The First Early Grade Reading Study

YEAR 7 REPORT

SUSTAINABILITY IMPACT EVALUATION OF THE EARLY GRADE READING STUDY (EGRS I) IN SOUTH AFRICA

basic education
Department:
Basic Education
REPUBLIC OF SOUTH AFRICA

USAID
FROM THE AMERICAN PEOPLE

AUTHORS

Dr. Jonathan Stern and Dr. Matthew Jukes, Research Triangle Institute International (RTI)
Margaret Roper, Leticia Taimo and Jennifer Bisgard, Khulisa Management Services
Benita Williams, Benita Williams Evaluation

CONTACT DETAILS

Margaret Roper
Khulisa Management Services
26 7th Avenue
Parktown North
Johannesburg, 2196
South Africa

Telephone: 011-447-6464
Email: mroper@khulisa.com
Web Address: www.khulisa.com

DISCLAIMER

The author's views expressed in this publication do not necessarily reflect the views of the United States Agency for International Development of the United States Government.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

GLOSSARY

ANA	Annual National Assessment
CAPS	Curriculum and Assessment Policy Statements
CRC	Community Reading Coach
cwpm	Correct words per minute
DBE	Department of Basic Education
EFAL	English First Additional Language
EGRA	Early Grade Reading Assessment
EGRS	Early Grade Reading Study
HL	Home Language
OLS	Ordinary least squares
ORF	Oral Reading Fluency
PED	Provincial Education Department
PIRLS	Progress on Reading Literacy Study
RCT	Randomized Control Trial
RSP	Reading Support Project
USAID	United States Agency for International Development

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, and end of Grade 2) were conducted by the Human Sciences Research Council (HSRC). The fourth wave of data (2018; end of Grade 4) 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.

To further evaluate the sustainability of the impact of EGRS I, a fifth wave of data was collected in 2021 by Khulisa. The purpose of this wave was to determine whether the results from the original cohort of learners who received the interventions would be sustained into Grade 7[1] in 2021, more than three years after they had the benefit of being taught by teachers who had received the EGRS. The overarching evaluation question:

What is the impact of the EGRS I interventions on reading outcomes in Setswana and EFAL for learners in Grade 7 (three years after the completion of the intervention)?

The results show evidence of a sustained impact from the coaching intervention on several Setswana (home language) outcomes. Although there was an average impact for all learners in the coaching intervention on only one of the Setswana oral reading passages (with marginally significant impacts on several other outcomes), the most consistent impacts were for learners who received the maximum dosage of the EGRS I intervention in Grades 1 to 3 (i.e., those who progressed, as intended, through the three years of the intervention from 2015 to 2017). Significant impacts were found for these learners on four of the seven Setswana reading outcomes (with marginally significant impacts on the remaining outcomes). This provides evidence that results from an early grade reading intervention can be sustained into upper grades but also points to the importance of dosage.

---

[1] Grade 7 was the intended grade for on track learners but some learners repeated grades and were assessed in their 2021 grade.

Interestingly, the coaching intervention also showed a statistically significant impact on retention, with a smaller proportion of attriting learners than the control group (38% versus 45%).

While English First Additional Language (EFAL) results followed a similar pattern to Setswana results (with the largest magnitude, positive impacts for the coaching intervention), there were no statistically significant long-term intervention effects on EFAL. Additionally, heterogeneous impacts were limited, with some evidence that learners with different baseline scores (from Grade 1) may have been impacted differently by the interventions.

Overall, the results are very promising for the EGRS I coaching intervention, particularly in light of the fact that few other research studies have shown such a sustained, long-term impact of an early grade reading program.

This report derives from the data collection and analysis for the Early Grade Reading Study (EGRS I), the Reading Support Project (RSP) and the Language Benchmarking study in two districts in North West, South Africa. This study was conducted for the United States Agency for International Development (USAID) and the Department of Basic Education (DBE).

INTRODUCTION

This report presents the findings from a 2021 sustainability evaluation of the Early Grade Reading Study (EGRS I). The EGRS I 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.

Purpose

This report, produced by Khulisa Management Services (Pty) Ltd. (Khulisa), is submitted under the Data Collection and Analysis for the Early Grade Reading Study (EGRS), the Reading Support Project (RSP) and Benchmarking to the United States Agency for International Development (USAID) under Indefinite Delivery Indefinite Quantity (IDIQ) Contract Number: 72067418D00001, Order Number: 72067421F00001 (Scope of Work in Annex 1).

This report derives from the data collection and analysis for the EGRS I, the RSP and the Language Benchmarking study in two districts in North West, South Africa.

A number of reports have already been published under this task order and are useful as background.

- Task Order 4 Data Collection and Analysis EGRS, RSP, Benchmark and COVID-19: Fieldwork Report [2]
- Methodology Plan and Study Protocol: Data Collection and Analysis for the EGRS, RSP and Benchmarking. [3]
- For the full instrument development process refer to the "Report on the Development of Learner Assessment Tools and Contextual Tools".[4]
- The Quality Assurance Surveillance Protocol (QASP). The QASP documents the quality assurance elements of both data collection and analysis.[5]

As part of this task order, Khulisa conducted additional research on COVID-19 in the schools and two reports were provided. The Preliminary COVID-19 Report submitted in 2021 enabled the DBE to consider the policy implications to prepare for the 2022 school year. [6] Thereafter, the Consolidated Final COVID-19 Report was submitted in 2022.[7]

---

[2] https://pdf.usaid.gov/pdf_docs/e4563ed819164a79956698c3a1998964.pdf

[3] https://pdf.usaid.gov/pdf_docs/42132810ec2c48809efe8ca11e155aff.pdf

[4] Bisgard, J. et al, 2021

[5] https://pdf.usaid.gov/pdf_docs/PA00Z8SX.pdf

[6] https://pdf.usaid.gov/pdf_docs/PA00XGST.pdf.

[7] https://pdf.usaid.gov/pdf_docs/PA00ZBHD.pdf

In addition to this impact evaluation, the data is being analysed to identify Setswana reading benchmarks and English First Additional Language (EFAL) reading benchmarks. The summary benchmark reports, learning briefs and final technical report will be published by June 30, 2022.

Background

To address the challenge of having large proportions of children not learning to read for understanding in the early grades in South Africa, the DBE initiated the EGRS in 2015.

The EGRS I was implemented in 230 quintile 1-3 schools in two districts in the North West province (districts of Ngaka Modiri Molema and Dr. Kenneth Kaunda). The EGRS I evaluated three Setswana Home Language interventions aimed at improving reading in the early grades: a teacher training intervention, an on-site teacher training and coaching intervention, as well as a parental intervention. Each intervention was implemented in a separate group of 50 schools with a further 80 control schools where ordinary schooling continued. A Randomized Control Trial (RCT) impact evaluation, 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 I 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 again 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. The interventions showed significant impacts on learner results when teachers benefited from training, coaching, and provision of learning materials, based on Wave 3 evaluation data. In 2018, a fourth wave of data was collected from the same sample of learners (who were then expected to be in Grade 4). The EGRS Wave 4 evaluation showed that the initial impacts of the EGRS I on learners' ability to read, continued one year beyond the end of the intervention.

