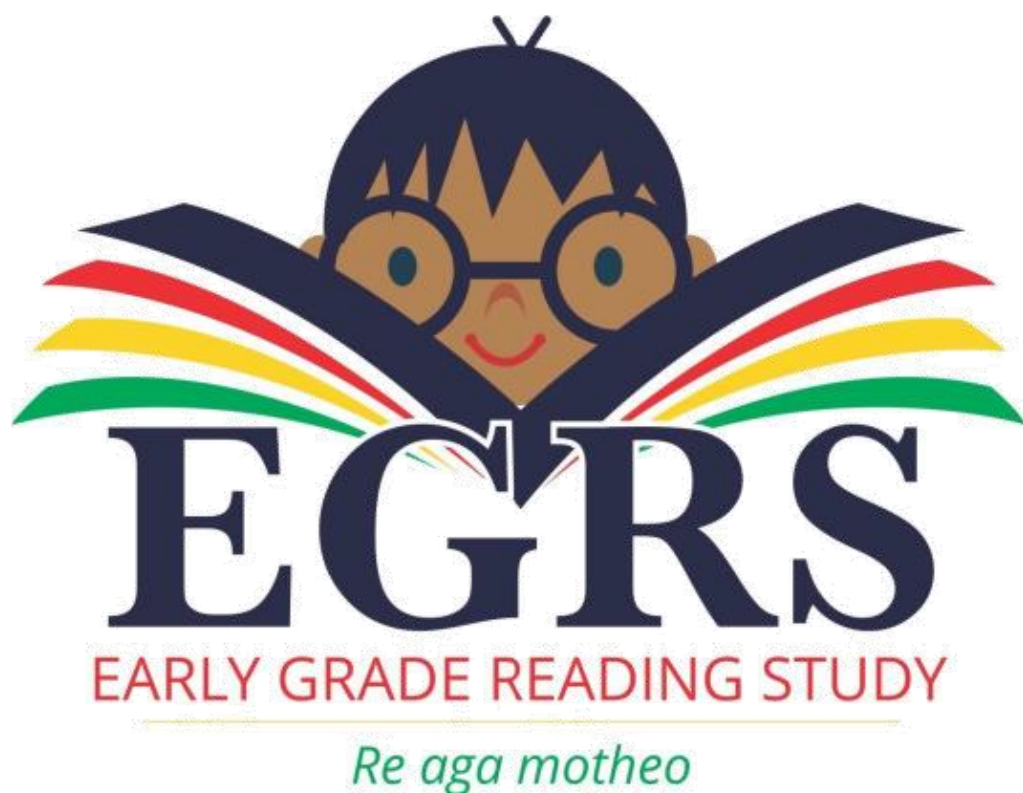


# THE EARLY GRADE READING STUDY SUSTAINABILITY EVALUATION

## Technical Report



IMPROVING EARLY GRADE READING IN SOUTH AFRICA

15 MARCH 2019

The Early Grade Reading Study team gratefully acknowledges the generous support from the donors and partners listed below



basic education

Department:  
Basic Education  
REPUBLIC OF SOUTH AFRICA



planning, monitoring  
& evaluation

Department:  
Planning, Monitoring and Evaluation  
REPUBLIC OF SOUTH AFRICA

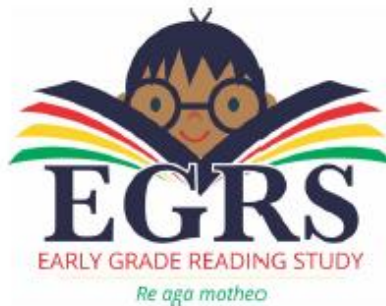


Education and Sport Development

Department of Education and Sport Development  
Departement van Onderwys en Sportontwikkeling  
Lefapha la Thuto le Tlhabololo ya Motshameko  
NORTH WEST PROVINCE



**USAID**  
FROM THE AMERICAN PEOPLE



International  
Initiative for  
Impact  
Evaluation



## CONTENTS

Acknowledgements	3
Executive Summary	4
Introduction	5
Background	6
Design	8
Intervention Design	8
Evaluation Design	9
Research Site and Sample Selection	10
Instruments	11
Grade 3 Learner Assessment	11
Grade 4 Learner Assessment	11
Teacher Questionnaire and Classroom Observation	12
Principal Questionnaire	12
School Functionality Assessment	12
Linking Form	13
Summary of Past Results	14
Baseline Results (Start of Grade 1)	14
Data Collection	14
Learner Results	14
Balance Tests	15
Correlation with Scores in Later Waves of Data Collection	15
Midline Results (End of Grade 1)	16
Data Collection	16
Learner Results	16
Correlation with Scores in Later Waves of Data Collection	17
Year 1 Results	17
Endline Results (End of Grade 2)	18
Data Collection	18
Learner Results	18
Correlation with Scores in Other Waves of Data Collection	19
Year 2 Results	20
Sustainability Evaluation of the EGRS	22
Data Collection: Grade 3 and Grade 4 learners	22
Supervisor Training and Tool Pilot Testing	22
Fieldwork	22
Attrition and Grade Repetition	23
Grade 3 Learner Sample	24
Sustainability Evaluation: Grade 4 Results	25
Summary Statistics	25
Descriptive Statistics by Intervention Group	27
Main Regression Results	31
Heterogeneous Impacts	33
Comparing Wave 3 Results with Wave 4 Results	38
Sustainability Evaluation: Grade 3 Results	40

Part 1: An Analysis of Grade 3 Learner Test Performance	40
Part 2: An Analysis of Teacher Instructional Practices	52
Conclusion	56
Way Forward	57

## ACRONYMS

Acronym	Definition
ANA	Annual National Assessments
CAPS	Curriculum and Assessment Policy Statements
DBE	South African Department of Basic Education
DPME	Department of Planning, Monitoring and Evaluation
EFAL	English First Additional Language
EGRA	Early Grade Reading Assessment
EGRS	Early Grade Reading Study
HOD	Head of Department
HSRC	Human Sciences Research Council
LTSM	Learning and Teaching Support Materials
NGO	Non Governmental Organisation
ODK	Open Data Kit
ORF	Oral Reading Fluency
PCA	Principal Component Analysis
PED	Provincial Education Department
PIRLS	Progress in International Reading Literacy Study
RAN	Rapid Automatized Naming
RCT	Randomized Control Trial
ReSEP	Research on Socio Economic Policy Unit, University of Stellenbosch
SES	Socio Economic Status
SMT	School Management Team
USA	United States of America
USAID	United States Agency for International Development
WCPM	Words Correct Per Minute

## LIST OF TABLES

Table 1: EGRS intervention design.....	9
Table 2: Summary statistics - Baseline sub-tests .....	14
Table 3: Baseline balance tests .....	15
Table 4: Correlations between baseline sub-tests and the composite scores of later waves.....	16
Table 5: Summary statistics - Midline sub-tests.....	16
Table 6: Correlations between midline sub-tests and the composite scores of the other waves.....	17
Table 7: Year 1 regression models with full controls .....	18
Table 8: Summary statistics – Endline sub-tests .....	19
Table 9: Correlation between endline sub-test and the composite scores of the other waves.....	19
Table 10: Year 2 regression models with full controls .....	20
Table 11: Instrument return rates .....	22
Table 12: Balance on reduced sample of non-attriters .....	24
Table 13: Summary statistics – Grade 4 sustainability assessment.....	26
Table 14: Correlation coefficients between waves.....	26
Table 15: Summary statistics by intervention group.....	27
Table 16: Effects on Setswana sub-tests (without controls) .....	31
Table 17: Effects on Setswana sub-tests (with controls) .....	32
Table 18: Effects on basic Setswana items (with controls) .....	33
Table 19: Effects on English and Mathematics sub-tests (with controls).....	33
Table 20: Intervention effect by learner level characteristics.....	35
Table 21: Intervention effect by school level characteristics .....	37
Table 22: Comparing learning gains between Grade 2 and Grade 4.....	39
Table 23: Summary statistics - Grade 3 sustainability assessment .....	41
Table 24: Averages on sub-tests by intervention, Grade 3 sustainability assessment.....	42
Table 25: Effects on Grade 3 Setswana sub-tests (without controls).....	44
Table 26: Effects on Grade 3 Setswana sub-tests (with controls) .....	45
Table 27: Effects on Grade 3 Setswana sub-tests (excluding multi-grade classes, with controls) .....	45
Table 28: Effects on Grade 3 English, mathematics and other sub-test outcomes (with controls) ...	46
Table 29: Intervention effects for Grade 3 sample, urban vs. rural.....	49
Table 30: Intervention effects for Grade 3 sustainability sample, male vs. female.....	49
Table 31: Intervention effects for Grade 3 sample, expected age vs. overaged.....	49
Table 32: Comparing Setswana learning levels between Grade 2 and Grade 3.....	51
Table 33: Teacher’s exposure to the programme by intervention, Grade 3 sustainability assessment (all teachers) .....	53
Table 34: Use of reading instructional practices by intervention, Grade 3 sustainability assessment .....	54
Table 35: Teaching perceptions by intervention, Grade 3 sustainability assessment. ....	54
Table 36: Work completed in learners’ DBE workbooks and exercise books .....	55
Table 37: Reading Support Project (RSP) intervention design .....	58

## LIST OF FIGURES

Figure 1: Proportion of attrited learners found in each grade .....	23
Figure 2: Median word recognition score across data collection waves and intervention. ....	28
Figure 3: Oral Reading Fluency of Setswana Text 1 by intervention.....	29
Figure 4: Oral Reading Fluency of Setswana Text 2 by intervention.....	30
Figure 5: Oral Reading Fluency of English Text by intervention.....	30
Figure 6: Graphical representation of estimated intervention .....	32
Figure 7: Quantile regression of training intervention .....	36
Figure 8: Quantile regression of coaching intervention .....	36
Figure 9: Grade 3 Setswana Oral Reading Fluency by intervention .....	42
Figure 10: Grade 3 English Oral Reading Fluency by intervention.....	43
Figure 11: Setswana composite Grade 3 test score by intervention.....	43
Figure 12: Graphical representation of estimated intervention effects (Grade 3 composite score)..	46
Figure 13: Setswana Oral Reading Fluency by intervention, Grade 3 sample.....	47
Figure 14: English Oral Reading Fluency by intervention, Grade 3 sample .....	48
Figure 15: Setswana composite score by intervention, Grade 3 sample .....	48

## APPENDICES

Appendix 1: EGRS Summary Report 2017

Appendix 2: EGRS Sustainability Assessment Tools

## ACKNOWLEDGEMENTS

This report is made possible by the generous support of the American people through the U.S. Agency for International Development (USAID), with thanks to the ZENEX Foundation, UNICEF and the Anglo American Chairman's Fund for funding the implementation of the three years of EGRS, and 3ie for funding the initial evaluation EGRS. The DBE particularly thanks USAID for their invaluable contribution and support in the design and delivery of the evaluation, and their input into the content and review of all reports.

The DBE would like to thank Khulisa Management Services for managing the sustainability evaluation, including data collection, data capturing, working with the research team on instrument design, and overseeing the development of this report.

We would also like to thank the North West Department of Education for partnering with us as the host province. We also recognize the support from officials in the districts of Ngaka Modiri Molema and Dr Kenneth Kaunda together with all the curriculum coordinators and area office staff who supported the evaluation.

Finally, the DBE appreciates the time and effort by all who participated in the evaluation, without which this evaluation would not be possible, including school principals, teachers, learners, and their parents across the North West province of South Africa.

### Contributing Authors

The South African Department of Basic Education (DBE) would like to acknowledge the work of the EGRS research team who provided oversight to the production of this sustainability evaluation report. Contributing authors include:

Stephen Taylor, Department of Basic Education\*

Janeli Kotze, Department of Basic Education

Gabrielle Wills, Research on Socio Economic Research Policy, University of Stellenbosch

Cobus Burger, Research on Socio Economic Research Policy, University of Stellenbosch

Jacobus Cilliers, Georgetown University

\*Corresponding author: [taylor.s@dbe.gov.za](mailto:taylor.s@dbe.gov.za)



## EXECUTIVE SUMMARY

From 2015, the Department of Basic Education (DBE), in collaboration with the University of Witwatersrand and other researchers, has conducted ongoing research on the acquisition of reading in the early grades in the North West Province of South Africa - Dr Kenneth Kaunda and Ngaka Modiri Molema districts.

The Early Grade Reading Study (EGRS) evaluated three Setswana Home Language interventions aimed at improving reading in the early grades: a teacher training intervention, an on-site training and coaching intervention, and a parental involvement intervention. These three interventions were implemented with the teachers of a cohort of learners in Grade 1 in 2015, the teachers of the same cohort of learners in Grade 2 in 2016, and the first two interventions were extended to the teachers of the same learners in Grade 3 in 2017. The EGRS was implemented by an organization called Class Act and the first three waves of data collection (start of Grade 1, end of Grade 1, end of Grade 2) were conducted by the Human Sciences Research Council (HSRC).

To evaluate the sustainability of the EGRS, a fourth wave of data was collected in 2018 with a focus on measuring two types of sustainability:

- i. Whether the results from the original cohort of learners who received the interventions would be **sustained into Grade 4** in 2018, one year after they had the benefit of being taught by teachers who had received the EGRS<sup>1</sup>.
- ii. Whether a **new cohort of learners in Grade 3** in 2018, whose teachers had received the EGRS interventions a year earlier, would benefit from a **sustained change in teacher instructional practices**.

The fourth wave of data was collected by Khulisa Management Services (Khulisa) and analysed by the DBE in collaboration with the Research on Socio Economic Policy unit (ReSEP) at the University of Stellenbosch. The analysis found that the impacts on the original cohort of learners persisted, with learners whose teachers received the “teacher training” and “training and coaching” interventions performing better in their Home Language than those in the control group. The magnitude of the advantage held by learners in the coaching group in 2018 was similar to that observed after two years of intervention. Whilst in 2016, the “training and coaching” intervention was estimated to be about twice as effective as the “teacher training” intervention (and therefore most cost-effective), by 2018 the gap narrowed placing the two interventions in a similar range of cost-effectiveness. Overall, the evidence of a positive impact appears clearer for the learners of the “training and coaching” intervention than the “teacher training” intervention. Moreover, a positive spill over impact on English was observed, confirming a similar trend observed in the initial impact evaluation findings.

The evaluation of the new Grade 3 sample also revealed positive estimated impacts of both the “teacher training” and “training and coaching” interventions, although the effect sizes were smaller than those seen in the original EGRS cohort and not consistently statistically significant across various learning domains. These positive estimates were larger and more robust for the coaching group compared to the training group. As per the original EGRS sample, there were also positive “spill over” effects on English outcomes.

Given these encouraging results, in 2019 and 2020 the DBE will proceed with a second phase to the EGRS – the Reading Support Program (RSP) – implemented by the Foundation for Professional Development (FPD) Consortium with funding support from USAID.

---

<sup>1</sup> Some learners were held back a year and were assessed in their repeat Grade



## INTRODUCTION

This technical report presents the findings from a **sustainability evaluation of the Early Grade Reading Study (EGRS)**. The EGRS is a large-scale impact evaluation – the biggest in South Africa – and aims to build evidence about what works to improve the teaching and learning of early grade reading in African languages.

The report is the most recent in a series of reports produced by the South African Department of Basic Education (DBE). The previous reports present the findings of the EGRS at the end of the first and second years of implementation<sup>2</sup>.

This report:

- Provides an overview of the background to the evaluation;
- Describes the EGRS design;
- Provides a summary of the past results (baseline, midline, and endline);
- Presents the process and methodology for the sustainability evaluation of the EGRS;
- Presents the findings of the sustainability evaluation;
- Presents conclusions based on these findings;
- Outlines the way forward for the EGRS.



USAID: Early Grade Learner

---

<sup>2</sup> 2017 Summary Report Available in Appendix 1 to this report

## BACKGROUND

One of the biggest developmental challenges facing South Africa is the high number of children who do not learn to read for meaning in the early years of school. This is the foundational skill upon which all others build and has, therefore, become a leading priority for the DBE. In order to address this challenge, in 2015, the DBE initiated the EGRS in collaboration with academics at the University of the Witwatersrand, the Human Sciences Research Council (HSRC), and Georgetown University (USA).

The EGRS was implemented in 230 quintile 1-3 schools in the North West province of South Africa. The core of the EGRS project is a comparison of the cost-effectiveness of three promising interventions to improve reading outcomes in learners' home language (Setswana<sup>3</sup>). Each intervention was implemented in a separate group of 50 schools with a further 80 control schools where ordinary schooling continued. A formal impact evaluation methodology known as a Randomised Control Trial (RCT), complemented by a 60-classroom observation study and eight detailed case studies, enabled the researchers to estimate the impact of each intervention on measures of reading, and to understand where, how and why different elements of the intervention models worked or did not work.

The three EGRS interventions include:

1. **A teacher training intervention:** The first intervention provided teachers with lesson plans aligned to the National Curriculum Statement Grades R-12 (NCS) including the Curriculum and Assessment Policy Statements (CAPS), as well as additional quality reading materials and training at centralized workshops twice a year.
2. **An on-site teacher training and coaching intervention:** The second intervention (implemented in a different group of 50 schools) provided teachers with the same set of lesson plans and reading materials as the first intervention but additionally provided ongoing support to teachers through specialist on-site coaching and small cluster training sessions.
3. **A parental intervention:** The third intervention (implemented in a different group of 50 schools) involved weekly meetings with parents to discuss the importance of learning to read in the early grades and to empower parents with the knowledge and tools to enable them to become more involved in their child's literacy development.

These three interventions were implemented with the teachers of a cohort of learners in Grade 1 in 2015, the teachers of the same cohort of learners in Grade 2 in 2016, and the first two interventions were extended to the teachers of the same learners in Grade 3 in 2017, covering the Foundation Phase. Baseline data collection (Wave 1) was conducted at the start of 2015 when learners had just begun Grade 1. Midline data collection (Wave 2) was conducted at the end of 2015 when the same learners had completed Grade 1. A third wave of data was collected at the end of 2016 when most of the learners were in Grade 2.

In 2018, the DBE extended the evaluation of the EGRS by collecting a fourth wave of data from the original cohort of learners whose teachers had received the interventions from Grade 1 in 2015 to Grade 3 in 2017, who were now mostly in Grade 4. The results, presented in this report, examine the sustainability of these interventions on learner outcomes by evaluating

---

<sup>3</sup> Setswana is the common Home Language in the North West province of South Africa

the long-term benefits of learners having received a higher quality of teaching in their Home Language in the Foundation Phase.

To assess whether the EGRS interventions had a sustained effect on teacher instructional practices, the DBE also collected data from a new cohort of Grade 3 learners in 2018, whose teachers had received the interventions a year earlier. The Grade 3 learner assessment results, also presented in this report, provide information on the sustainability of these interventions on teacher instructional practice by evaluating whether the impact of the interventions are seen on learner outcomes one year after the teachers received additional training and support.

This report focusses on the **sustainability of the EGRS interventions**, addressing the 2018 Grade 3 and Grade 4 results.



USAID: Learners

## DESIGN

### INTERVENTION DESIGN

As noted in the introduction, the EGRS evaluated three different interventions aimed at strengthening the teaching of early grade reading in learners' home language. The first two interventions supported teachers in teaching Setswana as Home Language. A third intervention supported parental involvement in reading outcomes. After the first two years of implementation, the evaluation results showed that the parental involvement intervention did not have a significant effect on learner performance and the intervention was subsequently stopped in the third year of implementation.

The two EGRS teacher support interventions are grounded in the educational theory of reading acquisition. As a starting point, a child needs to develop vocabulary and master decoding. Decoding relies on phonological and phonemic awareness and letter-sound knowledge, which form the basis for word recognition and oral reading fluency. Letter recognition, knowing the sound associated with letters, and phonemic awareness do not come naturally and need to be taught: they are mastered through systematic teaching and consistent practice.

To learn the basics of decoding following the Curriculum and Assessment Policy Statements (CAPS) balanced approach, a child requires a teacher who is present, capable, and motivated to deliver systematic reading instruction. For decoding to become fluent, the child requires suitable, graded materials and the discipline (perhaps imposed) and opportunities to practice substantially, both at school and at home.

The EGRS teacher interventions address these needs in various ways. Table 1 shows the difference in the materials and support received by teachers in the "teacher training" intervention (intervention 1) and "training and coaching" intervention (intervention 2) respectively.

**Scripted lesson plans**, provided to teachers in both intervention groups, deliver a structure to promote systematic teaching practice and teaching based on sound pedagogical theory. The lesson plans require little additional lesson preparation from teachers, so teachers are able to switch to more productive teaching practice with limited additional effort. The time freed up by limiting lesson preparation could improve reading acquisition if teachers allocate this time to productive teaching activities.

The literature shows that the opportunity to learn may also be hindered by a lack of suitable **Learning and Teaching Support Materials** (LTSM) to assist in the progression from one phase of reading acquisition to the next, and this is likely particularly true in African language schools. The accompanying graded reading materials that form an integral part of the scripted lessons provide material for learners to practice decoding and reading at their level of development. Teachers are required to regularly assess learners' reading proficiency in order to assign them the appropriate graded readers and assign them to smaller reading groups based on ability. Group guided reading also provides the teacher with opportunities to provide individualized attention to learners.

