over boys in Setswana. This study includes an analysis of learner outcomes by gender to reflect this reality observed in most South African classrooms.

All data collection tools were designed with a gender and age lens, enabling findings and observations to be disaggregated by gender and age and any emerging trends to be identified and analysed, thereby supporting a more equitable interpretation of results across the evaluation.

The schools sampled for this study fall into lower socioeconomic quintiles (1-3), are non-fee-paying, and face significant resource constraints, making them particularly vulnerable to the severe disruptions caused by the COVID-19 pandemic. From March 2020 to June 2021, the majority of primary learners experienced learning losses equivalent to 60% to a full year, with average attendance dropping from 97.9% in 2018 to 94.8% by April 2021 (Shepherd & Mohohlwane, 2021). To ensure an equitable understanding of these impacts, the evaluation employed regression modelling to assess how children's contextual factors contributed to differential effects, enabling a more nuanced interpretation of outcomes across diverse learner circumstances and ensuring that the needs of the most vulnerable are not obscured in the aggregate findings.

All protocols safeguarding children's rights, including assent, anonymity, and confidentiality, of all respondents were strictly followed throughout the study.

3.4 Theory of Change

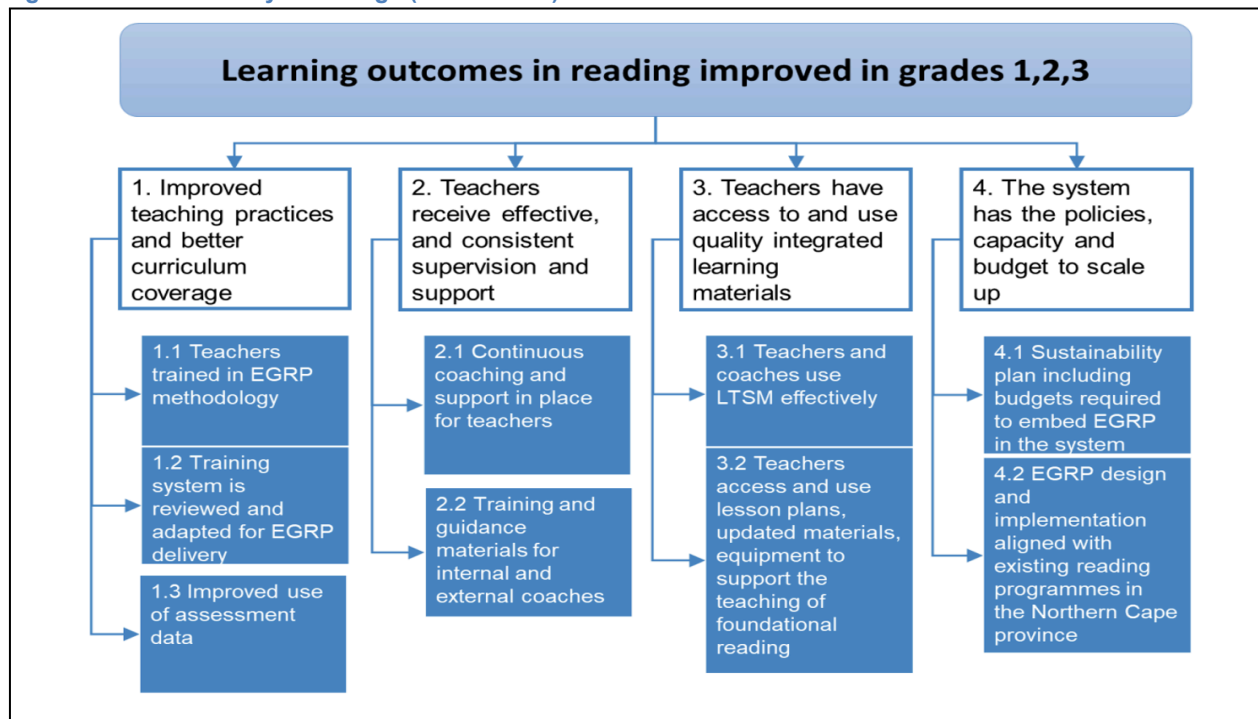
The Theory of Change (ToC) developed during the design phase envisages that improved reading outcomes for Foundation Phase learners can be achieved through four interconnected pathways. These pathways aim to enhance teachers' instructional skills, provide consistent support, ensure access to quality resources, and establish a sustainable framework for programme embedding and scaling.

- **Pathway 1: Improved teaching practices & curriculum coverage:** Focuses on enhancing teachers' reading instructional skills through policy-aligned training, continuous review, and encouraging the use of assessment data to address individual learner needs and improve outcomes.

- **Pathway 2: Consistent supervision and support for teachers:** Establishes a structured coaching system with regular supervision and targeted coach training. This ensures reinforcement of effective practices and alignment with programme goals, helping teachers refine skills and improve learner outcomes.
- **Pathway 3: Access to and use of quality integrated learning materials:** Emphasises providing quality reading resources, including lesson plans and updated materials. Training teachers to use these LTSMs effectively is expected to enhance classroom integration, teaching quality, and learner reading confidence.
- **Pathway 4: System capacity and budget for scaling up EGRP:** Envisages sustained programme effects by exploring pathways for embedding the intervention within the provincial education system, potentially through sustainability planning, ring fenced budgets, and alignment with existing provincial reading initiatives, ensuring the programme's long term operation and enduring impact.

The ToC shown in Figure 2 was developed during the design stage. A more detailed articulation of the ToC is provided in Annexure 6.

Figure 2 EGRP II Theory of Change (abbreviated)



3.5 Key stakeholders

The successful implementation of EGRP II required coordinated effort across multiple organisations and government departments, each with distinct roles and responsibilities. Table 8 outlines the key stakeholders involved in the programme, from funding and policy oversight to

implementation and evaluation. This multi-level governance structure ensured technical quality, effective coordination, and accountability throughout the programme lifecycle.

Table 8: Programme Stakeholders and their Roles

Organisation/Department	Role and oversight
UNICEF	Evaluation hiring agency, evaluation management, contact point for the evaluation team, and interviews of project staff. Member of the EGRP II Project Management Team.
Bill and Melinda Gates Foundation	Funder
DBE	DBE is the designated owner of the programme and undertakes due diligence and approvals of the programme design and various implementation modalities. DBE research and curriculum directorates are responsible for overseeing continuous operational and technical decision making. The Directorate provides policy direction and potential budgets to leverage the scale-up for greater system strengthening and scaled implementation. It oversees, considers, and analyses evidence and insights generated by the M&E function to provide timely feedback that informs improved implementation.
Project Management Team	The PMT comprises representatives from the DBE research team, curriculum and reading directorates, UNICEF and the ZENEX Foundation to function as the technical oversight forum to streamline and inform the implementation of the programme. The forum, chaired by the DBE Curriculum director, met quarterly to collectively oversee implementation and plan quality assurance.
Provincial Education Department (PED)	PED ensured timely and sustained coordination with DBE, district stakeholders and implementing partners around effective implementation. It supported districts with the necessary technical capacities to oversee implementation at the school level.
Districts	Districts oversaw direct implementation at the school level, ensuring coordination with target schools for effective delivery. They coordinated with the province to maintain oversight of teacher training and support activities delivered in schools.
Oxford University Press South Africa	Programme implementation agency
Social Impact Insights Africa	EGRP II Evaluation Agency

4. FINDINGS

This section presents findings on implementation fidelity, learner outcomes, teacher classroom practices, and sustainability. Sustainability findings are presented partially based on year one work (2024), as the current workstream remains in progress with key informant interviews and document reviews still ongoing. Complete sustainability findings will be included before the final report is completed.

4.2 Implementation fidelity

Implementation fidelity was examined across three key dimensions of the EGRP II Theory of Change: teacher and coach training, coaching support to teachers (which also included establishing/conducting PLCs and cluster workshops), and the distribution, availability, and use of LTSM. The assessment focused primarily on dosage and quality, with adherence and programme differentiation assessed where data permitted. Examining implementation fidelity is essential not only for understanding how and why an intervention produces its effects and how outcomes can be further improved (Dusenbury et al., 2003), but also for supporting the attribution of results to the intervention itself (Dane & Schneider, 1998).

Implementation contextual challenges: It is important to note that several contextual factors affected implementation fidelity during the early stages of the programme. These are documented here to provide an accurate account of the conditions under which implementation took place and to inform future planning. The finalisation of the contract between UNICEF and OUPSA was delayed, which had a knock-on effect on a number of key programme activities, including the delivery of Term 1 training, the provision of Term 1 LTSM, the design and distribution of LTSM content for multigrade schools, and the timely provision of rental vehicles to coaches, an essential operational resource.

Implementation activities were further affected by environmental factors outside the control of the implementation team. A labour dispute regarding compulsory teacher work during school holidays required training sessions to be rescheduled from the planned pre-term schedule to the first week after schools reopened, potentially affecting instructional time and teacher availability. Additionally, the "Work-to-Rule"¹ directive issued by teachers' unions in January and February 2025 caused delays in the distribution of tablets to teachers in the Francis Baard district.

4.1.1 Teacher Training

Teacher training was a core intervention component delivered to all programme teachers across both base and coaching cohorts. Training sessions were initially scheduled before each school term to minimise disruption to instructional time. However, due to a labour issue related to

¹ In late January 2025, the South African Democratic Teachers' Union (SADTU) and the Suid-Afrikaanse Onderwysunie (SAOU) in the Northern Cape, initiated a 'Work-to-Rule' action. This industrial action restricted teachers to working only their contracted hours, specifically the seven compulsory academic hours of the school day.

compulsory work for teachers during school holidays, training was shifted to the first week of each term. Given the vast distances in the Northern Cape, multi-day training sessions were conducted residentially, making them significant logistical undertakings for the implementing partner (responsible for training delivery and materials) and provincial education authorities (responsible for accommodation and transport). Training was facilitated by the implementation partner's instructional coaches. Important to note is that the training and LTSM support reached teachers beyond the RCT sample, as some schools were added by the PED and Districts. The findings reported here focus exclusively on RCT evaluation sample schools to maintain evaluation consistency, while OUPSA's training attendance data includes all schools that participated.

Coverage and dosage: Monitoring data for 2024 shows strong training reach, with 96% of programme teachers participating in at least some training sessions. However, the data did not allow for a breakdown of full versus partial attendance, limiting the ability to establish dosage per teacher. This limitation was also experienced in EGRP I and other studies, highlighting the need for system-level support to equip implementation partners with robust monitoring systems.

The 2025 training attendance data showed improved quality², enabling tracking of individual teacher attendance across terms as teachers could now be more reliably and uniquely identified. Training was delivered over one or two days per term, with Terms 1 and 3 consisting of two-day sessions and Terms 2 and 4 being one-day sessions. The implementing partner notes that the two-day residential training sessions resulted in better teacher attendance than one-day sessions, where teachers often arrived late or left early. The implementing partner's training data reports show that full attendance at all training days per term varied, ranging from 88% to 91% depending on term and format.³

Teacher survey data confirmed high overall participation in both years (95% in 2024, 100% in 2025), reflecting strong teacher commitment to the programme. However, there remains room for improvement in ensuring consistent attendance across all sessions, with teachers who reported full attendance (across all four terms) declining from 83% to 61% between 2024 and 2025 (Figure 3). A notable shift occurred between the two years, with multigrade teachers improving their full attendance rates from 2024 to 2025, while monograde teachers' full attendance declined. However, the small sample of multigrade teachers means findings should be interpreted with caution, as minor shifts in responses can disproportionately affect percentages. Attendance by coached teachers was more consistent (more full attendance) than among base-group teachers, except among the monograde teachers in 2024 where there was only a small difference.

² The evaluation team collaborated with the implementation partner through a data workshop in 2024 to strengthen data collection practices and improve the quality of implementation data.

³ The training attendance data could not be fully verified by the evaluators at the time of writing, since some data formatting issues remained in discussion with the implementing agency. This reinforces the importance of agreeing on standardised training record keeping formats for future programme iterations.

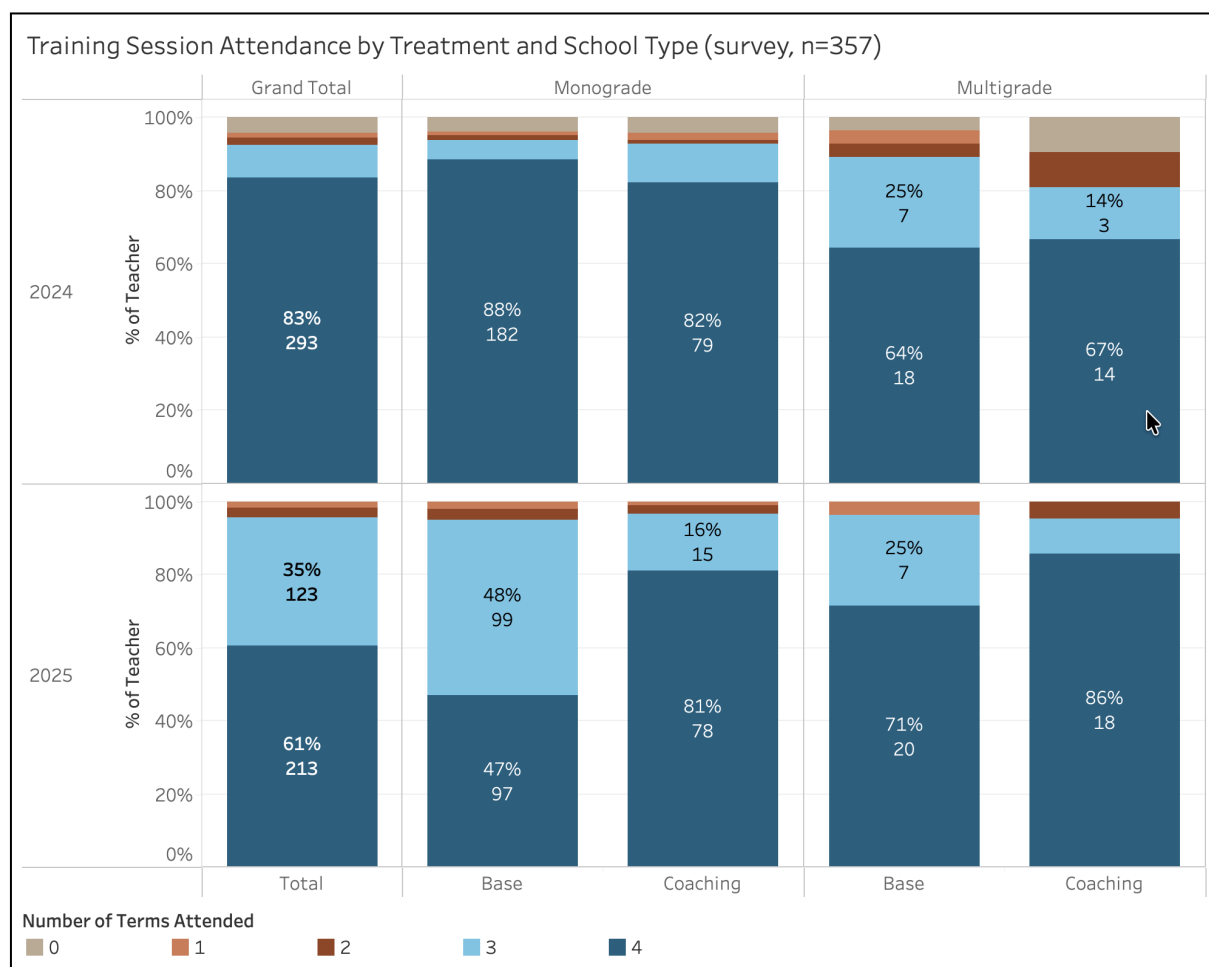


Figure 3: 2024 and 2025 teacher training attendance by Intervention and School Type

Adherence: Adherence to training content remained consistently high throughout the programme, demonstrating strong commitment from both trainers and participants. A summary of content covered is in Annexure 7. Trainers were well-prepared and actively encouraged teacher engagement throughout the observed sessions. All planned training sessions were held. However, implementation was affected by the late development of multigrade materials due to contracting delays. These materials were made available to teachers in the third term of 2024, potentially limiting the effectiveness of earlier training sessions for multigrade teachers who had to use monograde materials.

Teachers who missed training were expected to attend catch-up sessions conducted through cluster workshops, which served both coaching and non-coaching schools. Implementation data shows the programme's encouraging systematic efforts to support teachers who missed initial training. In Term 2, 37% (n=83) of teachers participated in catch-up sessions. The programme maintained these efforts throughout the year, reaching 17% (n=13) in Term 3 and a further 17% (n=8) in Term 4. Although catch-up training did not reach all intended teachers, the availability of catch-up training data represents a notable improvement over EGRP I, where such data was not collected.

Quality: Teacher training observations indicated strong fidelity to the training content covered. Survey results reflect very high teacher satisfaction with both the training content and delivery, particularly its ability to provide structure and support to multigrade teachers who had previously received little guidance. Some teachers found the training on grouping learners especially valuable, noting that it gave them both the permission and practical strategies to group learners by ability rather than solely by grade, offering a meaningful opportunity to address reading backlogs. Teachers noted:

*"In multigrade I didn't group learners by ability or move them between groups. Now my GGR (Group Guided Reading) groups are a mix of grades."
Setswana multigrade teacher 2, 2024*

"I didn't know that I should use a checklist, I use it to mark where learners are so that when they improve I can move them to different groups. Now I know it's okay for learners to read at their own level". Setswana multigrade teacher 1, 2025

A number of teachers reported an improved capacity for managing multigrade classrooms, specifically attributing it to the effective use of occupational tasks and differentiation strategies as covered in the training.

*"Before when we were at training I used to wonder how you can work with a small group while also not neglecting the rest of the class. I failed to do it before, but now I am excelling at it. I understand the usefulness of occupational tasks."
Setswana Grade 2 Teacher, 2024*

Case study observations saw effective multigrade teachers using these skills, with one teacher assigning Grade 3s to work in pairs on sentence construction, Grade 2s to practice handwriting, and Grade 1s to focus on letter formation while she worked with specific groups hearing them read.

4.1.2 Coach Training

Most EGRP II coaches are experienced educators and coaches who have worked on similar projects, including the predecessor of this project, the EGRP I programme.

Quality: Coaches' reflections on the 2024 training were mixed but largely positive: they found the OUP training comprehensive, well-structured, and memorable, valuing its focus on administrative systems, data capture, and contextualised coaching strategies. Coaches appreciated the training's responsiveness (Term 3 addressed handwriting and remediation on request of the coaches) and the practical utility of the GROW model for encouraging teacher self-reflection. One gap noted early in implementation was limited guidance on multigrade and mixed-ability classroom approaches, possibly reflecting some coaches' limited prior exposure to

these settings. However, this appeared to be an isolated early observation, as subsequent coach inputs to teachers demonstrated a strong depth of understanding of multigrade contexts.

4.1.3 LTSM

Availability and timely delivery

Implementation data for 2024 demonstrates strong LTSM coverage, with 98% of monograde teachers and 97% of multigrade teachers receiving materials. Implementing partner monitoring reports indicate that LTSMs for HL and EFAL were developed progressively throughout 2024. While Term 1 materials were finalised towards the middle of 2024, implementation delays meant teachers only began using them in Term 1 of 2025. Multigrade materials were delivered in Term 3 of 2024 due to delays in finalisation of the OUPSA contract, which affected the development of these materials.

The implementation partner maintained detailed data and source documents tracking the distribution of individual LTSM components to teachers. However, the format this data was in made it challenging for the evaluation team to verify if teachers received complete packs. This once again highlights an opportunity to strengthen monitoring systems and data management practices for implementing partners of education programmes.

Teacher survey data was used to close this gap. In the survey, teachers were given a full list of LTSM they should have received. Figure 4 confirms high coverage of distribution of HL LTSMs in 2024 though more sampled teachers in monograde schools received their full pack compared to their counterparts in multigrade schools in 2024. This discrepancy may be due to multigrade teachers reporting on whether they received the full multigrade-specific materials in 2024, which, as noted above, was not yet available. In terms 1 and 2, monograde materials were distributed to multigrade teachers for all grades they taught (e.g. a multigrade teacher covering Grades 1 and 2 would have received materials for both Grade 1 and 2), but it is possible that teachers did not refer to these in the survey.

It is important to note that LTSM distribution in 2025 consisted predominantly of multigrade materials, as most monograde teachers had already received their materials during 2024. Exceptions included top-up materials provided to newly appointed teachers, teachers who changed grades, and schools where additional classrooms were established due to enrollment growth. This context is essential for interpreting the 2025 survey findings, as the lower full pack receipt rates among base programme monograde teachers likely reflect this reduced distribution footprint rather than a systemic distribution failure. It is also possible that some teachers surveyed in 2025 reported on materials they had received in 2024 and were still using in 2025, explaining the high rates of coached monograde teachers reporting they received materials in 2025.

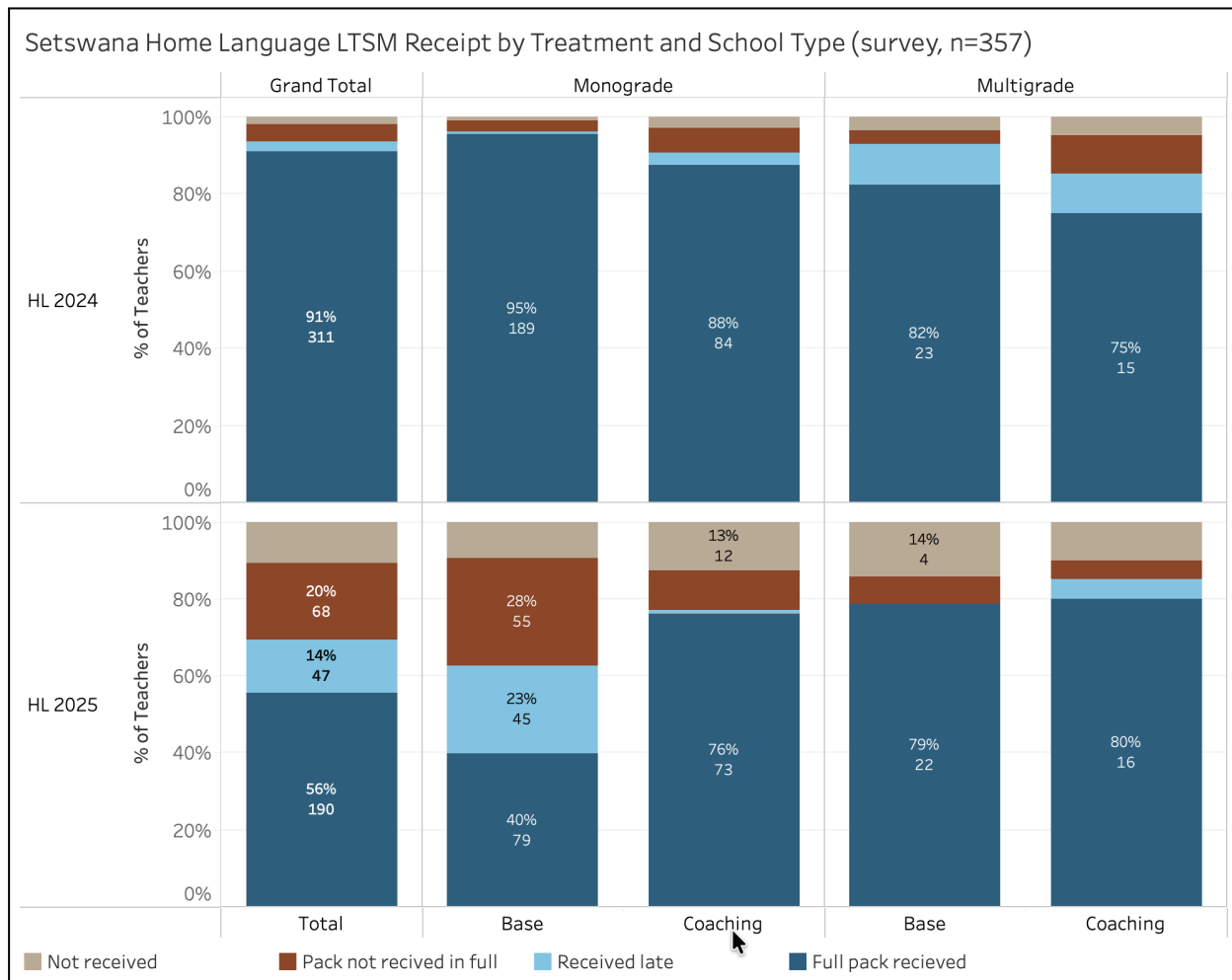


Figure 4: Receipt of EGRP HL LTSM by year, school type and intervention

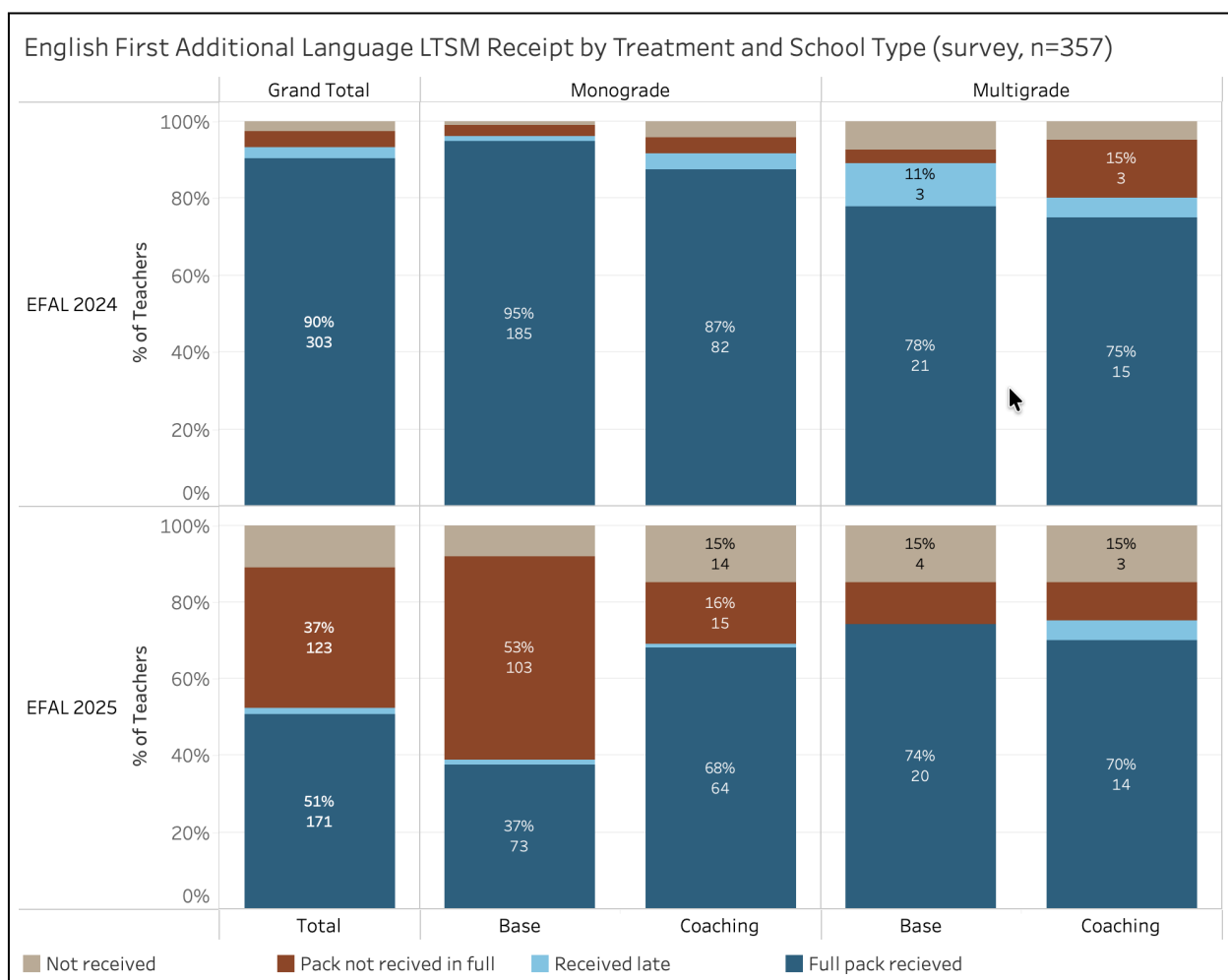


Figure 5: Receipt of EGRP HL LTSM by year, school type and intervention

Quality of LTSM

This element was assessed qualitatively and through surveys. Case study reflections are predominantly positive, with teachers and learners valuing the quality and relevance of the resources.

- **Support for multigrade teaching:** Coach shadowing found the materials helpful for managing different grade groups, particularly through the use of occupational tasks that kept one grade engaged while they taught another. Some teachers adapted the materials creatively, displaying resources for different grades side by side to create shared visual learning aids.
- **Better learner engagement:** Learners expressed a strong liking for the anthologies and the interactive nature of the Big Books during Shared Reading. Specific titles, such as *“Maatla a ga Neo”* and *“A Vet on the Farm”*, were cited as favourites by learners, who could vividly recall the storyline.
- **Transformation of classrooms:** The provision of posters, flashcards, and friezes has transformed classrooms into print-rich environments, which were previously bare according to teacher accounts. Teachers reported that the materials make their jobs

easier by providing clear sequencing for lesson content, enabling them to manage curriculum coverage better.

The positive sentiments around the quality of LTSM were very strong. Nonetheless, teachers expressed some concerns and requests for consideration in future programmes:

- **Language differences/errors:** From the outset, there was recognition that the variant of Setswana spoken in the Northern Cape differed from variants in other parts of the country, requiring contextualisation of LTSMs. Provincial authorities rightly advocated for this contextualisation and collaborated in reviewing the OUP materials during reference group sessions. However, Teachers raised concerns about language differences in some Setswana materials, citing the inclusion of Sesotho and Sepedi words. This appears to reflect the complexity of standardising materials across a linguistically diverse region, where the variant of Setswana referenced by education authorities during the review process had differences with the variants spoken in specific classroom contexts. While education authorities rightly hold responsibility for language standardisation, there may be value in supplementing the review process with input from teachers, helping to ensure that materials resonate with the linguistic realities of the classrooms they are designed for, without compromising standardisation principles.

Consequently, some teachers manually corrected words they perceived as errors in books, flashcards, and other materials because these contradicted what they taught. Teachers reported that this created confusion for learners.

- **Difficulty Level:** In some cases, teachers felt the provided readers were too difficult for struggling learners. Consequently, some teachers supplemented the EGRP II materials with easier readers from outside the project to support weaker learners. This practice is encouraging as it demonstrates teachers applying core methodologies to meet diverse learner needs.

Other teacher concerns related to resourcing and sustainability issues:

- **Durability and cost:** Teachers expressed concern about the lifespan of the posters and books. Since most schools lack ring-fenced budgets for foundation phase resources, many teachers reported spending their own money to laminate materials or buy printers to ensure the resources last.
- **Printing and reproduction:** Teachers noted that the digital resources (occupational tasks) are in colour, but schools often lack colour printers. Printing these resources in black and white renders them less effective or unusable, leading to additional administrative workloads as teachers try to recreate them with images or contrast that work better in black and white.

Teacher perceptions on multigrade LTSMs in their classrooms:

- **Multigrade versus monograde LTSM usefulness in multigrade classrooms:** Since multigrade teachers received their tailored materials only in Term 3 of 2024, they initially used monograde materials. Having experienced both, the survey asked them to compare their usefulness for multigrade contexts relative to monograde LTSM. Figure 6 shows

that coaching teachers were far more positive about multigrade LTSMs than base teachers, suggesting that coaching support may enhance teachers' ability to recognise and utilise these specialised materials effectively.