The findings presented in this report are designed to examine the sustainability of the EGRS I evaluation based on Wave 5 data collected in 2021. The goal of this data collection was to determine whether or not the results from the original cohort of learners who received the interventions would be sustained into Grade 7[8] in 2021, four years after they had the benefit of being taught by teachers who had received the EGRS I (and 3 years after the first sustainability evaluation in 2018).

This EGRS I sustainability impact evaluation, conducted under USAID PERFORMANCE Task Order 4, provides the last opportunity to track learners who benefited from the EGRS I before transitioning to

---

[8] Grade 7 was the intended grade for on track learners but some learners repeated grades and were assessed in their 2021 grade.

secondary schools9. With this additional round of data collection, the understanding of the longitudinal benefit of early grade reading interventions can be strengthened.

Intervention design

Description of the EGRS I

This EGRS I was set up as a randomized control trial. The EGRS I included three different interventions, all aimed at improving early grade reading in Setswana.

Treatment 1: Training, scripted lessons, graded readers

Treatments 1 and 2 aimed to apply the same set of instructional practices in the teaching of home language literacy in Grade 1, 2 and 3 classrooms working in one grade at a time. Both treatments therefore provided teachers with clearly scripted lesson plans, which were aligned to 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 learning support materials including the government-provided workbooks as well as certain additional materials (graded reading booklets, flash cards, posters, etc.), which were provided through the EGRS I. The graded reading booklets provided a key resource for the teacher to use in group-guided reading and individual work so as to facilitate reading practice at an appropriate pace and sequence of progression.

Treatment 1 trained the teachers on how to use the lesson plans and accompanying materials through central training sessions, each lasting two days, and occurring twice yearly.

Treatment 2: Reading Coaches, scripted lessons, graded readers

Exactly the same set of instructional materials (scripted lesson plans, graded reading booklets and other materials) were provided to Treatment 2 schools. However, instead of central training sessions, one day training/orientation was provided at the start of each term, accompanied by ongoing support to teachers consisting of regular (monthly) on-school coaching from specialist "reading coaches" visits. In addition to these on-site visits, there were occasional needs-based workshops with the coach and a small cluster of nearby Treatment 2 schools.

Treatment 3: Parental involvement

Treatment 3 was designed to promote parental involvement to support their children's reading progress. At each of the 50 schools in this treatment arm a Community Reading Coach (CRC) was recruited. The CRC was identified through communication with the school principal who recommended a suitably qualified, but available person in the community. The CRCs attended a one-day training session facilitated by the service provider (Class Act) at the start of each school term (quarterly). The CRCs were trained to deliver weekly training sessions for Grade 1 parents at their respective schools. A total of 30 sessions were scheduled for each year covering a total of 10 topics.

---

9 These data were initially intended to be collected with Grade 6 learners in 2020 but COVID-19 related delays led to a 2021 data collection with learners in Grade 7.

The parental involvement intervention arm was prematurely ended in 2016, after two years of implementation working first with Grade 1 parents and then Grade 2 parents. The parents of Grade 3 learners were not part of the intervention.

PRIOR EVALUATION FINDINGS

Findings from the EGRS I Wave 4 Report provided evidence of sustained impact of the 'training' and 'training and coaching' interventions, for Grade 4 learners one year after the end of the intervention. These learners were found to be outperforming control group learners in home language reading proficiency. The magnitude of the advantage held by learners in the coaching group in 2018 was similar to that observed after two years of intervention (in 2016). However, the "training and coaching" intervention was estimated to be about twice as effective as the "teacher training" intervention in 2016 but, by 2018, the gap had narrowed. This put the two interventions in a similar range of cost-effectiveness for sustained impacts (while initial impacts found the coaching intervention to be more cost-effective).

Overall, although Wave 4 results showed an impact for both interventions, the evidence was slightly clearer for the "training and coaching" intervention. Furthermore, there was evidence of a positive spill over effect on English reading scores (which confirmed trends from initial impact evaluation findings).

METHODOLOGY

Evaluation Design

The EGRS I was designed as an RCT with three treatment arms and one control arm. A total of 230 schools were randomly assigned to one of the intervention groups (50 schools each), or the control group (80 schools). To evaluate the effectiveness of the interventions, a random sample of 20 learners per school was selected in Grade 1 in 2015. These same learners were tracked into Grade 2 (Wave 3), then into Grade 4 (Wave 4) and finally into Grade 7 (Wave 5). Learners' reading proficiency was assessed at each wave.

Treatment Assignment and Random Selection for EGRS I

The sampling process for the EGRS I began with 458 primary schools registered in 2014 administrative data in the districts of Dr. Kenneth Kaunda and Ngaka Modiri Molema. The sampling started by excluding relatively affluent schools (those in quintiles 4 and 5). Next, schools were excluded in which the language of instruction in the Foundation Phase was not Setswana. Schools were also excluded if they were missing in the 2014 Annual National Assessment (ANA)[10] dataset. An additional eight schools were excluded because they took part in the EGRS I pilot. Small schools (fewer than 20 Grade 1 enrolments) were also excluded because many of them practice multi-grade teaching rendering the scripted lesson plans less appropriate. Larger schools (more than 180 Grade 1 enrolments) were also excluded to limit intervention costs. Three more schools were excluded after the North West Provincial Education Department (PED) checked the list of schools and found specific problems with these schools (e.g., the school had been closed down, or a particular conflict around school management was occurring in a school). After all of these exclusions, 235 eligible schools remained. Using a random number generator, five schools were selected as possible replacement schools, with the remaining 230 schools constituting the main sample.

To increase power and assure balance between treatment arms, stratified randomization was performed. Ten strata of 23 similar schools each were created, based on school size, socio-economic status, and previous performance in the Annual National Assessments. Within each stratum, five schools were randomly assigned to each treatment group and eight to the control group. In total, 50 schools were assigned to each treatment group and 80 to the control group. Sample size calculations indicated that this sample is sufficient to identify a minimum effect size of 0.21 standard deviations when comparing a treatment group with the control group and a minimum effect size of 0.23 standard deviations when comparing two treatment groups. These calculations are based on data collection from 20 Grade 1 learners and assume a 95% confidence interval, an alpha value of 0.8, an intra-class correlation coefficient (rho) of 0.3 and a correlation between pre- and post-test scores of 0.7.

Sampling for EGRS I Sustainability Impact Evaluation

Since the start of EGRS I, one school closed, leaving 229 of the original 230 schools. These 229 schools formed the proposed sample for data collection. Furthermore, data were expected to be collected from all learners for whom data was collected in Grade 1 and who were present at the school on the day of assessment. The original sample was 20 learners per school. In 2018, an average of 14.5 learners per school were tracked. The expectation was that this sample would be further reduced to 12 learners per school in 2021. The minimum detectable effect size for this reduced sample was estimated to be 0.23 standard deviations.

---

[10] The Annual National Assessments (ANA) are standardized national assessments for languages and mathematics in the senior phase (Grades 7 - 9), intermediate phase (Grades 4 – 6) and in literacy and numeracy for the foundation phase (Grades 1 – 3). They were administered between 2012 and 2015.