The **reading coach** intervention provides intensive training and support to improve teacher capacity. The underlying assumption is that, just like learning to read, the ability to teach is a skill that needs to be developed over time and may not be accomplished in one or two days of training. The literature shows that reading coaches can improve teacher motivation as teachers are frequently visited, provided with much-needed supplementary support, and may draw inspiration by learning from example.



Both teacher support interventions applied the same set of instructional practices in the teaching of Home Language literacy in Grade 1, 2, and 3 classrooms. They also both provided teachers with clearly scripted lesson plans aligned with the curriculum as specified in the Curriculum and Assessment Policy Statements (CAPS) for Home Language literacy in the Foundation Phase. The lesson plans incorporated the use of LTSM including government-provided workbooks and additional materials (graded reading booklets, flashcards, and posters). Graded reading booklets were a key resource for teachers to use in group guided reading and individual work to facilitate reading practice at an appropriate pace and sequence of progression.

Intervention 1 teachers were trained on how to use the lesson plans and accompanying materials through central training sessions, each lasting two days, and occurring twice per year. Intervention 2 teachers received one-day cluster training sessions at the start of each term and additionally received ongoing support consisting of regular (monthly) in-school coaching from specialist reading coaches. In addition to these on-site visits, intervention 2 teachers attended occasional needs-based training sessions.

*Table 1: EGRS intervention design*

<b>Programme elements</b>	<b>Intervention 1 (teacher training)</b>	<b>Intervention 2 (on-site coaching)</b>
<b>Daily lesson plans (aligned to CAPS in Foundation Phase)</b>	x	x
<b>Learning and teaching support material (Integrated with the lesson plans, including DBE workbooks, graded reading booklets, flashcards, and posters)</b>	x	x
<b>Teacher training:</b>	x (Centralized training for two days, twice a year)	x (Cluster training for one day, four times a year)
<b>Ongoing support: (Regular on-site coaching)</b>		x

An analysis of the effect of Interventions 1 and 2 on learner outcomes after only two years of implementation showed that the instructional practices regime had the potential to improve reading acquisition and that coaching, as a mode of teacher support, was an important component of effectiveness.

## EVALUATION DESIGN

To evaluate the effectiveness of the EGRS interventions, a random sample of 20 learners per school was selected in Grade 1 in 2015 and these same learners were tracked into Grade 2 in 2016. Using a battery of subtests, learners' reading proficiency was tested at the start of The evaluation results showed that the teacher training and on-site coaching interventions both improved decoding and reading comprehension, with the coaching intervention being the more cost-effective of the two. As noted previously in this report, the parental involvement intervention showed no significant impact after two years and the intervention was stopped at that point.

In 2018, the same learners were assessed in Grade 4 to evaluate whether the initial benefits of the intervention were sustained after learners received a further year of instruction from Grade 4 teachers who had not received the interventions. Furthermore, to assess the sustainability of the interventions on teacher instructional practice, data was collected from a new random sample of 10 Grade 3 learners per school. These Grade 3 learners comprise a

cohort whose teachers would have received training and support the previous year. Should there be a significant difference between Grade 3 learners' performance in the intervention schools relative to the control schools, this will provide evidence that the teachers continued implementing the methodologies that they were trained and supported on during the intervention.

## **RESEARCH SITE AND SAMPLE SELECTION**

The EGRS was implemented in the North West province of South Africa, in the districts of Dr Kenneth Kaunda and Ngaka Modiri Molema. The North West province was chosen on the basis of it being 1) a relatively poor province, thus making it relevant to the underperforming South African school system; 2) relatively homogenous in terms of home language (Setswana) making it more affordable to develop learning support materials in a single language; 3) within driving distance from Gauteng province where the national DBE is located, and 4) the senior management of the North West provincial education department was eager to partner with the DBE on this project. The district of Bojanala was excluded because another targeted intervention was taking place in that district at the same time. The district of Dr Ruth Segomotsi Mompati was excluded due to its distance from the Gauteng province and there were sufficient numbers of schools in the districts of Dr Kenneth Kaunda and Ngaka Modiri Molema to carry out the study.

The sampling frame was developed at the start of the project in 2015, starting with 458 primary schools registered in the districts of Dr Kenneth Kaunda and Ngaka Modiri Molema, per the 2014 administrative data. Relatively affluent schools (those in quintiles 4 and 5), and schools in which the language of instruction in the Foundation Phase was not Setswana were excluded. Schools that were not represented in the 2014 Annual National Assessment (ANA) dataset were further excluded. Eight schools in which instruments were piloted were also excluded. Further exclusions included very small schools (fewer than 20 Grade 1 enrolments) as many of these schools practice multi-grade teaching rendering the scripted lesson plans less appropriate. Large schools (more than 180 Grade 1 enrolments) were excluded to limit intervention costs. Three more schools were excluded after the North West Provincial Education Department (PED) checked the list and identified specific problems with these schools (e.g. the school had closed or there was a school management conflict). Following all of these exclusions, 235 eligible schools remained. Using a random number generator, five schools were excluded so that they could be retained as possible replacement schools.

The sampling frame thus included 230 schools, which were subsequently allocated randomly to the three intervention groups (50 schools each) and the control group (80 schools). To increase power and assure balance between the intervention groups, stratified randomization was applied. Stratification was based on 10 strata comprising 23 similar schools based on school size, socio-economic status, and previous performance in the ANAs. Within each stratum, 5 schools were randomly assigned to each intervention and 8 schools were randomly assigned to the control. Overall, 50 schools were randomly assigned to each of the intervention groups (teacher training, on-site coaching, and parental involvement) and 80 schools were randomly assigned to the control group.

In the 2018 Grade 4 sustainability assessment, within each sampled school the original cohorts of learners were tracked into Grade 4 and re-assessed. Data collection was not conducted in three of the schools in the original sample, with two schools having closed since the start of the study, and another not available for data collection. The Grade 3 sustainability assessment was restricted to 214 schools, excluding all identified multi-grade schools since the lesson plans are not appropriate for use in multi-grade classrooms. Within the sample of 214 schools, a random sample of 10 Grade 3 learners was selected to participate in the learner assessments. Furthermore, two Grade 3 teachers were selected to participate in a teacher interview. Finally, in all schools, the Principals were asked to participate in a short interview.

## **INSTRUMENTS**

For the sustainability assessments (Wave 4, 2018), the following instruments were administered in the schools in each sample:

1. 2018 Grade 3 oral and written learner assessment (individual and group);
2. 2018 Grade 4 oral and written learner assessment (individual and group);
3. Principal consent form;
4. Teacher consent form;
5. Teacher questionnaire and classroom observation;
6. Principal questionnaire;
7. School functionality assessment;
8. Linking form.

### **GRADE 3 LEARNER ASSESSMENT**

The Grade 3 learner assessment consisted of both an orally administered and a written assessment. The oral component of the assessment was built on the Setswana Early Grade Reading Assessment (EGRA) tasks, including the same letter recognition task as used in previous waves of data collection. The word recognition tasks included a different set of words to previous assessments and the set was extended to 70 words in total. The Grade 3 assessment included two completely new assessment tasks (Rapid Automatised Naming (RAN) of objects and letters) as measures of speed of lexical access. These tasks required learners to name out loud a set of highly familiar objects and letters within a short time period, thereby measuring phonological processing skills. RAN tasks are predictive of reading fluency and were therefore included to determine whether the underlying skills needed for reading fluency development were similar across control and intervention groups.

Further, the assessment included Setswana Oral Reading Fluency (ORF) tasks with comprehension questions. The Setswana ORF text and comprehension questions in the Grade 3 assessment was also used as the first Setswana text and comprehension questions in the Grade 4 assessment. The assessment included some English items to evaluate any spill over effects from the interventions to learners' English proficiency. This included an English word recognition task, an English ORF assessment, as well as an English receptive vocabulary assessment. The fieldworkers orally administered and recorded each learner's responses during the assessment. The learner assessment was administered one-on-one by fieldworkers, with learners, using an electronic tablet to capture the responses.

The written component of the learner assessment entailed learners individually completing an English and a Setswana reading comprehension test, as well as a one-minute, timed Mathematics task in pen and paper format. The written assessments were administered to the sampled learners in a group (i.e. one Grade 3 group). The fieldworkers monitored the written assessment to ensure that no copying occurred. The written assessment was captured by the fieldworkers using a memo for marking and thereby minimizing requirements for data capturing.

### **GRADE 4 LEARNER ASSESSMENT**

The Grade 4 learner assessment was similar to the Grade 3 assessment, with the inclusion of a grade level ORF passage to replace the ORF passage used in previous data collection.

The Grade 4 oral assessment comprised the following tasks:

- a. Setswana Rapid Automatized Naming of objects (15 seconds);
- b. Setswana Rapid Automatized Naming of letters (15 seconds);



- c. Setswana letter sound recognition (1 minute);
- d. Setswana word recognition (1 minute);
- e. Setswana oral reading fluency 1 (1 minute) followed by comprehension questions (untimed);
- f. Setswana oral reading fluency 2 (1 minute);
- g. English word recognition (1 minute);
- h. English oral reading fluency (1 minute) followed by comprehension questions (untimed).

The assessment included two ORF tasks to allow for an analysis of the sensitivity of the ORF results to the text in the assessment.

The Setswana passage and related questions used in the Grade 4 learner assessment were taken from the Progress in International Reading Literacy Study (PIRLS) Literacy narrative passage (2016; The Pearl).<sup>4</sup> Due to time constraints, only half the text was included in the assessment. The English comprehension task consisted of an informational text in English that dealt with a topic on the life cycle of plants that did not require a large amount of background knowledge. The questions on this text included both open-ended and multiple choice questions, as in the PIRLS passage. Finally, the written assessment included a short mathematics task.

## **TEACHER QUESTIONNAIRE AND CLASSROOM OBSERVATION**

The teacher questionnaire aimed to gather information about the teacher to be used in the measurement of heterogeneous treatment effects (i.e. differential impact across relevant sub-groups of schools or learners) and to measure changes in intermediate outcomes along the hypothesized causal chain for each intervention. The instrument, therefore, collected information on various teacher demographics, instructional practices, and beliefs about teaching. As part of this questionnaire, teachers were required to participate in a timed 2-minute oral and written English exercise made up of 10 questions.

## **PRINCIPAL QUESTIONNAIRE**

The principal questionnaire aimed to gather information about the school to be used in the measurement of heterogeneous treatment effects (i.e. differential impact across relevant sub-groups of schools or learners) and to measure changes in intermediate outcomes along the hypothesized causal chain for each intervention. The instrument, therefore, collected information on principal demographics, school resources, beliefs about instructional leadership, and other related topics.

## **SCHOOL FUNCTIONALITY ASSESSMENT**

The school functionality assessment was introduced as a result of the Department of Planning, Monitoring, and Evaluation (DPME) CAPS evaluation, which found major blockages to the implementation of the CAPS at schools. One of the recommendations of the CAPS evaluation was for the National and Provincial governments focus on improving school functionality. The DBE committed to implementing this recommendation in the DBE Improvement Plan. The diagnosis of school functionality in EGRS schools was the first step towards scoping the

---

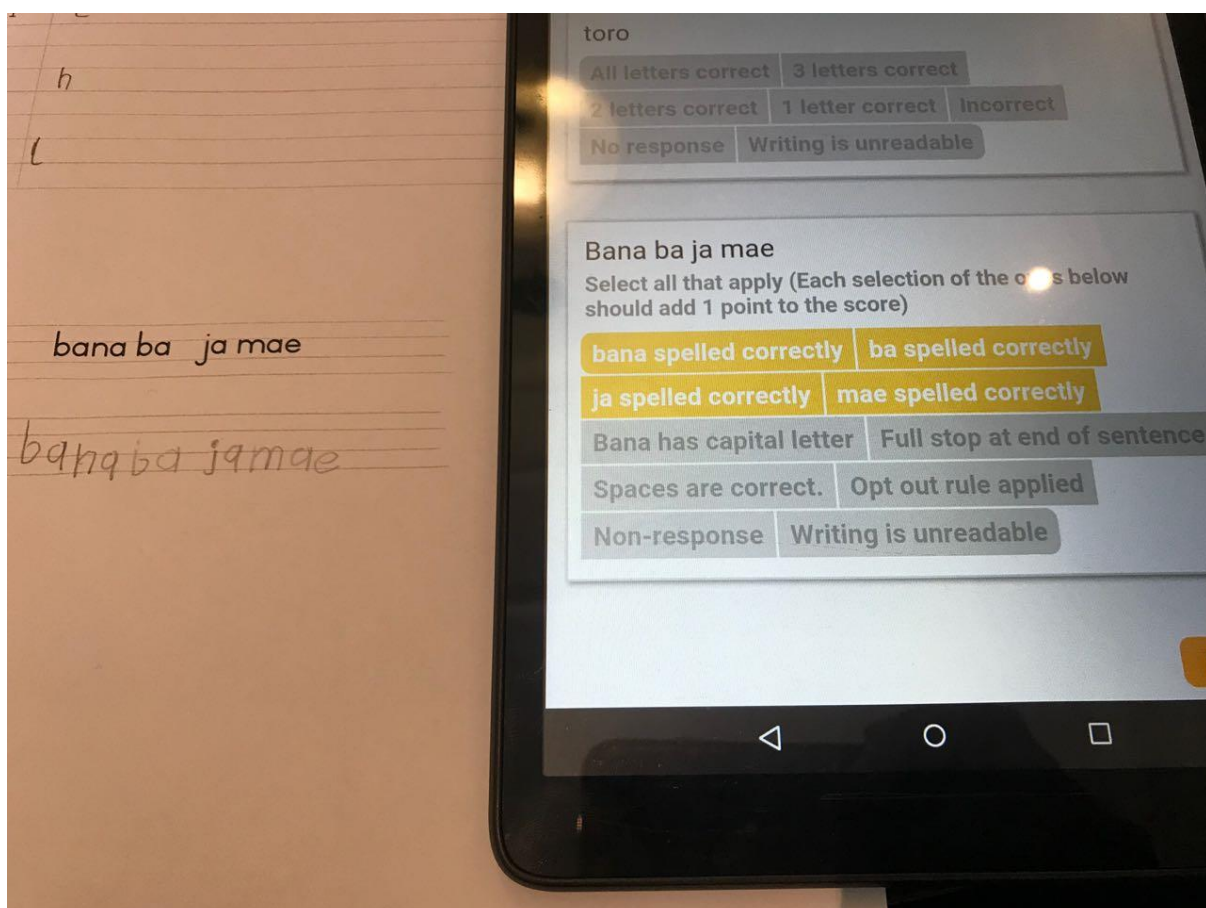
<sup>4</sup> The PIRLS Literacy passages underwent a rigorous translation process and the Setswana passage used in the assessment was received as the final PIRLS translations used by the Centre for Evaluation and Assessment (CEA). Permission for the use of this passage was granted by the International Association for Educational Achievement.

challenges associated with school functionality. The purpose of the tool was to provide a rapid appraisal of key school functionality elements such as food and nutrition, hygiene and healthcare, the school environment, teaching and curriculum delivery, school management, and learning and teaching materials. The instrument was dependent on observation and fieldworkers marked responses on a continuum.

## LINKING FORM

Linking forms were used to capture identifying information from all respondents allowing for the consolidation of databases. The linking forms were completed on the day of the school visit and the information collected during the baseline was used to pre-populate information to the extent possible. The linking forms were designed to ensure easy consolidation of databases and decrease the time required post-data collection for data screening and cleaning. Learners, teachers, and principals were assigned unique identifiers to aid with linking the databases.

The administration and completion of all the contextual instruments within a school were exclusively in English. These instruments were administered using an electronic tablet. Data capturing therefore occurred directly upon the administration of the questionnaires.



Khulisa: Learner Assessments

## SUMMARY OF PAST RESULTS

The EGRS intervention was implemented by an organization called Class Act and the baseline, midline, and endline evaluation data was collected by the Human Sciences Research Council (HSRC).

### BASELINE RESULTS (START OF GRADE 1)

#### DATA COLLECTION

Baseline data collection was conducted in all 230 sampled schools from 4 – 24 February 2015. Fieldwork monitoring confirmed that the random selection of 20 learners per school was carried out effectively and a final number of 4539 learners were assessed. Some problems with fieldwork quality were experienced during the baseline data collection, resulting in low response rates for the parent, teacher, and principal questionnaires.<sup>5</sup>

#### LEARNER RESULTS

At baseline, significant floor effects were found on the reading sub-tests, while learners performed well on the expressive vocabulary task and working memory tasks. The sub-tests related to reading were more difficult, but this was to be expected for learners at the start of Grade 1. On the letter recognition task, learners on average managed to read 5 letters correctly in a minute, and 42% of learners were not able to recognize a single letter correctly. Similarly, floor effects were seen on the word reading and sentence reading tasks.

As a measure of pre-reading skills, working memory was assessed using word span and digit span exercises. For the word span test, 89% of learners could successfully repeat a two-word sequence, with smaller proportions being able to repeat more than two words, down to only 12% who could repeat six words. In the numbers section, 93% of learners could successfully repeat a two-number sequence while only 14% could repeat a six-digit sequence. To further assess pre-literacy skills, a phonemic awareness sub-test was included. The results on this sub-test were quite low, but fieldwork monitoring revealed that the item was not always administered correctly by the fieldworkers.

Table 2: Summary statistics - Baseline sub-tests

	count	mean	min	p10	p25	p50	p75	p90	Max
Expressive vocabulary	4538	8.58	0	7	8	9	10	10	10
Letter recognition	4538	5.08	0	0	0	2	6	13	99
Working memory	4538	4.99	0	2	4	5	6	8	10
Phonological awareness	4538	2.17	0	0	0	1	3	7	12
Word recognition	4538	1.90	0	0	0	0	2	5	50
Sentence Comprehension	4538	0.73	0	0	0	0	1	3	3
Words in a sentence	4538	1.22	0	0	0	0	0	3	15
Composite score*	4538	0.00	-1.83	-.82	-.58	-.29	.27	1.13	5.40

\*The composite score is constructed to have a mean of zero and a standard deviation of 1.

---

<sup>5</sup> A full report on the data collection issues experienced is available in the report: Department of Basic Education. (2017). *Summary Report: Results of Year 2 Impact Evaluation - the Early Grade Reading Study (EGRS)*

## BALANCE TESTS

The random assignment of schools to the intervention groups, supported by the random selection of learners in schools, assumes some degree of balance between the intervention groups. Random assignment of schools to intervention, as well as the random assignment of learners in schools, happened with fidelity. Nevertheless, it is always possible for some degree of imbalance to exist. The baseline balance test indicated that one of the intervention groups (Intervention 1: Training) achieved statistically significantly lower scores on several of the baseline sub-tests. Table 3 shows the results of the balance test, which assesses whether the differences in the average scores in the learning outcomes between intervention groups are statistically significantly different from zero. Each column shows the result of a separate regression that was run on the intervention indicators, controlling for the strata fixed effects. The bottom three rows of the table show the p-values for the pair-wise tests comparing the means of the intervention groups. Of the 42 possible comparisons, there was a slight imbalance in six cases, all involving Intervention 1. Given this imbalance, it was decided that the baseline learner scores would be included in all future main model specifications.

Table 3: Baseline balance tests

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Expressive Vocabulary	Letter Recognition	Working memory	Phonological awareness	Word recognition	Sentence comprehension	Words in sentence	Combined score
T1	-0.272* (0.148)	-1.259 (1.100)	-0.490* (0.281)	-0.789** (0.346)	-0.681 (0.541)	-0.0672 (0.188)	-0.0619 (0.356)	-0.208* (0.119)
T2	-0.0829 (0.137)	0.417 (1.213)	-0.156 (0.295)	-0.192 (0.405)	0.199 (0.714)	0.248 (0.195)	0.597 (0.508)	0.0671 (0.146)
T3	-0.224 (0.147)	-0.743 (1.320)	-0.227 (0.336)	-0.190 (0.462)	0.272 (0.863)	-0.0749 (0.187)	0.944 (0.611)	-0.00297 (0.176)
Obs	4211	4211	4211	4211	4211	4211	4211	4211
T1=T2: p-value	0.233	0.105	0.288	0.151	0.214	0.153	0.200	0.069
T2=T3: p-value	0.774	0.659	0.455	0.202	0.268	0.971	0.103	0.252
T1=T3: p-value	0.371	0.369	0.846	0.997	0.942	0.143	0.630	0.728
Control mean	8.704	5.406	5.196	2.450	1.994	0.719	0.926	0.040

Notes: \*\*\* p<0.01; \*\*p<0.05; \*<0.1. Standard errors clustered at the school level and reported in parentheses.

