

The Second Early Grade Reading Study

Year 1 Report

Learner performance after one year of implementation



Contents

Executive Summary	7
Project overview.....	7
Evaluation findings	7
Next Steps in the project.....	8
1. Why focus on Early Grade Reading in English as First Additional Language?	9
2. Research Context	10
2.1. The evolution of Early Grade Reading Studies in South Africa	11
2.2. South African EFAL Curriculum.....	12
3. Study Design.....	14
3.1. Research Site	14
3.2. Interventions	15
3.3. Theory of Change	18
3.3.1. Common aspects of the two interventions	18
3.3.2. Intervention 1 specifically.....	19
3.3.3. Intervention 2 specifically.....	19
3.4. Evaluation.....	20
3.4.1. Overview of Evaluation Methods	20
3.4.2. Sample selection and intervention assignment.....	21
3.4.3. Instrument Design	23
3.4.4. Data Collection	24
3.5. Ethical Clearance	26
3.6. Sample Characteristics	26
3.7. Balance at Baseline.....	28
4. Implementation Fidelity	29
4.1. Intervention 1 and 2.....	29
4.2. Intervention 2 specifically.....	30
5. Midline Results	32
5.1. Attrition	32
5.2. Item level learner assessment results	33
5.2.1. Sub-task 1: Home Language expressive vocabulary	35
5.2.2. Sub-task 2: Home Language word recognition.....	36
5.2.3. Sub-task 3: English phoneme isolation	37
5.2.4. Sub-task 4: English letter recognition.....	37
5.2.5. Sub-task 5: English word recognition	39

5.2.6.	Sub-task 6: English Listening.....	39
5.2.7.	Sub-task 7 – 9: English Listening Comprehension, Vocabulary and Spelling	40
5.3.	Main Results	41
5.4.	Intervention effects on sub-groups of interest.....	43
5.4.1.	Learner level.....	44
6.	Changes in instructional practices	46
7.	Sensitivity Checks	49
7.1.	Switching schools	49
7.2.	Multi-grade and attrition.....	50
7.3.	Improvements due to increase in time spent teaching EFAL	51
7.4.	Crowding out of home language teaching time	53
8.	Cost-effectiveness discussion	54
9.	USAID indicators.....	55
10.	Next Steps in the project.....	57
	References.....	58

List of Figures

Figure 1: Year 2 results of EGRS I	12
Figure 2: Map of the research site	14
Figure 3: Did you receive any in-service training in the teaching of EFAL?	30
Figure 4: Average number of days taught, per week.....	31
Figure 5: Lesson plan coverage	32
Figure 6: Attrition rate by intervention group.....	32
Figure 7: Sub-task 1, HL vocabulary	35
Figure 8: Percentage of learners who could not read a single word correctly.....	36
Figure 9: Phoneme Isolation - Percentage correct by intervention group	37
Figure 10: Distribution of letters read correctly by intervention group	38
Figure 11: Percentage of learners correctly responding to each English instruction	40
Figure 12: Percentage correct on English Listening Comprehension and Vocabulary.....	41
Figure 13: Mean comparison of Index Scores	41
Figure 14: Differential impact based on Baseline performance	46
Figure 15: Frequency of English used in the classroom.....	47
Figure 16: Instructional practices in shared reading.....	48
Figure 17: Instructional support.....	49
Figure 18: Time spent on teaching EFAL per week	52

List of Tables

Table 1: Comparison of intervention 1 and intervention 2	17
Table 2: Sub-tests contained in the instruments at each point in time	23
Table 3: Percentage of learners tested during baseline data collection	25
Table 4: Percentage of teacher interviews conducted during midline data collection	26
Table 5: Percentage of Home Background Questionnaires returned	26
Table 6: Sample Characteristics	27
Table 7: Balance tests per sub-task	28
Table 8: Balance test on Teacher English Proficiency	29
Table 9: Teacher attendance at training sessions	29
Table 10: Using technology in the classroom	30
Table 11: Item descriptive statistics	34
Table 12: Item correlations between baseline and midline	35
Table 13: Comparing letter-sound recognition between Mpumalanga and North-West	39
Table 14: Sub-task mean comparison scores, by intervention group	42
Table 15: Main regression	43
Table 16: Heterogeneous effects - Baseline performance	45
Table 17: Robustness check - Re-allocating intervention status	50
Table 18: Sensitivity check - Multi-grade and Attrition	51
Table 19: Restricting to only include learners in 3-hour a week classes	52
Table 20: The effect of crowding out Home Language Teaching	53
Table 21: Specific USAID Indicators	55
Table 22: Custom USAID Indicators	56

Acronyms

ANA	Annual National Assessment
CAPS	Curriculum and Assessment Policy Statement
DBE	Department of Basic Education
EFAL	English as First Additional Language
EGRA	Early Grade Reading Assessment
EGRS	Early Grade Reading Study
EGRS II	Second Early Grade Reading Study
GPLMS	Gauteng Primary Literacy and Mathematics Strategy
HOD	Head of Department
ICT	Information Communication Technology
LoLT	Language of Learning and Teaching
LTSM	Learning and Teaching Support Material
NECT	National Education Collaboration Trust
PIRLS	Progress in International Reading and Literacy Study
RCT	Randomised Control Trial
RCUP	Reading Catch-Up Programme
SMT	School Management Team
USAID	United States Agency for International Development

Executive Summary

Project overview

The second Early Grade Reading Study (EGRS II) forms part of a series of studies conducted by the South African Department of Basic Education (DBE) in collaboration with the University of the Witwatersrand and the United States Agency for International Development (USAID). Building on the lessons learned through the first Early Grade Reading Study (EGRS), EGRS II aims to evaluate two models of supporting and strengthening the teaching of English as First Additional Language (EFAL) in the Foundation Phase. This study therefore aims to contribute to the evidence base about what works to improve the learning and teaching of early grade reading and will therefore guide the DBE in the implementation of programmes going forward.

The study is designed as a Randomised Control Trial (RCT) which randomly allocated schools to two intervention groups, and one control group where regular schooling continues. This means that on average the three groups of schools are exactly the same and that the only difference between them is the manner in which the teachers are trained and supported in the teaching of EFAL. At the end of the study, we will therefore be able to see whether the learners in the intervention groups learned more than the learners in the control group and thereby determine whether the interventions have been successful.

The interventions are intended to improve and strengthen teachers' implementation of the EFAL Curriculum and Assessment Policy Statement (CAPS) curriculum in the Foundation Phase. Both interventions consist of three components: (1) lesson plans, (2) integrated learning and teaching support material and (3) instructional coaching and training. The key differences between the two interventions lie in the delivery mode of coaching support that teachers receive and the format in which teachers receive the lesson plans. In both interventions, teachers receive initial block training, as well as one-day training at the start of each school term. In the first intervention, the teachers receive a paper-based version of the lesson plans and benefit from regular on-site coaching. In intervention 2, the teachers receive a tablet with an electronic version of the lesson plans, including various audio-visual resources, and are supported through a virtual coaching model that includes telephone calls and cell phone messaging to the teacher by a virtual reading coach.

Evaluation findings

To evaluate the success of each of the interventions after the first year of intervention, the language abilities of 20 learners in each school that forms part of the study sample was tested at the start of Grade 1 (January 2017). At the end of Grade 1 (October/ November 2017), we tested the exact same learners again to determine whether there is any difference in the learner performance.

The year 1 results seem to suggest that both interventions have been equally successful in improving the English oral language proficiency of the Grade 1 learners. Learners in the two intervention groups were found to be significantly ahead of their control group counterparts on the core instructional outcomes associated with the Grade 1 curriculum for learning English as an additional language. The scope of learning in the curriculum for Grade 1 English learning is oral language development, with reading and phonics being addressed through methods such as shared reading, listening to stories, and total physical response. All three of these practices essentially focus on receptive language proficiency and to a lesser extent on expressive language development practices.

For the three subtests that focus on English oral language development - English Listening (and following instruction with actions), English Listening Comprehension, and English vocabulary - the learners in the two coaching interventions did significantly better than their counterparts in the control schools. On the higher order skill subtest that assessed children comprehension in English, learners in the intervention classrooms did only marginally better than children in the control schools, suggesting that while vocabulary development is stronger as a result of the interventions, this has not yet translated into stronger comprehension skills. That said, if oral vocabulary development in the second language is a developmental building block for reading acquisition, then the interventions may be working to improve the basic skills targeted in the Grade 1 curriculum but are not yet impacting the higher order skills.

The lack of positive impact (and even weak indications of some negative effects) on home language subtests and on decoding outcomes (which is really only targeted in the home language curriculum) raises the possibility that the English intervention may be successful through crowding out teaching time in the other areas of the curriculum. The results on the other learning areas are, however, very tentative at this stage and warrant further investigation through subsequent rounds of data collection before any concrete statements can be made.

Next Steps in the project

The two interventions continued throughout 2018, with the initial training having been conducted with the Grade 2 teachers in January 2018. Teachers further attended clustered workshop training for one day at the start of each term, and the reading coaches provided on-going support throughout the year. The third round of data collection will take place from 22 October to 9 November 2018 and will provide the data necessary to determine the impact of the interventions after two years of implementation. The interventions are set to continue to Grade 3 teachers in 2019 and the same learners will be tested at the end of Grade 3 in 2019 and again at the Grade 4 in 2020. In October 2018 a set of case studies were conducted to gain more detailed information on the aspects of the intervention which may be driving the success. Similarly, lesson observations in 60 schools and a further set of case studies are planned for 2019.

1. Why focus on Early Grade Reading in English as First Additional Language?

Reading is often used as an indicator of how well an education system is delivering on its mandate. International assessments such as the Progress in International Reading Literacy Study (PIRLS) have shown that at Grade 4 and Grade 5 level, South African learners have not yet learned to read with meaning. In the 2011 prePIRLS assessment, 29% of learners did not even reach the low international benchmark and therefore, when reading information texts these learners could not make any inferences about logical connections to provide reasons or interpret obvious reasons and causes when reading literary texts (Howie, et al., 2012). The results of the PIRLS Literacy 2016 further confirmed the literacy crisis in the country where 78% of South African learners who were tested could not reach the same Low International Benchmark. By these grade levels learners are required to use these specific skills to accumulate all further knowledge, signifying that for the largest majority of South African learners, further learning will always be constrained.

South Africa, like many other countries on the African continent, has a rich linguistic heritage with eleven official languages. The language policy in South Africa explicitly promotes primary home language instruction in the foundation phase and the choice of the specific language of learning and teaching (LoLT) in the foundation phase is left to the governing bodies of schools. From Grade 4 onwards, learners are taught most subjects in either English or Afrikaans, but English is the language chosen by the majority of schools. Although African languages are not used as the LoLT from Grade 4 onwards, African language speakers are expected to continue studying their home language as a school subject until Grade 12. To ameliorate the language transition learners face in Grade 4, English is introduced as an additional language from Grade 1. The language policy therefore promotes an additive approach to bilingualism where the home language is developed together with the additional language (Matjila & Pretorius, 2004).

Additive bilingualism is an approach echoed in the EFAL Curriculum, where proficiencies in the first language are to be used as a base for developing English proficiency. However, it is not always the case that the CAPS is implemented as intended, which may delay the successful acquisition of English, the language that will ultimately be used for instruction from Grade 4. In South Africa, the literacy crisis therefore can be explained by the poor acquisition of literacy in the first language, compounded by the inadequate implementation of the EFAL curriculum.

Strengthening the teaching of English in the foundation phase is therefore critical to effecting improved learning performance in the later grades. In understanding the main constraints to more effective teaching, school effectiveness studies in South Africa have found that primary school classrooms are mainly characterised by a lack of print material, a lack of opportunities for reading and writing, chorusing practices, low levels of cognitive demand and slow pacing (Taylor, 2007; Hoadley, 2012). However, while a range of factors influences learning outcomes, it is widely recognised that instruction or instructional practice is critical to improving learning

outcomes (Coe, et al., 2014). One of the key characteristics of South African education is that the dualistic nature of learning outcomes between the wealthy and the poor is mirrored by dual types of instructional practice happening in the schools serving these communities (Hoadley, 2012). It is likely that weak instructional practices have a causal impact on learning outcomes in the poorly performing part of the school system. To substantially shift achievement in the weak part of the schooling system it may be necessary to apply a comprehensive instructional change intervention, involving a set of coherent and aligned instructional inputs. For this study, the instructional inputs include lesson plans, aligned learning materials and in-classroom support to teachers.

The study is designed as a randomised control trial (RCT) to allow the robust evaluation of the causal impact of the two interventions trialled. This impact evaluation will thus inform credible policy decisions around interventions that have been implemented on a relatively large scale. The main benefit of the RCT design is the inclusion of a randomly selected control group in the evaluation, which allows the measurement of the amount of learning that would have taken place in schools, should the interventions not have been implemented. This, in turn, provides a benchmark to which the learning gains in the intervention groups can be compared in order to establish the additional learning gained by implementing the interventions. Furthermore, the comparison of the two interventions with the control group also allows a direct comparison of the costs involved to affect the learning gains, and therefore allows the determination of the cost-effectiveness of each of the interventions.

The study aims to consider four main research questions:

1. Did the face-to-face support model improve learner English reading proficiency?
2. Did the technological support model improve learner English reading proficiency?
3. Did the impact on reading proficiency differ between the two models of training and support delivery?
4. Which model of delivery is the most cost- and resource-effective?

The primary implementing partner is the South African government, in particular, the Department of Basic Education. The national Department has partnered with the Provincial Education Department in Mpumalanga and the University of the Witwatersrand, to implement and evaluate the interventions in the province. Service providers have been appointed to assist with the implementation of the interventions, as well as the data collection for the evaluation of the programme. The evaluation side of the project is being supervised by the Research Team. The study is completely funded through USAID, with the funding housed and administered through the Wits Health Consortium.

2. Research Context

There is a desperate need to improve the teaching of early grade reading in South Africa, but very little robust evidence on programmes or interventions which have been able to shift

practice at a systemic level. The purpose of the government-led initiative, which has become known as the Early Grade Reading Studies, is to influence policy based on rigorous evidence. A series of three experiments, led by the Department of Basic Education in collaboration with academics, have aimed to evaluate the effectiveness of structured learning programmes through various modes of delivery.

2.1. The evolution of Early Grade Reading Studies in South Africa

The Gauteng Primary Language and Mathematics Strategy (GPLMS) was the first major intervention that made use of a structured pedagogic programme and instructional coaching at a systemic level. The programme was implemented across the Gauteng Province from 2011 to 2014 and included lesson plans, quality educational materials and instructional coaching in underachieving primary schools. Unfortunately, the programme was not rolled-out in a way that allowed a rigorous evaluation of its effectiveness, but using a regression discontinuity design, Fleisch et al (2017) found that the programme showed strong promise. The suggested effectiveness of the GPLMS provided the motivation for exploring structured pedagogic programmes further in the South African context.

The first experiment to evaluate the effectiveness of structure pedagogy was the “Reading Catch-Up Programme” (RCUP) study, which used on-site teacher coaches over an 11 week time period to boost the teaching and subsequent learning of English reading amongst Grade 4 students whose home language was not English.¹ The catch-up programme had reportedly been very successful in improving English reading in the Gauteng province (Hellman, 2012). Moreover, the short duration of this programme renders it relatively cost- and resource effective, since reading coaches could be rotated between different schools each term. The implementation of the catch-up programme in Gauteng was, however, not independently evaluated and had no counterfactual. An RCT was therefore conducted in the Pinetown district of KwaZulu-Natal to evaluate the impact of the programme. Fleisch et. al (2017) found that the average impact was not significantly different from zero, although children with a higher baseline English proficiency did register statistically significant gains. Two key lessons were gained through this experiment. Firstly, the language foundation of most students was too low to benefit from an additional support programme and pointed to the need for interventions prior to Grade 4. Secondly, the duration of coaching was too short to effect real change.

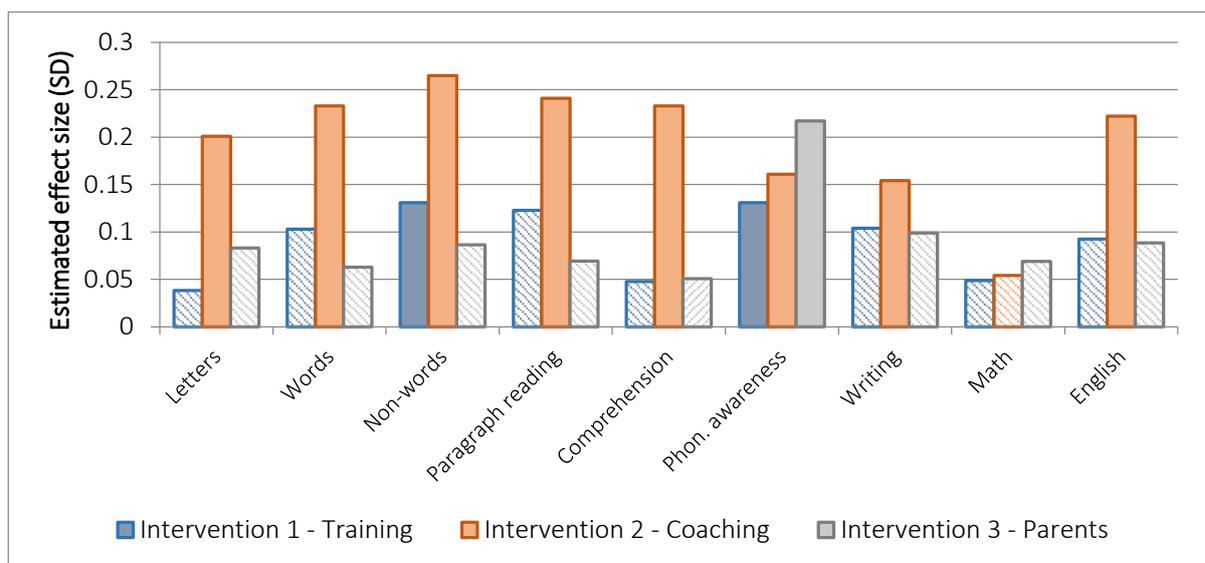
Building on these lessons, the first EGRS 1 was designed to introduce a structured pedagogic programme including reading coaches for an extended period in the Foundation Phase. Starting in 2015, EGRS I specifically set out to evaluate the effectiveness of the on-site coaching approach versus the traditional government training model - centralised teacher training. Both interventions were implemented in a group of 50 schools which received the same daily lesson

¹ In South Africa, most children learn in their home language as the language of instruction in the first three grades and then transition to English as the language of instruction in the fourth grade.

plans and additional reading materials. A further 80 schools served as the control in which schooling continued regularly. The interventions specifically targeted the teaching of reading and literacy in learners’ home language, which in the case of EGRS I was Setswana.² The centralised teacher training occurred for two days at a time, at the start of each year and again midway through each year. The major cost in this intervention was the travel and accommodation for teachers. The on-site coaching intervention included a half-day cluster meeting to disseminate lesson plans and reading materials at the start of each quarter, as well as on-site coaching visits about once a month. The ratio of coaches to schools was roughly 1:17 and the main cost in the coaching intervention was the salaries of coaches. The coaching intervention was therefore about 30% more expensive than the centralised training intervention. The impact evaluation showed that after two years of the implementation of the interventions, the on-site coaching intervention had returned an average home language test score impact of 0.24 standard deviations relative to control. The impact of the central training intervention was half as large at 0.12 standard deviations. More importantly, the coaching intervention had a statistically positive impact on all of the early grade subtests, as well as on the English items (Figure 1). Considering the cost-effectiveness of the two interventions, the coaching model proved to be more cost-effective with a 0.41 standard deviation increase in test scores per USD 100.