Data Collection and Final Sample

Data collection for this evaluation occurred from 7-30 September 2021. Two main types of data were collected for this study: 1) individual and written learner assessment data; and 2) contextual data. The contextual data provides background information on learners, teachers, principals, parents and schools.

Although the target sample for this evaluation was 229 schools, the team was able to successfully collect data from Grade 7 learners in 216 schools11. The final sample consisted of 2,439 learners, with the number of respondents for each tool shown in Table 1. All learner assessment data were collected via Tangerine, while all contextual data were collected via Kobo Toolbox.

Table 1: Final sample size for EGRS evaluation by instrument

Instrument	Number of observations
Grade 7 individual assessment	2,351
Grade 7 HL written assessment	2,260
Grade 7 EFAL written assessment	2,214
Learner asset list	2,357

Research Questions and Analysis

The main evaluation question for this sustainability impact evaluation is:

What is the impact of the EGRS I interventions on reading outcomes in Setswana and EFAL, for learners in Grade 7 (three years after the completion of all interventions)?

More specifically, we are interested in examining whether or not the initial impact of the program, which was found to be sustained for the coaching and training interventions through Grade 4, is further sustained until Grade 7.

For the analysis of EGRS I impact we conducted longitudinal analyses among a sample of learners who were available during the Wave 1 and Wave 5 data collections. Initial analyses are unadjusted for any covariates, while final models all include covariates that control for learner and school level

---

[11] The team was unable to collect data from 13 schools due to a combination of reasons: lack of Grade 7 at the school, uncooperative principals, rotational schedules with Grade 7 out during data collection, permanent school closure, temporary school closure and/or no matched learners from the linking sheet in Grade 7. https://pdf.usaid.gov/pdf_docs/e4563ed819164a79956698c3a1998964.pdf

characteristics that are unrelated to treatment assignment but may impact learner performance. The aim of these final models is to increase precision and statistical power, while accounting for any incidental differences that may exist between treatment groups.

Our main estimating equation is:

$$y_{isb1} = \beta_0 + \beta_1 T + X'_{isb0} \Gamma + \rho_b + \varepsilon_{isb1}$$

Where y_{isb1} is the outcome indicator of interest for learner i in school s and strata b; T is the treatment dummy which is equal to one in either of the treatment arms; ρ_b refers to strata fixed effects; X'_{isb0} is a vector of control variables; and ε_{isb1} is the error term clustered at the school level. To estimate the respective impacts of the three interventions, we restricted the sample to the control schools and the schools from the relevant treatment group.

We also examined the data for evidence of differential attrition. Attrition threatens validity of the evaluation if it differs between treatment groups (e.g., if struggling learners in the intervention group are less likely to drop out of school than the control group). We therefore assess attrition in two stages. In the first stage, we assess whether learners included in the final sample differ in baseline characteristics from those who were included at baseline. In the second stage, we look for evidence of differential attrition.

The primary outcomes of interest for this evaluation are reading proficiency in Setswana home language (HL) and EFAL. For reading proficiency we measured a range of reading skills: vocabulary, reading fluency, and reading with comprehension. Since a single composite score across subtasks will not provide meaningful interpretation, results throughout this study are presented for a simplified composite score (for orally administered tasks only), as well as for individual subtasks.

Instruments

The EFAL assessment tasks administered for this evaluation were adapted from existing instruments, such as the Early Grade Reading Assessment (EGRA), as well as released items from the Progress in Reading Literacy Study (PIRLS) and extended to account for the more advanced reading ability of learners in Grade 7. The Setswana home language assessment tasks were newly developed for this evaluation (and the accompanying benchmarking activity).

Broadly, the instruments used for this evaluation consisted of two types: 1) learner assessments (Table 2); and 2) contextual tools for learners, teachers, principals and parents (Table 3).

Table 2: List of learner assessment subtasks

Task #	Task Description
	Introduction to the Learner
	Learner Information
Task 1.1:	HL Oral Reading Passage 1 - *Ditshwanelo tsa Botho*
Task 1.2:	HL Oral Reading Text Comprehension 1
Task 2.1:	HL Oral Reading Passage 2 - *Bopelokgale jwa ga Bonolo*
Task 2.2:	HL Oral Reading Text Comprehension 2
Task 3.1:	EFAL Oral Reading Passage - *An Unbelievable Night*
Task 3.2:	EFAL Oral Reading Text Comprehension
	Introduction to Written Assessment
Task 4:	HL Text Comprehension - *Perele*
Task 5:	HL Vocabulary
Task 6:	EFAL Text Comprehension - *The life cycle of plants*
Task 7:	EFAL Vocabulary (3 sections)
Tool 18:	Learner Asset List
Tool 17:	Learner Wellbeing questionnaire

As the data collection had multiple objectives beyond the impact evaluation, instruments had to meet the different requirements. An important element was that of developing language benchmarks in Setswana and EFAL.

Prior to the collection of EGRS I / RSP Wave 5 data, the Waves 1-4 Oral Reading Fluency (ORF) assessments were not suitable for examining the relationship between reading fluency and comprehension. To effectively model this relationship, there must be a large enough sample that reads far enough in the text to be able to answer all reading comprehension questions. This requires that enough time is provided to learners to read the text, and that the text is not too long. In earlier EGRS / RSP waves (and many other reading studies) ORF assessments only allowed reading for a minute.12 As a result too few learners had the chance to answer all comprehension questions.

New instruments were developed and piloted for data collection in 2021, which allowed for both 1-minute and 3-minute timings for the oral reading assessment. As a desired outcome of the instrument development process (and piloting iterations), most learners should be able to finish the passage within 3-minutes so they can attempt all the comprehension questions.

---

[12] Learners are then only asked comprehension questions related to the parts of the passage that they have read within the time limit. This creates an artificial relationship between fluency and comprehension.

At each stage of the piloting process, the analysis team assessed passage length against the percentage of learners reading the entire passage in 3-minutes. Typically, the length of passages was found to be too long. Consequently, they were reduced with each pilot iteration and more appropriate passage lengths were finally achieved to support the availability of larger sample sizes reading the whole passage at final data collection. The piloting process was used to assess and adjust the difficulty level of the comprehension questions asked about the passage. Pilot data was checked for irregularities in results, to highlight any instrument wording that may be ambiguous. The pilot data was also checked to ensure that the scores across all the questions aligned with expectations in terms of the comprehension processes that each question was tapping into. Various alterations were made at each piloting stage to improve the appropriateness of the comprehension questions.

Contextual Tools

The purpose of the contextual tools was to collect information on implementation fidelity and mediators for the impact evaluation. The contextual tools also contain questions relating to the COVID-19 disruption in schools.