## CORRELATION WITH SCORES IN LATER WAVES OF DATA COLLECTION

The concerns around the quality of data collection and the validity of the baseline scores were tested by considering the correlations between the sub-tests and the combined scores in later waves of data collection (Table 4). Overall, the correlations among the baseline sub-tests were relatively low, suggesting that the baseline measure may reflect substantial 'noise'. The correlations between the baseline sub-tests and the composite scores of the later waves of data further support this conclusion, with the correlation of the baseline composite score and the year 2 composite score being 0.25, the correlation with the year 3 composite score being 0.22 and the correlation with the sustainability assessment (year 4) being 0.18. Although the 'noise' measure can partially be ascribed to the low quality of baseline data collection, it is also recognized that it is difficult to assess younger learners. The floor effects that were found on some of the sub-tests contributed to the 'noise' in the baseline.

Table 4: Correlations between baseline sub-tests and the composite scores of later waves

	[A]	[B]	[C]	[D]	[E]	[F1]	[F2]	[Y1]	[Y2]	[Y3]	[Y4]
A. Picture Comprehension	1										
B. Letter sound recognition	0.14	1									
C. Digit span (Working memory)	0.30	0.24	1								
D. Phonological awareness	0.23	0.34	0.46	1							
E. Word recognition	0.17	0.53	0.33	0.47	1						
F1. Sentence reading	0.15	0.18	0.22	0.37	0.35	1					
F2. Sentence comprehension	0.14	0.23	0.31	0.45	0.46	0.53	1				
Y1. Composite Baseline Score	0.39	0.58	0.62	0.76	0.76	0.63	0.72	1			
Y2. Wave 2 composite score	0.12	0.34	0.18	0.19	0.18	0.10	0.05	0.25	1		
Y3. Wave 3 composite score	0.11	0.26	0.19	0.18	0.13	0.09	0.06	0.22	0.72	1	
Y4. Wave 4 composite score	0.10	0.22	0.18	0.14	0.11	0.03	0.07	0.18	0.64	0.80	1

## MIDLINE RESULTS (END OF GRADE 1)

### DATA COLLECTION

Incorporating lessons learned from the baseline data collection, the midline data collection was of a higher quality. The additional measures implemented included fieldworker selection criteria, conditions around the approval of deliverables, functionality criteria for the fieldworker organization and more extensive fieldworker training. Furthermore, extensive revisions were made to the midline instruments to ensure shorter contextual questionnaires, with the intention of improving response rates. At midline, 4143 (91.3%) of the 4538 learners that were assessed at baseline were re-assessed. Of the 395 learners that were not re-assessed, 283 were absent and 85 left the schools. For 25 of these learners, no reasons for non-assessment were provided, and there were only partial results for two learners. The sample attrition was relatively similar across the intervention groups, and no significant differences were found among the groups.

### LEARNER RESULTS

The midline learner assessment was adapted from the Setswana EGRA and included three sub-tests from the baseline learner assessment (letter recognition, word recognition and four of the phonological awareness tasks). It is, therefore, possible to make direct comparisons between the baseline and midline scores on these sub-tests to quantify the learning gains made over the year. Per Table 5, letter recognition improved from an average of five letters per minute to approximately 23 letters per minute. Word recognition also improved from an average of approximately two words per minute to approximately seven words per minute. Floor effects in the midline assessment were more pronounced than expected, with only the writing sub-test not showing a floor effect. The writing sub-test, as well as the letter recognition sub-test, provided enough variation within the bottom 25% of learners.

Table 5: Summary statistics - Midline sub-tests

Sub-test	mean	min	p10	p25	p50	p75	p90	max
Letter recognition	22.7	0	0	4	16	38	54	110
Word recognition	6.9	0	0	0	3	9	22	50
Non-word decoding	4.5	0	0	0	0	6	18	50
Sentence reading	4.1	0	0	0	1	9	11	11

Sub-test	mean	min	p10	p25	p50	p75	p90	max
Paragraph reading	8.0	0	0	0	0	11	30	64
Comprehension	1.0	0	0	0	0	1	4	6
Writing	5.9	0	1	4	6	8	11	12
Phonological awareness	0.7	0	0	0	0	1	3	4
Combined score	0.0	-0.94	-0.87	-0.72	-0.44	0.49	1.69	3.65

## CORRELATION WITH SCORES IN LATER WAVES OF DATA COLLECTION

Table 6 depicts the correlation matrix for the various sub-tests in the midline assessment, with the composite scores for the baseline, endline, and sustainability assessments (year 4). The within sub-test correlations are much stronger than those seen at baseline, suggesting that the midline assessment provided more valid results. The correlation coefficients between the midline composite score and the endline composite score (0.72), as well as with the year 4 score (0.64) are also high - a further indication of the validity of the midline score.

*Table 6: Correlations between midline sub-tests and the composite scores of the other waves*

	[A]	[B]	[C]	[D]	[E1]	[E2]	[F]	[G]	[Y2]	[Y1]	[Y3]	[Y4]
[A] Letter recognition	1											
[B] Word recognition	0.71	1										
[C] Non-word recognition	0.68	0.91	1									
[D] Sentence reading	0.65	0.72	0.73	1								
[E1] Paragraph reading	0.65	0.89	0.89	0.73	1							
[E2] Comprehension	0.64	0.85	0.86	0.80	0.88	1						
[F] Writing	0.62	0.60	0.57	0.65	0.59	0.61	1					
[G] Phonological awareness	0.55	0.62	0.61	0.59	0.62	0.64	0.55	1				
[Y2] Wave 2 composite	0.80	0.93	0.92	0.86	0.92	0.92	0.75	0.75	1			
[Y1] Baseline composite	0.20	0.24	0.23	0.23	0.22	0.23	0.19	0.17	0.25	1		
[Y3] Wave 3 composite	0.68	0.63	0.60	0.66	0.61	0.62	0.64	0.52	0.72	0.22	1	
[Y4] Wave 4 composite	0.61	0.58	0.54	0.55	0.55	0.55	0.56	0.47	0.64	0.18	0.80	1

## YEAR 1 RESULTS

The results suggested, after one year of implementation, that both the training and coaching interventions had a small positive impact on learner reading performance (see Table 7). The impact translated to approximately 20% of a year of learning. The impact of the parental involvement intervention was small and null-impact could not be ruled out. Data on implementation fidelity indicated that parent attendance during the weekly meeting was very low, which could explain the zero impact.

The year 1 results further revealed stronger impact for boys than for girls, suggesting that the interventions may help boys “catch up” to girls in literacy outcomes. Furthermore, urban schools (33% of the sample) saw larger gains than rural schools and, consistent with this finding, quintile 1 schools showed no impacts for both Intervention 1 and 2. Finally, excluding multi-grade schools from the sample led to clearer positive impacts on the intervention



received. This was expected, given that the interventions are aimed at supporting specific grades rather than at the multi-grade context.

*Table 7: Year 1 regression models with full controls*

	Intervention 1 (Training)	Intervention 2 (Coaching)	Intervention 3 (Parents)
Intervention 1	0.130*		
	(0.078)		
Intervention 2		0.139*	
		(0.080)	
Intervention 3			0.053
			(0.073)
Observations	2,321	2,359	2,345
R-squared	0.190	0.208	0.243

**Notes:** \*\*\* p<0.01; \*\*p<0.05; \*<0.1. Standard errors clustered at the school level and reported in parentheses.

In terms of intermediate outcomes, the evidence indicated that some instructional practices changed among intervention 1 and intervention 2 teachers in line with what is to be expected based on the interventions. Both intervention 1 and 2 teachers were more likely to group learners according to their reading ability than the control teachers. Intervention 2 teachers were also more likely to conduct individualized assessments more frequently than the control group. Finally, a review of the written work in learners' books showed more written activities, as well as a larger range of written activities among intervention 1 and 2 learners.

## ENDLINE RESULTS (END OF GRADE 2)

### DATA COLLECTION

The same strategies that were employed to ensure reliable data collection during midline, were used for the endline data collection, again providing similar positive results. At endline, 3781 (83.3%) of the original 4539 learners were re-assessed. Half of the learners that were absent at the midline data collection were successfully re-assessed, suggesting that approximately 4% of the original sample may have left the school. Over and above this 4%, a further 13% of learners were either absent on the day of the endline data collection or had left the school after Grade 1. Of the learners that were successfully identified, 591 (16% of the original sample) were found to be repeating Grade 1. Importantly, there were no significant differences in attrition or grade repetition across the three intervention groups.

### LEARNER RESULTS

The learner assessment for the end of Grade 2 (Table 8) included the same letter recognition task used in the baseline and midline evaluation. On average, letter recognition improved from five letters per minute at the start of Grade 1, to approximately 23 letters per minute at the end of Grade 1. At the end of Grade 2, learners on average read 39.5 letters per minute.

Similarly, the word recognition task was exactly the same in all three waves, showing an improvement from approximately seven words per minute at the end of Grade 1 to 19 words per minute at the end of Grade 2. Learner performance on the ORF tasks showed a higher proportion of learners (36%) not managing to read a single word than in the word recognition task (16%). However, at each performance percentile, the average words read in context was higher than the single words read in the word recognition task. This could suggest that once learners master the skill of decoding, reading fluency quickly follows. That learners were able



to read more words in connected text than in single word list could suggest that readers reached a level of automaticity in their decoding such that they were able to utilise other cues in the text to enhance fluency (such as pre-processing words in the perceptual span, and using syntactic and morphological information). Learners could also potentially have read more words in the connected text due to how Setswana is written. Many of the function words (grammatical markers) are shorter than the words included in the word reading test, therefore leading to more words read in the ORF passage in Setswana.

Three items were included to test phonological awareness; however, this remained a difficult skill to assess. Several writing items were included and provided good variation in the learner scores, with learners in the 10<sup>th</sup> percentile at least scoring some items correct and only learners at the 90<sup>th</sup> percentile achieving full marks on the sub-tests. Finally, some English and mathematics items were included to evaluate the effect of the interventions on other subjects. On average, learners scored 0.6 out of 2 on the mathematics items and 3.14 out of 8 on the English items.

*Table 8: Summary statistics – Endline sub-tests*

	count	mean	min	p10	p25	p50	p75	p90	max
Letter recognition	3781	39.5	0	2	16	41	60	74	110
Word recognition	3781	19.4	0	0	3	17	34	45	50
Non-word recognition	3781	14.4	0	0	0	13	26	34	50
Oral Reading Fluency	3781	25.6	0	0	0	23	50	64	66
Reading comprehension	3781	1.27	0	0	0	1	2	3	4
Phonological awareness	3781	1.82	0	0	1	2	3	3	3
Writing	3781	5.97	0	3	4	6	8	9	9
Mathematics	3781	0.60	0	0	0	1	1	1	2
English	3781	3.14	0	0	0	2	6	8	8
Composite score (SD)	3781	0	-1.59	-1.23	-0.97	-0.01	0.89	1.37	2.16

## CORRELATION WITH SCORES IN OTHER WAVES OF DATA COLLECTION

Table 9 shows the correlation matrix of the various sub-tests in the endline assessment compared with the composite scores of the assessments in the other waves of data collection. The inter-test correlations of the Setswana sub-tests in the endline assessment were very well correlated. The English sub-test was also well correlated with the Setswana sub-tests, whereas the Mathematics sub-test had lower correlation coefficients. Finally, the high correlation coefficients between the endline composite score and the midline composite score (0.72), as well as the sustainability composite score (0.75), suggest that the endline assessment provided a valid indication of learner performance.

*Table 9: Correlation between endline sub-test and the composite scores of the other waves*

	[A]	[B]	[C]	[D1]	[D2]	[E]	[F]	[G]	[H]	[Y3]	[Y1]	[Y2]	[Y4]
[A] Letter recognition	1												
[B] Word recognition	0.73	1											
[C] Non-word recognition	0.71	0.94	1										
[D1] Paragraph reading	0.69	0.93	0.91	1									
[D2] Comprehension	0.61	0.82	0.80	0.83	1								
[E] Phonological awareness	0.39	0.47	0.47	0.47	0.46	1							

	[A]	[B]	[C]	[D1]	[D2]	[E]	[F]	[G]	[H]	[Y3]	[Y1]	[Y2]	[Y4]
[F] Writing	0.59	0.66	0.63	0.65	0.61	0.45	1						
[G] Math	0.37	0.42	0.41	0.42	0.42	0.29	0.45	1					
[H] English	0.63	0.87	0.84	0.86	0.78	0.46	0.62	0.42	1				
[Y3] Wave 3 composite	0.80	0.95	0.94	0.95	0.88	0.59	0.77	0.47	0.87	1			
[Y1] Wave 1 composite	0.20	0.19	0.19	0.19	0.16	0.15	0.17	0.12	0.19	0.21	1		
[Y2] Wave 2 composite	0.54	0.74	0.69	0.71	0.63	0.38	0.53	0.36	0.72	0.72	0.23	1	
[Y4] Wave 4 composite	0.64	0.79	0.76	0.78	0.68	0.41	0.60	0.41	0.75	0.80	0.18	0.64	1

## YEAR 2 RESULTS

After two years of implementation, the coaching intervention showed a substantial positive impact (Table 10). Learners who received two years of the coaching intervention were approximately 40% of a year of learning ahead of the learners in the schools that received no intervention (control schools). The training intervention and the parent involvement intervention showed a small positive impact, but the impact was not significantly different from zero. The coaching interventions registered statistically significant positive effects on all the Home Language literacy measures, as well as on the English sub-tests.

This finding is encouraging as it suggests that although the coaching intervention focused on Setswana as Home Language, it also had a positive effect on English. This can either be the result of the strengthened home language ability among the learners or simply due to improved classroom management and transferable instructional methods. No significant impacts were seen on the Mathematics sub-test, at least indicating that the increased focus of the teaching of Home Language had no negative consequences for Mathematics.

*Table 10: Year 2 regression models with full controls*

	Intervention 1 (Training)	Intervention 2 (Coaching)	Intervention 3 (Parents)
Intervention 1	0.112 (0.0814)		
Intervention 2		0.252*** (0.0792)	
Intervention 3			0.103 (0.0768)
Observations	2,121	2,140	2,140
R-squared	0.170	0.178	0.183

**Notes:** \*\*\* p<0.01; \*\*p<0.05; \*p<0.1. Standard errors clustered at the school level and reported in parentheses.

The endline assessment further revealed that the coaching intervention helped boys “catch up” to some degree to girls. The impact was also concentrated among the urban schools, with the impacts of all three interventions being larger in urban township settings. No measurable impacts were found in deep rural settings, suggesting that the interventions should be approached differently in these settings. Further, both the training and the coaching interventions (teacher support interventions) appear most beneficial for relatively large classes (38 to 45 learners). However, in very large classes (50 plus learners), the impact of the EGRSs was smaller. These results could be the result of the teacher support interventions having emphasized good classroom management practices, such as classroom organization, work in small groups, and predictability to the classroom.

In terms of teacher-level outcomes, teachers in the coaching intervention schools were more likely to report feeling a high level of professional support, compared to teachers in the control

schools. Teachers in the training intervention were also somewhat more likely to report having experienced high levels of professional support. The Learning and Teaching Support Materials (LTSM) provided through the interventions were similar in the teacher support interventions and, in both groups, learners were more likely to use the graded readers provided. This increase was substantially larger for teachers who received coaching relative to the teachers in the control schools. Similarly, teachers in both teacher support interventions were more likely to implement group guided reading, resulting in more opportunities for learners to do individualised reading. Again, this practice was observed more frequently in the coaching schools, relative to the training schools.



USAID: Teacher

## SUSTAINABILITY EVALUATION OF THE EGRS

### DATA COLLECTION: GRADE 3 AND GRADE 4 LEARNERS

#### SUPERVISOR TRAINING AND TOOL PILOT TESTING

Pilot testing for the Grade 4 learner assessment took place from 2-8 August 2018. Five fieldwork supervisors were trained on the protocols for fieldwork and each of the assessment instruments. Fieldwork supervisors were provided an overview of the study and were introduced to the research tools and data collection software. A substantial part of the training was dedicated to tool orientation and protocols for assessing learners.

Pilot testing was carried out at four schools, selected by the DBE, in Dr Kenneth Kaunda district in the North West province from 6–8 August 2018. The pilot provided valuable lessons for the instrument design, sequencing (i.e. the order of administration that made the most sense), and lessons to take forward into fieldworker training (i.e. which tools fieldworkers should be trained on first and when to introduce the tablet-based data collection instruments). Furthermore, the tools were reviewed using the pilot results to determine which questions needed to be dropped or adjusted.

#### FIELDWORK<sup>6</sup>

Fieldworker training took place from 20 – 24 August 2018. The five-day training workshop in the North West province was attended by 56 fieldworkers, of which 46 were selected for fieldwork and 10 were appointed as reserves. Two days of the training were dedicated to the tools, while the third and fourth day entailed in-venue and school-based fieldworker role play and fieldwork simulation, and the fifth focused on administration and logistical arrangements. Fieldwork was conducted from 27 August – 28 September 2018, in the third academic term.

During fieldwork, high attrition of Grade 4 learners was evident. These learners had either transferred to different schools within or outside the province. Data collection was not conducted at some schools for various reasons, mainly logistical. Table 11 below provides a summary of the return rates per research tool.

*Table 11: Instrument return rates*

Instrument Name	Number collected	Number expected	Percentage
Gr 3 Learner Assessment	2116	2140	99%
Gr 3 Written Assessment	2099	2116	99%
Gr 4 Learner Assessment	3304	4519	73%
Gr 4 Written Assessment	3304	4519	73%
Principal Questionnaire	221	228	97%
Teacher Questionnaire	633	NA	NA
School Functionality Tool	217	228	95%
Parent /Guardian Questionnaire (Gr 1 parents)	3462	4202	82%

---

<sup>6</sup> A more comprehensive fieldwork report is available upon request.

## ATTRITION AND GRADE REPETITION

In 2015, 4538 Grade 1 learners were assessed at the start of the year in the baseline data collection. At the end of 2015, 4143 of this sample of learners were re-assessed in the midline data collection, meaning that 9% of the sample were absent from school on the day of testing. At the end of 2016 (end of Grade 2), 3781 of the original sample of learners were re-assessed in the endline data collection. At the end of Grade 2, it was evident that 16% of the sample of learners who were successfully re-assessed were found to be repeating Grade 1, and 16% of learners were absent on the day of testing.

Figure 1 shows the attrition and grade repetition in 2018 when the original cohort of learners was expected to be in Grade 4. The graph shows the proportions of the original sample in Grade 4 and those that had repeated and were found to be either in Grade 3 or Grade 2. Overall, approximately 50% of the sample were successfully identified in Grade 4, whereas approximately 21% of the learners were repeating Grade 3. A small minority of learners were identified in Grade 2, which may either suggest that the learner repeated twice during the Foundation Phase, or that the learner may have been demoted to Grade R during the Grade 1 year. The first reason is not expected as the Admission Policy for Ordinary Public Schools<sup>7</sup> stipulates that the guideline for repetition is 'one year per school phase where necessary'. The latter reason is not uncommon, as learners often enter school without being adequately prepared. Regression analysis found no significant correlation between grade repetition and intervention status.

On average, 27% of the original sample of learners were not identified in the wave 4 data collection. A slightly larger proportion of the learners who were in the schools that received the training intervention were not observed in the wave 4 data collection. Regression analysis confirms that attrition is not significantly predicted by intervention status.

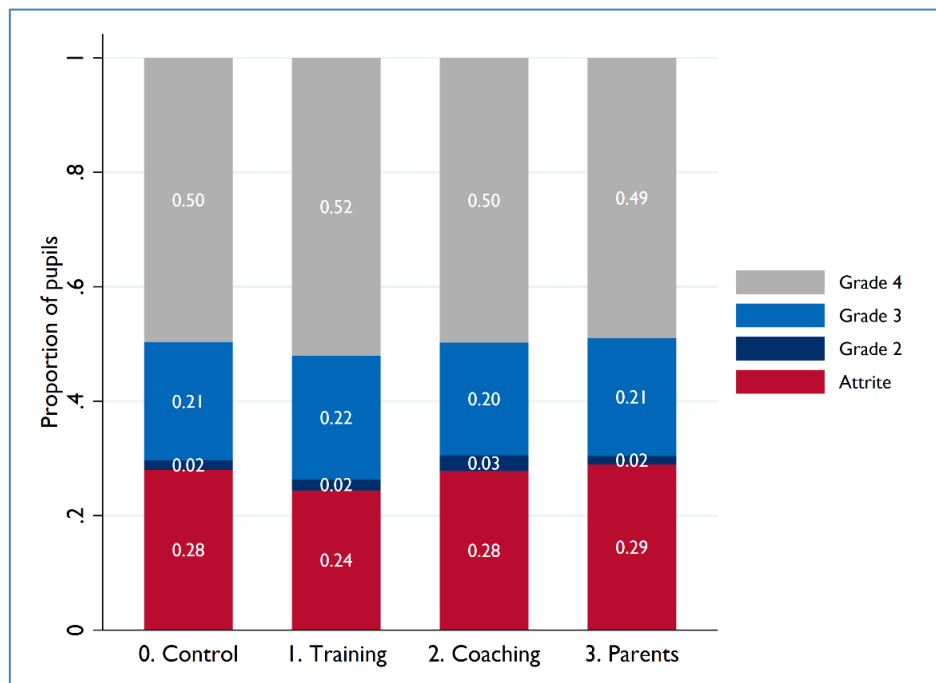


Figure 1: Proportion of attrited learners found in each grade

<sup>7</sup> Department of Education. 1998. Admission Policy for Ordinary Schools, as Published as Government Notice 2432, Government Gazette, Volume 200, No. 19377.

