

# THE SECOND EARLY GRADE READING STUDY



# YEAR 4 REPORT

Evidence one year after the end of implementation



**basic education**  
Department:  
Basic Education  
REPUBLIC OF SOUTH AFRICA

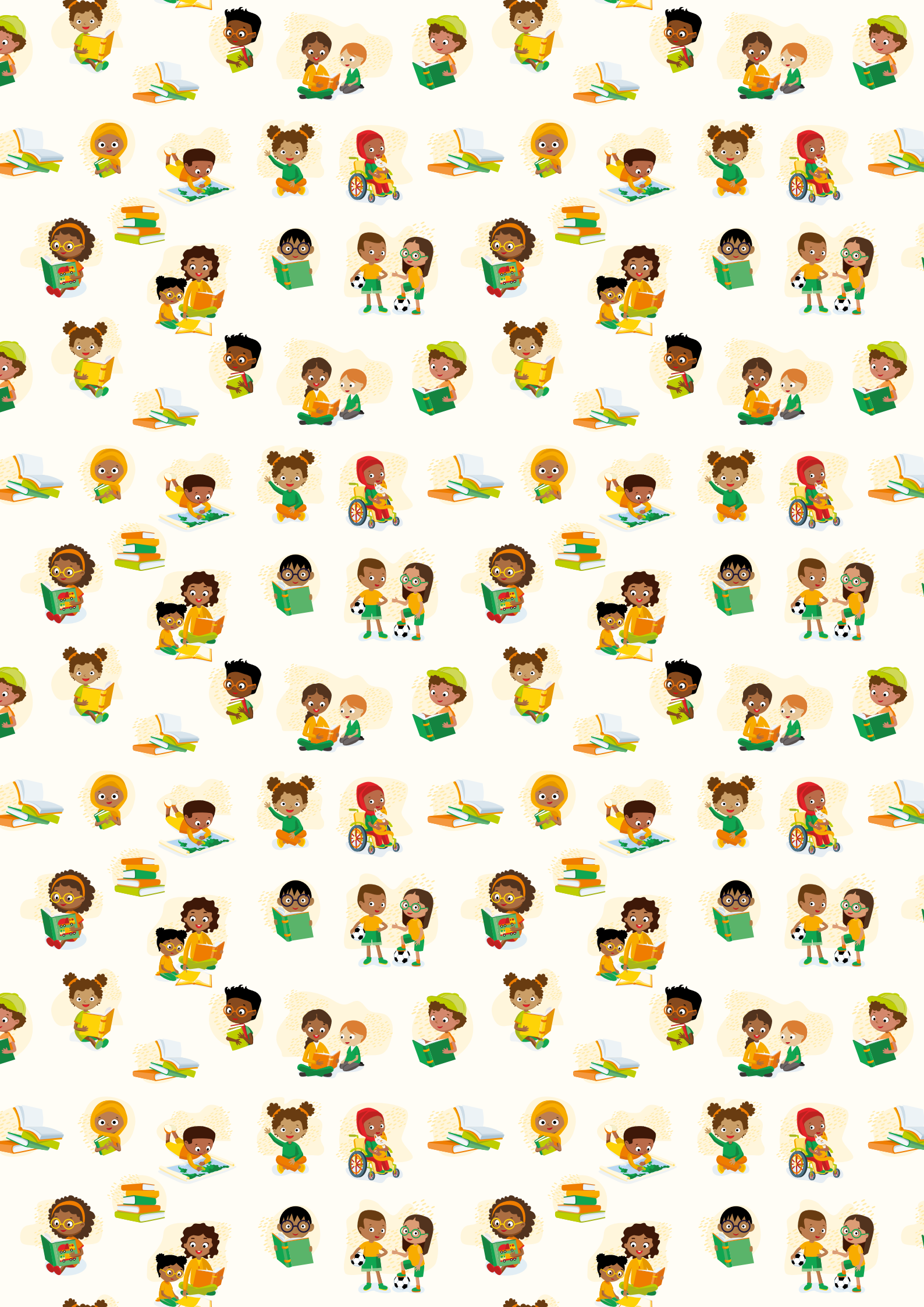


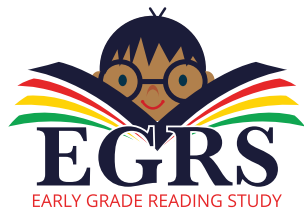
**education**  
DEPARTMENT: EDUCATION  
MPUMALANGA PROVINCE

UNIVERSITY OF THE  
WITWATERSRAND  
JOHANNESBURG



**USAID**  
FROM THE AMERICAN PEOPLE





# THE SECOND EARLY GRADE READING STUDY

# YEAR 4 REPORT

Evidence one year after the end of implementation



**basic education**  
Department:  
Basic Education  
REPUBLIC OF SOUTH AFRICA



**education**  
DEPARTMENT: EDUCATION  
MPUMALANGA PROVINCE

UNIVERSITY OF THE  
WITWATERSRAND  
JOHANNESBURG



**USAID**  
FROM THE AMERICAN PEOPLE



---

This work was led by the Research, Monitoring and Evaluation Directorate in the national Department of Basic Education with collaboration from the University of Witwatersrand and the University of Georgetown. The primary research team consists of Dr Stephen Taylor, Prof Brahm Fleisch, Dr Janeli Kotze, Prof Jacobus Cilliers and Ms Nompumelelo Mohohlwane. A special thank you to Ms Kholosa Nonkenge, Ms Carol Nuga Deliwe and Ms Tshegofatso Thulare who provided significant contributions to the work.

We are grateful for the financial support provided by the United States Agency for International Development (USAID). We are also grateful to the Mpumalanga Provincial Education Department as the research site partner.



# CONTENTS

|   |    |
|---|----|
| List of Figures   | ii |
| List of Tables  | ii |
| Acronyms and Abbreviations  | ii |
| Executive Summary   | 1  |
| 1. Background to the EGRS II  | 2  |
| 1.1. Intervention Design  | 2  |
| 1.2. Research Questions   | 4  |
| 2. The effect of the COVID pandemic   | 5  |
| 3. Year 4 Data Collection   | 6  |
| 3.1. Grade 4 Instruments  | 6  |
| 3.2. Data Collection  | 8  |
| 3.3. Impact of the COVID-19 pandemic on data collection                       | 8  |
| 3.4. Balance at Baseline  | 10 |
| 3.5. Attrition  | 11 |
| 4. Descriptive results: task-level learner performance                        | 13 |
| 4.1. English Word Recognition and Oral Reading Fluency                        | 15 |
| 4.2. English linguistic comprehension: Listening comprehension and Vocabulary | 17 |
| 4.3. English Oral and Written Reading Comprehension                           | 18 |
| 4.4. Home Language Reading outcomes   | 18 |
| 5. Main Results   | 20 |
| 5.1. Main Regression Findings on EFAL   | 20 |
| 5.2. Comparing the impact of the interventions between Grade 3 and Grade 4    | 22 |
| 5.3. Sub-Group Analysis   | 23 |
| 6. Secondary Outcomes   | 25 |
| 7. Sustained teacher practices  | 27 |
| Conclusion  | 31 |
| References  | 31 |

## List of Figures

|            |   |    |
|------------|---|----|
| Figure 1:  | Attrition rate by intervention group  | 11 |
| Figure 2:  | Improvements in oral reading fluency between Grade 2 and Grade 4  | 16 |
| Figure 3:  | Distribution of EFAL Word Reading and EFAL ORF scores by treatment group                                    | 16 |
| Figure 4:  | Learner performance in linguistic comprehension   | 17 |
| Figure 5:  | English reading and writing comprehension   | 18 |
| Figure 6:  | HL ORF for Grades 2 - 4   | 18 |
| Figure 7:  | Distribution of HL oral reading fluency   | 19 |
| Figure 8:  | HL reading outcomes by intervention group   | 19 |
| Figure 9:  | Learner performance on EFAL tasks – On-site coaching  | 20 |
| Figure 10: | Learner performance on EFAL tasks – Virtual coaching  | 21 |
| Figure 11: | Learning gains in oral reading fluency and word recognition between Grade 3 and Grade 4                     | 22 |
| Figure 12: | Comparing the impact of the on-site coaching intervention in Grade 3 and Grade 4 on EFAL outcomes           | 23 |
| Figure 13: | Girls performing better than boys in English proficiency  | 24 |
| Figure 14: | Impact of the interventions across the learning distribution  | 24 |
| Figure 15: | The impact of on-site coaching and virtual coaching on learner HL outcomes                                  | 25 |
| Figure 16: | Comparing the impact of the on-site and virtual coaching intervention in Grade 3 and Grade 4 on HL outcomes | 26 |
| Figure 17: | Proportion of classes in which print-richness was classified as adequate                                    | 27 |
| Figure 18: | Access to and use of EFAL graded readers  | 28 |
| Figure 19: | Correct teaching of phonics sounds  | 29 |
| Figure 20: | Continuous professional development in 2019 and 2020  | 29 |
| Figure 21: | On-site support from a Subject Advisor or Head of Department  | 30 |

## List of Tables

|           |  |    |
|-----------|--|----|
| Table 1:  | Comparison of Intervention 1 and Intervention 2                      | 3  |
| Table 2:  | Effective days of teaching lost in 2020                              | 5  |
| Table 3:  | Learner assessment tasks across the various waves of data collection | 7  |
| Table 4:  | Correlation between the assessments across the waves                 | 8  |
| Table 5:  | Instrument response rates  | 9  |
| Table 6:  | Balance tests per task   | 10 |
| Table 7:  | Learner progression in Wave 5 data collection                        | 12 |
| Table 8:  | Balance on Grade 4 sample  | 12 |
| Table 9:  | Item descriptive statistics  | 13 |
| Table 10: | Tasks means in Wave 5, by intervention group                         | 14 |
| Table 11: | Comparing EFAL word recognition across the waves of data collection  | 15 |
| Table 12: | Raw scores by treatment group  | 21 |
| Table 13: | EGRA results by treatment group                                      | 26 |

## Acronyms and Abbreviations

|         |   |
|---------|---|
| DBE     | Department of Basic Education             |
| EFAL    | English as First Additional Language      |
| EGRS    | Early Grade Reading Study                 |
| EGRS II | Second Early Grade Reading Study          |
| HL      | Home Language                             |
| ICT     | information and communications technology |
| LTSM    | learning and teaching support material    |
| ORF     | oral reading fluency                      |
| PCA     | principal component analysis              |
| RCT     | randomised control trial                  |
| WCPM    | words correct per minute                  |

# EXECUTIVE SUMMARY



The Second Early Grade Reading Study was conducted in Mpumalanga from 2017 – 2020, with the interventions running from 2017 – 2019 and a follow-up assessment one year after the interventions concluded.

The study was designed as a randomised control trial and focussed on strengthening the teaching of English as a First Additional Language in the Foundation Phase. The intervention was a structured pedagogy programme consisting of lesson plans with integrated teaching materials, teacher training and ongoing coaching. The learners who were assessed as part of the study were tracked from Grade 1 through to Grade 4, with the Grade 4 assessment allowing for the evaluation of the sustainability of the initial impacts one year after the learners were not exposed to the intervention anymore.

The study evaluated two different delivery models – the first being the traditional model of paper-based lesson plans and regular in-person coaching visits, and the second exploring the role that technology can play through the provision of the lesson plans on tablets and regular support to the teachers via a 'virtual coach'. The tablets provided in the second intervention were preloaded with the same lesson plans as in the first intervention, but also included audio and video clips to help teachers better understand the English phonic sounds and the teaching methodologies. The virtual coach was a specialist reading coach, similar to the coaches doing the in-person visits, but communicated with the teachers regularly via WhatsApp and phone calls.

This report presents the results of the Wave 5 data collection during which the Grade 4 learners were assessed at the end of 2020. This, however, was the year during which the COVID-pandemic disrupted schooling severely globally. The data collected was therefore particularly useful in measuring the impact of the severe loss of schooling on learning outcomes and the results of this evaluation can be found in Ardington, Wills & Kotze (2021). This report will specifically focus on whether the effects of the EGRS II interventions were sustained one year after the conclusion of the interventions. Nevertheless, all results should be interpreted while bearing in mind the significant loss in teaching time in 2020.

Similar to what was seen in the first Early Grade Reading Study in the North West province, the Wave 5 results show that the on-site coaching intervention had a sustained impact. In EGRS II the sustained impact was seen on learners' linguistic comprehension skills such as listening and reading comprehension, as well as English vocabulary. There was, however, some fade-out of the magnitude of the impact, with an 18% – 24% decrease in the impact on listening and reading comprehension. The small impact that was seen on the decoding skills in Grade 3 were not sustained one year later. Similarly, no evidence was found of any impact for the virtual coaching intervention in Grade 4. The negative spillover effects that were seen on the Home Language outcomes at the end of Grade 3 however, were still evident one year later.





# 1. BACKGROUND TO THE EGRS II

This report builds on the information provided in the year 1 - year 3 reports of the Second Early Grade Reading Study (EGRS II). These reports provide detailed information about the implementation of the project over the three years, as well as the impact that was seen on both the English First Additional Language (EFAL) and the Home Language (HL) reading outcomes at the end of each year of intervention. The focus of the year 4 report is on the 2020 data collection and the sustainability of the impact of the interventions on learners' reading outcomes one year after their exposure to the programme.

It is important to note that the year 4 results cannot be interpreted separately from the COVID-pandemic that caused major disruptions to teaching and learning for the majority of the academic year. Ardington, Wills & Kotze (2021) compared the EGRS II year 4 results with two other reading studies in South Africa to quantify the impact of the pandemic on learning losses. This report will therefore not comment on the impact of the pandemic, but will rather focus on the sustainability of the interventions on learner outcomes. Given that the pandemic and the ensuing school closures affected both intervention groups and the control group equally, we do not expect the sustainability of the interventions to have been impacted disproportionately in any of the groups. Nevertheless, the results in this report should be interpreted as the sustainability of the interventions after a year of very limited schooling.

Details of the study site, the school selection process and the evaluation design are included in the year 1 report. As a summary, the EGRS II is a randomised control trial (RCT) that evaluates two early grade reading interventions in 180 primary schools in two districts in the province of Mpumalanga, South Africa. The EGRS II was implemented with Grade 1 teachers in 2017, in 2018 the interventions were targeted at the Grade 2 teachers, and in 2019 at the Grade 3 teachers. Across all three years of the implementation, the study focused on measuring the causal impact on learner reading performance and unpacking the change mechanisms of the structured pedagogic programme. Two classroom observation studies and two case studies were also conducted over the study period to provide the research team with a deeper understanding of the mechanisms through which change happened.

## 1.1. Intervention Design

The EGRS II focused on the early learning of English as a second language (officially named English as First Additional Language, or EFAL, in the South African curriculum) by providing specific resources, training and ongoing coaching to teachers. The interventions trialled were based on the official government curriculum, formally referred to as the National Curriculum Statements Grades R-12. As such, the interventions were designed to improve and strengthen teachers' enactment of the official curriculum, and not to evaluate and comment on the curriculum.

Both interventions consisted of three components: (1) detailed lesson plans, (2) integrated learning and teaching support materials (LTSM), and (3) instructional coaching and training. The main difference between the two intervention models was in the delivery model of the lesson plans and the coaching support. In Intervention 1, the teachers received a paper-based version of the lesson plans and benefitted from regular on-site coaching with a specialised reading coach that visited the teachers in their classrooms. The coaches would also provide needs-based workshops after school with clusters of teachers within close proximity of each other. In Intervention 2, the teachers received a tablet with an electronic version of the lesson plans, including various audio-visual resources to provide teachers with additional guidance on the implementation of the lesson plans. The teachers in this

intervention were supported through an information and communications technology (ICT) coaching model that included telephone calls and cell phone messaging. The electronic lesson plans were provided on an application that was specifically developed for the study. The application was available offline to ensure functionality without data and connectivity concerns. This application contained additional electronic resources such as short training videos, sound clips of the phonics sounds, songs and rhymes and examples of learners' work.

Teachers from both interventions received training at the start of each term. The first training session was residential training and entailed two days of training for Intervention 1 and three days of training for Intervention 2, with the additional day spent on orientating the teachers to the tablets. The remaining training sessions were one-day cluster training with smaller groups of teachers at the start of each term. The on-site coaches trained the teachers that they were coaching, but because there was only one virtual coach, additional trainers were utilised to assist with the training of the Intervention 2 teachers. The trainers rotated so that, once during the year, all the teachers in this intervention would be trained by the virtual coach once. If teachers from either intervention group did not manage to attend the training session, the on-site coaches organised a catch-up session to make sure that the teachers had the new materials and understand the instructional practices which were covered during the training.

In Intervention 1, teachers received visits from specialist reading coaches about once a month. During these visits, coaches modelled, supported and evaluated teachers' practices and monitored implementation fidelity. Coaching in Intervention 2 involved a phone call to each teacher once every two weeks, regular text messaging and the establishment of virtual communities of practice. The virtual reading coach used text messaging to provide teachers with weekly teaching tips, to answer questions on the lessons and to run bi-weekly competitions to see evidence of teachers' enactment of the lesson plans.

**Table 1: Comparison of Intervention 1 and Intervention 2**

|                                  | Intervention 1  | Intervention 2  |
|----------------------------------|---|---|
| <b>Provision of lesson plans</b> | Paper-based   | Electronic<br>On an application on a tablet   |
| <b>Provision of LTSM</b>         | Paper-based:<br>Big Books<br>Posters<br>Flashcards<br>Writing frames  | Paper-based:<br>Big Books<br>Posters<br>Flashcards<br>Writing frames  |
| <b>Coaching</b>                  | Coach visits the teacher in her classroom.<br><b>Once every three weeks.</b>  | Coach contacts the teacher via telephone calls and instant messaging (WhatsApp).<br><b>Once every two weeks.</b>  |
| <b>Training</b>                  | <b>Initial training:</b><br>2-day block training<br><b>Quarterly training:</b><br>1 day at the start of each term<br>Needs-based training:<br>As required | <b>Initial training:</b><br>3-day block training<br><b>Quarterly training:</b><br>1 day at the start of each term<br>Needs-based training:<br>Bi-weekly competitions* |
| <b>Core methodologies</b>        | Paper-based instructional manual  | Application-based instructions,<br>Includes videos, sound clips and photos of example writing   |

\*The bi-weekly competitions provide a platform for teachers to showcase their 'good practice' in instructional techniques and how they build print-rich classroom environments.