Table 3: List of Contextual Tools

Instrument	Purpose
Principal consent form	To explain research to principal and obtain signed consent
Principal questionnaire	One per school, either Principal or other school leader. Provides important information about the school context. Also includes COVID-19 school-based research questions
Teacher consent form	One Grade 7 teacher to explain the research process and obtained consent
Grade 7 Teacher questionnaire	Two Grade 7 teachers provide information related to classroom practice, and COVID-19 issues
School Functionality	One per school. Conducted by a fieldworker that observes a set of indicators while walking around the school
Parent Home Questionnaire	Information about the research study for parents of learners selected for participation. It includes a consent form for learners to participate in the study and a questionnaire on socio-economic indicators in the household
Learner Asset List	Questionnaire administered to all learners that completed the learner assessments. The last section of the questionnaire elicits open-ended qualitative responses from learners but was only administered to 6 boys and 6 girls per grade per school.
Learner Wellbeing Questionnaire	Questionnaire administered to all learners that completed the learner assessments. It is designed to gather information on household assets for learners.

SUSTAINABILITY IMPACT FINDINGS

This section provides the main findings of the sustainability evaluation of EGRS I, as well as heterogeneous treatment effects of interest.

Attrition

There was evidence of high rates of attrition for Grade 7 learners during data collection. In 2015, 4,538 Grade 1 learners were assessed at the start of the year for the EGRS I baseline data collection. At the end of Grade 2, attrition was approximately 16%, while it increased to 27% by the end of Grade 4 (i.e., the Wave 4 data collection). Grade repetition and attrition results from 2021 (when the original cohort of learners was expected to be in Grade 7), are shown in Figure 1. This figure shows that approximately 30 to 40% of learners were found to be in the expected grade, while just over one-fifth of learners were found to be repeating an earlier grade (between Grade 4 and Grade 6). On average, approximately 43% of the original sample of learners were therefore not identified in the Wave 5 data collection[13].

The lowest rate of attrition was found for learners in schools that received the coaching intervention. Regression analyses confirmed that attrition was predicted by intervention status, with the coaching intervention showing a significantly smaller proportion of attriters than the control group. This is an interesting finding in itself (i.e., that coaching had a significant impact on retention in the sample). Furthermore, attrition estimates at the school level (i.e., for whole schools that were dropped from the Wave 5 sample), show that learners in the coaching intervention schools that were dropped had significantly higher baseline scores than learners in control schools that were dropped (see Annex A Table 16). These attrition estimates provide evidence that a larger proportion of coaching intervention learners were assessed in Wave 5 and that higher performing coaching intervention schools were dropped (as compared with control schools), which arguably means that any effect of the coaching intervention could be seen as lower-bound estimates of the intervention effect.

---

[13] These estimates are for learners in the 216 schools that were assessed in Grade 1 and Grade 7 (i.e., not including the 13 schools that were dropped from the Grade 7 sample). When examining all schools from the Wave 1 baseline, the overall attrition rate increases to 46%.

Figure 1: Proportion of attrition and grade repetition by treatment group
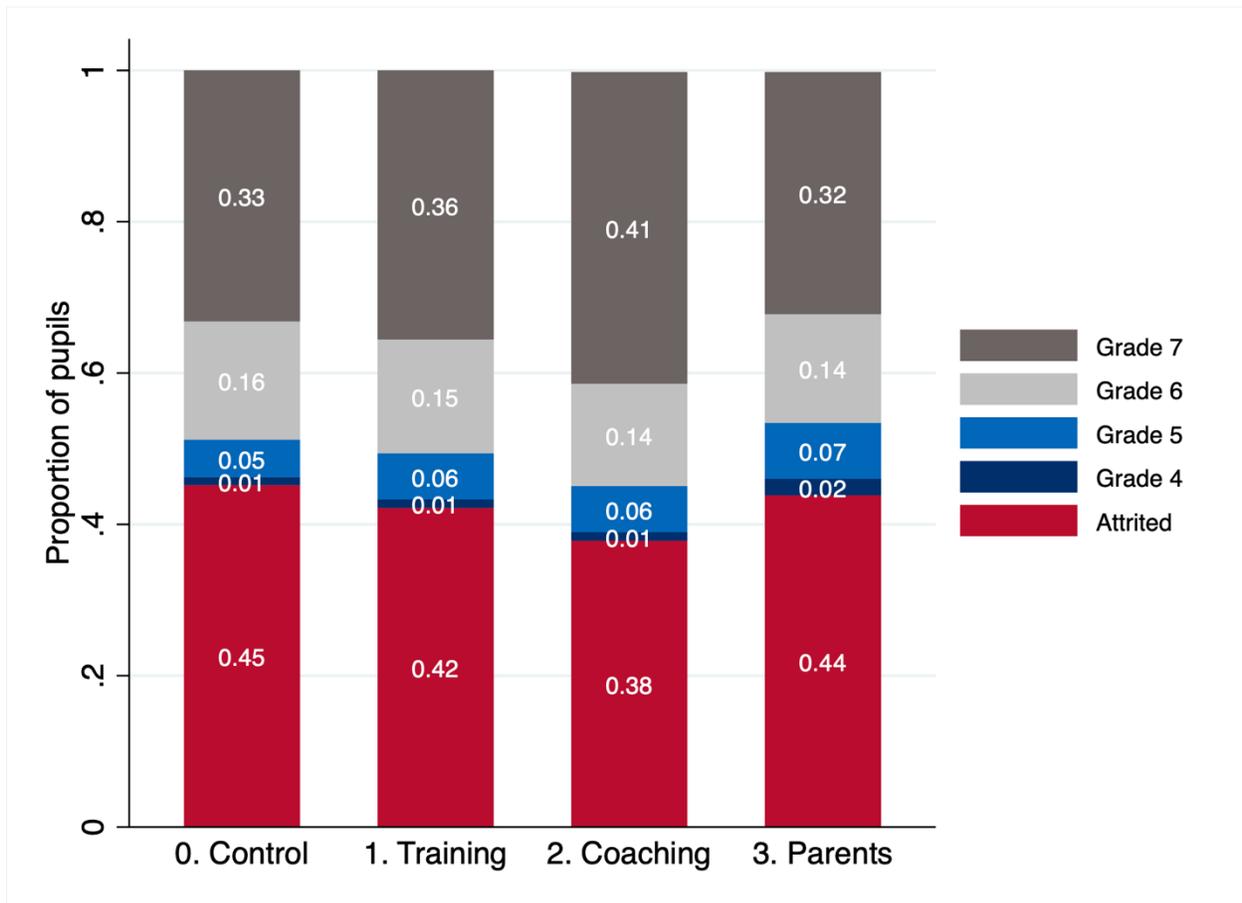
To further investigate the implications of attrition on sample balance, additional regression analyses were conducted, as shown in Table 4. Each column in the table represents a separate regression which was run on learner age, learner gender, and learner baseline score, respectively. The results suggest that in the Wave 5 sample, 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 parent group than the control group. As shown in the final column, the attrition did not cause a significant imbalance in terms of learner ability as measured at baseline. While we do control for baseline characteristics in all adjusted models throughout this report, this reassuring result confirms that additional adjustments are not needed to account for differential attrition.