To further investigate the implications of attrition on sample balance, subsequent analysis was conducted as depicted in Table 12. Each of the columns in the table represents a separate regression which was run on learner age, learner gender, and learner baseline score respectively. The results suggest that in the wave 4 sample, the learners in the training intervention group were on average slightly older than the learners in the control group and that there were more male learners in the training group than the control group. Fortunately, the attrition does not seem to have caused an imbalance in terms of learner ability as measured at baseline.

*Table 12: Balance on reduced sample of non-attriters*

	Age	Female	Combined Score
Training	0.141**	-0.051**	-0.189
	-0.055	-0.025	-0.120
Coaching	0.028	-0.025	0.105
	-0.052	-0.022	-0.155
Parents	0.077	-0.052**	-0.014
	-0.055	-0.021	-0.163
Observations	3,283	3,295	3,295
R-squared	0.012	0.003	0.065
Control mean	6.419	0.5	0.039
Std Dev	0.645	0.5	0.943

**Notes:** \*\*\* p<0.01; \*\*p<0.05; \*<0.1. Standard errors clustered at the school level and reported in parentheses.

## GRADE 3 LEARNER SAMPLE

Fieldworkers had to randomly select 10 learners from 214 of the original 230 schools. In the end, data was collected in 212 of the 214 schools, and of these, 72 were control schools, 47 received the training intervention, 46 were coaching schools and the remaining 47 had received the parent intervention. In 173 of the 212 schools exactly 10 learners were tested. But in 39 schools, less or more than 10 learners were tested where the minimum number of learners tested per school was 6 and the maximum was 13.

One-on-one test data orally administered and assessed by the fieldworker, using tablet-based software called Tangerine, was obtained for 2116 Grade 3 learners. Of these learners, 2081 also wrote the written component of the test. There were 35 learners who only wrote one-on-one tests, while 18 learners only wrote the written assessments.<sup>8</sup> It is important to note that the Grade 3 sample excludes any Grade 3 learners who may have repeated and thus are a part of the original baseline sample.

---

<sup>8</sup> Due to the number of learners that were assessed on a day, the oral assessments and written assessments were not necessarily all administered on the same day.



## SUSTAINABILITY EVALUATION: GRADE 4 RESULTS

This section of the report evaluates whether the impact of the interventions are sustained on learner outcomes one year after the learners were exposed to the interventions. To evaluate the sustained impact, this section focuses on the analysis of Grade 4 learner performance.

### SUMMARY STATISTICS

Overall, the summary statistics (Table 13) for the Grade 4 assessment attest to good variation in learner test scores, with items such as Object Naming, Letter Naming and Letter Recognition providing ample information at the lower end of the performance distribution. On average, learners read 41 Setswana letters per minute correctly, which is remarkably similar to the average of 39.5 letters per minute attained in the endline assessment (end of Grade 2). Encouragingly, the number of letters correctly identified improved at the lower end of the performance distribution, since the Grade 2 assessment. There does, however, appear to have been a slight decrease in the number of letters identified correctly at the higher end of the performance distribution. This could be due to the inclusion of more difficult letter sounds such as digraphs (e.g. 'ng') and trigraphs (e.g. 'ngw') later in the list of letters.

On average, learners read 30.5 words correct per minute (WCPM); an improvement from the 19.4 words on average at the end of Grade 2. Similarly, ORF improved from an average of 25 words per minute to between 49 – 56 words per minute (dependent on which of the Setswana texts is considered). Interestingly, the ORF results revealed that there were learners at the 10<sup>th</sup> percentile who still could not manage to read a single word correctly, whereas learners at the 25<sup>th</sup> percentile managed to read about 23-25 WCPM. This indicates the possibility of a fluency “tipping point” and, only after reached, learners become more fluent readers.

The average score for English word recognition was similar to the average score for Setswana word recognition, but this hides the difference in the distribution between these scores. Learners at the lower end of the performance distribution were less likely to be able to read the English words correctly compared to the Setswana words, whereas learners at the 75<sup>th</sup> and 90<sup>th</sup> percentile were reading more English words correctly. English text reading again indicates a tipping point, with learners at the 10<sup>th</sup> and 25<sup>th</sup> percentile not being able to read much English text, but learners at the 50<sup>th</sup> percentile reading on average 31 words per minute. From the 75<sup>th</sup> percentile upwards, learners appear to have read with the same reading speed in both languages.

Some floor effects were found on both the Setswana and English written comprehension test. On average, learners scored 1.6 out of 7 on the Setswana assessment and 1.1 out of 6 on the English assessment. For both the Setswana and English written comprehension, only learners at the 50<sup>th</sup> percentile managed to answer one question correctly.

A composite score was also constructed from all the Home Language literacy items, using Principal Component Analysis (PCA). PCA is a type of factor analysis which identifies the common underlying variation amongst a set of variables, which in the context of this study is Setswana reading literacy. The composite score was standardized to have a mean value of 0 and a standard deviation of 1. The purpose of the composite score is to provide a uni-dimensional measure of Setswana reading ability to assess the impact of the interventions on learner performance. For this reason, only the sub-test that related to the Setswana skills that learners were supposed to have acquired through the Setswana Home Language curriculum (CAPS) throughout the Foundation Phase were included in the composite score. Since the two rapid naming speed tests measure ‘inherent’ reading ability, and should therefore not be affected by the interventions, they were excluded from the composite score. Similarly, the English sub-tests were excluded.



Table 13: Summary statistics – Grade 4 sustainability assessment

	N	Mean	Min	p10	p25	p50	p75	p90	Max
Object Naming	3302	14.6	0	10	12	14	17	20	35
Letter Naming	3302	20.3	0	10	15	20	26	31	36
Setswana Letter Recognition	3302	41.1	0	14	28	41	55	66	110
Setswana Word Recognition	3302	30.5	0	2	16	34	44	51	70
Setswana Text 1 Reading	3302	48.9	0	0	23	50	68	92	159
Setswana Text 1 Comprehension	3302	2.5	0	0	1	3	4	5	8
Setswana Text 2 Reading	3302	56.5	0	0	25	56	83	107	220
English Word Reading	3302	30.0	0	0	6	27	48	64	102
English Text Reading	3302	39.1	0	0	1	31	66	92	142
English Text Comprehension	3302	1.3	0	0	0	0	2	4	8
Setswana Written Comprehension	3321	1.6	0	0	0	1	2.5	4	7
English Written Comprehension	3321	1.1	0	0	0	1	2	3	6
Mathematics	3321	2.4	0	0	1	2	4	5	9
Composite score	3201	0.05	-1.76	-1.51	-.63	.16	.71	1.27	3.45

Overall, the learner test information collected in waves 2 (midline), 3 (endline) and 4 (sustainability) appear to have been of better quality than that obtained through the baseline. Table 14 provides evidence of the low correlation between the baseline score and the composite scores in the later waves of data collection. The wave 4 score, however, is highly correlated with the scores in waves 2 and 3, signalling a more reliable measure of learner performance. Cronbach's alpha is a further measure that can be used to evaluate the validity of the assessment. Cronbach's alpha provides a measure of how well the various sub-tests fit together as measures of a single underlying construct, where a value closer to 1 is better. The value for the Grade 4 assessment is 0.87 – further substantiating the validity of the assessment.

Table 14: Correlation coefficients between waves

	[A]	[B]	[C]	[D]	[E]	[F]	[G]	[H]	[I]	[J]	[K]	[L]	[Y4]	[Y1]	[Y2]	[Y3]
[A] Object Naming	1															
[B] Letter Naming	0.41	1														
[C] Setswana Letter Recog.	0.35	0.75	1													
[D] Setswana Word Recog.	0.37	0.60	0.66	1												
[E] Setswana Text 1 Reading	0.34	0.55	0.62	0.90	1											
[F] Setswana Text 1 Comp.	0.33	0.50	0.56	0.75	0.80	1										
[G] Setswana Text 2 Reading	0.35	0.53	0.59	0.88	0.89	0.74	1									
[H] English Word Reading	0.34	0.53	0.58	0.85	0.84	0.71	0.82	1								
[I] English Text Reading	0.33	0.50	0.54	0.82	0.85	0.72	0.82	0.89	1							
[J] English Text Comp.	0.27	0.41	0.42	0.60	0.63	0.61	0.60	0.72	0.75	1						
[K] Setswana Written Comp.	0.10	0.11	0.11	0.15	0.15	0.14	0.14	0.15	0.16	0.12	1					
[L] English Written Comp.	0.05	0.09	0.12	0.13	0.12	0.10	0.13	0.14	0.14	0.13	0.53	1				
[Y4] Wave 4 composite	0.39	0.65	0.75	0.95	0.95	0.87	0.93	0.86	0.85	0.65	0.20	0.16	1			
[Y1] Baseline composite	0.06	0.17	0.17	0.16	0.15	0.15	0.13	0.17	0.16	0.15	0.06	0.06	0.17	1		
[Y2] Wave 2 composite	0.28	0.42	0.45	0.57	0.58	0.52	0.56	0.61	0.61	0.55	0.13	0.12	0.60	0.23	1	
[Y3] Wave 3 composite	0.32	0.52	0.55	0.76	0.76	0.66	0.73	0.75	0.76	0.60	0.14	0.12	0.78	0.21	0.72	1

## DESCRIPTIVE STATISTICS BY INTERVENTION GROUP

To get an initial perspective on the possible sustained impacts of the intervention, a simple comparison of means is reported in Table 15. From the table, it is evident that the average scores for the learners in the coaching group are higher than the learner performance in the other groups. No remarkable differences in learner performance are clear between any of the other intervention groups and the control group.

*Table 15: Summary statistics by intervention group*

	N	Control	Training	Coaching	Parents
Object Naming	3301	14.61	14.74	14.71	14.41
Letter Naming	3301	20.52	19.60	20.65	20.23
Setswana Letter Recognition	3301	40.33	40.25	43.65	40.48
Setswana Word Recognition	3301	29.89	30.92	32.98	28.52
Setswana Text 1 Reading	3301	47.36	50.19	53.72	45.29
Setswana Text 1 Comprehension	3301	2.38	2.57	2.83	2.30
Setswana Text 2 Reading	3301	55.73	56.60	60.97	52.88
English Word Reading	3301	29.41	30.45	32.30	28.02
English Text Reading	3301	39.13	39.25	41.86	36.22
English Text Comprehension	3301	1.33	1.32	1.51	1.13
Setswana Written Comprehension	3321	1.51	1.46	1.77	1.56
English Written Comprehension	3321	1.11	1.04	1.14	0.94
Mathematics	3321	2.41	2.24	2.39	2.38
Composite score	3201	0.00	0.05	0.20	-0.04

The next four figures present further descriptive evidence of the differences in learner performance among the intervention groups. Figure 2 firstly shows the progression in single word recognition across the four waves of data collection by looking at the median WCPM. As expected, learners from all intervention groups could hardly read any words correctly at the start of Grade 1 (baseline). By the end of Grade 1, learner performance improved marginally, and by the end of Grade 2, larger learning gains were seen. These learning gains were sustained through the two years with the median WCPM ranging from about 31 to 35 words (dependent on intervention group) in Grade 4. In terms of differences between the various intervention groups, the learners in the coaching group had much higher learning gains than the other intervention groups by the end of Grade 2 (Wave 3), but that the other intervention groups closed that gap marginally by the end of Grade 4 (Wave 4). Nevertheless, learners in the coaching group still had the highest median WCPM, followed by learners in the training group.

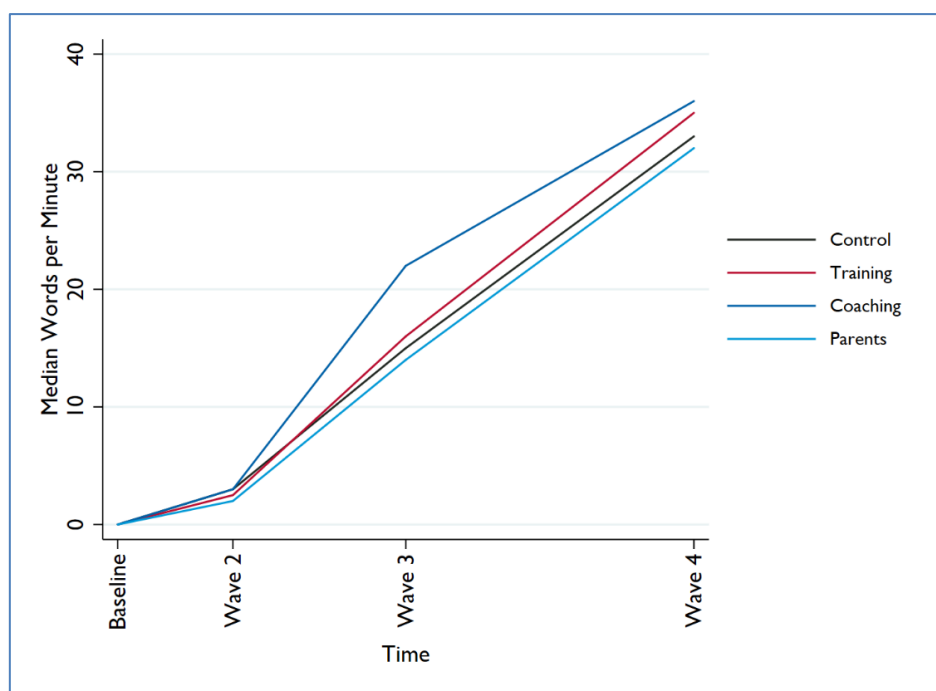


Figure 2: Median word recognition score across data collection waves and intervention.<sup>9</sup>

The following three figures (Figures 3, 4 and 5) demonstrate the percentage of learners achieving above a specific threshold (measured in correct WCPM). The first two graphs present the results for the first and second Setswana ORF texts, whereas the third graph shows the results for the English ORF text. Considering the first Setswana text (Figure 3), it can be seen that in all four groups, at least 86% of learners could read at least one word correctly (with about 14% of learners not being able to identify any words correctly). Considering a threshold of 40 words read correctly<sup>10</sup>, approximately 60% of the learners in the control and parent intervention groups could read at this level. In the training intervention group, this percentage is slightly higher at 65% and in the coaching group, 70% of the learners could read at this level. The percentage of learners reaching certain thresholds differ between the three texts considered, but the ranking between the intervention groups remains constant. Learners in the coaching intervention group were consistently able to read more words correctly per minute than learners in the other groups, followed by learners in the training intervention group. Moreover, the pattern in all three graphs suggests that the impact of the coaching intervention was largest for learners in the mid to lower range of the performance distribution.

<sup>9</sup> Wave 2 was conducted at the end of Grade 1, Wave 3 at the end of Grade 2 and Wave 4 during the third term of Grade 4.

<sup>10</sup> This is an arbitrary threshold set for the purposes of explaining the interpretation of the graph. Once reading norms have been developed for Setswana, more meaningful thresholds could be used.

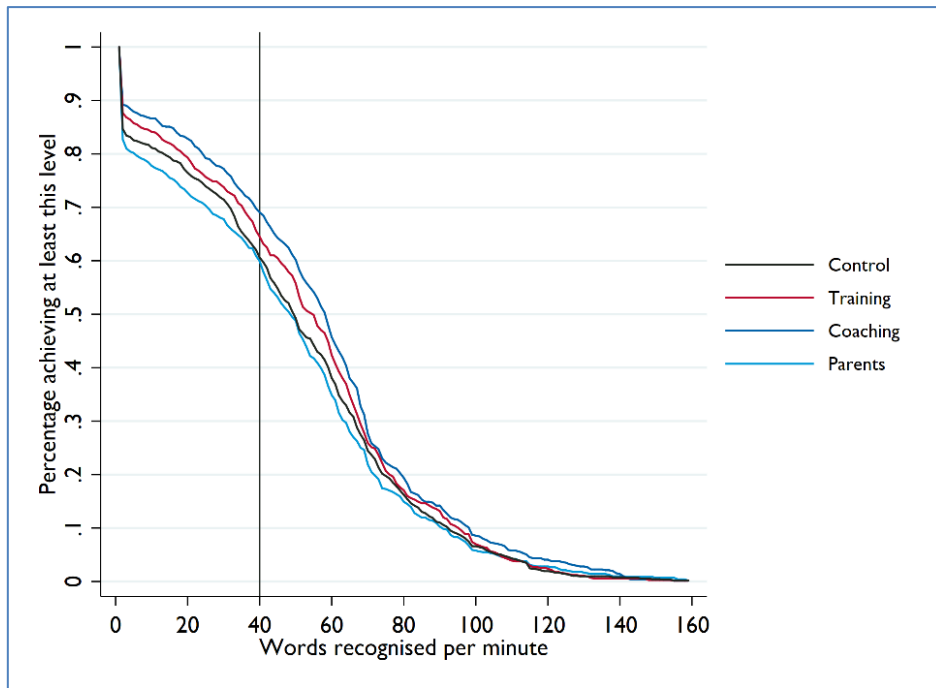


Figure 3: Oral Reading Fluency of Setswana Text 1 by intervention

Figure 4 presents the same type of graph as above, applied to the second Setswana text. From the mean summary statistics, learners performed slightly better on this text relative to the first text. Further linguistic analysis will be conducted to understand whether this result is driven by the second text having been easier, or whether this may have been a result of learners being more familiar with the task in the second text. Again, at least 88% of learners could read at least one word correctly (that is about 12% of learners could not identify a single word correctly). Approximately 65% to 70% of learners in the control, training, and parent intervention groups were able to read at least 40 WCPM, and 73% of learners in the coaching group reached this threshold.



USAID: Learner Writing

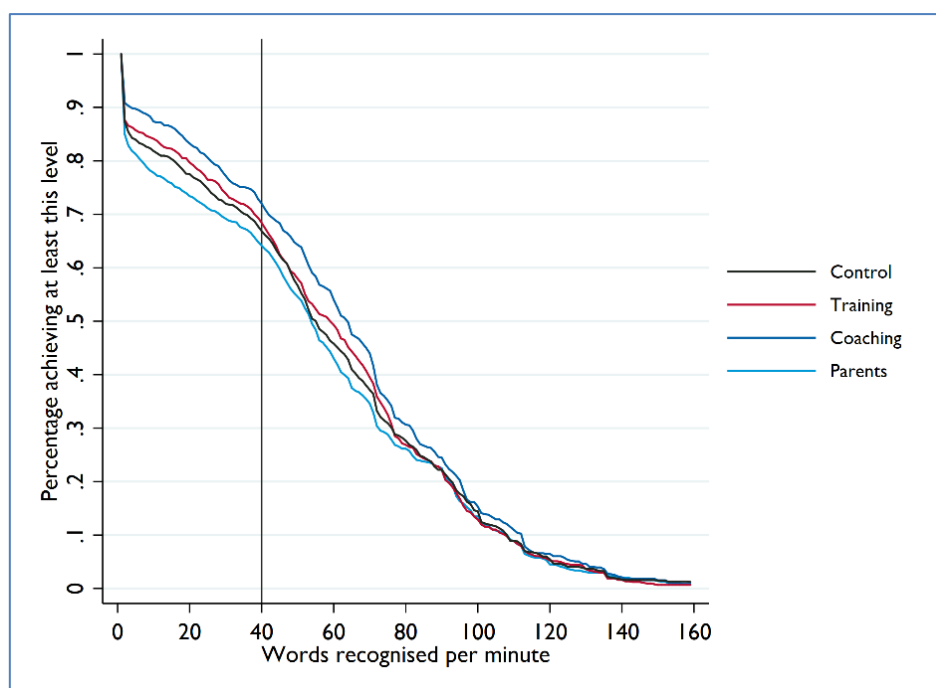


Figure 4: Oral Reading Fluency of Setswana Text 2 by intervention

Figure 5 shows the percentages of learners reading at the various thresholds on the English ORF text. Approximately 75% of learners in the control and parent intervention groups managed to read at least one word correctly (23% and 28% of learners in these groups respectively could not read a single word correctly), whereas approximately 80% of learners in the training intervention and coaching intervention managed to reach this threshold. Again, at each threshold of words read correctly, a higher proportion of learners in the coaching intervention managed to read at that level.