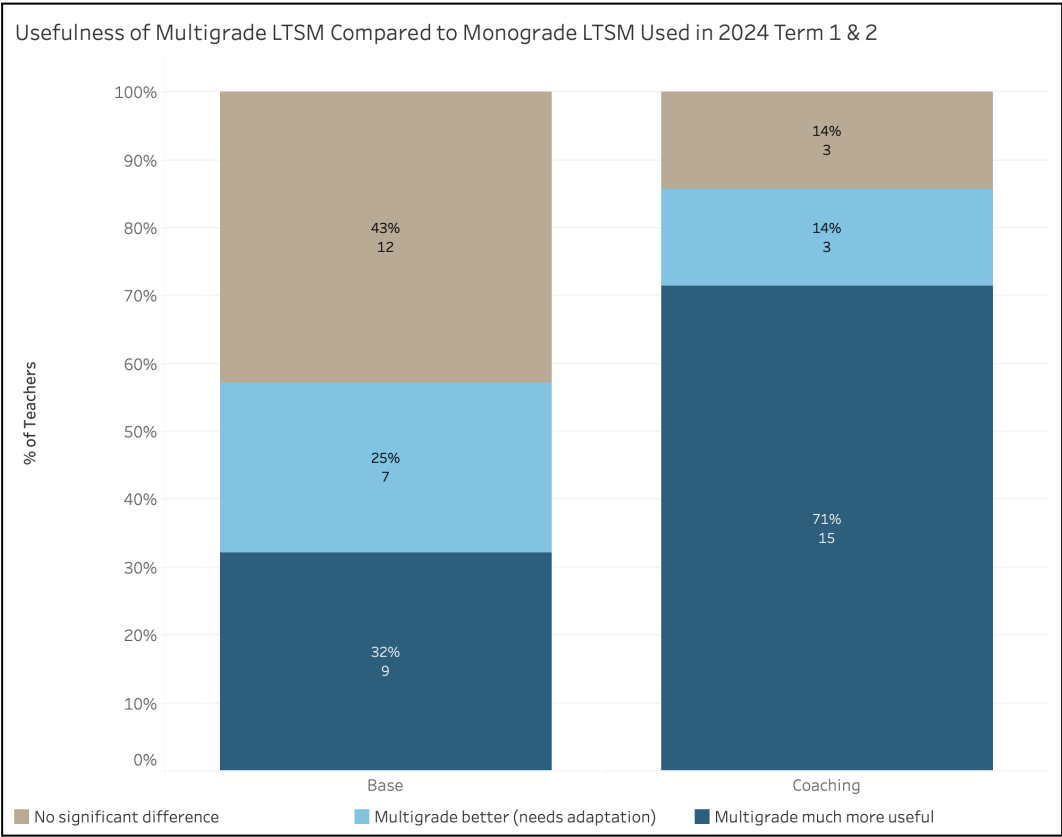


Figure 6: Usefulness of multigrade LTSMs compared to monograde LTSMs.

While 71% (n=15) of multigrade coaching teachers found the multigrade materials much more useful than monograde materials, only 32%(n=9) of base teachers shared this view. Notably, 43% (n=12) of base teachers reported no significant difference between the two, compared to only 14% (n=3) of coaching teachers. This suggests that without adequate coaching support, teachers may not fully recognise or leverage the unique features of multigrade materials for differentiated instruction. The strong appreciation among the majority of coached teachers validates the investment in multigrade-specific materials, while also suggesting that their effectiveness is closely tied to the quality and consistency of accompanying instructional support from coaches.

- Multigrade teachers' reported efficiency gains from using LTSMs:** As expected, the provision of scripted lesson plans significantly reduced teacher preparation time and streamlined instructional planning. This was particularly valuable for multigrade teachers, who typically struggle with classroom management while teaching multiple grades simultaneously. The survey asked teachers how much time they spent on lesson preparation, occupational tasks, and assessments compared to before EGRP II.

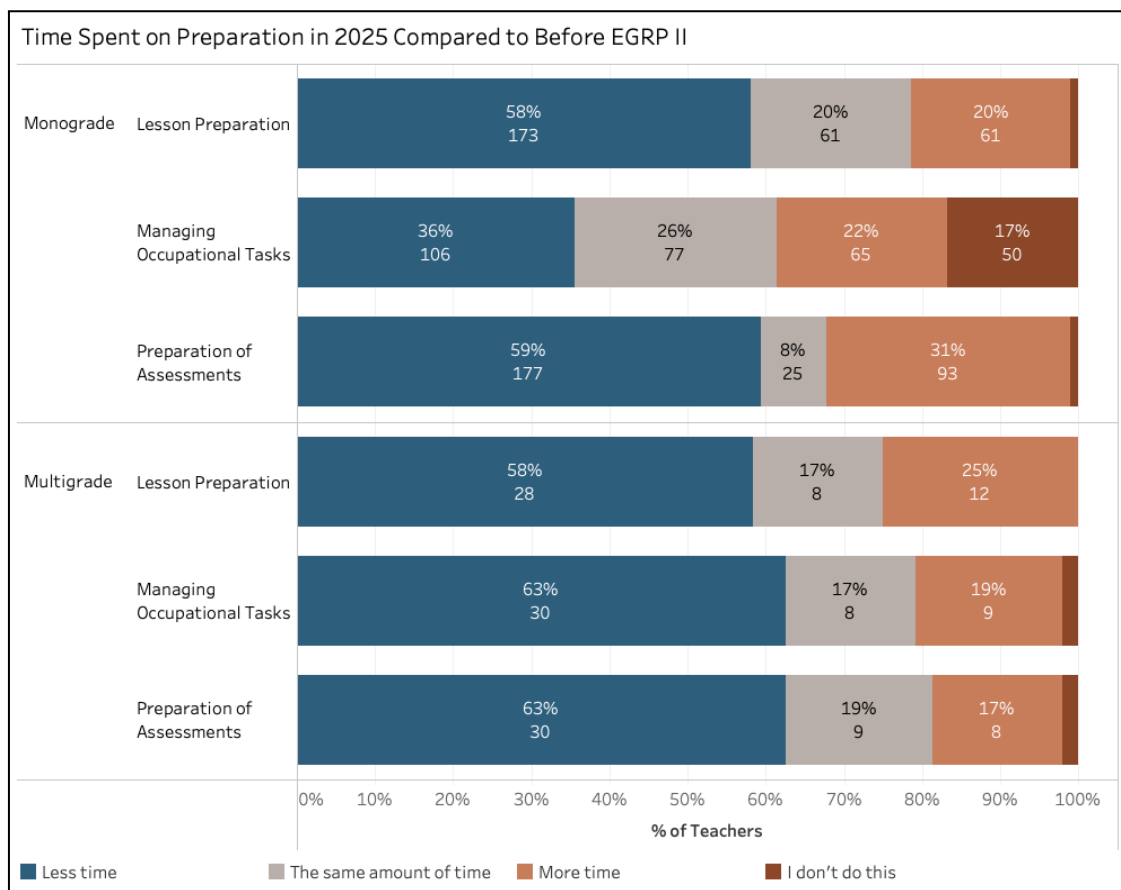


Figure 7: Time spent on lesson and assessment preparation, as well as occupational task

Figure 7 shows that more multigrade coached teachers reported efficiency gains from LTSM use, with 58% (n=28) spending less time on lesson preparation, 63% (n=30) on managing and preparing occupational tasks, and 63% (n=30) on preparing assessments. Fewer uncoached multigrade teachers reported similar gains, particularly in managing and preparing occupational tasks (36%, n=106), though lesson preparation (58%, n=173) and assessment preparation (59%, n=177) were comparable. Monograde teachers, both coached and uncoached, reported notably fewer efficiency gains across all three areas.

These patterns suggest a compounding effect between LTSMs and coaching. Multigrade teaching carries inherently greater planning and administrative demands, meaning structured materials make a proportionally larger difference to workload. However, the pronounced gap between coached and uncoached multigrade teachers, particularly in managing occupational tasks, indicates that coaching plays a critical role in helping teachers fully utilise LTSMs. Without that support, the materials alone appear insufficient to unlock the same gains. The comparatively modest efficiency gains among monograde teachers likely reflect the fact that their planning demands are less complex to begin with, reducing the marginal benefit of LTSM use.

Receipt and use of tablets

Every programme teacher was expected to receive a tablet containing essential LTSMs, including lesson plans, and serving as the primary tool to access further resources from the EGRP platform. Tablets in JTG were supplied by the Sishen Iron Ore Company's Community Development Trust (SIOC-CDT) and distributed by the implementation partner, whilst those in FB were supplied by UNICEF and distributed directly by provincial authorities. Tablet distribution occurred during Term 1 2025 training. However, multigrade and Grade 1 teachers in FB received their devices only in late February and early March 2025 due to delays caused by the 'Work-to-Rule'⁴ directive issued by teachers' unions in January and February 2025.

Implementation data shows that all teachers received these tablets, confirmed with equally high percentages of teachers surveyed.

- **Teachers who received tablets:** Figure 8, confirms that overall tablet receipt was high across both districts, though slightly more teachers in JTG (5%, n=267) reported not receiving tablets compared to FB. These slight differences from monitoring data likely reflect teacher attrition and movement rather than non-distribution of tablets by the implementing agency. Teachers who left may not have returned their tablets, leading newly appointed teachers to report not receiving devices.
- **Condition of tablets:** Figure 9 shows that 70% (n=241) of surveyed teachers indicate that tablets are working perfectly, whilst 22% (n=76) have major issues, primarily freezing or unresponsiveness. At district level, JTG reported better functionality, with 76% (n=192) of teachers having perfectly functioning tablets compared to 55% (n=49) in FB. Critically, 29% of tablets in FB were either non-functional or severely impaired, compared to 21% in JTG. Tablet functionality was significantly higher among multigrade teachers compared to their monograde counterparts
- **Sharing of tablets:** The majority of teachers do not share tablets (Figure 10), though analysis of tablet sharing at cohort levels showed it is slightly higher among multigrade base teachers.

⁴ In late January 2025, the South African Democratic Teachers' Union (SADTU) and the Suid-Afrikaanse Onderwysunie (SAOU) in the Northern Cape, initiated a 'Work-to-Rule' action. This industrial action restricted teachers to working only their contracted hours, specifically the seven compulsory academic hours of the school day.

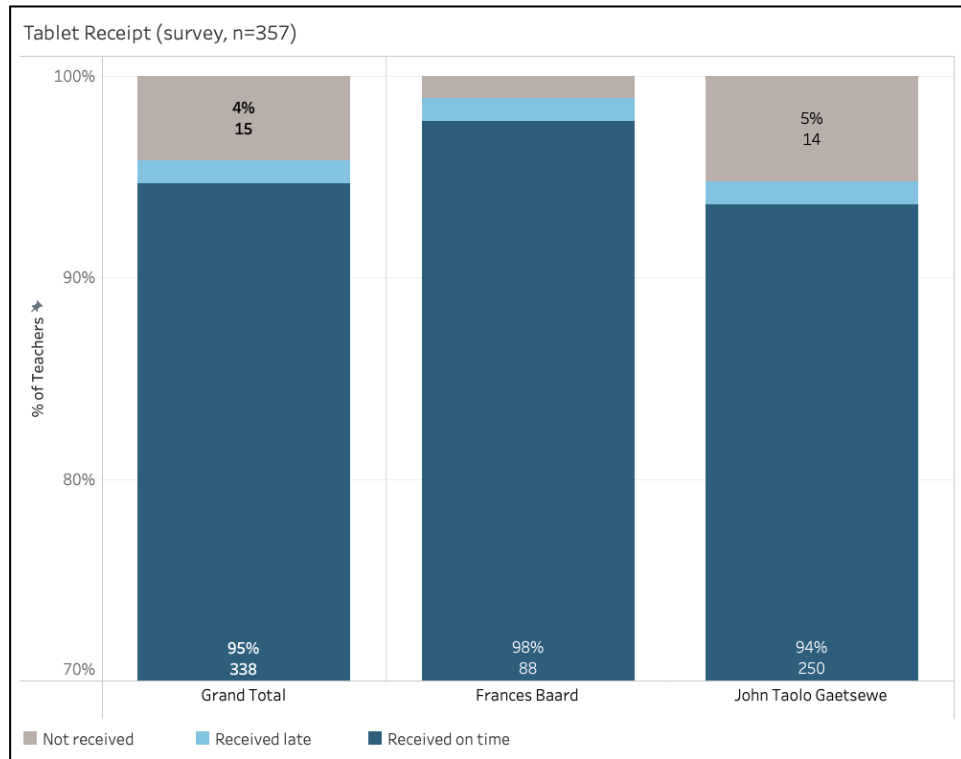


Figure 8: Tablet distribution

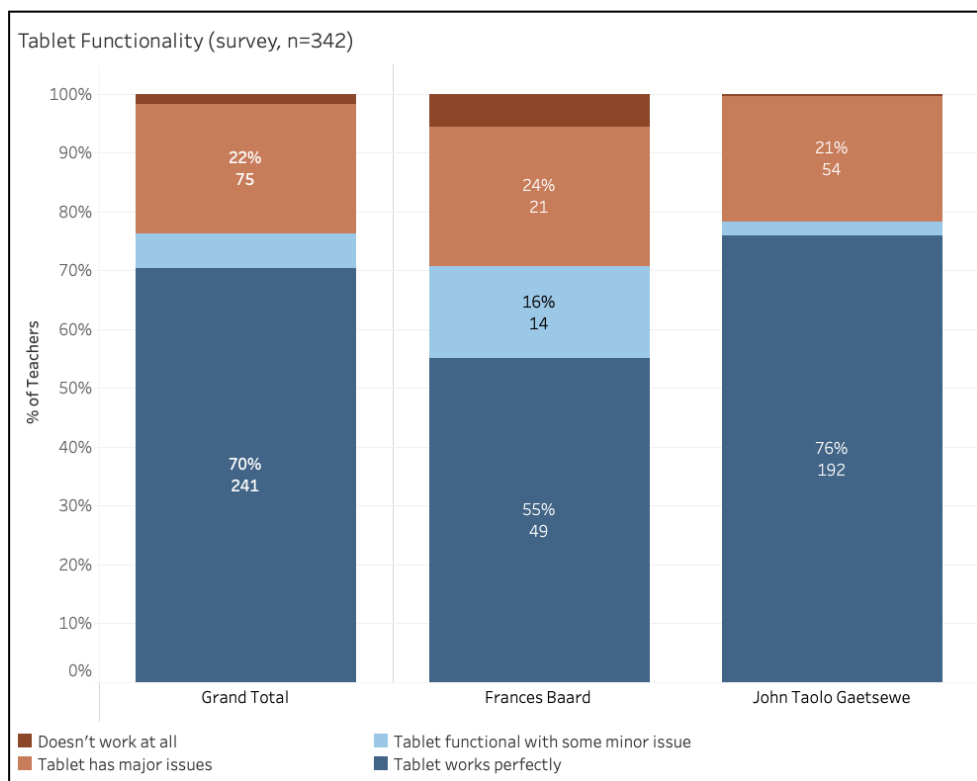


Figure 9: Condition of tablets

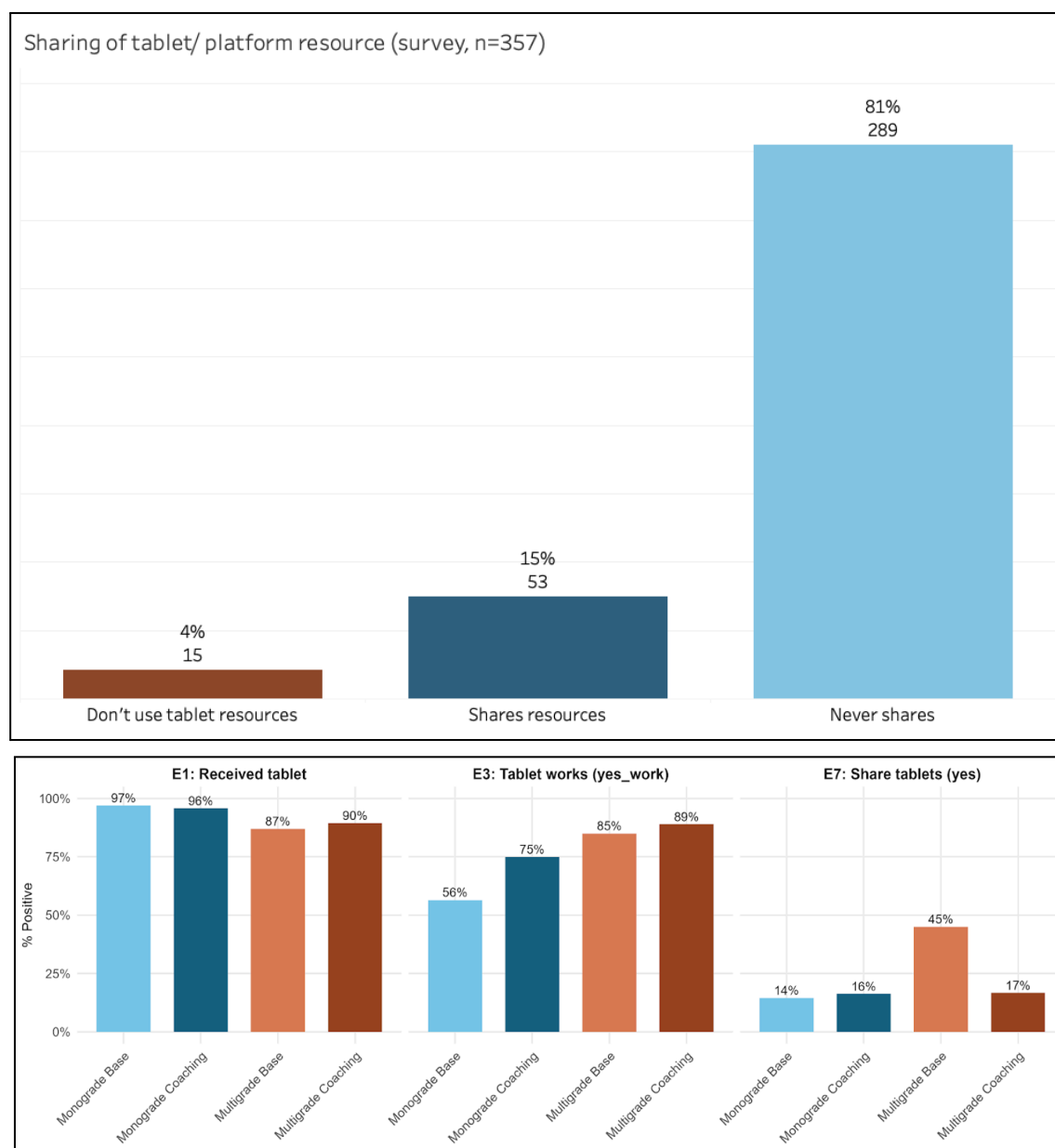


Figure 10 Tablet sharing

4.1.4 Coaching

Coaches were expected to support teachers for at least 3.5 hours per term, translating to approximately 3 visits per teacher per term.

Coverage: Similar to 2024, monitoring data indicates strong coaching reach in 2025, with 96% (n=217) of monograde teachers and 99% (n=21) of multigrade teachers receiving a coaching visit during the year.

Dosage: Though prescribed dosage was not fully met across all teachers, according to the implementing partner's monitoring data, coaching dosage showed a notable improvement over time, with the proportion of teachers meeting or exceeding the prescribed coaching dosage increasing from 2024 to 2025. The increase was evident across both teacher types with monograde teachers rising from 8% (n=22 of 267) in 2024 to 18% (n=40 of 227) in 2025. Multigrade teachers also improved, from 51% (n=18 of 35) to 60% (n=12 of 22). This shows that multigrade teachers received the intended coaching dose at a considerably higher rate than their monograde counterparts, though the multigrade sample is notably smaller and comparisons should be interpreted with this in mind.(Figure 11)

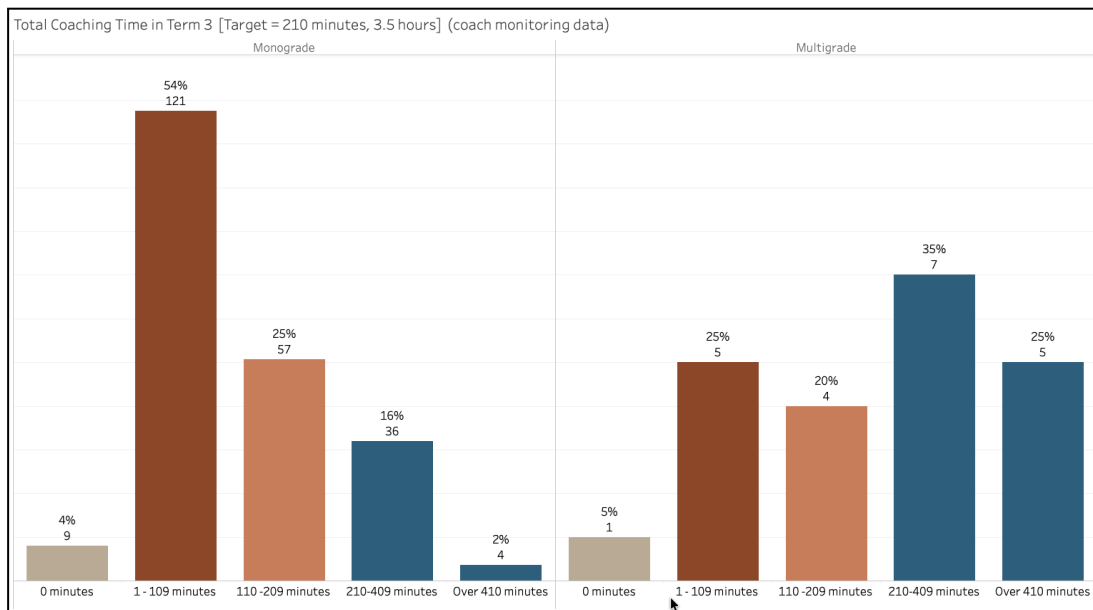


Figure 11 Distribution of coaching time in Term 3 2025 by school type

A further 20% (n=4) of multigrade teachers and 25% (n=57) of monograde teachers received at least half the target dosage (1.8 to 3.5 hours). However, 54% (n=121) of monograde teachers received less than half the prescribed dosage (under 1.8 hours), compared to 25% (n=5) of multigrade teachers, suggesting that while coaching reached the majority of teachers, dosage adequacy remains an area for improvement, particularly for monograde teachers. Figure 12 shows the same dosage in terms of visits not amount of time spent on coaching.

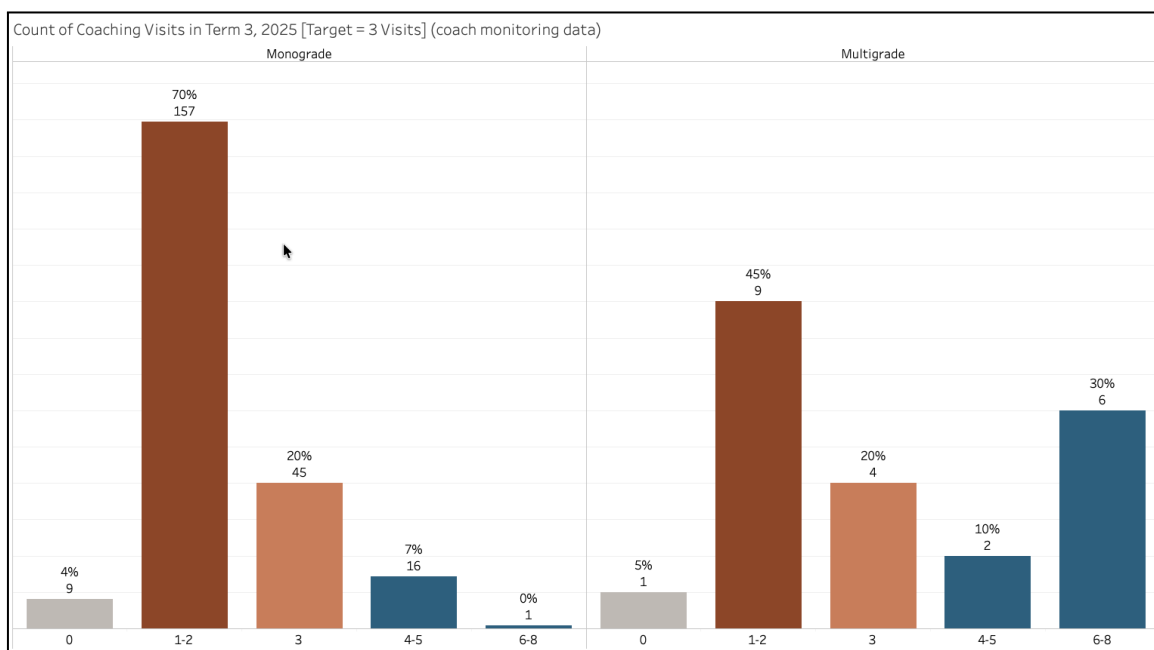


Figure 12 Distribution of coaching visits in Term 3 2025 by school type

The consistency of coaching visits differed notably between school types, and there was also considerable variation in the support received across individual teachers. This dispersed distribution likely reflects logistical challenges in standardising coaching delivery across a geographically spread programme.

Quality of coaching provided: Quality of coaching was assessed through case studies and survey questions examining whether coaches were implementing core principles of developmental instructional coaching. Key dimensions included frequency of reflection and feedback sessions, alignment of feedback with lesson plans and learning materials, opportunities for teacher reflection before receiving feedback, and inclusion of improvement strategies in coaching conversations.

The survey results in Figure 13 show coaching quality was generally high. Further analysis shows that it was comparable between school types, indicating that developmental coaching principles were largely implemented across both monograde and multigrade contexts.

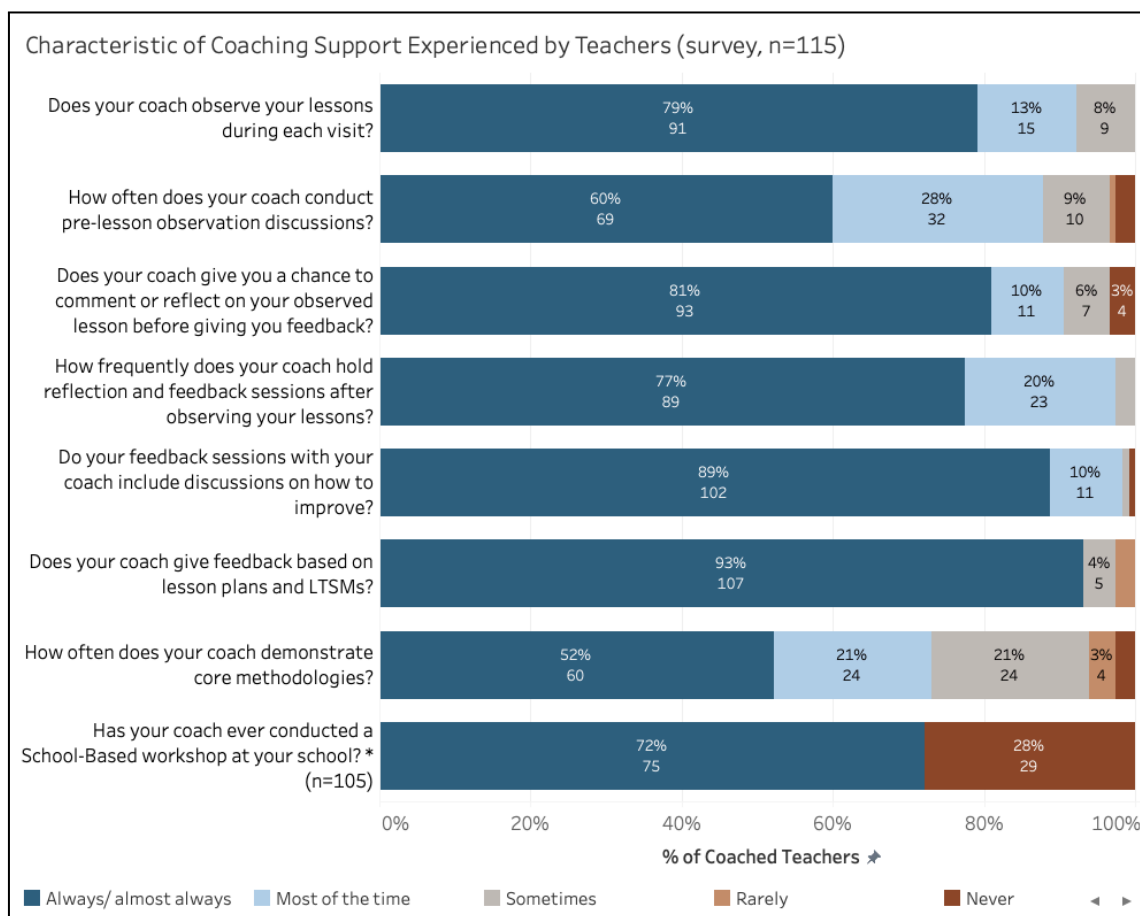


Figure 13 Teacher reported quality of coaching

An important area for growth lies in the consistent demonstration of core teaching methodologies, which, if strengthened, would provide teachers with concrete models to observe and emulate in their own practice.

Insights from the 2024 EGRP II case studies revealed that while most coaches maintained consistent support for EGRP II core methodologies and built strong teacher relationships, the majority regularly missed opportunities to facilitate deeper reflection during post-observation discussions. Although most opened feedback sessions by asking teachers about their lesson experience, most typically moved quickly to offering advice rather than scaffolding a substantive reflective process. As a result, the teacher reflection section had limited evidence of deeper self-assessment or teacher driven insights.

The second round of case studies in 2025 confirmed the persistence of this pattern. However, the teacher survey results present an encouraging picture that reflective practices are happening widely across schools, with 81% of teachers reporting opportunities to reflect during feedback sessions. This suggests coaches are creating space for the teacher's voice. The key opportunity now lies in deepening the quality of these reflective conversations to move from surface level exchanges to more analytical, teacher led self-assessment.

4.1.5 Cluster workshops

Coaches were also expected to facilitate cluster workshops. These sessions provided targeted training on topics where teachers needed additional support, as identified through coaching interventions, teacher feedback, and consultation with district officials. They were open to both coaching schools and base programme schools.

Cluster workshops demonstrated substantial progress in EGRP II, reaching 50% of programme teachers. Given that these workshops are voluntary and take place in the afternoon after school hours, participation reflects significant commitment from teachers to their professional development. This represents an important evolution from EGRP I, where this component did not materialise in a consistent manner. There has also been progress over time within EGRP II: while in 2024 only two coaches conducted workshops with no implementation data collected, the 2025 results show not only expanded reach but also systematic tracking of engagement.

Implementation data from 2025 showed particularly promising engagement in places where the model was fully adopted. In multigrade schools, 73% (n=41) of teachers participated in at least one workshop, with nearly a third attending three sessions, demonstrating strong teacher buy-in to this professional learning approach. The primary growth opportunity lies in expanding workshop reach within monograde schools, where only 47% (n=214) of teachers have engaged with at least one session. Addressing logistical barriers around accessibility and scheduling could significantly increase participation to levels comparable with the successful multigrade implementation.

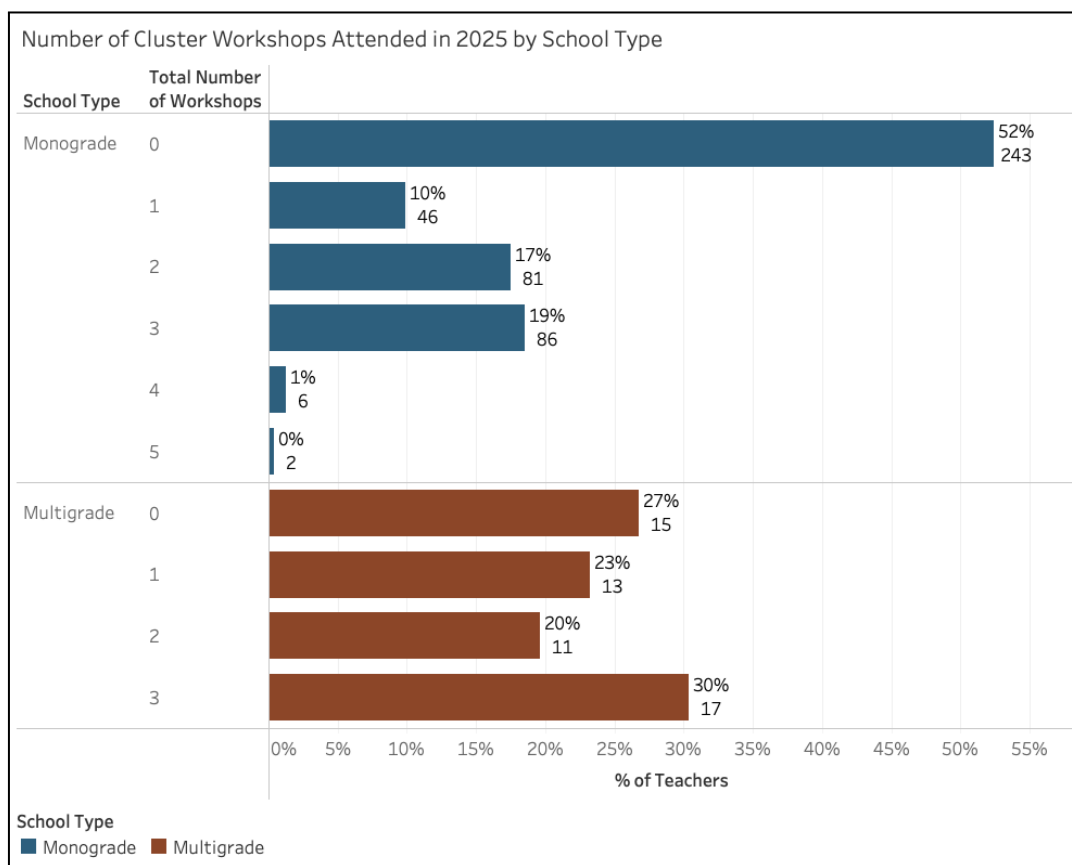


Figure 14 Cluster workshop attendance by cohort

4.1.6 Systemic reflections on training, LTSM and coaching dosage challenges

A recurring challenge across evaluations of similar programmes has been the difficulty implementing partners face in meeting prescribed dosage targets. In the EGRP I, coaching dosage targets were substantially missed, with the DH coaching arm achieving only 23% of the target and the external coaching arm 18%. Teacher training achieved comparatively higher coverage, though it also fell short on expected dosage with 72% of teachers attended all scheduled training sessions. The Reading Support Programme (RSP), implemented in 2019 across 230 schools in two North West Province districts, similarly found that coaching dosage was insufficient to support differential impact, a pattern that extended to School Based Workshops (SBW), training attendance, and delivery of LTSM.