Table 4: Regression estimates for balance on sample of non-attriters

	Age	Female	Baseline Composite Score
Training	0.101[t]	-0.0224	-0.148
	(0.0528)	(0.0311)	(0.136)
Coaching	-0.00576	-0.0200	0.110
	(0.0552)	(0.0283)	(0.165)
Parents	0.0593	-0.0591**	-0.0538
	(0.0529)	(0.0277)	(0.185)
Observations	2,431	2,439	2,439
R-squared	0.008	0.005	0.077
Control mean	6.407	0.508	0.131
Std Dev	0.632	0.500	1.040

Notes: ***$p<0.01$; **$p<0.05$; [t]$p<0.1$. Standard errors are clustered at the school level and reported in parentheses.

Lastly, while we did not collect data on the reason for attrition, it is possible to further explore some aspects of attriting learners. For example, after removing schools with no Grade 7 data, there were 1,132 learners in the sample who were assessed in Wave 1 but were labelled as attriters in Wave 4. Of those, 893 (79%) were also attriters in Wave 5, while 239 (21%) had returned to the Wave 5 sample. This provides evidence that at least one-fifth of attrition was likely due to absenteeism or another non-permanent attrition status.

Initial Results

Overall summary statistics

Summary statistics from the reading assessment provide evidence that there was strong variation in learner test scores. On average, learners read approximately 56 correct words per minute (cwpm) on the first Setswana passage and 76 cwpm on passage two, while they read a similar 77 cwpm on the English passage, as shown in Table 5. For Setswana, written comprehension rates were higher than oral reading passage comprehension rates but the opposite was true for English. Learners were able to correctly identify approximately one-quarter of English vocabulary items on the first part of the task (A) but significantly less so on parts B and C.

Table 5: Summary statistics for Grade 7 reading assessment by subtask

	N	Mean	p25	Median	p75
Setswana Text 1 Reading Fluency	2,357	56.4	37.7	58.3	76.8
Setswana Text 2 Reading Fluency	2,355	75.9	51.7	77.3	101.6
Setswana Text 1 Comprehension (% correct)	2,357	29.2	0	28.6	42.9
Setswana Text 2 Comprehension (% correct)	2,356	41.3	12.5	37.5	62.5
Setswana Written Comprehension (% correct)	2,260	49.4	27.3	54.5	72.7
Setswana Vocabulary (% correct)	2,264	62.7	40	70	90
English Text Reading Fluency	2,353	76.8	46	82.9	109.2
English Text Comprehension (% correct)	2,353	51	30	50	80
English Written Comprehension (% correct)	2,214	39.5	15.4	38.5	61.5
English Vocabulary (A) (% correct)	2,213	24.1	5.6	16.7	38.9
English Vocabulary (B) (% correct)	2,210	8.6	0	0	0
English Vocabulary (C) (% correct)	2,210	1.3	0	0	0

As shown in Table 6, zero scores were relatively low but non-negligible for the majority of subtasks (generally ranging from about 6% to 28%, aside from the final two vocabulary tasks). There were large floor effects on English Vocabulary (B) and English Vocabulary (C), with at least 90% of learners being unable to identify even a single item correctly on either task. As a result, these two tasks will not be a focus of the impact analyses throughout this report.

Table 6: Summary statistics for Grade 7 reading assessment zero scores by subtask

	N	Percentage scoring zero
Setswana Oral Passage Reading 1	2,357	7%
Setswana Oral Passage Reading 2	2,355	7%
Setswana Reading Comprehension (Passage 1)	2,357	28%
Setswana Reading Comprehension (Passage 2)	2,356	16%
Setswana Vocabulary	2,264	9%
Setswana Written Comprehension	2,260	6%
English Oral Passage Reading	2,353	11%
English Reading Comprehension	2,353	13%
English Written Comprehension	2,214	7%
English Vocabulary (A)	2,213	22%
English Vocabulary (B)	2,210	91%
English Vocabulary (C)	2,210	98%

The percent of learners unable to identify a single item per task ranges from 7% to 98% for English Vocabulary (C).

In addition to good score variation within subtasks, there was also strong evidence of test reliability for the Wave 5 assessments. Results from pairwise correlation analyses are presented in Table 7. The correlation coefficients are moderate too large for all pairwise estimates, except for English Vocabulary B and English Vocabulary C (which is not surprising, given the minimal variation in those subtasks). Interestingly, the correlations between comprehension and oral reading fluency were similar across subtasks and languages. Internal consistency was additionally calculated via Cronbach's alpha. The overall test scale was 0.94.

Table 7: Correlation coefficients for Wave 5 subtasks

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
(1) Setswana Text 1	1											
(2) Setswana Text 2	0.90	1										
(3) English Text	0.67	0.62	1									
(4) Setswana Comp 1	0.72	0.69	0.72	1								
(5) Setswana Comp 2	0.69	0.66	0.57	0.67	1							
(6) English Comp	0.72	0.69	0.55	0.64	0.68	1						
(7) English Written Comp	0.92	0.87	0.67	0.74	0.71	0.73	1					
(8) English Vocab A	0.69	0.67	0.68	0.74	0.67	0.66	0.79	1				
(9) English Vocab B	0.60	0.57	0.55	0.64	0.65	0.59	0.67	0.71	1			
(10) English Vocab C	0.67	0.63	0.59	0.65	0.65	0.60	0.73	0.70	0.73	1		
(11) Setswana Vocab	0.30	0.28	0.27	0.26	0.28	0.26	0.32	0.30	0.33	0.45	1	
(12) Setswana Written Comp	0.15	0.14	0.18	0.17	0.13	0.11	0.14	0.14	0.17	0.24	0.16	1

Descriptive statistics by intervention group

As a preliminary approach for assessing the sustained impact of EGRS I on learners in 2021, mean estimates for each subtask are presented by treatment group in Table 8. Two general trends are apparent in this table. First, the average scores among learners in the coaching group are slightly higher than those in other groups. Second, the average scores among learners in the parental involvement group tend to be lower than in the other groups. However, these differences are small.

Table 8: Average reading performance by subtask and treatment group (mean scores)

	N	Control	Training	Coaching	Parents
Setswana Text 1 Reading Fluency	2,357	56.9	55.8	59.0	53.2
Setswana Text 2 Reading Fluency	2,355	75.6	75.7	80.4	71.7
Setswana Text 1 Comp (% correct)	2,357	29.7	26.3	31.9	28.4
Setswana Text 2 Comp (% correct)	2,356	41.2	39.8	44.3	39.7
Setswana Written Comp (% correct)	2,260	49.2	49.1	51.6	47.7
Setswana Vocabulary (% correct)	2,264	62.0	62.7	65.8	60.4
English Text Reading Fluency	2,353	77.9	76.7	79.5	72.2
English Text Comp (% correct)	2,353	52.2	49.7	53.7	47.2
English Written Comp (% correct)	2,214	40.3	37.5	42.1	37.3
English Vocabulary (A) (% correct)	2,213	25.7	23.8	25.0	20.7
English Vocabulary (B) (% correct)	2,210	11.3	7.9	8.5	4.9
English Vocabulary (C) (% correct)	2,210	1.1	0.9	1.4	2.0

Similar trends are found in Table 9, which shows the percent of learners who were unable to correctly identify a single item on a given subtask (i.e., zero scores). The largest differences for learners in the parental involvement group appear to be related to the oral reading passages (both in terms of reading fluency and comprehension, in both Setswana and English).