The teachers in Intervention 2 were also supplied with videos, to assist the virtual coach in 'modelling' lessons to the teachers. The video technology operated in three different ways. Firstly, there were videos demonstrating core methodologies that were pre-loaded onto the tablet. The teacher could access these videos through the tablet, and they showed a teacher or reading coach demonstrating the methodology in an authentic classroom context. The second type of videos was not pre-loaded onto the tablet but was rather utilised in an organic way within the WhatsApp group. For example, at the beginning of each week, the virtual coach filmed herself saying the phonic sounds and words for the coming week. Finally, videos were used in the bi-weekly competitions. In this case, teachers were asked to take pictures or film videos in their classrooms of their practices. This was an opportunity for the virtual coach to gain 'eyes' into the classrooms she was supporting and to see how teachers were implementing the instructional practices and core methodologies. The video (or picture) submissions, in turn, were designed to help the coach better assess the areas she should concentrate on when coaching the teachers.

## 1.2. Research Questions

The rationale for supporting the teaching of EFAL in the Foundation Phase is to prepare learners better for the language transitional that they face in Grade 4. The theory of change for intervening in EFAL is that the improved teaching of reading and language comprehension in English in the Foundation Phase will enable learners to have a greater command of the English language in Grade 4 when most of their subjects will be taught in English. From a sustainability perspective, we would therefore expect that the stronger foundations laid during Grades 1 – 3 will translate into greater reading fluency and English comprehension by the end of Grade 4.



**The EGRS II is designed as an RCT that evaluates the difference between two different coaching models. Our main research questions for the first four waves were:**

1. **Did on-site coaching improve learning outcomes in EFAL?**
2. **Did virtual coaching improve learning outcomes in EFAL?**
3. **Did the impact on reading proficiency differ between the two coaching models?**
4. **Which model is the most cost-effective?**

**The data collection at the end of Grade 4 looked specifically at the sustainability of the programmes on learning outcomes, one year after implementation. The research questions in the year 4 report are therefore:**

1. **Is the impact on reading outcomes due to the on-site coaching sustained one year after implementation?**
2. **Is the impact on reading outcomes due to the virtual coaching sustained one year after implementation?**
3. **Which model is the most cost-effective, factoring in sustainability?**



## 2. THE EFFECT OF THE COVID PANDEMIC



The year 4 results cannot be disentangled from the severe impact of the COVID-19 pandemic. On 18 March 2020, all South African schools closed in an attempt to curb the spread of the COVID-19 virus. As international research emerged showing that learners were less likely than adults to become infected, as well as to infect others, school systematically started re-opening.

However, since research on the impact of the virus was still ongoing and public perceptions of the risk of the virus were highly variable, the re-opening plans had to be very adaptable and fluid. The Department of Basic Education (DBE) opted for two complementary models to the reopening of schools: The first was a phased-in approach to the returning of different grades and the second was a rotational model once learners had returned to school. The phased-in approach entailed bringing in the grades deemed most critical in primary and secondary schools first, while learners in other grades still remained at home. The other grades were subsequently brought in on an incremental basis. The purpose of the phased-in approach was to allow school management teams the time to trial COVID protocols with fewer numbers of learners in attendance per day. This played out practically with the Grade 12 and Grade 7 learners returning to school on 8 June 2020 and the Grade 11 and Grade 6 learners returning a month later on 6 July 2020. However, all public schools were closed again on 27 July 2020 after which the second round of phased reopening commenced in August 2020 (DBE, 2020).

In 2019, the total number of teaching days for all grades totalled 199 days. With the closing of schools in 2020 and the subsequent phased reopening, between 44 and 97 days of teaching were lost for the different grades. In 2020, the Grade 4 learners in our sample officially lost 42% of the school days of a regular school year.

Over and above the number of official teaching days lost, further 'unaccounted' for learning was lost due to the social-distancing regulations. On 24 August 2020, all grades returned to school, but the social distancing regulations required learners to keep a distance of 1.5 meters between them. The large majority of South African schools do not have the floor capacity to fit all learners with the 1.5-meter requirement and were forced to implement rotational teaching to lower the learner numbers attending school at the same time.

The DBE proposed three different rotational models (bi-weekly rotation, daily rotation or platooning) and schools could choose which of the rotational models best-suited learners in their context. Regardless of the option chosen, the implications of the rotational model in low resourced schools or in schools with larger class sizes were that learners lost a further half of the remaining school days. Table 2 below shows estimates of the effective days of learning as a result of the phased-in approach to reopening in conjunction with the rotational approach. The Grade 4 learners in our sample therefore effectively **lost 59% of their school year**, assuming that learners were not absent and attended school on the days allocated to them.

**Table 2: Effective days of teaching lost in 2020**

| Grades       | Jan – 18 Mar | June | July | Aug | Sep | Oct | Nov | Total: 2020 | Total: 2019 | % of school days lost* |
|--------------|--------------|------|------|-----|-----|-----|-----|-------------|-------------|------------------------|
| Gr 7 & 12    | 46           | 7    | 10   | 10  | 11  | 9   | 9   | 101         | 199         | 49%                    |
| Gr R, 6 & 11 | 46           |      | 8    | 10  | 11  | 9   | 9   | 91          | 199         | 54%                    |
| Gr 3 & 10    | 46           |      | 5    | 10  | 11  | 9   | 9   | 89          | 199         | 56%                    |
| Gr 1 & 2     | 46           |      | 3    | 10  | 11  | 9   | 9   | 86          | 199         | 57%                    |
| Gr 4 & 9     | 46           |      |      | 8   | 11  | 9   | 9   | 81          | 199         | 59%                    |
| Gr 5 & 8     | 46           |      |      | 1   | 11  | 9   | 9   | 74          | 199         | 63%                    |

\* relative to 2019



## 3. YEAR 4 DATA COLLECTION

The purpose of the year 4 data collection was to assess the panel of learners a year after they were exposed to the intervention. If learners were repeating a grade, they were still assessed on the same instruments, since the EGRA instruments are not strictly grade-specific and can assess a range of ability levels.

The data collection entailed both an oral and written assessment of the panel of learners. Furthermore, we also interviewed the Grade 2 and Grade 3 teachers, to better understand the extent to which they were still making use of the methodologies they were trained on one year after the intervention.<sup>1</sup> Finally, a document review was also conducted of the exercise and workbooks on one learner in the class of the teacher interviewed.

### 3.1 Grade 4 Instruments

The learner assessments throughout the study were designed to evaluate learners' language and literacy abilities at the end of each grade but were not designed to necessarily benchmark learner performance against curriculum requirements. Given this focus, care was taken to minimise floor effects. As was the case with the wave 1 – 4 testing, the tests were designed to take no longer than 15 minutes to administer. The evaluation team piloted the instruments in five schools one month before the main data collection, after which the required amendments were made to ensure that the assessments are fair and appropriate. The written assessment was designed to be administered in a group setting and the written assessments were marked and captured on SurveyCTO by the fieldworkers afterwards. The oral assessments were administered one-on-one and captured on SurveyCTO directly.

**Table 3** on the following page shows the various assessment tasks that have been included across the five waves of data collection. The learner assessment at the end of Grade 4 included an oral assessment comprising five tasks assessing HL and EFAL linguistic comprehension and reading proficiency. The written assessment included both written comprehension and vocabulary assessments in both HL and EFAL, as well as a very short mathematics task.



<sup>1</sup> Given the major disruptions to schooling during 2020, it was not expected that teachers will still be using the lesson plans provided. For this reason, the teacher interviews were designed with realistic expectations of the teaching methodologies that teachers would have been able to continue with.

**Table 3: Learner assessment tasks across the various waves of data collection**

| Construct      |                                      | Baseline     |      | Year 1     |      | Year 2     |      | Year 3     |      | Year 4     |      |
|----------------|--------------------------------------|--------------|------|------------|------|------------|------|------------|------|------------|------|
| Language Comp. |                                      | Start - Gr 1 |      | End - Gr 1 |      | End - Gr 2 |      | End - Gr 3 |      | End - Gr 3 |      |
|                |                                      | HL           | EFAL | HL         | EFAL | HL         | EFAL | HL         | EFAL | HL         | EFAL |
|                | Receptive vocabulary                 |              | x    |            | x    |            | x    |            |      |            |      |
|                | Expressive vocabulary                | x            | x    | x          | x    |            | x    |            | x    |            | x    |
|                | Listening comprehension              | x            |      |            | x    |            | x    |            | x    |            | x    |
| Decoding       | Phonological working memory          | x            |      |            |      |            |      |            |      |            |      |
|                | Phonological awareness               | x            |      |            | x    |            |      |            |      |            |      |
|                | Rapid letter naming                  |              |      |            |      | x          |      | x          |      |            |      |
|                | Letter-sound recognition             | x            |      | x          |      | x          |      | x          |      |            |      |
|                | Word reading fluency                 | x            |      | x          | x    |            | x    |            | x    |            | x    |
|                | Sentence reading fluency             | x            |      |            |      |            |      |            |      |            |      |
|                | Oral reading fluency (ORF)           |              |      |            |      | x          | x    | x          | x    | x          | x    |
|                | Reading comprehension                |              |      |            |      | x          | x    | x          | x    | x          | x    |
| Writing        | Written comprehension                |              |      |            |      |            |      | x          | x    | x          | x    |
|                | <sup>1</sup> Spelling of a CVC word* |              |      |            | x    |            |      |            |      |            |      |
|                | Writing two words                    |              |      |            |      |            | x    |            |      |            |      |
|                | Written receptive vocabulary         |              |      |            |      |            |      |            |      | x          | x    |
|                | Written vocabulary                   |              |      |            |      |            |      |            |      |            | x    |

\*A CVC word is made up of a consonant, vowel and consonant sound, e.g. dog.

The end of Grade 4 assessments focussed mostly on reading fluency and reading comprehension since learners are expected to be able to read for meaning by the end of Grade 3. Extensive EFAL and HL vocabulary tasks were also included in the written assessment to enable us to identify whether challenges in reading comprehension lie with reading fluency, or rather comprehension of English as a second language.

The end of Grade 4 assessments focussed mostly on reading fluency and reading comprehension since learners are expected to be able to read for meaning by the end of Grade 3. Extensive EFAL and HL vocabulary tasks were also included in the written assessment to enable us to identify whether challenges in reading comprehension lie with reading fluency, or rather comprehension of English as a second language

Home Language (i.e. Siswati and isiZulu) and Mathematics tasks were included to examine possible spill-over and crowding-out effects. In this context, crowding-out would involve teachers using more time for EFAL than the amount the curriculum has earmarked for teaching literacy in the HL and/or Mathematics. Positive spill-overs, on the other hand, would be the result of learners or teachers transferring skills taught in EFAL to their HL.

**Table 4** shows the correlation between the overall index scores between the assessments and it is evident that the correlations between the later waves are stronger than the correlations with the baseline assessment. This is mostly due to the difficulty of measuring literacy and language outcomes at the start of formal schooling.

**Table 4: Correlation between the assessments across the waves**

|                                       | Wave 1 | Wave 2 | Wave 3 | Wave 4 EFAL oral | Wave 4 EFAL reading | Wave 5 |
|---------------------------------------|--------|--------|--------|------------------|---------------------|--------|
| Wave 1                                | 1      |        |        |                  |                     |        |
| Wave 2                                | 0.398  | 1      |        |                  |                     |        |
| Wave 3                                | 0.378  | 0.751  | 1      |                  |                     |        |
| Wave 4 EFAL oral language proficiency | 0.371  | 0.534  | 0.658  | 1                |                     |        |
| Wave 4 EFAL reading proficiency       | 0.339  | 0.656  | 0.839  | 0.694            | 1                   |        |
| Wave 5                                | 0.383  | 0.659  | 0.828  | 0.744            | 0.897               | 1      |

## 3.2 Data Collection

The Wave 5 data collection was conducted by external service providers and entailed the data collection of reading tests to the same group of 20 learners per school, now in Grade 4, across 180 schools in the two districts of Ehlanzeni and Gert Sibande, Mpumalanga. Sixteen teams of two fieldworkers visited at least one school per day over three weeks from 9 November to 27 November 2020 with a limited number of teams continuing into the fourth week for final mop-up data collection.

Fieldworkers were trained for four days, of which the first two days focused solely on the learner assessment tools and the third day of training was dedicated to in-school simulations. The final day was dedicated to an assessor competency assessment, the contextual questionnaires and final logistical arrangements.

## 3.3 Impact of the COVID-19 pandemic on data collection

Given the COVID-19 pandemic, the data collection teams were faced with several challenges, including an altered school calendar, school activities and learner attendance. To reduce the risk of infection, fewer learners attended school per day with the exact arrangements being left to each school. This meant that on any given day only a certain proportion of learners were present. In some schools, different grades attended on different days though this was sometimes limited to certain weeks. Given that there were only three weeks of fieldwork planned, at some

schools this meant that learners would only be present for one of the three weeks. In other cases, certain classes came to school and in others, only certain learners attended each day regardless of their class. This led to many learners being marked absent on the first visit to the school.

Fortunately, teams were often able to make arrangements with the school to return on a day when those absent learners were scheduled to attend. Rather than make all return visits during the last week of fieldwork, return visits were interspersed throughout the fieldwork period with some initial visit dates having to be shifted. School officials were always informed about changes to the schedule. The limited number of learners also made it possible for teams to split up and visit multiple schools per day especially in cases where only half of the sampled learners were present. One of the schools was often a school that had already been visited.

Despite schools being instructed not to administer exams to learners during the final term, the practice was found to be quite common especially during the last 2 weeks of fieldwork. The exam schedule had both benefits and drawbacks. The benefits were that exams were administered on a strict schedule, so teams were able to plan a time to return and were often able to visit a nearby school to look for previously absent learners. In addition, because exams were mandatory, teams could expect most learners to be present. The drawback was that less time was available to do assessments. In general, the amount of time to conduct assessments was reduced by two hours.

Due to time constraints, fieldworkers were often forced to choose between conducting additional EGRAs and conducting the written assessment. Because classes were not being conducted, there was reduced supervision of learners. Although learners were identified and asked to remain at school to be assessed, some learners left the school either before being assessed or before the written assessment was conducted. In cases where fieldworkers chose to conduct EGRAs, a return visit to conduct the written assessment was scheduled.

**Table 5** shows the final response rates for each of the instruments administered. Given the complexities mentioned above, the response rate for the learner assessments was higher than expected. Schools were visited on at least two occasions in most cases. This allowed for the maximum number of learners to be assessed given regular absences and the platooning of learners as a result of COVID-19 guidelines. At several schools, however, the second visit was not possible because schools had finished exams and learners were found to no longer be attending.

No learners could be assessed at two of the schools. In one of these schools, the teachers were still present and could be interviewed, however, the second school had already closed down for the December holidays and no assessments or interviews could be arranged.

**Table 5: Instrument response rates**

| Research tool                     | Total expected | Total collected | Response rate |
|-----------------------------------|----------------|-----------------|---------------|
| <b>Learner oral assessment</b>    | 3,327          | 2,405           | 72%           |
| <b>Learner written assessment</b> | 2,405          | 2,352           | 98%           |
| <b>Gr 2 teacher interview</b>     |                | 173             |               |
| <b>Gr 3 teacher interview</b>     |                | 168             |               |
| <b>School observation</b>         | 180            | 177             | 98%           |

## 3.4 Balance at Baseline

As reported in the baseline report as well as subsequent reports, the sample was balanced on the baseline assessment at the start of Grade 1. There was a slight imbalance on one of the sub-tasks, but since we are making 20 comparisons below, this is in line with what is expected. Moreover, the p-value of the F-test shows that we cannot reject the null for the joint significance across all the indicators. There is therefore no evidence of imbalance at baseline.<sup>2</sup>

**Table 6: Balance tests per task**

|   | Control          | On-site          | Virtual           | On-site<br>vs C | Virtual<br>vs C |
|---|------------------|------------------|-------------------|-----------------|-----------------|
|   | Mean/SE          | Mean/SE          | Mean/SE           | (1)-(2)         | (1)-(3)         |
| <b>Naming animals in HL</b>                   | 7.155<br>[0.127] | 7.310<br>[0.155] | 7.501<br>[0.154]  | -0.155          | -0.346*         |
| <b>Word recall</b>                            | 9.981<br>[0.084] | 9.953<br>[0.093] | 10.081<br>[0.092] | 0.028           | -0.099          |
| <b>Non-word recall</b>                        | 4.208<br>[0.049] | 4.179<br>[0.052] | 4.237<br>[0.082]  | 0.029           | -0.030          |
| <b>Phoneme isolation</b>                      | 1.129<br>[0.087] | 1.037<br>[0.092] | 1.161<br>[0.107]  | 0.092           | -0.032          |
| <b>Story comprehension</b>                    | 2.179<br>[0.045] | 2.154<br>[0.050] | 2.263<br>[0.047]  | 0.025           | -0.084          |
| <b>Letter sounds correct</b>                  | 6.978<br>[0.447] | 6.784<br>[0.590] | 7.019<br>[0.610]  | 0.194           | -0.041          |
| <b>Words read correct</b>                     | 0.387<br>[0.096] | 0.347<br>[0.103] | 0.510<br>[0.148]  | 0.039           | -0.123          |
| <b>Sentence words read correct</b>            | 0.051<br>[0.012] | 0.027<br>[0.011] | 0.034<br>[0.012]  | 0.024           | 0.018           |
| <b>Visual perception</b>                      | 1.460<br>[0.082] | 1.597<br>[0.111] | 1.651<br>[0.109]  | -0.137          | -0.192          |
| <b>English items</b>                          | 0.836<br>[0.044] | 0.789<br>[0.063] | 0.839<br>[0.045]  | 0.047           | -0.003          |
| <b>N</b>                                      | <b>1459</b>      | <b>924</b>       | <b>944</b>        |                 |                 |
| <b>Clusters</b>                               | <b>80</b>        | <b>50</b>        | <b>50</b>         |                 |                 |
| <b>F-test of joint significance (p-value)</b> |                  |                  |                   |                 | <b>0.782</b>    |
| <b>F-test, number of observations</b>         |                  |                  |                   |                 | <b>2383</b>     |