All three programmes, including the current one, were structured around the “triple cocktail” approach and share notable continuity in implementing partners. Molteno participated in the EGRP I, while both Oxford University Press and Molteno formed part of the RSP consortium. These organisations have accumulated considerable experience across successive programmes and have engaged with the implementation challenges each evaluation has surfaced. The persistence of dosage-related constraints across this body of work likely reflects the structural and contextual nature of the barriers involved, rather than just issues of partner capacity or programme design. A dedicated implementation study exploring the conditions under which dosage targets can realistically be met would help to better understand these constraints

and inform the design of future programmes.

Implementing partner reflection on the EGRP II shares some known insights, some of which are specific to the Northern Cape.

Achieving aspirational coaching hours depends on a predictable school environment and reliable access to educators and school leadership. Several systemic and contextual realities constrained dosage during implementation, not as a reflection of failure in design or intent, but as a consequence of the complex environment in which the programme operated. Within the Northern Cape, competing departmental demands, human resource pressures, and significant geographical constraints all reduced the time available for scheduled coaching across the term.

- **Competing departmental initiatives:** Mandatory activities initiated by the Northern Cape Department of Education, including subject-specific training, curriculum workshops, and administrative directives, frequently ran concurrently with planned coaching sessions. While these activities serve legitimate school improvement purposes, their scheduling created unavoidable overlaps that reduced the number of coaching sessions achievable within any given term.
- **Leadership and management obligations:** District meetings, school improvement planning sessions, and other external engagements further limited educator availability. This was especially pronounced in multigrade schools, where a single teacher often carries responsibility for multiple grades while also serving as school principal. When these educators are called away, learning programmes pause entirely, making coaching impossible and creating recurring structural constraints on consistency.
- **Teacher absenteeism:** An increase in teacher absenteeism, largely due to extended sick leave, also contributed to the shortfall. In smaller multigrade and rural schools, the absence of even one educator can suspend the timetable altogether, eliminating coaching opportunities for that visit and reducing overall session frequency across the term.
- **Geographic distances and logistical constraints:** The Northern Cape's vast geography means coaches regularly travel hundreds of kilometres to reach a single school, substantially reducing the net time available for direct coaching. This also limits the feasibility of rescheduling cancelled sessions within the same term, as return visits carry significant logistical implications. Weather, road conditions, and transport delays add further uncertainty in rural areas.

Individually, each of these factors presents a manageable obstacle; collectively, they create conditions in which consistently achieving the aspirational target of 3.5 coaching hours per teacher per term is highly unlikely. The implementing partner's experience indicates that approximately 1.5 hours per term more accurately reflects what is realistically attainable in this context.

4.3 Teacher classroom practices

The key programme components focused on influencing teachers' classroom practices, as they are primary agents in improving learner outcomes. This section evaluates if observed classroom practices aligned with training methodologies and if differences existed between monograde and multigrade settings. Evidence was gathered from surveys, interviews, and two rounds of classroom observations conducted during the case study visits to 10 schools (62 observations across 30 teachers).

Adoption/adaption of instructional methodologies and reading protocols: the case studies found that teacher adherence to reading routines improved across both Shared Reading (SR) and Group Guided Reading (GGR). More proficient teachers followed protocols closely, reading Big Books expressively and drawing learners into meaningful discussion, while weaker practitioners continued to default to chorus reading rather than developing individual reading during GGR. Confidence in teaching EFAL also grew, with teachers reducing code switching and increasingly sustaining English-only instruction as coached. Interestingly, by 2025, teachers at base schools were often applying EGRP II protocols more consistently than those receiving dedicated coaching support.

Teacher perception of ease of implementation of methodologies: Teacher surveys (n=342) indicate that teachers are most confident implementing foundational home language methodologies, particularly Phonics and Spelling. Confidence declines notably for Shared Reading, Group Guided Reading, and Writing, a pattern consistent with EGRP I, EGRS, and similar evaluations. This reflects a broader systemic challenge rooted in the realities of most low quintile South African classrooms, where large class sizes and wide learner ability ranges make differentiated instruction particularly demanding. Compounding this is the policy that allows learners to be retained only once per phase, meaning learners who are not yet ready are nonetheless progressed to the next grade. This results in classrooms where learners span a wide range of literacy levels, effectively making most monograde classrooms function similarly to multigrade settings, but without commensurate teacher training or resources to support this context. These conditions make methodologies requiring intensive classroom management and learner grouping inherently more difficult to implement with fidelity. This points to a need for more sustained, contextually responsive coaching support focused specifically on these methodologies, alongside consideration of whether current programme models adequately account for the structural constraints teachers face.

The survey results in Figure 15 show that HL monograde teachers consistently report higher ease of implementation than their multigrade counterparts, with the gap smallest for Phonics and Spelling, where all cohorts report above 90% ease. The divide is most pronounced for GGR, where multigrade base teachers reported only 54% (n=15) ease compared to 80% (n=158) for monograde base. Writing is universally the most challenging methodology across all cohorts and classroom types. A consistent association is observed between coaching and higher reported ease of implementation, particularly in multigrade settings. Coached multigrade teachers reported noticeably higher ease across all methodologies, most notably in Shared

Reading (90%, n=18 vs 79%, n=22), GGR (70%, n=14 vs 54%, n=15), and Writing (70%, n=14 vs 61%, n=17), in many cases nearly matching their monograde peers.

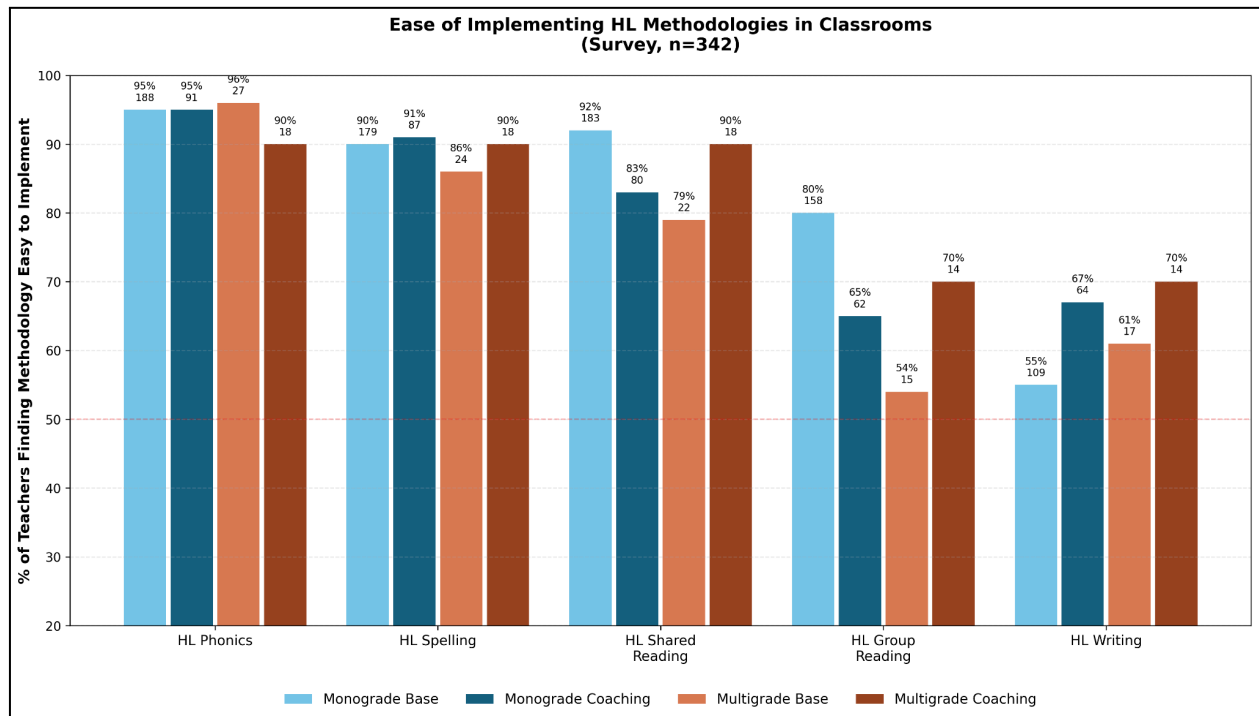


Figure 15 HL Methodologies ease of implementation

Teacher responses in Figure 16 show that EFAL methodologies are broadly harder to implement than HL methodologies across all cohorts, possibly reflecting the added complexity of teaching in a language that is not the learners' (or often teachers') home language. Foundational EFAL methodologies ease of implementation data from the teacher survey shows considerably more variation than in HL, with monograde base Spelling dropping to 58% (n=115) and GGR to just 29% (n=58). The association between coaching and higher ease of implementation holds in EFAL as well, and is particularly pronounced in multigrade settings, where coached multigrade teachers not only outperform their uncoached counterparts but in some cases exceed monograde base teachers in Phonics and Spelling.

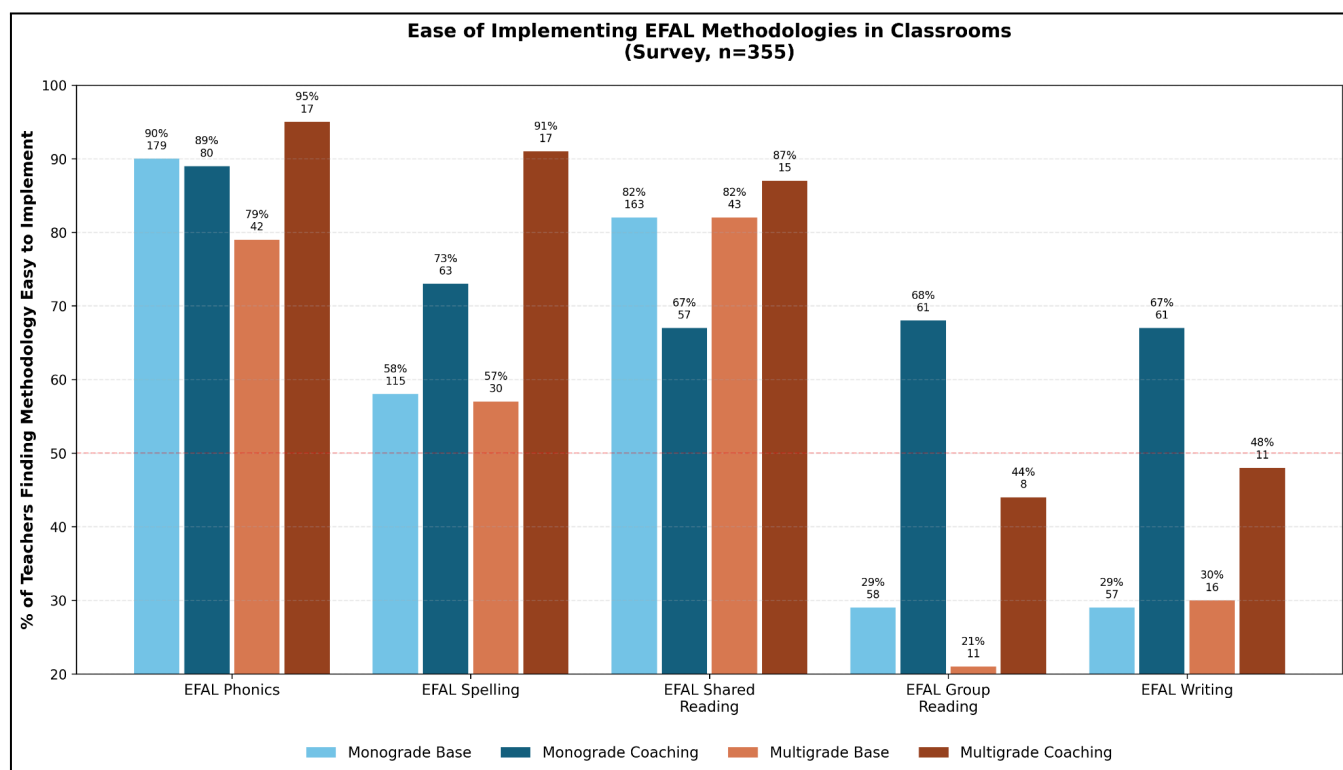


Figure 16 EFAL Methodologies ease of implementation

Classroom management and grouping: A significant shift occurred toward ability-based grouping, particularly in multigrade settings, where teachers increasingly grouped learners by reading ability rather than grade level, enabling targeted instruction. Teachers managed lesson timing effectively, moving from complaints about timing in 2024 to executing well-paced, organised lessons in 2025. Monograde classrooms typically used grouped desk arrangements for collaboration, while multigrade classrooms maintained row seating by grade, with some teachers innovating with U-shaped arrangements for better monitoring.

Multigrade teachers reported that prior to the intervention, they frequently taught only one grade at a time while leaving others disengaged, lacking clear methodology for simultaneous instruction. The training provided specific structures, including sequencing lessons within extended timeframes and managing occupational tasks across grades, fundamentally transforming their teaching approach.

Assessment and written work: Written work remained inadequate in quantity and quality, with partial workbook completion, insufficient corrections, and misalignment between assigned work and lesson plans. Time constraints frequently prevented adequate completion of writing exercises, a reflection the implementation partner also noted in its monitoring work. Writing tasks were flagged as the most difficult by most teachers in the survey.

Occupational tasks: Teachers managed occupational tasks alongside GGR, with significantly better outcomes in classrooms supported by Reading Champions who assisted with discipline and resource management, allowing teachers to focus on reading groups.

Use of resources and technology: Most teachers navigated tablets confidently for lesson planning, with resistance limited to veteran teachers who found technology challenging or relied on Reading Champions for support. Classrooms featured print rich environments with posters, flashcards, and reading corners, though overcrowded multigrade classrooms sometimes sacrificed reading corners for desk space. Multigrade teachers effectively shared resources across grades while differentiating vocabulary instruction by level.

Multigrade teachers effectively used LTSMs, such as Listening and Speaking posters and vocabulary flashcards, with the entire class simultaneously. They employed differentiated questioning by directing grade appropriate questions to learners at different levels while addressing what they could see on the same poster.

The teacher survey explored the access and use of the EGRP platform as well as what format of lesson plans they preferred to use.

- **Paper based or electronic lesson plans:** The largest proportion of teachers (44%, n=154) prefer a hybrid approach, combining both paper and electronic formats, reflecting a gradual integration of tablets alongside familiar traditional methods, and to some extent practical constraints such as tablet functionality issues raised by teachers. This finding resonates with evidence from the EGRP I, where teachers showed a preference for paper-based lesson plans.

Multigrade base teachers were considerably more reliant on paper (39%) compared to their monograde counterparts (15%), pointing to the compounding challenges of managing multi-grade instruction while navigating new technology.

Coached multigrade teachers show a notably different pattern compared to their uncoached counterparts, with far fewer relying exclusively on paper (18% vs 39%) and more using combined formats (36% vs 25%), suggesting coaching may be associated with greater capacity to integrate electronic resources.

Overall, the data points to an early stage of digital transition, highlighting the need for sustained digital literacy support if electronic lesson planning tools are to become a consistent part of teachers' daily practice.

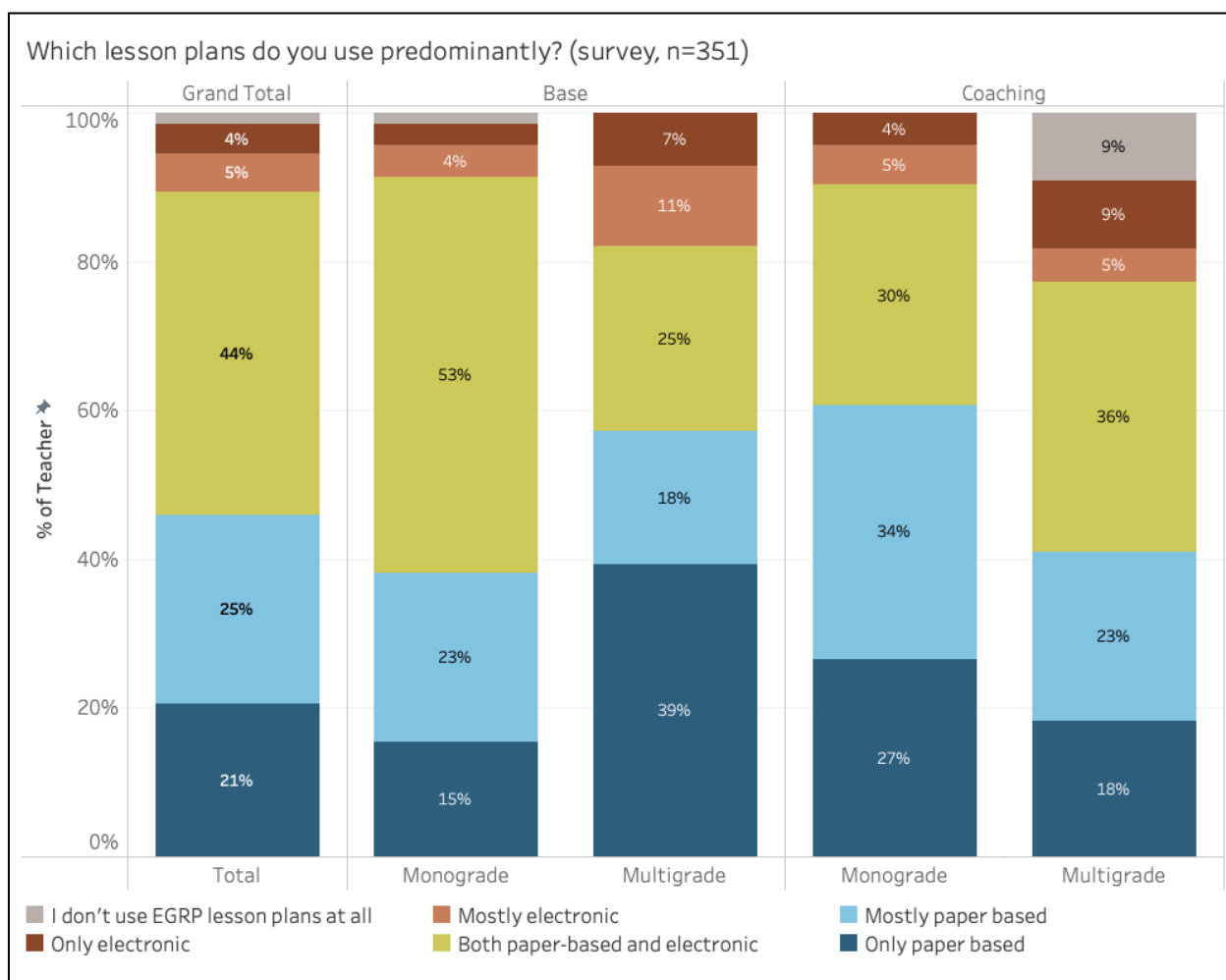


Figure 17 Adoption of electronic vs paper based lesson planning

- EGRP Platform use:** The EGRP II introduced a platform allowing teachers to access LTSMs and other resources to support their instructional activities, with each teacher assigned a unique login. Hosting has since been transferred to the DBE, having initially been managed by the implementation partner. Survey results in Figure 18 show varying teacher experiences across platform access, navigation, and technical support.

Overall, teachers found the platform relatively easy to access and navigate, though monograde teachers reported greater ease than their multigrade counterparts, particularly in the base group (accessing: 85% (n=175) vs 68% (n=19); navigating: 84% (n=174) vs 64% (n=18)). Where coaching was provided, teachers especially in multigrade settings reported greater ease of access (77% (n=17) vs 68% (n=19)), though improvements were less consistent for navigation and technical support, pointing to a continued need for more targeted technical assistance to support reliable platform use.

Technical support proved to be a challenging area, with monograde base teachers reporting difficulty with getting technical support, (22%, n=46) though this should be

interpreted with caution as the majority of monograde teachers reported never encountering platform issues (monograde: 54% (n=112); multigrade: 43% (n=12)), suggesting that demand for support was limited in the first place. This also suggests that the platform was generally functional when teachers accessed it.

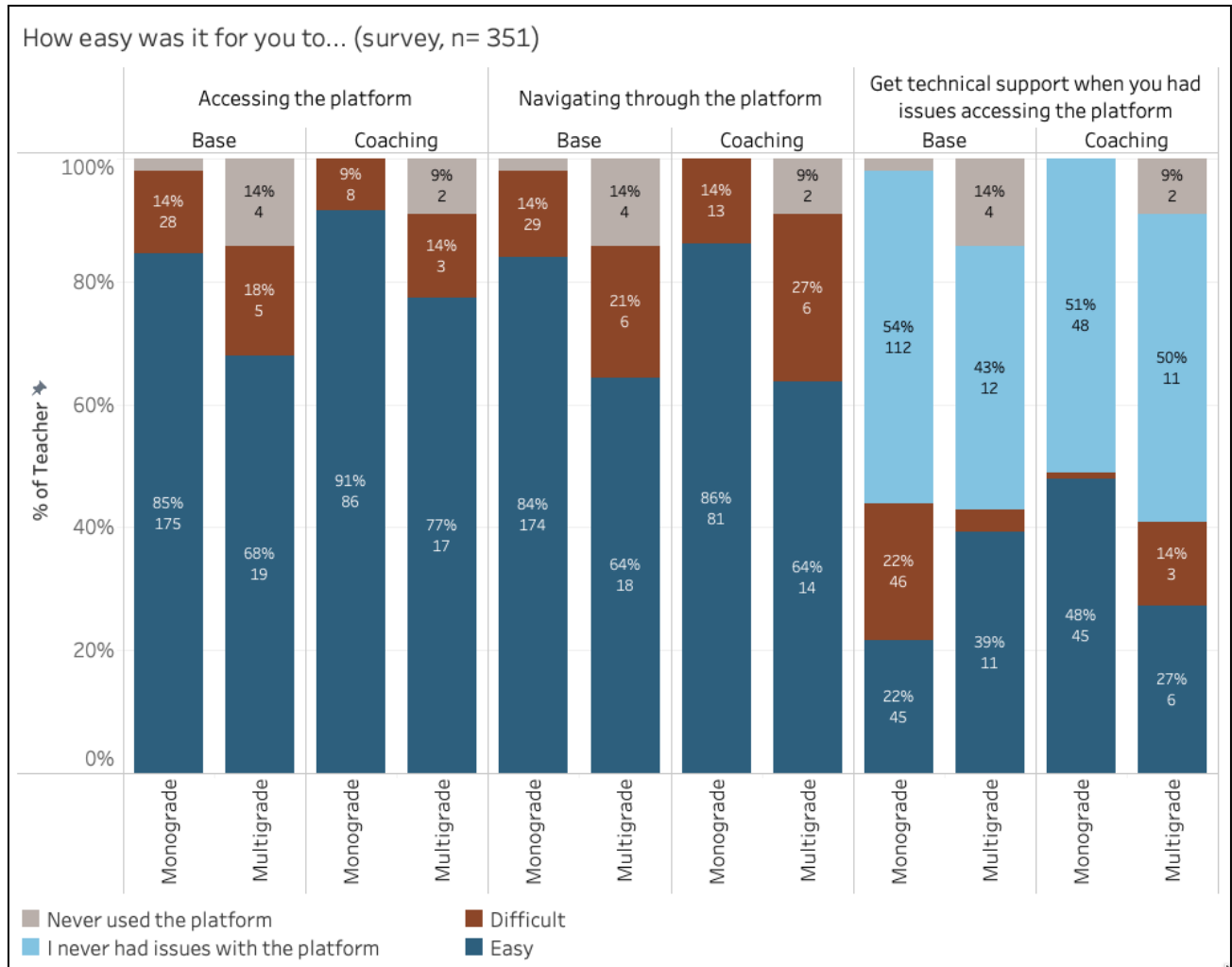


Figure 18 Teacher experiences in accessing EGRP resources

Use of curriculum trackers and assessment checklists: To support curriculum coverage across schools, the EGRP II introduced curriculum trackers as tools teachers could use to monitor their pacing and planning against prescribed curriculum requirements. However, case study findings indicate that teachers struggled to integrate curriculum trackers and assessment checklists into their practice, with many viewing them as an additional administrative burden rather than a useful planning support. Uptake remained limited even among early adopters, and engagement with these tools had declined by 2025. Survey findings reflect a similar pattern, with multigrade teachers in particular showing low uptake (multigrade base: 30% vs monograde base: 19%). Where trackers were used, timing varied considerably, pointing to an inconsistent practice across schools. Teachers most commonly cited time constraints as the key barrier, suggesting

the tools may need to be better integrated into existing planning routines to support more meaningful uptake.

Systemic reflections on the introduction of technology in educational innovations: Reflection on the introduction of tablets in EGRP I and the EGRP I App suggests that the EGRP II represents a meaningful improvement. The EGRP II benefited from more deliberate planning, better training support, and a more polished and accessible interface, addressing several of the development challenges that constrained the EGRP I. Where the EGRP I was hampered by delays in lesson plan loading and other technical difficulties that discouraged teacher engagement, the EGRP II platform arrived as a more stable and refined product.

Nevertheless, the challenges associated with technological innovation in education remain considerable, and the uptake issues observed in the EGRP II are not isolated. In the EGRP I, app data showed that only 36% of registered teachers accessed the app to view core methodologies over an eight-month period, and coaches reported that in some instances only one teacher would engage with the tablet content, with others copying that teacher's activities rather than accessing the material independently. Similar patterns have been observed in the EGRP II, suggesting that barriers to meaningful technology adoption persist across both programmes despite improvements in design and delivery. This points to deeper behavioural and contextual factors that technical refinements alone are unlikely to resolve.

These patterns are consistent with what the technology adoption literature describes as shallow adoption, where a technology is nominally present but not engaged with as intended, and reflect the well documented gap between access to devices and meaningful pedagogical integration that characterises early stage diffusion of innovation in education contexts. This suggests that future programme iterations should invest not only in the quality of the technology itself, but in the structured pedagogical support and behaviour change processes needed to move teachers from passive access to active and independent use.

4.4 Learner outcomes

This section presents learner assessment results based on the administration of standardised EGRA and EGMA subtasks across both Grade 2 and Grade 3. Learner outcome findings will be cautiously cross-referenced with the Funda Uphumulele National Survey (FUNS) Northern Cape Provincial Report (2025) to contextualise the results. The FUNS was established to create a national baseline for reading proficiency in South Africa, using reading benchmarks across all 11 official languages. Its goal was to measure the percentage of learners meeting foundational reading standards in Grades 1–3, ensuring they are on track to read for meaning by Grade 4. The study assessed learners in Grades 1 to 4 from 59 public schools across the Northern Cape in Setswana and Afrikaans, with the former accounting for 35% of learners. This report will only reference the Setswana results. The evaluation will also reference EGRP I results where appropriate. Both studies were implemented in monograde schools meaning that the only comparison will be with monograde schools in the evaluation.

4.4.1 Description of learner assessment subtests

To measure learner outcomes across different literacy skills in EGRP II, a broad assessment EGRA/EGMA battery was administered at the endline. Table 9 summarises the individual and group assessments used, including the language of administration, applicable grade levels, and duration of each task. Several assessments used identical tasks across both grades to allow for direct comparison of performance. This evaluation and the FUNS share similar subtests. Important to note is that non-readers (as per FUNS) refers to learners who scored above zero (implying they read at least one word) but whose proficiency is too low to be classified as pre-readers.

Table 9 EGRP II Assessment tasks administered at the endline

Test Name	Language	Grade 2	Grade 3	Time (seconds)	
Individual Assessments					
Home Background	HL	X	X	180	
Rapid Object Naming	HL	X		20	
Letter Sound Knowledge	HL	X	X	60	Same task
Complex Consonants and Diacritic Knowledge	HL		X	60	
Word Reading	HL	X	X	60	Same task
Oral Reading Fluency (ORF)	HL	X	X	180	Same story
Oral Reading Fluency Comprehension	HL	X	X	90	Same task
Word Reading	EFAL	X	X	60	Same task
Oral Reading Fluency (ORF)	EFAL	X	X	180	
Oral Reading Fluency Comprehension	EFAL		X	90	
Group Assessments					
Complex Consonant Sound Knowledge	HL	X	X	50	
Word Recognition and Writing	HL	X		50	
Written Comprehension	HL		X	900	
Early Grade Mathematics Assessment	HL		X	120	

4.4.2 Learner Test Scores

Grade 2 Test Scores

Thirty-three percent of learners in HL (Setswana) ORF scored zero, indicating that they were unable to read a single word in their home language. For EFAL, 28% of learners were unable to read a single word on the word reading assessment. Among HL learners who demonstrate some reading ability, performance remains limited. The typical learner reads only 16 words correctly per minute (cwpm) in their home language, which is far below the DBE benchmark targets of 40 cwpm. The average in English is only 5 cwpm. The strongest readers reach

approximately 58 cwpm in their home language, while in English the top performers achieve just 31 cwpm. Of the learners who attempted the comprehension task, half answered fewer than five out of eleven questions correctly.

Taken together, the data reveal a concerning continuum that a third of learners cannot read a passage at all, most of those who can read some words do so well below expected fluency levels, and even among the most capable readers, comprehension remains weak. **Reading difficulty in this cohort is not confined to a struggling minority but reflects a system-wide challenge.**

The following table summarises the reading performance of the Grade 2 cohort.

Table 10 Grade 2 Summary statistics per sub-task

Subtask	Observations	% zero	Score attained at each percentile					Min	Max	items in the subtask
			p10	p25	p50	p75	p90			
HL Rapid Object Naming cwpm	1,930	0.31	30.00	36.00	48.00	54.00	63.00	0	110.53	36
HL Letter Sound Knowledge (LSK) clpm	1,930	2.33	11.90	31.00	47.00	63.27	76.60	0	128.89	60
HL Word Reading cwpm	1,930	9.64	1.00	2.00	9.00	21.00	27.00	0	85.71	60
HL Oral Reading Fluency cwpm	1,926	33.28	0.00	0.00	16.33	41.78	57.73	0	131.38	128
HL Comprehension (% score)	1,926	39.72	0.00	0.00	0.25	0.45	0.59	0	1.00	11
EFAL Word Reading cwpm	1,929	28.46	0.00	0.00	5.00	20.00	31.20	0	78.00	80

Grade 3 Test Scores

A similar analysis was done for the Grade 3 learners. The results show improvement from Grade 2 but persistent challenges remain. Between 15% to 30% of learners still cannot read a single word, with the highest zero scores rates (nearly 30%) appearing in English comprehension tasks. The percentage of zeros scores for HL is similar to, though marginally higher than the FUNS score of 15% of learners who cannot read a single word. The EGRP I score was marginally higher at 17%. Among those learners who can read, the typical learner reads 42 words per minute in Setswana, while EFAL fluency remains considerably lower.

Even high-performing learners (top 10%) show mixed results with the strongest readers achieving approximately 79 words per minute in home language fluency, they struggle with

complex phonics elements, achieving only 57% proficiency in areas such as diacritics and complex consonants.

These results indicate uneven progress across the cohort and present the teacher with a challenge of widely divergent abilities, with some learners demonstrating strong foundational skills while a substantial minority remain unable to decode text at all.

Table 11 Grade 3 Summary Statistics per sub-task

Subtask	Observations	% zero	Score attained at each percentile					Min	Max	items in the subtask
			p10	p25	p50	p75	p90			
HL Letter Sound Knowledge (LSK) clpm	1,946	1.95	20	39.00	56.00	68.24	80.45	0	139	70
HL Word Reading cwpm	1,946	5.45	1	8.00	20.00	28.00	34.00	0	66	60
HL Complex Consonant Diacritic Knowledge clpm	1,946	6.01	2	9.00	23.00	33.00	40.00	0	112	5
HL Oral Reading Fluency cwpm	1,944	15.90	0	15.92	42.12	60.96	78.95	0	142	128
HL Oral Reading Fluency Comprehension (% score)	1,944	18.83	0	0.21	0.41	0.55	0.64	0	1	11
EFAL Word Reading cwpm	1,946	16.19	0	4.00	21.00	37.00	51.00	0	89	80
EFAL Oral Reading Fluency cwpm	1,944	26.03	0	0	28.23	60.00	86.40	0	166.8	72
EFAL Oral Reading Fluency Comprehension (% score)	1,944	29.99	0	0.00	0.30	0.60	0.80	0	1.00	10

4.3.3 Grade 2 and Grade 3 subtest correlation matrix

The Grade 2 HL correlation matrix shows that HL letter sound knowledge is a pivotal foundational skill strongly linking to both HL word reading (0.73) and EFAL word reading (0.66). HL word reading and fluency are also very tightly connected (0.93), and both strongly predict EFAL reading performance (0.85-0.88), showing that home language skills effectively transfer across languages. Comprehension shows moderate relationships with these skills (0.40-0.53) suggesting that it is still developing as a secondary outcome at this stage.