Having shown that coaching plays an important role in shifting teacher instructional practice to improve learner reading proficiency, it was necessary to move the research agenda to focus on questions about the financial viability of rolling-out coaching system-wide. Further questions also remain about the ability to recruit, train and manage the larger number of reading coaches, and whether the structured learning programme is effective in teaching EFAL in the foundation phase. EGRS II therefore sets out to consider these research questions.

Figure 1: Year 2 results of EGRS I



² During the first three grades there are four learning areas required by the South African curriculum: Home language literacy, First Additional Language (which is usually English), Numeracy and Life Orientation.

As mentioned above, the intervention lesson plans intend to strengthen the implementation of the South African EFAL curriculum. The EFAL curriculum comprises 4 learning areas in the Foundation Phase. These are:

1. Listening and Speaking
2. Reading and phoneme awareness/phonics
3. Supported writing and handwriting
4. Language use

Thinking and reasoning, as well as language structure and use, are integrated into the domains above during Grade 1 and Grade 2, but language use becomes a full learning area in Grade 3.

The EFAL curriculum takes an additive bilingualism approach which intends for EFAL proficiencies to be built on home language proficiencies. For example, if a learner can identify the letter “m” in their home language, it is easier to build the English knowledge of “m” onto this. More time can then be spent on areas where there are not similarities e.g. the voiceless dental click “c” in *cula* (*sing* ~ isiZulu) vs the voiceless velar plosive “c” in *cat* (English).

In Grade 1, the EFAL curriculum prioritises the development of English oral proficiency. Learners are only expected to gain incidental reading exposure, which gradually transitions to more explicit reading and writing instruction in Grades 2 and 3. The curriculum further highlights the need for maximum time devoted to EFAL instruction especially for those learners who have to transition to English as the LoLT in Grade 4.

The time allocated for all languages in Grades 1 and 2 is 10 hours each with 11 hours in grade 3. In Grades 1 and 2, the time allocation for EFAL can vary between 2 hours minimum to 3 hours maximum and this affects the time allocated for home language instruction. Thus, if 3 hours maximum time is allocated to EFAL, 7 hours are allocated to home language in Grade 1. In Grade 4, EFAL can vary between 3 and 4 hours, and again affects the time allocated to home language instruction. Based on this time allocation, it is important for teachers to maximize similarities between the home language and English so that they can use their time effectively on vocabulary and reading and writing development in English.

The EFAL curriculum therefore rests on the following assumptions:

- English oral proficiency is necessary for English literacy acquisition
- 3 hours a week of English instruction for three years is sufficient for learners to develop the necessary conceptual vocabulary and understanding to transition to English medium instruction in year 4
- There is a strong home language literacy foundation to springboard English literacy proficiencies off
- Teachers are able to use English as the main language of instruction in EFAL lessons to increase learners’ exposure to English

It is within this framework that the EGRS II lesson plans were developed.

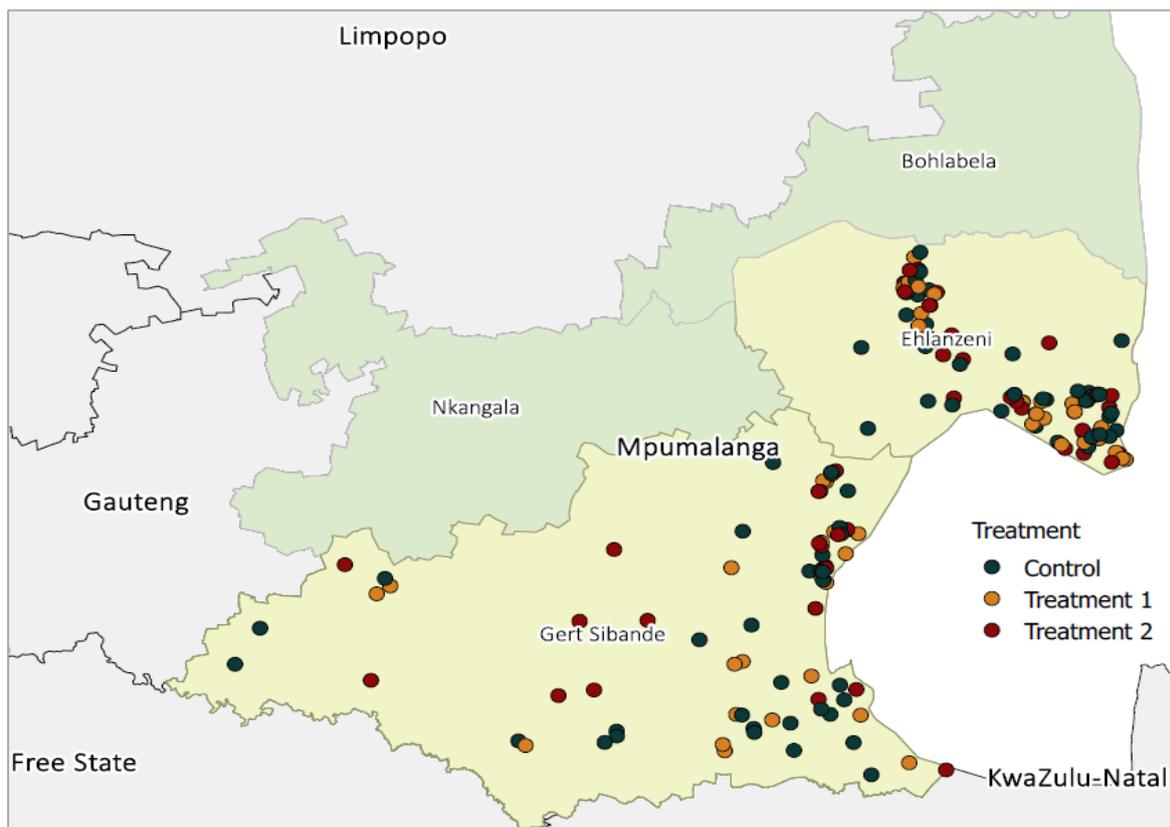
3. Study Design

The EGRS II is designed as a Randomised Control Trial (RCT), which aims to measure the effects of each of the interventions and compare it to the situation among a comparison group of learners where typical teaching is taking place. The study entails the implementation of two early grade reading interventions in South Africa in 180 primary schools in the Mpumalanga Province from 2017 to 2019. The EGRS II focuses on the early learning of EFAL and the interventions consist of lesson plans, additional reading resources and instructional coaching and training. Two alternative strategies to training and coaching are used: (i) the traditional face-to-face format, (ii) a combination of face-to-face training and an Information and Communication Technology (ICT) component that includes scripted lesson plans on tablets and cell phone messaging to the teacher. The sections to follow will elaborate on the research site, the specific interventions, the theory of change, the sample selection, the instrument design and the midline data collection.

3.1. Research Site

The study is being implemented in the Gert Sibande and Ehlanzeni districts of Mpumalanga. Relative to the other provinces in South Africa, Mpumalanga is a mid- to low- performing

Figure 2: Map of the research site



province. In the 2016 Matric Examination³, Mpumalanga ranked fifth out of the nine provinces. Unfortunately, South Africa does not currently have a standardised assessment with which to compare schools, districts and provinces in the primary phase, but various other indicators can provide an indication of the current schooling conditions. The 2016 General Household Survey reports that 94.3% of 5 to 6-year-olds in Mpumalanga are currently attending an educational institution and 91.4% of Grade 1 learners reported having attended Grade R prior to Grade 1. Poverty analysis shows that 28.4% of learners attending schools in Mpumalanga fall below the food poverty line (monthly per capita income is below R442.00) and 47% below the Lower Bound Poverty Line (monthly per capita income is below R660.00).

3.2. Interventions

The EGRS II focuses on the early learning of EFAL by providing specific resources, training and on-going support to teachers. The interventions aim to support Foundation Phase teachers in their instruction of EFAL, in accordance with the DBE's National Curriculum Statement, including the CAPS. The interventions are intended to improve and strengthen teachers' implementation of the EFAL CAPS curriculum in the Foundation Phase. Both interventions consist of three components: (1) lesson plans, (2) integrated learning and teaching support material and (3) instructional coaching and training.

The key differences between the two interventions lie in the delivery mode of coaching support that teachers receive and the format in which teachers receive the lesson plans. In both interventions, teachers receive initial block training, as well as one-day training at the start of each school term. In the first intervention, the teachers receive a paper-based version of the lesson plans and benefit from regular on-site coaching. In Intervention 2, the teachers receive a tablet with an electronic version of the lesson plans, including various audio-visual resources, and are supported through an ICT coaching model that includes telephone calls and cell phone messaging to the teacher. The electronic lesson plans in the second intervention are delivered using an application which is specifically developed for the study, and the additional electronic resources include short training videos, sound clips of the phonics sounds, songs and rhymes and examples of learners' work. The focus of both the interventions is explicitly on how to deliver EFAL instruction and on how to effectively use the new materials as well as other available materials - especially the government-provided workbooks - to most effectively teach reading in accordance with the National Curriculum Statement Grades R – 12.

Intervention 1 is implemented in 50 randomly selected schools and provides teachers with regular face-to-face coaching, as well as quarterly training workshops. The training workshops in both interventions are similar and focus specifically on supporting teachers with the integration of the lesson plans and Learning and Teaching Materials (LTSM) into the teaching of EFAL, as well as on various methodologies for the teaching of reading and literacy in English.

³ The final school leaving examination in South Africa. This examination is standardised across all schools in South Africa, and is the only credible standardised assessment which can be used to make cross-province comparisons.

Unlike typical training workshops, this is best understood as ‘just-in-time’ training: not on general principles, but specifically on how to implement the EFAL literacy programme in the upcoming weeks. The provision of specialist reading coaches is an alternative model of support, involving regular (once a month) in-school coaching. The reading coaches play a number of distinct and overlapping roles, firstly in modelling new practices and secondly in monitoring implementation fidelity. As the teachers become more familiar with the new practices, the coach’s role shifts from introducing new practices towards that of ‘critical friend’, observing and providing real-time input to teachers as they begin to master the new practice. The coaches also play an important monitoring role, providing regular feedback to the project management team on the level and quality of programme implementation. The monitoring is always done in an encouraging manner, which therefore contributes to teachers being more motivated and encouraged to implement the programme throughout the year. The reading coaches also initiate needs-driven clustered workshops throughout the school term with identified teachers invited to attend. Finally, teachers in both interventions are supplied with sets of LTSM including graded reading booklets, posters, big books, sentence strips and vocabulary words.

Intervention 2 is implemented in a different group of 50 randomly selected schools and uses a different model of teacher support and mode of delivering the lesson plan. The technology-supplemented intervention aims to provide new forms of support and guidance on teaching strategies through a range of resources, electronically available lesson plans and interactive support platforms that are available at all times to the teacher. The application further includes various resources such as video clips on best practice, and audio clips of English sounds, the songs and the rhymes in the lesson plans. The intervention makes use of WhatsApp/text messaging to create virtual reading coaching practices and virtual communities of practice. The virtual reading coach uses instant messaging to communicate with teachers regularly, providing them with teaching tips on a weekly basis, and answering any questions they may have on the lessons that they have taught or are preparing to teach the following day. The regular communication is important in building trust in the virtual reading coach as well as in the overall intervention. Secondly, the virtual reading coach also helps facilitate virtual teacher groups using an appropriate platform such as WhatsApp. This platform assists in building teacher networks where they can share their experiences and assist each other with useful tips and techniques.

Table 1. Comparison of intervention 1 and intervention 2

	Intervention 1	Intervention 2
Provision of lesson plans	Paper-based	Electronic On an application on a tablet
Provision of LTSM	Paper-based: <ul style="list-style-type: none"> - Big books - Posters - Flashcards - Writing frames 	Paper-based: <ul style="list-style-type: none"> - Big books - Posters - Flashcards - Writing frames
Coaching	Coach visits the teacher in her classroom. Once every three weeks.	Coach contacts the teacher via telephone calls and instant messaging (Whatsapp). Once every two weeks.
Training	Initial training: 2-day block training Quarterly training: 1 day at the start of each term Needs-based training: As required	Initial training: 3-day block training Quarterly training: 1 day at the start of each term Needs-based training: None
Core methodologies	Paper-based instructional manual	Application-based instructions, Includes videos, sounds clips and photos of example writing

The training and coaching for teachers in both interventions aim to improve teacher content knowledge as well as their instructional practice in the classrooms. The core methodologies which are covered in EGRS II training and coaching are centred on:

- The use of display boards
- Daily activities for listening and speaking
- Shared reading
- Phonemic awareness
- Writing
- Assessment
- Classroom management and environment

During training and coaching, the coaches explain the purpose of each methodology and model for the teacher how to implement it in the classroom based on the lesson plans and provided LTSM. These core methodologies are in line with the learning areas in the CAPS.

3.3. Theory of Change

EGRS I showed that the combination of lesson plans, integrated LTSM and coaching was particularly important in affecting sustained behaviour change in the teaching of home language. In EGRS II, we are evaluating whether the same results can be obtained when applying the same programme in a different province and in a different subject (EFAL). Furthermore, we are evaluating whether on-site, face-to-face coaching is essential, or technology can be utilised to reduce the costs of coaching with the same effects.

The Early Grade Reading Study II aims to effect behaviour change among teachers at a large scale, in line with both the curriculum and methodologies in which teachers were trained during the teacher training at the start of the programme. The lesson plans form the foundation of the interventions by integrating the different components of the programme. The lesson plans provide a focus for the coaching interactions and ensure the appropriate use of the learning and teaching support materials. The lesson plans therefore provide a mechanism to prompt the enactment of the behaviour change, whereas the coaching serves as an additional mechanism to encourage fidelity to the programme. The role of the learning materials is to provide the appropriate resources to ensure that learners are able to develop and consolidate knowledge and skills related to reading fluency, vocabulary development and shared reading.

Given the perceived high costs of direct in-class coaching, the study tests the relative cost-effectiveness of two kinds of capacity building: the traditional face-to-face model and the new combined face-to-face model supplemented by electronic support in the form of a tablet with video demonstration lessons and cell phone messaging support. The technology-supplemented intervention aims to provide new forms of support and guidance on teaching strategies through a range of materials, teaching guides, videos and interactive support platforms that are available at all times to the teacher. While the Early Grade Reading Programme cannot address all the learning challenges, particularly for learners with a severe learning disability, it aims to strengthen English reading performance for the majority of learners across the performance distribution.

3.3.1. Common aspects of the two interventions

The lesson plans provide specifications of the new instructional practice including faster-paced instruction, more appropriately sequenced content, and dramatically expanded pedagogic repertoires. In the primary school teaching of reading in EFAL, the newly expanded repertoires include the systematic teaching of phonemic awareness and phonics, strategies that focus on increased reading speeds or fluency, shared reading strategies, vocabulary development and strategies that improve comprehension. The lessons require little additional lesson preparation from teachers, which makes the take-up of a more productive teaching practice more manageable. The faster-paced instruction (relative to typical practice) also ensures that the teachers cover the full prescribed curriculum for the year. Furthermore, lesson plans free up

teachers' time, because they no longer need to allocate as much time to planning. This could improve reading acquisition if teachers allocate this time to productive teaching activities.

The role of the learning materials is to provide the appropriate resources to ensure that learners are able to develop and consolidate knowledge and skills related to English language proficiency, English reading fluency, English vocabulary development and guided reading. Flashcards, Big Books and Posters are provided to teachers to use in an integrated manner when teaching vocabulary, phonics and reading. From Grade 2 onwards, ten titles of graded reading books will be provided for each classroom. The accompanying graded reading materials provide ample material for learners to practice decoding and reading at their level of development. It is recognized in South Africa that the opportunity to learn EFAL may be hindered by a lack of suitable materials to assist in the progression from one phase of reading acquisition to the next, and this is likely to be particularly true in African language schools, the focus of our intervention. Furthermore, teachers are required to provide regular assessment of learners' reading proficiency in order to assign learners to the appropriate graded readers and small reading groups, based on ability.

In addition to the above, coaching is provided as more intensive training to improve teacher capacity. In essence, the role of coaching support is to fuse capacity building and accountability. The assumption is that, just like learning to read, the ability to teach is a skill that needs to be developed over time and might not be accomplished in one day of training. Furthermore, the coach also plays the role of a 'champion' who keeps the teacher accountable for implementing the programme through encouragement and motivation.