Table 9: Percentage scoring zero by subtask and treatment group (mean percent scores)

	N	Control	Training	Coaching	Parents
Setswana Text 1 Reading	2,357	6%	7%	6%	11%
Setswana Text 2 Reading	2,355	10%	10%	10%	10%
Setswana Text 1 Comp	2,357	27%	28%	23%	35%
Setswana Text 2 Comp	2,356	17%	16%	13%	20%
Setswana Written Comp	2,264	10%	8%	8%	9%
Setswana Vocabulary	2,260	6%	7%	5%	5%
English Text Reading	2,353	10%	11%	9%	14%
English Text Comp	2,353	13%	14%	11%	17%
English Written Comp	2,214	8%	9%	7%	7%
English Vocabulary (A)	2,213	18%	22%	25%	24%
English Vocabulary (B)	2,210	88%	92%	91%	95%
English Vocabulary (C)	2,210	99%	99%	98%	97%

Complimenting the mean scores presented in Table 8 and Table 9, we also present the full distribution of oral reading fluency scores for each intervention group, in the following three figures. As shown in Figure 2, the light blue line for learners in the parental involvement group is below the other three lines for much of the distribution until approximately 90 correct words per minute (which aligns with approximately 0.1 on the y-axis). This means that while the top 10% of learners in the parental involvement group are scoring similarly to their counterparts in other groups, the bottom 90% are scoring below their counterparts. The higher coaching performance shown in Table 8, appears to come primarily from the middle of the distribution (before and after which point the coaching distribution overlaps training and control considerably).

*Figure 2: Distribution of English oral reading fluency scores by intervention*
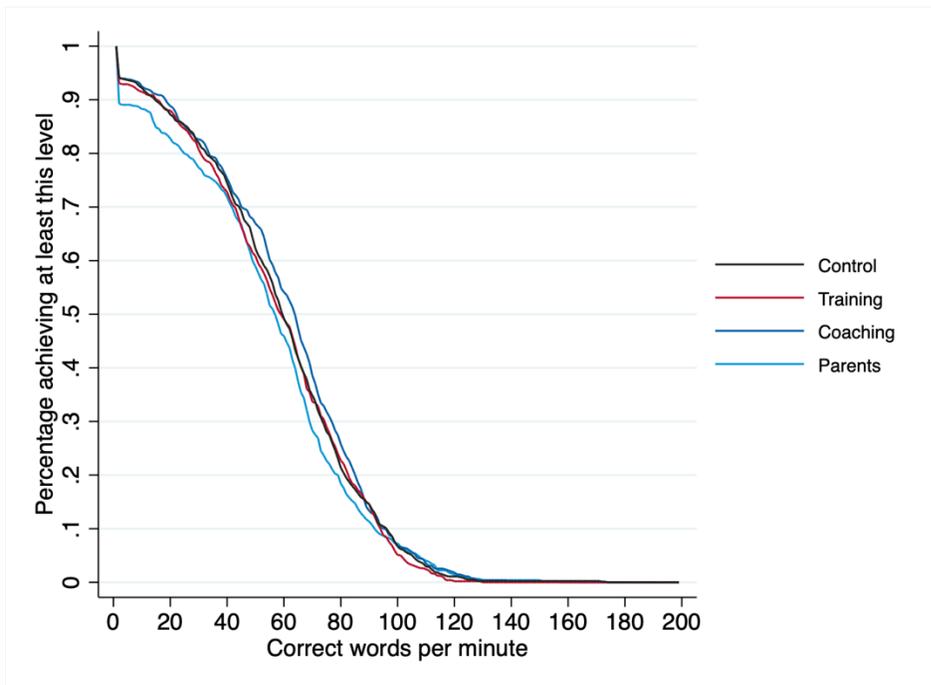
Figure 3 and Figure 4 show similar trends. In each of these cases, the lower performance among learners in the parental involvement intervention seems to span a larger part of the distribution (for all but the top performing readers).

Figure 3: Distribution of Setswana 1 text oral reading fluency scores by intervention
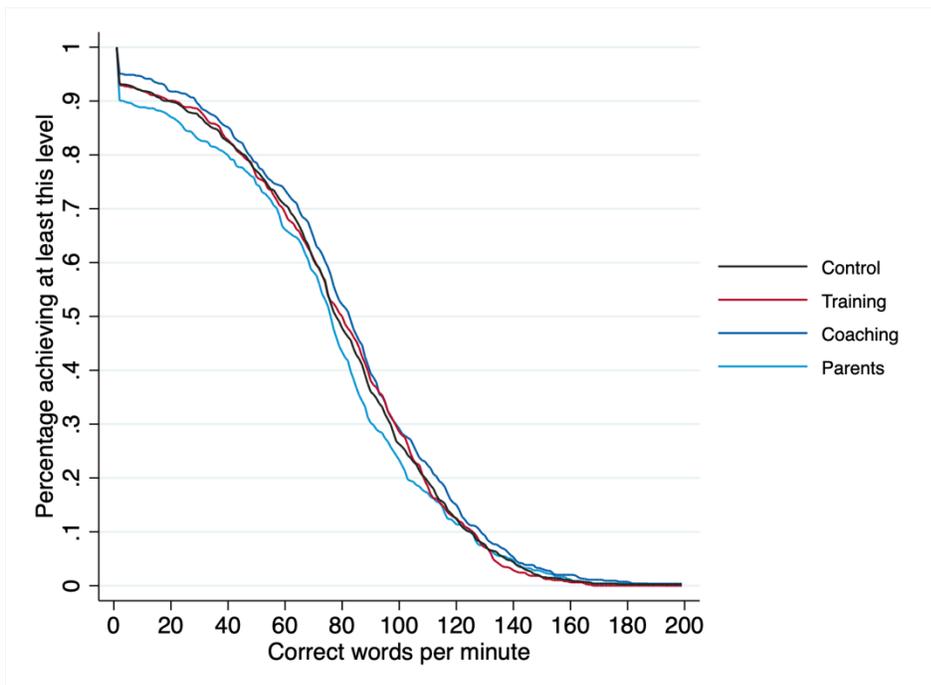
Figure 4: Distribution of Setswana 2 text oral reading fluency scores by intervention
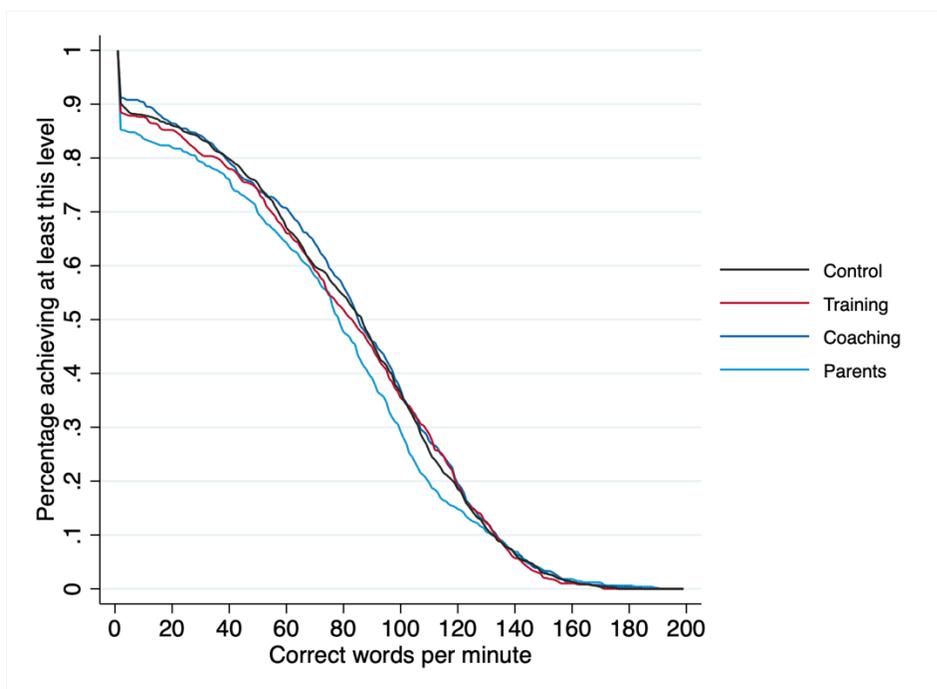
Main Outcome Results