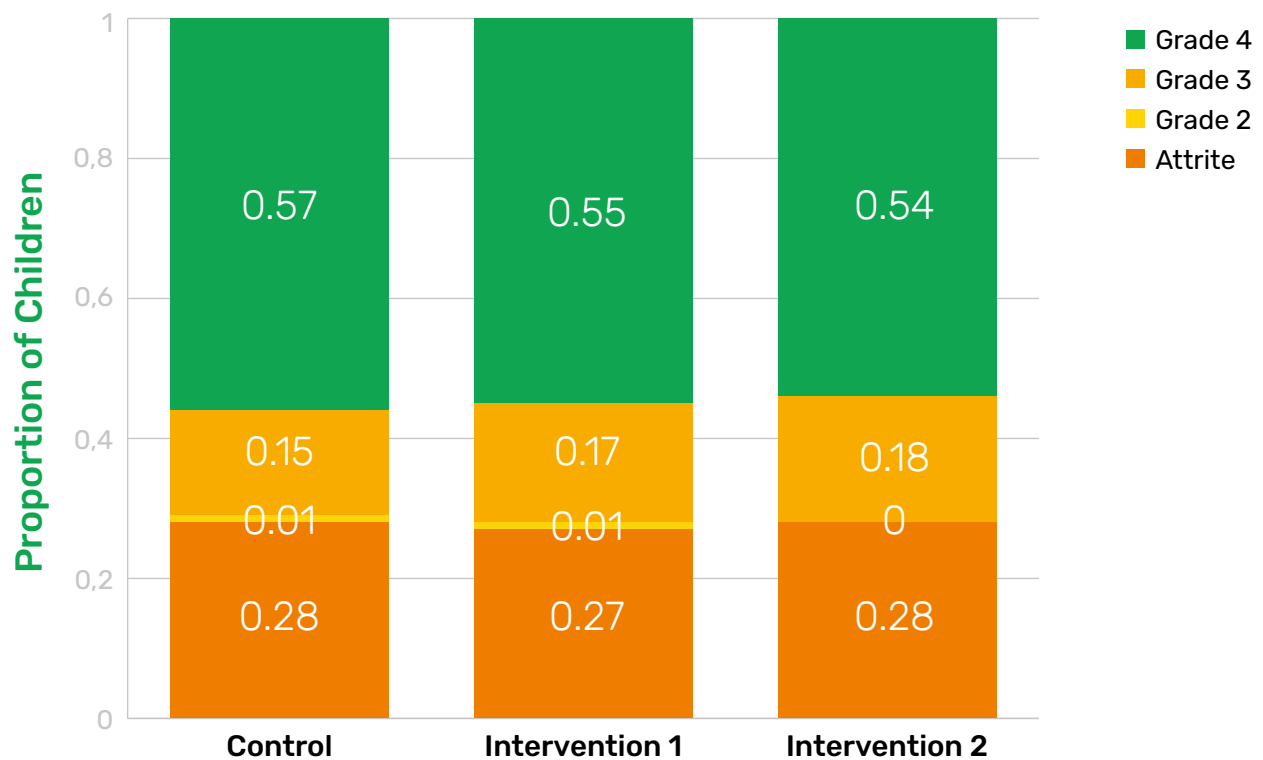
Note. The value displayed for t-tests is the differences in the means across the groups. Standard errors are clustered at the school level. \*\*\*, \*\*, and \* indicate significance at the 1, 5 and 10 percent critical level.

<sup>2</sup> Previous reports provide further information regarding the balance on school and learner characteristic at baseline as well.

### 3.5 Attrition

During the year 4 data collection, 2,352 of the 3,327 learners who were tested during the baseline data collection were re-tested on both the oral and written assessment and were successfully matched to their baseline results. The overall attrition rate of 29% is slightly higher than the rate in previous studies, but is still encouraging given that this is the fifth wave of data collection, and taking into account the additional challenges introduced by the pandemic. When breaking down the attrition rate by intervention group, the differences are not statistically significant, with the attrition rate of learners in the control schools (28%) being very similar to the attrition rate among learners in the virtual coaching (28%) and on-site coaching schools (27%).

**Figure 1: Attrition rate by intervention group**



**Table 7** further explore whether there is any relationship between the likelihood of repeating a grade and the treatment groups for the sample of learners who were assessed in wave 5. The results show that there are no significant imbalances in the intervention groups in terms of learners who repeated a year or two. The first row suggests that male learners, learners in the isiZulu schools and learners with lower scores at baseline seem to be more likely to have repeated a year. Although learners in the virtual coaching arm in wave 5 seem to have had slightly higher baseline learning scores, the interaction between having repeated and virtual coaching is very small.

**Table 7: Learner progression in Wave 5 data collection**

|                | (1) Delayed | (2) Age | (3) Male | (4) Zulu | (5) BL learning |
|----------------|-------------|---------|----------|----------|-----------------|
| Delayed        |             | -0.075  | 0.238*** | 0.106**  | -0.533***       |
|                |             | (0.082) | (0.036)  | (0.050)  | (0.082)         |
| On-site coach  | 0.024       | 0.041   | 0.007    | -0.029   | 0.030           |
|                | (0.029)     | (0.063) | (0.025)  | (0.075)  | (0.089)         |
| Virtual coach  | 0.035       | 0.080   | 0.009    | -0.015   | 0.169**         |
|                | (0.032)     | (0.062) | (0.031)  | (0.079)  | (0.079)         |
| Delayed x T1   |             | 0.057   | -0.070   | 0.004    | -0.042          |
|                |             | (0.110) | (0.052)  | (0.070)  | (0.131)         |
| Delayed x T2   |             | -0.122  | -0.020   | -0.092   | -0.029          |
|                |             | (0.111) | (0.054)  | (0.073)  | (0.122)         |
| Observations   | 2,405       | 2,405   | 2,405    | 2,405    | 2,405           |
| R-squared      | 0.006       | 0.020   | 0.034    | 0.151    | 0.083           |
| Mean attrition | 0.217       |         |          |          |                 |

Nevertheless, we checked the possible impact of attrition and grade repetition on the resultant sample of Grade 4 learners. **Table 8** shows the balance on the sample of learners who were assessed in Grade 4 in 2020. It seems as if the virtual coaching group may have had slightly higher baseline learning scores than the control group. There were no other imbalances in the key learner characteristics. To control for this imbalance, the baseline learning outcomes will be included as controls in the regressions.

**Table 8: Balance on Grade 4 sample**

|               | Age     | Male    | BL learning | Zulu    |
|---------------|---------|---------|-------------|---------|
| On-site coach | 0.054   | -0.005  | 0.008       | -0.025  |
|               | (0.059) | (0.022) | (0.076)     | (0.079) |
| Virtual coach | 0.044   | 0.010   | 0.142*      | -0.034  |
|               | (0.062) | (0.026) | (0.075)     | (0.080) |
| Observations  | 2,410   | 2,410   | 2,410       | 2,410   |
| R-squared     | 0.016   | 0.001   | 0.027       | 0.144   |
| Control mean  | 6.082   | 0.539   | 0.0213      | 0.304   |



## 4. DESCRIPTIVE RESULTS: TASK-LEVEL LEARNER PERFORMANCE

**Table 9** provides information on the descriptive statistics of the assessment tasks administered at the end of year 4. The scores in the table include the averages, an indication of the performance distribution, and the percentage of learners that scored zero on the task. The purpose of the table is to provide insights on how learners on average performed relative to the maximum score in the task. It further shows that there was a good distribution of scores in the Grade 4 assessment that will provide sufficient information to differentiate learner performance across the sample.

The zero scores indicate that the percentage of non-readers (i.e. the learners that could not read a single word correctly) was still remarkably high at the end of Grade 4. At the end of Grade 1, about 48% of learners could not read a single word correctly in the HL. Two years later, at the end of Grade 3, 18% of the Grade 3 learners still did not read a single word correctly in the HL. The proportion of non-readers declined slightly to 15% in 2020, despite the little schooling due to the pandemic. The percentage of non-readers in EFAL was about the same (18%), even though the word length in the English language is shorter than in isiZulu and Siswati.

**Table 9: Item descriptive statistics**

|                   | N    | Mean | s.d. | Min. | Max. | p10  | p25  | p50 | p75 | p90 | % zero score |
|-------------------|------|------|------|------|------|------|------|-----|-----|-----|--------------|
| HL ORF 60s        | 1839 | 28   | 17.5 | 0    | 62   | 0    | 15   | 30  | 41  | 51  | 15%          |
| HL ORF 180s       | 1839 | 46.7 | 22.4 | 0    | 62   | 0    | 45   | 59  | 60  | 61  | 15%          |
| HL ORF Compr.     | 1839 | 4    | 2.4  | 0    | 8    | 0    | 2    | 4   | 6   | 7   | 12%          |
| EFAL Word Recog.  | 1825 | 33.8 | 22.7 | 0    | 105  | 0    | 14   | 36  | 51  | 63  | 16%          |
| EFAL ORF 60s      | 1839 | 44.4 | 33.5 | 0    | 126  | 0    | 11   | 47  | 71  | 88  | 16%          |
| EFAL ORF 180s     | 1839 | 83.3 | 47.6 | 0    | 126  | 0    | 44   | 109 | 119 | 124 | 18%          |
| EFAL ORF Compr.   | 1839 | 2.4  | 2.4  | 0    | 8    | 0    | 0    | 2   | 4   | 6   | 31%          |
| EFAL Prod. Vocab. | 1839 | 18.9 | 6.6  | 0    | 30   | 8    | 16   | 20  | 24  | 26  | 1%           |
| EFAL L. Compr.    | 1839 | 1.5  | 1.2  | 0    | 4    | 0    | 1    | 1   | 2   | 3   | 23%          |
| HL W. Compr.      | 1809 | 3.1  | 2    | 0    | 6    | 0    | 1    | 3   | 5   | 6   | 19%          |
| HL Recep Vocab.   | 1809 | 3    | 2.3  | 0    | 10   | 0    | 1    | 3   | 4   | 6   | 15%          |
| EFAL W. Compr.    | 1809 | 1.6  | 1.7  | 0    | 7    | 0    | 0    | 1   | 3   | 4   | 37%          |
| EFAL Recep Vocab. | 1809 | 4.3  | 2.4  | 0    | 8    | 1    | 2    | 4   | 6   | 8   | 7%           |
| EFAL W. Vocab.    | 1809 | 2.1  | 1.9  | 0    | 8    | 0    | 1    | 2   | 3   | 5   | 24%          |
| Maths             | 1809 | 3.2  | 2    | 0    | 10   | 1    | 2    | 3   | 4   | 6   | 5%           |
| EFAL Prof.        | 1795 | 0.7  | 2.4  | -3.4 | 7.1  | -2.3 | -1.3 | 0.4 | 2.5 | 4.2 |              |
| HL Prof.          | 1809 | 0.3  | 1.7  | -2.7 | 4    | -2.5 | -0.9 | 0.6 | 1.6 | 2.2 |              |

Notes: Sample only includes the Grade 4 learners. The statistics for the full sample are in Table A1 in the Appendix. The construction of the indices is further explained in section 5.1

It should be noted that unless otherwise specified, the descriptive statistics shown in this section only include the scores of the learners who were still on track (that is, in Grade 4 at the end of year 4). In the Appendix, the same tables are shown for the full sample of learners.

**Table 10** shows the average scores for each of the sub-tasks by intervention group. Column 4 suggests that the learners in the on-site coaching intervention group performed better than the control group learners in HL letter recognition, EFAL word recognition, EFAL oral reading fluency (ORF), reading comprehension, EFAL vocabulary and English listening comprehension. For learners in the virtual coaching group, the results were less significant, with some negative effects on the HL writing comprehension task.

**Table 10: Tasks means in Wave 5, by intervention group**

|                              | Control (1)       | On-site (2)<br>Coaching | Virtual (3)<br>Coaching | Difference<br>(1)-(2) | Difference<br>(1)-(3) |
|------------------------------|-------------------|-------------------------|-------------------------|-----------------------|-----------------------|
| HL ORF 60s                   | 29.136<br>[0.960] | 27.688<br>[0.965]       | 26.320<br>[1.359]       | 1.448                 | 2.816*                |
| HL ORF 180s                  | 48.424<br>[1.182] | 46.365<br>[1.262]       | 44.215<br>[1.817]       | 2.059                 | 4.209*                |
| EFAL Word Recognition.       | 32.991<br>[1.117] | 35.673<br>[1.318]       | 33.347<br>[1.932]       | -2.681                | -0.355                |
| EFAL ORF 60s                 | 43.915<br>[1.606] | 45.355<br>[1.903]       | 44.190<br>[2.796]       | -1.440                | -0.275                |
| EFAL ORF 180s                | 84.718<br>[2.402] | 83.196<br>[2.672]       | 81.042<br>[3.561]       | 1.522                 | 3.677                 |
| HL ORF Comprehension         | 4.123<br>[0.127]  | 3.941<br>[0.128]        | 3.751<br>[0.173]        | 0.182                 | 0.372*                |
| EFAL ORF Comprehension       | 2.252<br>[0.112]  | 2.612<br>[0.144]        | 2.476<br>[0.196]        | -0.360**              | -0.225                |
| EFAL Productive Vocabulary   | 18.581<br>[0.351] | 19.569<br>[0.393]       | 18.891<br>[0.570]       | -0.988*               | -0.311                |
| EFAL Listening Comprehension | 1.341<br>[0.054]  | 1.600<br>[0.069]        | 1.551<br>[0.096]        | -0.259***             | -0.210*               |
| HL Written Comprehension     | 3.156<br>[0.100]  | 2.968<br>[0.107]        | 2.978<br>[0.151]        | 0.187                 | 0.178                 |
| HL Receptive Vocabulary      | 3.141<br>[0.106]  | 2.942<br>[0.119]        | 2.720<br>[0.139]        | 0.198                 | 0.421**               |
| EFAL Written Comprehension   | 1.542<br>[0.078]  | 1.607<br>[0.102]        | 1.699<br>[0.122]        | -0.065                | -0.158                |
| EFAL Receptive Vocabulary    | 4.109<br>[0.109]  | 4.583<br>[0.148]        | 4.431<br>[0.183]        | -0.474**              | -0.322                |
| EFAL Written Vocabulary      | 2.042<br>[0.089]  | 2.119<br>[0.107]        | 2.092<br>[0.127]        | -0.077                | -0.050                |
| Maths                        | 3.246<br>[0.096]  | 3.014<br>[0.121]        | 3.194<br>[0.166]        | 0.232                 | 0.052                 |
| EFAL Proficiency             | 0.538<br>[0.117]  | 0.871<br>[0.146]        | 0.728<br>[0.216]        | -0.333*               | -0.191                |
| HL Proficiency               | 0.419<br>[0.088]  | 0.247<br>[0.094]        | 0.116<br>[0.127]        | 0.173                 | 0.303*                |

The value displayed for t-tests is the differences in the means across the groups. Standard errors are clustered at variable NatEmis. \*\*\*, \*\*, and \* indicate significance at the 1, 5, and 10 percent critical level. The mean scores do not include any controls. The statistics for the full sample are in Table A2 in the Appendix.

## 4.1 English Word Recognition and Oral Reading Fluency

The simple view of reading theorises that reading comprehension is a function of decoding and linguistic comprehension. The first skill that we therefore consider is the ability to decode words, or reading fluency, measured through English word recognition and English ORF. Although Foundation Phase learners spent the bulk of their time learning how to read in their home language, the EFAL curriculum does allocate time for learners to learn to read in English.

In the word recognition task, learners were given a chart of 104 words in English arranged from the simplest two-letter words to complex multi-syllable ten-letter words. Learners were asked to correctly name each of the words on the chart in sixty seconds.

At the end of Grade 3, we noted that learners in the on-site coaching intervention read about 4 more words correct per minute (wcpm) than the control group, whereas the learners in the virtual coaching intervention performed at the same level as the learners in the control group. At the end of Grade 4, learners in the on-site coaching group on average still read more words correctly per minute than the learners in the virtual coaching and control group, but this has decreased from 4 more words correctly to 2 more words correctly. It is important to note that there were more non-readers (learners who could not read a single word correctly) in the virtual coaching group than in the other two groups.

**Table 11: Comparing EFAL word recognition across the waves of data collection**

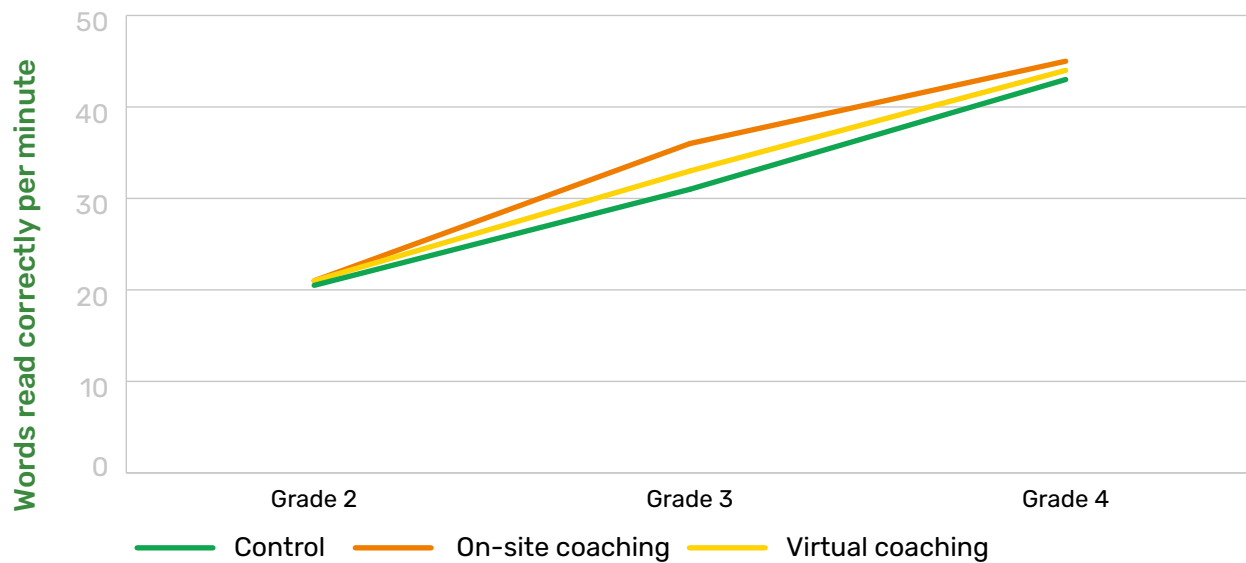
|                  | End of Grade 1        |                   | End of Grade 2        |                   | End of Grade 3                   | End of Grade 4 | End of Grade 4 |
|------------------|-----------------------|-------------------|-----------------------|-------------------|----------------------------------|----------------|----------------|
|                  | Decodable words (30s) | Sight words (30s) | Decodable words (30s) | Sight words (30s) | Decodable + Sight Combined (60s) |                | % zero         |
| Control          | 2.5                   | 2.7               | 18.5                  | 18.0              | 26.4                             | 33.0           | 11%            |
| On-site coaching | 2.7                   | 2.8               | 20.0                  | 19.7              | 30.8                             | 35.7           | 11%            |
| Virtual coaching | 2.3                   | 2.3               | 18.0                  | 17.9              | 26.9                             | 33.3           | 15%            |

Notes: Statistics shown are the mean scores and does not include any controls. The sample excludes all repeaters. The statistics for the full sample are in Table A3 in the Appendix. In Grades 1 and 2, learners were assessed on decodable and sight words separately, and each assessment was only done for 30 seconds. In Grades 3 and 4 the decodable and sight word lists were combined and learners were assessed for 60 seconds.