By Grade 3, reading skills become more integrated and sophisticated. HL word reading and fluency remain tightly linked (0.94), while complex consonant diacritic knowledge emerges as

central, correlating strongly with word reading (0.87), fluency (0.84), and basic letter sounds (0.71) showing the importance of handling of complex orthographic features. Cross-language transfer to English stays robust (0.78 to 0.84), with multiple home language skills contributing.

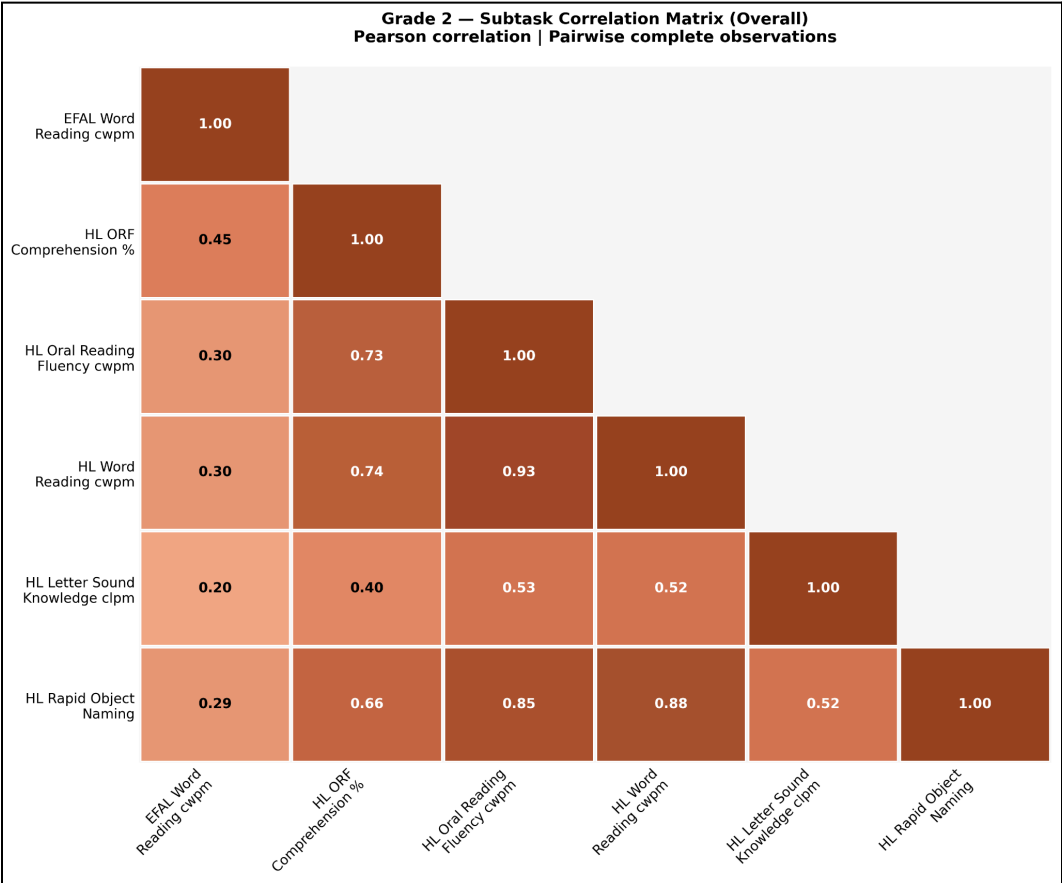


Figure 19 Grade 2 Subtask Correlation Matrix

The most significant shift is in comprehension, which strengthens considerably with correlations to EFAL word reading (0.74), HL fluency (0.66), HL word reading (0.65), and complex diacritics (0.63). This marks a transition from Grade 2's loose connections between the earlier sub-tasks and comprehension to Grade 3's tighter integration, where comprehension increasingly depends on both home language fluency and emerging English decoding abilities.

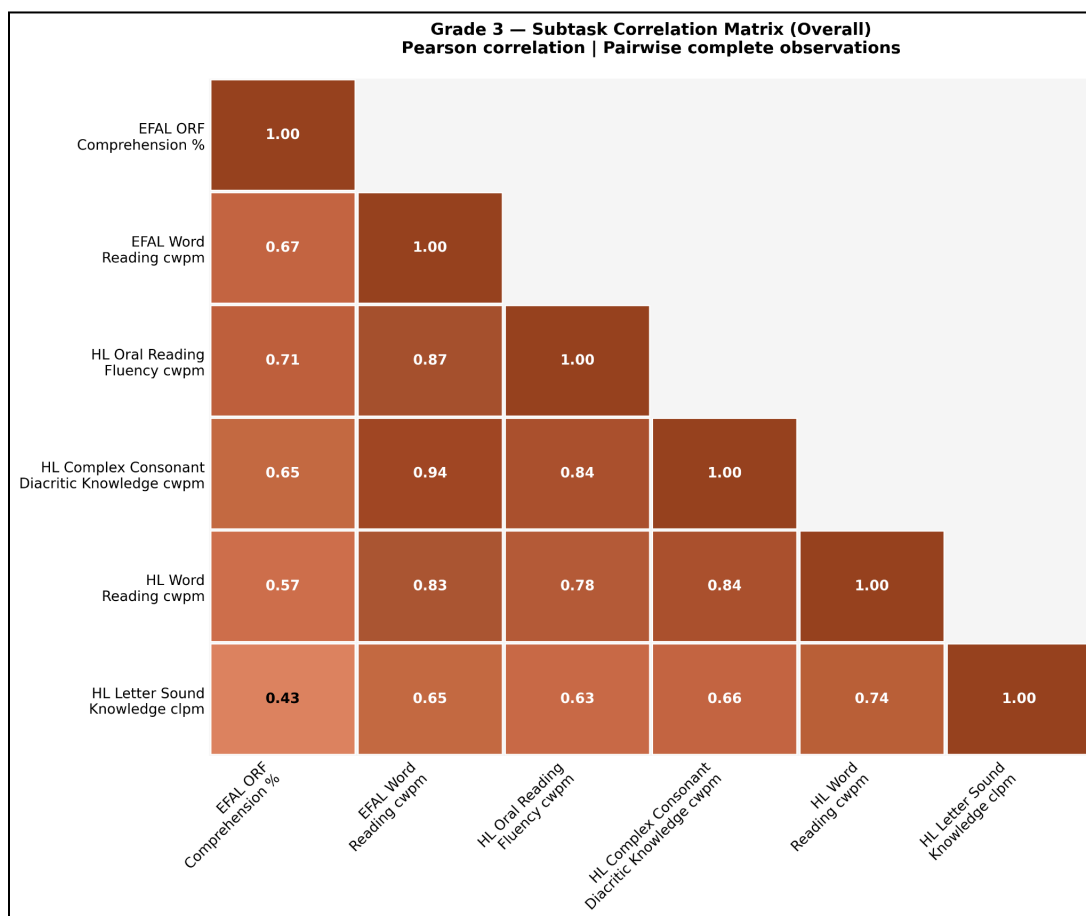


Figure 20 Grade 3 Subtask Correlation Matrix

These patterns suggest that while early Grade 2 interventions should prioritise basic phonics and word reading to build the base (with strong HL-to-EFAL spill over), Grade 3 efforts benefit most from reinforcing complex decoding features and fluent text reading in the home language to support both languages and lift emerging comprehension outcomes in this bilingual context.

4.3.4 Learner Results against Reading Benchmarks

To contextualise learner performance results we compare them to the DBE reading benchmarks for both Setswana and EFAL for Grades 2 and 3. The benchmarks are:

- Grade 2 LSK = used as a Grade 1 benchmark which is 40 correct letters per minute
- Grade 2 HL ORF = 40 cwpm when reading a passage
- Grade 3 HL ORF = 40 cwpm when reading a passage
- Grade 3 EFAL ORF = 50 cwpm when reading a passage

Grade 2 learner benchmark achievements

The analysis done by intervention group and school type shows interesting results for Grade 2 (Figure 21). In general, monograde classrooms outperform multigrade classrooms on Letter Sound Knowledge (64% vs 56% meeting benchmark), confirming the inherent challenges of

teaching different ability levels in one classroom. However, the two contexts respond differently to the introduction of coaching. Multigrade coaching schools show substantial gains with LSK rising from 50% to 63%, while ORF nearly doubles from 18% to 35% and zero scores drop from 47% to 27%. In monograde classrooms, coaching groups perform virtually identically to base groups for both skills (LSK at 63 to 64%, ORF even regressing from 29% to 27%).

This suggests coaching may be particularly beneficial in multigrade contexts, potentially narrowing the achievement gap, while regular instruction in monograde settings may already adequately support literacy development, especially in phonics instruction at Grade 2. The significance of the coaching effect on multigrade versus monograde classrooms is explored further below through regression analyses.

A comparison with other similar studies is useful for interpreting the significance of the EGRP II findings. The FUNS study, which was conducted in monograde schools, reveals that 29% of learners met or exceeded the reading benchmark, a figure comparable to the EGRP II monograde base group, though slightly lower than the monograde coaching group. For the EGRP I, 15% of learners met the benchmark, which is much lower than the EGRP II results. The EGRP II zero scores are similar to the 32% zero scores from the FUNS result, and much lower than the 59% zero scores documented by EGRP I in the North West province.

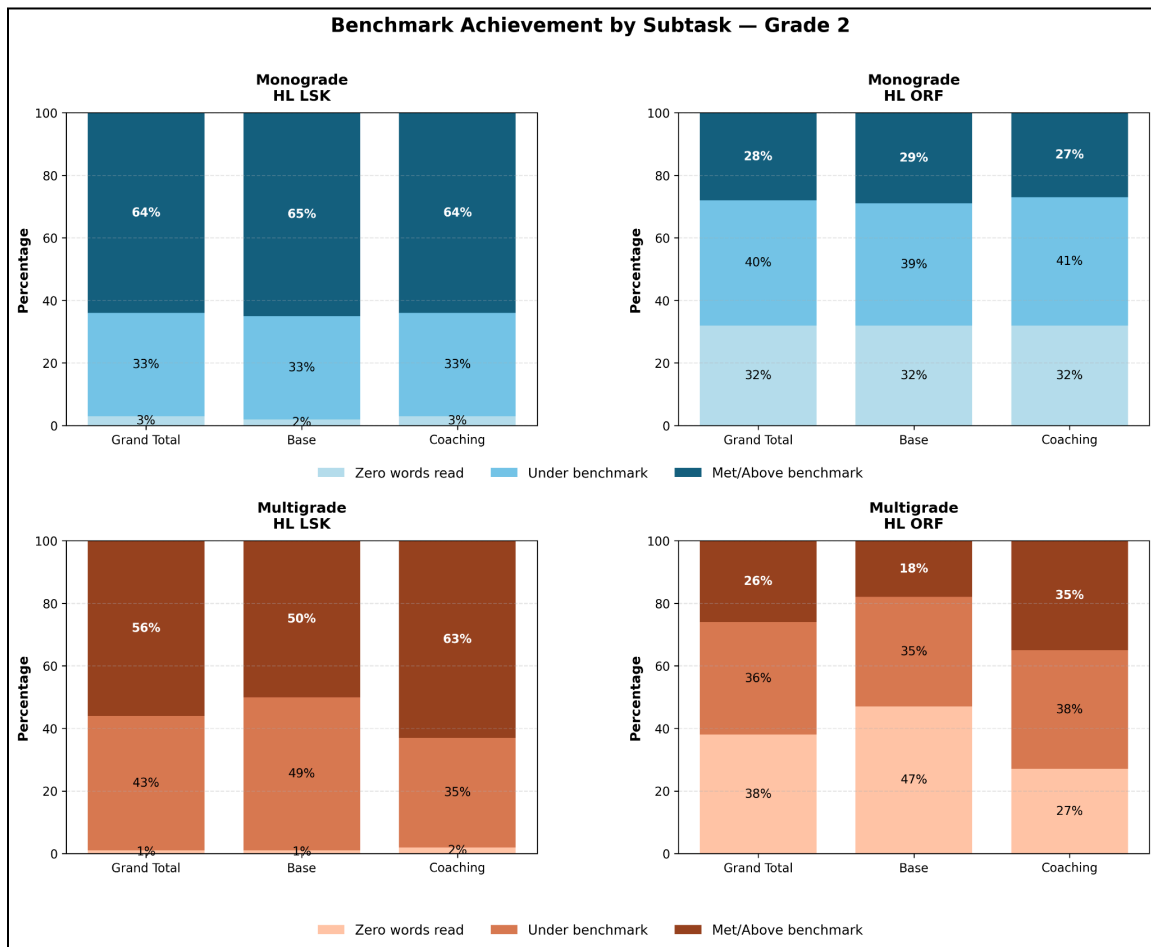


Figure 21 Grade 2 Benchmark Achievement

Grade 3 reading benchmark achievements

A similar analysis of Grade 3 results reveals a consistent pattern (Figure 22). Monograde classrooms maintain their performance advantage, with 74% to 75% meeting the LSK benchmark compared to 68% to 74% in multigrade settings. Coaching again differentiates sharply by context for LSK but not for ORF. In multigrade classrooms, the coaching group LSK zero scores drop from 3% to 0%, however ORF benchmark achievement rates regress from 28% to 23% and zero scores increase by 1%. In monograde schools, ORF for coaching groups remains marginally increased from base to coaching groups from 25% to 29%.

The FUNS comparison shows that 28% of learners met or exceeded the Grade 3 HL ORF benchmark, placing them within a similar range to the EGRP II monograde coaching group (29%). Notably, zero scores were consistent across both studies, with 15% of learners in each failing to read a single word. The EGRP I study achieved stronger results than EGRP II for Grade 3 ORF: EGRP I had two ORF stories in which 32% and 42% of learners met or exceeded the benchmark, compared to the 29% benchmark achievement in EGRP II's single ORF story. However, this was accompanied by higher zero scores of 18% and 17% respectively,

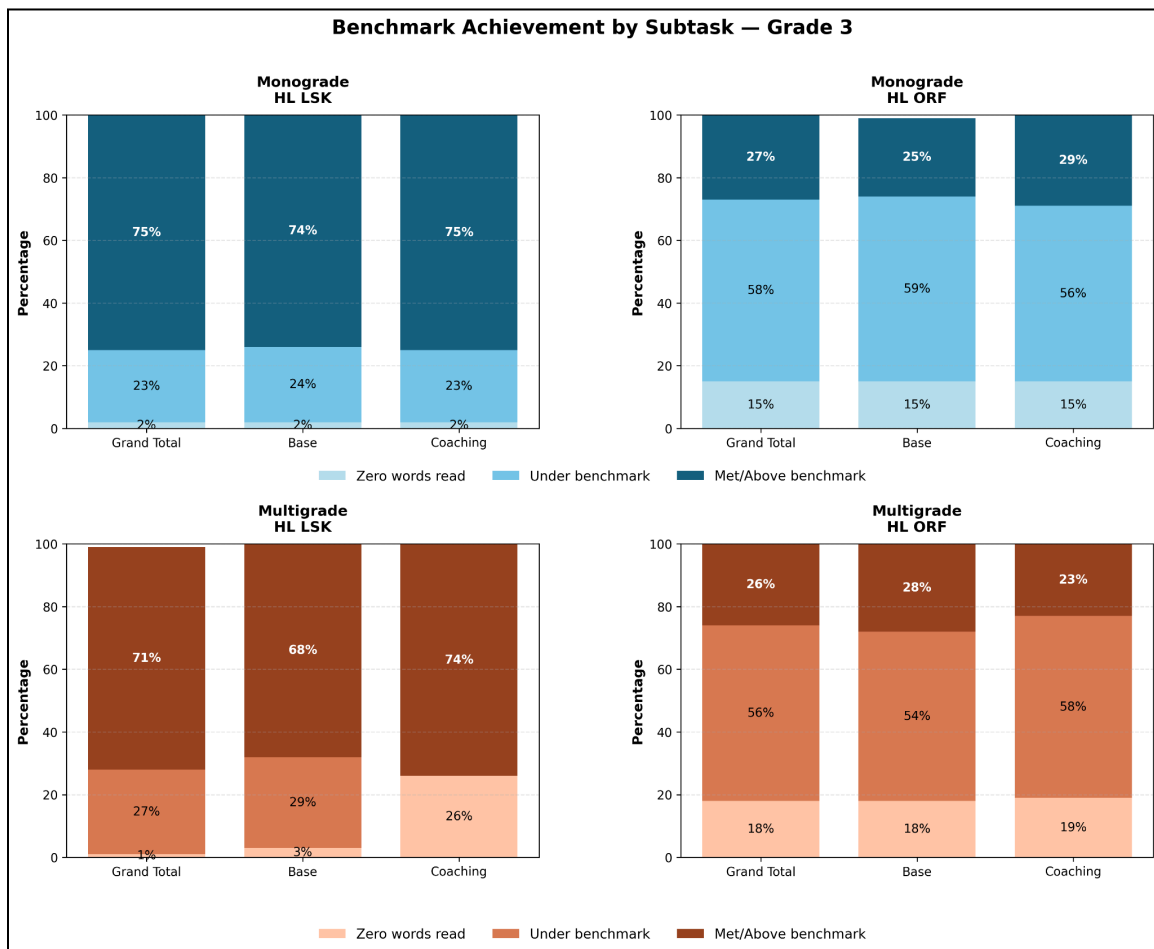


Figure 22: Grade 3 HL Benchmark Achievement

Similar to the preceding HL sub-tasks benchmark analysis, EFAL Word Recognition in Grade 2 shows monograde classrooms with weak performance (12% to 14% meeting benchmark, 27% to 28% zero scores) and negligible coaching differentiation (Figure 22). Multigrade classrooms continue to perform better, with fewer readers who can't read a single word with the coaching group showing notable improvement, reducing non-readers from 41% to 23% and increasing those who meet the benchmark to 14%. This reinforces the pattern that coaching may be particularly beneficial in multigrade contexts for both home language and EFAL literacy development.

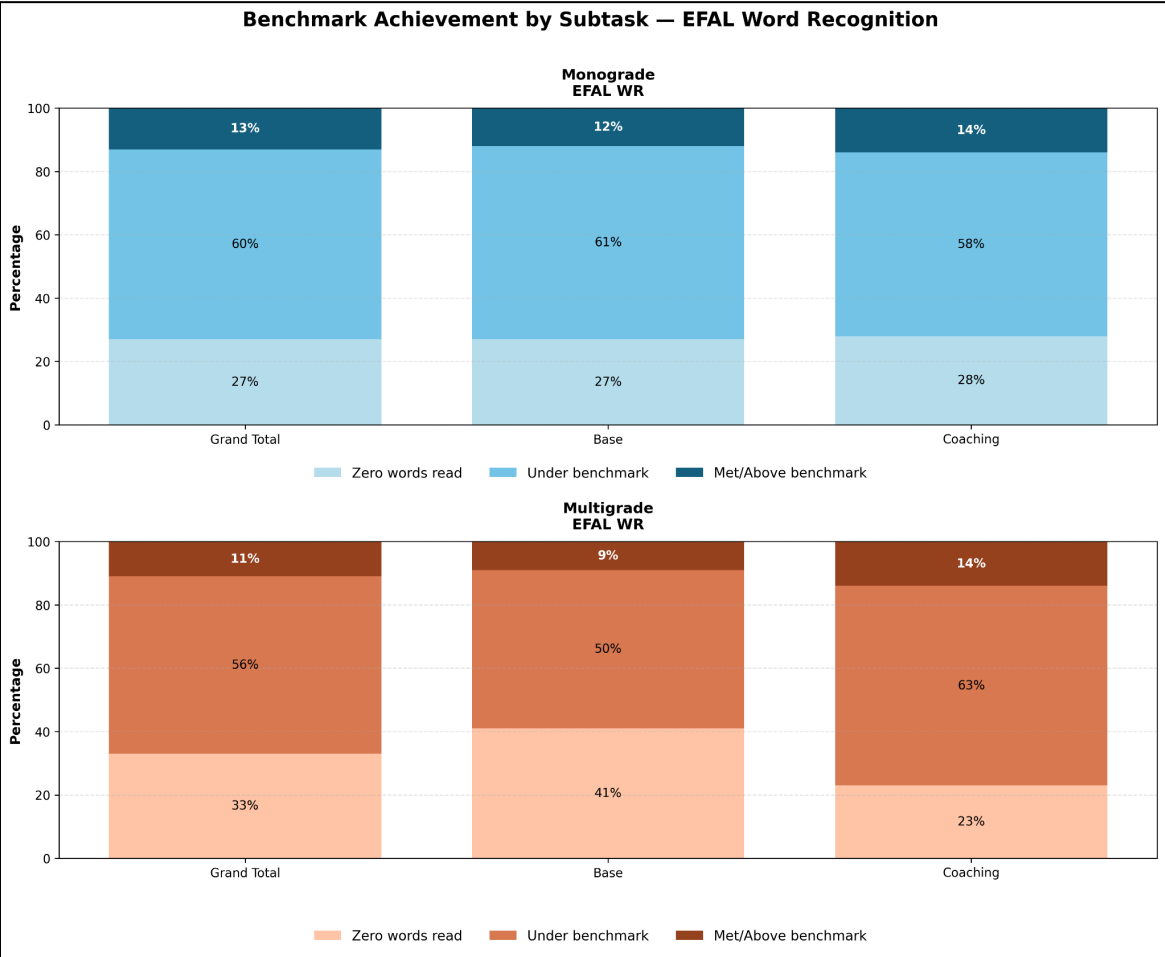


Figure 23: Grade 2 EFAL Benchmark achievement by treatment and school type

Grade 3 benchmarks show no meaningful difference between the monograde and multigrade cohorts, though the multigrade coaching group demonstrates a 5% increase in above-benchmark achievement compared to the base programme. However, this improvement is tempered by a rise in zero scores alongside a reduction in learners falling below the benchmark, suggesting that some learners who previously attempted the assessment dropped off entirely (Figure 24). The EGRP II recorded a higher proportion of learners meeting or

exceeding the benchmark at 36%, though this was accompanied by notably larger zero scores of 22% among monograde classes.

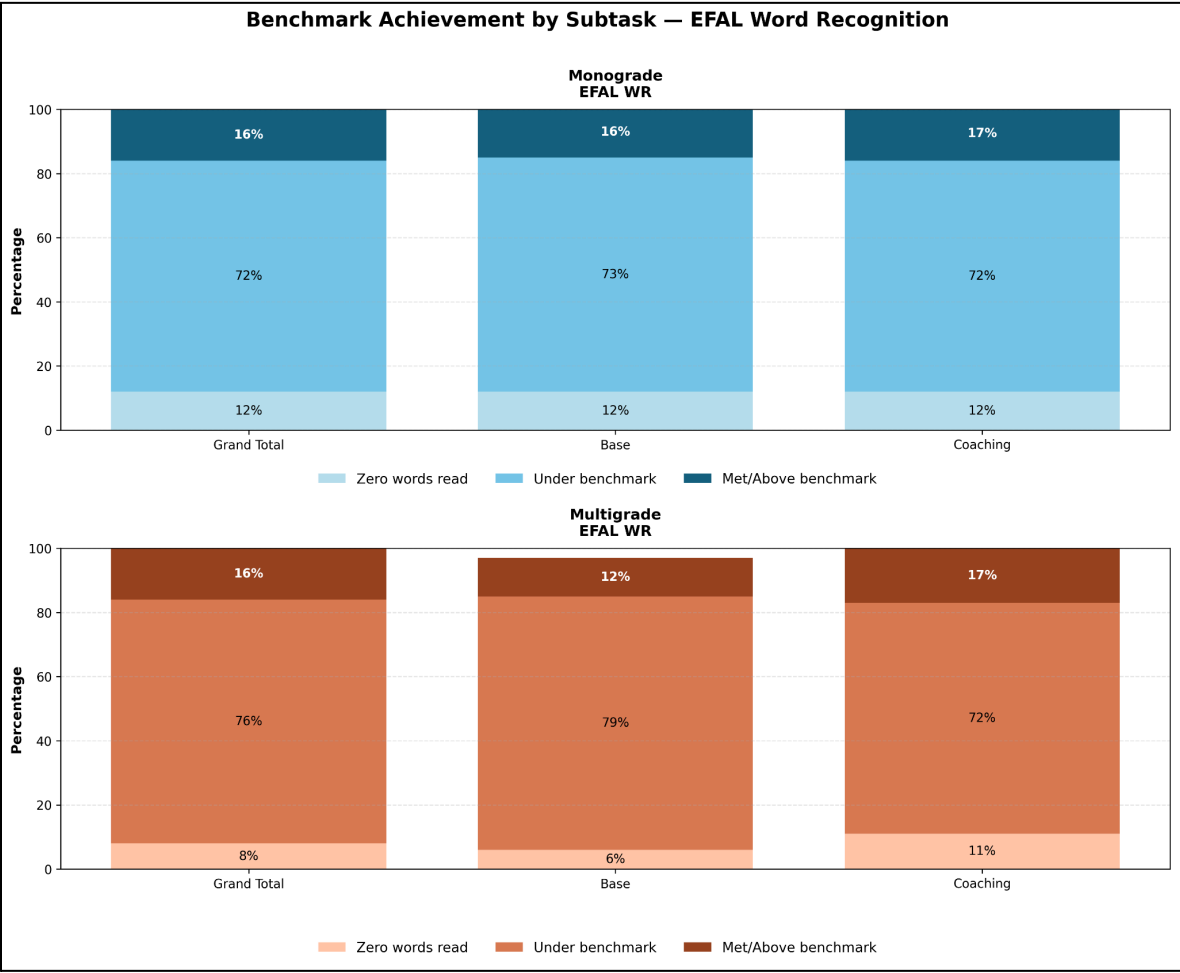


Figure 24: Grade 3 EFAL Benchmark achievement

The preceding benchmark analyses show consistently poor performance, with most learners failing to meet literacy standards. It shows Grade 2 learners have strong phonics knowledge but struggle to transition into fluent reading, with only 27% to 29% meeting oral reading fluency benchmarks despite 62% to 64% meeting letter-sound benchmarks. Coaching seems most effective in multigrade contexts, substantially associated with improving outcomes and reducing non-readers by nearly half, while monograde classrooms show minimal response to coaching.

4.3.5 Distribution of learner test scores

A further analysis of the test score was carried out to see the distribution of composite learner scores for Setswana Home Language for both Grades 2 and 3. Composite learner scores are combined measures that aggregate multiple individual test results into a single overall performance indicator. Instead of examining each reading subtask separately (such as letter sounds, word reading, oral reading fluency, and comprehension), a composite score combines

these elements into one number, providing a holistic view of a learner's overall reading ability rather than isolated skill performance.

For HL, both Grade 2 and Grade 3 show nearly identical patterns between Base and Coaching groups, with most learners clustered around middle-range scores, suggesting coaching did not produce substantially different home language reading outcomes compared to base programme support alone (Figure 26). Grade 3 learners perform slightly better overall than Grade 2, indicating modest improvement as learners progress through grades, though the difference is relatively small. Both Grade 2 and Grade 3 EFAL distributions also show nearly identical patterns between Base and Coaching groups, suggesting coaching did not produce substantially different English reading performance outcomes (Figure 25).

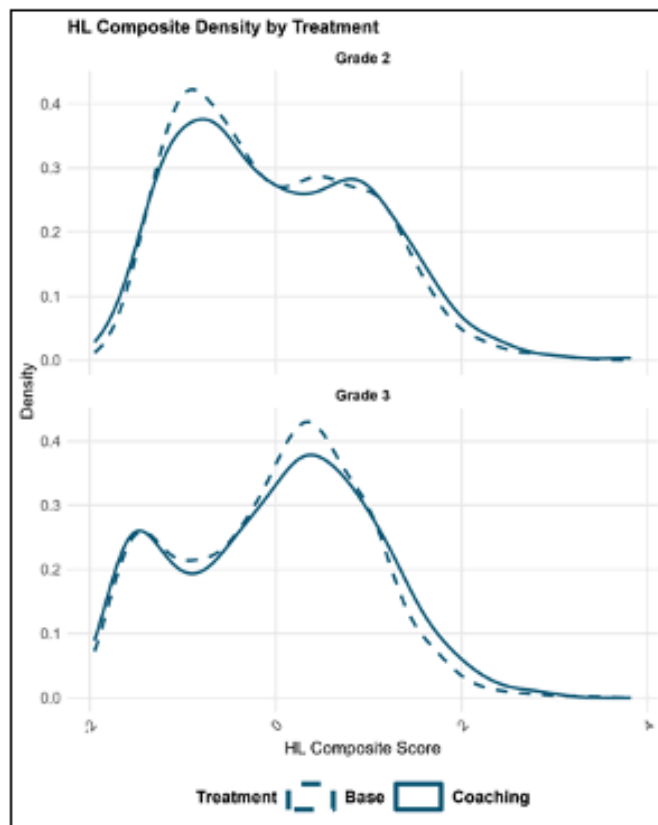


Figure 26: Home Language Composite Score

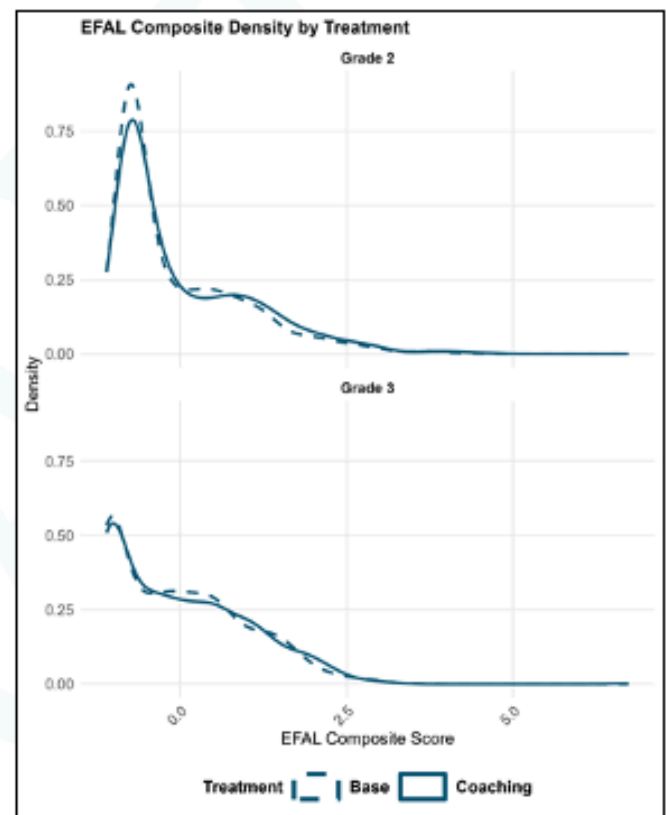


Figure 25: EFAL Composite Score

Unlike home language results, these distributions show most learners concentrated at very low score levels, with a sharp peak near zero and a long tail indicating fewer learners are spread out across higher scores. Grade 3 shows slight improvement over Grade 2, but both grades show the concerning pattern of most learners struggling with EFAL reading proficiency.

When analysed by school type (Figure 27), small differences emerge between Base and Coaching groups with Grade 3 showing slight differences suggesting positive effects from coaching, while Grade 2 multigrade coaching (rust line) shows fewer learners at low scores and

more at high scores, though the distribution remains bimodal, with two distinct performance groups.