3.3.2. Intervention 1 specifically

In intervention 1, specialist reading coaches visit each intervention school about once a month. The in-class support allows for the modelling of the new practice on site and the gradual development of teachers in the new practice from novice to expert. The in-class support also allows teachers to manage the emotional labour, i.e. stress, insecurity and anxiety associated with developing a new professional practice mid-career. The presence of the in-class support allows for the development of professional accountability in an environment of trust, where the coach monitors and evaluates the teachers' teaching practices in order to encourage more productive teaching practices. The on-going support from the coach also encourages the teacher to keep up with the increased pace of the lesson plans throughout the course of the year.

3.3.3. Intervention 2 specifically

The technology-supplemented intervention aims to provide new forms of support and guidance on teaching strategies through a range of materials, teaching guides, videos and interactive support platforms that are available at all times to the teacher. These resources are intended to encourage more productive teaching practices among the teachers. Lesson plans are provided in an electronic format on a tablet and are integrated with various audio and

visual resources to support teachers in the teaching of EFAL. The resources are supplemented with virtual coaching through phone calls and cell phone messaging. Similar to intervention 1, the on-going support from the virtual coach intends to encourage the teacher to keep up with the increased pace of the lesson plans throughout the course of the year. However, given that the virtual coach will not be in the classrooms, the role of the coach will focus more on encouragement and less on monitoring.

3.4. Evaluation

The study is designed as an RCT and consists of an implementation and evaluation side. Implementation service providers were appointed to ensure the effective implementation of the interventions. The evaluation component of the study is conducted independently of the implementation service providers to ensure that the impact of the interventions to minimize service provider bias. Data collection service providers will be recruited for each round of data collection and the Project Management Team is ultimately responsible for the instrument design and data analysis.

3.4.1. Overview of Evaluation Methods

The EGRS II is being evaluated using a mixed methods approach, whereby quantitative and qualitative methods are integrated to ensure both robustness and depth in evaluating the impact of the interventions. The evaluation analysis for the midline report therefore relies on three rounds of research: the case studies in September 2017, the lesson observations in September 2017 and the midline learner assessments in November 2017.

i) Quantitative estimation of impact:

The design is based on an RCT approach, where a sample of around 3600 learners in the 180 schools is tracked throughout the duration of the study. The baseline performance of the learners was taken at the start of Grade 1 (February 2017) and the follow-up testing on which this report is based was conducted at the end of Grade 1 (November 2017). Three more rounds of data collection are planned at the end of Grade 2, the end of Grade 3 and again at the end of Grade 4, allowing us to track the learners throughout the Foundation Phase and into Grade 4 where the language of learning and teaching switches to English. The RCT design will allow for the comparison between the average learner performance among the intervention and control groups to establish the relative effectiveness of the interventions respectively.

ii) Qualitative estimation of intermediate outcomes.

This approach combines evidence from the lesson observation study and the case studies, in order to understand the key channels through which the interventions influence teacher instructional practice. The information gained through these studies can also shed some light on any programme ineffectiveness, specifically in identifying weaknesses in programme implementation. The research questions that will be considered using this data will be in line with the research questions specified in the pre-analysis plan.

iii) Costing study

The costing study is intended to investigate the cost of the programme, as well as other resource requirements necessary in the scale-up of the interventions. The costing is largely based on spending data from the implementation of the programme.

3.4.2. Sample selection and intervention assignment

Through a process of elimination, we developed a sampling frame of 180 eligible schools. Beginning with 731 primary schools registered in the 2016 administrative data in the districts of Ehlanzeni and Gert Sibande, we firstly excluded relatively affluent schools (those in quintiles 4 and 5). Next, we excluded schools in which the language of instruction in the Foundation Phase was neither Siswati nor isiZulu. We also excluded schools which were missing in the 2014 Annual National Assessment (ANA) dataset.⁴ We further excluded particularly small schools (fewer than 30 Grade 1 enrolments) since many of these schools would practice multi-grade teaching rendering the grade-specific lesson plans less appropriate. We also excluded particularly large schools (more than 160 grade 1 enrolments, or more than three classes in Grade 1, or classes with more than 60 learners in) to limit intervention costs. After all of these exclusions, 193 eligible schools remained. Using a random number generator, we then excluded three further schools to remain with a sample of 190 schools. The 190 school sample included a sample of ten replacement schools (one in each of the strata) should the need arise to drop one of the sample schools. We thus obtained the sampling frame of 190 schools.

To increase power and assure balance between intervention arms, we performed stratified randomisation. We created ten strata of 19 similar schools based on school size, socio-economic status, and previous performance in the Annual National Assessments. Within each stratum, we then randomly assigned five schools to each intervention group, eight to the control group and one as a replacement school. Thus we randomly assigned 50 schools to each intervention and 80 to the control. Given that we aim to collect data on 20 Grade 1 learners per school, this sample should be sufficient to identify a minimum effect size of 0.21 standard deviations when comparing an intervention group with the control group and a minimum effect size of 0.23 standard deviations when comparing two intervention groups. These calculations assume a 95% confidence interval, an alpha value of 0.8, an intra-class correlation coefficient (ρ) of 0.3 and a correlation between pre- and post-test scores of 0.7.

Classroom Observation Study

The Classroom Observation Study was conducted with a sample of 60 schools (1 classroom from each school) from the larger 180 school sample. This comprises of 20 schools from the control, intervention 1 and intervention 2 group respectively. Sixty schools were considered to be the optimal number of schools in which it would be possible to conduct the fieldwork given

⁴ The Annual National Assessments have not been administered since 2014. This is therefore the most recent standardised measure of school performance we have for the full population of schools.

the budget constraints and cost of collecting such in-depth data. The collection of in-depth data also required very specialized fieldworkers which further drove up the cost. Having 20 schools in each of the three groups allowed for an analysis of trends emerging as a result of the interventions implemented. After data collection was completed, however, it was found that 21 Intervention 1 schools had been visited and 19 control schools had been visited. This mistake occurred because an intervention 1 and a control school shared the same name.

The following criteria were used to select the schools from each group:

- 5 top-performing urban schools (based on baseline results at the start of Grade 1);
- 5 additional urban schools;
- 5 schools in 25% of top performing rural schools (based on baseline results at the start of Grade 1); and
- 5 schools in the bottom performing 75% of rural schools

An urban/rural distinction was made since EGRS I had shown that the intervention had the largest effect in urban rather than rural schools. Since there were so few urban⁵ schools in the 180 school sample, for the most part, all urban schools sampled at the start of the year were visited for an EGRS II classroom observation. Top performing schools were over-sampled for this study because, according to the EGRS I findings, it is the top performing schools that are more likely to follow the EGRS II program. The Classroom Observation Study thus examines the changes in the instructional practice of teachers who are more likely to take up new practices.

Case studies

A sample of six schools was drawn for in-depth qualitative observations. The method employed aimed to find two average performing schools in each of the Intervention groups and Control group, one rural and one urban school, and to sample schools such that the schools in each pair (urban, rural) were close enough to each other to allow for school visits on consecutive days. Unfortunately, the researcher was unable to visit one of the sample Intervention 1 (I1) urban schools as planned due to service delivery unrest. A replacement I1 urban school therefore had to be found that was close enough to one of the other schools that were to be visited so that the researcher could cover two schools in one day.

Below are the basic descriptors of the six schools actually visited for the in-depth case studies (the names of the schools have been changed to protect the identity of participants):

1. Lerato Primary School: a rural Intervention 1 (I1) school near Piet Retief.
2. Thami Primary school: an urban Intervention 1 (I1) [replacement] school near Nelspruit.
3. Busi Primary School: a rural Intervention 2 (I2) school near Malelane.
4. Thabiso Primary School: an urban Intervention 2 (I2) school near Nelspruit.
5. Nowazi Primary School: a rural Control school near Nelspruit.
6. Siyabonga Primary School: an urban Control school near Malelane.

⁵ as classified by the principals in a questionnaire at baseline

3.4.3. Instrument Design

Learner Instrument

Both the baseline and midline learner assessment were adapted from the Early Grade Reading Assessment (EGRA). The tests were designed in collaboration with linguists and were based on the premise that certain basic literacy skills acquired in a first language are transferred to a second language. The tests therefore comprised of both home language (in this case isiZulu and siSwati) and English items. Given that the baseline data collection was to be conducted at the start of Grade 1, the baseline test was designed to test learners’ home language literacy skills at the start of Grade 1, as well as skills that are predictive of future literacy development. The midline test was designed to assess learners’ English literacy skills at the end of Grade 1, but also included home language items. The main purpose of both assessments was not to benchmark learner performance against curriculum requirements, but rather to determine learners’ literacy abilities at the start and end of Grade 1. To this end, it was important to ensure that the tests discriminated well among children at all levels of proficiency.

Both tests were designed to be orally administered by the fieldworkers and to be captured electronically on the Tangerine software. In order to test the targeted 20 learners within one school day, the tests were designed to take no longer than 15 minutes to administer. The baseline test included various sub-tasks including word recall, non-word recall, phonemic awareness, letter sound recognition, word recognition, listening comprehension, sentence reading and English vocabulary. The midline assessment included expressive vocabulary and word reading in both home language and English, as well as listening comprehension, letter sound recognition, phonological awareness and spelling in English.

The home language sections of the tests were originally designed in English and then versioned to SiSwati and isiZulu. In the versioning, specific care was taken to use words which are similar in the two languages, to minimise any bias that can be introduced through the language used in the assessment.

Table 2: Sub-tests contained in the instruments at each point in time

	Construct	Baseline – start of Gr 1		Midline – end of Gr 1	
		L1	English	L1	English
Language	Receptive Vocabulary		x		x
Comprehension	Expressive Vocabulary	x	x	x	x
	Listening Comprehension	x			x
Decoding	Phonological working memory	x			
	Phonological Awareness	x			x
	Letter-sound recognition	x			x
	Word reading fluency	x		x	x
	Sentence reading fluency	x			
Spelling	Spelling of a CVC ⁶ word				x

⁶ Consonant – vowel – consonant word such as “cat”

Contextual Questionnaires

During both the baseline and the midline data collection contextual questionnaires were administered to parents, the Grade 1 teachers and the school principal. The contextual questionnaires were designed by adapting the instruments which were administered in the EGRS in North West to the context in Mpumalanga. An additional focus on the frequency of using English both at home and in the classroom was included in both the Home Background Questionnaire and the Teacher Questionnaire.

Lesson observation study questionnaires

Three observation based instruments were developed for the lesson observation data collection to capture elements of teacher instructional practice. These instruments were developed for the South African context to capture information on the classroom level factors highlighted as affecting learner academic achievement in South African classrooms (Hoadley, 2016). Because of the focus on factors specific to the South African context, other classroom observation instruments such as Classroom Assessment Scoring System (CLASS) (Pianta, La Paro, & Hamre, 2008), Stallings classroom snapshot method (World Bank Group, 2015) and Individualising Student Instruction (ISI) (McDonald Connor, et al., 2009) were not used.

The list below outlines the sub-sections of each instrument designed for this study:

- *Lesson Observation Schedule* comprising 11 sections focusing on: teaching and learning environment; time on task and pacing; discipline; use of LTSM; language of instruction; listening and speaking; literacy and language; phonics and vocabulary development; opportunities to write; cognitive demand; and assessment.
- *Classroom Document Review Schedule* comprising 5 sections focusing on: a review of – the teacher’s classroom book collection; work schedule and work plan; records of assessment; school timetable; and learners’ workbooks.
- *Teacher Interview Schedule* comprising 8 sections focusing on: teacher and lesson details; class size; time on task; planning and curriculum coverage; availability and use of learning material; approaches to English language teaching; assessment; and teaching support received for EFAL.

3.4.4. Data Collection

Both the baseline and midline data collection were conducted by external service providers. Each round employed fifteen pairs of fieldworkers over a three week period. During baseline, a random sample of 20 learners was selected to be tested in each school, and the exact same learners were re-tested again at midline.

For each round of data collection, each school was visited by a pair of two fieldworkers with one fieldworker being responsible for the individual administration of the learner assessment, and the second fieldworker being responsible for the structured questionnaires. The fieldworker responsible for the learner assessment was also responsible for arranging the

completion of a structured contextual questionnaire (Home Background Questionnaire) by the parents or caregivers of all the sampled learners. The contextual questionnaires were taken home by the learners and fieldworkers arranged to retrieve them from the schools again via the learners' teachers within the three week fieldwork period. The second fieldworker was responsible for administering a structured questionnaire to all the Grade 1 teachers, as well as the school principal. Both these instruments also involved completing some school and classroom facility observations.

The learner assessments, teacher and principal questionnaires were administered using the Tangerine software. The home background questionnaire was sent home with the learners for a parent or guardian to fill out. These questionnaires were subsequently collected by the data collection company and captured and cleaned. The final data set, as well as .do files used to clean the data in the statistical software, Stata, was provided to the project management team at the conclusion of the data collection contract.

The quality of the fieldwork at baseline and the issues that were faced during the data collection are described in the Baseline Report. The main problem was caused by a teacher strike which made it more difficult for the fieldworkers to test 20 learners in one day. For this reason, the baseline sample does not consist of the expected 3,600 learners, but rather of 3,482 learners.

The midline data collection experienced fewer problems, with all fieldwork teams having been well received in schools. Some teams encountered minor problems at first, but none of which led to serious constraints. During the first week of data collections, various teams reported that some communities experienced strikes. These strikes were mainly local transportation and union strikes. However, the schools affected by the strikes were very understanding and willing to reschedule the visits. Apart from the strikes, absenteeism of learners and teachers was the biggest challenge experienced during fieldwork.

Table 3: Percentage of learners tested during baseline data collection

	Intended sample	Baseline		Midline	
		<i>Tested Number</i>	<i>Percentage Tested</i>	<i>Number re-tested & matched</i>	<i>Percentage re-tested & matched</i>
Control	1,600	1,459	94%	1,347	92%
Intervention 1	1,000	924	95%	820	89%
Intervention 2	1,000	944	98%	873	92%
Total	3,600	3,482	97%	3,040	91%

The teacher questionnaires were linked to the learner assessments and the principal questionnaires by means of a linking form that was filled out by the fieldworkers. Unfortunately, the linking forms were filled out very poorly, specifically with regards to the teacher names and surnames. This has meant that the linking of the teacher questionnaire to

the learner questionnaires has been challenging, and only 82% of the teacher interviews could be linked to the learner assessments during the midline.

Table 4: Percentage of teacher interviews conducted during midline data collection

	Number of teacher interviews	Number of teacher exercises	Number of learners matched to teachers	Percentage Questionnaires merged
Control	141	133	1204	83%
Intervention 1	90	90	747	81%
Intervention 2	88	88	786	83%
Total	319	315	285	82%

The principal interview asked general questions about the school to gain a better understanding of the context in which the Grade 1 teachers and learners function. The instrument was administered at each school, with a deputy-principal or another school management team (SMT) member standing in for the principal where necessary.

The final instrument administered was the home background questionnaire which learners took home. The purpose of the questionnaire was to collect information about the home circumstances from which the learner are from. 3,271 of the returned questionnaires were successfully matched with learners in the sample.

Table 5: Percentage of Home Background Questionnaires returned

	Number returned	Number matched	% Learner matched to PBQ
Control	1,459	1,422	97%
Intervention 1	924	914	99%
Intervention 2	944	935	99%
Total	3,327	3,271	98%

3.5. Ethical Clearance

The research methodology, with the intended instruments and research consent forms related to the baseline data collection, was formally submitted to the University of Witwatersrand Human Research Ethics Committee (Non-medical) in May 2016. The project was approved on 3 November 2016 and was granted ethical clearance for the duration of the study, up to 2 November 2019. During the baseline data collection, consent forms were provided to all principals and teachers interviewed. Consent forms were also sent with the home background questionnaires to parents.

3.6. Sample Characteristics

The sample population more or less mirror the larger school population in the province. The average age of learners in the study was 6.7 years at the end of Grade 1 and 28.6% of learners reported themselves as first language isiZulu speakers, with the other learners reporting

SiSwati as their home language. 56% of learners are from households where the responding parent did not complete any secondary schooling, and a further 23% only completed secondary schooling, but have not pursued any further education. The parents were also asked some questions on how regularly they speak English to their child and 60% of the parents in this sample reported that they sometimes speak English to their child. However, 27.5% reported that they never speak English to their child.

As with the country as a whole, the teachers in the study had an average age of close to 50 years, with over a quarter of all teachers being older than 55 years. All were female teachers. Given the ICT focus of intervention 2, all teachers were asked about the technological devices they have in their household, as well as how regularly they use certain technological features. It was interesting to note that 72% of teachers owned a smartphone and 54% owned a computer. However, only 44% of teachers made use of email, only 28% accessed educational resources on the internet and only 25% made use of Microsoft Word.

More than half the schools were classified as Quintile 1 schools (poorest), with 32.8% in Quintile 2 and only 13.5% Quintile 3. The majority of the principals in the survey described the location of the schools as remote rural, with a small number in small villages, townships or informal settlements in a city.

Overall the sample seems balanced on observable characteristics, with the only significant differences seen in the baseline score, class size and principal age. The virtual coaching group performed slightly better at the baseline and had smaller class sizes relative to the control group. The on-site coaching group, on the other hand, had a lower average principal age, but a slightly higher proportion of teachers who are older than 55 years.