The second measure for reading fluency entails the reading of an English passage (the traditional ORF measure). There is a growing consensus about the validity and reliability of ORF as a key measure of reading and it is now widely accepted that the ability to read connected texts rapidly, accurately and with expression, is a critical competency required for successful reading for understanding.

A similar pattern that was seen in English word recognition is seen in English ORF. At the end of Grade 2, there was very little difference in the reading fluency between the different intervention groups in EFAL (**Figure 2**). The EFAL curriculum introduces an increased focus on reading in Grade 3 as can be seen in the increased reading fluency in the control group. It was encouraging to note that learners in both intervention groups saw a slightly higher rate of increase in their EFAL reading fluency than the control group in Grade 3. However, the higher rate of learning that was experienced by the intervention groups during Grade 3, diminished in Grade 4 and there was some catch-up by the control group.

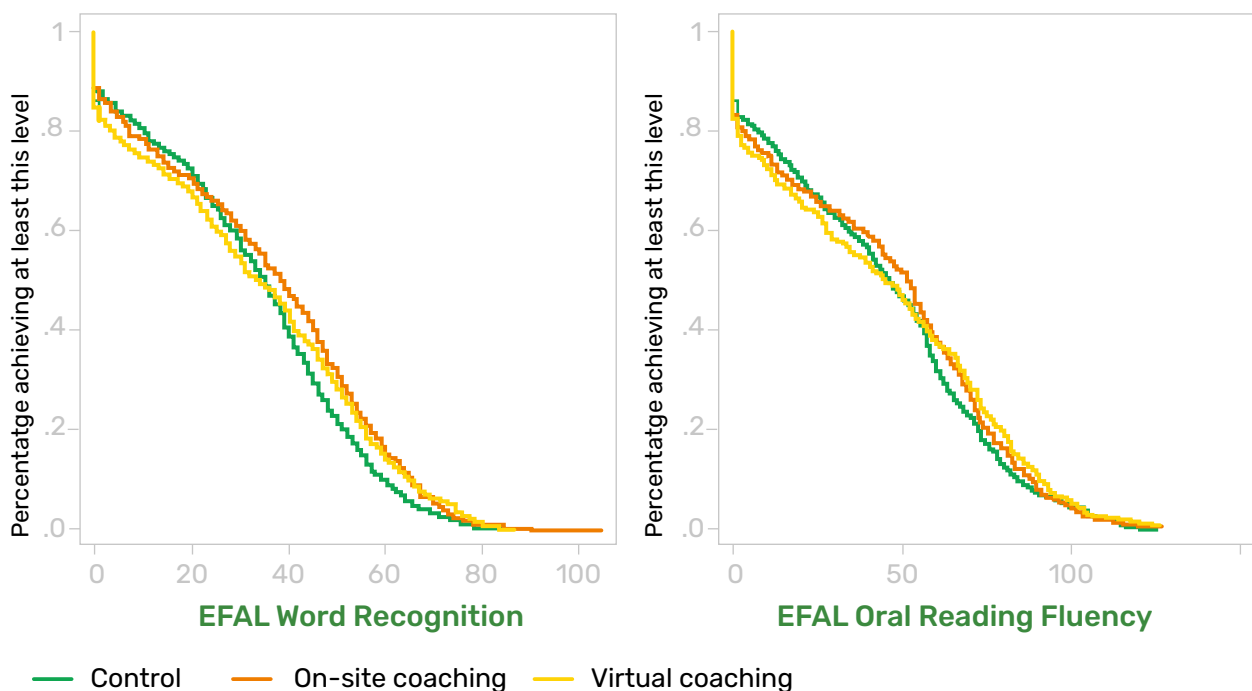
**Figure 2: Improvements in oral reading fluency between Grade 2 and Grade 4**



Notes: Sample excludes all repeaters and the mean scores do not include any controls. The statistics for the full sample are in Figure A1 in the Appendix.

**Figure 3** shows the performance distribution of the Grade 4 learners in the three different groups for both word recognition and oral language fluency. Both graphs seem to tell a similar story, with learners in the control group performing better than the learners in the two interventions groups at the bottom-end of the distribution, but at the top-end of the distribution, learners in the intervention groups seem to have been performing better than the control group learners. The point at which the learners in the on-site intervention group start to out-perform the learners in the control group is much lower (at around 30 wcpm), than for the virtual coaching group (at around 40 wcpm for word recognition, and 50 wcpm for ORF).

**Figure 3: Distribution of EFAL Word Reading and EFAL ORF scores by treatment group**

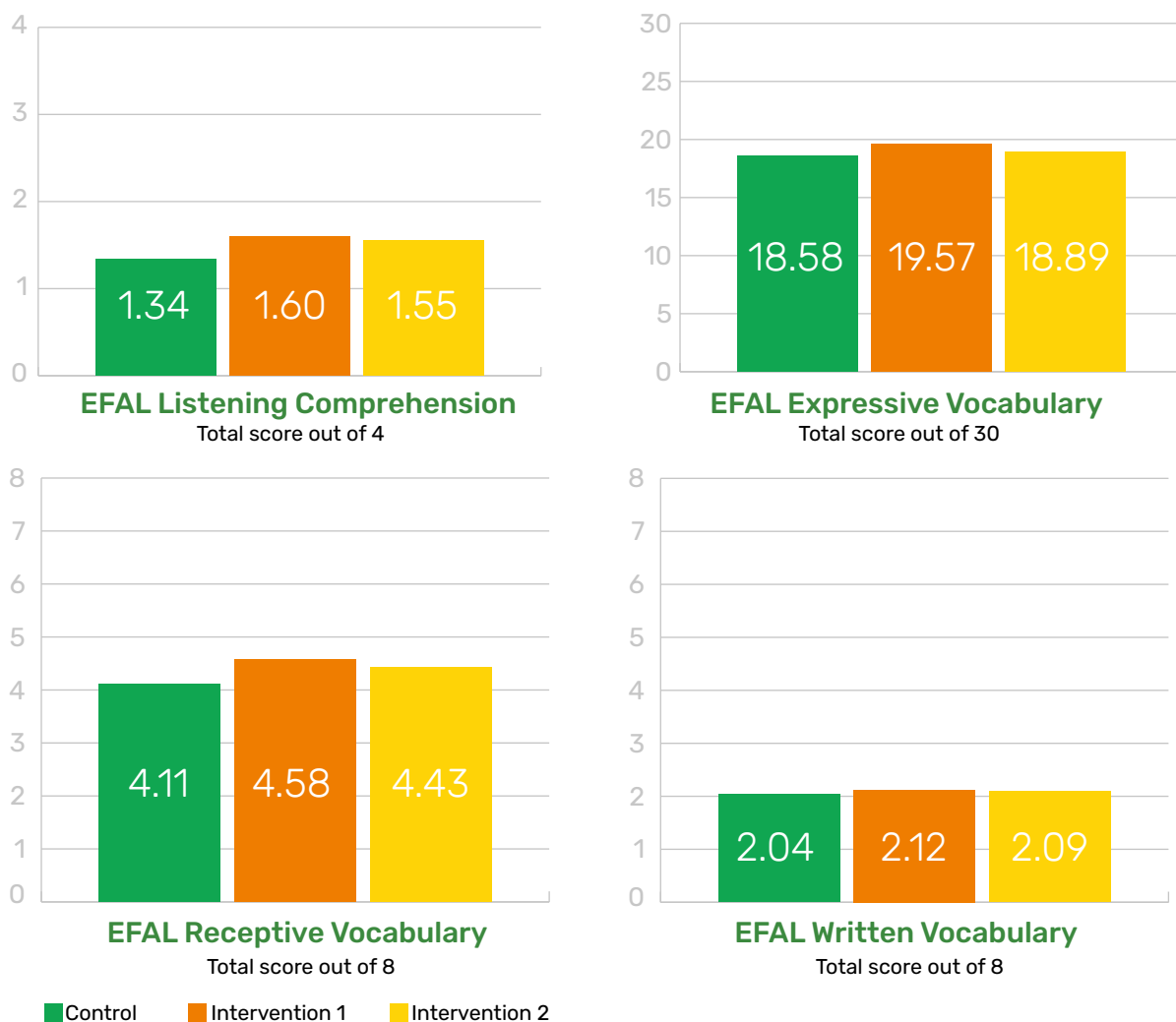


## 4.2 English linguistic comprehension: Listening comprehension and Vocabulary

The second skill required for reading with meaning is linguistic comprehension, measured in our assessment through a listening comprehension task, an English expressive vocabulary task, an English receptive vocabulary task and an English written vocabulary task. To measure listening comprehension, learners needed to respond to a set of questions related to a short story that was read to the learner in English. The English expressive vocabulary task entailed showing the learners a series of 30 pictures and asking the child to name the item in English. In the written English receptive vocabulary task, learners were expected to fill in a missing word or a sentence using one of three options provided to them. They were given six minutes to complete eight sentences. The English written vocabulary task required learners to match a descriptive sentence to one word from a list of thirteen words and learners were given eight minutes to match eight descriptions to the thirteen words.

One of the striking findings in the year 1 analysis was the similarity in the gains for learners in both intervention groups relative to the control on English productive and expressive vocabulary. At the end of Grade 3, learners in the intervention groups had a more extensive English vocabulary than their peers in the control group, but they only had a marginal benefit in the listening comprehension task. At the end of Grade 4, we see learners in the intervention groups performing better than learners in the control group in the listening comprehension and receptive vocabulary task. However, in terms of magnitude, the differences between the groups are still very small.

**Figure 4: Learner performance in linguistic comprehension**

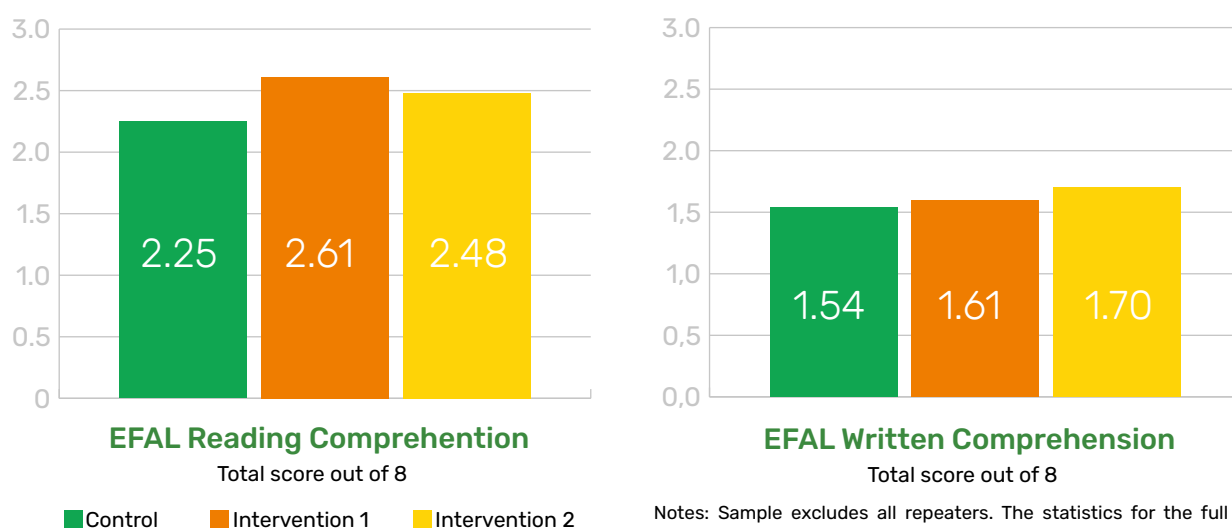


Notes: Sample excludes all repeaters. The statistics for the full sample is in Table A4 in the Appendix.

## 4.3 English Oral and Written Reading Comprehension

Having considered both reading fluency and language comprehension, the final skill for English that we consider is the function of these two: reading comprehension. To assess reading comprehension, we included an oral and written reading comprehension task. The oral reading comprehension task followed the ORF passage, whereas the written comprehension task required learners to read a passage independently and answer the questions in a written format. There was no time limit for the responses to the five questions that followed the ORF passage, and learners were given three minutes to read the passage. Learners were subsequently only asked questions relevant to the section up to where they read. For the written comprehension task, learners were given eight minutes to read the passage and answer the six questions in a written format. The descriptive comparison of the average scores between the three groups does not suggest any significant differences between any of the groups for either of the two tasks.

**Figure 5: English reading and writing comprehension**

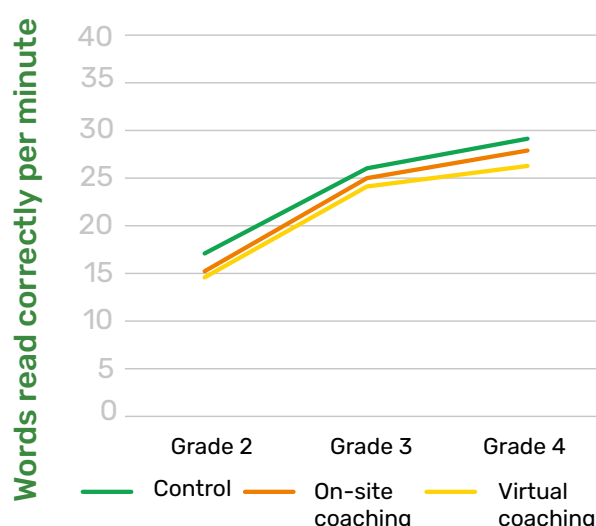


## 4.4 Home Language Reading outcomes

Although HL teaching was not part of either the on-site or virtual coaching interventions, the assessment of HL literacy skills provides important insights into the foundations on which learners build their second language skills. The HL items also allow us to evaluate whether there were any positive or negative spill-over or crowding-out effects as a result of the interventions. We evaluate the impact on HL reading outcomes by looking at HL ORF, HL reading and written comprehension and HL receptive vocabulary.

**Figure 6** shows the progression in HL reading outcomes for the HL ORF task since Grade 2. Across all three years, we note that the control group learners on average performed better than the intervention group learners. The average control group learner read about 29.1 wcpm, whereas the average on-site coaching group learners read about 27.7 wcpm and the average virtual coaching group learner read about 26.3 wcpm.

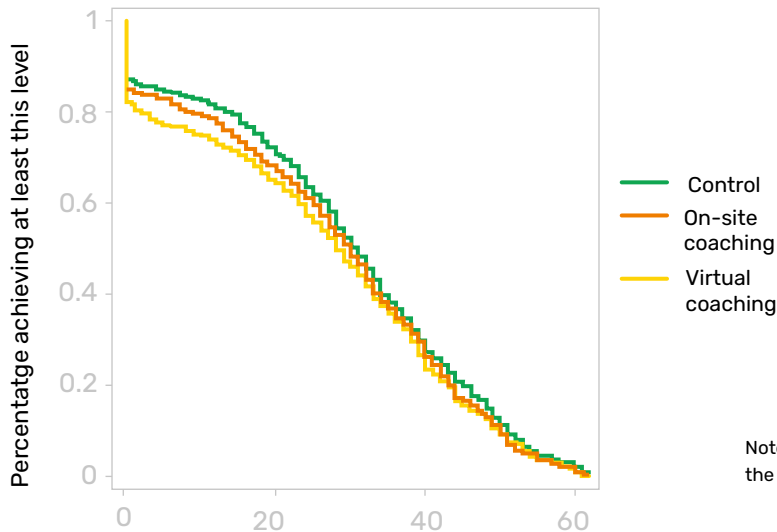
**Figure 6: HL ORF for Grades 2 - 4**



**Figure 7** shows the distribution of HL reading outcomes for the three groups and it is evident that the control group learners were more likely to have a better reading fluency throughout the full distribution. It is further clear that the virtual coaching group seem to be performing worse than both the control and the on-site coaching group throughout the distribution.