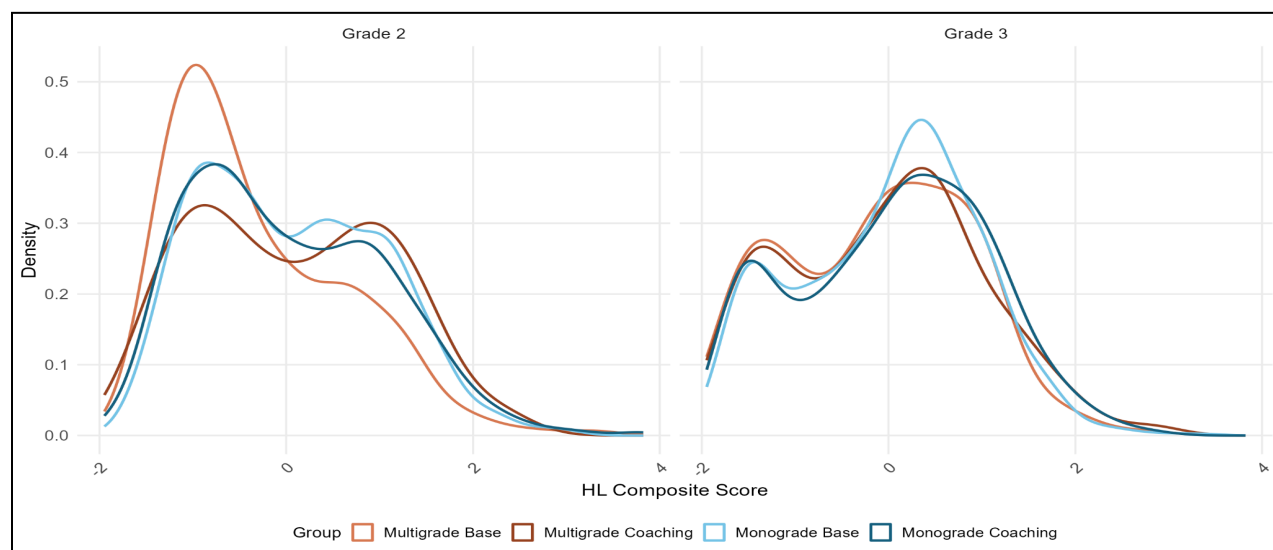


Figure 27 HL Composite Density by treatment and school type

Differences between multigrade and monograde learners in the coaching cohort become more pronounced when analysed by school type (Figure 28) **suggesting coaching may have a stronger influence in multigrade schools than in monograde schools.**

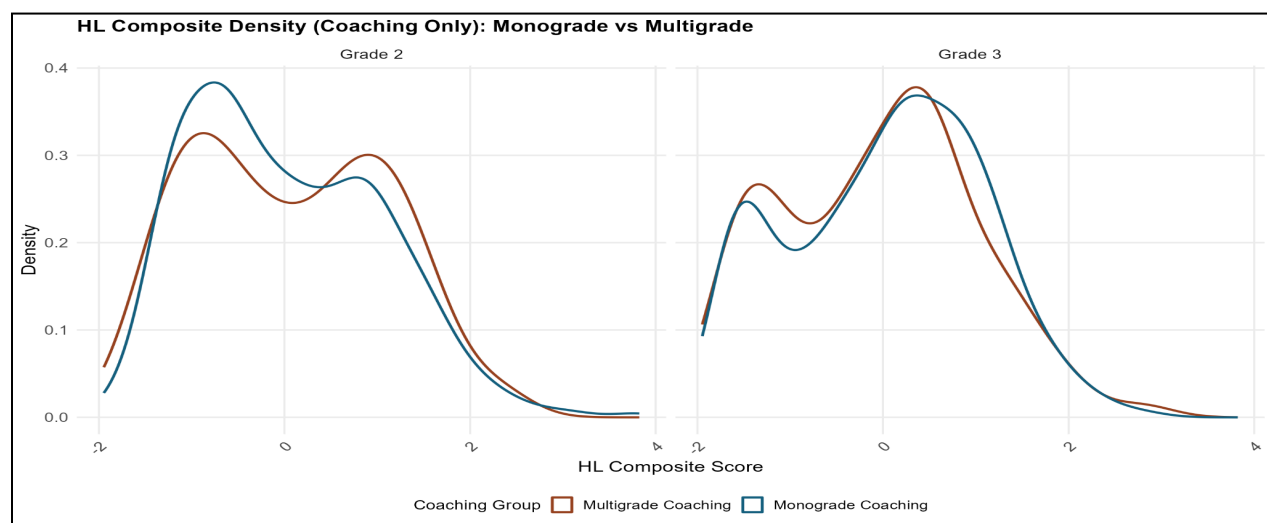


Figure 28: HL Composite Density (coaching only): Monograde vs Multigrade

Further analysis of Grade 2 results at the sub-task level, comparing mono- and multigrade learners who were coached, shows a slight decrease in zero scores for LSK and larger reductions in zero scores for HL WR and ORF and EFAL WR, suggesting coaching may have had an effect (Figures 29 to 32).

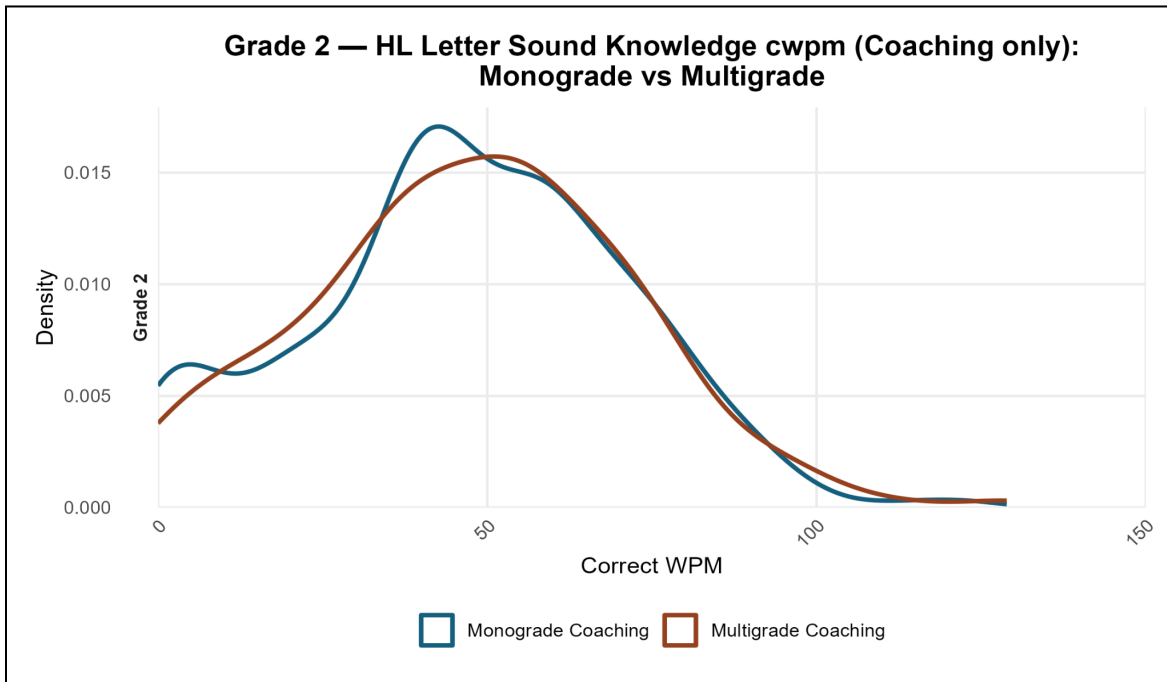


Figure 29: HL Letter Sound Reading (coaching only): Monograde vs Multigrade

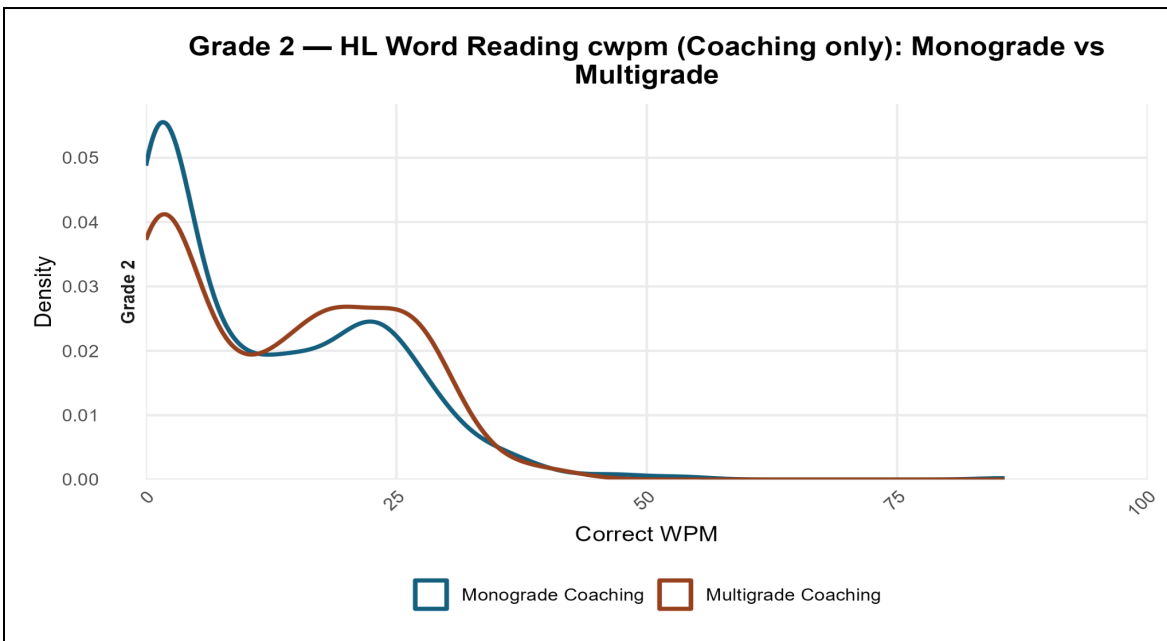


Figure 30: HL Word Reading (coaching only): Monograde vs Multigrade

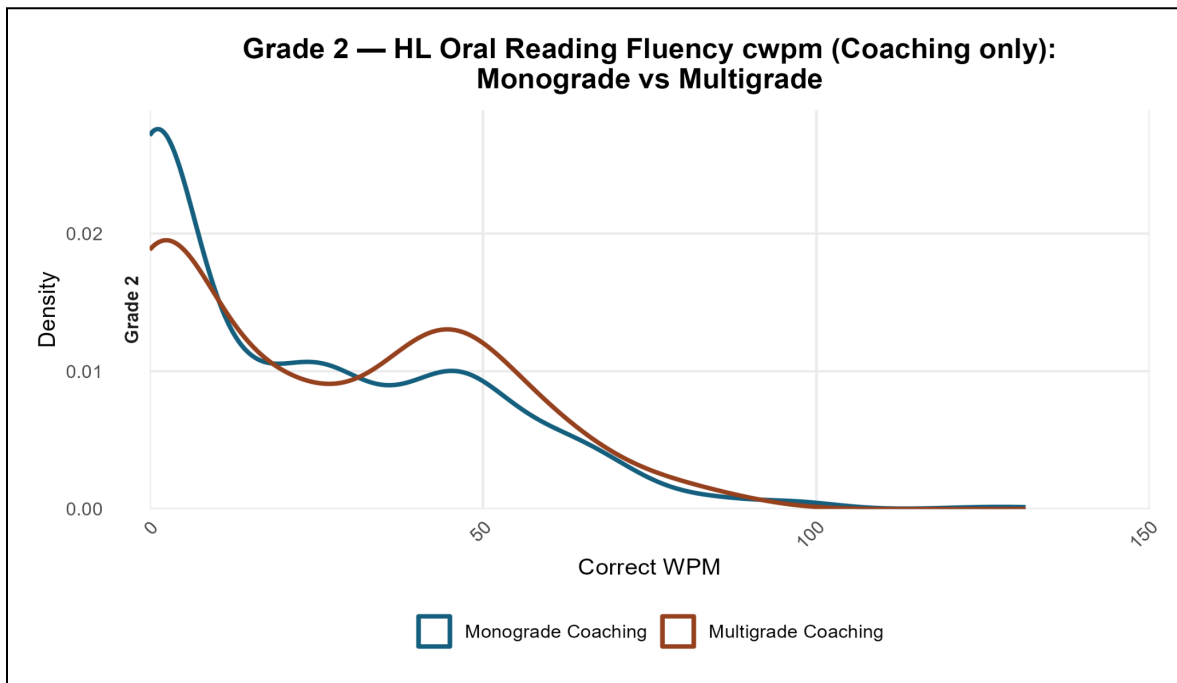


Figure 31: Grade 2 HL ORF (coaching only) Monograde vs Multigrade

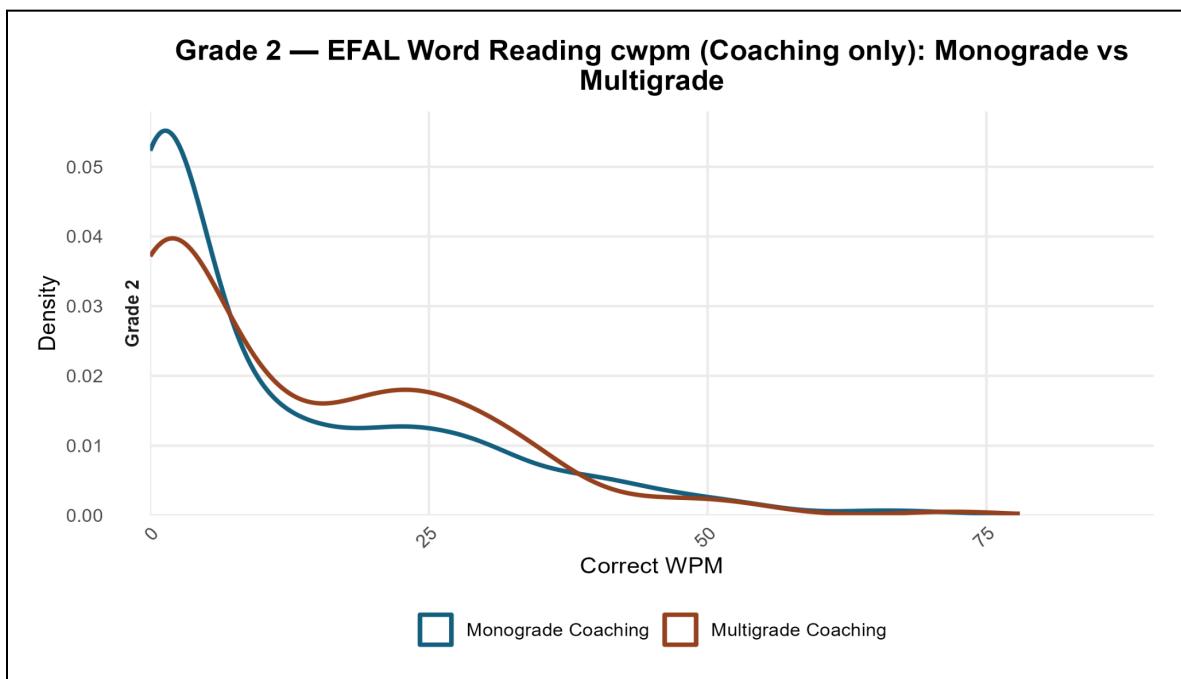


Figure 32: Grade 2 EFAL Word Reading cwpm (coaching only): Monograde vs Multigrade

In Grade 3, however, there were no significant differences between mono- and multigrade learners in the coaching stream, with nearly identical density curves. This suggests that

coaching had no effect on HL letter sound knowledge, word reading, oral reading fluency or comprehension (Figure 33.1 and 33.2). For EFAL, Grade 3 monograde learners performed slightly better, but not significantly so (Figure 34).

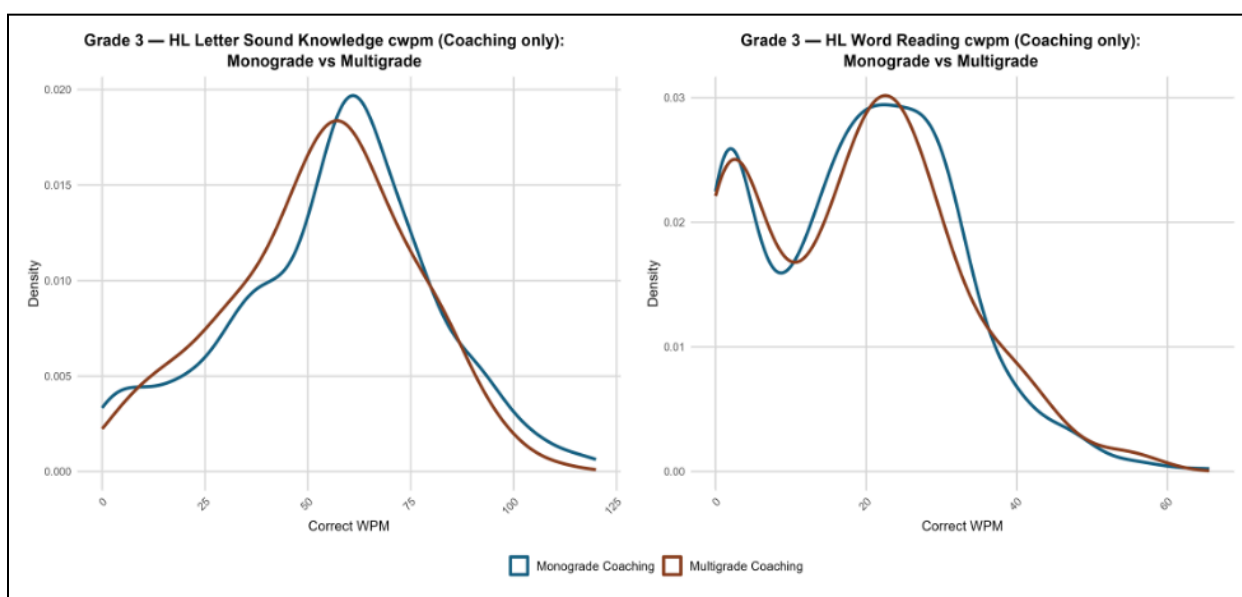


Figure 33.1: Grade 3 HL Subtask Densities LSK & WR (coaching only): Monograde vs Multigrade

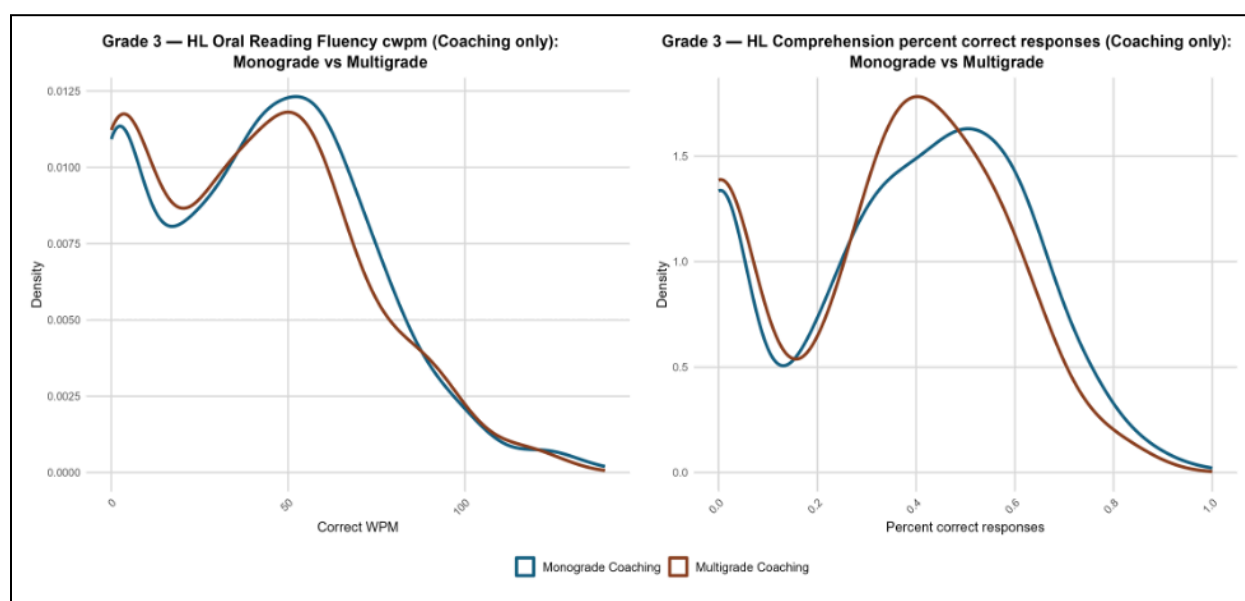


Figure 33.2: Grade 3 HL Subtask Densities ORF & Comprehension⁵ (coaching only): Monograde vs Multigrade

⁵ Learners who scored 0 for oral reading fluency (could not read a single word) were not asked the comprehension questions. However, to ensure the inclusion of the full learner sample in the comprehension results, the graphs in this report recode those learners as having a zero-score and combine them with learners who were asked the comprehension questions and answered 0 correctly.

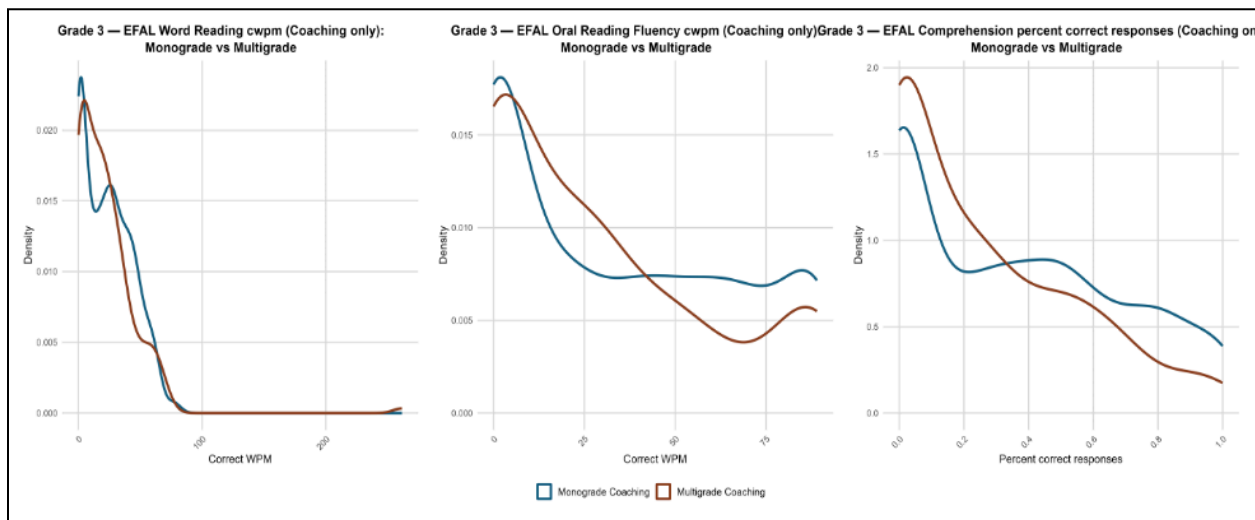


Figure 34: Grade 3 EFAL Subtask Densities (coaching only): Monograde vs Multigrade

4.4.3 Comparison of intervention streams

As part of the comparison of intervention streams, it is important to establish that the findings for each stream are reliable and consistent before drawing conclusions about their relative effectiveness. To further confirm the reliability of the preceding findings, a series of additional checks were carried out.

Robustness and Sensitivity Checks

A sensitivity analysis was conducted by removing schools with unusually high or low performance from the analysis to test whether the overall findings were being influenced by a small number of exceptional cases. This did not meaningfully change the results, as treatment effects remained statistically significant, confirming that the findings reflect a genuine and consistent pattern. The analysis was repeated using different ways of measuring learning outcomes (standardised and composite outcome models), and the results remained consistent across all approaches, including the finding that the programme did not work equally well across all schools, further supporting the conclusion that the findings are well-founded. The analysis also used statistical methods that accounted for the way learners are organised within schools, adding to the reliability of the results.

Balance Tests

To ensure that comparisons between intervention streams are valid, balance tests were conducted to check whether the group of schools that received the intervention and the group that did not were comparable in terms of age, Setswana as home language, gender, access to books at home, and homework assistance at home. The two groups were found to be very similar across all these characteristics, confirming that the random assignment worked as intended i.e. the base and coaching treatments had similar learner profiles and were thus comparable. While the study does not have a baseline measure, this similarity between the two

groups suggests that observed differences in outcomes are more likely to reflect the effect of the programme than pre-existing differences between the groups. This conclusion is nonetheless drawn cautiously, recognising that the absence of a baseline limits the strength of any causal claims.

Composite Outcomes

Performance across multiple literacy areas was combined into a single overall score and the analysis was repeated using this composite measure. This was done to test whether the findings remained consistent when looking at literacy more broadly rather than through individual measures alone.

Two aggregate performance scores were generated using principal component analysis (PCA), one for Setswana home language and one for EFAL.⁶ For home language, the PCA was conducted across all relevant subtasks, with oral reading fluency and word reading emerging as the most compatible components. Since only one EFAL subtask was available, it was standardised and used directly as the outcome variable.

4.4.3.1. Grade 2 Home Language and EFAL Composite scores

Multigrade schools without coaching performed significantly below monograde schools across most reading skills in Grade 2, with gaps of 5.98 cwpm in HL oral reading fluency, 13.63 cwpm in letter sound knowledge, 5.36 cwpm in word reading, and 10.19 cwpm in oral reading fluency. Similarly in EFAL, multigrade base programme learners were 5.53 cwpm below monograde base programme learners in word reading. These findings confirm the persistent performance disadvantage faced by multigrade schools.

Where coaching was introduced in multigrade schools, however, learners showed meaningful and statistically significant improvements in fluency and accuracy in both their home language and English. **Learners in coached multigrade schools significantly outperformed their counterparts in non-coached multigrade schools.** Four sub-tasks showed statistically significant differences (Table 12).

- HL Letter Sound Knowledge (+7.19 cwpm),
- HL Word Reading (+4.49 cwpm),
- HL Oral Reading Fluency (+9.41 cwpm), and
- EFAL Word Reading (+5.52 cwpm).

⁶ Before combining the subtasks, robustness checks were carried out to identify any floor or ceiling effects that could distort the results. No ceiling effects were detected, however floor effects were present in both Grade 2 and Grade 3. In Grade 2, these were found in home language oral reading fluency and EFAL word reading. In Grade 3, floor effects were present in home language oral reading fluency, EFAL word reading, and EFAL comprehension. These floor effects were not considered significant enough to invalidate the results and so all items were included in the analysis.

Table 12 Grade 2 Setswana HL and EFAL Regression Results (mean scores)

Grade 2 Sub-tasks	Coaching effect on Monograde Schools			Coaching effect on Multigrade schools			R ²
	Mean of Base	Coaching Effect	Mean of Coaching	Mean of Base	Coaching Effect	Mean of Coaching	
HL Rapid Object Naming (cwpm)	48.07	0.27	48.34	42.09	-1.25	40.84	0.033
HL Letter Sound Knowledge (cwpm)	50.18	-0.92	49.26	36.55	7.19*	43.74	0.037
HL Word Reading (cwpm)	13.26	-0.68	12.59	7.90	4.49**	12.39	0.031
HL Oral Reading Fluency (cwpm)	26.07	-0.89	25.18	15.88	9.41**	25.29	0.026
EFAL Word Reading (cwpm)	12.24	0.04	12.27	6.71	5.52**	12.23	0.026
Oral Reading Comprehension (% score)	0.26	-0.01	0.26	0.16	0.10***	0.27	0.028

*Note: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Stratum controls included but not shown. Scores are expressed in correct words per minute (cwpm), except Reading Comprehension which is a proportion score.⁷*

Coaching in monograde schools, however, produced no significant differences across any reading sub-task, with effects close to zero throughout. The coaching programme, therefore, had its greatest effect in the multigrade setting, where the structured support it provided appears to have made a tangible difference to learner outcomes.

The table below shows the same results as above but in standard deviations. This standardises everything onto the same scale, making it easier to compare the size of effects across different reading skills that were measured differently. It also allows comparisons with other studies that report their findings in the same way.

The table shows that multigrade schools performed between 0.4 and 0.6 standard deviations below monograde schools, confirming their well-documented performance disadvantage. Where coaching was introduced, however, learners showed meaningful gains of between 0.35 and 0.47 standard deviations across home language word reading, reading fluency, and English word reading, suggesting that the programme helped narrow this gap. Compared with other education interventions studied globally, an effect size of 0.35 and above is considered to be large (Hill, et al 2028, Kraft 2020).

⁷ The R² values per item are low, ranging from 0.026 to 0.037. The relatively low R² values observed in these regressions are consistent with prior education and randomized controlled trial research. In randomized experiments, causal identification relies on random assignment rather than model explanatory power, and low R² values are common and expected (Angrist & Pischke, 2009; Duflo et al., 2007). Previous education RCTs report similar explanatory power, reflecting substantial individual-level variation in learning outcomes. Therefore, low R² values do not undermine the validity of causal treatment effect estimates.

Table 13 Grade 2 Setswana HL and EFAL Regression Results (z-scores)

	HL Rapid Object Naming	HL Letter Sound Knowledge clpm	HL Word Reading cwpm	HL Reading Fluency (Story 2) cwpm	EFAL Word Reading cwpm	HL Reading Comprehension
Coaching in monograde schools (treatment)	0.019	-0.0039	-0.062	-0.037	0.003	-0.030
Multigrade School (no coaching)	-0.434**	-0.579**	-0.488*	-0.425	-0.388	-0.379
Coaching in Multigrade Schools	-0.110	0.345*	0.471**	0.430**	0.386**	0.460*
No. of Learners	1851	1851	1851	1847	1850	1847

*Note: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Stratum controls included but not shown.⁸ Scores are standardised to allow comparison across skills.*

It is furthermore important to note that the size of the coaching effect for multigrade learners enabled them to almost 'catch up' with average monograde learner performance. If taking an **equity perspective** and compare mean results per item to show the differences between (rather than within) intervention groups (Table 14) - e.g. comparing multigrade base programme with monograde base programme learners and multigrade coaching with monograde coaching learners - we see that the gap between multi and monograde learners was greatly reduced across all items except Rapid Object Naming, almost achieving parity for HL word reading, HL ORF, EFAL word reading and HL oral reading comprehension. While this finding does not address the underlying challenge of low reading skills in monograde classrooms, it does very importantly show that performance inequities between multi- and monograde classrooms can be addressed.

Table 14 Grade 2 Setswana HL and EFAL Regression Results (mean scores)

Grade 2 Sub-tasks	Base Programme Comparison of Monograde and Multigrade Schools			Coaching Programme Comparison of Monograde and Multigrade Schools		
	Mean of Mono Base	Mean of Multi Base	Difference	Mean of Mono Coached	Mean of Multi Coached	Difference
HL Rapid Object Naming (cwpm)	48.07	42.09	-5.98	48.34	40.84	-7.5
HL Letter Sound Knowledge (cwpm)	50.18	36.55	-13.63	49.26	43.74	-5.52
HL Word Reading (cwpm)	13.26	7.90	-5.36	12.59	12.39	-0.2

⁸ The stratum controls were included in the model as robustness checks to account for the necessary design effects that lead to differences between the geographic and administrative areas in which schools are located. These are included as controls for the regression to ensure that any performance differences between areas do not distort the programme's results. The inclusion of strata do not change the overall finding that coaching made a meaningful difference in multigrade schools.

HL Oral Reading Fluency (cwpm)	26.07	15.88	-10,19	25.18	25.29	-0.11
EFAL Word Reading (cwpm)	12.24	6.71	-5.52	12.27	12.23	-0.04
Oral Reading Comprehension (% score)	0.26	0.27	-0.01	0.25	0.27	-0.02

*Note: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Stratum controls included but not shown. Scores are expressed in correct words per minute (cwpm), except Reading Comprehension which is a proportion score.*

4.4.3.2. Grade 3 Home Language and EFAL Composite scores

Grade 3 results tell a markedly different story from Grade 2, with coaching in multigrade schools producing no statistically significant differences across any reading skill. The effects are small and inconsistent in direction, with some skills showing marginal gains and others slight declines, none of which are meaningful or reliable (statistically significant).

Interestingly, coaching in monograde schools showed a significant improvement in Letter Sound Knowledge (+4.49 cwpm), which was not the case in Grade 2. However, this was the only significant finding and did not extend to other reading skills.⁹

Unlike Grade 2, multigrade schools at Grade 3 did not show a statistically significant performance disadvantage compared to monograde schools across any skill, whether in the base or the coaching programme, suggesting that the gap between school types may narrow as learners progress through the grades.

Overall, the Grade 3 findings suggest that the coaching programme had little measurable effects at this level, raising questions about whether the intervention needs to be adapted to meet the needs of older learners.

Table 15 Grade 3 Setswana HL and EFAL Regression Results (Raw Scores)

⁹ As for the Grade 2 results, the R^2 values of Grade 3 results remain low across most skills, ranging from 0.020 to 0.064, consistent with the complexity of factors influencing learner performance. The higher R^2 for Letter Sound Knowledge and Reading Comprehension suggests these skills may be better explained by the factors captured in the model.

Grade 3	Coaching effect on Monograde Schools			Coaching effect on Multigrade schools			R ²
Outcome correct words per minute total score	Mean of Base	Coaching Effect	Mean of Coaching	Mean of Base	Coaching Effect	Mean of Coaching	
HL Letter Sound Knowledge	52.45	4.49*	56.94	48.22	2.95	51.17	0.064
HL CCDK	21.43	0.19	21.62	20.81	1.33	22.14	0.022
HL Word Reading	19.05	0.32	19.37	18.07	0.92	18.99	0.020
HL Oral Reading Fluency	40.93	1.48	42.41	41.64	-1.65	39.99	0.029
HL Oral Reading Comprehension (% score)	0.35***	-0.003	0.35	0.34	-0.015	0.33	0.027
EFAL Word Reading	21.69	0.63	22.32	21.38	1.76	23.14	0.022
EFAL Oral Reading Fluency	32.87	0.25	33.12	34.27	-1.63	32.64	0.025
EFAL Oral Reading Comprehension (% score)	0.27***	-0.01	0.27	0.29	-0.01	0.27	0.057

Note: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Stratum controls included but not shown. Scores are expressed in correct words per minute (cwpm), except Reading Comprehension which is a proportion score.

As for Grade 2, the table below shows the same results as above but expressed as standard deviations. This confirms the lack of significant effects, apart from HL LSK, and shows the very small effect sizes for all items.