Table 6: Sample Characteristics

		Control	On-site Coaching (I1)		Virtual Coaching (I2)		N	I1 vs I2
		Mean	Coeff	s.e.	Coeff	s.e.		P-value
Learner Characteristics	Learner = Boy	0.522	0.020	(0.019)	0.026	(0.022)	3,040	0.820
	Learner Age	6.658	-0.004	(0.048)	0.004	(0.046)	3,039	0.885
	Learner Language = isiZulu	0.296	-0.038	(0.052)	-0.023	(0.054)	3,327	0.809
	Baseline Score	-0.020	-0.016	(0.067)	0.113*	(0.068)	3,327	0.075
Teacher Characteristics	Teacher Age	49.58	0.815	(1.144)	-0.252	(1.090)	306	0.372
	Older than 55	0.241	0.108*	(0.064)	0.000	(0.060)	306	0.130
	Teacher Language = Zulu	0.299	-0.02	(0.059)	-0.021	(0.052)	306	0.984
	Teacher Language = SiSwati	0.672	0.025	(0.056)	0.019	(0.052)	306	0.918
	Multi-grade Classroom	0.044	0.025	(0.034)	0.002	(0.035)	306	0.542
	Class Size	42.75	-2.368	(1.531)	-4.007***	(1.476)	306	0.317
	At least a bachelor's degree	0.547	0.006	(0.077)	-0.019	(0.071)	306	0.779
	Teacher English Proficiency	3.821	-0.083	(0.284)	0.436	(0.289)	268	0.112
Principal and School Characteristics	Principal Age	52.55	-2.338**	(0.948)	-1.372	(0.944)	180	0.360
	Older than 55	0.375	-0.177**	(0.081)	-0.034	(0.080)	180	0.112
	Gr 1 teacher vacancies	0.063	-0.021	(0.042)	0.005	(0.042)	180	0.590
	Quintile 1 school	0.537	-0.014	(0.054)	0.001	(0.053)	180	0.801

3.7. Balance at Baseline

Furthermore, balance tests were conducted on each sub-task at baseline to ensure that the sample is balanced. Table 7 shows the results based on regression analysis to test for balance. These tests evaluate whether the differences in learning outcomes among the different intervention groups are statistically significantly different from zero. Each column in Table 7 is a separate regression run for each sub-task on intervention indicators, controlling for strata and district fixed effects. The significant differences in the table will be represented by stars. For instance, the one statistically significant result in Table 7 is the performance on the ‘naming the animals’ sub-task of the learners in the intervention two group, relative to the control group. The two stars indicate that this difference is significant at a 5% level. The final row in the table shows the p-value for the pair-wise test comparing the means between the two intervention groups (i.e. not the means of the intervention groups with the control group). A p-value of less than 0.05 would indicate that there is an imbalance between the two groups for the specific learning outcome. There are only two slight imbalances visible, but overall the samples are clearly balanced.

Table 7: Balance tests per sub-task

	Animals	Word Recall	Non-word Recall	Phoneme Isolation	Comprehension	Letter Sounds	Words Correct	Sentence Reading	Visual Perception	English Vocabulary
Paper-Based	0.15 (0.152)	-0.02 (0.118)	-0.02 (0.067)	-0.06 (0.12)	-0.02 (0.064)	-0.16 (0.484)	-0.05 (0.136)	-0.03* (0.016)	0.17 (0.121)	-0.04 (0.066)
Tablet-Based	0.39*** (0.149)	0.12 (0.118)	0.03 (0.094)	0.04 (0.131)	0.1 (0.063)	0.07 (0.48)	0.14 (0.165)	-0.02 (0.016)	0.19 (0.124)	0.01 (0.057)
Control Mean	7.155	9.981	4.208	1.129	2.179	4.652	0.387	0.051	1.46	0.836
N	3,327	3,327	3,327	3,327	3,327	3,327	3,327	3,327	3,327	3,327
Paper=Tablet: p-value	0.158	0.245	0.573	0.448	0.092	0.66	0.27	0.518	0.857	0.511

Note: Each column represents a separate regression on intervention dummies and stratification dummies. Standard errors are clustered at school level. * for $p < .1$; ** for $p < .05$; *** for $p < .01$

A short English proficiency task was administered to teachers. The main purpose of the task is to serve as a control in the learner regressions, but it also allows an additional balance check. Table 8 shows that the teachers in the various intervention groups are similar in their English proficiency. Although teacher English proficiency is not the focus of the interventions, it is likely that improved English proficiency might be the result of increased use of English during the lessons. The balance between the groups at the baseline means that any differences in teacher English proficiency in the wave 2 data collection, could be ascribed as a secondary outcome of the interventions.

Table 8: Balance test on Teacher English Proficiency

	Teacher Score
Paper-Based	-0.4 (0.256)
Tablet-Based	0.14 (0.267)
N	2963
Paper=Tablet: p-value	0.063

Note: Teacher score regressed on intervention and stratification dummies. Standard errors are clustered at school level.
* for p<.1; ** for p<.05; *** for p<.01

4. Implementation Fidelity

As a first step to checking the success of the interventions, it is useful to consider the quality of implementation. Successful implementation would entail teachers attending the teacher training sessions, as well as teachers using the methodologies and materials in their classrooms. Administrative data, as well as information from both the teacher questionnaire and the lesson observation study, are used to evaluate the fidelity of implementation.

4.1. Intervention 1 and 2

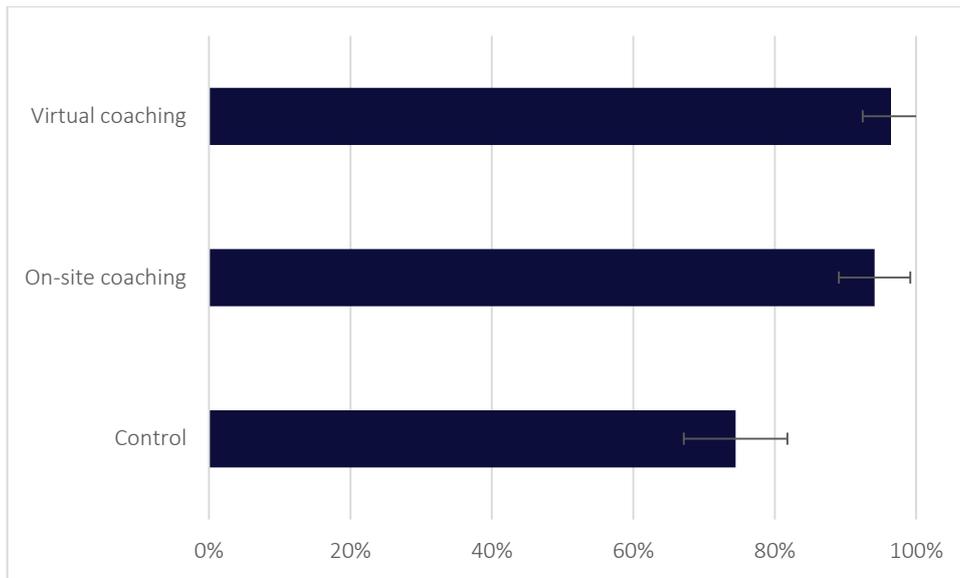
Table 9 presents the teacher attendance statistics at the training sessions for each term. Overall teacher attendance was high and remained high throughout the year. In term 3 the teacher training had a lower attendance rate, but there is no real difference between the attendance rates of the two intervention groups.

Table 9: Teacher attendance at training sessions

	Term 1		Term 2		Term 3		Term 4	
	No	%	No	%	No	%	No	%
<i>On-site coaching</i>	82	86%	90	95%	75	79%	90	95%
<i>Virtual coaching</i>	88	95%	91	98%	73	78%	89	96%

Various questions were included in the teacher questionnaire to gain a better understanding of the fidelity of implementation. The first question under consideration asked teachers whether they received any in-service training during the year. 94% and 96% of teachers in the intervention 1 and intervention 2 schools respectively responded that they received in-service training on the teaching of EFAL during the year. About 74% of teachers in the control schools also responded that they received in-service training in the teaching of EFAL during the year. This training was most likely conducted by either the province or the district, and it will be interesting to compare the effect of the regular training model to the EGRS training and coaching model.

Figure 3: Did you receive any in-service training in the teaching of EFAL?



4.2. Intervention 2 specifically

The virtual coaching intervention lends itself to further monitoring of implementation by considering teachers' usage of the tablets in the classroom. Two questions in the teacher questionnaire aim to gauge the usage of technology and tablets in the classroom. It is necessary, however, to keep in mind that these responses are self-reported, and teachers in intervention may have been more likely to respond favourably knowing that they were supposed to be using the tablets. To further investigate the actual usage of the tablets, the timestamp data from the tablets were analysed to determine the proportion of teachers that opened the lesson plans at least 3 times a week, and the proportion of teachers that covered at least 75% of lessons.

Table 10 reports the results from the questions included in the teacher questionnaire relating to using technology in the classroom. 92% of the teachers responded that they use technology in the classroom, and similarly, 92% of teachers responded that they make use of a tablet in the classroom. In the control and intervention 1 schools, a small proportion of teachers reported using a tablet in the classroom, and a slightly higher proportion indicated that they use technology. A radio was the most prevalent form of technology teachers in the control and intervention 1 schools reported as using in the classroom, followed by a smartphone.

Table 10: Using technology in the classroom

	% of teachers that use technology in class	% of teachers that use a tablet in class
Control	24% (0.037)	7% (0.022)
On-site coaching	19% (0.043)	5% (0.023)
Virtual coaching	92% (0.030)	92% (0.030)

Using the time-stamp data which is captured by the application, figures 4 and 5 provide further information regarding the time a teacher spent using the tablet. The first graph considers the average number of days per week that a teacher used the application for at least 15 minutes. The lesson plans scheduled 30 minutes of teaching time each day, however, a lower threshold⁷ was chosen to calculate the average number of times the teacher taught a lesson a week. It is necessary to note that the time-stamp data only provides us with information on when the teacher entered and exited a lesson. Using this information it is possible to derive when the teacher accessed the lesson, as well as the amount of time the teacher spent in a lesson. A day ‘taught’ was calculated by considering whether a teacher spent at least 15 minutes during school hours in the application, regardless of the number of lessons she accessed. Using this measure, it is evident that 70% of the teachers in the virtual coaching schools on average taught at least three lessons a week using the lesson plans on the application. About a quarter of teachers taught at least 4 days a week, and finally, 7% of teachers managed to teach every day of the week.

Figure 4: Average number of days taught, per week

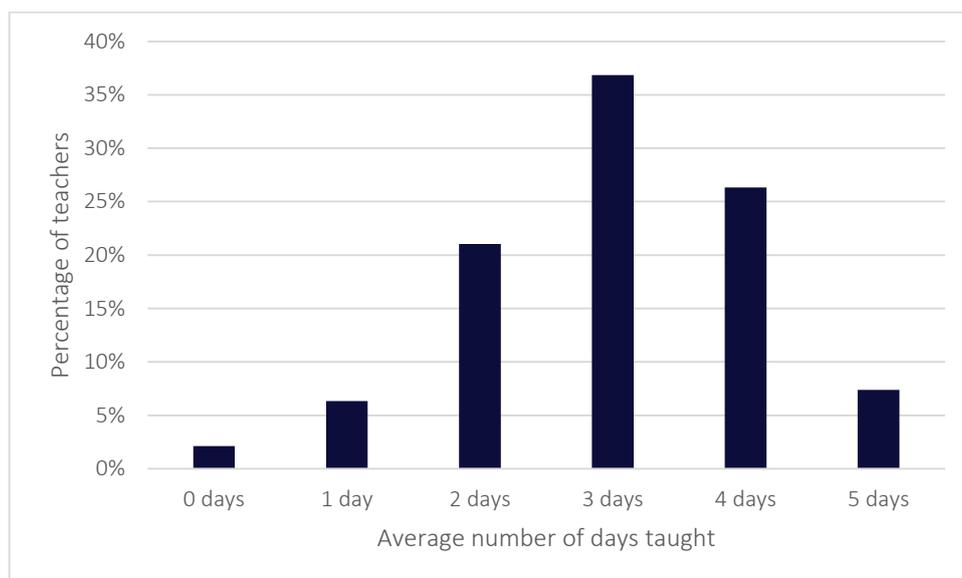
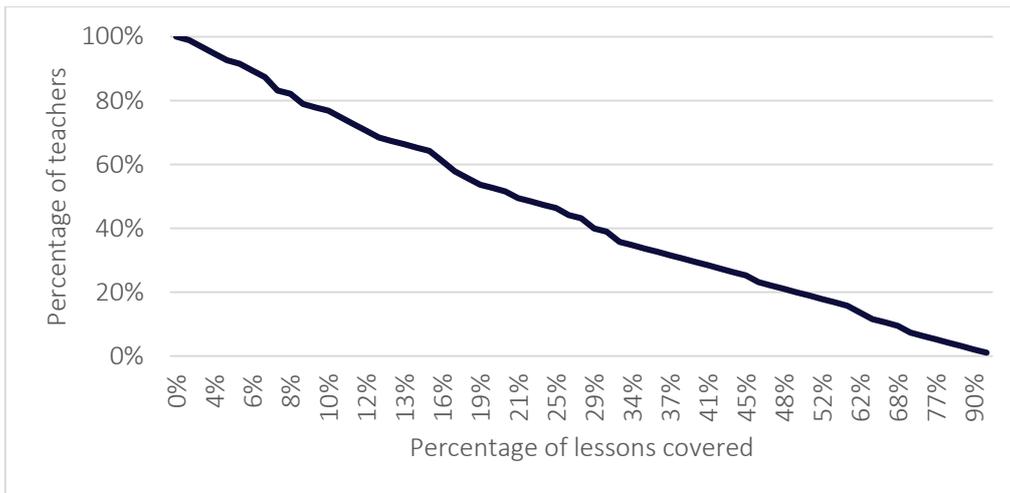


Figure 5 shows the lesson plan coverage for the teachers in the virtual coaching schools. A lesson was considered as covered if a teacher spent more than 10 minutes in total on a specific lesson on a specific day. Using this measure, figure 5 shows that only 20% of the teachers managed to cover 50% of the lessons, whereas only 10% of teachers managed to cover at two-thirds of the lessons. The information from figure 4 and figure 5 seem to suggest that the teachers managed to spend regular time using the application (and therefore most likely teaching English), but that they did not necessarily manage to cover the required number of lessons.

⁷ It is recognised that the school day is often disrupted in South African schools by various external factors. To give the teachers the benefit of the doubt, the threshold for “a lesson taught” was therefore set at 15 minutes.

Figure 5: Lesson plan coverage



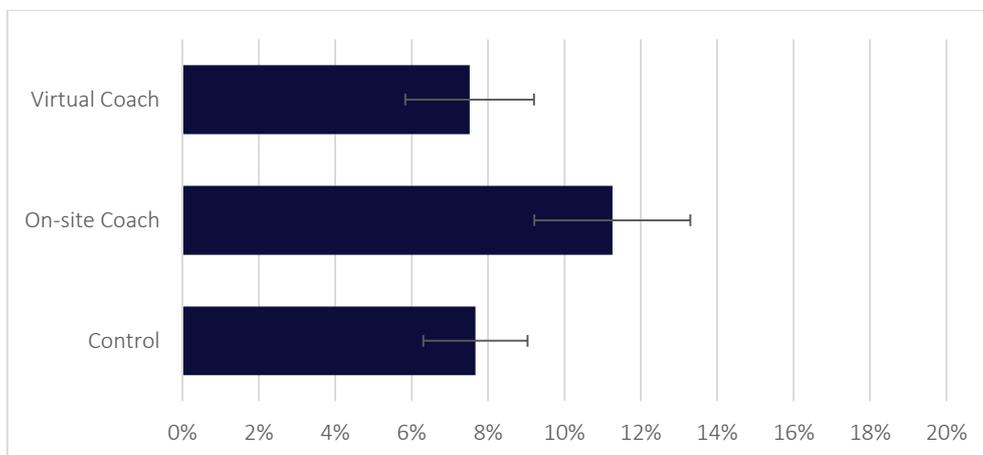
5. Midline Results

The sections to follow report on the data collected through the midline data collection. This will entail considering the attrition rate for each of the intervention and control groups, a detailed analysis of the learner performance on each of the sub-tests, as well as the main impact evaluation results.

5.1. Attrition

During the midline data collection, 3,040 of the 3,327 learners who were tested during the baseline data collection were re-tested and successfully matched to their baseline results. The overall attrition rate of 9% is in line with what other longitudinal studies in South Africa have found. When breaking down the attrition rate by intervention group, the differences are not statistically significant, but from figure 6 it is clear that the attrition rate of learners in the on-site coaching schools (11%) was slightly higher than the attrition rate among learners in the virtual coaching and control schools (each at about 8%).

Figure 6: Attrition rate by intervention group



The only learner or school characteristic that was significantly correlated to the likelihood of a learner having had attrited was the learners' baseline score, which suggests that poorer performing learners were more likely to not have been in the midline sample. There was no significant interaction effect between intervention assignment and the baseline score. Nevertheless, the final results will be evaluated using inverse probability weights as a final sensitivity check.

5.2. Item level learner assessment results

The midline learner assessment consisted of nine different tasks that assessed various home language and EFAL reading, writing and comprehension skills. Six of these assessment tasks were repeated from the baseline assessment to further allow the measurement of the learning that took place throughout the Grade 1 year.

Table 11 provides information on the descriptive statistics of both the baseline and the midline assessment tasks. The scores in the table have not been standardised and the averages should, therefore, be interpreted relative to the minimum and maximum value in the range column. For example, in the baseline expressive vocabulary task, the average performance was 7.2 items correct out of 12. Learners at the 10th percentile of the performance distribution on this item managed to score 4 items correct on average, whereas learners at the 90th percentile managed to score 10 out of the 12 correct. The table shows that in some of the sub-tasks there were floor effects (eg. baseline visual perception), and two items had a ceiling effect (eg. baseline non-word recall and midline expressive vocabulary). Knowledge of English vocabulary was low with only 25% of the sample correctly identifying at least one animal in English. Overall, both the assessments seem to provide enough information to discriminate among learners in various parts of the performance distribution.

In general, there is variability in learners' vocabulary levels, phonological working memory, and listening comprehension ability at the start of Grade 1. Learners have varying levels of letter-sound knowledge with half the learners naming 5 or fewer letters correctly in a minute, and 18% of learners not being able to identify a single letter correctly. Most learners were unable to identify the first phoneme of a two-syllable word in their home language, with learners only at the 75th percentile able to identify 2 of 6 phonemes correctly. Half the learners could not name or point to farm animals in English and even at the 90th percentile learners could identify only 2 of 6 animals.