**Figure 7: Distribution of HL oral reading fluency**

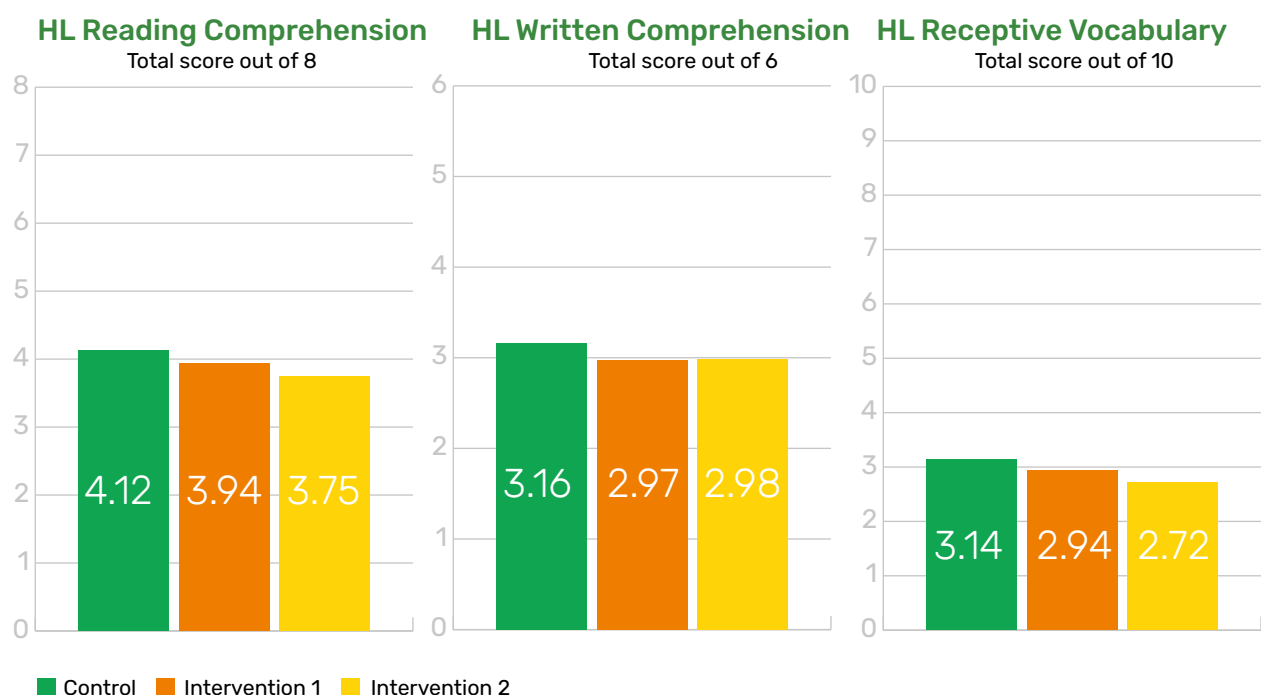
**HL Oral Reading Fluency**



Notes: Sample excludes all repeaters. The graph for the full sample is in Figure A3 in the Appendix.

The average scores for the three groups on the HL reading comprehension, the HL written comprehension and the HL vocabulary task all seem to corroborate the findings of the HL ORF task, with the outcomes in the intervention groups being lower than that of the control group. The year 3 report explored the reasons behind the lower performance in HL outcomes for the intervention groups extensively. The results above suggest that the effects of the crowding out of HL teaching in the Foundation Phase could possibly have been sustained into Grade 4.

**Figure 8: HL reading outcomes by intervention group**



Notes: Sample excludes all repeaters. The statistics for the full sample are in Table A6 in the Appendix.



## 5. MAIN RESULTS

### 5.1 Main Regression Findings on EFAL

In previous waves of data collection, we evaluated the overall impact of the interventions using two indices that are based on the two language constructs that learners of a second language have to master in the Foundation Phase. The first construct is language proficiency as it relates to English vocabulary development and the second relates to decoding skills. In the first grade, learners are only taught language proficiency skills during the English lessons, whereas the decoding skills are already taught during the Home Language lessons in the first grade. Decoding skills are only introduced in the English lessons from the second half of the second grade and build on the skills that learners were already taught in their HL. By the third grade, both language proficiency and decoding skills are consolidated and learners should be able to read for meaning. By the end of the fourth grade, we therefore combine the two language constructs to only consider a single EFAL index.

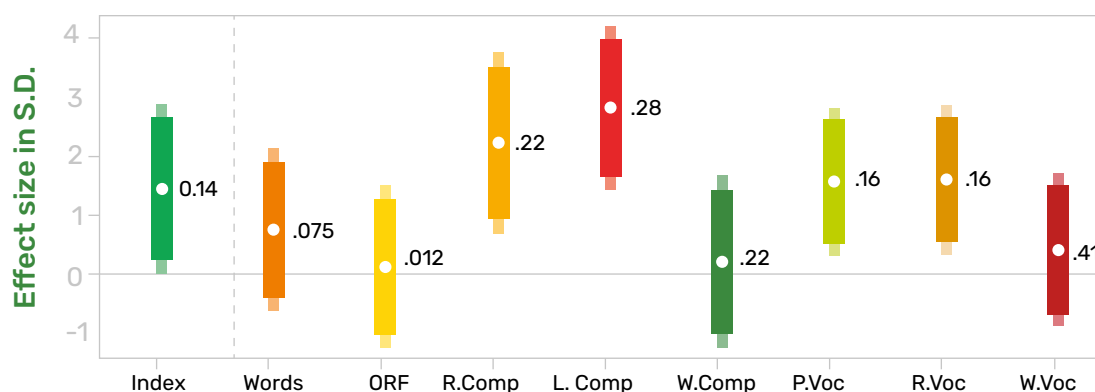
The index is constructed using principal component analysis (PCA), which is a statistical method that combines the various sub-tasks under each construct into one single score. Intuitively, the PCA creates an index that is reflective of the most common underlying construct of the sub-tasks included in the index. The index was constructed using only the control group's scores as they serve as our reference group. The index was then standardised on the control group mean and standard deviation so that the results can be interpreted in standard deviations relative to the control group (whose mean will now be zero).

**Figure 9** shows the estimated impacts for the Grade 4 learners who were in the schools that received the on-site coaching intervention. The first columns indicate the coefficient for the EFAL index, whereas the rest of the columns to the right of the dotted line shows the coefficients for the underlying sub-tasks. The point estimate is indicated by the white diamond, the confidence interval at the 95% level is shown by the darker coloured bar and the confidence interval at the 90% level is shown by the lighter colour bar.

The coefficients were derived from separate regressions run on each variable, controlling for the learners' scores on the baseline sub-tasks, learner gender, learner age, the stratification dummies and the fieldworker dummies. We decided which controls to include, based on the controls which explained the most variation in regressions run only on the control group. The regression table is shown in Table A7 in the Appendix.

It is clear from **Figure 9** that a year after the learners were exposed to the intervention, the learners who were in the on-site coaching intervention group still had a positive and significant advantage over their control group

**Figure 9: Learner performance on EFAL tasks – On-site coaching**

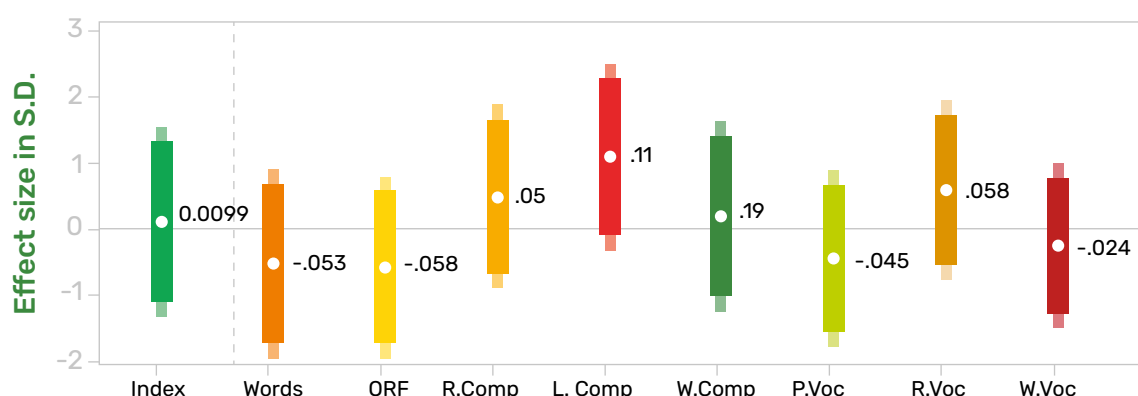


Notes: Sample excludes all repeaters. The regression outputs for full sample are in Table A8 in the Appendix.

peers. The coefficients on the sub-tasks indicate which sub-tasks are driving these results. Learners in the on-site coaching intervention did significantly better than their control group peers in the ORF reading comprehension, the listening comprehension, as well as in the productive and receptive English vocabulary tasks. The advantage in the decoding skills (word recognition and ORF), as well as in the written vocabulary tasks were negligible.

At the end of Grade 3, the impacts of the virtual coaching intervention were much less pronounced than the impact of the on-site coaching intervention. A small, and barely significant, positive impact was seen on the oral language proficiency score, and this was mostly driven by a positive impact on the listening comprehension score. Figure 10 shows that one year after the intervention concluded, there is no evidence of a sustained impact of the virtual coaching intervention. The positive impact that was seen on listening comprehension has also diminished to 0.11 standard deviations and is not statistically significant.

**Figure 10: Learner performance on EFAL tasks – Virtual coaching**



Notes: Sample excludes all repeaters. The regression outputs for full sample are in Table A8 in the Appendix.

Although standard deviation change is a useful measure of intervention impact, it is often difficult to understand the magnitude of the impact in terms of actual learning gains. **Table 12** shows the item scores on each of the English tasks, after having controlled for the baseline scores and other basic controls. The asterisks indicate whether the differences between the groups are statistically significant. For example, on average, learners in the on-site coaching group scored 2.7 out of 5 for the EFAL comprehension task, whereas learners in the control group scored on average 2.3 and learners in the virtual coaching group scored on average 2.4.

**Table 12: Raw scores by treatment group**

|               | Words correct per minute |              | Items correct   |                   |                |                |                   |                |
|---------------|--------------------------|--------------|-----------------|-------------------|----------------|----------------|-------------------|----------------|
|               | EFAL Word Recog.         | EFAL ORF 60s | EFAL ORF Compr. | EFAL Prod. Vocab. | EFAL L. Compr. | EFAL W. Compr. | EFAL Recep Vocab. | EFAL W. Vocab. |
| Control       | 33.0                     | 43.9         | 2.3             | 18.6              | 1.3            | 1.5            | 4.1               | 2.0            |
| On-site coach | 34.7                     | 44.3         | 2.7***          | 19.7***           | 1.7***         | 1.6            | 4.5***            | 2.1            |
| Virtual coach | 31.8                     | 42.0         | 2.4             | 18.2              | 1.5            | 1.6            | 4.2               | 2.0            |

\*\*\*, \*\*, and \* indicate significance at the 1, 5, and 10 percent critical level. Showing results for raw scores. Controlling for learner gender, learner age, baseline scores, district, school quintile, fieldworker and stratification dummies. Only includes Grade 4 learners. The statistics for the full sample are in Table A9 in the Appendix.

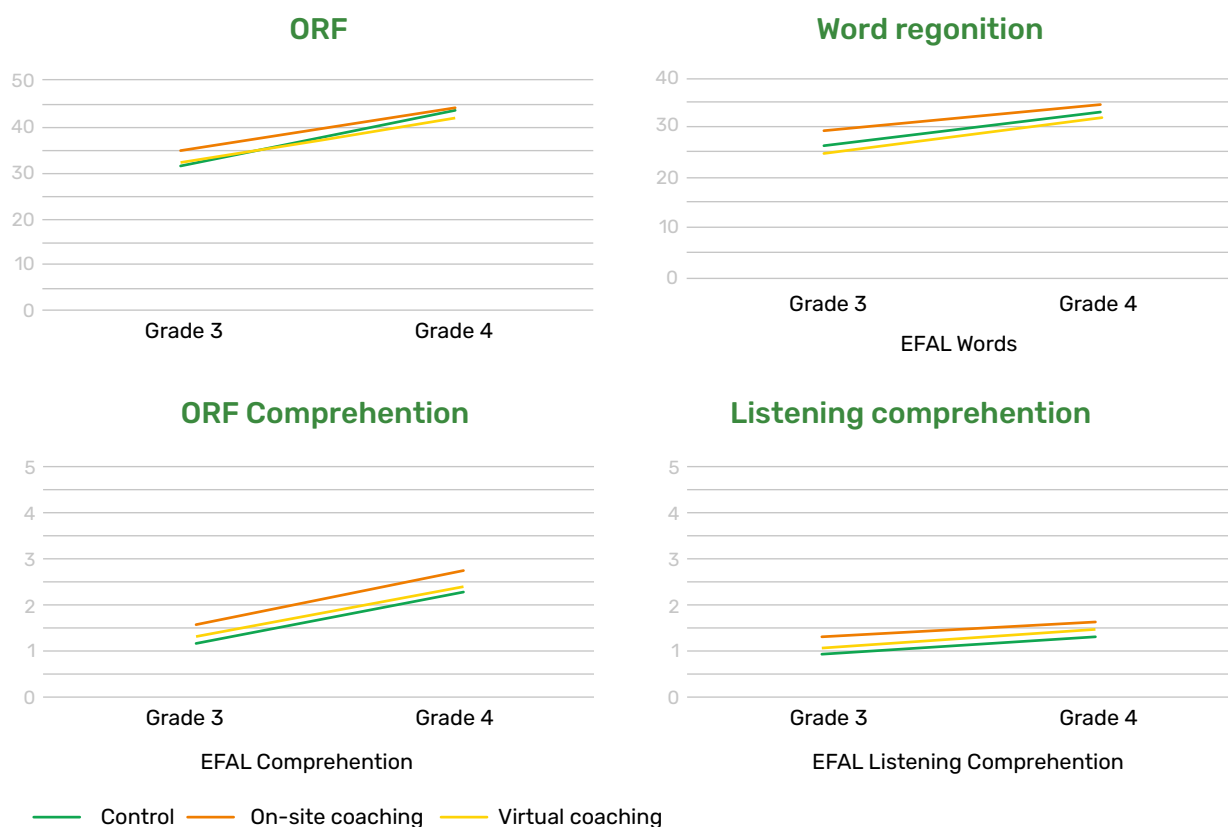
## 5.2 Comparing the impact of the interventions between Grade 3 and Grade 4

One of the key questions regarding the sustainability of an intervention is whether the initial impact that was experienced due to the interventions was sustained one year later. Figure 11 shows the mean scores for the different sub-tasks, controlling for the baseline scores and the other basic controls. These graphs illustrate how the learning gains that were made at the end of Grade 3 were sustained through to the end of Grade 4. These results should be interpreted from the perspective that schooling was severely impacted by the COVID-pandemic and that all the Grade 4 learners were exposed to very little learning.

In the EFAL word recognition task, the learners in the control group read 6.5 more words correctly within a minute in Grade 4 than in Grade 3. For the same task, the on-site coaching group gained 5.8 wcpm during Grade 4 and the virtual coaching group gained 6.8 wcpm. For the ORF task, the control group learners read 12 more words correctly in Grade 4 than in Grade 3, compared to 9.3 more words read correctly by learners in the on-site coaching group and 9.8 more words by learners in the virtual coaching group. In interpreting the lower magnitude of learning gains in the wcpm by the learners in the on-site coaching group it is important to notice that they started from a higher base in Grade 3 and performed at the same level as the control learners at the end of Grade 4. This suggests that the control and virtual coaching group learners may have caught up on the advantage that the on-site coaching learners experienced due to the interventions.

Similarly, we can track the learning gains in the English listening comprehension task, as well as the reading comprehension task based on the ORF passage. In both tasks, we saw some learning gains over the year, but these were similar across the different groups with no evidence of convergence.

**Figure 11: Learning gains in oral reading fluency and word recognition between Grade 3 and Grade 4**

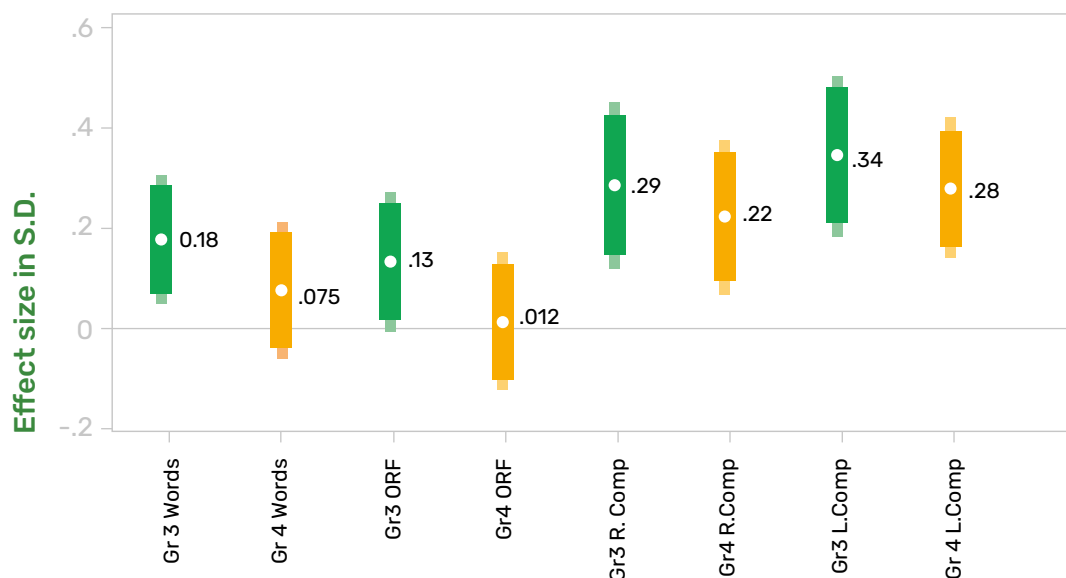


Notes: Sample excludes all repeaters. The statistics for the full sample is in Table A10 in the Appendix.

The figure below compares the effect sizes of the on-site intervention in Grade 3 (directly after learners were exposed to three years of learning) with the sustained impact in Grade 4 (one year after the interventions have stopped). Since the virtual coaching intervention did not show any significant impacts at the end of Grade 3, there is very little reason to expect that they will improve one year later. Nevertheless, we double-checked whether any impact of virtual coaching was experienced in Grade 4 and did not find evidence of such (see Figure A4 in the appendix).

In Grade 3 we already noted that the interventions have not been specifically effective in terms of improving reading fluency skills. For the on-site coaching group, we note that any impacts that were seen at the end of Grade 3 have disappeared by the end of Grade 4. A larger impact was seen on the reading and listening comprehension tasks and it is evident that these impacts were also more likely to have been sustained, with the effect size only decreasing 24% and 18% for reading and listening comprehension respectively. This seems to suggest that the on-site coaching intervention may have managed to prepare learners better for reading for meaning in English in Grade 4.

**Figure 12: Comparing the impact of the on-site coaching intervention in Grade 3 and Grade 4 on EFAL outcomes**



Notes: Sample excludes all repeaters. The statistics for the full sample is in Table A11 in the Appendix.