Table 16 Grade 3 Setswana HL and EFAL Regression Results (z score)

	HL LSK	HL CCDK	HL Word Reading cwpm	HL Oral Reading Fluency cwpm	HL Comprehension	EFAL Word Reading cwpm	EFAL ORF 1	EFAL Comprehension
Coaching in monograde schools (treatment)	0.191*	0.013	0.026	0.050	-0.014	0.031	0.007	-0.018
Multigrade School (no coaching)	-0.180	-0.043	-0.080	0.024	-0.040	-0.015	0.040	0.063
Coaching in Multigrade Schools	-0.065	0.079	0.049	-0.106	-0.066	0.056	-0.054	-0.044
No. of Learners	1847	1847	1847	1845	1845	1847	1845	1845
R ²	0.064	0.022	0.020	0.029	0.027	0.022	0.025	0.057

*Note: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Stratum controls included but not shown. Scores are standardised to allow comparison across skills.*

4.4.3.3 Distribution of effects by learner performance

Quantile regressions were run for Grades 2 and 3 learner results to explore whether the programme affected learners differently depending on their performance level.¹⁰ Quantile regression analysis shows that there is an improvement in the middle group of performers in the Grade 2 multigrade coached vs non-coached schools. No significant differences were observed in Grade 3 performance groups.

The quantile regression plot in Figure 35 & 36 below for Grade 2 shows the treatment effect of coaching in multigrade schools for HL and EFAL. For Grade 2 Setswana, the intervention showed a small positive effect even for the weakest learners, roughly the bottom 10 to 25% of the class, though this effect was modest and not conclusive. The benefit grew steadily for average performers, those in the middle 25% to 75% of the class, reaching its peak for learners performing around the 65th to 70th percentile, meaning those doing better than about two thirds of their peers. The EFAL results follow a similar pattern, with one key difference being that the intervention showed almost no benefit for the weakest learners in the bottom 25% of the class, only becoming meaningful for those performing around the middle of the class and above.

The programme thus had its strongest positive effect on middle to upper-middle performing learners, while offering comparatively less benefit to the lowest performers. Notably, the confidence intervals remain wide throughout, indicating some uncertainty in the estimates, though the positive trend across the middle quantiles is consistent with the finding that coaching differentially benefited learners who were neither the weakest nor the strongest in Grade 2.

The programme has a solid foundation to build on, but may need to be adapted or supplemented with more targeted support strategies to ensure that the most vulnerable learners are not left behind.

¹⁰ The quantile regressions for the multigrade learners were run using two approaches. The first used the same approach as for the monograde treatment group, including all sample strata as controls. The second approach used only strata 1, 2 and 3 as all multigrade schools fell into these strata. The results for both methods were substantively similar, with the reduced-strata model shows slightly higher effects for better-performing learners (top quintiles), while the full-strata model showed effects concentrated in both middle and high-performing quintiles.

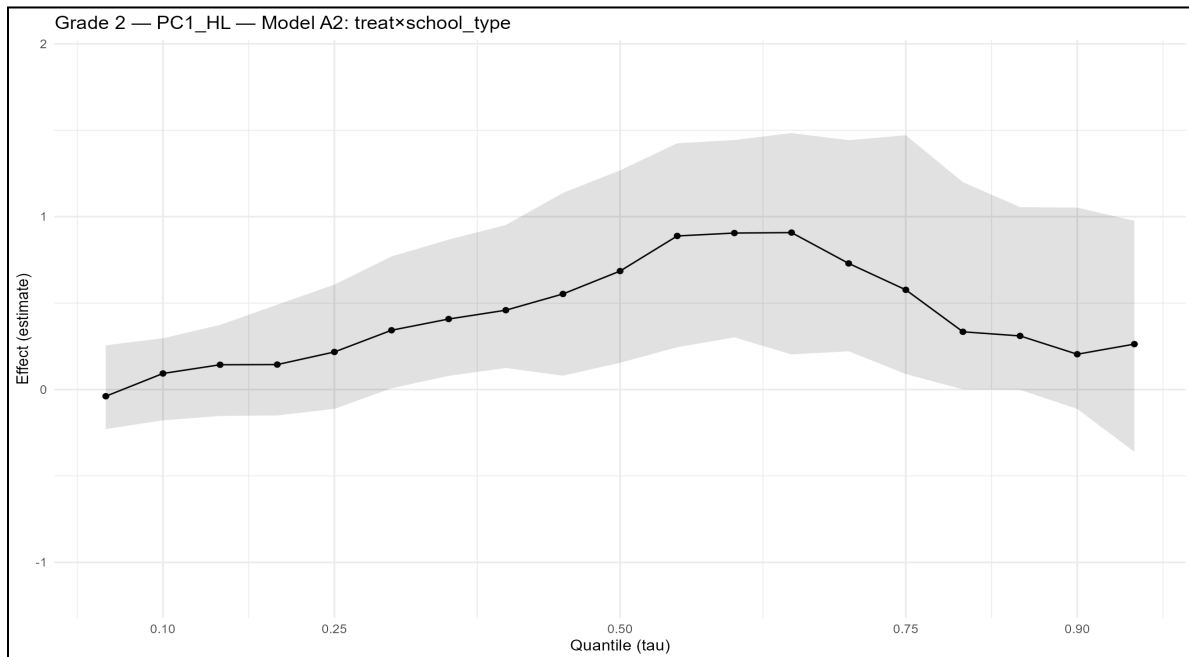


Figure 35: Grade 2 HL Quantile Plot

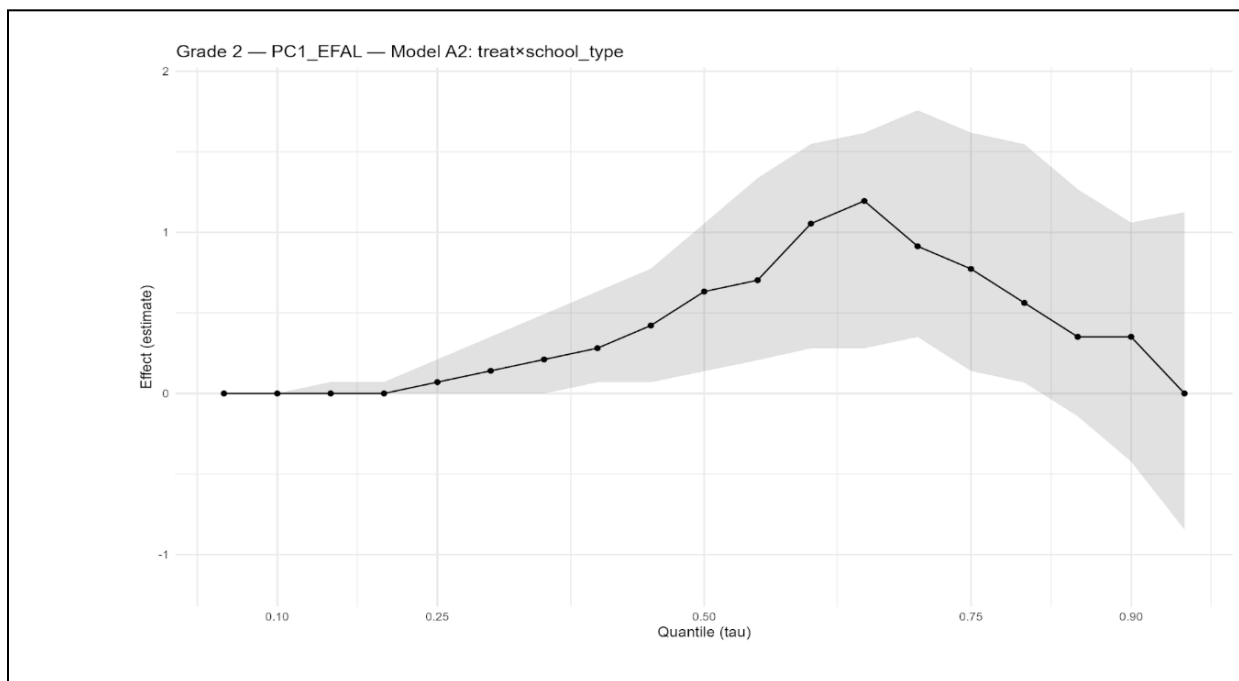


Figure 36: Grade 2 EFAL Quantile Plot

4.4.3.4 Interpretation of learner outcomes

It is important to note that multigrade schools faced implementation challenges from the outset, with customised materials and training only arriving in mid to late 2024, providing approximately one academic year for the intervention to take effect. These delays raised concerns about whether meaningful outcomes could be realised within such a compressed timeframe. Despite this challenging start, the findings reveal encouraging results, with **the most substantial learner outcomes occurring among Grade 2 learners in multigrade schools**. Several explanations may account for this positive finding, including:

- a. **Growth from a lower base:** Multigrade teachers represented an underserved priority within the education system. No dedicated intervention for multigrade teachers has been implemented in the Northern Cape since the District Development Support Project ended over twenty years ago. The specialised attention and resources provided by EGRP II may have generated heightened motivation and engagement to receive context-specific support tailored to their unique instructional challenges. This recognition and investment in their practice likely contributed to stronger implementation and enhanced instructional quality.
- b. **Larger coaching dosage:** multigrade teachers received more coaching support than monograde teachers. Although differential dosage was not part of the original intervention design (which foresaw equal coaching dosage across school types), and may have impacted negatively on the evaluation's ability to measure coaching effects in the Northern Cape's monograde context, it may have been an appropriate adaptation given the programme focus on multigrade schools as a previously under-supported school type. Coach prioritisation of multigrade schools, which they acknowledged finding more complex to support, meant that their most intensive efforts were concentrated where the need for specialised guidance was greatest.
- c. **Grade 1 advantage:** since the EGRP II was a two-year programme, the Grade 2 learners at the endline had been exposed to quality teaching support since Grade 1, while the Grade 3 learners did not have the benefit of supported teaching in this core foundational learning year. This may have impacted on the cumulative learning journey.
- d. **Grade 2 learners may have benefited disproportionately due to typical attention patterns in multigrade classrooms.** Teachers managing Grades 2 and 3 often prioritise Grade 3 learners transitioning to the next phase, while those teaching Grades 1, 2, and 3 typically focus on Grade 1 learners requiring intensive foundational support and Grade 3s preparing for transition. In both configurations, Grade 2 learners often receive less direct attention. The EGRP II training emphasised ability-based grouping and differentiated instruction, which may have disproportionately benefited Grade 2 learners, particularly the midrange to better performing ones (a claim partly confirmed by the quantile regressions analysis), as they gained increased teacher engagement and exposure to Grade 3 content, addressing a previously underserved group within multigrade contexts.

Perhaps the most surprising result is that after intensive coaching support, Grade 2 learners in multigrade coached classes performed at a higher level in reading than their peers in multigrade base schools, which was expected, and those in monograde schools, both coached and base, which was not anticipated. Deep diving into this difference indicates that teachers in multigrade coaching schools responded particularly well to GGR and Shared Reading protocols. It seems that implementing ability-based grouping with these reading protocols made a substantive difference to their teaching and so to improving learner performance.

The lack of differential progress in coached monograde schools compared to base schools may be attributed to the compressed implementation timeframe. Previous iterations of EGRS and EGRP demonstrated that in-class support by professional coaches produces meaningful impact on learner performance, suggesting that longer exposure may be necessary for effects to materialise in monograde contexts. Additionally, coaches may have prioritised multigrade classes, a claim partly supported by implementation fidelity findings, which they acknowledged finding more challenging to support, potentially concentrating their most intensive efforts and innovative strategies in these complex settings where the need for specialised guidance was greatest. This may also reflect coaches understanding the strategic focus of the EGRP II programme on multigrade classes and therefore focusing their attention on this group.

All learner performance data should be understood within the broader context of literacy challenges across monograde and multigrade schools in the two Northern Cape districts. A substantial proportion of Grade 2 learners in base multigrade schools (47%) cannot yet read a single word of grade-appropriate text in their home language, while 41% face similar challenges in EFAL. In Grade 2 monograde schools, these proportions are 32% and 27% respectively. While the percentage of emergent readers decreases somewhat in Grade 3 base schools, with approximately 25% to 30% of learners not yet able to read grade-appropriate text in both monograde and multigrade classes, this remains a significant concern. These foundational literacy gaps help contextualise Northern Cape's declining matriculation results over the past two decades, which have moved from leading the country to requiring intensive support.

Finally, these learner reading results cannot be fully interpreted due to the absence of baseline data. Without knowing how learners were performing before the intervention, it is difficult to measure improvement or confidently attribute changes to the programme. For instance, it is possible that Grade 2 learners in multigrade coached schools were already outperforming those in base schools prior to the intervention. While this is unlikely given the sampling process, the lack of baseline data limits definitive conclusions about programme impact and the magnitude of change achieved. Most importantly, the lack of a baseline precludes identifying the level of change brought about by the base programme (compared to no intervention).

4.5 Sustainability Findings

This section presents findings on the embedding and sustainability of EGRP II, including the conditions that supported or hindered this process. Sustainability plans were initially documented in 2024. The current reports update these findings through documentation review

and interviews conducted in 2026 upon the conclusion of EGRP II implementation. The methodology, sources, and key foregrounding literature are detailed in Annexure 10.

The EGRP II theory of change includes a pathway for embedding and sustaining the programme within the education system, with the aim of ensuring the system has the policies, capacity, and budget to independently scale up core elements of the EGRP II design. This means aligning the programme with existing provincial reading initiatives, securing provincial budget commitments, and reducing dependence on external support to ensure lasting impact.

Before presenting results, we summarise the findings of the literature review conducted during the first year of the sustainability evaluation. This review grounds the interpretation of results and the sustainability evaluation process followed in an empirical and theoretical framework. The review found that there is a broad consensus that, despite numerous interventions to address poor literacy outcomes in South Africa, these have failed to produce systemic change, largely because they were not integrated into the structures and practices of the system (Meiklejohn et al., 2021; Metcalfe, as cited in Christie & Monyokolo, 2018). The review further found that this was a global trend. The literature identifies two categories of levers for successful system-level change, namely enabling programme design factors and enabling embedding process factors. These are briefly explained in the table below.

Table18 Enabling factors for system-level change

Category	Enabling factors
Design factors	• Embedded professional development design: School-based, iterative professional development that is peer-facilitated and participatory, educationally effective and scalable (Hassler et al., 2021).
	• Context-responsive: the intervention design adapts to local cultural, economic, and social conditions.
	• Policy alignment: Alignment with national education policy with high-level government support (Burns, 2014; McKinsey, 2024).
Embedding process factors	• Stakeholder alliance: documented collaboration across government units, levels, and NGOs, built into procedures.
	• Financial sustainability: long-term financing strategies and up-front resourcing agreements; translate external costs into internal system costs (Piper & Dubek, 2024).
	• Clarity on core design elements: distinguish core impact activities from secondary ones to avoid partial adoption undermining impact.
	• Clear, simple messaging: compelling, communicable narrative to prevent breakdowns during scaling (Al-Ubaydli et al., 2019).
	• Defined roles and mandates: explicit roles integrated into accountability and attached to positions (UNICEF, 2024).
	• Implementation & management capacity: continuous, embedded capacity building tied to immediate practice problems; use Professional Learning Communities (City et al., 2010).

	• Data-driven management: coordinated monitoring and data-informed adaptive decisions (UNICEF, 2024).
	• Aligned responsibility and authority: implementers must have a mandate, resources, and authority for sustainable internal accountability.

In the local context, Metcalfe's work, as cited in Christie and Monyokolo (2018), identified several necessary conditions for the successful embedding of an educational program in KwaZulu-Natal. We organise these according to the eleven factors listed above, showing their application within the South African context

These factors are applied in a rubric to assess the level of embedded design and emergent embedded practice within EGRP II, and what they imply for the prospects of sustainability. This assessment was first conducted in 2024, and the current report presents an updated position.

EGRP II design factors

It is important to note that EGRP II was designed with embedding in mind from the outset. The programme incorporated many of the key factors identified in the literature as enabling system-level change, and government ownership was always central to its conception. The PMT made genuine and sustained efforts to realise this, involving key provincial and district officials from the beginning and maintaining their engagement throughout implementation. **The findings of this evaluation should therefore be read not as a critique of programme intent or design, but as a reflection of the considerable difficulty of translating well-conceived embedding principles into durable institutional practice within a system facing significant structural and fiscal pressures.** Where gaps exist, they speak more to the complexity of the enabling environment than to a failure of programme ambition. Where such failures existed, this is articulated.

4.5.1 School-based professional development that is iterative and participatory

School-based professional development deepened over the programme, with the extension period into the first half of 2026, beyond the original program period ending December 2025, promising the consolidation of gains for multigrade schools that had shown the most progress but remained vulnerable relative to the broader system. The introduction of "champion teachers", experienced teachers in a peer-mentoring role within cluster workshops, is a step toward institutionalising professional development within the system and reducing reliance on external facilitation, though this remains at an early stage. Cluster meetings have strengthened over the course of implementation and become the primary vehicle for ongoing professional development, with SAs integrating EGRP II content into regular monthly meetings within existing system structures. In the JTG district, some PLCs show encouraging signs of organic ownership, with teachers meeting independently after hours, setting joint assessment tasks, moderating scripts, and organising district-wide spelling bees. District officials attribute this to

teacher enthusiasm, a PLC culture cultivated by external coaches, and logistical support from school principals, including food and transport.

However, SAs frequently have time constraints and lack transport to attend cluster meetings, limiting their oversight and support role. PLC sustainability depends heavily on the goodwill of uncompensated champion teachers, which may not be a reliable long-term foundation. Where PLCs are less developed, collaboration remains confined to individual schools rather than functioning as cross-school networks. The withdrawal of EGRP II coaches poses a real risk to momentum, as no equivalent driver has yet been established within the system, and the underlying SA shortage remains unresolved.

4.5.2 Context-responsive resources and local adaptation

Interviews across 2024 and 2026 reinforced the importance of provincial and district involvement in the material review process. SAs worked through lesson plan content term by term alongside the implementing partner, grounding materials in both the curriculum and the practical realities of Northern Cape classrooms. This process built deep familiarity with the materials among provincial and district officials and created a sense of ownership that is important for sustainability. The Setswana language materials are widely acknowledged as the first major investment in structured reading instruction in the Northern Cape, generating genuine enthusiasm among teachers and officials.

The provision of multigrade materials remains the most readily adopted element of EGRP II and is likely to endure beyond the project, as teachers recognised their utility and the absence of equivalent resources in the system. Interviews and case study reflections show that teachers were passionate about having something *'homegrown'* in their specific linguistic and pedagogical context, something that felt like *'theirs'*. Stakeholders interviewed suggest that this emotional ownership and proven utility of the LTSMs are important for continued use after external support ends. This reflects a fundamental implementation reality, captured by Metcalfe (as cited in Christie & Monyokolo, 2018), who asserts that

"Policy intentions and instructions do not change behaviour. These are filtered through what educators and officials already understand and believe, what they are convinced will work in their context, and what they think is possible." (p.22)

A telling counter-example is teachers' rejection of curriculum trackers, which they did not perceive as contextually relevant or practically useful to their circumstances.

4.5.3 Policy alignment and government leadership

The active involvement of district and provincial authorities established a strong foundation for government ownership. This was reinforced by a deliberate effort to keep the PED in the lead, with provincial officials consistently chairing PMT meetings and SAs actively aligning materials with the curriculum, local dialects, and ATPs, collectively strengthening the conditions for

long-term sustainability. A Provincial Reading Strategy (2025–2030)¹¹ that draws on some elements of the EGRP II is a welcome development for the embedding process. Interviewed provincial and district officials expressed a genuine commitment and proactive planning. The district officials demonstrated a deep level of engagement with EGRP II's methods that would not have existed without sustained commitment and deliberate capacity development.

A key learning from EGRP II is the importance of protecting the pre-implementation phase for genuine co-creation. In practice, stakeholders overwhelmingly reported that this period was compressed, with implementation activities beginning before contracts were signed and the designated preparation time absorbed into delivery rather than used for joint planning, relationship-building, and aligning on shared responsibilities. Stakeholders reflected that meaningful co-creation requires a sustained period in which government partners are positioned as architects rather than facilitators or hosts. The sequencing of funding, procurement, and implementation timelines can inadvertently crowd out the very processes that determine whether a programme becomes embedded in the system or remains dependent on external support.

4.5.4 Stakeholder alliance

The EGRP II PMT served as an effective coordination mechanism, bringing together national and provincial officials, funders, and implementing partners, with relationships deepening substantially over the implementation period. The provincial education department proactively sought to extend its partnerships, including monthly engagements with a local university and discussions with technology companies around digital resource access. However, the introduction of the National Home Language Reading Programme¹² (NHLRP) across three Northern Cape districts outside the EGRP II footprint risks creating a bifurcated provincial approach¹³, inadvertently positioning EGRP II as an externally funded initiative in contrast to a government-owned alternative.

While the two programmes can be seen as complementary rather than competing, given that the Setswana scale in the province has largely been achieved through EGRP II and that the NHLRP addresses different languages in different districts. Nevertheless, there remains the risk of the gradual gravitational pull of a government-owned and funded programme drawing planning energy, budget allocations, and stakeholder commitment away from an intervention that, however effective, continues to carry the weight of being externally funded. If EGRP II is

¹¹ The evaluation team had not received nor reviewed the Provincial Reading Strategy at the time of writing and is therefore unable to independently verify its content.

¹² The NHLRP emerged from a national stocktaking workshop in early 2025, at which major literacy programmes were reviewed with the aim of developing a single, coordinated national approach to early grade reading that draws on the best elements of existing programmes, including EGRP.

¹³ This position is taken cautiously and without the benefit of having reviewed the Provincial Reading Strategy, which may well address how these two approaches are integrated. Stakeholder interviews did not provide any indication that this is the case. The evaluation acknowledges this as a limitation.

not explicitly integrated into the provincial reading strategy, its long-term embedding beyond the current funding period remains vulnerable.

The preceding analysis suggests that the stakeholder alliance was genuinely consolidated within a defined inner circle (EGRP II PMT) but remained fragile and was likely never fully consolidated at the broader provincial system level or across the relevant national education department directorates, lending credence to the need for sustained relationship building and engagement as an important condition for embedding.

4.5.5 Financial resources and funding mechanisms

A key assumption underpinning EGRP II's sustainability model was that the PED would progressively absorb programme costs, including LTSM printing and training. By the second year of implementation (2024/2025), the PED was placed under severe austerity measures that significantly constrained its operational budget. Stakeholders described the austerity measures as a structural binding constraint, not a temporary challenge to be managed around, but a systemic condition that fundamentally limits what is achievable. Given South Africa's long-term fiscal outlook, a constrained public education budget should be treated as a baseline condition, not an exception. Therefore, Any future externally funded intervention must consider whether it is plugging resource gaps that immediately reopen upon withdrawal, or genuinely making existing resources more effective.

The practical consequences of these austerity measures are significant as full teacher training cannot be sustained, expansion to additional districts has been ruled out, and material replenishment remains uncertain. Stakeholders asserted that even ring-fencing school LTSM allocations for reading resources is challenging when schools themselves face budget cuts and late disbursements. Austerity, therefore, does not merely limit what is possible; it undermines the planning and resourcing logic on which sustainability depends.

4.5.6 Clarity on which design elements drive embedding and sustainability

At the outset of the programme, it was not sufficiently clear which design elements were essential to embedding and long-term sustainability, and which were secondary. Though it was too early to expect clarity on that, this ambiguity had practical consequences as different components carry fundamentally different embedding barriers, and technology-dependent elements, for instance, require distinct resources and levels of political commitment compared to practitioner-led structures such as PLCs. Without clarity on the essential package, it would have been difficult to align stakeholders around shared priorities or to develop realistic sustainability plans within the constraints of available resources.

Over the course of implementation, greater clarity has since emerged. It is now understood that coaching in its current EGRP form is not replicable within the existing system, and thinking has shifted toward decoupling this function so that it is absorbed by individuals already operating within the system. There is now clearer consensus around prioritising cluster workshops as a more sustainable and affordable model of teacher support, alongside lead teacher and subject advisor support. While teacher training proved impactful, it is recognised that it cannot continue

in the form delivered during EGRP implementation, especially given prevailing austerity measures.

4.5.7 Clear messaging, compelling narrative

Clarity on core design elements is also a matter of messaging to create alignment across multiple stakeholders, from school to district to province level. District and provincial officials interviewed in 2026 were well-informed about the EGRP II programme, but it was not clear whether officials less actively involved in the EGRP governance structures had a reliable sense of what problem the programme aimed to solve and its core elements.

4.5.8 Clear implementation planning with well-defined mandates

Although no formal sustainability plan was developed as envisaged by the EGRP II Performance Management Framework, provincial education department officials demonstrated clear and considered thinking about what continuation could realistically look like under current constraints. Rather than replicating the full coaching model, they proposed an adapted approach that decouples key programme functions from their original delivery mechanisms, reflecting pragmatic institutional intent to preserve what drove results. For such planning to be credible, however, it must be grounded in a realistic assessment of system capacity, including binding constraints such as SA shortages, LTSM budgets, and inter-departmental coordination..

4.5.9 Professional daily practices (implementation capacity)

At school level, there is reliable evidence of strong methodology adoption, particularly among multigrade teachers, with encouraging uptake also observed in uncoached schools where principal support was strong. While variability persists, cluster workshops may offer a meaningful lever for sustaining and spreading practice, with targeted support directed at lower-performing schools. At district and provincial level, a full systemic shift may be premature to expect after only two years, though stakeholders note that existing monitoring tools still lack specific indicators to track EGRP II classroom practices. Any monitoring and evaluation strengthening should be practical and integrated into existing workflows, equipping SAs to collect meaningful data and make evidence-based decisions about school support.

4.5.10 Management capacity for data-driven improvement

The training of SAs and other officials in monitoring and evaluation skills and data use represents a meaningful step toward data-driven management, but a significant gap remains between having the skills and having the human resources to apply them systematically given the SA shortage. A further constraint is data quality, as school-based assessment data correlates poorly with independent standardised assessments and tends to overrepresent learner achievement, limiting its reliability as a management tool. Strengthening this will require decisions on assessment standardisation and clarity on which data sources will anchor performance monitoring. Much of the necessary technological infrastructure already exists through the Data-Driven Districts programme, meaning that investment in more standardised

school-based assessment data could quickly translate into improved targeting of Subject Advisor support..

4.5.11 Alignment of responsibility and authority

For embedding to be meaningful, sustainability functions need to be formally assigned in performance contracts at district and provincial level. While it may be too early to expect formal shifts after only two years of implementation, no stakeholder interviews confirmed such plans, and no performance contract currently identifies EGRP II continuation as an accountability target. The provincial Reading Strategy provides a useful framework, but without clearly assigned responsibility it risks remaining aspirational.

A donor-funded sustainability coordinator role within the provincial education department is planned to galvanise embedding efforts. For this to be effective, it must engage with existing roles that carry formal sustainability responsibilities, with clear decision-making authority and strong linkages to system structures. Without this, the coordinator risks operating in isolation and embedding remains a shared but unowned responsibility, present in planning documents but absent in institutional routines.

EGRP II Key Successes

Several elements of EGRP II reflect genuine system-building work that extends beyond the programme's immediate footprint and timeline.

- **The reference group model for LTSM development functioned as an authentic co-creation mechanism rather than a consultative exercise.** Subject Advisors worked through lesson plan content term by term alongside the implementing partner, ensuring materials were grounded in both the curriculum and the practical realities of Northern Cape classrooms. Stakeholders identified this process as a key lever for sustainability, not only because it produced contextually relevant materials, but because it built the capacity and agency of Subject Advisors to adapt and continue using those materials independently.
- **Monitoring and evaluation capacity building:** The M&E capacity-building work reached an uncommonly advanced level in externally funded programmes. A provincial workshop convened during the EGRP II period engaged over 120 officials from across the Northern Cape, extending well beyond the two EGRP districts. Officials were supported in developing monitoring tools, building data literacy, and establishing reporting processes, reflecting intentional system-building that went beyond the narrow programme boundary.
- **Adoption of classroom practices:** Teacher-level adoption in EGRP II schools has been notably strong, particularly in multigrade schools. Teachers expressed a sense of ownership over the Setswana materials, describing them as contextually relevant and pedagogically effective. This sense of emotional ownership is an important precondition for continued use after external support ends.
- **Multigrade LTSM:** The multigrade LTSM developed through EGRP II addresses a gap that no other current South African programme fills. Structured lesson plans, readers,

and methodological guides adapted to the multigrade context are a resource that schools can draw on long after the programme concludes. Stakeholders noted that the value of these materials is likely to outlast the programme itself and that their potential was only fully appreciated during the extension period.

- **Training of SAs and provincial officials in coaching methodology, M&E processes and data-driven school oversight represents a durable system-level investment.** These officials carry EGRP II knowledge and practice in their professional repertoire regardless of whether the programme continues formally, and represent perhaps the most lasting human capital legacy of the intervention.

Critical sustainability challenges

The preceding analysis presents a mixed picture of sustainability. Provincial and district officials are broadly positive about programme achievements and their intentions to continue key elements beyond the funding period. However, several structurally critical themes emerge that temper these claims. The following section identifies and examines six challenges that bear most directly on the prospects for long-term embedding.

- **Austerity measures:** they represent a structural and systemic condition rather than a temporary challenge, fundamentally limiting what the PED can commit to. LTSM replenishment, sustained teacher training, SA deployment, and the absorption of programme costs, the very actions on which sustainability depends, are all at risk under current and foreseeable fiscal conditions.
- **Partial Impact and its Implications for embedding:** A fundamental question concerns the premise of embedding itself. The case for deeper embedding must be grounded in evidence of sufficient, broad-based impact, which the quantitative findings only partially support. EGRP II's most demonstrable gains are concentrated in Grade 2 multigrade schools, and the coaching component did not consistently improve outcomes for Grade 3 learners in the same classrooms. The absence of coaching effects in monograde schools is partly explained by dosage shortfalls, and should be read alongside the substantial body of evidence from prior studies confirming the efficacy of coaching in monograde contexts. Without negating the significance of the Grade 2 multigrade gains, embedding the full EGRP II package more broadly risks institutionalising an approach whose conditions for success have not been sufficiently established beyond this specific Grade 2 multigrade setting.
- **The unresolved issue of the coaching element:** Coaching remains the most impactful yet least sustainable element of the intervention. The proposed alternative model, using SA-led clusters and champion teachers, is context-sensitive and reduces workload demands, but its effectiveness in sustaining instructional quality remains unproven. A further structural constraint is that SA relationships with teachers are widely perceived as evaluative rather than developmental, posing a barrier to the trusting, growth-oriented support that drove EGRP II's results. Without negating the value of the coaching training SAs received, shifting these deeply held perceptions is likely to take time and sustained effort beyond what a single programme can achieve. Cluster workshops and champion

teachers can play a supporting role, but neither replicates the sustained, individualised, in-classroom coaching that produced meaningful outcomes.

What is most likely to continue

The evidence points to several elements with genuine sustainability potential, though, in most cases, their continuation depends on conditions that remain only partially resolved.

4. **Teacher practice improvements are among the most promising:** In schools with motivated teachers and supportive leadership, the methodological practices introduced by EGRP II show signs of taking root. The programme's status as the first major investment in Setswana home language reading instruction has generated authentic ownership and enthusiasm among teachers, providing an intrinsic motivational foundation that persists without external support.
5. **Structured lesson plan use represents the single most portable and durable element of the EGRP II package:** SLPs are not tied to specific materials, do not require ongoing external facilitation, and can be monitored through existing local structures. They present the clearest case of a practice capable of outlasting the programme itself.
6. **Cluster-based professional development offers a cost-effective and structurally compatible mechanism for continued engagement with EGRP II content:** integrating programme refreshers into existing monthly cluster meetings aligns with provincial budget constraints and established routines. It should be noted, however, that the quality of support this model can provide is substantially lower than the intensive, in-classroom coaching delivered during implementation, and this difference has implications for how much instructional improvement can realistically be sustained through this channel alone.
7. **The multigrade LTSMs are arguably the most significant and enduring material legacy of EGRP II.** Prior to the programme, no structured reading resources existed for multigrade classrooms in the Northern Cape, leaving teachers to navigate a complex instructional environment without dedicated support. The materials developed through EGRP II fill a gap that no other current South African programme addresses. Teachers and officials have expressed strong enthusiasm for these resources, describing them as practically effective and contextually relevant in ways that generic national materials are not. This combination of proven utility, emotional ownership, and the absence of any equivalent alternative makes the multigrade LTSMs the element of EGRP II most likely to remain in active use beyond the programme, provided that material replenishment is secured and a deliberate strategy exists to sustain teacher engagement with them over time.

The table below summarises the different embedding criteria

Table 19: Sustainability Factors Assessment

Sustainability Criterion	Year 1 Status	Year 2 Status
1. School-Based PD that is Iterative and Participatory	Enabling	Enabling
2. Context-Responsive Resources & Local Adaptation	Enabling	Enabling
3. Policy Alignment & Government Leadership	Enabling	Enabling
4. Stakeholder Alliance	Emerging	Emerging
5. Financial Resources & Funding Mechanisms	Emerging	Needs Attention
6. Clarity on Design Elements Driving Sustainability	Needs Attention	Emerging
7. Clear Communication of a Simple Message	Needs Attention	Emerging
8. Clear Implementation Planning with Well-Defined Mandates	Emerging	Emerging
9. Professional Daily Practice (Implementation Capacity)	Emerging	Emerging
10. Management Capacity for Data-Driven Improvement	Needs Attention	Emerging
11. Alignment of Responsibility & Authority	Needs Attention	Needs Attention

In conclusion the overall sustainability findings point to sustainability foundations in teacher ownership of the multigrade LTSMs, structured lesson plan use, capacity built in SAs and provincial officials, and the integration of EGRP II content into cluster-based professional development. These elements are portable, compatible with existing system routines, and do not require continued external funding to remain in use.