By the end of Grade 1, most of these skills showed improvements, but there was still a large proportion of learners who scored zero on some of the tasks. With regards to word reading, learners at the 50th percentile could only read one home language word correctly; while this was zero below the 50th percentile, at the 90th percentile learners could read 17 words per minute correctly. Similar results were found for English word reading. Learners below the 50th percentile could not perform English phoneme identification tasks correctly by the end of Grade 1. Learners could recognise more letter-sounds correctly on average moving from a

mean of 6.98 at the start of the year to a mean of 17.62 by the end of the year. However, learners at the 10th percentile were still unable to identify a single letter correctly in the given time at the end of Grade 1. Overall, learner knowledge of English vocabulary is moderate with regards to the listening to instructions task and learners on average managed to correctly respond to about three of the five instructions given in English. The listening comprehension and the English expressive vocabulary tasks proved to be more difficult. Only at the 90th percentile did learners manage to correctly answer one of the listening comprehension questions and at the 75th percentile learners managed to give the correct English word for a picture. Finally, only learners at the 90th percentile were able to spell the English word *dog*.

Table 11: Item descriptive statistics

	N	Mean	s.e.	p10	p25	p50	p75	p90	Min.	Max.	% zero score	
Baseline	1. HL Naming the Animals	3327	7.3	0.41	4	6	7	9	10	0	12	1.3%
	2. HL Word Recall	3327	10.0	0.04	7	9	10	12	13	0	14	0.0%
	3. Nonword Recall	3327	4.2	0.02	3	4	5	5	5	0	5	1.5%
	4. Phoneme Isolation	3327	1.1	0.03	0	0	0	2	4	0	6	62.9%
	5. HL Story Comprehension	3327	2.2	0.02	1	1	2	3	4	0	4	8.7%
	6. HL Letter Sound Recognition	3327	6.9	0.13	0	2	5	9	18	0	30	18.7%
	7. HL Words Correct	3327	0.4	0.02	0	0	0	0	1	0	6	88.3%
	8. HL Sentence Words Correct	3327	0.0	0.01	0	0	0	0	0	0	6	98.5%
	9. HL Visual Perception	3327	1.6	0.03	0	0	1	3	4	0	10	35.9%
	10. EFAL Vocabulary	3327	0.8	0.02	0	0	0	1	2	0	6	50.4%
Midline	10. HL Expressive Vocabulary	3067	4.9	0.01	5	5	5	5	5	0	5	0.0%
	10. HL Words Correct	3066	5.0	0.12	0	0	1	11	17	0	18	48.3%
	11. EFAL Phoneme Isolation	3066	3.6	0.06	0	0	3	6	8	0	9	28.2%
	14. Letters Correct	3068	16.7	0.31	0	3	11	26	40	0	80	12.1%
	15. EFAL Words Correct	3062	5.1	0.13	0	0	2	7	16	0	36	35.6%
	16. EFAL Listening	3062	2.9	0.02	1	2	3	4	4	0	5	8.3%
	17. EFAL Listening Comprehension	3062	0.2	0.01	0	0	0	0	1	0	3	82.1%
	18. EFAL Vocabulary	3060	0.7	0.02	0	0	0	1	2	0	2	51.1%
	19. EFAL Writing	3056	0.2	0.01	0	0	0	0	1	0	1	76.5%

Table 12 shows the inter-item correlations between the sub-tasks assessed at baseline and at midline. Overall, there was a correlation of 0.4 between the baseline and midline assessment index scores. Letter sound recognition and phoneme awareness at baseline had the strongest correlation to the midline sub-tasks, followed by the English vocabulary sub-task at baseline. The home language word recognition sub-task had a very low correlation to the midline sub-tasks, but this is due to the severe floor effects on this task at the baseline.

Table 12: Item correlations between baseline and midline

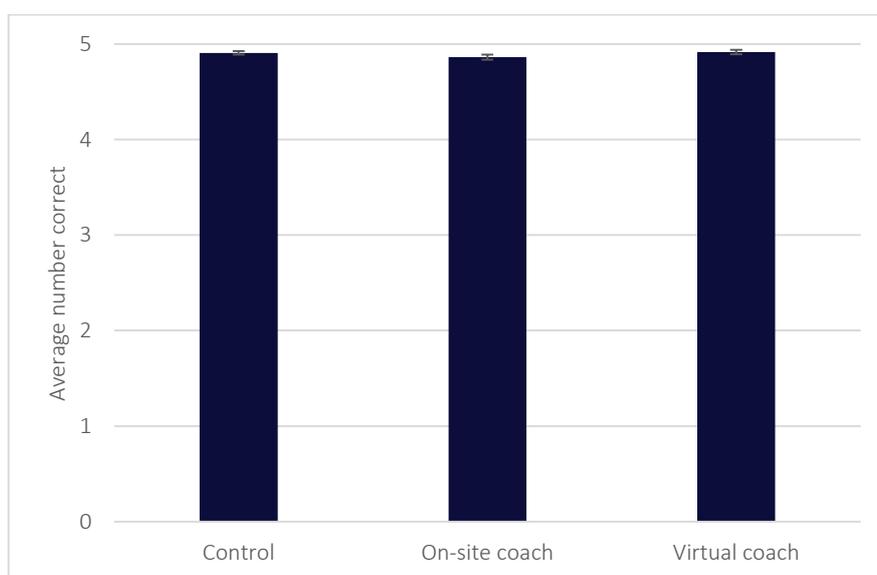
			Baseline							EFAL
			Home Language							
			Naming Animals	Word Recall	Non-word Recall	Phoneme Isolation	Story Compre- hension	Letter Sound Recognition	HL Words Correct	
Midline	Home Language	Expressive Vocabulary	0.09	0.08	0.06	0.05	0.07	0.06	0.04	0.02
		Words Correct	0.11	0.11	0.09	0.25	0.14	0.33	0.02	0.16
		Phoneme Isolation	0.12	0.15	0.08	0.27	0.19	0.23	0.00	0.19
	EFAL	Letters Correct	0.12	0.10	0.10	0.26	0.15	0.34	0.00	0.15
		English Words Correct	0.14	0.09	0.05	0.20	0.10	0.27	0.02	0.15
		Listening	0.16	0.12	0.09	0.19	0.21	0.20	0.01	0.28
		Listening Comprehension	0.09	0.07	0.08	0.09	0.12	0.12	-0.04	0.23
		Vocabulary	0.19	0.07	0.06	0.22	0.15	0.24	0.02	0.29
		Writing	0.11	0.08	0.06	0.21	0.10	0.28	0.03	0.20

5.2.1. Sub-task 1: Home Language expressive vocabulary

A short home language vocabulary task was developed for the midline assessment which required learners to identify the actions which were being performed by characters in pictures shown to the learners. The limitations of assessing vocabulary in this way are noted, but unfortunately, more comprehensive vocabulary tests such as the Peabody Picture Vocabulary Test and the Woodcock-Muñoz Language Survey are too time-consuming for the purposes of evaluating the EGRS II.

This task was intentionally included as a very easy task to enable some discrimination among the lowest performing learners. It is therefore to be expected that the learners performed very well overall, with the majority of learners having been able to correctly identify the actions. Given the ceiling effects, no significant differences were observed between the control and intervention groups.

Figure 7: Sub-task 1, HL vocabulary



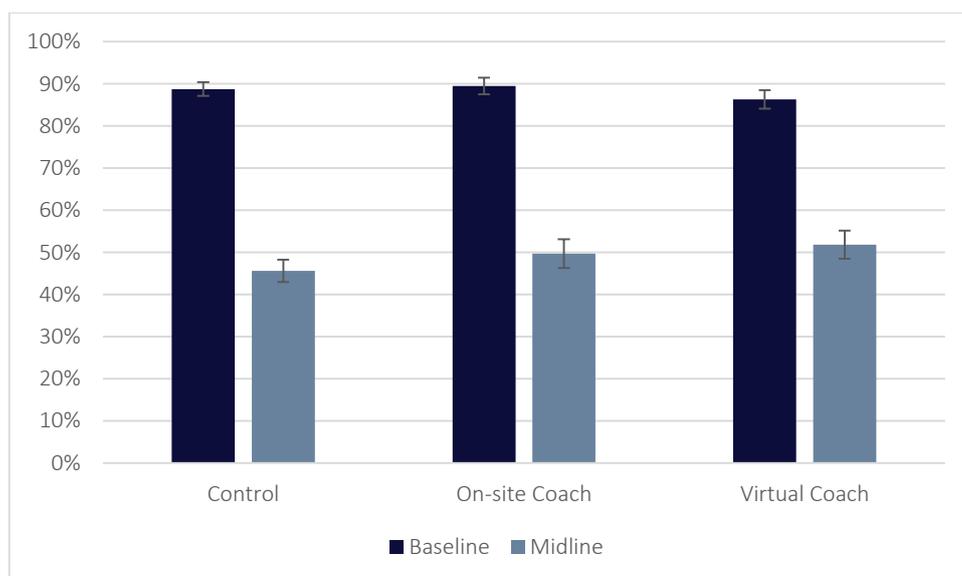
5.2.2. Sub-task 2: Home Language word recognition

At baseline, learners were asked to read an isiZulu or SiSwati word list (6 words) depending on their home language. The words were all two syllables long and similar in structure in both languages. Learners were required to complete a longer home language word reading task at midline (18 words), with the first six words to this list being the exactly the same words that were used at baseline. During both the baseline and midline administration of this sub-task, learners were timed for one minute per word list and errors were recorded and subtracted from the total number of words attempted.

Learners are not expected to be able to read words at the start of Grade 1 and therefore the severe floor effects on this item at baseline is not surprising. At the end of Grade 1, learners are however expected to be able to read basic decodable words. The midline results show that learners on average read five words correctly. This is slightly lower than expected, relative to the Setswana readers in EGRS I at the end of Grade 1 who on average read seven words correctly (Taylor et al. 2016).

Figure 8 below shows the percentage of learners who could not read a single word correctly at the start and end of Grade 1, disaggregated by intervention group. The percentage of learners who did not read any words correctly decreased by the end of Grade 1, but is still quite high with around half of the Grade 1 sample of learners still not being able to read a single word correctly in their home language by the end of Grade 1. The intervention disaggregation further shows that there are no significant differences between the control and intervention groups on when doing a descriptive mean comparison. However, the control group had a slightly lower proportion of learners who could not read any words correctly.

Figure 8: Percentage of learners who could not read a single word correctly

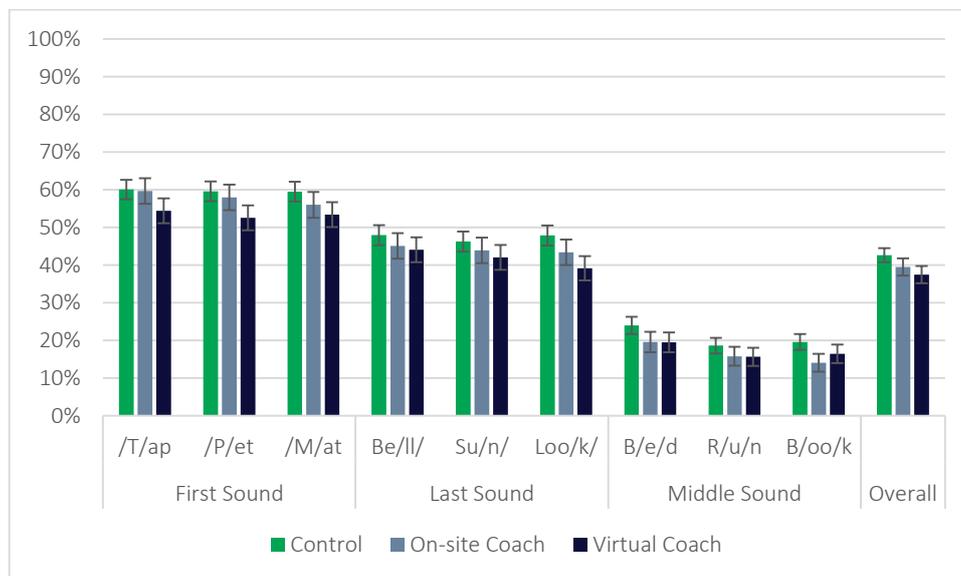


5.2.3. Sub-task 3: English phoneme isolation

A home language phonological awareness task was administered at the start of Grade 1. After two examples, learners were asked to identify the first phoneme of two syllable real words (for example, what is the first sound of *gogo* ~ *grandmother*?). These words were similar in structure and meaning in isiZulu and SiSwati. At the end of Grade 1 learners were administered phoneme identification tasks where they were asked to identify the first, last or middle sound of CVC English words (e.g. *cat*). Three words were presented per condition. In order to reduce the cognitive load learners were given bottle tops. For each sub-task, learners were presented with two examples, using the bottle tops to represent sounds at the start, end and middle of the word. Corrective feedback was provided in the examples, but not when the learner had to complete the non-example tasks.

At the start of Grade 1, learners identified 1.1 phoneme isolation tasks, whereas learners at the end of Grade 1 could identify 3.6 of the 9 phoneme isolation sounds at the end of Grade 1. It is concerning that learners below the 50th percentile could not perform any of the English phoneme identification tasks correctly by the end of Grade 1. A further concern is that the control group learners performed slightly better than both of the intervention groups, with the difference being statistically significant between the control and the virtual-coaching group.

Figure 9: Phoneme Isolation - Percentage correct by intervention group



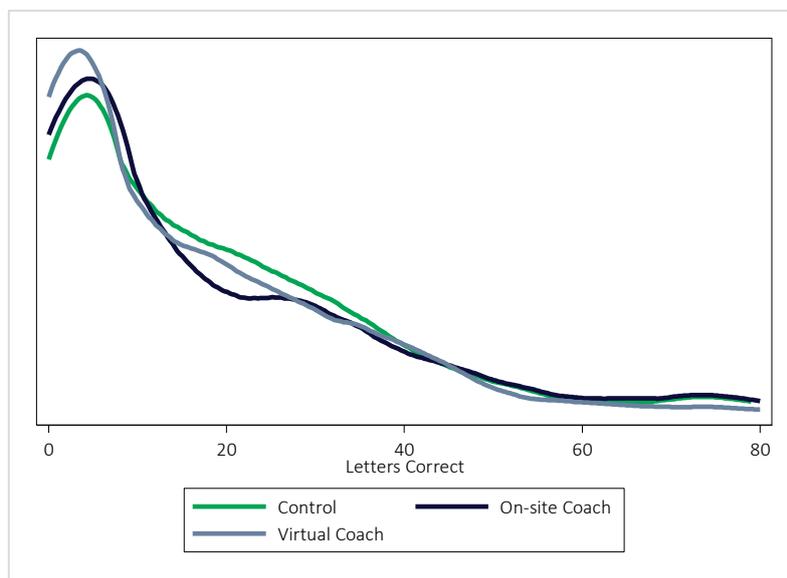
5.2.4. Sub-task 4: English letter recognition

A letter-sound knowledge task was administered to learners at both the start and end of Grade 1 to assess their automatic knowledge of letter-sound correspondences. At the start of Grade 1, learners were asked to provide the isiZulu or SiSwati (depending on their home language) letter sound for a maximum of 20 letters presented in a chart. At the end of Grade 1, learners were required to provide English letter-sounds for maximum 80 letters presented in a chart. Some English letters which are not pronounced the same as the home language (such as “c” pronounced [k] in English but as a voiceless dental click in isiZulu and SiSwati) were included,

but were arranged to appear later in the letter list. For the most part, the English and home language letter sounds are pronounced the same. Letters which sounded different in English compared to isiZulu and SiSwati were only included 17th, 23rd, 28th, 34th and 58th in the chart. Learners were timed for one minute and errors were recorded and subtracted from the total number of letters attempted. Learners' scores are reported as letters correct per minute.

Learners could recognise more letter-sounds correctly on average moving from a mean of 7 letter-sounds correct at the start of Grade 1 to 16.7 letter-sounds correct by the end of Grade 1. However, learners at the 10th percentile were still unable to identify a letter correctly at the end of Grade 1. Figure 10 shows the distribution of the number of letter sounds read correctly by the intervention group. Once again it is evident that the control group learners performed marginally better than both the intervention group learners in this task.

Figure 10: Distribution of letters read correctly by intervention group



Comparing learner performance on the letter recognition between the North-West and Mpumalanga samples, it is evident that the Mpumalanga sample of learners performed better at the start of Grade 1, but that they were performing worse than the North-West sample by the end of Grade 1. This is the case for learner performance in both the control and intervention schools, which could therefore reflect the instructional practices in each province or the effect of learning the more transparent Setswana language. With regards to the effect of on-site coaching between the two samples, we notice that there was a statistically significant positive difference between the learner performance in the on-site coaching schools and the control schools in North-West, but that this was not the case in Mpumalanga.

Table 13: Comparing letter-sound recognition between Mpumalanga and North-West

		Start of Grade 1	(s.e.)	End of Grade 1	(s.e.)
EGRS I	<i>Control</i>	5.40	(0.28)	22.70	(0.60)
	<i>On-site coaching</i>	5.80	(0.29)	25.10	(0.72)
EGRS II	<i>Control</i>	6.98	(0.01)	17.67	(0.19)
	<i>On-site coaching</i>	6.78	(0.01)	16.72	(0.25)
	<i>Virtual coaching</i>	7.02	(0.01)	15.06	(0.25)

5.2.5. Sub-task 5: English word recognition

Learners completed two English word reading lists timed for 30 seconds each. One list contained CVC regularly spelled nouns and one list contained high-frequency words from the Dolch sight word list (Dolch 1936). The number of words read correctly in each English word reading task were added to give a words correct per minute score.

On average learners could read 4.9 of the decodable words and 5.1 of the sight words per minute correctly, with there being no statistically significant difference between the two types of words. No statistically significant difference was observed between the control and intervention groups.