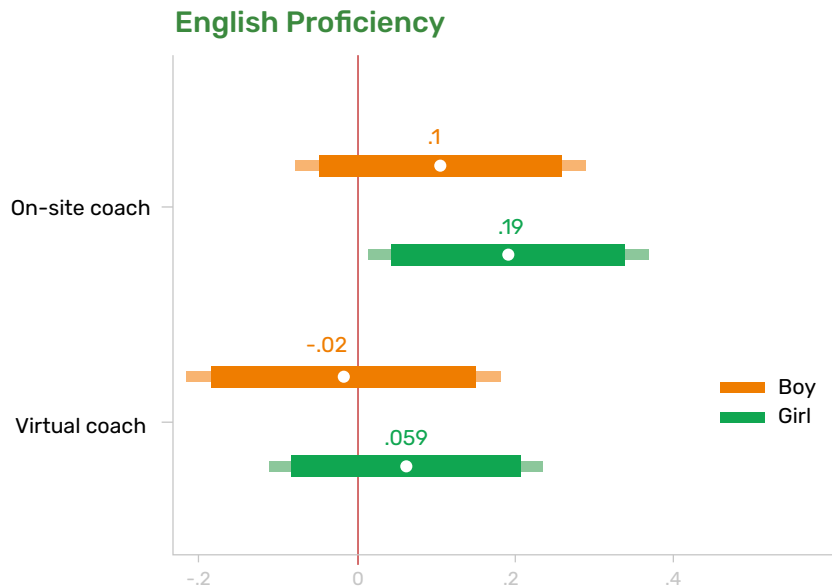
## 5.3 Sub-Group Analysis

In our pre-analysis plan for waves 1 – 4, we specified that we would evaluate whether the interventions impacted learners differently based on four different characteristics, namely the districts where the programme is implemented, learner gender, learner ability and the home language of the learner. We will evaluate the same for the wave 5 results. No noteworthy sub-group effects were evident based on the district in which the interventions were implemented or the home language of the learner. The effects seen based on gender and learner ability are discussed below.

In South Africa, we find that girls outperform boys in reading outcomes from a very young age. The baseline assessment confirmed this and showed that girls performed better than the boys in tasks such as non-word recall, letter recognition and phoneme isolation. An interesting finding from the first Early Grade Reading Study was that the interventions seemed to help boys catch up some of the way to girls. However, in the EGRS II we saw girls outperforming boys in both English reading proficiency and English oral language proficiency, and they also

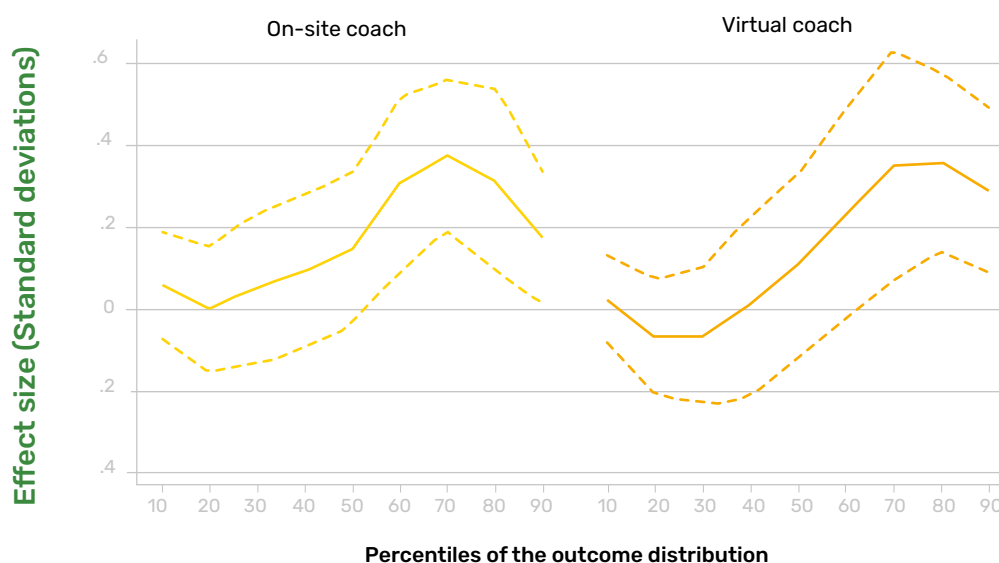
benefitted more from the intervention than boys (although this result was not statistically significant). **Figure 13** below shows the estimated effects of each intervention, by gender, at the end of Grade 4. The same held true at the end of Grade 3.

**Figure 13: Intervention effects for boys and girls**



The second heterogeneous effect that we want to consider is whether the interventions are supporting learning improvements among the bottom, middle or top-performing learners. Evaluating any differential impact based on baseline ability is, however, problematic because the correlation between the Wave 1 and Wave 5 scores are particularly weak. For this reason, we look at the impact of the interventions across the learning distribution. **Figure 14** shows that the on-site coaching intervention only sustainably benefitted learners in the top half of the distribution, whereas the virtual coaching intervention only had sustained benefits for learners in the top 20% of the learning distribution.

**Figure 14: Impact of the interventions across the learning distribution**





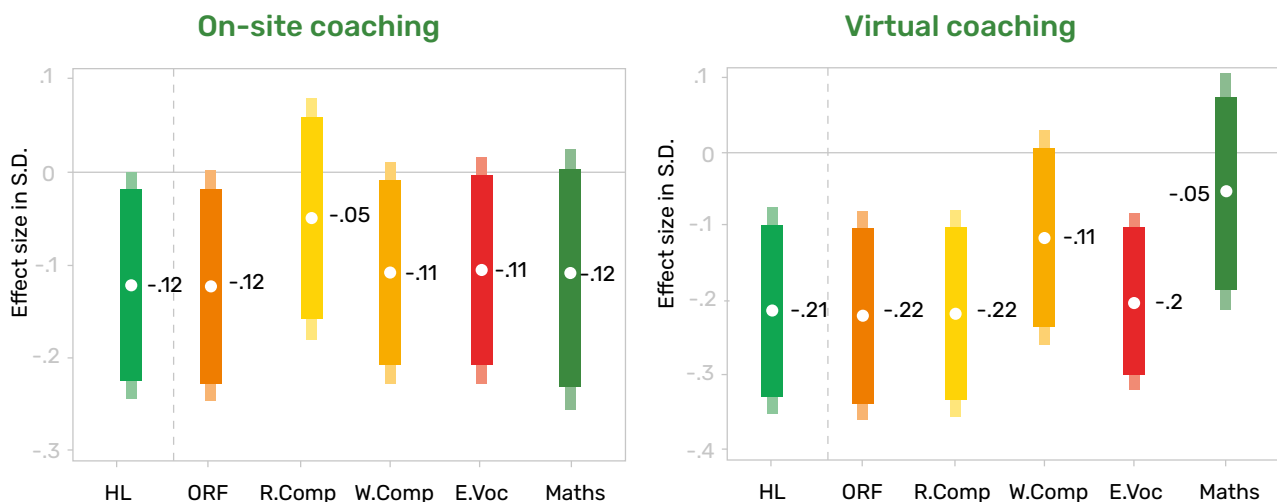
## 6. SECONDARY OUTCOMES

The South African curriculum specifies that four subjects should be taught during the Foundation Phase: HL, EFAL, Mathematics and Life Skills.

The interventions only supported teachers with the teaching of EFAL, but there are two ways in which the teaching of the other subjects could have been influenced by the interventions: teachers could either have dedicated more time and effort to the teaching of EFAL, at the cost of time and effort spent on teaching the other subjects (crowding-out), or teachers could have applied the more effective teaching methodologies that they were taught in the EFAL intervention to the teaching of the other subjects (positive spill-over). At the end of Grade 3, we noted that there were negative spillover effects on the HL outcomes and that this may have been driven by an increased focus on the teaching of EFAL at the cost of teaching HL. To evaluate whether this negative impact was sustained one year later, the learners were assessed on a few HL reading tasks again at the end of Grade 4. We also included the same short mathematics task as in Grade 3.

The HL tasks included an ORF passage followed by eight comprehension questions, a written comprehension assessment and an expressive vocabulary task. Figure 15 below shows that overall the negative spillover effects of the intervention on HL were sustained one year after the interventions concluded and that it was more pronounced for the virtual coaching intervention than the on-site coaching intervention.

**Figure 15: The impact of on-site coaching and virtual coaching on learner HL outcomes**



Notes: Sample excludes all repeaters. The regression results that both includes and excludes repeaters are in tables A12 and A13 in the Appendix.

**Table 13** shows the magnitude of the negative spillover of the interventions on the HL outcomes. In the HL ORF passage, learners in the on-site coaching group read on average 2 words fewer correctly in a minute and learners in the virtual coaching group read on average 4 words fewer correctly in a minute. Although not statistically significant, it seems that learners in the intervention groups also did poorer on the mathematics task. However, this negative effect was not observed at the end of Grade 3 and should be interpreted with caution.

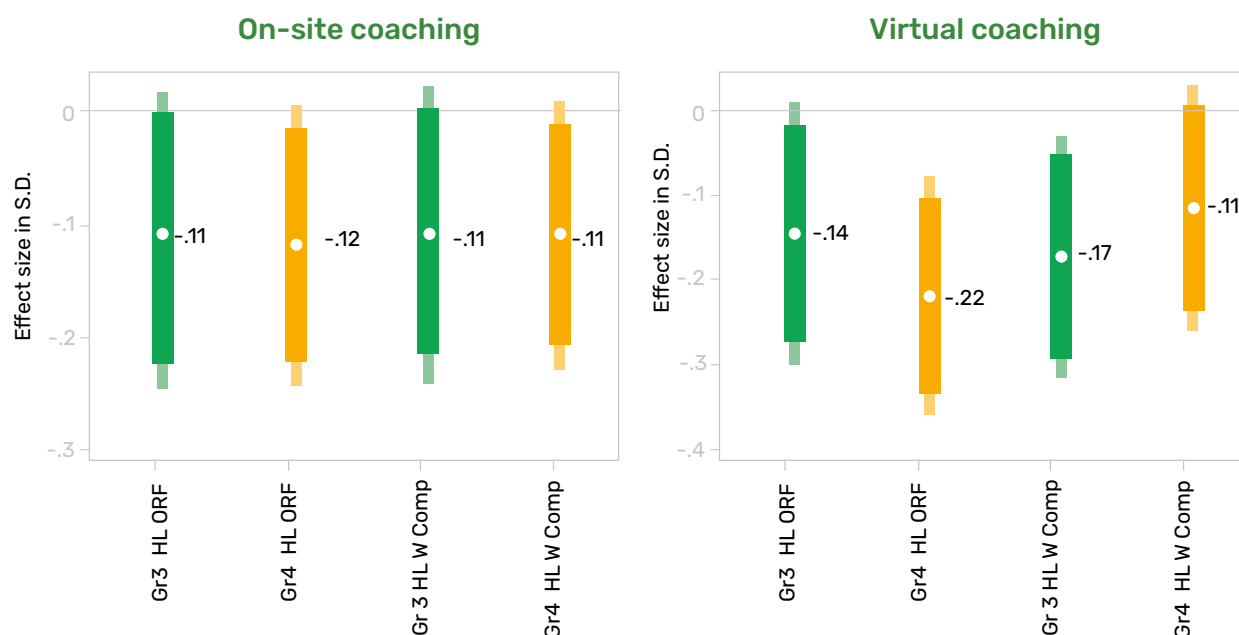
**Table 13: EGRA results by treatment group**


|               | HL ORF 60s | HL ORF Compr. | HL W. Compr. | HL Recep Vocab. | Maths |
|---------------|------------|---------------|--------------|-----------------|-------|
| Control mean  | 25.09      | 3.59          | 2.69         | 2.73            | 2.88  |
| On-site coach | 22.95*     | 3.46          | 2.46*        | 2.49*           | 2.65  |
| Virtual coach | 21.10***   | 3.04***       | 2.46         | 2.28***         | 2.77  |

\*\*\*, \*\*, and \* indicate significance at the 1, 5, and 10 percent critical level. Showing results for raw scores. Controlling for learner gender, learner age, baseline scores, district, school quintile, fieldworker and stratification dummies. Only includes Grade 4 learners.

Comparing the impacts of the interventions between Grade 3 and Grade 4, it is clear that the negative spillover effect that was seen at the end of Grade 3 was sustained in both interventions. The magnitude of the negative spillover effect remained very similar for the on-site coaching group. In the virtual coaching group, it seemed as if the learners may have fallen further behind on the HL ORF task, but may have recovered slightly in the HL written task.

**Figure 16: Comparing the impact of the on-site and virtual coaching intervention in Grade 3 and Grade 4 on HL outcomes**





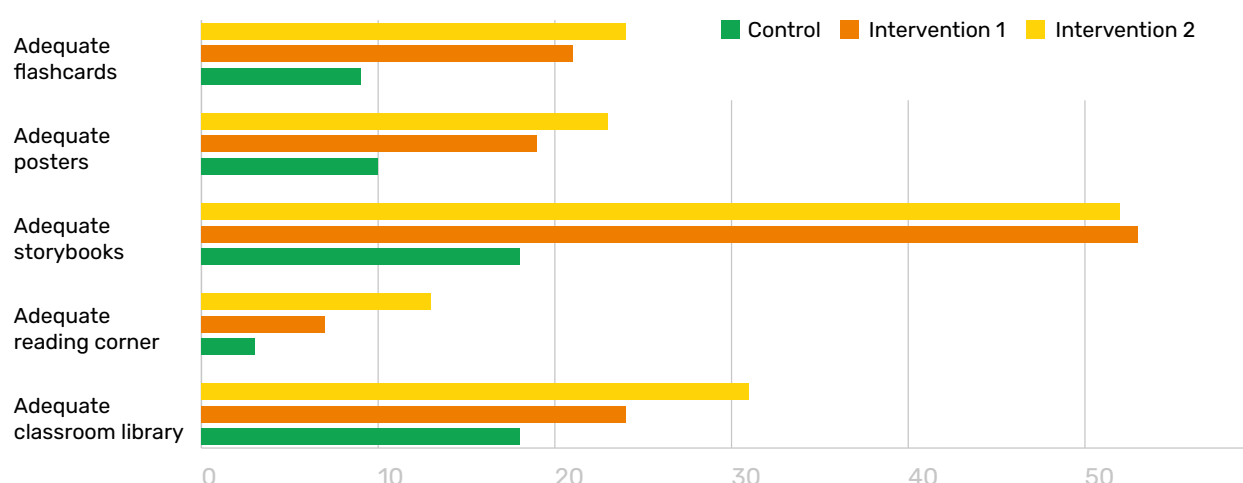
# 7. SUSTAINED TEACHER PRACTICES

Given the school closures and rotational teaching time in 2020, it is quite challenging to evaluate whether teachers sustained any of the intervention practices. The lesson plans were based on the premise of a weekly routine and the coverage of the full curriculum. In response to the significant loss in teaching time, the DBE developed revised Annual Teaching Plans based on a trimmed curriculum and the lesson plans that were used in 2017-2019 would not have been suited to the trimmed curriculum.

Similarly, the rotational teaching time has meant that the weekly routine of activities was not suitable for the 2020 year. Nevertheless, interviews were conducted with the Grade 2 and Grade 3 teachers to see whether there is any evidence of the sustainability of the interventions a year (Grade 3 teachers) or two years (Grade 2 teachers) after the training and support that they received.

Learning and teaching support materials are the simplest aspect of the interventions to evaluate from a sustainability lens. Regardless of the teaching timetable, one would expect to still see evidence of the materials provided in the classrooms, and hopefully the use of the materials as well. Figure 17 shows the proportion of the classrooms in which the different resources provided through the interventions were deemed adequate. To measure adequacy, a grid with descriptions was used by fieldworkers to rate the quantity and quality of the resource. A rating of three and higher was considered 'adequate'. Classrooms of teachers who received the intervention were much more likely to have adequate flashcards and posters against the walls, as well as to have eleven or more storybooks in their classroom. These are all resources that were provided through the intervention and their sustained presence in the classrooms are encouraging. Interestingly, it is only the classrooms of the teachers in intervention 2 (virtual coaching) that were statistically significantly more likely to have a reading corner or a reading mat, than the control group teachers. Organising the classroom to have a dedicated space for reading suggests a changed behaviour, rather than merely having a resource still available in the classroom.

**Figure 17: Proportion of classes in which print-richness was classified as adequate**

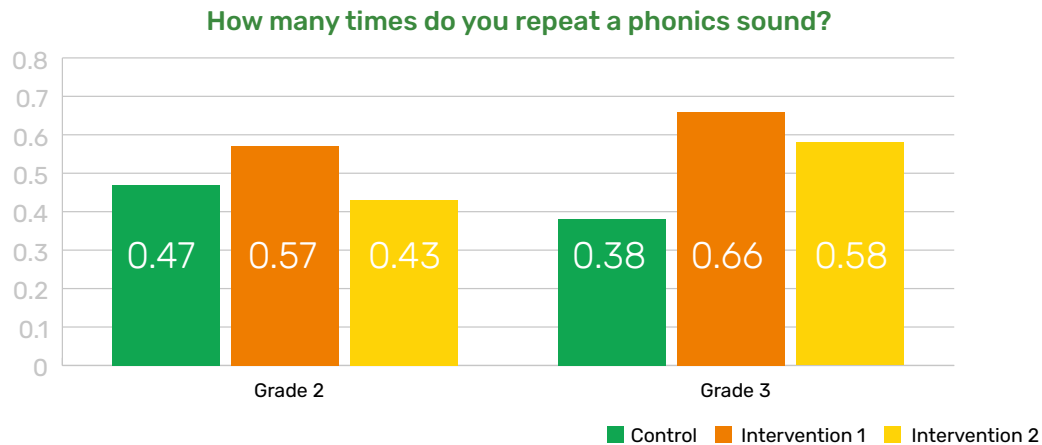


One of the specific elements of the EGRS II intervention was the provision of small graded readers, to allow learners access to reading material appropriate to their level of reading ability. Figures 18a-c below indicate that the classrooms in schools that participated in the interventions were more likely to have EFAL graded readers and to have a sufficiency number of EFAL graded readers (more than 20). There is also no difference between the Grade 2 and Grade 3 teachers' responses, indicating that 'fade-out' due to the loss of the booklets is not a major problem two years after the interventions have stopped. However, teachers across all three groups are equally likely to use the EFAL graded readers, if they had graded readers in their class. This suggests that at a very basic level, merely the provision of the graded readers may have led to increased opportunities for reading in the classroom.

## Figure 18: Access to and use of EFAL graded readers

The only measure that we have of correct teaching practices was asking teachers the number of times they repeat a phonics sound when teaching a new sound. The intervention train teachers that this should only be three times. Figure 19 shows that only the Grade 3 intervention 1 teachers were statistically significantly more likely to respond correctly. The difference in the treatment groups between Grade 2 and Grade 3 also suggests that there may have been some fade-out of this practice over the years.