However, the conditions required to build on these foundations are significantly constrained. Austerity measures are not a temporary challenge but a structural reality that limits material replenishment and makes sustained teacher training effectively unaffordable. SA shortages represent a binding constraint that undermines the human infrastructure on which any continuation model depends. Sustainability responsibilities remain formally unassigned with the possibility that austerity may itself be suppressing the institutional appetite to take them on.

The introduction of the NHLRP across Northern Cape districts outside the EGRP II footprint adds a further risk. A government-owned initiative carries institutional gravity that an externally funded programme cannot match, and in a fiscally stressed system, planning energy and budget allocations will likely gravitate toward it. Austerity is likely to sharpen rather than ease this tension. Whether EGRP II's gains endure depends less on what the programme built and more on whether the system can protect and resource what remains, a question that current fiscal and structural conditions leave genuinely open.

5. CONCLUSIONS

The EGRP II evaluation reveals critical insights about intervention design, implementation, and the potential for targeted support to address educational inequities in challenging contexts. While the absence of baseline data limits causal attribution, the patterns observed suggest important directions for future programming and systemic reform.

Targeted coaching as a potential equaliser for multigrade schools: The most striking finding is the strong performance of Grade 2 learners in coached multigrade schools, who achieved literacy outcomes comparable to their monograde peers despite historically facing structural disadvantages. Multigrade schools in the Northern Cape have long contended with complexity and limited targeted support, yet coached multigrade learners demonstrated proficiency levels that matched or exceeded those in monograde settings across several literacy subtasks. This pattern suggests that the combined approach of coaching, training, and customised materials may offer a viable pathway to address the structural inequalities facing rural, farm-based multigrade schools. The question for future interventions is whether this model can be sustained and scaled to consistently support multigrade environments.

Pedagogical flexibility may improve learner outcomes more than standardisation: The observed success in multigrade classrooms appears connected to fundamental pedagogical shifts rather than solely resource provision. The intervention encouraged multigrade teachers to move away from rigid grade-level teaching toward ability-based grouping. Teachers in multigrade settings reported using this approach to address learning gaps more flexibly than may have been possible in monograde classrooms bound by strict curriculum pacing expectations. This raises an important hypothesis: instructional adaptability and differentiation may be more powerful drivers of literacy improvement in diverse classrooms than standardised, prescriptive curriculum delivery. This raises the question of whether tools like curriculum trackers failed to gain traction because they were poorly suited to complex teaching environments that require flexibility rather than rigid adherence to prescribed pacing. Future research should explore whether deliberate cultivation of pedagogical flexibility yields stronger outcomes than adherence to uniform pacing schedules.

Digital tools require stronger infrastructure and support systems: The programme's effort to digitalise lesson planning and curriculum tracking through tablets encountered some substantial challenges. Tablet functionality was inconsistent, technical support was not fully accessible, and many teachers favour paper-based methods despite near-universal tablet distribution. However, this pattern may partly reflect current teacher demographics, with resistance likely concentrated among older teachers while younger teachers demonstrated greater digital comfort. As generational turnover occurs, adoption patterns may shift naturally, raising questions about optimal timing for technology integration.

The literacy crisis demands sustained, multi-year commitment: While the relative superior performance observed in coached multigrade schools offers encouragement, absolute literacy

levels across both districts remain critically low. Substantial proportions of learners, ranging from 25% to 47% depending on context, cannot read grade-appropriate text even after this two-year intervention. These foundational gaps indicate that short-term, project-based interventions, regardless of relative success, will struggle to address the depth of literacy challenges facing the Northern Cape - or South Africa more generally. Meaningful progress toward ensuring all learners meet minimum literacy benchmarks will likely require sustained, intensive support over multiple years rather than time-limited programmes. Education authorities should consider how successful intervention components might be institutionalised within regular district support systems to enable the long-term engagement necessary for transformative change. The value of Reading Champions in this and earlier early grade literacy projects should also be factored into planning.

Systemic data collection must improve for effective monitoring: A critical constraint on this evaluation was the absence of reliable baseline data and the inadequacy of existing School-Based Assessment systems for monitoring learner progress. Without pre-intervention performance data, the evaluation could only compare endline differences across groups rather than definitively measuring growth or attributing changes to programme activities. Additionally, some implementation data contained gaps requiring triangulation with surveys and case studies. This situation highlights an urgent need for systemic reform in how learner performance data is collected, archived, and used for decision-making. Before scaling interventions province-wide, education authorities should invest in strengthening data systems to enable rigorous monitoring, evaluation, and continuous improvement. Without reliable data infrastructure, programmes cannot adequately assess impact, adjust strategies, or demonstrate accountability.

Implementation quality shapes programme outcomes: Beyond intervention design, the quality and consistency of implementation emerged as critical factors potentially influencing outcomes. While training attendance was high and materials were well received and coaching quality improved, several implementation challenges arose. Most importantly, coaching visit targets were frequently unmet (albeit for systemic contextual reasons). In addition, only a few coaches encouraged adequate reflective practice among the teachers they coached. Further, most multigrade teachers in the base schools had challenges in the use of the multigrade LTSM, given the survey findings that coached multigrade teachers report much higher comfort with the multigrade-specific LTSM and the core methodologies expressed therein. These implementation inconsistencies may have constrained the programme's potential effects and possibly created unintended inequities between treatment groups. Future programming should prioritise implementation fidelity alongside intervention design, ensuring adequate resources for logistics, ongoing teacher support, and quality assurance systems (e.g digitised M&E systems) that can identify and address implementation challenges in real time.

Sustainability and embedding processes are largely emergent but face significant constraints: EGRP II leaves behind portable assets in the form of multigrade LTSMs, structured lesson plan use, and capacity developed in SAs and provincial officials. This forms a promising foundation for embedding and sustainability. However, austerity constrains LTSM replenishment and training, SA shortages limit any realistic continuation model, and no formal sustainability responsibilities have been assigned, a gap likely compounded by fiscal pressure. The

emergence of the NHLRP poses an additional risk, drawing institutional focus and planning energy toward a government-owned programme in an already stretched system. The durability of EGRP II gains will ultimately depend on whether the system actively resources and protects what remains.

6. LESSONS LEARNT AND RECOMMENDATIONS

6.1 Lessons learnt

Baseline data is essential for credible impact evaluation: The absence of baseline data significantly limited the evaluation's ability to make definitive causal claims about programme impact, highlighting that future interventions must prioritise establishing reliable baseline measures before implementation begins.

Targeted support for underserved teachers can catalyse rapid pedagogical change: Real shifts in teaching practice occurred within one year when interventions focused on multigrade teachers who had historically received minimal specialised support, suggesting that addressing long-standing neglect can generate strong motivation and uptake.

Multigrade contexts can achieve parity with monograde outcomes when provided appropriate resources: With tailored materials and coaching support, multigrade teachers and learners demonstrated performance levels comparable to their monograde counterparts, challenging assumptions about inherent limitations in multigrade settings.

Pedagogical flexibility may yield better outcomes than rigid curriculum compliance: The intervention was most successful in contexts where teachers had autonomy to group learners by ability and adjust pacing, suggesting that strict adherence to grade-level curriculum coverage may constrain literacy development in diverse classrooms.

Structured, problem-focused training drives stronger engagement: Teachers and coaches in the Northern Cape responded more positively to training than in previous North West implementations, attributed to more structured content and practical problem-solving approaches rather than theoretical frameworks.

Program timing and delivery: Insufficient lead time for contracting, materials development, and pre-piloting can result in cascading delays that affect implementation fidelity from the outset. The year preceding full-scale implementation is critical for finalising agreements, developing and testing materials, and resolving operational challenges early. Where this period is not adequately utilised, the programme risks entering full rollout before its foundational elements are sufficiently in place, heightening the risk of compromised fidelity. Delays in implementation limit the opportunity for programme effects to develop and be observed.

6.2 Recommendations

For the Department of Basic Education and Provincial Education Departments

1. Prioritise targeted support for multigrade early grade teachers. The DBE should assist PEDs in developing and delivering specialised training for multigrade teachers, with particular emphasis on mastering Group Guided Reading and Shared Reading methodologies. These approaches enabled multigrade teachers to fundamentally rethink classroom management and simultaneously instruct learners across multiple grades and ability levels.
2. Extend implementation timeframes for literacy interventions. Future programmes should plan for minimum two-year implementation periods, particularly for monograde contexts where coaching effects require sustained engagement to materialise. Compressed timelines constrain the intervention's ability to shift established teaching practices and demonstrate measurable learner outcomes.
3. Establish reliable baseline data collection protocols. All future interventions should prioritise collecting credible baseline learner performance data before programme commencement. The absence of baseline data in this evaluation severely limited the ability to make definitive causal claims about programme impact and measure actual growth over time.

For Future Research and Programme Design

1. Investigate coaching effects in multigrade contexts through rigorous experimental design. The promising results in coached multigrade classrooms warrant further exploration through studies employing baseline measures and experimental designs to validate or challenge these findings. Research should examine mechanisms driving efficiency gains, pedagogical shifts, and learner outcomes, as well as explore whether multigrade-adapted approaches could benefit monograde classrooms with diverse learner ability groups.
2. Integrate ability-based grouping strategies with structured literacy instruction. Combining standard literacy improvement approaches with differentiated instruction and ability-based grouping (key elements of Teaching at the Right Level) proved particularly effective in multigrade classrooms. Programme designers should intentionally incorporate these complementary strategies rather than treating them as separate interventions.
3. Refine coaching dosage and quality standards. Future programmes should establish clear coaching visit targets with accountability mechanisms, while also strengthening coach training to develop facilitative rather than directive approaches that cultivate genuine teacher reflection and independent problem-solving capabilities.

For Technology Integration

1. Sequence technology introduction with adequate support infrastructure. Digital tool integration should only proceed when accompanied by robust technical support systems, sufficient teacher preparation time, and demonstrated practical utility. Premature digitalisation without these conditions generates resistance and undermines adoption, as evidenced by widespread rejection of curriculum trackers and preference for paper-based planning.

6. REFERENCES

- Al-Ubaydli, O., List, J., Mackevicius, C., Lee, M. S., & Suskind, D. (2019). *How can experiments play a greater role in public policy? 12 proposals from an economic model of scaling*. Artefactual Field Experiments 00679, The Field Experiments Website. <https://www.fieldexperiments.com> (Accessed 7 March 2025)
- Angrist, J. D., & Pischke, J.-S. (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton University Press.
- Bamberger, M. (2010). Institutionalising impact evaluation: A key element in strengthening country-led monitoring and evaluation systems. In M. Segone (Ed.), *From policies to results: Developing capacities for country monitoring and evaluation systems*. UNICEF.
- Brown, B. A. (2010). *Multigrade teaching: A review of issues, trends and practices: Implications for teacher education in South Africa*. CEPD.
- Burns, M. (2014). *The myths of scaling-up*. Global Partnership for Education. <https://www.globalpartnership.org/blog/myths-scaling> (Accessed 7 March 2025)
- Cornish, L., & Taole, M. J. (2021). *Perspectives on multigrade teaching: Research and practice in South Africa and Australia*. Springer.
- Department of Basic Education. (2018). *Second Early Grade Reading Study: Year 1 Report*.
- Department of Basic Education. (2022). *Synthesis report of EGRS I and EGRS II: The second Early Grade Reading Study. Evidence one year after the end of implementation. Year 4 report*. <https://www.education.gov.za/Portals/0/Documents/Reports/EGRS/EGRS%20Indaba%202022/Synthesis%20Report%20digital.pdf>
- Department of Basic Education. (2023). *School realities*. <https://www.education.gov.za/Portals/0/Documents/Publications/School%20Realities%20December%202023.pdf?ver=2024-02-21-110237-047>
- Department of Basic Education. (2025a). *Funda Uphumelele National Survey: Northern Cape Provincial Report*.
- Department of Basic Education. (2025b). *Key findings from a series of studies on structured learning programmes in South Africa*. <https://www.education.gov.za/Portals/0/Documents/Publications/EGRS/EGRS%2025/EGRS%20Series-Synthesis%20Report%202025.pdf>
- Duflo, E., Glennerster, R., & Kremer, M. (2007). Using randomization in development economics research: A toolkit. In T. P. Schultz & J. Strauss (Eds.), *Handbook of development economics* (Vol. 4, pp. 3895–3962). Elsevier.
- Fleisch, B. (2018). *The education triple cocktail: System-wide instructional reform in South Africa*. University of Cape Town Press.
- Fraser, T. (2023). *The art of scaling deep*. The Systems Sanctuary. <https://systemsanctuary.com/scale-deep> (Accessed 7 March 2025)

- Graham, A., Powell, M., Taylor, N., Anderson, D., & Fitzgerald, R. (2013). *Ethical research involving children*. UNICEF Office of Research – Innocenti.
- Haßler, B., Bennett, G., & Damani, K. (2021). Teacher professional development in sub-Saharan Africa: Equity and scale. In C. McNaught & S. Gravett (Eds.), *Embedding social justice in teacher education and development in Africa*. Routledge. <https://docs.opendeved.net> (Accessed 7 March 2025)
- Haßler, B., Hennessy, S., & Hofmann, R. (2020). *OER4Schools: Outcomes of a sustained professional development intervention in sub-Saharan Africa*.
- Hill, C. J., Bloom, H. S., Black, A. R., & Lipsey, M. W. (2008). Empirical benchmarks for interpreting effect sizes in research. *Child Development Perspectives*, 2(3), 172–177. <https://doi.org/10.1111/j.1750-8606.2008.00061.x>
- Joubert, J. (2010). *Multigrade teaching in South Africa*. Commonwealth Education Partnership.
- Kraft, M. A. (2020). Interpreting effect sizes of education interventions. *Educational Researcher*, 49(4), 241–253. <https://doi.org/10.3102/0013189X20912798>
- Little, A. (1995). *Multigrade teaching: A review of practice and research* (Serial No. 12). Overseas Development Administration.
- Little, A. (2001). Multigrade teaching: Towards an international research and policy agenda. *International Journal of Educational Development*, 21(6), 481–497.
- Little, A. (2005). *Learning and teaching in multigrade settings*. Paper prepared for the UNESCO 2005 EFA Monitoring Report. www.skoleplassen.hisf.no/Rapportar/LearningandTeachinginMultigrade
- Little, A. (2006). *Education for all and multigrade teaching: Challenges and opportunities*. <https://wcedportal.co.za/multigrade-teaching-and-learning>
- Maher, E. J., Jackson, L. J., Pecora, P. J., Schultz, D. J., Chandra, A., & Barnes-Proby, D. S. (2009). Overcoming challenges to implementing and evaluating evidence-based interventions in child welfare: A matter of necessity. *Children and Youth Services Review*, 31, 555–562. <https://doi.org/10.1016/j.childyouth.2008.10.01>
- McKinsey & Company. (2024). *Spark & sustain: How all of the world's school systems can improve learning at scale*. <https://www.mckinsey.com/industries/education/our-insights/spark-and-sustain-how-school-systems-can-improve-learning-at-scale> (Accessed 7 March 2025)
- Mohohlwane, N. L. (2016). *The contribution of randomised control trials (RCTs) to improving education evaluations for policy: Evidence from developing countries and South African case studies*. <http://wiredspace.wits.ac.za/handle/10539/22682> (Accessed 7 February 2022)
- Moir, S. (2018). Why is implementation science important for intervention design and evaluation within educational settings? *Frontiers in Education*. <https://www.frontiersin.org/journals/education/articles/10.3389/feduc.2018.00061/full> (Accessed 7 March 2025)

Mullis, I., Martin, M. O., Foy, P., & Drucker, K. (2012). *PIRLS 2011 international results in reading*. International Association for the Evaluation of Educational Achievement (IEA), TIMSS & PIRLS International Study Center.

Muralidharan, M., & Niehaus, P. (2017). Experimentation at scale. *Journal of Economic Perspectives*, 31(4), 103–124. <https://doi.org/10.1257/jep.31.4.103>

OECD. (2010). *Quality standards for development evaluation*. DAC Guidelines and Reference Series, Development Assistance Committee, OECD.

Oxford University Press Southern Africa. (2025). *Monitoring & evaluation report: Early Grade Reading Programme 2 implemented in the Frances Baard and John Taole Gaetsewe districts of the Northern Cape PDE for the period June 2025 – December 2025*.

Parlett, M., & Hamilton, D. (1977). Evaluation as illumination: A new approach to the study of innovatory programmes. In D. Hamilton, D. Jenkins, C. King, B. MacDonald, & M. Parlett (Eds.), *Beyond the numbers game: A reader in educational evaluation*. MacMillan Education Ltd.

Piper, B., & Dubeck, M. (2024). Responding to the learning crisis: Structured pedagogy in sub-Saharan Africa. *International Journal of Education and Development*, 109, 103095. <https://doi.org/10.1016/j.ijedudev.2024.103095>

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

Social Impact Insights Africa. (2025). *Evaluation of the Early Grade Reading Programme II (EGRP II) Northern Cape, South Africa, 2024–2025: Final endline case study report*. Prepared for UNICEF.

Stern, J. M. B., et al. (2021). Responding to the learning crisis: Structured pedagogy in sub-Saharan Africa. *PMC*. <https://pmc.ncbi.nlm.nih.gov>

Strauss, J. (2022). *The pitfalls of scaling up educational interventions*. Loughborough University. <https://blog.lboro.ac.uk/cmc/2022/04/06/the-pitfalls-of-scaling-up-educational-interventions/> (Accessed 3 March 2025)

Taole, M. J. (2014). Quality basic education for all: Challenges in multi-grade teaching in rural schools. *Mediterranean Journal of Social Sciences*, 5, 531. <https://doi.org/10.5901/mjss.2014.v5n1p531>

Taylor, S., Cilliers, J., Prinsloo, C., Fleisch, B., & Reddy, V. (2017). *The Early Grade Reading Study: Impact evaluation after two years of interventions* [Technical report]. Department of Basic Education. <https://www.jet.org.za/clearinghouse/printed/resources/language-and-literacy-resources-repository/egrs-technical-report-13-oct-2017.pdf>

UNICEF. (2015). *Procedure for ethical standards in research, evaluation, data collection and analysis* (CF/PD/DRP/2015-001).

UNICEF. (2019). *Guidance note for conducting evaluability assessments in UNICEF*.

UNICEF Innocenti – Global Office of Research and Foresight. (2024). *Implement: Bridging the gap between policy and practice – Synthesis of UNICEF Innocenti research on education in Africa*. <https://www.unicef.org>

United Nations Evaluation Group. (2016). *Norms and standards for evaluation*. UNEG.

Van Broekhuizen, H., & Spaull, N. (2017). *The 'Martha effect': The compounding female advantage in South African higher education* (Stellenbosch Economic Working Papers: WP14/2017). <https://ideas.repec.org/p/sza/wpaper/wpapers290.html> (Accessed 7 February 2022)

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.

7. ANNEXURES

To access the downloadable annexures, please contact us at research@socialimpactinsights.co.za

ANNEX 1: Terms of Reference

Please see here the EGRP II terms of reference

ANNEX 2: Evaluation Framework

Please see the evaluation matrix attached here

ANNEX 3: Evaluability Assessment Findings

Please see attached here the evaluability assessment findings on School Based Assessment here

ANNEX 4: EGRP II Learner Assessment and Context Tools

The detailed tool for learner assessment and context assessments are found here.

ANNEX 5: EGRP II Case Study Manual

The EGRP II Case Study Manual containing the methodology and tools are found here

ANNEX 6: Training Observations tools

Tools for observing training delivery are found here.

ANNEX 7: EGRP II Detailed Theory of Change

These pathways provide the envisaged integrated approach to improving reading outcomes through teacher training, consistent supervision, resource provision, and system sustainability. The evaluation team developed "IF-THEN" statements to illustrate the anticipated causal relationships in these pathways, noting these are simplified representations of more complex, interwoven relationships. The statements are detailed below:

IF teachers are provided with and utilise updated lesson plans and LTSMs for both multigrade and monograde classes, which have been reviewed and integrated with existing system materials

AND

IF teachers are trained in CAPS-aligned methodologies/EGRP methodologies and the integrated use of LTSM, including lesson plans

AND

IF the system for training teachers is reviewed and adapted for EGRP delivery, ensuring that it is contextually and appropriately targeted for both mono-grades and multigrade classes.

AND

IF continuous coaching and support are provided to teachers by coaches and other stakeholders in the system, such as SMTs, district and provincial officials whose roles are clearly defined,

AND

IF appropriate training and guidance materials are provided to coaches

THEN

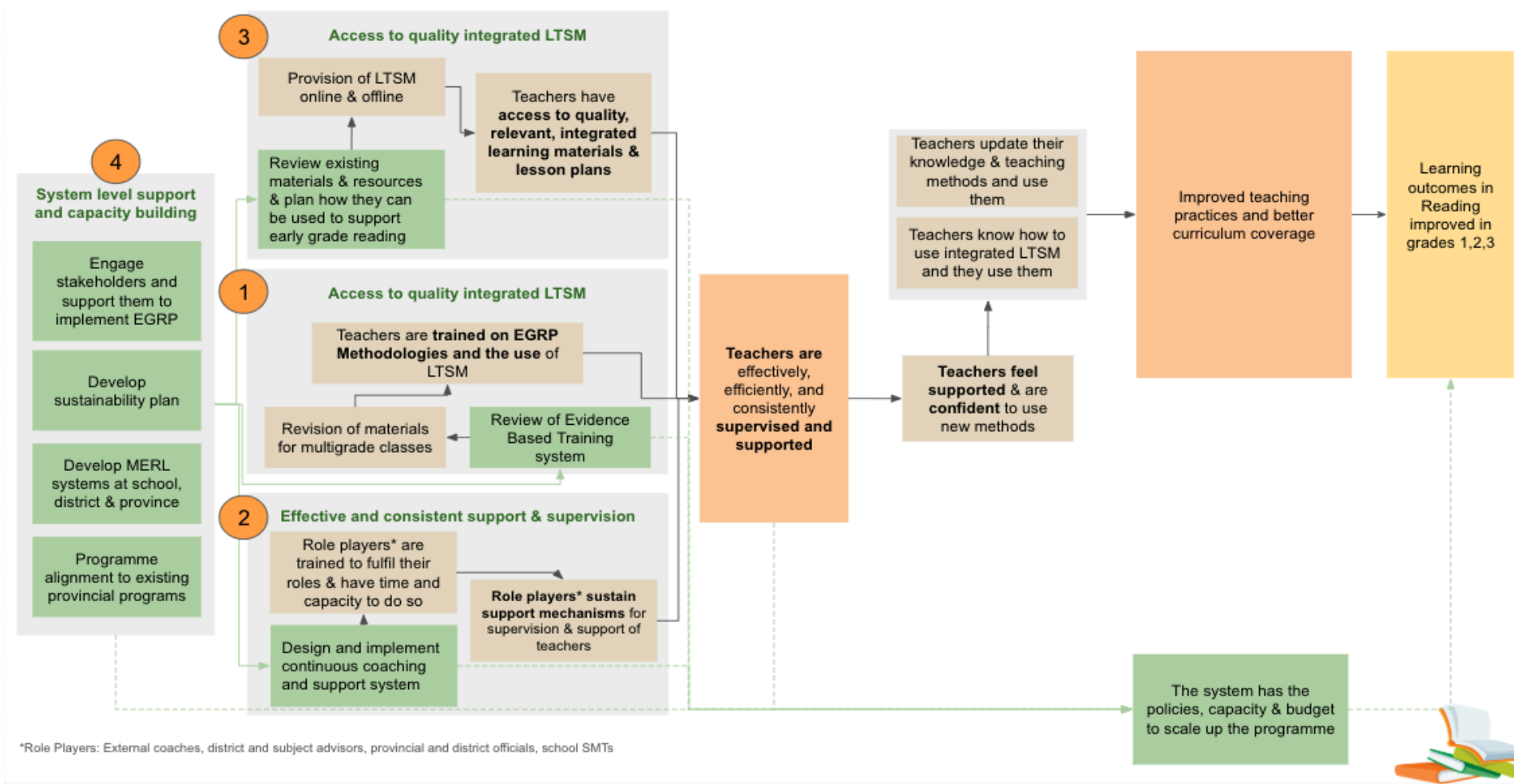
Teachers will be consistently supported through coaching, supervision, and resources to improve their teaching skills. This ongoing professional development will help teachers feel supported and enhance their knowledge, skills, classroom practices, and curriculum coverage, leading to better learning outcomes for grades 1, 2, and 3.

AND

IF the provincial education system is supported by aligning the EGRP to existing provincial reading initiatives, building capacity in officials and involving them in delivering the programme, supporting them in establishing a well-defined sustainability plan, MERL systems, and policies, including a provincial budget for expansion and independent operation beyond initial DBE oversight,

THEN

The programme can operate independently and be scaled across the province, and can be embedded as a self-sustaining part of the educational infrastructure, ensuring its benefits are both far-reaching and long-lasting



ANNEX 8: Teacher Training topics covered

(Source: OUP 2025)

Roadmap of Training Topics

2 Days of Training (8 days)

EFAL: Differentiation, types of reading (HL vs EFAL), shared reading techniques and preparation.

Setswana HL: Growth mindset, listening/speaking, phonics, group guided/shared reading, handwriting, and writing.

Multigrade: same topics, however, focused and applied to instruction in a multigrade classroom, i.e. how to implement whole class / grade-specific / ability-group-specific teaching.

1 Day of Training (4 days)

EFAL: Question of the day and daily activities during Listening & Speaking, Creative writing with focus on the writing process and strategies.

Setswana HL: Group Guided Reading, Helping struggling readers.

Multigrade: Orientation on T4 Lesson Plans and LTSM; same topics, focused and applied to instruction in a multigrade classroom, i.e. how to implement whole class / grade-specific / ability-group-specific teaching.

1 Day of Training (4 days)

The training focused on building a solid understanding of **reading for comprehension**. It covered key components such as **accuracy, fluency, automaticity**, and strategies to enhance comprehension. Teachers were equipped with practical tools to support learners in becoming confident, skilled readers.

Multigrade: Orientation on T2 Lesson Plans and LTSM; same topics, focused and applied to instruction in a multigrade classroom.

Term 1
'24

1 Day of Training (4 days)

Monograde: Orientation on LTSM material for monograde classrooms, covering **key methodologies for structured HL and EFAL learning programmes**.

Multigrade: Workshop with teachers on classroom management strategies, including national support, best practices, needs analysis, challenges in the Northern Cape, and sharing innovations.

Term 2
'24

Term 3
'24

2 Days of Training (8 days)

EFAL: Reading science, 5 pillars, phonemic awareness, phonics teaching methodology.

Setswana HL: Classroom management, timetable, Bloom's Taxonomy, listening/speaking, guided/shared reading, assessment, handwriting, lesson prep, tracker, and dictionary use.

Multigrade: Introducing the tailored MTG programme (T1), same topics, focused and applied to instruction in a multigrade classroom, i.e. how to implement whole class / grade-specific / ability-group-specific teaching.

Term 4
'24

Term 1
'25

2 Days of Training (8 days)

Baseline Assessment: Foundation Phase learners, booklet use, group placement (animal names/colours).

EFAL: Scarborough's Reading Rope, fluency, comprehension components.

Setswana HL: Guided/Shared reading, tasks, policy steps, informal assessment.

Multigrade: Orientation on T1 Lesson Plans and LTSM; same topics, focused and applied to instruction in a multigrade classroom.

Tablets: Asset management, care and rules, EGRP Platform onboarding, navigation for lesson planning

Term 2
'25

EGRP I Programme Elements

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 L and EFAL Discussion Posters HL and EFAL Phonic Friezes	Paper-based HL and EFAL Big Books HL and EFAL Sight Words Flash Cards HL and EFAL Discussion Posters HL and EFAL Phonic Friezes
Training of teachers	Teachers receive training at the beginning of each term throughout the intervention	Teachers receive training at the beginning of each term throughout the intervention	Teachers receive training at the beginning of each term throughout the intervention
Coaching of teachers	None	External Coach visits each teacher in the classroom from Year 1 to Year 3	External Coach visits teacher in the classroom in year 1
Coaching of DHs	None	None	2 External coaches coach DHs remotely at a ratio of 1 coach to 10 schools
Coaching of Teachers by DHs	None	None	Teachers coached by DHs in Years 2 and 3
Virtual coaching support	None	Ongoing remote support by external coach from year 1 to Year 3	None
Establishing PLCs	None	Established in year 2 and implemented throughout	Established in year 2 and implemented throughout
School-based workshops	None	Carried out with each coach visit	Carried out with each coach visit
Geographical Location	North West Province		
Period of Implementation	2021-2023		

ANNEX 9: Regression Models

Model A2 Regression Results (Grade 2 and Grade 3)

Specification: $Y \sim \text{treatment} * \text{school_type}$

Treatment coding: Base = 0, Coaching = 1

School type reference category: monograde

Grade 2: Model A2 (all outcomes)

Grade 2: Model A2 (all outcomes)

	hl_ron_correct _wpm	hl_lsk_correct _wpm	hl_wr_correct _wpm	hl_orf2_correct t_wpm	efal_wr3_correct t_wpm	hl_orf2_compreh ension
(Intercept)	48.069*** (1.541)	50.182*** (4.507)	13.262*** (1.934)	26.066*** (4.702)	12.238*** (2.687)	0.388*** (0.029)
treatment	0.266 (1.185)	-0.924 (2.157)	-0.675 (0.985)	-0.886 (2.098)	0.036 (1.338)	-0.008 (0.019)
school_type mul tigrade	-5.981** (2.525)	-13.628** (6.099)	-5.361* (2.914)	-10.186 (6.620)	-5.525 (4.037)	-0.087* (0.045)
stratum2	4.581* (2.627)	6.274 (5.617)	1.185 (2.620)	2.020 (5.987)	1.603 (3.535)	-0.006 (0.046)
stratum3	0.787 (3.359)	1.495 (6.044)	-0.032 (2.769)	-0.667 (6.335)	-0.081 (3.900)	-0.021 (0.063)
stratum4	5.188** (2.248)	5.894 (6.001)	1.986 (2.929)	2.431 (6.405)	2.794 (4.034)	0.043 (0.041)
stratum5	-1.089 (2.103)	1.986 (5.270)	1.041 (2.491)	0.835 (5.874)	-0.538 (3.353)	-0.002 (0.038)
stratum6	-0.942 (2.155)	-3.322 (5.454)	-1.275 (2.323)	-3.230 (5.340)	-2.335 (3.071)	-0.023 (0.041)
stratum7	-2.927 (2.129)	-3.746 (5.329)	-2.115 (2.331)	-4.818 (5.418)	-2.047 (3.206)	-0.027 (0.042)
stratum8	0.329 (1.988)	1.934 (5.362)	1.847 (2.391)	3.302 (5.672)	3.194 (3.501)	0.031 (0.039)
stratum9	-5.302** (2.523)	-7.838 (4.935)	-2.281 (2.148)	-5.709 (5.108)	-0.946 (2.981)	0.018 (0.036)

	hl_ron_correct _wpm	hl_lsk_correct _wpm	hl_wr_correct _wpm	hl_orf2_correc t_wpm	efal_wr3_correc t_wpm	hl_orf2_compreh ension
stratum10	-1.762 (2.351)	-7.255 (5.280)	-1.878 (2.278)	-5.025 (5.379)	-0.753 (3.113)	0.003 (0.040)
treatment × school_typed multigrade	-1.520 (2.428)	8.109* (4.208)	5.168*** (1.904)	10.298** (4.191)	5.487** (2.409)	0.068 (0.042)
Num.Obs.	1851	1851	1851	1847	1850	1228
R2	0.033	0.037	0.031	0.026	0.026	0.027

Grade 3: Model A2 (all outcomes)