The descriptive analyses point to potential, small differences in outcomes across intervention groups. Although random assignment to intervention groups theoretically allows for mean comparisons to measure impact, precision can be increased by using regression models that account for factors which differed at baseline. Therefore, in order to obtain more precise estimates of sustained impact, we use regression models which additionally control for factors that could impact scores but are unrelated to the interventions. Specifically, these models include controls at the learner level (i.e., baseline reading scores, gender, parental education, assets at home, and books at home), school level (i.e., location of school and average performance on 2014 ANAs), and community level (i.e., wealth index and attendance rates).

Tables 10 through 13 show the results of separate regression models for all reading outcomes, with the coefficients on training, coaching and parents signifying the estimated impact of each intervention relative to the control group. Table 10 shows the results of regressions run on each of the Setswana reading outcomes (with separate regression estimates shown in each column). The first column uses a composite score for the individually administered Setswana subtasks[14]. Although the magnitude and direction of impacts is similar to the descriptive analyses above, there is only one significant estimate in the table (with 95% confidence): an approximately 5.7 cwpm increase in Setswana reading fluency on the second passage for learners in the coaching intervention. There are

---

[14] The Setswana composite score was created using the first component loadings from a principal component analysis of the four individually administered Setswana reading subtasks.

also a number of marginally significant estimates (as noted by the single stars in the table), all for the coaching intervention.

Impact estimates for zero scores in Setswana are presented in Table 11. Since the 'zero score' outcome variables in each of these models are binary, logistic regression estimates were calculated. The coefficients in the table therefore represent odds-ratios (with numbers smaller than 1 signifying a lower likelihood of scoring zero and numbers larger than 1 signifying a greater likelihood of scoring zero). These results show that learners in the parental involvement intervention were significantly more likely to score zero on the first Setswana reading passage (~1.7 times the odds). The only other significant finding from these models was that learners in the training intervention schools were about half as likely to score zero on Setswana vocabulary (odd ration = 0.57), as compared with learners in control schools.

The same models were also run for all English outcomes, with results presented in Table 12 and Table 13. These tables show similar trends as the Setswana data (with the largest magnitude estimates for the coaching intervention) but there was insufficient evidence to show statistically significant impacts on any of the English subtasks (including the composite[15]). It is interesting to note that overall, the explained variance across all models (Setswana and EFAL) remained surprisingly low, given that they included baseline estimates as covariates.

---

[15] The English composite score was created using the first component loadings from a principal component analysis of the two individually administered English reading subtasks.

Table 10: OLS Regression estimates for impact of intervention on Setswana reading outcomes with controls

VARIABLES	(1) Setswana Composite	(2) Setswana Text 1 Reading	(3) Setswana Text 2 Reading	(4) Setswana Text 1 Comp	(5) Setswana Text 2 Comp	(6) Setswana Written Comp	(7) Setswana Vocabulary
Training	0.0353	1.190	2.816	-1.156	1.800	3.336	3.561
	(0.0731)	(2.138)	(2.827)	(2.180)	(2.264)	(2.096)	(2.541)
Coaching	0.134^t	2.730	5.670**	3.048	4.001^t	3.429^t	4.220^t
	(0.0687)	(1.928)	(2.706)	(2.207)	(2.137)	(1.913)	(2.259)
Parents	0.0241	-0.623	0.0953	1.158	2.087	1.441	1.247
	(0.0723)	(2.153)	(2.837)	(2.302)	(2.183)	(2.070)	(2.679)
Observations	2,353	2,356	2,354	2,356	2,355	2,193	2,196
R-squared	0.192	0.177	0.157	0.146	0.156	0.158	0.151
P-value	0.216	0.481	0.327	0.0983	0.398	0.967	0.777
Control mean	0.006	56.86	75.61	29.67	41.17	49.24	61.95

Notes: ***$p<0.01$; **$p<0.05$; $^t p<0.1$. Standard errors are clustered at the school level and reported in parentheses. P-value represents statistical significance for the difference between training and coaching interventions.

OLS = ordinary least squares, a type of linear least squares method for estimating unknown parameters in a linear regression model.

Table 11: Logistic regression estimates for impact of interventions on Setswana zero scores with controls

VARIABLES	(1) Setswana Text 1 Reading	(2) Setswana Text 2 Reading	(3) Setswana Text 1 Comprehension	(4) Setswana Text 2 Comprehension	(5) Setswana Written Comprehension	(6) Setswana Vocabulary
Training	1.053	0.920	0.856	0.757	0.975	0.571**
	(0.287)	(0.246)	(0.143)	(0.152)	(0.226)	(0.151)
Coaching	1.142	0.708	0.766	0.751	0.919	0.819
	(0.329)	(0.220)	(0.134)	(0.156)	(0.241)	(0.198)
Parents	1.675**	1.261	1.199	0.988	0.713	0.740
	(0.422)	(0.318)	(0.194)	(0.191)	(0.206)	(0.207)
Observations	2,356	2,354	2,356	2,355	2,193	2,196
Control mean	0.0559	0.0681	0.272	0.166	0.0579	0.100

Notes: ***p<0.01; **p<0.05; ᵗp<0.1. Estimates are odds ratios. Standard errors are clustered at the school level and reported in parentheses.

Table 12: OLS Regression estimates for impact of interventions on English reading outcomes with controls

VARIABLES	(1) English Composite	(2) English Text Reading	(3) English Text Comprehension	(4) English Written Comprehension	(5) English Vocabulary (A)
Training	0.0715	3.431	1.821	1.169	1.547
	(0.0769)	(3.298)	(2.503)	(2.067)	(1.810)
Coaching	0.0878	3.334	2.867	3.293[t]	0.932
	(0.0669)	(2.967)	(2.223)	(1.885)	(1.762)
Parents	-0.0118	-0.536	-0.324	0.883	-2.046
	(0.0715)	(3.200)	(2.301)	(1.843)	(1.564)
Observations	2,353	2,353	2,353	2,152	2,151
R-squared	0.225	0.199	0.215	0.188	0.177
P-value	0.840	0.977	0.698	0.348	0.758
Control mean	0.034	77.94	52.19	40.28	25.66

Notes: ***p<0.01; **p<0.05; [t]p<0.1. Standard errors are clustered at the school level and reported in parentheses. P-value represents statistical significance for the difference between training and coaching interventions.