Interestingly, comparing these results to sub-task 2 (home language word recognition), it is evident that there is also no statistically significant difference between learners' ability to read home language and English words. That home language and English word reading fluency are developing almost in parallel rather than home language reading far surpassing English reading could indicate that learners have not automated their letter-sound correspondence knowledge needed to automatically decode the very transparent isiZulu and SiSwati words. Measures of text reading fluency were not included. Nevertheless, with 50% of learners reading less than two words in a minute, it can be expected that learners would be very slow text readers, and would likely not be able to understand what they read.

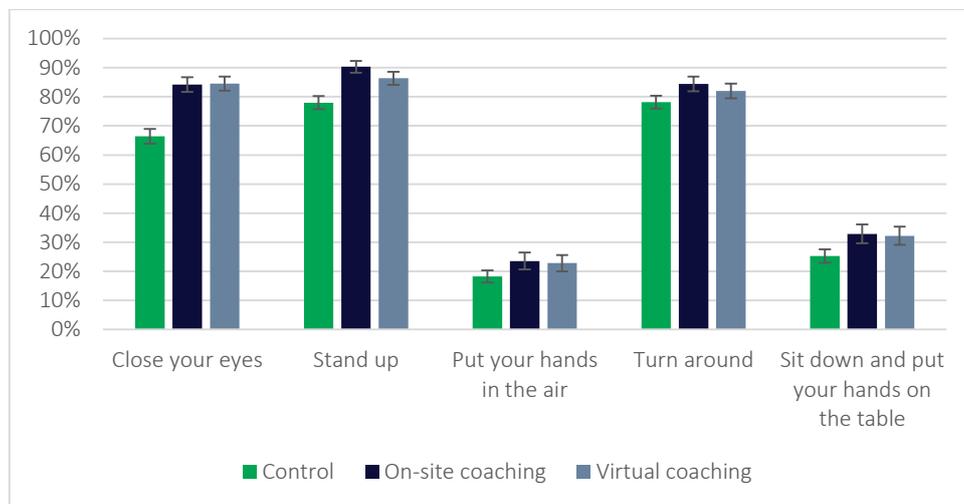
5.2.6. Sub-task 6: English Listening

At midline, learners' knowledge of basic English instructions was assessed using a custom-made instructions task. This task required learners to perform a number of English instructions such as "please stand up". This task has not been used in EGRAs before but was specifically designed to evaluate basic English comprehension.

Figure 12 shows the percentage of learners in each of the control and intervention groups which managed to respond correctly to the English instructions that were asked of them. Overall, it is clear that there was a significant difference between the control and intervention groups, but that there were no clear differences between the two intervention groups. On average, learners in the control group managed to respond correctly to 2.6 out of the five

English instructions, whereas learners in the intervention groups responded correctly to three out of the five instructions.

Figure 11: Percentage of learners correctly responding to each English instruction



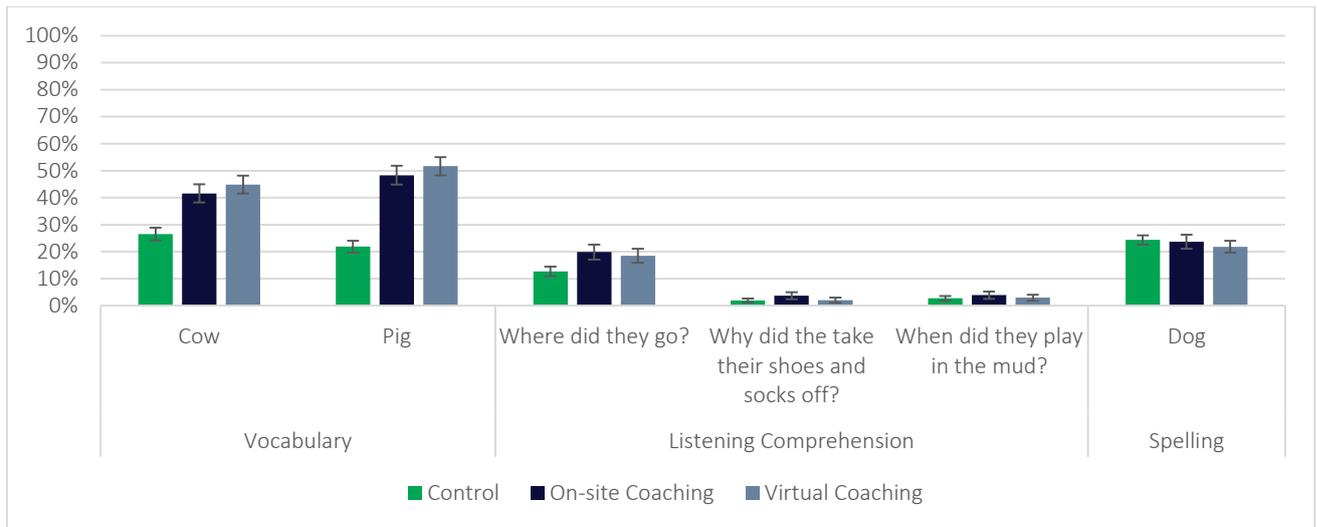
5.2.7. Sub-task 7 – 9: English Listening Comprehension, Vocabulary and Spelling

A further three small sub-tasks were included in the midline assessment to evaluate various aspects of learners’ English oral and writing proficiency. These sub-tasks included English listening comprehension, English expressive vocabulary and English spelling. For all three tasks, instructions and examples were given in the learners’ home language to ensure that we can be confident that the learner understood the ‘rules of the game’.

A short English vocabulary task which assessed farm animal knowledge was developed for the baseline and midline assessment. To assess learners’ English receptive vocabulary, learners were shown a picture of farm animals and asked: “what do we call this animal in English?” (2 items). Furthermore, an English listening comprehension task was administered using the same story that was used during the home language listening comprehension that was administered at the start of Grade 1. The enumerator read a short paragraph twice with expression about a girl playing in the rain. Learners were then asked three inferential questions about the story. At the end of Grade 1, an English version of the same paragraph was used for the English listening comprehension. Finally, learners were shown a picture of a *dog*, told it was a *dog* in English and were asked to write the word *dog* on a piece of paper. Learners were given a maximum of two minutes to complete this task.

Figure 12 shows the learner performance on each of these tasks, disaggregated by intervention group. Learners performed moderately on the English receptive vocabulary tasks, with learners in the two intervention groups performing significantly better than the learners in the control schools. The average performance on the listening comprehension and the spelling task was very poor, with learners in the intervention groups only doing marginally better than the control group learners on the first listening comprehension question.

Figure 12: Percentage correct on English Listening Comprehension and Vocabulary



5.3. Main Results

The first major findings of the study relate to the relative performance of the three groups on the midline learner assessment. A combined test score was derived using Principal Component Analysis to do a simple mean comparison of the average learner performance. The index score was standardised relative to the control group, with the control group having a mean of zero and a standard deviation of 1. Table 13 shows the mean performance of each of the control and intervention groups on both the baseline and midline learner assessment. From a descriptive perspective, there is no significant difference between the control and intervention groups. Although not statistically significant, the virtual coaching group had a slight advantage over the other two groups at baseline, but that this advantage was not evident at midline.

Figure 13: Mean comparison of Index Scores

	Baseline	(s.e.)	Midline	(s.e.)
Control	0.00	(0.03)	0.00	(0.03)
On-site coaching	-0.03	(0.03)	-0.02	(0.04)
Virtual coaching	0.10	(0.03)	-0.07	(0.03)

Given the nature of learning a second language, the ratio of English to home language items increases with each wave of data collection. We therefore control separately for each domain of reading proficiency: vocabulary, letter recognition, phonological awareness, word recognition and English oral proficiency. Table 14 below shows the mean scores for each of the intervention groups on each of the reading domains, as well as whether the mean scores are statistically significant between the control and intervention groups. This significance is indicated by the three stars next to the mean score.

This table suggests that at a descriptive level, both the intervention 1 and intervention 2 learners performed slightly worse than the control group learners on the reading of home

language words. It further suggests that the intervention 2 learners also performed slightly worse than the control group learners on the phonemic awareness task and the letter recognition tasks. Home language word reading, phonemic awareness and letter recognition are all skills which are developed mainly through the teaching of home language. The slightly worse performance may be indicative of potential crowding-out effects of the intervention on the teaching of home language. This will be explored further in the analysis.

Table 14: Sub-task mean comparison scores, by intervention group.

	Control N=1459		Intervention 1 N=924		Intervention 2 N=944	
	Mean	(s.e.)	Mean	(s.e.)	Mean	(s.e.)
HL Expressive Vocab (/5)	4.9	(0.01)	4.9	(0.01)	4.9	(0.01)
HL: Word Recognition (/18)	5.6	(0.19)	4.6 ***	(0.23)	4.7 ***	(0.22)
EFAL: Phonemic Awareness (/9)	3.8	(0.09)	3.6	(0.11)	3.4 ***	(0.11)
EFAL: Letters Recognition (/80)	17.6	(0.46)	16.7	(0.64)	15.1 ***	(0.53)
EFAL: Word Recognition (/36)	5.2	(0.20)	5.2	(0.27)	4.6	(0.24)
EFAL: Listening (/5)	2.7	(0.04)	3.2 ***	(0.04)	3.1 ***	(0.04)
EFAL Comprehension (/3)	0.2	(0.01)	0.3 ***	(0.02)	0.2	(0.02)
EFAL Vocabulary (/2)	0.5	(0.02)	0.9 ***	(0.03)	1.0 ***	(0.03)
EFAL Writing (/1)	0.2	(0.01)	0.2	(0.01)	0.2	(0.01)
English Oral Proficiency	3.1	(0.06)	3.8 ***	(0.08)	4.0 ***	(0.07)

Notes: *** indicates a significant difference between the mean score of the control and specific intervention group

To take into account the stratification variables, separate regressions were run on each subtest. For the sake of comparability, the scores were standardised to a mean of zero and a standard deviation of 1. The skills in columns 1 and 2 in Table 15 relate to home language proficiency and decoding skills, the skills in columns 3, 4 and 5 relate to English decoding skills, the skills in columns 6, 7 and 8 indicate English oral language proficiency skills, and, finally, column 9 refers to English spelling.

The coefficients for both intervention groups are negative for home language word recognition, English phonemic awareness and English letter recognition, but is only statistically significant for the virtual coaching group. Given that the students would have been taught decoding skills in their home language, and that the skill of decoding is considered transferable between the languages, this further confirms the possibility of a crowding-out effect to the teaching of the home language. In future rounds of data collection, this is something which will be evaluated rigorously. If the problem persists, it will warrant stronger claims of the consequences of the crowding-out.

Nevertheless, we see from columns 6, 7 and 8 that both interventions seem to have had an equally large positive impact on the various English language proficiency skills. The Curriculum and Assessment Policy Statement (CAPS) specifies that a maximum of three hours per week be given to the teaching of English (as the additional language) in Grade 1, with a maximum of one and a half hours for listening and speaking, one and a quarter hours for reading and

phonics, and 15 minutes for writing. The scope of learning in the curriculum for Grade 1 English learning is oral language development, with reading and phonics being addressed through methods such as shared reading, listening to stories, and total physical response. All three of these practices essentially focus on receptive language proficiency, and to a lesser extent on expressive language development practices.

For the three sub-tasks that focus on English oral language development – English listening (following instructions with actions), English listening comprehension, and English vocabulary – the students in the two coaching interventions did significantly better than their counterparts in the control schools. On the higher-order skill sub-tasks that assessed students’ comprehension in English, students in the intervention classrooms did only marginally better than students in the control schools, which suggests that while vocabulary development is stronger as a result of the interventions, this has not yet translated into stronger comprehension skills. That said, if oral vocabulary development in the second language is a developmental building block for reading acquisition, then the interventions may be working to improve the basic skills targeted in the Grade 1 curriculum, but they are not yet impacting the higher-order skills.

Table 15: Main regression

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	HL Vocabulary	HL Word Recognition	Phonemic Awareness	ESL Letter Sounds	ESL Word Recognition	ESL Listening	ESL Compre- hension	ESL Vocab	ESL Writing	English Oral Proficiency
On-site coach	-0.129** (0.054)	-0.133 (0.085)	-0.072 (0.091)	-0.049 (0.087)	0.030 (0.079)	0.365*** (0.082)	0.204*** (0.075)	0.526*** (0.071)	-0.011 (0.075)	0.529*** (0.083)
Virtual coach	0.014 (0.056)	-0.162* (0.083)	-0.171* (0.094)	-0.180** (0.077)	-0.104 (0.069)	0.288*** (0.083)	0.105 (0.065)	0.547*** (0.073)	-0.083 (0.067)	0.456*** (0.084)
Obs	3,067	3,066	3,066	3,068	3,062	3,062	3,062	3,060	3,056	3,060
R²	0.023	0.140	0.130	0.138	0.106	0.148	0.059	0.205	0.113	0.218
P-value	0.0191	0.742	0.328	0.122	0.100	0.383	0.236	0.780	0.303	0.434

Note: Each outcome variable is standardised to a mean of zero and a standard deviation of one. Learner gender, baseline score, learner age, learner home language, district and stratification dummies controlled for. Standard errors are clustered at school level. * for p<.1; ** for p<.05; *** for p<.01

5.4. Intervention effects on sub-groups of interest

Understanding the differential effect of the interventions on various sub-groups is important from a policy-making perspective, although we recognise the risks of the multiple comparison problem. This problem stems from an increased probability of finding at least one statistically significant result, related to the increased number of sub-groups analysed. Nevertheless, we are also aware that it is necessary to have a comprehensive understanding of the effect of an intervention on a different group of learners and schools. To preserve the integrity of the research study, we pre-specified the theory underlying our decision to investigate various sub-

groups in our pre-analysis plan.⁸ Furthermore, we are also cautious in interpreting any heterogeneous effects at this point in time, however, we will be more confident in our interpretation should these effects persist in future waves of data collection.

Another choice which straddles the fine balance between understanding the dynamics of the interventions, while heeding against the multiple comparisons problem, relates to the outcome indicator which will be used to evaluate the heterogeneous effects. Given that the interventions targeted the teaching of EFAL, an argument can be made to only consider the EFAL items. A similar argument can be made to only include the EFAL oral language proficiency items, since the EFAL Grade 1 curriculum solely only focusses on oral language proficiency. Since the main results suggested that the interventions have impacted learners' English oral language proficiency, an index including the three sub-tasks will be used at the outcome variable in testing for heterogeneous effects. Caution therefore needs to be taken not to over-interpret these results and the results will in no way be regarded as definitive. The sole purpose of these heterogeneous effects is to inform further research questionnaire design, as well as focus qualitative research efforts to further understand the full impact of the interventions.

5.4.1. Learner level

At the learner level, we expect two opposing heterogeneous intervention impacts based on midline learner reading proficiency. The scripted lesson plans require streaming by ability *within* the same classroom which provides opportunities for individualized attention. It is therefore expected that the interventions will benefit children who have otherwise been left behind. However, at the same time, the scripted lesson plans are aligned to the national curriculum, which prescribes an ambitious pace in the South African context. The worst-performing learners might actually benefit less if the teachers who follow the scripted lesson plans now progress at too fast a pace. Furthermore, boys/girls might benefit more/less from the individualized attention. Finally, the emphasis on individualized attention and tracking means that learners might benefit more from the scripted lesson plans when the class size is large, and the worse-performing learners, in particular, will benefit more.

Given these expectations, heterogeneous effects were first investigated for learner gender and learner age (see whether under/ overage learners benefited more/ less). No heterogeneous effects were evident for learner gender or learner age, which suggest that none of these groups was differentially impacted by the interventions.

The next consideration was whether the interventions impacted stronger or weaker learners differently. Although the interventions are designed to specifically support the non-fee paying schools, it is recognized that there is still large variance in learner performance in these schools. If the impact of the interventions is greater among the weaker learners, these interventions

⁸ The pre-analysis plan was registered on the American Economic Associations RCT registry on 26 September 2017.

can be considered as equity-enhancing. No differential effect based on baseline performance was found for the virtual coaching intervention, but the on-site coaching intervention seems to have benefitted learners who had a higher baseline performance score marginally more (Table 16). As a sensitivity check, these models were also run by including variables based on parents reporting their own English proficiency, as well as the frequency of speaking English at home. The inclusion of these variables did not have a remarkable effect on the coefficients of the interaction variables.

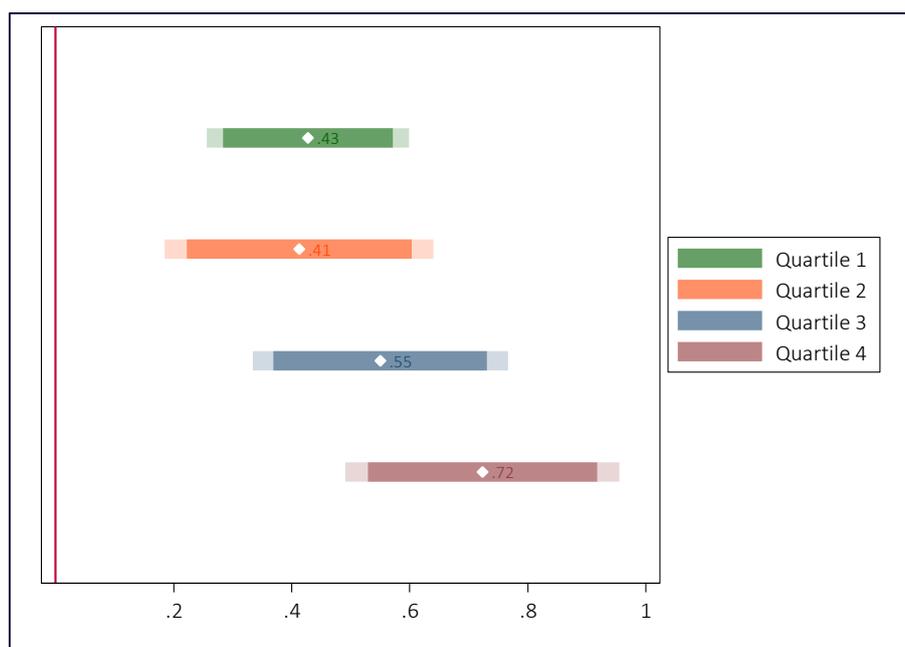
Table 16: Heterogeneous effects - Baseline performance

	(1) English Oral Proficiency	(2) English Oral Proficiency
On-site coach	0.529*** (0.083)	0.531*** (0.083)
Virtual coach	0.456*** (0.084)	0.454*** (0.083)
Baseline score	0.390*** (0.022)	0.325*** (0.030)
BL score x T1		0.151*** (0.051)
BL score x T2		0.080 (0.049)
Observations	3,060	3,060
R-squared	0.218	0.221
P-value	0.434	0.414

Note: Learner gender, learner age, learner home language, district and stratification dummies controlled for. Standard errors are clustered at school level. * for $p < .1$; ** for $p < .05$; *** for $p < .01$

Figure 14 assists us to better understand which part of the baseline performance distribution benefitted the most from the on-site coaching intervention. To construct the graph, the sample was divided into four sub-samples, based on the performance distribution of baseline scores. Separate regressions were run on each of the four sub-samples and figure 14 shows the intervention coefficient in each of the regressions run. This figure firstly shows that learners from all parts of the performance distribution did benefit from the on-site coaching distribution, but that the top 50% of learners benefitted marginally more than the bottom 50% of the performance distribution.