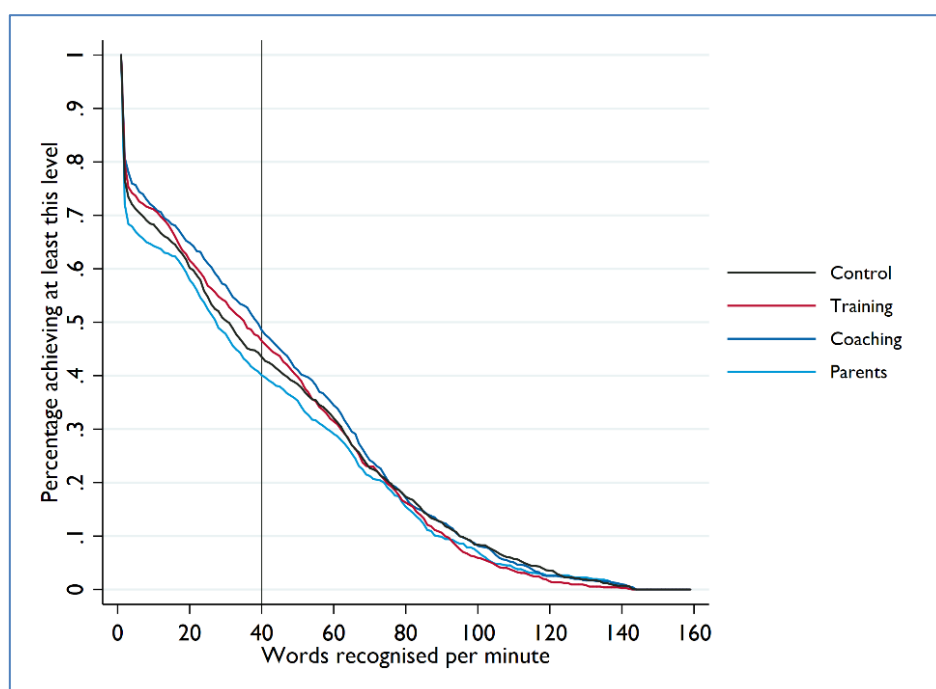


Figure 5: Oral Reading Fluency of English Text by intervention

## MAIN REGRESSION RESULTS

As an initial perspective, the descriptive results suggest that the benefits of the coaching intervention may have been sustained as learners progressed to the intermediate phase, where teachers did not receive additional support. Although the learner performance in the training group was also higher than the control group, it is less clear from the descriptive analysis whether this is statistically significant. Random assignment should allow for a simple mean comparison, but regression modelling has some additional benefits to obtaining more precise estimates.

The first regression model considered only controls for stratification, which assists in reducing the size of the standard errors based on the sampling procedure used. The results (Table 16) confirm the initial results from the descriptive analysis, with learners in the coaching group performing statistically significantly better on each of the Setswana language sub-tests. The p-value reported in the table provides an indication of the statistical difference of the impact between the training and coaching intervention groups. This indicator suggests that there was a significant difference on the Letter Recognition and the Written Comprehension tasks, but that we cannot say with confidence that the learners in the coaching group outperformed the learners in the training group on the other sub-tasks.

*Table 16: Effects on Setswana sub-tests (without controls)*

VARIABLES	(1) Overall	(4) Letter recognition	(5) Word recognition	(6) Text 1 reading	(7) Text 1 comprehension	(8) Text 2 reading	(9) Written comprehension
Training	0.054 (0.077)	-0.004 (0.085)	0.058 (0.076)	0.085 (0.071)	0.105 (0.0715)	0.022 (0.072)	-0.035 (0.071)
Coaching	0.199*** (0.073)	0.166** (0.082)	0.175** (0.068)	0.190*** (0.065)	0.247*** (0.0729)	0.132** (0.064)	0.165** (0.067)
Parents	-0.039 (0.077)	0.008 (0.084)	-0.077 (0.073)	-0.062 (0.069)	-0.0453 (0.070)	-0.072 (0.069)	0.030 (0.070)
Observations	3,201	3,301	3,301	3,301	3,301	3,301	3,321
R-squared	0.007	0.005	0.008	0.008	0.012	0.005	0.005
P-value	0.049	0.0388	0.110	0.136	0.046	0.108	0.008
Control mean	0.499	-0.038	-0.035	-0.047	-0.067	-0.019	-0.034

**Notes:** \*\*\*  $p < 0.01$ ; \*\*  $p < 0.05$ ; \*  $p < 0.1$ . Standard errors clustered at the school level and reported in parentheses. The P-value indicates the statistical difference of the impact between the training and coaching interventions.

Given the prior differences among the intervention groups evident at baseline, further controls were included to increase the precision of the estimates. The controls include learner gender, the district in which the school is located, learner performance on the baseline sub-tests, a school's average performance on the 2014 ANAs and some community-level controls. The inclusion of these controls did not have a major impact on either the size or precision of the coefficients on the coaching intervention, thereby confirming the robustness of the sustained impacts of this intervention group. However, the controls influenced the size of the coefficients of the training intervention quite remarkably. Table 17 shows that the inclusion of the controls led to a significant impact of the training intervention on the composite score, as well as most of the Setswana sub-tests. Using this model specification, the sustained impact of the training intervention appeared very similar in size to that of the coaching intervention.

Table 17: Effects on Setswana sub-tests (with controls)

VARIABLES	(1) Overall	(2) Letter recognition	(3) Word recognition	(4) Text 1 reading	(5) Text 1 comprehension	(6) Text 2 reading	(7) Written Comprehension
Training	0.172** (0.075)	0.091 (0.081)	0.180** (0.072)	0.190*** (0.069)	0.212*** (0.072)	0.121* (0.071)	0.030 (0.068)
Coaching	0.209*** (0.065)	0.158** (0.076)	0.192*** (0.062)	0.194*** (0.058)	0.244*** (0.066)	0.136** (0.060)	0.194*** (0.059)
Parents	0.0571 (0.073)	0.069 (0.079)	0.029 (0.070)	0.039 (0.068)	0.048 (0.067)	0.018 (0.067)	0.067 (0.066)
Observations	3,201	3,301	3,301	3,301	3,301	3,301	3,321
R-squared	0.170	0.113	0.153	0.161	0.132	0.140	0.035
P-value	0.568	0.354	0.857	0.942	0.633	0.809	0.021
Control mean	0.499	-0.038	-0.035	-0.047	-0.067	-0.019	-0.034

**Notes:** \*\*\*  $p < 0.01$ ; \*\*  $p < 0.05$ ; \*  $p < 0.1$ . Standard errors clustered at the school level and reported in parentheses. The P-value indicates the statistical difference of the impact between the training and coaching interventions.

Figure 6 shows the graphical representation of the impact of the interventions on the composite score. The graph differentiates between the full sample of learners and the sample restricted to only include the learners that successfully progressed to Grade 4 in the allotted timeframe (maximum dosage sample). This restriction considers the sample of learners who would have received the maximum dosage of the interventions, through having been in the classes during the year in which the teachers received the support. The solid lines represent the 95% confidence interval around the coefficients, whereas the coloured bars indicate the 90% confidence interval. Regardless of the sample considered, it is clear that both the training and coaching interventions had sustained impacts (0.17 and 0.21 respectively) that were significantly different from zero. Naturally, the maximum dosage sample showed a stronger impact, with an increase in effect sizes of 0.07 and 0.08 for each group respectively.

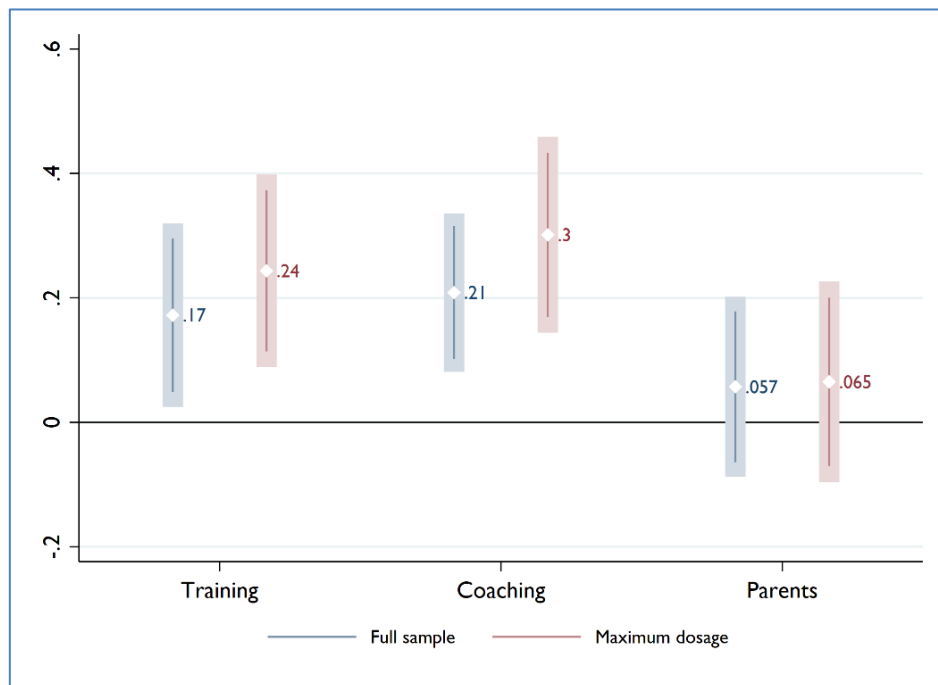


Figure 6: Graphical representation of estimated intervention

As mentioned in the description of the instruments, the Object Naming and Letter Naming sub-tests are measures of speed of lexical access, a skill essential for the development of reading fluency. As there is no consistent evidence that RAN skills can be improved and that the



interventions did not focus on this specific skill, the interventions were not expected to have had any effects on these sub-tests. Table 18 confirms that this was the case.

*Table 18: Effects on basic Setswana items (with controls)*

VARIABLES	(1) Object Naming	(2) Letter Naming
Training	0.067 (0.076)	-0.017 (0.089)
Coaching	0.023 (0.065)	0.016 (0.076)
Parents	-0.026 (0.080)	0.024 (0.084)
Observations	3,301	3,301
R-squared	0.036	0.111
P-value	0.545	0.655
Control mean	-0.000	0.030

**Notes:** \*\*\* p<0.01; \*\*p<0.05; \*<0.1. Standard errors clustered at the school level and reported in parentheses. The P-value indicates the statistical difference of the impact between the training and coaching interventions.

Finally, the regression models were run on the English and Mathematics sub-tests to see whether the interventions had any significant impacts on these subject areas. At Grade 2 endline, the coaching intervention had a significant impact on learners' English proficiency, suggesting possible positive spill-over effects. The Grade 4 assessment suggests that both the training and the coaching interventions had a sustained impact on learner English proficiency. At this point, we do not have enough information to speculate whether the improvements in English proficiency are due to improved English performance in Grade 3, or whether the stronger Home Language foundation has helped the learners in the acquisition of English as First Additional Language (as suggested by the additive-bilingual approach). No significant effects were seen on Mathematics outcomes in Grade 4.

*Table 19: Effects on English and Mathematics sub-tests (with controls)*

VARIABLES	(1) English Word Reading	(2) English Text Reading	(3) English Text Comprehension	(4) English Written Comprehension	(5) Mathematics
Training	0.165** (0.068)	0.116* (0.068)	0.102* (0.060)	-0.001 (0.067)	-0.030 (0.075)
Coaching	0.125** (0.058)	0.083 (0.053)	0.106* (0.055)	0.0481 (0.057)	-0.008 (0.062)
Parents	0.045 (0.062)	0.016 (0.061)	-0.018 (0.055)	-0.089 (0.056)	0.035 (0.074)
Observations	3,301	3,301	3,301	3,321	3,321
R-squared	0.158	0.146	0.113	0.045	0.045
P-value	0.543	0.597	0.946	0.500	0.770
Control mean	-0.023	0.000	0.003	0.034	0.027

**Notes:** \*\*\* p<0.01; \*\*p<0.05; \*<0.1. Standard errors clustered at the school level and reported in parentheses. The P-value indicates the statistical difference of the impact between the training and coaching interventions.

## HETEROGENEOUS IMPACTS

Similar to previous analysis conducted, and as was specified in the pre-analysis plan, learner-level intervention heterogeneity based on learner gender, learner age, and the initial performance of the learner at the start of Grade 1 was investigated. These effects were

investigated for both the training and coaching groups. Given the lack of any impact on the Parental Involvement group, this analysis was not conducted for this intervention group.

After two years of the intervention, differential effects were found on four characteristics. It was found that the effective coaching intervention helped boys catch up some way to girls; that the impact of all three interventions were more concentrated in the urban township areas; that the middle to top performing learners benefitted the most from the coaching intervention; and that the teacher support interventions had the largest influence on relatively large classes (38 – 45 learners).

The same heterogeneous effects were investigated in Grade 4, as shown in Tables 20 and 21. Each column in these tables shows the results of separate regressions run on the composite score, but each including a specific characteristic, as well as the interaction effect between the intervention group and the specific characteristic. For example, column (1) in Table 21 shows the results of a regression run on the composite score, including the intervention dummies, a variable accounting for learner gender, as well as the interaction variables between the intervention groups and the learner gender. The results suggest that although female learners were doing better in the training and coaching intervention groups, we cannot say with any certainty whether the boys have been catching up with them. Table 21 further suggests that the only possible differential impacts based on learner level characteristics are with regard to the learners' literacy abilities as measured at baseline.

The second consideration was whether the interventions benefitted weaker or stronger learners differently. There is often large variation in performance among learners in the same grade and the intention of the interventions was to benefit learners at all levels. The regressions that were run in columns (2) and (3) of Table 20 suggest that there may have been some differential impact based on learners' literacy proficiency at baseline.



Khulisa: Assessment

Table 20: Intervention effect by learner level characteristics

VARIABLES	(1) Learner: Female	(2) Learner: Baseline Score	(3) Learner: Baseline Score Squared
Training	0.219** (0.084)	0.168** (0.075)	0.218*** (0.074)
Coaching	0.225*** (0.081)	0.207*** (0.066)	0.206*** (0.065)
Parents	0.027 (0.082)	0.058 (0.074)	0.070 (0.075)
Training x group	-0.102 (0.090)	-0.024 (0.087)	0.034 (0.107)
Coaching x group	-0.033 (0.099)	0.028 (0.056)	0.077 (0.102)
Parents x group	0.067 (0.092)	0.041 (0.057)	0.094 (0.097)
Training x group squared			-0.074** (0.038)
Coaching x group squared			-0.005 (0.028)
Parents x group squared			-0.008 (0.026)
Observations	3,201	3,201	3,201
R-squared	0.171	0.170	0.179

To further investigate whether weaker or stronger learners benefitted more from the interventions, a quantile regression approach was taken. This approach estimates the effects of the interventions at various points in the distribution of the Grade 4 learner performance. This approach evaluates the impact, for example, on the 10<sup>th</sup> percentile, on the 20<sup>th</sup> percentile, the 30<sup>th</sup> percentile, and so on. Figures 7 and 8 show the differential impact of the training and coaching interventions on the various points in the performance distribution, respectively. The solid line in these graphs shows the overall average intervention effect, whereas the green line shows the intervention effect at a particular point in the performance distribution. The grey areas depicts the 95% confidence interval around the estimates. Both graphs show the differential impact for both the full sample of learners, as well as the sample of learners who are considered to have received maximum dosage. Figure 7 shows that the training intervention had a sustained positive impact for all learners on the performance distribution, but the impact may have been marginally stronger for learners at the middle to lower end of the performance distribution. Only considering learners that received maximum dosage suggests that the sustained impacts are concentrated among the lower half of the performance distribution, with the poorest performing 20-30% of learners having benefitted the most. The top 40% of learners may not have shown any sustained positive benefits from the intervention.

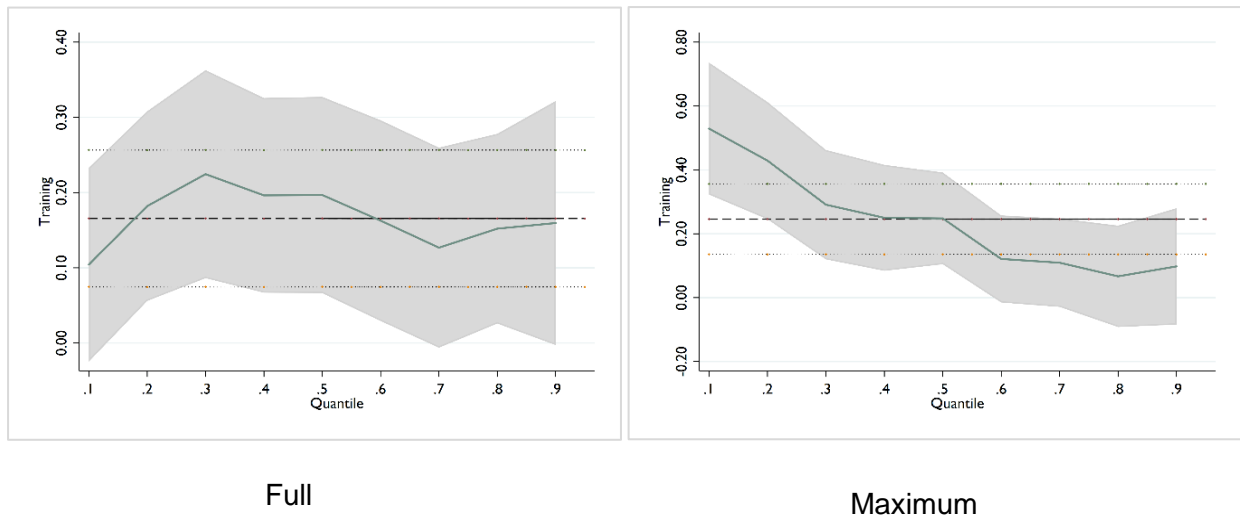


Figure 7: Quantile regression of training intervention<sup>11</sup>

Figure 8 tells a similar story to Figure 7, with learners at the lower end of the performance distribution displaying the strongest sustained benefits of the coaching intervention. However, the difference lies with the top performing learners, who also showed significant positive sustained benefits. These results seem to suggest that both the training and the coaching interventions seemed to have been equity-enhancing, by benefitting the weaker performing learners marginally more.

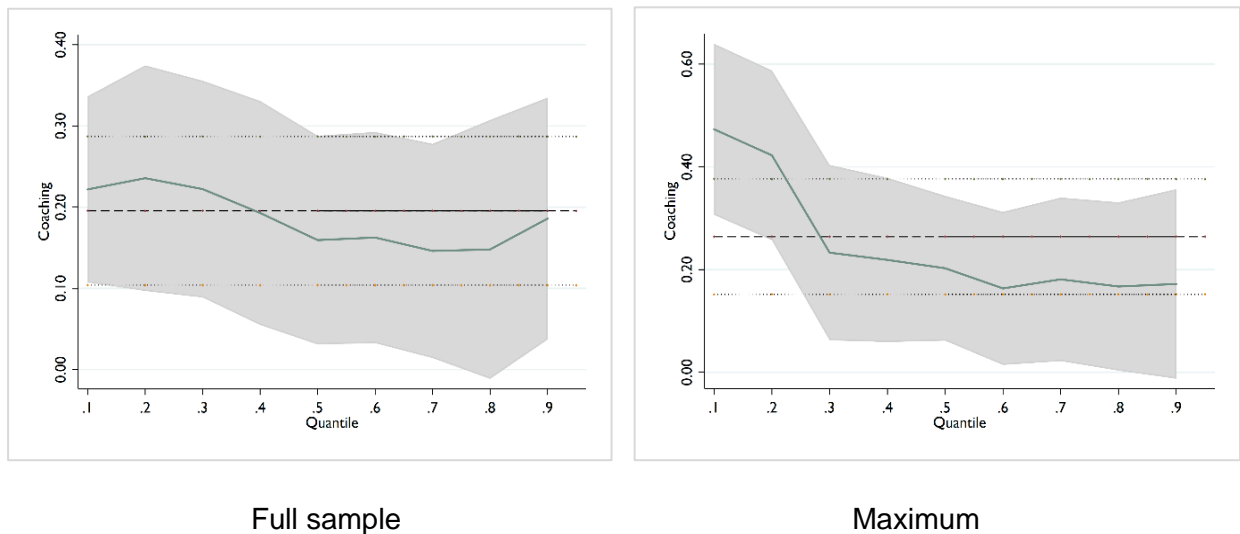


Figure 8: Quantile regression of coaching intervention

Further differential impacts were also investigated based on various school-level characteristics. These included whether a school is based in an urban or rural area, the district in which a school is based and whether the school is in a good condition based on a school functionality index.

In Wave 4, a new school functionality instrument was administered which required fieldworkers to observe and rate various aspects of the school. The responses provided reflect the judgments of the fieldworker, where each fieldworker's ratings and responses are informed by

<sup>11</sup> The horizontal line shows the average effect size.

his or her own frame of reference as to acceptable quality standards. Therefore, these are not objective measures of quality. Nevertheless, we explore whether any differential effects are evident, based on school functionality. To quantify school functionality, we construct an index based on the following observed factors:

- The school has flush toilets;
- The school periphery is secure;
- security is working well;
- Learners and teachers are safe and secure;
- The school area is clean;
- All or most classes have teachers actively teaching and engaging learners; free play is supervised;
- Teaching and learning materials are well organized and accessible;
- There is available drinking water (a water fountain, tap water, or water buckets and cups);
- There is evidence that the school governing body meets regularly, has meeting minutes, and they make meaningful decisions<sup>12</sup>.