Grade 3: Model A2 (all outcomes)

	hl_lsk_cor rect_wpm	hl_ccdk_co rrect_wpm	hl_wr_cor rect_wpm	hl_orf2_co rrect_wpm	efal_wr3_c orrect_wpm	efal_orf1_c orrect_wpm	efal_orf1_co mprehension
(Intercept)	52.447*** (4.204)	21.428*** (2.303)	19.048*** (2.217)	40.931*** (4.685)	21.691*** (3.928)	32.868*** (5.575)	0.274*** (0.042)
treatment	4.492* (2.546)	0.190 (1.214)	0.318 (1.064)	1.483 (2.569)	0.630 (1.668)	0.251 (2.835)	-0.006 (0.029)
school_typed multigrade	-4.225 (5.565)	-0.618 (2.210)	-0.983 (1.891)	0.706 (4.352)	-0.307 (3.411)	1.402 (5.331)	0.020 (0.040)
stratum2	6.909 (5.055)	0.369 (2.490)	1.072 (2.323)	1.170 (4.917)	-1.334 (3.706)	-1.565 (6.097)	0.000 (0.046)
stratum3	1.929 (5.281)	-1.749 (2.312)	-1.576 (1.982)	-3.888 (4.622)	-4.235 (3.564)	-6.599 (5.156)	-0.067 (0.045)
stratum4	-1.167 (4.925)	-0.842 (2.602)	-0.599 (2.348)	-2.634 (4.943)	0.802 (4.465)	0.147 (6.270)	0.013 (0.051)
stratum5	4.127 (6.095)	2.267 (2.776)	1.000 (2.458)	3.506 (5.398)	1.677 (4.248)	4.092 (6.115)	0.072 (0.058)
stratum6	-2.354 (5.377)	-0.267 (2.760)	-0.547 (2.559)	-1.083 (5.751)	-0.960 (4.250)	-0.153 (6.382)	0.003 (0.048)
stratum7	4.705 (5.336)	2.205 (2.769)	0.930 (2.628)	1.772 (5.591)	2.808 (4.560)	4.426 (6.794)	0.071 (0.050)

	hl_lsk_cor rect_wpm	hl_ccdk_co rrect_wpm	hl_wr_cor rect_wpm	hl_orf2_co rrect_wpm	efal_wr3_c orrect_wpm	efal_orf1_c orrect_wpm	efal_orf1_co mprehension
stratum8	6.688 (5.366)	4.346* (2.641)	3.220 (2.570)	9.689* (5.800)	7.781* (4.375)	15.242** (6.788)	0.195*** (0.058)
stratum9	-7.080 (4.795)	-0.414 (2.521)	-1.047 (2.375)	-2.723 (5.340)	1.042 (4.118)	4.428 (6.241)	0.178*** (0.053)
stratum10	-10.807** (5.218)	-2.897 (2.761)	-2.998 (2.457)	-8.861* (5.322)	-1.026 (4.151)	-2.498 (5.951)	0.106* (0.062)
treatment × school_typ emultigrade	-1.537 (4.629)	1.137 (2.304)	0.605 (1.973)	-3.131 (4.665)	1.130 (3.820)	-1.877 (5.449)	-0.014 (0.053)
Num.Obs.	1847	1847	1847	1845	1847	1845	1845
R2	0.064	0.022	0.020	0.029	0.022	0.025	0.057

Model A2 comparison tables (z-scored outcomes)

Grade 2: Model A2 (all z-scored outcomes)

Grade 2: Model A2 (all z-scored outcomes)

	hl_ron_corre ct_wpm_z	hl_lsk_corre ct_wpm_z	hl_wr_corre ct_wpm_z	hl_orf2_corre ct_wpm_z	efal_wr3_corr ect_wpm_z	hl_orf2_comp rehension
(Intercept)	0.122 (0.112)	0.153 (0.192)	0.116 (0.176)	0.128 (0.196)	0.047 (0.189)	0.050 (0.142)
treatment	0.019 (0.086)	-0.039 (0.092)	-0.062 (0.090)	-0.037 (0.088)	0.003 (0.094)	-0.038 (0.095)
school_type multigrade	-0.434** (0.183)	-0.579** (0.259)	-0.488* (0.265)	-0.425 (0.276)	-0.388 (0.284)	-0.429* (0.220)
stratum2	0.333* (0.191)	0.267 (0.239)	0.108 (0.239)	0.084 (0.250)	0.113 (0.248)	-0.030 (0.228)
stratum3	0.057 (0.244)	0.064 (0.257)	-0.003 (0.252)	-0.028 (0.264)	-0.006 (0.274)	-0.105 (0.307)
stratum4	0.377** (0.163)	0.251 (0.255)	0.181 (0.267)	0.101 (0.267)	0.196 (0.283)	0.210 (0.201)

	hl_ron_corre ct_wpm_z	hl_lsk_corre ct_wpm_z	hl_wr_corre ct_wpm_z	hl_orf2_corre ct_wpm_z	efal_wr3_corr ect_wpm_z	hl_orf2_comp rehension
stratum5	-0.079 (0.153)	0.084 (0.224)	0.095 (0.227)	0.035 (0.245)	-0.038 (0.236)	-0.010 (0.188)
stratum6	-0.068 (0.157)	-0.141 (0.232)	-0.116 (0.212)	-0.135 (0.223)	-0.164 (0.216)	-0.115 (0.199)
stratum7	-0.213 (0.155)	-0.159 (0.226)	-0.193 (0.212)	-0.201 (0.226)	-0.144 (0.225)	-0.131 (0.208)
stratum8	0.024 (0.144)	0.082 (0.228)	0.168 (0.218)	0.138 (0.237)	0.224 (0.246)	0.151 (0.192)
stratum9	-0.385** (0.183)	-0.333 (0.210)	-0.208 (0.196)	-0.238 (0.213)	-0.066 (0.209)	0.090 (0.176)
stratum10	-0.128 (0.171)	-0.308 (0.224)	-0.171 (0.207)	-0.210 (0.225)	-0.053 (0.219)	0.015 (0.196)
treatment × school_type multigrade	-0.110 (0.176)	0.345* (0.179)	0.471*** (0.173)	0.430** (0.175)	0.386** (0.169)	0.332 (0.205)
Num.Obs.	1851	1851	1851	1847	1850	1228
R2	0.033	0.037	0.031	0.026	0.026	0.027

Grade 3: Model A2 (all z-scored outcomes)

Grade 3: Model A2 (all z-scored outcomes)

	hl_lsk_corr ect_wpm_z	hl_ccdk_corr ect_wpm_z	hl_wr_corre ct_wpm_z	hl_orf2_corr ect_wpm_z	efal_wr3_corr ect_wpm_z	efal_orf1_corr ect_wpm_z	efal_orf1_co mprehension
(Intercept)	-0.040 (0.179)	-0.037 (0.160)	0.008 (0.180)	-0.007 (0.159)	-0.071 (0.195)	-0.082 (0.160)	-0.229* (0.131)
treatment	0.191* (0.108)	0.013 (0.084)	0.026 (0.086)	0.050 (0.087)	0.031 (0.083)	0.007 (0.081)	-0.018 (0.091)
school_type multigrade	-0.180 (0.237)	-0.043 (0.153)	-0.080 (0.153)	0.024 (0.148)	-0.015 (0.169)	0.040 (0.153)	0.063 (0.124)
stratum2	0.294 (0.215)	0.026 (0.173)	0.087 (0.188)	0.040 (0.167)	-0.066 (0.184)	-0.045 (0.175)	0.001 (0.145)

stratum3	0.082	-0.121	-0.128	-0.132	-0.210	-0.189	-0.209
	(0.224)	(0.160)	(0.161)	(0.157)	(0.177)	(0.148)	(0.141)
stratum4	-0.050	-0.058	-0.049	-0.090	0.040	0.004	0.042
	(0.209)	(0.181)	(0.190)	(0.168)	(0.221)	(0.180)	(0.158)
stratum5	0.175	0.157	0.081	0.119	0.083	0.117	0.226
	(0.259)	(0.193)	(0.199)	(0.184)	(0.211)	(0.175)	(0.182)
stratum6	-0.100	-0.019	-0.044	-0.037	-0.048	-0.004	0.010
	(0.229)	(0.191)	(0.207)	(0.196)	(0.211)	(0.183)	(0.150)
stratum7	0.200	0.153	0.075	0.060	0.139	0.127	0.222
	(0.227)	(0.192)	(0.213)	(0.190)	(0.226)	(0.195)	(0.157)
stratum8	0.284	0.301*	0.261	0.330*	0.386*	0.437**	0.609***
	(0.228)	(0.183)	(0.208)	(0.197)	(0.217)	(0.195)	(0.182)
stratum9	-0.301	-0.029	-0.085	-0.093	0.052	0.127	0.557***
	(0.204)	(0.175)	(0.192)	(0.182)	(0.204)	(0.179)	(0.167)
stratum10	-0.459**	-0.201	-0.243	-0.301*	-0.051	-0.072	0.331*
	(0.222)	(0.191)	(0.199)	(0.181)	(0.206)	(0.171)	(0.194)
treatment × school_type multigrade	-0.065	0.079	0.049	-0.106	0.056	-0.054	-0.044
	(0.197)	(0.160)	(0.160)	(0.159)	(0.189)	(0.156)	(0.167)
Num.Obs.	1847	1847	1847	1845	1847	1845	1845
R2	0.064	0.022	0.020	0.029	0.022	0.025	0.057

ANNEX 10: Sustainability Methodology and Literature Review

Methods and Data Sources

The sustainability assessment draws on evidence collected across two rounds of data gathering, initially in 2024 and again in 2026 following the conclusion of EGRP II implementation. Primary data includes interviews with seven members of the EGRP II PMT and two additional interviews with representatives of the implementing partner, supplemented by a review of key programme documents including the OUP implementation plans and the UNICEF EGRP II Performance Management Framework. Findings were further enriched by case study work conducted in 2024 and 2025, which involved classroom observations, and conversations with teachers, coaches, and principals, as well as teacher survey data collected in late 2025. District and provincial authorities were interviewed in 2025 rather than 2024, reflecting a deliberate decision to allow sufficient implementation experience to accumulate before soliciting their reflections. Together, these sources provide a layered and longitudinal evidence base that tracks the evolution of embedding conditions over the course of the programme.

Documents reviewed in 2025 included formal letters of commitment between the DBE and UNICEF, as well as meeting minutes from key programme structures, providing additional evidence of institutional intent and the progress of embedding activities during the final phase of implementation.

Study Limitations

Efforts to obtain additional key documents, including the Northern Cape Provincial Reading Plan, were unsuccessful at the time of concluding this report. Where references to this document appear, they are drawn exclusively from accounts provided by provincial respondents and should be read with appropriate caution, as the evaluation team was unable to independently verify its existence and contents or confirm the context in which it was developed.

Literature Review and Reflections from a South African Case Study

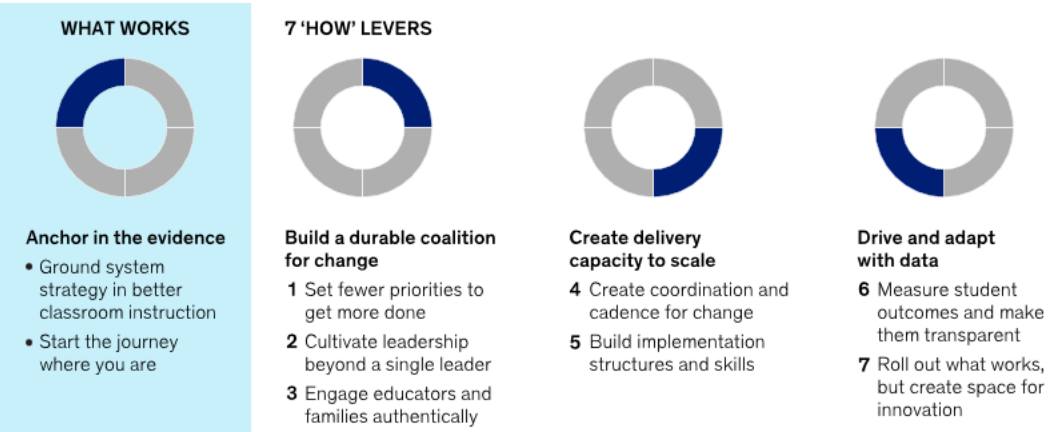
This section summarises global, African and South African literature on key attributes that should be considered when planning and implementing the embedding of educational innovations within broader public education systems. This review offers both a theoretical and empirical basis for the sustainability review of the EGRP II.

Integrating donor-funded education interventions into mainstream public schooling systems at scale has been a subject of extensive global evaluation (McKinsey 2024). Many interventions, even if they can show successful innovation towards improving learning outcomes, face challenges in embedding and scaling effectively.

This is due to a combination of intervention level and system level challenges. McKinsey's 2024 report on improving learning at scale in public school systems emphasises that very few countries and sub-national education systems have succeeded in achieving sustained learning

improvements in recent decades. The report identifies seven levers that enable a schooling system to improve learning overall (Figure 1).

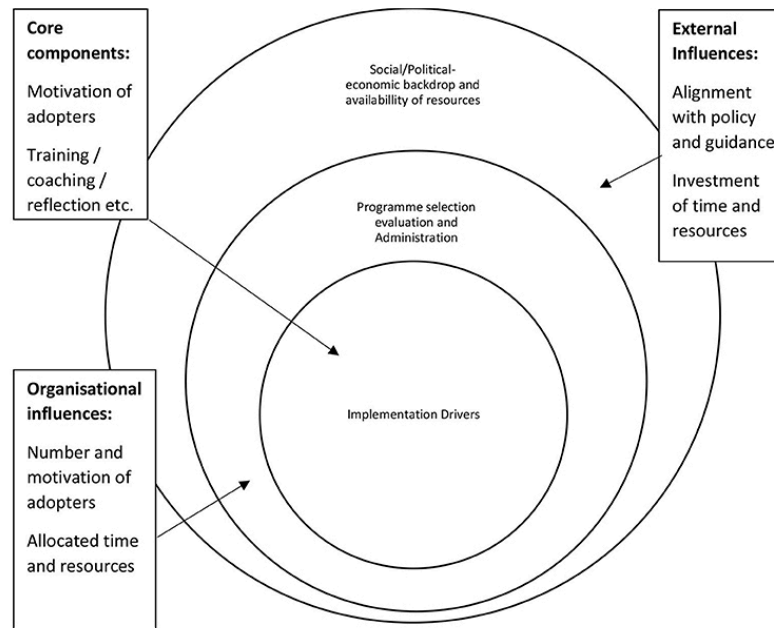
Figure 1: Reinforcing Strategies to Create a Virtuous Cycle of Learning Gains at System Level (Source: McKinsey 2024)



All of these same levers also apply at the intervention level, as discussed below.

The growing field of ‘implementation science’ in education is engaging with the intervention level and why interventions rarely succeed in being embedded in large-scale education systems, or show reduced impacts when they are. Moir (2018) sets out how implementation science brings in a systems perspective that links a specific intervention with “wider political, social and cultural influences” and emphasises that “key internal components of the programme have to be compatible with external influences for full implementation to occur.” Figure 1 (Moir 2018, based on Maher et al., 2009) shows the system levels to be considered.

Figure 2: Conception Model of the Ecology of Scaled Implementation (Source: Moir 2018)



Two types of factors influence the probability that a successful innovation will be sustained and embedded in a public schooling system. The first type relates to how the intervention is designed (enabling design factors located at the ‘implementation drivers’ level in Figure 1) and the second relates to the processes through which the actors responsible for the innovation engage with the wider set of system actors (enabling embedding process factors located at the outer two levels of Figure 1). The second set of factors is predominantly about the nature of institutional relationships, which is why we foreground the term ‘embedding’.

Enabling design factors:

2. **School-Based Professional Development that is Iterative and Participatory:** Programmes that are school-based, peer-facilitated, and focus on active learning have shown promise in shifting learning outcomes systemically because such models are not only educationally effective but also cost-effective and scalable, increasing equitable participation in education (Hassler et al 2021). Engaging teachers and local stakeholders in the iterative development of programmes ensures relevance and acceptance and, therefore, sustainability (Hassler et al 2020).
3. **Context-Responsive Resources & Local Adaptation:** In the international context, when interventions are often ‘imported’ from one country context to another, interventions that do not adequately consider local contexts and fail to adapt to the specific needs of the communities they serve are less likely to be sustained (Burns 2014). In a national context such as South Africa, the same applies to transferring practices across provincial contexts or from urban or rural contexts. Tailoring programmes to fit local cultural, economic, and social conditions is essential for long-term success.

4. **Policy Alignment & Government Leadership:** Interventions not aligned with national education policies or lacking high-level government support often struggle to be integrated into mainstream systems (McKinsey 2024, Piper & Dubek 2024).

Enabling embedding process factors:

5. **Stakeholder Alliance:** Most effective education interventions have multiple integrated components, which cannot all be implemented by a single agency or unit within a government department. Explicit collaboration processes among stakeholders, whether among different units within a government department, different levels of government, or between governmental and non-governmental agencies, must be part of the documented and negotiated operating procedures for an education innovation when scaled.
6. **Financial Resources & Funding Mechanisms:** Inadequate cost estimation, inaccurate translation of 'external' costs (previously held by implementing partners) into 'internal' costs within system actors, and lack of up-front agreements on securing financial resources, whether from governments or donors, often hinder the scale-up and embedding process. Establishing long-term financial strategies is vital for the longevity of educational interventions (Piper & Dubek 2024).
7. **Clarity on which Design Elements Drive Embedding and Sustainability:** complex education interventions may have many different elements, such as materials design and distribution, teacher training, coaching, teaching assistants, etc. They may focus on specific teaching methods or introduce new ways of using languages or teaching aids. The key insight for introducing a complex package of activities and materials into a public schooling system is to
 - a. be clear on which core activities (or which specific combinations of activities) drive impact and which are secondary, and
 - b. be clear on which activities are most likely to be attractive or affordable within the broader system and which may be more challenging.

The most impactful activities may not be the most attractive, and the most attractive may not be the most impactful. Therefore, the relative importance of intervention elements may look very different from the perspective of system actors than from the perspective of intervention designers (and evaluators). This creates the danger that some elements of the intervention design are adopted and sustained but with gaps that undermine the scale of impact. Creating alignment requires intentional discussions on what is important to whom and why.

8. **Clear Communication of a Simple Message (Focused Scope of Change):** While complex intervention designs with many elements may be necessary and appropriate for complex education contexts, they are difficult to communicate and build enthusiasm and commitment around. The implementation of an innovative intervention in its test phases

is, by definition, done by professionals who are committed or contracted to do the work, but in an embedded large scale environment that commitment must first be created and then maintained. This means thinking about aspects of embedding as a 'campaign' with a simple message about the change that is intended and the pathway to that change. It may require a catchy project name or a catchphrase that is easily recognisable and repeatable across multiple levels of the education system. If there are too many elements of change, it is difficult to convey a straightforward, compelling narrative that is focused and coherent. Al-Ubaydli et al. (2019) highlight that major breakdowns during scaling often stem from ineffective communication strategies, underscoring the need for clear and concise messaging to ensure successful adoption.

9. **Clear Implementation Planning with Well-Defined Mandates:** Effective embedding requires detailed plans that specify roles, resources, and monitoring approaches. This includes clarity in roles at all levels from national or provincial officials to district education officials, school leaders, teachers and community members/parents (UNICEF 2024). Roles in relation to the intervention must not only be explicit but also integrated in accountability mechanisms for that position so that the function is not only attached to a specific individual within the position, but it is attached to the definition of the position (e.g. the job description of a Subject Advisor) no matter who is in the position.
10. **Professional Daily Practice (Implementation Capacity):** It is necessary to identify and, where necessary, plan ways of building the required capacities within the system to implement all elements of the intervention. Capacity building cannot be a once-off activity but must itself be a set of processes that are embedded in the system to enable continuous re-generation of capacity as individuals cycle through critical roles and functions. For capacity building to be successful, it is important to recognise that "professional development must be connected to tangible and immediate problems of practice" (City et al., 2010). At the school level, there are limitations that restrict teachers' access to continuing professional development, such as resource constraints, distance, and other challenges. Since schools are where teachers engage professionally, they provide an opportunity to share professional challenges and potential solutions. This sets the stage for professional learning communities to serve as effective avenues for ongoing professional development. Consequently, supporting systems that enhance the functioning of Professional Learning Communities (PLCs) are essential.
11. **Management Capacity for Data-Driven Improvement:** Implementation capacity includes not only the capacity to implement distinct elements of the intervention (e.g. teacher training or LTSM distribution) but to coordinate multiple elements, monitor adherence and quality, and take management decisions based on the monitoring data. Management capacity includes the capacity and political will for **Data-Driven Improvement** so that data is not only used for monitoring but for continuously adapting approaches based on what works in different contexts (UNICEF 2024).
12. **Alignment of Responsibility & Authority = Accountability:** The final aim of a well-structured stakeholder alliance, clearly defined mandates, strong implementation

and management capacity and sufficient financing is that the responsibility and authority for implementing the education innovation sit with the right people and are aligned. If an intervention expects people (e.g. district officials or teachers) to carry out activities but the mandate, funding or performance criteria for those activities lie elsewhere or are not present, the activities are unlikely to be sustained. It is not reasonable or fair to hold system actors (e.g. school managers or teachers) accountable if they do not have the authority and resources to carry out the required activities (reciprocal accountability). They are also unlikely to develop systems to hold themselves and each other accountable in those circumstances (internal accountability).

South African education debates have only recently started engaging in a structured way with the criteria for effective sustainability, embedding and scaling in our context. Few education innovation interventions go beyond **stating the intention** to be sustainable or to embed activities in the public schooling system to **putting in place the activities, resources, time and monitoring systems** necessary to make embedding happen. An exception is the process PILO (an NGO) employed to integrate the Jika iMfundo (JiM) campaign within the KwaZulu-Natal Department of Education (KZNDoE). This process was conceptualised as a distinct project 'phase' (the embedding phase) and the embedding process was independently evaluated. This case provides a valuable opportunity for reflecting on the activities that may also apply for embedding EGRP II within the system.

The JiM campaign was implemented in KwaZulu-Natal (KZN) from 2015-2021 and aimed to improve learning outcomes by enhancing school curriculum coverage. With the backing of multiple donors and a comprehensive consultation process involving a wide range of education practitioners between 2011 and 2013, PILO initiated a collaborative design effort in KZN in 2013, engaging both officials and teacher unions. This collaboration culminated in completing the JiM design framework, with the implementation beginning in 1,200 schools across two districts in KZN in 2015 (Metcalf, as cited in Christie and Monyokolo, 2018).

The embedding process was introduced in 2021 to integrate JiM into the KZNDoE structures and routines (JET Education Services, 2023). This process has focused on the provincial and district levels of the education system. Key aspects of the PILO embedding process included:

4. Provincial Level: The embedding process involved promoting and integrating the JiM initiative into the KZNDoE planning and monitoring systems. This effort included collaborating with key directorates responsible for curriculum, district development, and monitoring and evaluation. PILO aimed to ensure the official adoption of JiM tools and practices by the KZNDoE across these directorates. Activities included participating in planning meetings, sharing documentation to support JiM practices, and collaborating with the Strategic Management, Monitoring & Evaluation (SMME) unit to incorporate curriculum coverage indicators. However, some planned initiatives, such as the JiM steering committee, failed to gain momentum, and funding challenges affected PILO's embedding activities in the province.

5. District Level: PILO concentrated on assisting the District Management Team (DMT) with district planning and leadership development. This involved enhancing the capabilities of the District Training Team (DTT), circuit managers (CMs), and subject advisors (SAs) in curriculum monitoring and support. PILO facilitated integrated planning, provided capacity-building for district leadership, and nurtured relationships with district leaders.
6. School Level: PILO's engagement at the school level was primarily through support to the district, with no direct engagement with schools in the embedding phase. However, the JiM programme developed school-level monitoring tools aligned with the Monitoring, Reporting, and Responding (MRR) process. District officials are expected to support schools in implementing JiM.

The JiM intervention focuses on curriculum coverage as its predominant pathway towards learning improvements, while EGRP II focuses on adherence to a structured learning programme (as expressed through dedicated learning and teaching materials) combined with improved teaching practices supported by coaching. The specific directorates, core activities, success metrics and forms of adherence monitoring at school, district and provincial levels are therefore different, but the broad considerations remain the same.

Metcalf, as cited in Christie and Monyokolo (2018), put forward a number of necessary conditions for the successful embedding of the JiM campaign in KZN. We organise these according to the eleven factors listed above, showing their application within the South African context:

- **School-Based Professional Development** that is **Iterative and Participatory**: The change programme must focus on incorporating new behaviours into the professional practices of educators and officials in ways that respect their professional identities. This gives the officials and teachers the agency to exercise their daily choices rather than simply responding to directives. Metcalfe (as cited in Christie & Monyokolo, 2018)

"Policy intentions and instructions do not change behaviour. These are filtered through what educators and officials already understand and believe, what they are convinced will work in their context, and what they think is possible" (p.22)

- **Context-Responsive Resources & Local Adaptation**: Staying flexible and responsive to the challenges that arose during the embedding process was crucial for sustaining the process. Addressing issues such as leadership vacancies and funding constraints helped maintain momentum, although this is not always straightforward, as these matters are subject to system-wide budget priorities that are beyond the control of involved stakeholders. This meant acknowledging that embedding practices may not occur uniformly across all levels. Tailored support at the school level, for example, was critical to ensure that innovative practices are effectively adopted at the coalface of instructional practices.

- **Policy Alignment & Government Leadership:** JiM was designed to be consistent with government policy priorities and facilitate the integration of these policy intentions into the routine practices of officials and schools. Doing so enhances buy-in and reduces resistance to change, as the intervention is perceived as integral to the officials' everyday work and priorities.
- **Stakeholder Alliance:** A broad coalition of key stakeholders is vital for initiating and sustaining change, particularly in education. The more that various centres of influence align with a shared purpose and sense of urgency, the greater their contribution to success across provincial, district, and school levels. Key stakeholders, such as unions, were part of the JiM alliance, and there was evidence that stakeholders were prepared and ready for change. For the JiM campaign, consultation and engagement with key stakeholders commenced early, and they were actively involved in identifying problems, designing the intervention, and implementing it. This ensured that there was buy-in and ownership from the outset.
- **Financial Resources & Funding Mechanisms:** The programme design was meticulously costed to ensure it could be replicated at scale within the government's resource limitations. This included being explicit about intensity and dosage of all intervention elements to determine whether the intervention will be feasible at a larger scale.
- **Clarity on which Design Elements Drive Embedding and Sustainability:** for JiM, the curriculum tracking tools were the key driving element, even though they were designed to support (rather than drive) more nuanced forms of professional reflection and relationships between teachers and school managers. Nonetheless, those practices would have been much less likely to survive and grow at scale if simple tools, routines and reporting lines had not been developed to frame those practices.
- **Clear Communication of a Simple Message (Focused Scope of Change):** The scope of the change was consistently focused and aligned with the key elements of the theory of change. For the JiM programme, the clear narrative was: "We are improving learning outcomes by enhancing curriculum coverage" (Christie & Monyokolo, 2018).
- **Clear Implementation Planning with Well-Defined Mandates:** PILO established clearly delineated roles and responsibilities within the system to enhance accountability. Placing responsibility for leadership of the process with the province and districts was a crucial principle that should be safeguarded.
- **Professional Daily Practice (Implementation Capacity):** Investing in ongoing professional development for all parties involved was crucial to the embedding process. Enhancing the capabilities of teams at various levels of the system was essential to ensuring the effective implementation and sustainability of practices. Professional Learning Communities (PLCs) are essential and a priority for PILO. Personal commitment to a vision of improved practice was backed by tools, training, and systems

that encourage these behaviours to become routine. These routines were made integral to the district and school operations, reinforced by successful experiences.

- **Management Capacity for Data-Driven Improvement:** At a management level, monitoring and tracking must be integral to the change process. This must occur at all levels; for instance, at the school level, school management teams should monitor teachers' curriculum coverage reports to provide timely support at the appropriate level. A comparable process at the district level involved district teams consolidating school-level curriculum coverage metrics to prioritise responses at the school level. This data-driven approach optimised the otherwise typically constrained district resources, such as subject advisors.
- **Alignment of Responsibility & Authority = Accountability:** Those who will continue to sustain the programme should be responsible for implementing the change. For instance, Metcalfe mentions that the districts and schools themselves own the JiM campaign in King Cetshwayo and Pinetown.

PILO's approach to change is based on two forms of accountability required for school improvement: reciprocal and internal accountability. The former means that for every unit of change in performance that is required, an equivalent level of support and capacity should be offered to those responsible for implementing the change, while the latter is the level of coherence within the organisation regarding norms, values, expectations, and processes for accomplishing tasks, making it a key form of accountability to achieve school-level improvements (Elmore, 1997, 2006, 2010). Internal accountability must be established before external accountability can be effective in well-functioning institutions, suggesting that schools are unlikely to respond to external accountability unless they have established practices of internal accountability (Metcalfe, as cited in Christie & Monyokolo, 2018).

The evaluation of the JiM embedding process also noted that systematic **documentation and communication** of the intervention's core practices and tools and especially information about successful outcomes and impacts can reinforce the importance of the initiative, promote wider adoption and replicate successes. Finally, the PILO evaluators note that embedding requires **persistence and long-term commitment**. Embedding innovations is a long-term process requiring sustained effort, resources, and commitment from all stakeholders involved.

ANNEX 11: Evaluation Team

Tendayi Zhou: Team Lead: Tendayi provided project and technical oversight and management, coordinating and supervising the evaluation team's work, while liaising with the UNICEF evaluation manager, EGRP II PMT, and key stakeholders to ensure coherence throughout the project. He was responsible for ensuring the quality of the process, outputs, methodology, and timely delivery of all products from start to finish, including adherence to ethical standards at all stages of the evaluation. Tendayi also led in report writing and participated in instrument design, testing, and data analysis. Tendayi has 15 years of experience in large-scale multi-country MERL systems design and oversight. In the past seven years, he has led several large-scale education evaluations including the EGRP I. He holds a Master's degree in Public Policy, a postgraduate diploma in Public Sector and NGO Sector M&E, both from Witwatersrand University and an Honours degree double major in Economics and Education from University of Zimbabwe.

Lebogang Shilakoe: Senior Data Manager: Lebogang has managed data collection and data management on large-scale evaluations in both education and non-education projects for more than 8 years, with the latest being the EGRP I Evaluation (2021-2024). She has experience with designing and implementing data quality and assurance plans. She has led the planning and training of data collection teams in EGRA assessments including designing and marking IRRs to ensure that only competent trainees join project teams. Lebo has led the programming and versioning of assessment tools in different software including Tangerine and Survey CTO. She will continue with this role in EGRP II. Lebo is fluent in Sesotho and isiZulu, and has a working knowledge of most South African languages. She holds a BSc (Hons) in Cognitive Psychology and Neuroscience from Jacobs University, Bremen Germany and a PGD Public Sector Monitoring and Evaluation from Witwatersrand University. She is currently studying for her Masters at Witwatersrand University.

Dr Tara Polzer Ngwato: Senior Methodology Advisor: Tara is one of the SII Africa Directors and acted as Senior Methodology Advisory for this project. She has extensive experience in methodology design and brings a system-level perspective to the study, interpreting policy and influencing national systems. She has experience with large-scale mixed methods education evaluations including the NECT Foundation Phase evaluation and EGRP I. She led the team in the development of the National Reading Barometer. Tara has a strong problem-solving orientation, simplifying complex phenomena through collaborative engagements with teams and writing up research findings for diverse policy and academic audiences. She continued with these roles in EGRP II. She holds a PhD in Development Studies from the London School of Economics. She has an extensive publication record.

Dr. Ling Ting: Data Analyst: Ling has worked as a technical expert reviewer for the EGRP I quantitative analysis of learner assessments for both the midline and endline phases of the project. Ling subsequently led the EGRP I study at the endline for descriptive and inferential analysis, in EGRP I, ensuring continuity with the previous EGRS studies and other related studies. She has strong quantitative analysis expertise in statistical analysis, econometric

modelling, data wrangling, data analysis, and report writing. Ling holds a PhD in Behavioural Economics from University of Johannesburg. She has developed a strong technical ability in the analysis of education data and learner assessment interpretation.

Dr Martin Prew: Senior Technical Education Advisor: Martin is an education specialist with broad experience in early-grade reading assessment, school management, and curriculum development. His work encompasses programme evaluation, policy development, and capacity building, making him a valuable asset to the team on this project. He served as the education expert for EGRP I, where he led the design and adaptation of EGRA instruments, developed case study methodologies and data collection tools, and oversaw coach shadowing, case study data analysis, and reporting. He continued in this role for EGRP II, including training field researchers. He led the case study data collection. Martin is well-published and holds a PhD from the School of Education at Exeter, UK.

Maggie Kgang: Education Researcher: Maggie has extensive experience in data collection within educational and school settings. She has conducted classroom observations, and interviews with learners, facilitators, teachers, and school management teams, and has written reports on her findings. She has also participated in similar evaluations. In this programme, her role involved data collection during two rounds of case studies, observing teacher and coach training, and leading focus group discussions with learners. Maggie is a native speaker of the Setswana dialect spoken in the Northern Cape and is fluent in Zulu, Afrikaans, and Sesotho. She is currently pursuing her PhD in Humanities Education (University of Pretoria) and holds a BA Psychology (University of Johannesburg), B.ED Honours in education Psychology and, M.ED in Gender Education from University of KwaZulu Natal

Siziwe Sangulukani: Administrator & Logistics Manager: Siziwe has been an administrator of research projects for the past 23 years with extensive experience in large-scale education projects including the NECT Foundation Phase Learning Program Evaluation and EGRP I Evaluation. She managed to contact schools, and field logistics, schedule data collection visits, and communicate any changes to them. She has led the recruitment of field researchers, contracting them, and managing payments in addition to organising the hiring of cars, accommodation, and other needs FR may have whilst in the field. She worked closely with field managers and the project manager. Siziwe is fluent in isiXhosa, Sesotho, Setswana, and isiZulu. Sizi holds a diploma in financial management.