Figure 14: Differential impact based on Baseline performance



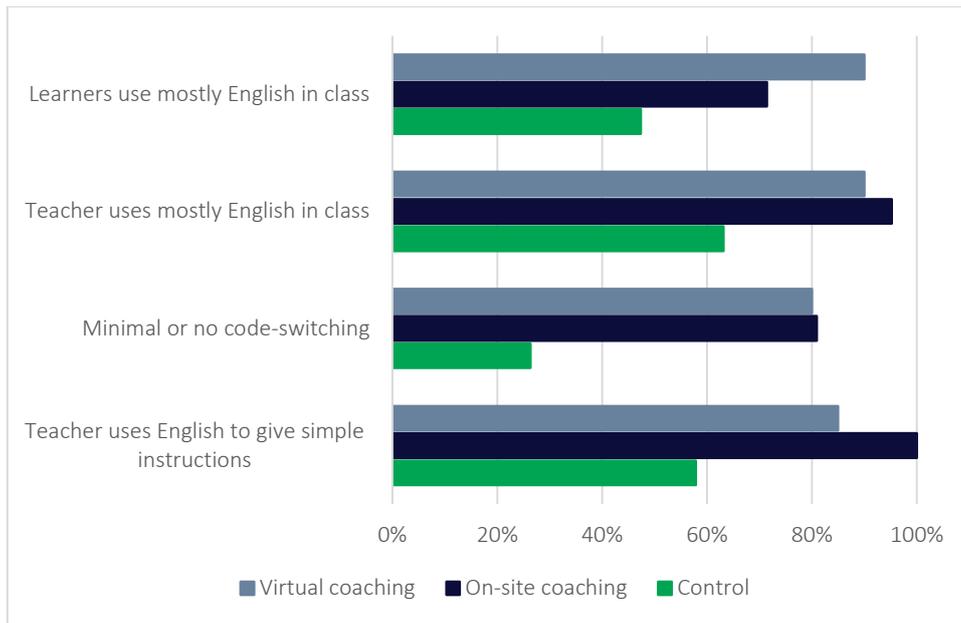
6. Changes in instructional practices

As discussed earlier, a classroom observation study was conducted on a sub-sample of 60 schools participating in the EGRS II study to provide a more qualitative perspective on the intermediate outcomes of the two interventions. In addition to the lesson observations, an education specialist did six case studies in a further sub-sample of schools, to provide an educationalist's perspective of the difference between the schools in the intervention and the comparison schools. The research design, process and results are reported in more comprehensive reports separate to this report. We will, however, shortly discuss the main findings of these studies here.

To ensure integrity in the reporting of the classroom observation results, differences in instructional practices are only reported if observed in both the lesson observation study and the case studies. Three main differences were observed between the intervention and the control classrooms, namely the frequency of English usage by both the students and the teachers, systematic teaching of oral language proficiency in English using extended texts and more efficient use of time, which allows for greater student exposure to the English curriculum.

Given the South African context of teachers who are not always confident English speakers being responsible for teaching students English, one potentially important achievement was the increased use of English during the teaching of the English lesson. The observations in the control classrooms attest to the English lesson being taught in the home language, with limited use of English during these lessons. It is evident from figure 14 that learners and teachers in the intervention classrooms were more likely to use English during the English lesson and the teachers were also less likely to make use of code-switching.

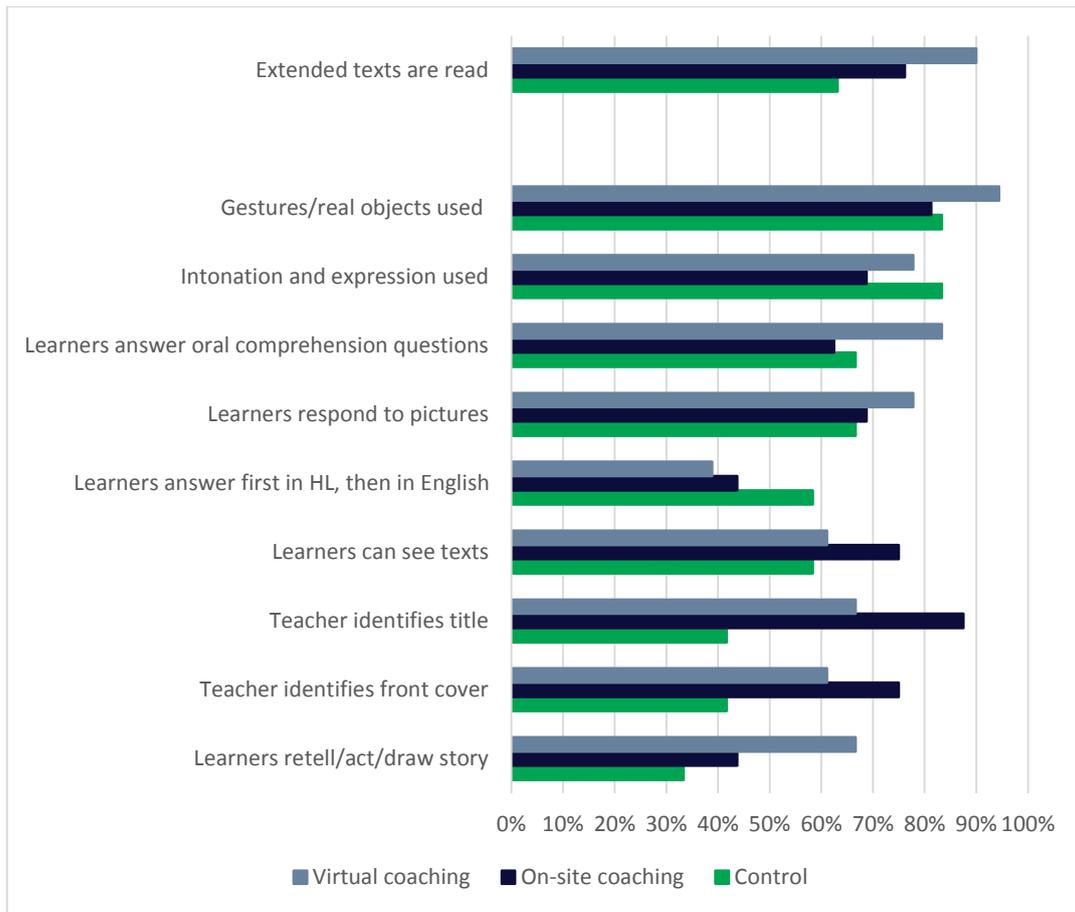
Figure 15: Frequency of English used in the classroom



Shared reading forms a substantial part of the Grade 1 EFAL curriculum and is instrumental in teaching learners English oral proficiency. A wide array of instructional practices should be employed in the shared reading lessons to enable the more effective learning of new vocabulary (see Figure 15). Although teachers in the control schools, on average, introduced more vocabulary words on a given day, these words were often only repeated, without any emphasis on the meaning. By contrast, the intervention schools teachers used the new vocabulary words in the context of a sentence, so as to promote understanding. Differences were also seen in the shared reading practices, with 90% of the intervention 2 teachers and 76% of the intervention 1 teachers having used extended texts during their lessons. Of the teachers that made use of extended texts, fewer intervention 1 teachers were observed using chorsing. This finding suggests that intervention 1 teachers may have made substantial progress in the use of the methodology.

In addition to reading aloud, shared reading should involve teachers asking students questions about the story, so as to build students' English oral language proficiency and increase their receptive English vocabulary. Intervention 2 teachers were more likely to ask their students comprehension questions based on the story as compared to either intervention 1 or control teachers. Further differences in groups arose in getting students to retell, act out or draw in response to the story, where more intervention 2 teachers (67%) than control (33%) or intervention 1 (44%) teachers got students involved in retelling, acting out or drawing responses to stories. The interventions therefore seem to be expanding the teachers' pedagogic repertoires, through systematic teaching of English vocabulary and English language proficiency. The greater exposure to vocabulary development and language development through the shared reading of extended texts should therefore lead to students having a stronger command of the English language.

Figure 16: Instructional practices in shared reading

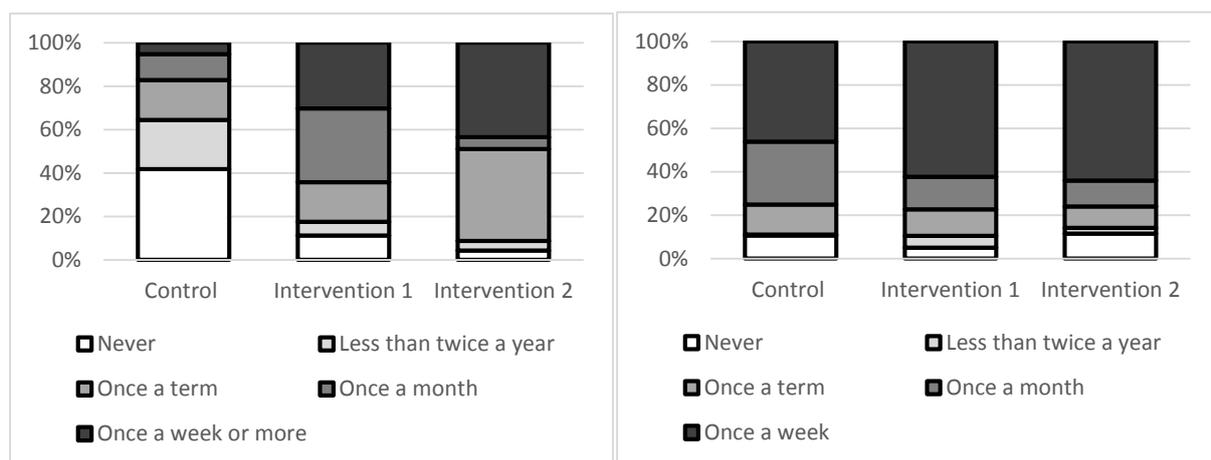


It is well recognised that Foundation Phase teachers lack ongoing instructional support in South Africa. The 2011 School Monitoring Survey indicated that Foundation Phase teachers received less monitoring and support from the school principal, the school management team, and the governance structures than their secondary school counterparts (Wills, 2016). Only 45% of the 4,128 teachers interviewed in the survey indicated that they had been visited during the year by a district official responsible for curriculum advice (Wills, 2016). In the districts that are participating in the EGRS II study, the curriculum advisor-to-school ratio is around 130 schools per advisor. This means that teachers, on average, receive a visit from the curriculum advisors once every two years. Heads of department (HODs) in schools are further responsible for providing instructional support to teachers. However, a recent evaluation of implementation of the curriculum found that the fact that HODs have full teaching loads, and the fact that inappropriate promotion practices are followed when appointing HODs, often means that HODs do not have the capacity or the ability to provide effective support to teachers (Department of Planning, Monitoring and Evaluation/Department of Basic Education, 2017).

The presence of the individualised support provided through both the on-site coaches and the ‘virtual’ coach in the EGRS II study therefore fills a critical gap in providing instructional support. The coaches promote the development of professional accountability in an environment of trust, where the coach monitors and evaluates the teachers’ teaching practices in order to

encourage more productive teaching practices. The coach-to-school ratio for the on-site coaches is designed to allow each teacher to be visited by a coach at least three times a term. Moreover, by removing the barrier of needing to be physically present in the classroom, the ‘virtual’ coach communicates with teachers on a weekly basis. Fig. 6 suggests that teachers in both intervention groups are much more likely than the control group to have contact with a coach or mentor regarding their English teaching practices, with 42% of control group teachers indicating that they never meet with a coach or mentor regarding their English teaching. When asked how often teachers meet with their heads of department (HODs) regarding their English teaching, teachers from both intervention groups reported meeting with their HODs more regularly. Sixty-six percent and 63% of intervention 1 teachers and intervention 2 teachers, respectively, reported meeting with the HODs at least once a week, while only 50% of control group teachers reported meeting their HOD weekly. This indicates that coaching is not only providing more instructional support to teachers directly, through their presence, but also indirectly, through encouraging the HODs to support teachers more regularly.

Figure 17: Instructional support.



7. Sensitivity Checks

7.1. Switching schools

During the first term of implementing the interventions in 2017, two issues emerged which may affect the sample balance. Firstly, one school had to be dropped from intervention 2 due to trouble caused by the principal. Secondly, in inviting the schools to the initial teacher training, confusion arose regarding two schools with the same name. This has meant that the incorrect school attended the teacher training.

With regards to the first issue, the project management team decided to still include the school in the data collection for the ensuing waves of data collection, but that the school will be considered as a non-compliant school. With regards to the second issue, the mistake was only recognised later in the study after implementation was well underway, and it was therefore decided to continue with the implementation as has been done for the first part of the year.

The project management team determined that the mix-up was merely an administrative mistake and therefore still consider the school allocation to be random. The concern, however, lies with the impact that this will have on future analysis. Both schools are in the Gert Sibande District, both schools are Quintile 1 schools and both schools were classified as low performers for the stratification. The original school was in stratum 6 (large Quintile 1 schools who have weaker performers), whereas the new intervention school is in stratum 2 (medium-sized schools from all quintiles who have weaker performers).

Sensitivity analyses will be conducted in all future reports to gauge the effect that this new allocation has on the final results. Table 20 below shows the balance test results on the baseline data, once the intervention status of the two schools was swapped.⁹ It is evident that this change does not affect the balance of the sample significantly, with only a minor additional imbalance in the visual perception task. This difference, however, does not influence the overall balance of the sample. A further test was done by dropping both schools from the sample, but this did not have any significant difference in the sample balance.

Table 17: Robustness check - Re-allocating intervention status

	Naming Animals	Word Recall	Non- word Recall	Phoneme Isolation	Compre- hension	Letter Sounds	Words Correct	Sentence Reading	Visual Perception	English Vocabulary
Paper- Based	0.15 (0.154)	-0.01 (0.118)	-0.02 (0.068)	-0.06 (0.12)	-0.01 (0.064)	-0.19 (0.486)	-0.06 (0.137)	-0.03* (0.016)	0.20* (0.12)	-0.02 (0.067)
Tablet- Based	0.39*** (0.149)	0.12 (0.118)	0.03 (0.095)	0.04 (0.13)	0.1 (0.063)	0.06 (0.477)	0.13 (0.164)	-0.02 (0.016)	0.21 (0.125)	0.01 (0.057)
Control Mean	7.155	9.981	4.208	1.129	2.179	4.652	0.387	0.051	1.46	0.836
N	3327	3327	3327	3327	3327	3327	3327	3327	3327	3327
Paper= Tablet: p-value	0.154	0.259	0.557	0.448	0.101	0.649	0.26	0.507	0.965	0.594

As a further check, the main results at the end of Grade 1 was run while excluding the two schools that were switched. This did not cause a significant change in the results.

7.2. Multi-grade and attrition

As noted under section 5.1., the attrition rate of intervention 2 learners was slightly higher than in the control of intervention 1 group. Although the difference between the groups was not statistically significant, a model was run including inverse probability weights to see whether this may influence the intervention coefficients.

Another check involves determining whether the inclusion of multi-grade schools may be affecting the results. The daily lesson plans are specifically designed to follow the Grade 1

⁹ The stratum allocation of the two schools were also swapped around. This decision essentially makes the assumption that the allocation of these schools to their new intervention status was random, and could have been based on the stratum they were allocated to. This ensures that the strata also remain equally sized.

curriculum and are therefore not appropriate for multi-grade settings. Although specific care has been taken to exclude multi-grade schools from the sample, there are still some teachers that responded that they were teaching a multi-grade school in 2017. Table 19 below shows that neither of these issues had any significant effect on the intervention effects in the main results.

Table 18: Sensitivity check - Multi-grade and Attrition

	(1)	(2)	(3)
	Main	Multi-grade	Attrition
On-site coach	0.529*** (0.083)	0.501*** (0.090)	0.558*** (0.086)
Virtual coach	0.456*** (0.084)	0.440*** (0.094)	0.451*** (0.084)
Observations	3,060	2,583	3,059
R-squared	0.218	0.208	0.232
P-value	0.434	0.434	0.434

7.3. Improvements due to increase in time spent teaching EFAL

As discussed earlier, the CAPS curriculum provides teachers with the choice of teaching EFAL for two or three hours a week. The lesson plans used in EGRS II are based on a maximum time (3 hours) allocation. One sensitivity check is therefore to determine whether the improvements in the English Oral Reading Proficiency items are due to improved teaching practices as a result of the interventions, or merely because of the increase in time spent on teaching EFAL.

The concern in trying to disentangle these mechanisms lies in controlling for teacher selection effects. There may be some unobservable characteristics that leads certain teachers to deciding to teach the maximum English time. These unobservable characteristics may also be correlated with learner performance and therefore influence the coefficients of interest. To disentangle the two mechanisms, we therefore need to control for these unobservable characteristics.