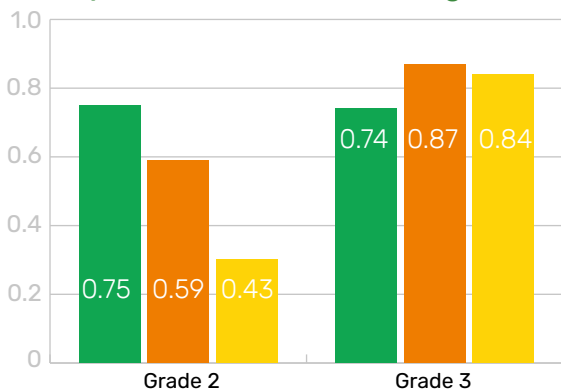


**Figure 19: Correct teaching of phonics sounds**

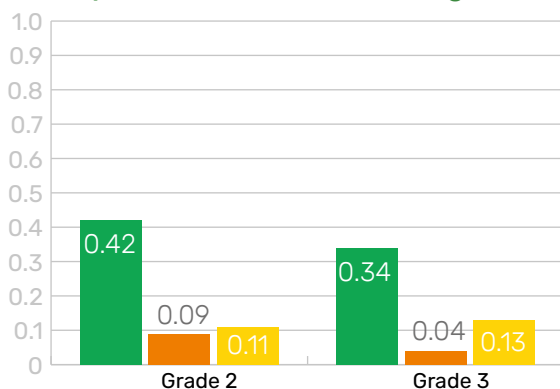
Teachers were also asked about the EFAL and HL training opportunities that they have had in 2019 and 2020. Some interesting things to note from figures 19a-d: Firstly, given the attendance records from the training sessions, one would expect all of the Grade 3 teachers in interventions 1 and 2 to have responded that they received EFAL training in 2019. Secondly, as expected, all teachers were much less likely to have attended either EFAL or HL training in 2020. Thirdly, the intervention 1 and 2 teachers were more likely than the control group teachers to respond that they did not receive training either HL or EFAL the years after they participated in the interventions (Grade 2 teachers in 2019 and 2020, and the Grade 3 teachers in 2020). The difference in responses, however, are only statistically significant for EFAL training. Cilliers et al noted a similar response trend in the first Early Grade Reading Study in the North West province and theorised that this may be due to teachers having adjusted their expectations on what constitutes training after having participated in the EGRS interventions. The same may likely be true here, especially given the response rates that we see regarding training received in HL.

**Figure 20: Continuous professional development in 2019 and 2020**

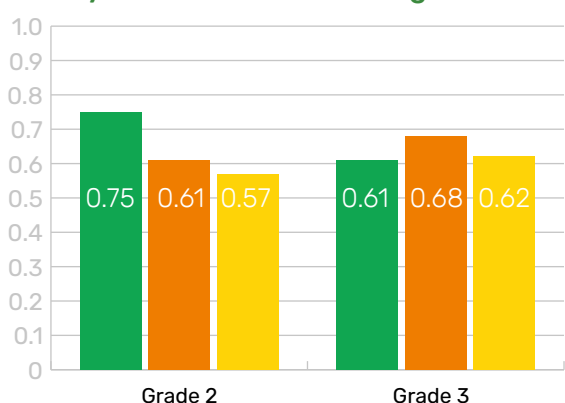
Have you received EFAL training in 2019?



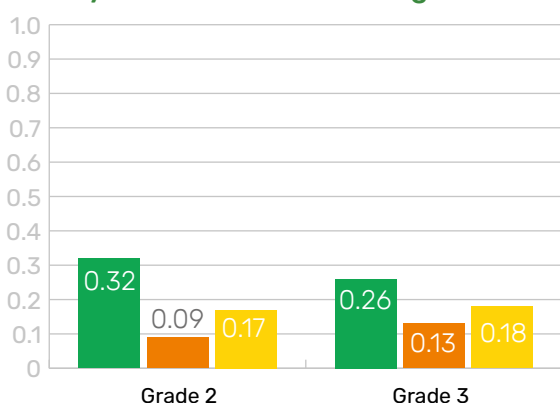
Have you received EFAL training in 2020?



Have you received HL training in 2019?

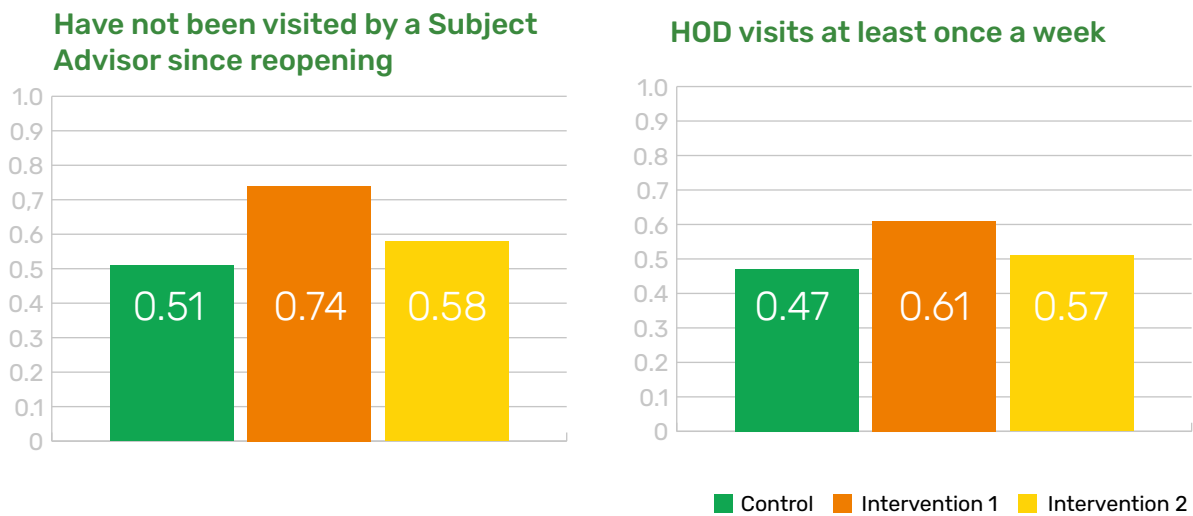


Have you received HL training in 2020?



Finally, we consider the support that teachers have received since the reopening of the schools in August 2020. Teachers were asked how frequently they have received support from their Subject Advisors (from the District Office), their principals, their Heads of Departments (HODs) or from an external coach or mentor. No differences were reported regarding the support received from their principals or external coaches, however, teachers in the Intervention 1 group were more likely to respond that they have not been visited by a Subject Advisor since the reopening of schools. They were also more likely to respond that their HOD has visited them at least once a week since the reopening of schools.

**Figure 21: On-site support from a Subject Advisor or Head of Department**



# CONCLUSION



This study compared the effectiveness of a structured pedagogy programme that was implemented through two different delivery models. The first was through providing teachers with paper-based lesson plans and support by an on-site coach. The second was through providing teachers with lesson plans on a tablet and support by a virtual coach. The main research question of the study was whether an alternative form of coaching could work, and this report focussed specifically on whether the impact could be sustained.

From a sustainability perspective, we would expect that the stronger foundations laid during Grades 1 – 3 will translate into great reading fluency and English comprehension by the end of Grade 4. We therefore considered the following three questions:

## 1. Is the impact on reading outcomes due to the on-site coaching sustained one year after implementation?

Yes, a year after the learners were exposed to the on-site coaching intervention, they still had a positive and significant advantage over their control group peers in the ORF reading comprehension, the listening comprehension, as well as in the productive and receptive English vocabulary tasks. The advantage in the decoding skills as well as in the written vocabulary tasks was, however, negligible.

## 2. Is the impact on reading outcomes due to the virtual coaching sustained one year after implementation?

No. At the end of Grade 3, the impacts of the virtual coaching intervention were already very small and

one year after the intervention concluded, there is no evidence of a sustained impact of the virtual coaching intervention.

## 3. Which model is the most cost-effective, factoring in sustainability?

Given that the impacts of the virtual coaching intervention were not sustained, the on-site coaching intervention is the only intervention that showed effectiveness one year after implementation. There was, however, some evidence of fade-out of the impacts on learning outcomes, with the effect size on the reading and listening comprehension tasks a decreasing by 24% and 18% respectively. This seems to suggest that the on-site coaching intervention may have managed to prepare learners better for reading for meaning in English in Grade 4.

It is not possible to draw any strong conclusions on teacher sustainability, given the complexity of measuring sustained teacher practices during 2020. The evidence on the sustained accessibility and use of the learning and teaching resource provided is, however, encouraging.

Evidence of the effectiveness of on-site coaching has been building over the past couple of years and this study further contributes to this evidence base.

# REFERENCES

- Ardington, C., Wills, G. and Kotze, J., 2021. COVID-19 learning losses: Early grade reading in South Africa. *International Journal of Educational Development*, 86, p.102480.
- Cilliers, J., Fleisch, B., Kotze, J., & Mohohlwane, M., 2019. *The sustainability of early grade education interventions: Do learning gains and improved teacher practices persist?* Gui2de Working Paper Series No. 6. Washington, DC: Georgetown University Initiative on Innovation, Development and Evaluation Department of Economics and the McCourt School of Public Policy.
- Cilliers, J., Fleisch, B., Prinsloo, C., & Taylor, S., 2020. How to improve teaching practice? An experimental comparison of centralized training and in-classroom coaching. *Journal of Human Resources*, 55(3), pp. 926–962.
- DBE. (2011). *Curriculum and Assessment Policy Statement (CAPS) Grades 1–3: English First Additional Language*. Pretoria: DBE.
- Fleisch, B., 2018. *The education triple cocktail: System-wide instructional reform in South Africa*. Cape Town: University of Cape Town Press & Juta.
- Kotze, J., Fleisch, B., & Taylor, S., 2019. Alternative forms of early grade instructional coaching: Emerging evidence from field experiments in South Africa. *International Journal of Educational Development*, 66, pp. 203–213.

# APPENDIX

**Table A1: Item descriptive statistics – full sample**

|                   | N    | Mean | s.d. | Min. | Max. | p10  | p25  | p50  | p75 | p90 | % zero score |
|-------------------|------|------|------|------|------|------|------|------|-----|-----|--------------|
| HL ORF 60s        | 2405 | 23.7 | 18.3 | 0    | 62   | 0    | 4    | 25   | 38  | 49  | 22%          |
| HL ORF 180s       | 2405 | 41.1 | 25.3 | 0    | 62   | 0    | 10   | 57   | 60  | 61  | 22%          |
| HL ORF Compr.     | 2405 | 3.4  | 2.6  | 0    | 8    | 0    | 0    | 4    | 6   | 7   | 26%          |
| EFAL Word Recog.  | 2390 | 27.8 | 23.5 | 0    | 105  | 0    | 2    | 26   | 47  | 60  | 20%          |
| EFAL ORF 60s      | 2405 | 36.2 | 34.1 | 0    | 126  | 0    | 1    | 30   | 64  | 83  | 24%          |
| EFAL ORF 180s     | 2405 | 70.2 | 52.1 | 0    | 126  | 0    | 1    | 98   | 117 | 123 | 23%          |
| EFAL ORF Compr.   | 2405 | 1.9  | 2.3  | 0    | 8    | 0    | 0    | 1    | 3   | 6   | 43%          |
| EFAL Prod. Vocab. | 2405 | 17.1 | 7.4  | 0    | 30   | 5    | 12   | 19   | 23  | 26  | 1%           |
| EFAL L. Compr.    | 2405 | 1.3  | 1.2  | 0    | 4    | 0    | 0    | 1    | 2   | 3   | 28%          |
| HL W. Compr.      | 2352 | 2.6  | 2.1  | 0    | 6    | 0    | 0    | 3    | 4   | 5   | 28%          |
| HL Recep Vocab.   | 2352 | 2.6  | 2.3  | 0    | 10   | 0    | 1    | 2    | 4   | 6   | 21%          |
| EFAL W. Compr.    | 2352 | 1.3  | 1.6  | 0    | 7    | 0    | 0    | 1    | 2   | 4   | 46%          |
| EFAL Recep Vocab. | 2352 | 3.8  | 2.5  | 0    | 8    | 0    | 2    | 4    | 6   | 7   | 11%          |
| EFAL W. Vocab.    | 2352 | 1.8  | 1.8  | 0    | 8    | 0    | 0    | 1    | 3   | 4   | 31%          |
| Maths             | 2352 | 2.8  | 2    | 0    | 10   | 1    | 1    | 2    | 4   | 6   | 9%           |
| EFAL Prof.        | 2331 | 0.1  | 2.4  | -3.5 | 7.1  | -2.7 | -1.9 | -0.4 | 1.8 | 3.9 |              |
| HL Prof.          | 2346 | -0.1 | 1.8  | -2.7 | 4    | -2.7 | -2   | 0.1  | 1.3 | 2.1 |              |

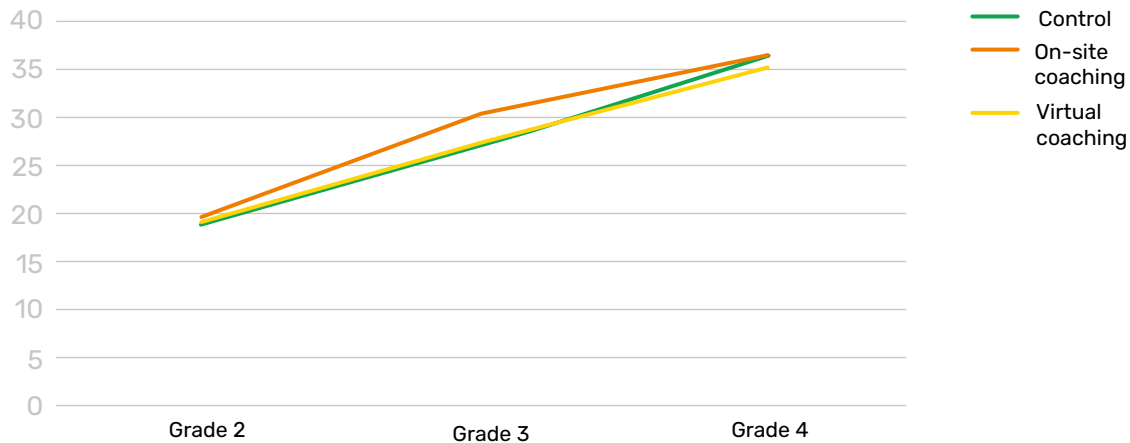
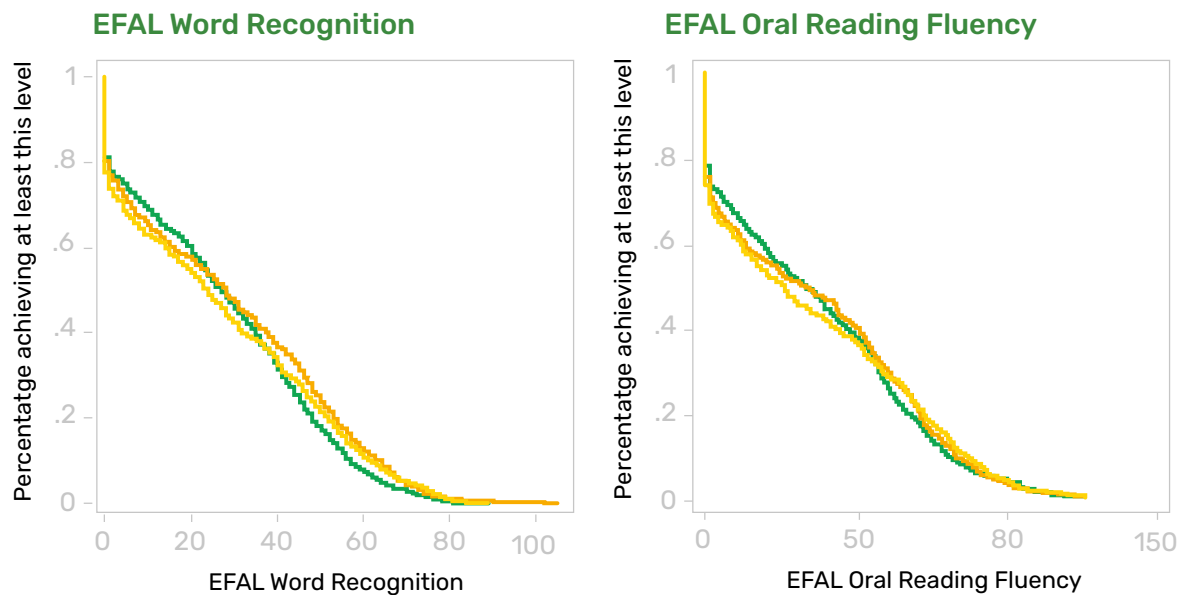
**Table A2: Tasks means in Wave 4, by intervention group – full sample**

|                   | Control (1)       | On-site Coaching (2) | Virtual Coaching (3) | Difference (1)-(2) | Difference (1)-(3) |
|-------------------|-------------------|----------------------|----------------------|--------------------|--------------------|
| HL ORF 60s        | 25.093<br>[0.880] | 23.115<br>[0.811]    | 22.161<br>[1.224]    | 1.978*             | 2.932*             |
| HL ORF 180s       | 43.050<br>[1.177] | 40.164<br>[1.246]    | 38.964<br>[1.734]    | 2.886*             | 4.086*             |
| EFAL Word Recog.  | 27.535<br>[1.020] | 29.039<br>[1.121]    | 27.063<br>[1.819]    | -1.503             | 0.473              |
| EFAL ORF 60s      | 36.480<br>[1.475] | 36.658<br>[1.548]    | 35.351<br>[2.604]    | -0.179             | 1.129              |
| EFAL ORF 180s     | 72.430<br>[2.279] | 69.670<br>[2.345]    | 67.129<br>[3.640]    | 2.760              | 5.302              |
| HL ORF Compr.     | 3.585<br>[0.125]  | 3.358<br>[0.112]     | 3.243<br>[0.164]     | 0.227              | 0.342*             |
| EFAL ORF Compr.   | 1.841<br>[0.099]  | 2.083<br>[0.110]     | 1.967<br>[0.178]     | -0.241             | -0.126             |
| EFAL Prod. Vocab. | 16.829<br>[0.354] | 17.538<br>[0.402]    | 17.142<br>[0.567]    | -0.709             | -0.313             |
| EFAL L. Compr.    | 1.199<br>[0.050]  | 1.404<br>[0.066]     | 1.365<br>[0.087]     | -0.204**           | -0.166*            |
| HL W. Compr.      | 2.686<br>[0.093]  | 2.479<br>[0.095]     | 2.468<br>[0.134]     | 0.207              | 0.219              |
| HL Recep Vocab.   | 2.732<br>[0.103]  | 2.548<br>[0.099]     | 2.265<br>[0.115]     | 0.184              | 0.466***           |
| EFAL W. Compr.    | 1.287<br>[0.067]  | 1.336<br>[0.085]     | 1.337<br>[0.108]     | -0.050             | -0.051             |
| EFAL Recep Vocab. | 3.698<br>[0.096]  | 3.986<br>[0.131]     | 3.857<br>[0.163]     | -0.288*            | -0.159             |
| EFAL W. Vocab.    | 1.766<br>[0.076]  | 1.787<br>[0.086]     | 1.756<br>[0.114]     | -0.021             | 0.010              |
| Maths             | 2.878<br>[0.090]  | 2.623<br>[0.106]     | 2.764<br>[0.140]     | 0.256*             | 0.114              |
| EFAL Prof.        | -0.000<br>[0.108] | 0.248<br>[0.128]     | 0.086<br>[0.200]     | -0.248             | -0.086             |
| HL Prof.          | 0.000<br>[0.086]  | -0.180<br>[0.082]    | -0.309<br>[0.114]    | 0.180              | 0.309**            |

The value displayed for t-tests are the differences in the means across the groups. Standard errors are clustered at variable NatEmis. \*\*\*, \*\*, and \* indicate significance at the 1, 5, and 10 percent critical level.