Table 21 reports the heterogeneous impacts on the school level characteristics. No differential treatment effects were found on any of these characteristics, except on the school functionality index. The results could suggest that the sustainability of the training intervention may have been stronger in schools with a higher school functionality score, though this was only significant at a 90% confidence level.

*Table 21: Intervention effect by school level characteristics*

VARIABLES	(1) School: Rural	(2) School: District	(3) School: Good condition	(4) School: Functionality Index
Training	0.214 (0.140)	0.130 (0.082)	0.144 (0.091)	0.186** (0.079)
Coaching	0.320*** (0.105)	0.184*** (0.070)	0.155** (0.079)	0.198*** (0.068)
Parents	0.106 (0.109)	0.014 (0.083)	0.007 (0.083)	0.069 (0.074)
Training x group	-0.056 (0.155)	0.256 (0.194)	0.087 (0.147)	0.139* (0.075)
Coaching x group	-0.148 (0.128)	0.151 (0.163)	0.177 (0.131)	0.087 (0.068)
Parents x group	-0.066 (0.137)	0.233 (0.165)	0.214 (0.171)	0.041 (0.077)
Observations	3,201	3,201	3,167	3,061
R-squared	0.171	0.171	0.171	0.180

**Notes:** \*\*\* p<0.01; \*\*p<0.05; \*<0.1. Standard errors clustered at the school level and reported in parentheses. The P-value indicates the statistical difference of the impact between the training and coaching interventions.

<sup>12</sup> It is not clear how fieldworkers determined whether SGBs make 'meaningful decisions'.

## COMPARING WAVE 3 RESULTS WITH WAVE 4 RESULTS

Table 22 repeats the above analysis with the preferred specification but restricts the sample to the same set of learners who were assessed at the end of Grade 2 (wave 3) and the end of Grade 4 (wave 4). This allows for direct comparisons of the relative magnitude of effect sizes. The results of the sub-tests are reported in their raw score, rather than a standardized score; the comprehension test is coded as the proportion of the questions answered correctly. The bottom row in the table reports the average scores in the control group, which also provides some indication of progress among learners in the control group over the same period.

For example, at the end of Grade 2 learners in the control group on average read 18.9 WCPM. In the Grade 4 assessment, these learners read 29.8 WCPM, signalling learning gains of about 11 words, between Grade 2 and Grade 4. The coefficients on the different interventions help quantify the additional learning gains made in the intervention groups, over and above the expected gains (as measured by the control group). The coefficients on the training group indicate that the learners in this group read about 2.3 words more than the control group (i.e. about 22 words per minute) at the end of Grade 2, and about 3.3 words more (i.e. about 33 words per minute) at the end of Grade 4. For learners in the coaching group the gains were about 3.7 words more (i.e. about 23 words per minute) at the end of Grade 2, and 3.8 words more at the end of Grade 4.

Table 22, therefore, shows that learners in the three different intervention groups experienced very different trends between these two years, which also varied by indicator of reading proficiency. The learning gains that learners in the coaching intervention experienced by the end of Grade 2 did not increase or decrease by Grade 4. In contrast, learners in the training intervention seem to have learned more relative to learners in the control group by the end of Grade 4, compared to Grade 2. The small gain in the parental involvement intervention had eroded by half at the end of Grade 4.

Turning to the sub-domains of reading proficiency, training had somewhat “caught up” with coaching in the domains of word recognition, paragraph reading, and comprehension, although the effects continue to be slightly smaller compared to coaching. Perhaps the benefits of the coaching intervention had plateaued, which allowed learners in the Training group to eventually catch up; or the lower quality of the training intervention just meant that it took longer for learners to master reading proficiency.



Table 22: Comparing learning gains between Grade 2 and Grade 4

VARIABLES	(1) Aggregate	(2)	(3) Letter recognition	(4)	(5) Word recognition	(6)	(7) Paragraph reading	(8)	(9)	(10) Comprehension	(11)
Training	0.137* (0.082)	0.184** (0.077)	0.955 (2.447)	1.761 (1.668)	2.293 (1.412)	3.275** (1.291)	3.633* (1.917)	6.691*** (2.388)	5.166* (2.864)	0.032 (0.026)	0.048*** (0.017)
Coaching	0.238*** (0.078)	0.229*** (0.067)	4.521* (2.558)	3.352** (1.561)	3.708*** (1.285)	3.819*** (1.146)	5.433*** (1.705)	7.232*** (2.039)	6.433*** (2.458)	0.078*** (0.024)	0.054*** (0.016)
Parents	0.119 (0.078)	0.060 (0.076)	2.518 (2.381)	1.454 (1.618)	1.409 (1.355)	0.881 (1.295)	1.977 (1.850)	1.943 (2.390)	1.277 (2.815)	0.025 (0.024)	0.012 (0.016)
Grade	2	4	2	4	2	4	2	4	4	2	4
Observations	2,946	2,946	2,946	2,946	2,946	2,946	2,946	2,946	2,946	2,946	2,946
R-squared	0.174	0.167	0.149	0.113	0.160	0.149	0.155	0.159	0.139	0.123	0.128
Control mean	7.09e-10	-8.82e-10	39.04	40.26	18.91	29.80	24.48	47.22	55.56	0.309	0.297

**Notes:** \*\*\* p<0.01; \*\*p<0.05; \*<0.1. Standard errors clustered at the school level and reported in parentheses. The P-value indicates the statistical difference of the impact between the training and coaching interventions.

## SUSTAINABILITY EVALUATION: GRADE 3 RESULTS

This section of the report evaluates whether the impact of the interventions are sustained on learner outcomes one year after teachers received additional training and support. The evaluation of programme sustainability in this section involves two parts:

- 1) **An analysis of Grade 3 learner test performance.** The Grade 3 learner sample are a distinct cohort from the original baseline sample but are an informative group to follow. These learners in the training and coaching intervention schools would likely have been taught by teachers previously exposed to intervention programmes. In a sense these learners are indirectly impacted by the interventions. Evaluating Grade 3 performance therefore provides information on the sustainability of the interventions on teacher instructional practice.
- 2) **An analysis of measures of teacher instructional practice.** If interventions have had sustained impacts on teachers a year after implementation, one would expect them to engage in better practices or at least have a better knowledge of good instructional practices in the teaching of reading compared with teachers who have never been exposed to the interventions. Analysing responses to various questions administered to teachers, and observations of work done in learner workbook and exercise books, yields suggestive evidence of changes in practice.

### PART 1: AN ANALYSIS OF GRADE 3 LEARNER TEST PERFORMANCE

#### Summary Statistics

Although this report is primarily concerned with intervention effects, the new Grade 3 sample data provides a novel data source to understand to what extent learners have mastered Home Language literacy skills just before a significant grade transition involving being taught in English language in Grade 4. The curriculum assumes that by the end of Grade 3 learners have mastered basic literacy skills, reading in Home Language and in English. The summary statistics for the entire Grade 3 sample's performance in Table 23 highlights that this is not the case.

On average, these Grade 3 learners only read 39 WCPM from the Setswana text reading passage. In English, the same learners on average read 23 WCPM. Using a threshold<sup>13</sup> of 40 WCPM in English to distinguish readers from non-readers, then three quarters (76%) of the Grade 3 sample remain non-readers in English towards the end of Grade 3.

Significant floor effects were observed on the English oral reading fluency (ORF) tests, both the Setswana and English ORF comprehension and written comprehension, with floor effects most pronounced in English ORF and the related oral reading comprehension test. Low levels of oral reading fluency and comprehension are expected, given learners' inability to master the basics of decoding. Within the minute allocated for the letter recognition task, learners were on average able to correctly read only 42 Setswana letters, and on average they correctly recognised 26 familiar words in Setswana.

---

<sup>13</sup> This is an arbitrary threshold set for the purposes of explaining the interpretation of the table. Once reading norms have been developed for EFAL, thresholds that are more meaningful could be used.

Table 23: Summary statistics - Grade 3 sustainability assessment

	N	Mean	Min	p10	p25	p50	p75	p90	Max <sup>14</sup>
Object Naming	2116	17.9	0	12	15	18	21	24	36
Letter Naming	2116	25.0	0	13	20	27	32	36	36
Setswana Letter Recognition	2116	42.3	0	15	28	43	56	67	110
Setswana Familiar Word Recognition	2116	26.0	0	1	10	29	39	47	68
Setswana Text Reading (ORF)	2116	38.7	0	0	14	40	59	72	159
Setswana Text (ORF) Comprehension	2116	2.0	0	0	0	2	3	4	8
English Word Reading	2116	20.2	0	0	2	16	32	47	104
English Text Reading (ORF)	2116	23.4	0	0	0	14	39	63	126
English Text (ORF) Comprehension	2116	0.6	0	0	0	0	1	2	7
English Receptive Vocabulary	2116	3.8	0	2.5	4	4	5	5	5
Setswana Written Comprehension	2099	1.7	0	0	0	2	3	4	7
English Written Comprehension	2099	1.0	0	0	0	1	2	2	4
Mathematics	2099	1.8	0	0	1	2	3	4	8

### Descriptive Statistics by Intervention Group

There is a suggestion of marginally higher reading fluency in both Setswana and English for learners in the coaching schools relative to the control schools, as seen in Table 24. Grade 3 learners in the coaching schools on average read 3 additional words correctly from a Setswana passage and 1 additional word from the English passage when compared with learners in the control group. Figure 9 and Figure 10 provide a graphical representation of the percentage of learners who were reading a certain number of words correctly per minute from the Setswana text and the English reading passage respectively.

However, there are few other sub-test areas where educationally significant differences in reading or literacy skills were observed across intervention groups and the control group. It is then not surprising that a composite score<sup>15</sup> of Grade 3 performance in Setswana reading is on average very similar across the control and three intervention groups. However, as shown in Figure 11, learners in the coaching schools who are performing between the 20<sup>th</sup> and 60<sup>th</sup> percentiles performed slightly better than learners in the same performance percentiles in the control schools.

<sup>14</sup> The maximum score obtained is the same as the maximum possible score in all but one case. For Setswana familiar word recognition the maximum possible words correct is 70.

<sup>15</sup> The composite score is calculated using a principal components analysis to combine total scores on Setswana letter recognition, Setswana familiar word recognition, Setswana reading passage 1 and related comprehension passages as well as the total score on the written assessment.

Table 24: Averages on sub-tests by intervention, Grade 3 sustainability assessment

Mean	N	Control	Training	Coaching	Parents
Object Naming	2116	17.95	17.29	18.08	18.15
Letter Naming	2116	25.36	24.02	25.07	25.56
Setswana Letter Recognition	2116	42.71	40.61	43.42	42.41
Setswana Familiar Word Recognition	2116	25.90	25.35	26.88	25.96
Setswana Text 1 Reading (ORF)	2116	38.13	38.94	41.21	37.15
Setswana Text (ORF) Comprehension	2116	1.88	2.04	2.24	1.96
English Word Reading	2116	20.69	18.57	21.55	19.79
English Text Reading (ORF)	2116	24.12	20.91	25.60	22.61
English Text (ORF) Comprehension	2116	0.64	0.52	0.69	0.62
English Receptive Vocabulary	2116	3.91	3.72	3.95	3.73
Setswana Written Comprehension	2099	1.72	1.70	1.89	1.65
English Written Comprehension	2099	1.06	0.89	1.07	0.98
Mathematics	2099	1.84	1.79	1.8	1.96
Setswana Composite Score	2081	0.00	0.00	0.12	0.01

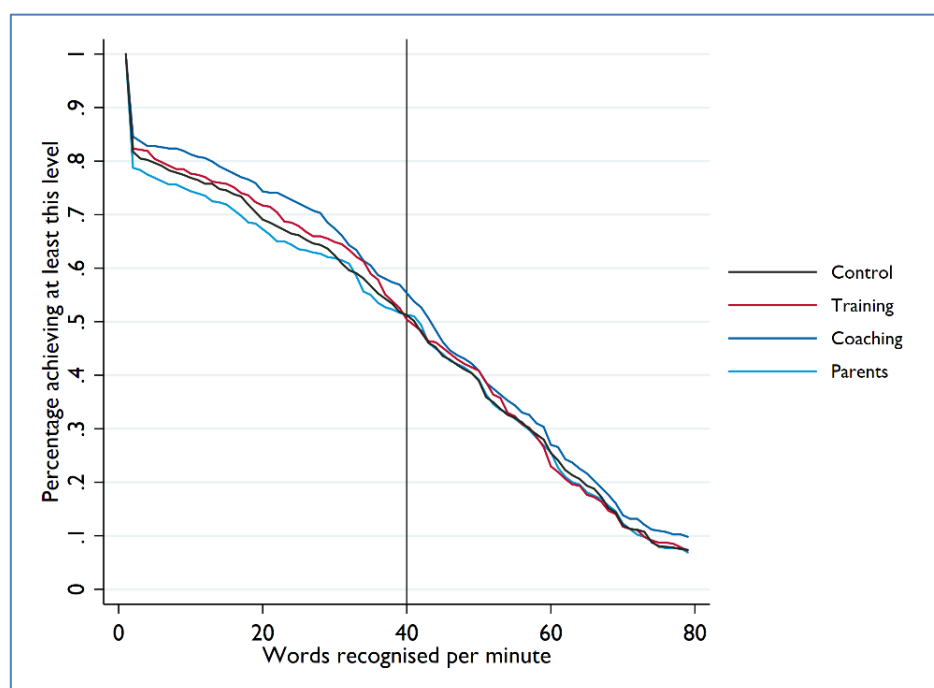


Figure 9: Grade 3 Setswana Oral Reading Fluency by intervention

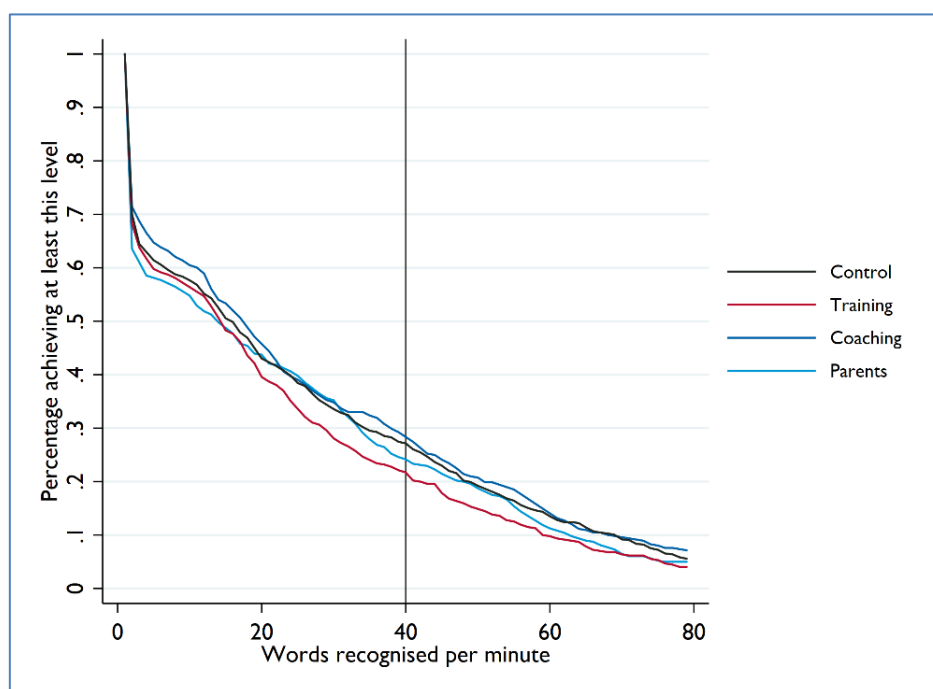


Figure 10: Grade 3 English Oral Reading Fluency by intervention

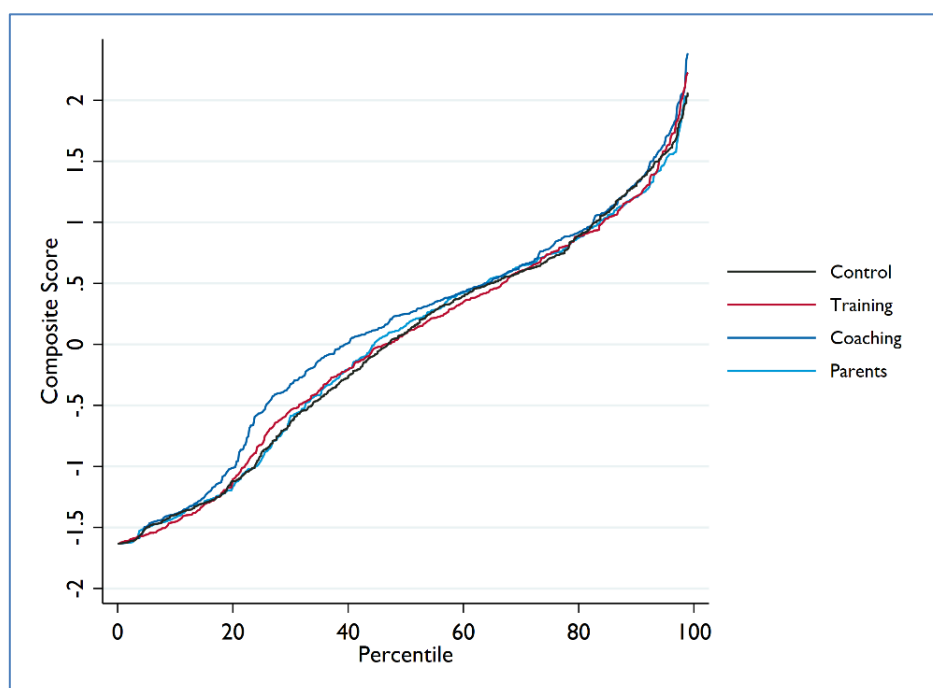


Figure 11: Setswana composite Grade 3 test score by intervention

## Main Regression Results

The descriptive results indicate that coaching potentially had sustained positive, but small effects on this new learner cohorts' reading and literacy outcomes. Regression modelling provides a more formal method for evaluating the impacts on outcomes, allows one to control for any incidental sample differences (which are bound to exist due to the limited sample size) and has the advantage of slightly reducing the size of standard errors (i.e. better precision) as

more variation in the outcome can be accounted for by control variables. Table 25 shows the results of an Ordinary Least Squares regression model estimating Setswana outcomes, controlling only for the stratification dummies. Table 26 includes school level controls including learner gender, learner age, a 2011 Census index of the wealth of the school area, rates of school enrolment in the area, and indicators for district and school performance on the Annual National Assessments. Since we do not have prior test scores for the Grade 3 sample, we are not able to control for prior literacy skills.

As expected, the estimates in Table 25 and Table 26 are suggestive of positive effects of the coaching intervention on Setswana reading and literacy outcomes, although effects are seldom significant even using a 90% confidence interval.

The most positive effect sizes for the coaching intervention are observed on estimates of Setswana ORF, Setswana oral reading comprehension and written comprehension. While the largest (and only statistically significant effect) is observed for Setswana oral reading comprehension, with a weakly significant positive effect also noted for the training intervention schools, it is difficult to interpret this result. The number of comprehension questions asked of a learner is dependent on how far a child progresses in the reading passage so that disentangling ORF from related comprehension questions is not possible.

The coefficients, expressed in standard deviations, suggest Grade 3 learners in coaching schools have a 0.1 to 0.15 standard deviation advantage in Setswana reading and literacy (as expressed in a composite score) over their peers in control schools. We note that estimated effects are typically robust to the exclusion of multi-grade schools (see Table 27). However, it is not possible to rule out that the effect sizes of estimates of both Setswana outcomes could be closer to zero. This is graphically represented in Figure 12, which shows how probable intervention effect estimates of composite measures of Setswana literacy performance are to include the value zero.

The estimated effects of the coaching programme on Grade 3 English outcomes are almost as large as on Setswana outcomes (see Table 28), suggesting positive spill-over effects of the programme beyond Setswana instruction to English. But, importantly, effects remain insignificant at the 90% confidence level. Positive but insignificant coefficients of 0.8 to 0.1 standard deviations are observed on the following English sub-tests: text reading (ORF), text (ORF) comprehension, receptive vocabulary, and written comprehension.

*Table 25: Effects on Grade 3 Setswana sub-tests (without controls)*

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	Composite	Letter recognition	Word recognition	Setswana Text Reading	Setswana Text comprehension	Written Comprehension
Training	0.002 (0.097)	-0.102 (0.102)	-0.033 (0.093)	0.028 (0.090)	0.097 (0.085)	-0.0153 (0.095)
Coaching	0.119 (0.096)	0.035 (0.087)	0.059 (0.089)	0.106 (0.091)	0.222** (0.099)	0.096 (0.093)
Parents	0.006 (0.0959)	-0.015 (0.095)	0.004 (0.095)	-0.034 (0.091)	0.047 (0.094)	-0.045 (0.088)
Observations	2,081	2,116	2,116	2,116	2,116	2,099
R-squared	0.002	0.002	0.001	0.002	0.007	0.002
P-value	0.240	0.168	0.326	0.400	0.205	0.258
Control mean	0.021	0.019	-0.006	-0.021	-0.079	-0.007

**Notes:** All tests are expressed in z-scores. \*\*\* p<0.01; \*\*p<0.05; \*<0.1. Standard errors clustered at the school level and reported in parentheses. P-value is obtained from a test of whether the coefficient on 'coaching' is significantly different to the coefficient on 'training'. Estimates only control for stratification indicators.