In the baseline questionnaire, teachers were asked how much time they plan to spend on teaching EFAL per week. Since the baseline questionnaire was conducted before the schools received the training, there is no reason to expect that any teachers in the intervention groups would be systematically more likely to have planned to teach the maximum hours EFAL curriculum. We can therefore compare the learner performance in the intervention groups to the learner performance of control learners in classrooms where the teachers responded that they are planning teach for 3 hours.

Figure 18 suggest that the teachers in the intervention groups may have been more likely than the teachers in the control group to have planned to teach EFAL for three hours a week. To test for this, a balance test was run on the variable which captures the amount of time that

teachers plan to spend on teaching EFAL. The results from this test supports the theory that the teachers in the different groups were not systematically more likely to have planned to teach for 3 hours.

To determine whether the intervention effects are a result of the increased time spent on teaching EFAL, the sample was restricted to only include learners that were in classrooms in which the teachers responded that they were planning of teaching EFAL for 3-hours a week in 2017. Table 19 shows that the intervention effects have remained similar in size and significance despite this restriction, suggesting that the effects are not only caused by the increased time spent on teaching EFAL, but that the intervention itself is successful in improving learning outcomes.

Figure 18: Time spent on teaching EFAL per week

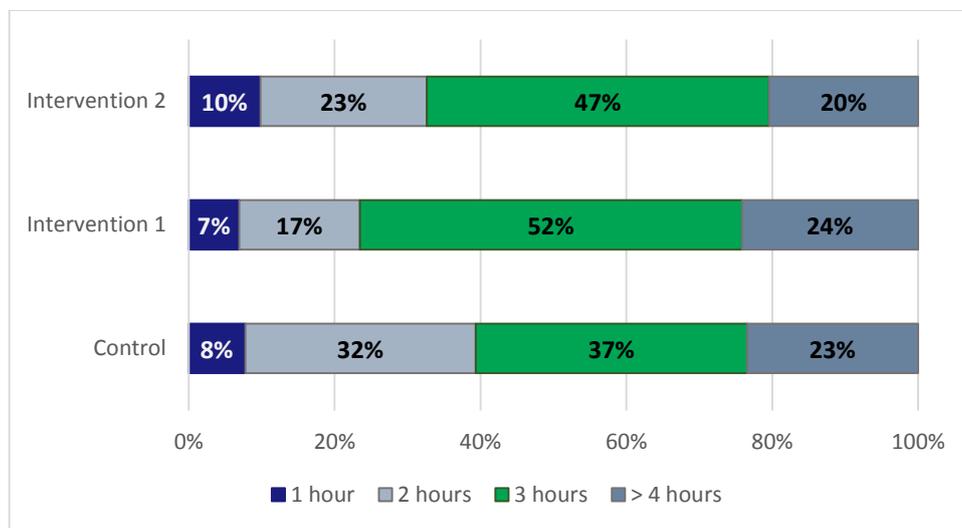


Table 19: Restricting to only include learners in 3-hour a week classes

	(1) ESL Listening	(2) ESL L Comprehension	(3) ESL Vocabulary	(5) English Oral Proficiency
On-site coach	0.452*** (0.109)	0.271** (0.116)	0.609*** (0.097)	0.641*** (0.116)
Virtual coach	0.307** (0.120)	0.039 (0.115)	0.591*** (0.104)	0.454*** (0.133)
Observations	1,214	1,214	1,212	1,212
R-squared	0.187	0.079	0.205	0.228
P-value	0.189	0.0531	0.874	0.179

7.4. Crowding out of home language teaching time

The increased time spent on teaching EFAL can also have a negative effect on the teaching of Home Language through a crowding-out effect. The crowding-out effect can work through two different mechanisms, firstly, opting for teaching EFAL for 3 hours a week means teaching home language for 7 hours a week instead of the 8 hours under the minimum EFAL time option. The second mechanism flows from teachers receiving additional support in EFAL and for this reason also giving more attention to teaching EFAL than home language. This may particularly be the case when teachers have lost some teaching time and need to decide which subject to prioritise.

Evaluating the first effect is very tricky, since it is not possible to control for the unobservable teacher characteristics that would have caused teachers to switch from the minimum to maximum EFAL time. The second mechanism is easier to measure, since we would like to determine whether there are any negative effects on the home language items, regardless of teachers having planned to spend 3 hours teaching EFAL a week. That is, comparing the learner performance in the control and intervention schools to each other, for learners who are in classrooms where the teachers upfront already planned to teach 3 hours EFAL a week. The sample was therefore once again restricted to only include learners in classes of teachers who responded that they spent 3 hours a week teaching EFAL. The model was then run on the sub-tasks that related to the skill which learners would have been taught during their home language lessons. Table 20 shows that while the coefficients on the intervention variables are still negative, they are mostly no longer significant. The negative coefficient on the first sub-tasks remains significant, but given the ceiling effects in this sub-tasks, this is most likely caused by a data anomaly. More concerning is the negative and significant coefficient on the Letter Sound Recognition tasks which seem to suggest some negative spill-over effect.

Table 20: The effect of crowding out Home Language Teaching

	(1)	(2)	(3)	(4)	(5)
	HL Vocabulary	HL Word Recognition	Phonemic Awareness	Letter Sounds	ESL Word Recognition
On-site coach	-0.175** (0.072)	0.004 (0.125)	-0.005 (0.131)	-0.005 (0.123)	0.052 (0.123)
Virtual coach	-0.026 (0.076)	-0.151 (0.129)	-0.216 (0.136)	-0.307** (0.125)	-0.131 (0.127)
Observations	1,217	1,215	1,215	1,217	1,214
R-squared	0.040	0.162	0.186	0.190	0.121
P-value	0.126	0.239	0.0991	0.0313	0.182

8. Cost-effectiveness discussion

The results of the EGRS I study found that instructional coaching as the professional development component of a structured pedagogic programme is more cost-effective than the centralised training model. Building on the findings of EGRS I, the current study investigates the sustainability of alternative models, which would allow government to take coaching to scale. The evidence after one year of implementation suggests that the on-site coaching and virtual coaching interventions are equally effective in improving student English oral proficiency. The resources required would therefore determine whether the virtual coaching model would be a more viable method.

For cost estimates, the programme budget for the first year of implementation was taken, excluding any costs that were involved in the development of the programme. These estimates should therefore provide a realistic per-student cost if these models of delivery are scaled up. Based on these preliminary estimates, the per-student costs of the on-site coaching and the virtual coaching models do not differ dramatically and are US\$48 and US\$43, respectively, per year.¹⁰ The main cost item in the on-site coaching model is the salary cost of the three coaches, while the additional night of residential training, the tablets, and cellular data for teachers are the main cost items in the virtual coaching model. A critical resource to the quality of both the on-site coaching and the virtual coaching models is the coaches. To support the 98 Grade 1 teachers in the 50 intervention 1 schools, three specialist reading coaches were employed, while one 'virtual' coach was employed in intervention 2 to support a similar number of Grade 1 teachers in 50 intervention 2 schools. The availability of expert reading coaches in each of the country's 11 home languages is therefore an important resource constraint that will need to be taken into account in decision-making regarding the feasibility of taking the coaching model to scale.

¹⁰ The cost for implementing the on-site coaching in 50 schools totalled US\$182,920, while the cost for implementing the virtual coaching programme totalled US\$164,673. Assuming an average class size of 76.8 students per school at the start of the programme, per-student spending is US\$48 and US\$43, respectively.

9. USAID indicators

Table 21: Specific USAID Indicators

	Indicator no	Indicators	FY October 2015 - September 2016				FY October 2016 - September 2017			
			Target	Actual	Male	Female	Target	Actual	Male	Female
Specific Indicators	ES. 1-1	Proportion of students who, by the end of two grades of primary schooling, demonstrate that they can read and understand the meaning of grade level text.	0%	0%	0%	0%	0%	0%	0%	0%
	ES. 1-1a;b	Percent learners	0%	0%	0%	0%	0%	0%	0%	0%
	ES. 1-1c;d	Numerator	0	0	0	0	0	0	0	0
	ES. 1-1e;f	Denominator	0	0	0	0	0	0	0	0
	ES. 1-5	Number of learners reached in reading programs at the primary level	0	0	0	0	7000	7600	4114	3486
	ES. 1-7	Number of primary school educators who complete professional development activities on implementing evidence-based reading instruction with USG assistance	0	0	0	0	212	188	0	188
	ES. 1-11	Number of primary school classrooms that receive a complete set of essential reading instructional materials with USG assistance	0	0	0	0	212	188	0	188

Table 22: Custom USAID Indicators

	Indicator no	Indicators	FY October 2015 - September 2016				FY October 2016 - September 2017			
			Target	Actual	Male	Female	Target	Actual	Male	Female
Custom Indicators	2.1.	Proportion of teachers that attended each compulsory training session	0	0	0	0	95%	94%	0%	100%
	2.2.	Number of principals/HOD's that attended each training session	0	0	0	0	50	27	1	26
	2.3.	Number teachers who attended needs driven workshops	0	0	0	0	50	66	0	66
	2.4.	Average number of interactions between the virtual coach and a teacher	0	0	0	0	1 per week	1 per week	0	93
	2.5.	Average number of visits a teacher receives from a reading coach.	0	0	0	0	1 per month	0.75 per month	0	93
	2.6.	Comparison of baseline and endline scores	0	0	0	0	0	0	0	0
	3.1.	Number of learner background questionnaires completed, as well as actual data	0	0	0	0	3,600	3,327	1,801	1,526
	3.2.	Number of teacher questionnaires completed, as well as actual data	0	0	0	0	360	320	0	320
	3.3.	Number of principal questionnaires completed, as well as actual data	0	0	0	0	180	180	98	82
	4.1.	Number and details of key stakeholders involved in planning	10	0	0	0	5	4	2	2
	4.2.	Number and details of research dissemination sessions with provinces and schools	0	0	0	0	5	4	.	.
	4.3.	Number and details of reports, journal articles and conferences	0	0	0	0	0	0	0	0

10. Next Steps in the project

Interventions continued throughout 2018, with the initial training having been conducted with the Grade 2 teachers in January 2018. Teachers further attended clustered workshop training for one day at the start of each term, and the reading coaches provided on-going support throughout the year. The third round of data collection will take place from 22 October to 9 November 2018 and will provide the data necessary to determine the impact of the interventions after two years of implementation. The interventions are set to continue to Grade 3 teachers in 2019 and the same learners will be tested at the end of Grade 3 in 2019 and again at the Grade 4 in 2020. In October 2018 a set of case studies were conducted to gain more detailed information on the aspects of the intervention which may be driving the success. Similarly, lesson observations in 60 schools and a further set of case studies are planned for 2019.

References

- Adams, M., 1990. *Beginning to Read: Thinking and Learning about Print*. Cambridge: MIT Press.
- Aploon-Zokufa, K., 2013. Locating the difference: A comparison of pedagogic strategies in high and low performing schools. *South African Journal of Childhood Education*, 3(2), pp. 112-130.
- Coe, R., Aloisi, C., Higgins, S. & Major, L., 2014. *What makes great teaching? Review of the underpinning research*, London: Sutton Trust.
- Diemer, M., Van der Merwe, K. & De Vos, M., 2015. The development of phonological awareness literacy measures for isiXhosa. *South African Linguistics and Applied Language Studies*, 33(3), pp. 325-341.
- DPME, forthcoming. *CAPS Implementation Evaluation*, Pretoria: Department of Planning, Monitoring and Evaluation.
- Duke, N., 2000. For the Rich It's Richer: Print Experiences and Environments Offered to Children in Very Low- and Very High-Socioeconomic Status First-Grade Classrooms. *American Educational Research Journal*, 37(2), pp. 441-478.
- Fleisch, B. & Christie, P., 2004. Structural Change, Leadership and School Effectiveness/Improvement: Perspectives from South Africa. *Discourse: Studies in the Cultural Politics of Education*, 25(1), pp. 95-112.
- Fleisch, B. & Schoer, V., 2014. Large-scale instructional reform in the Global South: Insights from the mid-point evaluation of the Gauteng Primary Language and Mathematics Strategy. *South African Journal of Education*, 34(3), pp. 1-11.
- Fleisch, B., Taylor, S., Schöer, V. & Mabogoane, T., 2017. Failing to catch up in reading in the middle years: The findings of the impact evaluation of the Reading Catch-Up Programme in South Africa. *International Journal of Educational Development*, Volume 53, pp. 36-47.
- Gathercole, S., 2006. Nonword repetition and word learning: The nature of the relationship. *Applied Psycholinguistics*, Volume 27, pp. 513-543.
- Hellman, L., 2012. *GPLMS Interses Catchup Programme: Analysis of Results*, s.l.: s.n.
- Hill, H., Charalambos, Y. & Kraft, M., 2012. When Rater Reliability Is Not Enough: Teacher Observation Systems and a Case for the Generalizability Study. *Educational Researcher*, pp. 56-64.
- Hoadley, U., 2012. What do we know about teaching and learning in South African primary schools?. *Education as Change*, 16(2), pp. 187-202.
- Hoadley, U., 2016. *A review of the research literature on teaching and learning in the foundation phase in South Africa*, Stellenbosch: Research on Socioeconomic Policy, Department of Economics, Stellenbosch University.
- Hoadley, U. & Galant, J., 2016. Specialisation and school organisation: Investigating pedagogic culture.. *British Journal of Sociology and Education*, 37(8), pp. 1187-1210.
- Howie, S. et al., 2017. *PIRLS 2016: South African Highlights Report*, Pretoria: University of Pretoria.
- Howie, S. et al., 2012. *PIRLS 2011 South African Children's reading literacy achievement: Summary Report*, Pretoria: Centre for Evaluation and Assessment, University of Pretoria.

- Macdonald, C., 2002. Are children still swimming up the waterfall? A look at literacy development in the new curriculum. *Language Matters*, 33(1), pp. 111-141.
- Macdonald, C., 2006. The properties of mediated action in three different literacy contexts in South Africa. *Theory and Psychology*, 16(1), pp. 51-80.
- Matjila, D. & Pretorius, E., 2004. Bilingual and Biliterate? An exploratory study of Grade 8 reading skills in Setswana and English. *Per Linguam*, 20(1), pp. 1-21.
- McDonald Connor, C. et al., 2009. The ISI Classroom Observation System: Examining the literacy instruction provided to individual students. *Educational Researcher*, pp. 85-99.
- McDonald, C. et al., 2009. The ISI Classroom Observation System: Examining the Literacy Instruction Provided to Individual Students. *Educational Researcher*, 38(2), pp. 85-99.
- NEEDU, 2013. *National Report 12. The state of literacy teaching and learning in the Foundation Phase*, Pretoria: Department of Basic Education .
- Nonoyama-Tarumi, Y. & Bredenberg, K., 2009. Impact of school readiness program interventions on children's learning in Cambodia. *International Journal of Educational Development*, 29(1), pp. 39-45.
- Pianta, R. C., La Paro, K. M. & Hamre, B. K., 2008. *Classroom Assessment Scoring System, Manual K-3*. Baltimore: Paul H Brookes Publishing.
- Piper, B., Zuilkowski, S. & Mugenda, A., 2014. Improving reading outcomes in Kenya: First-year effects of the PRIMR Initiative. *International Journal of Education Development*, 37(11-21), pp. 11-21.
- Piper, B., Zuilkowski, S. & Mugenda, A., 2014. Improving reading outcomes in Kenya: First-year effects of the PRIMR Initiative.. *International Journal of Educational Development*, Volume 37, pp. 11-21.
- Pretorius, E. & Currin, S., 2010. Do the rich get richer and the poor poorer? The effects of an intervention programme on reading in the home and schools language in a high poverty multilingual context.. *International Journal of Educational Development*, 30(1), pp. 67-76.
- Pretorius, E. & Lephala, M., 2011. Reading Comprehension in High-Poverty Schools: How should it be taught and how well does it work?. *Per Linguam*, 27(2), pp. 1-24.
- Pretorius, E. & Mokhwesana, M., 2009. Putting reading in Northern Sotho on track in the early years: Changing resources, expectations, and practices in a high poverty school. *South African Journal of African Languages*, 1(1), pp. 54-73.
- Raudenbush, S., 2005. Learning from attempts to improve schooling: The contribution of methodological diversity. *Educational Researcher*, 34(5), pp. 25-31.
- Reeves, C. et al., 2008. *Evaluation of literacy teaching in primary schools in Limpopo*, Polokwane: Limpopo Department of Education.
- Snilstveit, B. et al., 2016. *The impact of education programmes on learning and school participation in low- and middle-income countries: a systematic review summary report, 3ie Systematic Review Summary 7*, London: International Initiative for Impact Evaluation (3ie).
- Spaull, N., 2016. *Learning to Read and Reading to Learn*. Policy Brief. Research on Socioeconomic Policy (RESEP): (Online). Available: www.resep.sun.ac.za.

Taylor, N., 2007. Equity, efficiency and the development of South African schools. In: T. Townsend, ed. *International Handbook of School Effectiveness and Improvement. Springer International Handbooks on Education*. Dordrecht: Springer.

Taylor, N., Van der Berg, S. & Mabogoane, T., 2013. *Creating Effective Schools*. Cape Town: Pearson.

Taylor, S., 2011. Uncovering indicators for effective school management in South Africa using the National School Effectiveness Study. *Stellenbosch Working Paper Series: 10/11*.

Taylor, S., 2013. Modelling educational achievement. In: N. Taylor, S. Van der Berg & T. Mabogoane, eds. *Creating effective schools*. Cape Town : Pearson Education South Africa, pp. 63-101.

Taylor, S. & Watson, P., 2015. The impact of the study guides on "matric" performance: Evidence from a randomised experiment. *Stellenbosch Economic Working Papers: 13/15*.

Van der Berg, S., 2009. Fiscal incidence of social spending in South Africa. *Stellenbosch Working Paper Series: 10/09*.

World Bank Group, 2015. *User Guide: Conducting Classroom Observations*. s.l.:s.n.