**Table A3: Comparing EFAL word recognition across the waves of data collection – full sample**

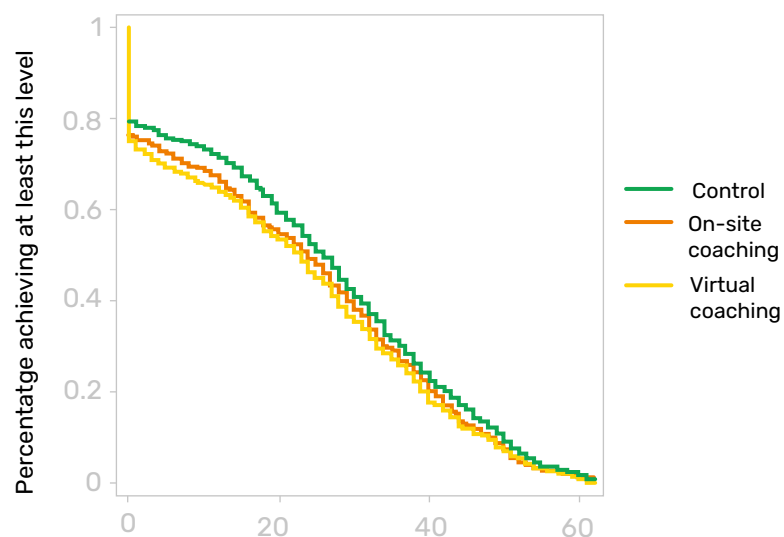
|                  | End of Grade 1        |                   | End of Grade 2        |                   | End of Grade 3                   | End of Grade 4 | End of Grade 4 |
|------------------|-----------------------|-------------------|-----------------------|-------------------|----------------------------------|----------------|----------------|
| incl repeaters   | Decodable words (30s) | Sight words (30s) | Decodable words (30s) | Sight words (30s) | Decodable + Sight Combined (60s) |                | % zero         |
| Control          | 2.5                   | 2.7               | 17.0                  | 16.5              | 23.1                             | 27.5           | 19%            |
| On-site coaching | 2.7                   | 2.8               | 17.8                  | 17.6              | 25.9                             | 29.0           | 20%            |
| Virtual coaching | 2.3                   | 2.3               | 16.5                  | 16.3              | 22.7                             | 27.1           | 22%            |

**Figure A1: Improvements in oral reading fluency between Grade 2 and Grade 4 – full sample****Figure A2: Performance distribution for EFAL word recognition and in oral reading fluency – full sample****Figure A4: EFAL learner performance by intervention group – full sample**

|                  | Reading Comp | Written Comp | Listening Comp | Expressive Vocab | Receptive Vocab | Written Vocab |
|------------------|--------------|--------------|----------------|------------------|-----------------|---------------|
|                  | (Out of 8)   | (Out of 7)   | (Out of 4)     | (Out of 30)      | (Out of 8)      | (Out of 8)    |
| Control          | 1.84         | 1.29         | 1.20           | 16.83            | 3.70            | 1.77          |
| On-site coaching | 2.08         | 1.34         | 1.40           | 17.54            | 3.99            | 1.79          |
| Virtual coaching | 1.97         | 1.34         | 1.37           | 17.14            | 3.86            | 1.76          |

**Table A5: Comparing HL word reading – full sample**

|                  | Grade 2 | Grade 3 | Grade 4 | % zero |
|------------------|---------|---------|---------|--------|
| Control          | 15.7    | 23.1    | 25.1    | 20%    |
| On-site coaching | 13.8    | 21.5    | 23.1    | 24%    |
| Virtual coaching | 13.5    | 20.7    | 22.2    | 25%    |

**Figure A3: Distribution of HL oral reading fluency – full sample****HL Oral Reading Fluency****Table A6: HL reading outcomes by intervention group**

|                  | Reading Comprehension | Written Comprehension | Receptive vocabulary |
|------------------|-----------------------|-----------------------|----------------------|
|                  | (Out of 8)            | (Out of 7)            | (Out of 10)          |
| Control          | 3.58                  | 2.69                  | 2.73                 |
| On-site coaching | 3.36                  | 2.48                  | 2.55                 |
| Virtual coaching | 3.24                  | 2.47                  | 2.27                 |

**Table A7: Regression results on the EFAL learning outcomes – Grade 4 sample**

|               | (1) EFAL Prof.     | (2) EFAL Word Recog. | (3) EFAL ORF 60s  | (4) EFAL ORF Compr. | (5) EFAL Express. Vocab. | (6) EFAL L. Compr.  | (7) EFAL W. Compr. | (8) EFAL Recep Vocab. | (9) EFAL W. Vocab. |
|---------------|--------------------|----------------------|-------------------|---------------------|--------------------------|---------------------|--------------------|-----------------------|--------------------|
| On-site coach | 0.143**<br>(0.072) | 0.075<br>(0.069)     | 0.012<br>(0.069)  | 0.221***<br>(0.077) | 0.156**<br>(0.063)       | 0.280***<br>(0.070) | 0.022<br>(0.073)   | 0.158**<br>(0.064)    | 0.041<br>(0.065)   |
| Virtual coach | 0.010<br>(0.073)   | -0.053<br>(0.073)    | -0.058<br>(0.069) | 0.050<br>(0.070)    | -0.045<br>(0.068)        | 0.109<br>(0.072)    | 0.019<br>(0.073)   | 0.058<br>(0.069)      | -0.024<br>(0.062)  |
| Observations  | 1,795              | 1,811                | 1,825             | 1,825               | 1,825                    | 1,825               | 1,809              | 1,809                 | 1,809              |
| R-squared     | 0.277              | 0.236                | 0.235             | 0.247               | 0.256                    | 0.235               | 0.164              | 0.182                 | 0.174              |
| P-value       | 0.103              | 0.116                | 0.375             | 0.0457              | 0.00491                  | 0.0282              | 0.970              | 0.178                 | 0.358              |
| Control mean  | 0.233              | 0.247                | 0.226             | 0.189               | 0.236                    | 0.123               | 0.160              | 0.172                 | 0.152              |

Robust standard errors in parentheses. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ . Controlling for learner gender, learner age, baseline scores, fieldworker and stratification dummies. Only includes Grade 4 learners.

**Table A8: Regression on the English learning outcomes – full sample**

|               | (1) EFAL Prof.    | (2) EFAL Word Recog. | (3) EFAL ORF 60s   | (4) EFAL ORF Compr. | (5) EFAL Express. Vocab. | (6) EFAL L. Compr.  | (7) EFAL W. Compr. | (8) EFAL Recep Vocab. | (9) EFAL W. Vocab. |
|---------------|-------------------|----------------------|--------------------|---------------------|--------------------------|---------------------|--------------------|-----------------------|--------------------|
| On-site coach | 0.097<br>(0.065)  | 0.012<br>(0.065)     | -0.037<br>(0.063)  | 0.154**<br>(0.065)  | 0.123*<br>(0.067)        | 0.225***<br>(0.064) | 0.016<br>(0.065)   | 0.080<br>(0.059)      | 0.010<br>(0.054)   |
| Virtual coach | -0.047<br>(0.066) | -0.107<br>(0.066)    | -0.116*<br>(0.062) | 0.001<br>(0.062)    | -0.049<br>(0.067)        | 0.044<br>(0.063)    | -0.054<br>(0.063)  | -0.018<br>(0.060)     | -0.054<br>(0.053)  |
| Observations  | 2,331             | 2,375                | 2,390              | 2,390               | 2,390                    | 2,390               | 2,351              | 2,351                 | 2,351              |
| R-squared     | 0.314             | 0.279                | 0.275              | 0.262               | 0.284                    | 0.251               | 0.184              | 0.215                 | 0.196              |
| P-value       | 0.05              | 0.11                 | 0.25               | 0.03                | 0.02                     | 0.01                | 0.30               | 0.14                  | 0.27               |

Robust standard errors in parentheses; \*\*\* p<0.01, \*\* p<0.05, \* p<0.1; Controlling for learner gender, learner age, baseline scores, fieldworker and stratification dummies. Includes repeaters.

**Table A9: Raw scores by treatment group – full sample**

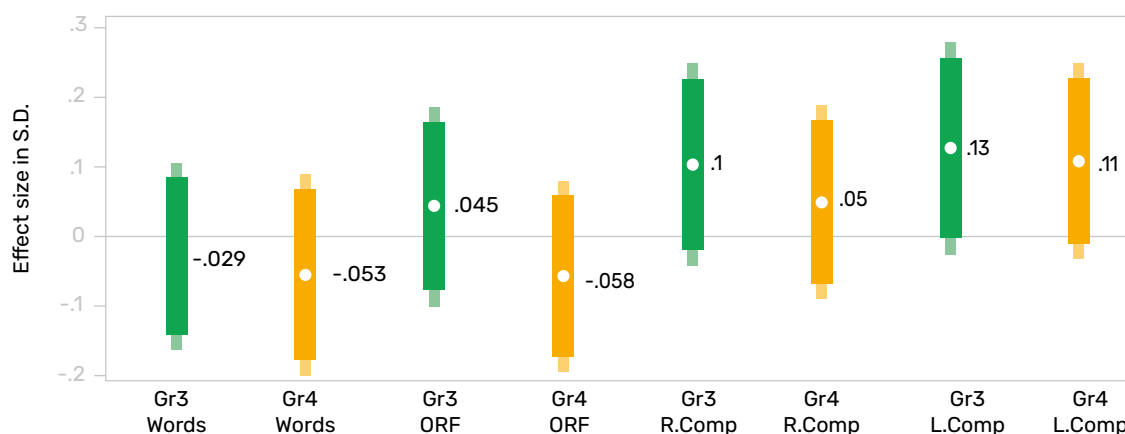
|               | Words correct per minute |              | Items correct   |                   |                |                |                   |                |
|---------------|--------------------------|--------------|-----------------|-------------------|----------------|----------------|-------------------|----------------|
|               | EFAL Word Recog.         | EFAL ORF 60s | EFAL ORF Compr. | EFAL Prod. Vocab. | EFAL L. Compr. | EFAL W. Compr. | EFAL Recep Vocab. | EFAL W. Vocab. |
| Control       | 27.5                     | 36.5         | 1.8**           | 16.8*             | 1.2***         | 1.3            | 3.7               | 1.8            |
| On-site coach | 27.8                     | 35.3*        | 2.2             | 17.7              | 1.5            | 1.3            | 3.9               | 1.8            |
| Virtual coach | 25.2                     | 32.7         | 1.8             | 16.5              | 1.3            | 1.2            | 3.7               | 1.7            |

\*\*\*, \*\*, and \* indicate significance at the 1, 5, and 10 percent critical level. Showing results for raw scores. Controlling for learner gender, learner age, baseline scores, district, school quintile, fieldworker and stratification dummies. Includes repeaters.

**Table A10: Learning gains between Grade 3 and Grade 4 – full sample**

|               | ORF     |         | Word recognition |         | Reading Comprehension |         | Listening Comprehension |         |
|---------------|---------|---------|------------------|---------|-----------------------|---------|-------------------------|---------|
|               | Grade 3 | Grade 4 | Grade 3          | Grade 4 | Grade 3               | Grade 4 | Grade 3                 | Grade 4 |
| Control       | 27.3    | 36.5    | 23.1             | 27.5    | 1.0                   | 1.8     | 0.9                     | 1.2     |
| On-site coach | 29.6    | 35.3    | 25.4             | 27.8    | 1.3**                 | 2.2***  | 1.2***                  | 1.5***  |
| Virtual coach | 26.5    | 32.7    | 21.0*            | 25.2    | 1.0                   | 1.8     | 1.0                     | 1.3     |

\*\*\*, \*\*, and \* indicate significance at the 1, 5, and 10 percent critical level. Showing results for raw scores. Controlling for learner gender, learner age, baseline scores, district, school quintile, fieldworker and stratification dummies. Includes repeaters.

**Figure A4: Comparing the impact of the virtual coaching intervention in Grade 3 and Grade 4 on EFAL outcomes**

**Table A11: Comparing the impact of the interventions in Grade 3 and Grade 4 on EFAL outcomes – Full sample**

|               | Word Recognition  |                   | ORF               |                    | Reading Comprehension |                    | Listening Comprehension |                     |
|---------------|-------------------|-------------------|-------------------|--------------------|-----------------------|--------------------|-------------------------|---------------------|
|               | Grade 3           | Grade 4           | Grade 3           | Grade 4            | Grade 3               | Grade 4            | Grade 3                 | Grade 4             |
| On-site coach | 0.109<br>(0.068)  | 0.012<br>(0.065)  | 0.084<br>(0.068)  | -0.037<br>(0.063)  | 0.260***<br>(0.071)   | 0.154**<br>(0.065) | 0.369***<br>(0.071)     | 0.225***<br>(0.064) |
| Virtual coach | -0.103<br>(0.066) | -0.107<br>(0.066) | -0.027<br>(0.065) | -0.116*<br>(0.062) | 0.052<br>(0.063)      | 0.001<br>(0.062)   | 0.096<br>(0.067)        | 0.044<br>(0.063)    |
| Observations  | 2,654             | 2,375             | 2,654             | 2,390              | 2,654                 | 2,390              | 2,654                   | 2,390               |
| R-squared     | 0.251             | 0.279             | 0.272             | 0.275              | 0.261                 | 0.262              | 0.208                   | 0.251               |
| P-value       | 0.00311           | 0.112             | 0.115             | 0.253              | 0.00466               | 0.0336             | 0.000384                | 0.00895             |
| Control mean  | 4.19e-09          | 9.69e-09          | -1.25e-08         | -1.38e-09          | 1.55e-08              | -4.39e-09          | 9.09e-09                | 2.49e-09            |

Robust standard errors in parentheses. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1. Showing results for Grade 3 and Grade 4. Controlling for learner gender, learner age, baseline scores, district, school quintile, fieldworker and stratification dummies. Includes repeaters

**Table A12: Regression results on the HL and Maths learning outcomes – Grade 4 sample**

|               | HL Prof.             | HL ORF 60s           | HL ORF Compr.        | HL W. Compr.       | HL Recep Vocab.      | Maths             |
|---------------|----------------------|----------------------|----------------------|--------------------|----------------------|-------------------|
| On-site coach | -0.122*<br>(0.062)   | -0.118*<br>(0.063)   | -0.050<br>(0.066)    | -0.109*<br>(0.060) | -0.106*<br>(0.061)   | -0.116<br>(0.071) |
| Virtual coach | -0.213***<br>(0.070) | -0.219***<br>(0.071) | -0.218***<br>(0.069) | -0.114<br>(0.072)  | -0.200***<br>(0.058) | -0.055<br>(0.079) |
| Observations  | 1,809                | 1,825                | 1,825                | 1,809              | 1,809                | 1,809             |
| R-squared     | 0.267                | 0.260                | 0.203                | 0.213              | 0.189                | 0.172             |
| P-value       | 0.199                | 0.173                | 0.0161               | 0.946              | 0.155                | 0.443             |
| Control mean  | 0.241                | 0.222                | 0.215                | 0.229              | 0.179                | 0.186             |

Robust standard errors in parentheses. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1. Controlling for learner gender, learner age, baseline scores, district, school quintile, fieldworker and stratification dummies. Only includes Grade 4 learners.

**Table A13: Regression results on the HL and Maths learning outcomes – Full sample**

|               | HL Prof.             | HL ORF 60s           | HL ORF Compr.        | HL W. Compr.        | HL Recep Vocab.      | Maths               |
|---------------|----------------------|----------------------|----------------------|---------------------|----------------------|---------------------|
| On-site coach | -0.144***<br>(0.055) | -0.160***<br>(0.057) | -0.102*<br>(0.061)   | -0.127**<br>(0.053) | -0.106**<br>(0.053)  | -0.138**<br>(0.064) |
| Virtual coach | -0.230***<br>(0.061) | -0.239***<br>(0.060) | -0.225***<br>(0.063) | -0.152**<br>(0.063) | -0.229***<br>(0.049) | -0.106<br>(0.065)   |
| Observations  | 2,346                | 2,390                | 2,390                | 2,351               | 2,351                | 2,351               |
| R-squared     | 0.300                | 0.294                | 0.235                | 0.250               | 0.211                | 0.197               |
| P-value       | 0.153                | 0.199                | 0.0517               | 0.698               | 0.0245               | 0.635               |
| Control mean  | -3.52e-09            | 9.72e-09             | 1.40e-08             | -6.71e-09           | -7.14e-09            | 4.21e-10            |

Robust standard errors in parentheses. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1. Controlling for learner gender, learner age, baseline scores, district, school quintile, fieldworker and stratification dummies. Includes repeaters.