Table 26: Effects on Grade 3 Setswana sub-tests (with controls)

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	Composite	Letter recognition	Word recognition	Setswana Text Reading	Setswana Text comprehension	Written Comprehension
Training	0.082 (0.092)	-0.053 (0.100)	0.032 (0.088)	0.096 (0.085)	0.149* (0.083)	0.082 (0.088)
Coaching	0.150 (0.097)	0.056 (0.086)	0.087 (0.092)	0.137 (0.093)	0.248** (0.097)	0.148 (0.093)
Parents	0.055 (0.092)	0.009 (0.091)	0.044 (0.090)	0.013 (0.089)	0.079 (0.092)	0.018 (0.085)
Observations	2,081	2,116	2,116	2,116	2,116	2,099
R-squared	0.109	0.089	0.104	0.103	0.058	0.080
P-value	0.464	0.276	0.554	0.651	0.312	0.497
Control mean	0.021	0.019	-0.006	-0.021	-0.079	-0.007

**Notes:** All tests are expressed in z-scores. \*\*\* p<0.01; \*\*p<0.05; \*<0.1. Standard errors clustered at the school level and reported in parentheses. P-value is obtained from a test of whether the coefficient on 'coaching' is significantly different to the coefficient on 'training'. Estimates control for learner age and gender, district, previous ANA scores, school enrolment levels and wealth of the surrounding area around the school.

Table 27: Effects on Grade 3 Setswana sub-tests (excluding multi-grade classes, with controls)

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	Composite	Letter recognition	Word recognition	Setswana Reading	Setswana comprehension	Written Comprehension
Training	0.076 (0.101)	-0.069 (0.108)	0.022 (0.097)	0.079 (0.095)	0.142 (0.092)	0.124 (0.098)
Coaching	0.131 (0.110)	0.061 (0.097)	0.078 (0.106)	0.120 (0.108)	0.183* (0.106)	0.158 (0.104)
Parents	0.090 (0.099)	0.058 (0.097)	0.097 (0.097)	0.049 (0.095)	0.084 (0.098)	0.0548 (0.093)
Observations	1,742	1,770	1,770	1,770	1,770	1,757
R-squared	0.108	0.089	0.107	0.103	0.059	0.085
P-value	0.629	0.252	0.607	0.706	0.709	0.761
Control mean	0.017	0.022	-0.005	-0.010	-0.058	-0.014

**Notes:** All tests are expressed in z-scores. \*\*\* p<0.01; \*\*p<0.05; \*<0.1. Standard errors clustered at the school level and reported in parentheses. P-value is obtained from a test of whether the coefficient on 'coaching' is significantly different to the coefficient on 'training'. Estimates control for learner age and gender, district, previous ANA scores, school enrolment levels and wealth of the surrounding area around the school.

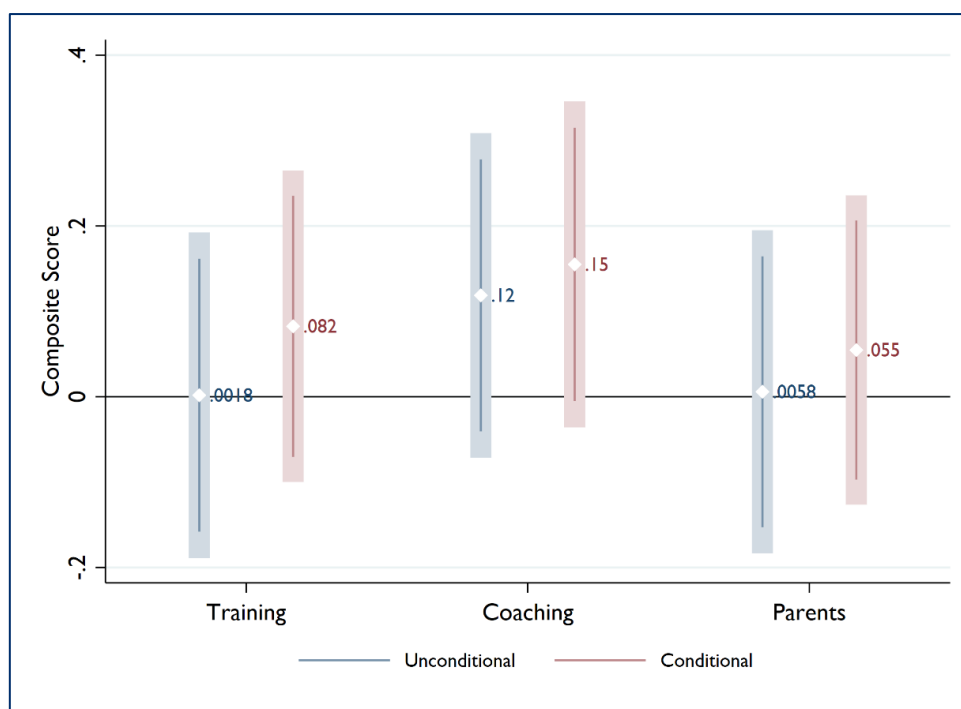


Figure 12: Graphical representation of estimated intervention effects (Grade 3 composite score)

Table 28: Effects on Grade 3 English, mathematics and other sub-test outcomes (with controls)

VARIABLES	(1) English word recognition	(2) English Text Reading	(3) English Text comprehension	(4) English Receptive Vocabulary	(5) English Written	(6) Mathematics
Training	-0.113 (0.099)	-0.037 (0.083)	-0.049 (0.078)	-0.056 (0.074)	-0.081 (0.099)	-0.074 (0.077)
Coaching	-0.020 (0.090)	0.084 (0.087)	0.089 (0.084)	0.0815 (0.087)	0.108 (0.087)	0.037 (0.081)
Parents	0.043 (0.099)	-0.002 (0.082)	-0.013 (0.077)	0.014 (0.072)	-0.100 (0.095)	-0.023 (0.086)
Observations	2,116	2,116	2,116	2,116	2,116	2,099
R-squared	0.073	0.084	0.077	0.046	0.085	0.083
P-value	0.332	0.171	0.111	0.138	0.106	0.176
Control mean	0.036	0.025	0.027	0.017	0.070	0.055

**Notes:** All tests are expressed in z-scores. \*\*\*  $p < 0.01$ ; \*\*  $p < 0.05$ ; \*  $p < 0.1$ . Standard errors clustered at the school level and reported in parentheses. P-value is obtained from a test of whether the coefficient on 'coaching' is significantly different to the coefficient on 'training'. Estimates control for learner age and gender, district, previous ANA scores, school enrolment levels and wealth of the surrounding area around the school.

### Heterogeneous Impacts

We also explored whether differences by intervention group are more or less pronounced for sub-samples of schools. Figures 13 to 15 suggest that the coaching intervention has had more sustained effects on learners' Setswana and English language proficiency in *rural* than urban schools. This is observed when comparing for example, the number of single words read correctly per minute in Setswana, English and using a composite score across the Setswana assessments. The regression results confirm the descriptive analysis. In rural schools, Grade 3 learners in the coaching schools experience a significant literacy advantage compared to their peers in the control group and other intervention arms. In urban areas, there is no evidence to suggest sustained impacts of any of the intervention programmes on learning outcomes.

Initially, the rural-urban results appear to be surprising given that larger gains at endline (Wave 3) were observed in urban schools compared with rural schools. However, the urban advantage for the original cohort sample (i.e. current Grade 4s) was not sustained in Wave 4. There are also few urban schools in the sample. Taken together, we caution against drawing conclusions about the differential efficacy of the programme across rural and urban classroom settings.

We also considered heterogeneous effects by district, learner gender and learner age. There are no notable differences across districts in terms of intervention impacts. By gender, there is weak evidence that the training programme has better outcomes for girls' reading than boys' reading (see Table 30). A very promising result is that the coaching intervention has much larger (0.3 standard deviations) and significantly positive effects on overaged learners' Setswana reading outcomes (see Table 31). Being overaged is likely to proxy for having lower baseline reading levels, suggesting that the coaching intervention has sustained positive effects for the weakest Grade 3 learners.

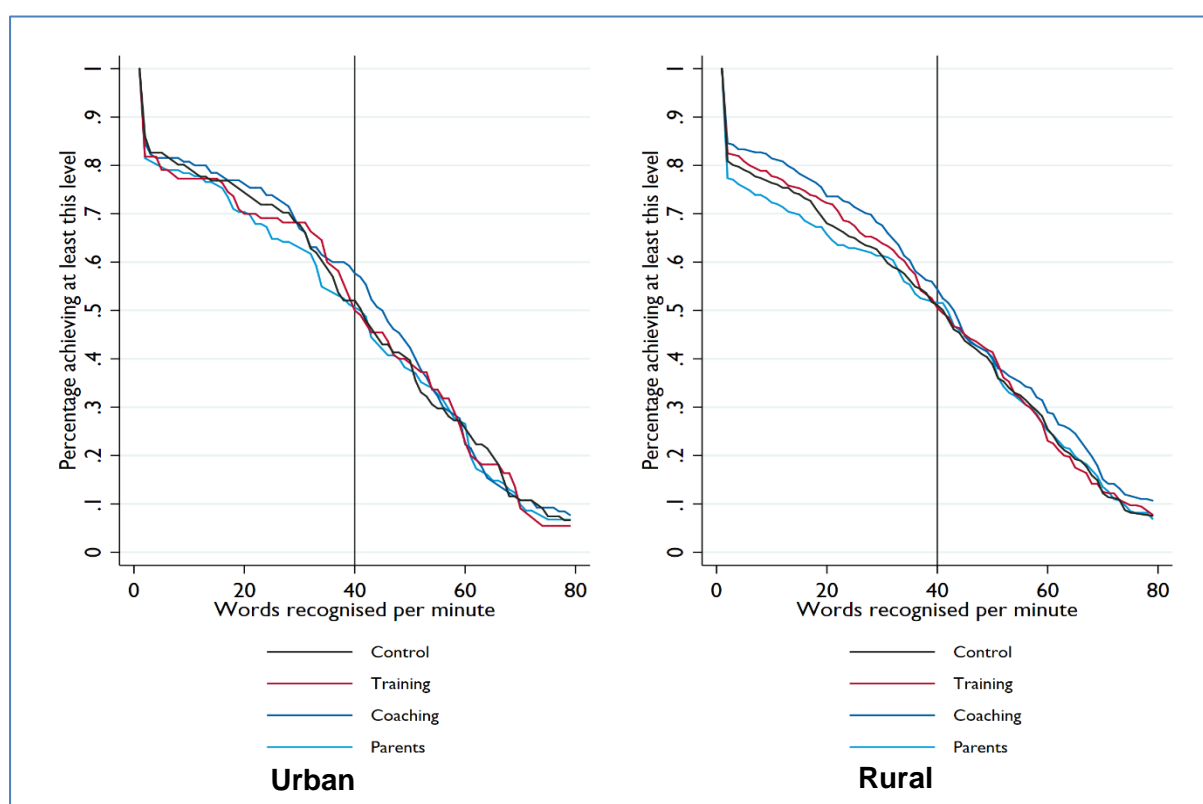


Figure 13: Setswana Oral Reading Fluency by intervention, Grade 3 sample

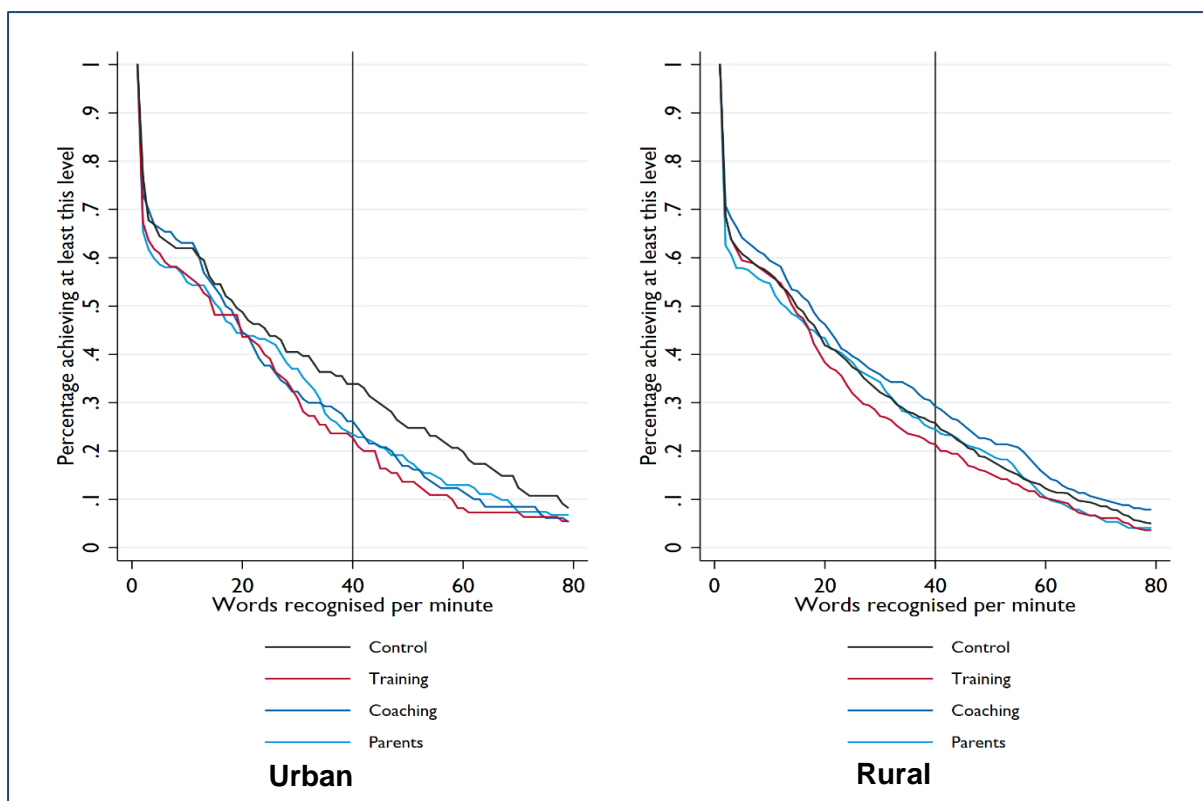


Figure 14: English Oral Reading Fluency by intervention, Grade 3 sample

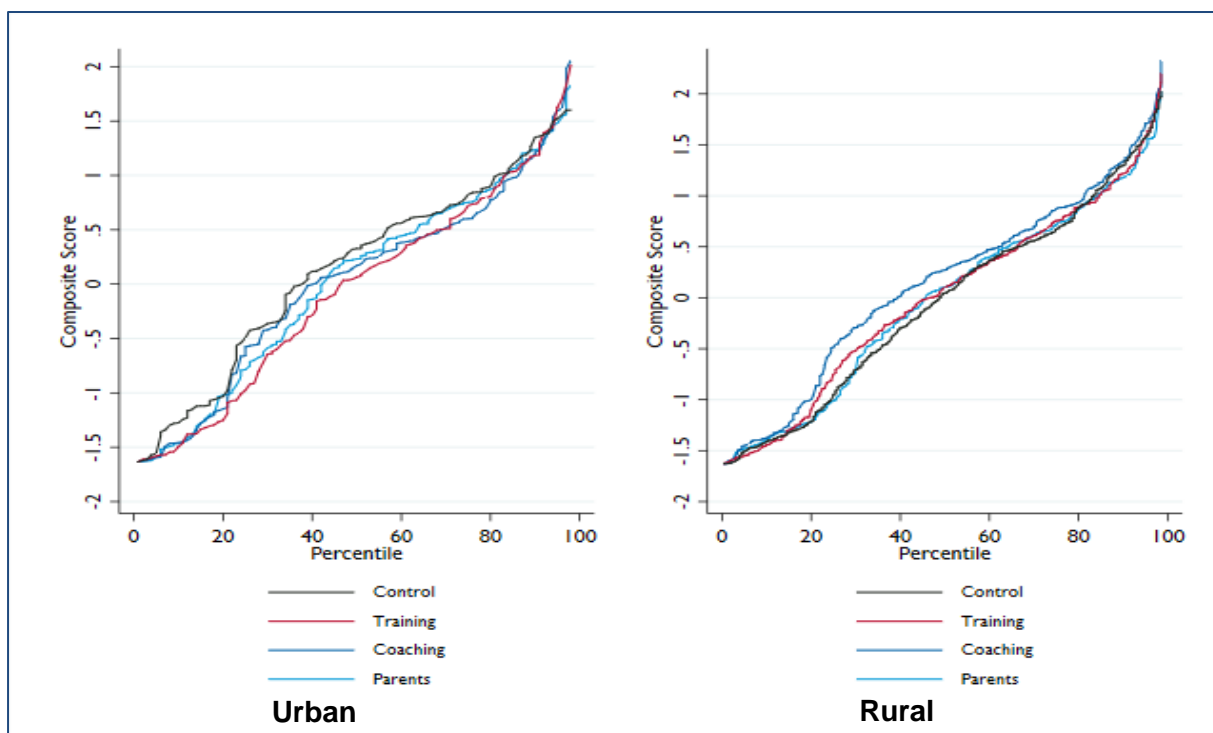


Figure 15: Setswana composite score by intervention, Grade 3 sample

Table 29: Intervention effects for Grade 3 sample, urban vs. rural.

VARIABLES	(1) Composite Score	(2)	(3)	(4)	(5)	(6)
	Urban	Rural	Urban	Rural	Urban	Rural
Training	-0.063 (0.252)	0.148 (0.097)	1.757 (5.706)	3.510 (2.736)	-8.838 (5.450)	0.060 (2.373)
Coaching	0.092 (0.178)	0.247** (0.115)	5.911 (4.664)	4.755 (3.257)	-2.003 (5.357)	4.783* (2.886)
Parents	-0.000 (0.181)	0.097 (0.120)	0.285 (4.454)	0.870 (3.429)	-6.743 (5.071)	1.054 (2.677)
Observations	517	1,564	523	1,593	523	1,593
R-squared	0.127	0.121	0.125	0.109	0.099	0.084
P-value	0.476	0.371	0.408	0.692	0.150	0.106
Control mean	0.021	0.165	0.161	0.805	0.165	0.814

**Notes:** All tests are expressed in z-scores. \*\*\* p<0.01; \*\*p<0.05; \*<0.1. Standard errors clustered at the school level and reported in parentheses. P-value is obtained from a test of whether the coefficient on 'coaching' is significantly different to the coefficient on 'training'. Estimates control for learner age and gender, district, previous ANA scores, school enrolment levels and wealth of the surrounding area around the school.

Table 30: Intervention effects for Grade 3 sustainability sample, male vs. female

VARIABLES	(1) Composite Score	(2)	(3)	(4)	(5)	(6)
	Male	Female	Male	Female	Male	Female
Training	-0.008 (0.102)	0.172 (0.113)	-0.287 (2.839)	5.762* (3.051)	-2.813 (2.267)	0.431 (3.025)
Coaching	0.145 (0.106)	0.150 (0.120)	2.758 (3.073)	4.766 (3.404)	2.543 (2.602)	2.486 (3.389)
Parents	0.034 (0.107)	0.0758 (0.115)	-0.645 (2.979)	1.419 (3.427)	-0.694 (2.472)	-0.259 (2.905)
Observations	1,048	1,033	1,069	1,047	1,069	1,047
R-squared	0.047	0.064	0.036	0.053	0.048	0.049
P-value	0.174	0.850	0.317	0.764	0.038	0.546
Control mean	0.021	0.486	0.476	0.490	0.486	0.494

**Notes:** All tests are expressed in z-scores. \*\*\* p<0.01; \*\*p<0.05; \*<0.1. Standard errors clustered at the school level and reported in parentheses. P-value is obtained from a test of whether the coefficient on 'coaching' is significantly different to the coefficient on 'training'. Estimates control for learner age and gender, district, previous ANA scores, school enrolment levels and wealth of the surrounding area around the school.

Table 31: Intervention effects for Grade 3 sample, expected age vs. overaged

VARIABLES	(1) Composite Score	(2)	(3)	(4)	(5)	(6)
	Expected Age	Over Aged	Expected Age	Over Aged	Expected Age	Over Aged
Training	0.0673 (0.0995)	0.136 (0.128)	2.686 (2.723)	3.270 (3.381)	-1.673 (2.473)	0.706 (2.872)
Coaching	0.109 (0.108)	0.327** (0.139)	3.109 (3.071)	6.908* (3.911)	2.452 (2.795)	2.666 (3.149)
Parents	0.00412 (0.0980)	0.283** (0.138)	-0.665 (2.805)	5.497 (3.574)	-1.865 (2.350)	5.196 (3.354)
Observations	1,633	444	1,656	456	1,656	456
R-squared	0.078	0.104	0.083	0.109	0.062	0.095
P-value	0.711	0.224	0.892	0.381	0.150	0.569
Control mean	0.0205	0.789	0.779	0.187	0.789	0.191

**Notes:** All tests are expressed in z-scores. \*\*\* p<0.01; \*\*p<0.05; \*<0.1. Standard errors clustered at the school level and reported in parentheses. P-value is obtained from a test of whether the coefficient on 'coaching' is significantly different to the coefficient on 'training'. Estimates control for learner age and gender, district, previous ANA scores, school enrolment levels and wealth of the surrounding area around the school.

### Benchmarking with Effect Size on the Original Cohort

Table 32 compares the intervention impact for the “later cohort” – i.e. the Grade 3 learners who entered school the year after the teachers were trained or coached – with the “initial cohort” – i.e. the learners who were present in the year that the teachers received the training or coaching. The sample is restricted to the same set of schools. This is not a perfect comparison for two reasons. First, the assessments were slightly different between years. Second, the original cohort was assessed at the end of Grade 2, whereas the “later cohort” was assessed during the third term of Grade 3. Nonetheless, it provides a better comparison than simply comparing the effect sizes from the different studies.

It is clear from Table 32 that the learning impacts are substantially smaller for the “later cohort”: The reduction is less pronounced for paragraph reading and comprehension. Again, this is encouraging, since these are the indicators that really matter for reading proficiency. However, the impacts still more than halved between the two cohorts.





Table 32: Comparing Setswana learning levels between Grade 2 and Grade 3

VARIABLES	(1) Composite	(6)	(2) Letter recognition	(7)	(3) Word recognition	(8)	(4) Paragraph reading	(9)	(5) Comprehension	(10)
Training	0.130 (0.082)	0.085 (0.093)	1.854 (2.503)	-1.063 (2.063)	1.956 (1.389)	0.649 (1.475)	3.437* (1.902)	2.772 (2.466)	0.025 (0.025)	0.030* (0.017)
Coaching	0.378*** (0.082)	0.155 (0.097)	8.268*** (2.558)	1.163 (1.780)	5.784*** (1.339)	1.475 (1.530)	8.465*** (1.786)	3.979 (2.696)	0.116*** (0.025)	0.050** (0.019)
Parents	0.116 (0.081)	0.057 (0.092)	2.746 (2.461)	0.217 (1.878)	1.241 (1.382)	0.839 (1.492)	2.076 (1.911)	0.367 (2.570)	0.021 (0.023)	0.016 (0.018)
Cohort Grade	Initial 2	Later 3	Initial 2	Later 3	Initial 2	Later 3	Initial 2	Later 3	Initial 2	Later 3
Observations	3,507	2,081	3,507	2,116	3,507	2,116	3,507	2,116	3,507	2,116
R-squared	0.092	0.109	0.088	0.089	0.085	0.101	0.081	0.103	0.066	0.058
Control mean	0	0	39.04	42.71	18.91	25.90	24.48	38.13	0.309	0.235

**Notes:** All tests are expressed in z-scores. \*\*\* p<0.01; \*\*p<0.05; \*p<0.1. Standard errors clustered at the school level and reported in parentheses. P-value is obtained from a test of whether the coefficient on 'coaching' is significantly different to the coefficient on 'training'. Estimates control for learner age and gender, district, previous ANA scores, school enrolment levels and wealth of the surrounding area around the school.

## PART 2: AN ANALYSIS OF TEACHER INSTRUCTIONAL PRACTICES

In this section, we explore whether teachers sustain their improved knowledge and teaching practice the year following the intervention. We surveyed Grade 3 teachers in all the schools where we assessed Grade 3 learners. Most of these teachers were exposed to the program the prior year. If we observe a positive significant impact on knowledge and practice, that suggests that the benefits of the program can be sustained, even after only one year of exposure. Specifically, we explore teachers' use of lesson plans and graded readers; knowledge and adherence to the correct routine; engagement in activities related to group guided reading; and their perceptions of the difficulty of instructional practices. We also explore whether teachers' practice results in evidence of more work completed by learners.

We have teacher data for 270 teachers, from 198 schools: 67 schools in the Control, 43 in the Training arm, and 44 in the Coaching and Parents arms. 251 of these teachers are teaching Grade 3 this year and also taught Grade 3 learners in the same school the previous year. We restrict analysis to these teachers, since they are the teachers who were exposed to the program the previous year.

We only report results for Coaching and Training interventions, since these are the teachers who should be affected by the program.

### Exposure to Training, Use of Lesson Plans and Graded Readers

As a starting point, Table 33 shows that Grade 3 teachers in the year following the intervention are still more likely to be exposed to the programme. Compared to teachers in the control schools, they are almost three times more likely to use lesson plans prepared by an NGO; and they are 72-76 percentage points more likely to use the Vula Bula graded readers. This is an encouraging finding.

Compared to teachers in the control schools, teachers in the Training and Coaching schools are three times and one-and-a-half times more likely to use lesson plans prepared by an NGO, and they are over four times more likely to use the Vula Bula graded readers. This is an encouraging finding. They are also more likely to state that they received training in teaching Setswana in 2017. There is also a statistically significant difference between the Training and Coaching schools in use of the lesson plans.

However, teachers in both the Training and Coaching schools state that they are *less* likely to have received professional in-serve teacher training on how to teach Setswana in 2018, with the largest reduction for teachers that received Coaching. It is possible that government is prioritising non-intervention schools to receive training in the years following the intervention. The Early Grade Reading Study was very salient and well known by provincial government. It is thus not surprising that they decided to re-allocate resources towards the control schools. Moreover, it is possible that the schools themselves, aware that they were controls schools in a study, lobbied to additional support and training.

If control schools were exposed to other useful training after the intervention, this may limit the identification of larger differences in learning and teacher practice between the control and intervention schools a year after the intervention. Any estimates should therefore be interpreted as a *lower bound*.

Table 33: Teacher's exposure to the programme by intervention, Grade 3 sustainability assessment (all teachers)

VARIABLES	(1) Use NGO lesson plans	(2) Use Vula Bula	(3) Trained in 2017	(4) Trained in 2018	(5) Feel supported- strongly agree
Training	0.430*** (0.0770)	0.747*** (0.0590)	0.137** (0.0588)	-0.194** (0.0856)	0.430*** (0.0770)
Coaching	0.199** (0.0781)	0.730*** (0.0609)	0.143** (0.0595)	-0.366*** (0.0790)	0.199** (0.0781)
Teacher observations	249	251	249	249	249
R-squared	0.227	0.627	0.047	0.123	0.227
Mean in control	0.126	0.180	0.750	0.761	0.126
P-value Train = Coach	0.0156	0.764	0.914	0.0768	0.0156

Notes: \*\*\* p<0.01; \*\*p<0.05; \*<0.1. Standard errors clustered at the school level and reported in parentheses.

## Teachers' Knowledge and Engagement in Activities Related to Reading

### Teacher Survey Questions

In the teacher survey we asked teachers how often a week they perform different reading activities. The government curriculum documents, CAPS, prescribes specific weekly frequencies of these activities: group guided reading should take place on a daily basis, shared reading three to five times a week, writing three days a week, and phonics four to five times a week. Stating that they are implementing these activities at the correct frequency suggests that they are implementing the curriculum more effectively. A critical reader might be concerned about social desirability bias. If this is the case then, at the very least, it means that they have better knowledge of what *should* take place in the classroom. As a further test of teacher knowledge, we asked teachers how often they should repeat a word when teaching phonics, and coded which teachers provided the correct answer of three. We also asked questions directly related to group guided reading: whether teachers listen to a learner read aloud on a daily basis, and whether they group learners in their class by ability.

### Results

Results are shown in Table 34. Teachers that received Coaching are more likely to implement group guided reading and shared reading at the appropriate level of frequency. We see no impacts on these indicators for teachers that received Training. However, both the Trained and Coached teachers are far more likely to get the question on phonics correct.

These results mean that teachers that received Coaching are far more likely to state that they implement group guided reading on a daily basis. This reading activity was highly correlated with learning and predictive of treatment impact for the initial cohort. However, we do not see an impact to other activities that should take place if teachers correctly implement group guided reading: individually listening to learners read aloud, and group learners by ability. Note that 88% teachers in the control already state that they group learners by ability, but our results from the classroom observations conducted in a previous study showed that teachers rarely do this in practice. It is thus possible that the teachers that received Coaching are more likely to *effectively* implement it – something we cannot test for with our data

Table 34: Use of reading instructional practices by intervention, Grade 3 sustainability assessment

	(1)	(4)	(5)	(2)	(3)	(6)	(7)	(8)
	Correct weekly frequency of performing activity							
VARIABLES	Group guided reading	Shared reading	Creative writing	Spelling	Phonics	Phonics knowledge	Read out loud daily	Stream
Training	0.064 (0.091)	0.084 (0.088)	0.041 (0.091)	0.077 (0.062)	0.010 (0.088)	0.242*** (0.082)	-0.036 (0.084)	-0.040 (0.060)
Coaching	0.151* (0.091)	0.191** (0.075)	0.153* (0.091)	0.090 (0.060)	0.110 (0.089)	0.387*** (0.079)	0.074 (0.081)	0.066 (0.047)
Observations	251	251	251	251	251	251	251	251
R-squared	0.064	0.051	0.051	0.036	0.042	0.166	0.029	0.032
Mean in control	0.326	0.517	0.461	0.798	0.461	0.393	0.618	0.876
P-value Training = Coaches	0.401	0.258	0.251	0.838	0.319	0.116	0.246	0.0744

### Perceptions of Instructional Practices

In Table 35, we looked at teachers' perceptions of different instructional practices. We hypothesized that one of the reasons that coaching was more effective than training, is that teachers who received coaching were more likely to conduct group guided reading properly, which is a very difficult technique to master. This hypothesis is supported by the data. As a starting point, we looked at the average difficulty rating for the different techniques in the control group. Creative writing is considered the most difficult, followed by group guided reading, encouraging learners to read aloud, shared reading, and phonics. Furthermore, Table 35 shows that teachers in the training arm find group guided reading more difficult relative to teachers in the control schools; but this is not the case for teachers who received coaching. Why do teachers who received training find group guided reading more difficult compared to teachers in the control schools? A possible explanation is that in attempting to implement it properly, they had to exert more effort and engage with the challenges of implementing this effectively in the classroom.

We do not find that teachers who received training or coaching the previous year are more likely to state that they enjoy teaching Setswana, but this is compared to a high base of 61% in the control.

Table 35: Teaching perceptions by intervention, Grade 3 sustainability assessment.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	How difficult to you find performing...?					Do you enjoy teaching?	
VARIABLES	Group guided reading	Phonics	Shared reading	Creative writing	Reading out loud	Setswana	English
Training	0.304** (0.136)	0.175 (0.142)	0.316** (0.153)	0.221 (0.161)	0.255 (0.164)	0.00342 (0.0843)	-0.0232 (0.0875)
Coaching	0.00220 (0.140)	-0.0616 (0.126)	0.181 (0.118)	0.0483 (0.148)	0.0580 (0.149)	-0.0593 (0.0902)	-0.0509 (0.0835)
Observations	250	249	250	250	250	250	250
R-squared	0.056	0.057	0.074	0.049	0.065	0.028	0.018
Mean in control	2.225	1.742	1.854	2.607	2	0.614	0.455
P-value Training = Coaches	0.0342	0.113	0.391	0.302	0.277	0.527	0.767

## Coverage of Work in Learners' Books

There is strong evidence that in coaching schools, learner's more evidence of work completed in their exercise books: roughly about 16-18 additional days of exercises in their books compared with learners in control schools. However, the results relating to coverage of work in DBE workbook suggest less work in DBE workbooks in intervention schools compared with control schools. Teachers exposed to the EGRS training may be using exercise books more than DBE workbooks.

Finally, in Table 36, we consider whether teachers' exposure to structured planning of time for learning is accompanied by more evidence of work in learners' DBE workbooks and exercise books. Teachers were asked to choose one learner whose workbooks and exercise books could be evaluated by fieldworkers. There is strong evidence that in coaching schools, learner's more evidence of work completed in their exercise books: roughly about 16-18 additional days of exercises in their books compared with learners in control schools. However, the results relating to coverage of work in DBE workbook suggest less work in DBE workbooks in intervention schools compared with control schools. Teachers exposed to the EGRS training may be using exercise books more than DBE workbooks.

Table 36: Work completed in learners' DBE workbooks and exercise books

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
	No. of pages with any exercises	No. of pages with any exercises	No pages showing at least one full sentence	No pages showing at least one full sentence	No. of days with any exercise	No. of days with any exercise
Training	-4.294* (2.283)	-4.182* (2.374)	-3.295** (1.371)	-3.514** (1.425)	3.917 (5.272)	5.932 (5.471)
Coaching	-1.127 (2.551)	-0.141 (2.643)	1.696 (2.863)	2.098 (3.032)	18.27** (7.967)	16.10** (8.117)
T3	-0.443 (3.021)	-0.347 (3.301)	0.658 (1.782)	0.995 (1.955)	3.076 (4.330)	5.499 (4.573)
Observations	261	242	262	243	262	243
R-squared	0.050	0.054	0.059	0.071	0.057	0.056
Mean in control	19.81	19.65	10.36	10.42	36.46	35.82
P-value Training = Coaches	0.150	0.0875	0.0920	0.0724	0.0848	0.227

Notes: \*\*\* p<0.01; \*\*p<0.05; \*<0.1. Standard errors clustered at the school level and reported in parentheses.

## CONCLUSION

After two years, the impact evaluation of the EGRS showed positive impacts on reading outcomes for two of the interventions – “teacher training”, and “training and coaching” -, both of which provided support to teachers through structured lesson plans aligned to the curriculum, integrated reading materials and professional development. The largest impact was observed for the group that received on-site coaching as the specific form of professional development, rather than centralized teacher training. After two years, the parent involvement intervention was not found to have a significant positive impact and was therefore discontinued, whereas the two teacher support interventions (“training” and “coaching”) were extended for a third year covering the Grade 3 learners of 2017.

Unfortunately, no data collection was undertaken at the end of 2017, which would have allowed one to observe the impact of the two teacher interventions after three years of implementation. However, this report presents the results of a follow-up data collection that was administered towards the end of 2018 – nearly one year after the interventions were completed. The original sample of learners who had been tracked from the start of grade 1 in 2015 were re-tested in 2018 in order to answer the research question of whether the positive impacts that were observed in 2016 had persisted. About 50% of this original sample of learners were observed to be in grade 4 (not having repeated); nearly a quarter were observed to be in grade 3 or grade 2 (having repeated at least once); and just over a quarter were not identified (probably due mainly to changing schools or being absent on the day of the survey).

In addition, a fresh sample of ten Grade 3 learners per school in 214 of the original 230 EGRS schools was tested in order to assess whether the next cohort of children also benefited from lasting changed teaching practices as a result of the EGRSs. The main reason for excluding 16 of the original schools was that these schools were by this time using multigrade teaching in the Foundation Phase meaning that the grade-specific EGRS lesson plans would not be appropriate for use.

The primary finding of this follow-up evaluation is that the impacts on the original cohort of learners have persisted, with both the “training” and “coaching” groups of children still performing better in their Home Language than those in the control group. The magnitude of the advantage held by children in these groups is roughly similar to that observed after two years of intervention. Using one method of benchmarking the effect size, one can say that those children in the “coaching” group are approximately 40% of a year of learning ahead of those in the control group. It is also encouraging that a positive spillover impact on English was again observed in the follow-up evaluation, confirming what was found in the Year 2 evaluation.

Whilst in 2016, the “coaching” intervention was estimated to be about twice as effective as the “training” intervention (and therefore most cost-effective), by 2018 the gap between “training” and “coaching” appears to have narrowed somewhat placing the two interventions in a similar range of cost-effectiveness. Overall, when one weighs the combined evidence across the various waves of data collection and across the different learning sub-domains, the evidence of a positive impact seems clearer for coaching than for training – for instance in 2018 the coaching intervention is estimated to have significant positive effects on all Setswana sub-tests but the training intervention did not register significant effects on letter sound recognition or the written comprehension test.

The evaluation on the Grade 3 sample also revealed positive estimated impacts of both the training and coaching interventions, although the effect sizes were somewhat smaller and were not consistently statistically significant across the learning domains. These positive estimates were again slightly larger and more robust to model specification for the coaching group compared to the training group. The largest effect sizes for coaching were observed on Setswana ORF, Setswana reading comprehension and Setswana written comprehension. As with the original EGRS sample, there were also positive “spill over” effects on English outcomes. There was no clear and consistent evidence of any differential effects on specific sub-groups of children or schools within the Grade 3 sample.



The interviews held with the Grade 3 teachers provided some insights into the above-mentioned impacts on learning. Firstly, teachers in the “coaching” and “training” schools were almost three times more likely, compared to the control group, to report using lesson plans provided by an external service provider and also much more likely to report using the Vula Bula reading series, which was provided by EGRS. This indicates a sustained use of the materials provided through EGRS, or at the very least (if one is suspicious of socially desirable responses) a sustained awareness of the materials they should have been using. Teachers in intervention schools were also significantly more likely to say they conduct group guided reading on a daily basis, and more likely to follow the correct routines for group guided reading and creative writing. These instructional activities are prescribed in the official curriculum but classroom observation studies indicate that they are either missing or poorly implemented in most classrooms. Supporting teachers to implement these activities were a core element of the EGRS programme.

However, it was also observed that teachers in schools that had received EGRSs were also significantly less likely to have received in-service training in 2018 (the year after EGRS finished) than teachers in the control group. This may indicate a prioritization of control schools in the support provided by district officials or by other school support initiatives such as the Primary School Reading Improvement Programme. If any such training activities had a positive impact on learning it is possible that this may have contributed to the somewhat smaller estimated effect sizes of the EGRSs on the next cohort of learners.

A first policy implication of these findings is that effective early interventions in reading may have benefits that last and can contribute to long-term improvements in educational outcomes. Secondly, the provision of support to teachers through a structured learning programme with integrated materials, which fill key gaps in the African language learning classroom learning environment, can make a significant difference. In the absence of other formal impact evaluation evidence of what makes a positive impact in South Africa’s primary schools, this is an important finding. It also confirms that a number of the Department of Basic Education’s initiatives, such as the Primary School Reading Improvement Project and the work being done by the National Education Collaboration Trust, which also make use of similar structured lesson plans, are on the right track.

However, both the training and the coaching programmes included substantial direct additional support to teachers, whether through four days of residential training per year or through monthly on-site coaching visits. This level of support is often lacking in existing support to teachers (provided through district officials and the School Management Team), and in other external initiatives. It cannot be expected, for example, that orientating district officials to core instructional methodologies or new materials, would have a similar impact on the classroom practice of teachers and hence the reading outcomes of children. Although district-level subject advisors have an important role to play in providing systemic support to schools, the low ratio of subject advisors to schools and their wider range of job responsibilities means that more direct forms of support will need to be explored. The critical importance of addressing the reading challenge in South Africa means that this is imperative, especially in the light of the evidence produced through the Early Grade Reading Study.

## **WAY FORWARD**

The positive findings of the EGRS evaluation led the DBE to make the decision to implement a second phase to the first EGRS, called the Reading Support Project (RSP), and evaluate whether the results will hold if the interventions are rolled out at scale. In addition to the Grade 3 and 4 data collected in 2018, Grade 1 baseline data for the RSP was collected during the same period. In 2019 and 2020, the two EGRS teacher interventions will be rolled out to all quintile 1 – 3 Setswana schools (263 schools) in the same two districts in the North West province. Instead of staggering the roll-out (as was done in the first phase), the interventions will be rolled out to all Foundation Phase teachers at the same time, and focus on the teaching of both Home Language (Setswana) and English as First Additional Language (EFAL).

All schools in the sample will receive the scripted lesson plans, with the integrated learning and teaching materials, as well as quarterly training on the use of the lesson plans. More than half of the

schools (n = 140) will also receive regular on-site coaching by a specialist reading coach. Eighty-two (82) of the coaching intervention schools will additionally receive professional development of school Principals and the Heads of Department for the Foundation Phase. Finally, as a layer of cross-randomization, classroom libraries will be randomly assigned to 110 schools across the two districts. This information is summarized in Table 37. The RSP will be evaluated using an impact evaluation design.

*Table 37: Reading Support Project (RSP) intervention design*

		All Schools (263)	
		Coaching Schools (140 of the schools)	
			SMT Schools (82 of the schools)
Daily lesson plans (aligned to CAPS in Foundation Phase)	x	x	x
Learning and teaching support material (Integrated with the lesson plans, including DBE workbooks, graded reading booklets, flashcards & posters)	x	x	x
Quarterly cluster training	x	x	x
Regular on-site coaching		x	x
Principal and SMT training			x
Classroom library		Random selection	Random selection



Shutterstock: Learners