Table 13: Logistic regression estimates for impact of interventions on English zero scores with controls

VARIABLES	(1) English Text Reading	(2) English Text Comprehension	(3) English Written Comprehension	(4) English Vocabulary (A)
Training	1.027	0.876	0.964	1.030
	(0.249)	(0.176)	(0.241)	(0.188)
Coaching	0.948	0.853	0.842	1.489^t
	(0.240)	(0.192)	(0.210)	(0.307)
Parents	1.246	1.102	0.750	1.143
	(0.305)	(0.229)	(0.206)	(0.213)
Observations	2,353	2,353	2,152	2,151
Control mean	0.0973	0.125	0.0760	0.183

Notes: ***p<0.01; **p<0.05; ^{t}p<0.1. Estimates are odds ratios. Standard errors are clustered at the school level and reported in parentheses.

In order to explore these results further, we calculated intervention impacts separately for subgroups of learners. First, we examined whether or not there were differential impacts for three samples of learners: 1) the full sample of learners who were tracked from Wave 1 to Wave 5; 2) the subsample of learners who were 'on track' in Wave 5 (i.e., those in Grade 7); and 3) the subsample of learners who had progressed to Grade 4 by the time of the Wave 4 data collection (i.e., those who were deemed as having received 'maximum dosage' of the intervention by the time of the Wave 4 report). There is significant overlap in these two subgroups but there was also variation. Of the maximum dosage learners in the Wave 5 sample, 80% were on track (i.e., in Grade 7).

Estimates of sustained impacts on the Setswana composite score for the three intervention groups across each of these three learner samples are displayed in Figure 5. The impact estimates are shown by the numbers next to the circles (note that the full sample estimates are identical to those displayed in column 1 of Table 10), while the confidence intervals are represented by the vertical lines (90%) and vertical bars (95%). In other words, if an entire bar sits above or below the horizontal line at zero, that represents a statistically significant impact with at least 95% confidence (p<0.05). It is clear from this figure that the impact of the coaching intervention was only marginally significant overall and for learners who were on track but it was 50% larger and statistically significant for those learners who received the maximum dosage. This trend was similar for the majority of subtasks, with the impacts of coaching being larger for maximum dosage samples than for the full sample of learners, as shown in Table 14. There was only one subtask (written comprehension) with significant impacts for the training intervention, for on-track and maximum dosage learners.

Figure 5: Impact on Setswana composite score by intervention and subsample
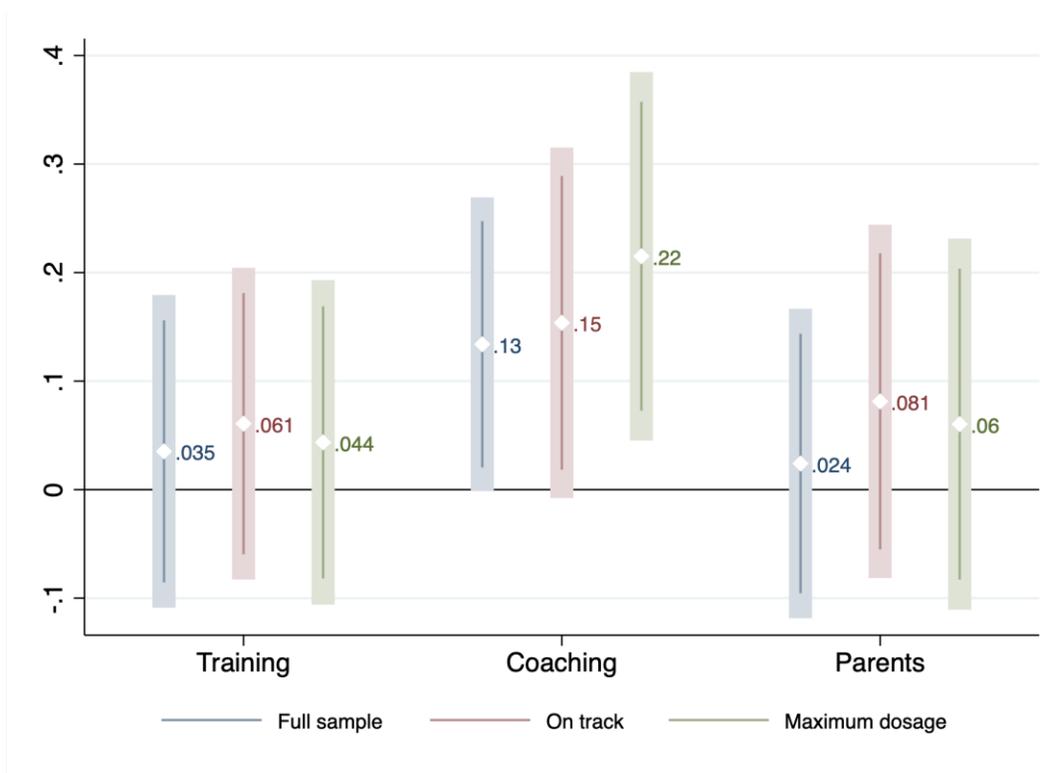
Table 14: Impact on Setswana reading outcomes by intervention and subsample

	Training			Coaching			Parental Involvement		
	Full	On Track	Max Dose	Full	On Track	Max Dose	Full	On Track	Max Dose
Setswana Composite	0.0353	0.0608	0.0435	0.134^t	0.154^t	0.215**	0.0241	0.0813	0.0604
	(0.130)	(0.129)	(0.135)	(0.122)	(0.146)	(0.153)	(0.129)	(0.147)	(0.154)
Setswana Text 1 ORF	1.190	1.242	2.060	2.730	3.290	5.471**	-0.623	0.902	0.549
	(2.138)	(2.025)	(2.207)	(1.928)	(2.184)	(2.227)	(2.153)	(2.344)	(2.472)
Setswana Text 2 ORF	2.816	4.122	4.223	5.670**	6.410^t	9.703**	0.0953	2.546	1.559
	(2.827)	(2.709)	(2.941)	(2.706)	(3.453)	(3.761)	(2.837)	(2.946)	(3.273)
Setswana Text 1 Comp	-1.156	-1.705	-2.523	3.048	3.060	4.767^t	1.158	1.554	1.751
	(2.180)	(2.766)	(2.561)	(2.207)	(2.852)	(2.863)	(2.302)	(3.025)	(2.973)
Setswana Text 2 Comp	1.800	3.972	1.988	4.001^t	4.986^t	4.942^t	2.087	4.137	2.793
	(2.264)	(2.427)	(2.451)	(2.137)	(2.656)	(2.627)	(2.183)	(2.647)	(2.740)
Setswana Written Comp	3.336	6.504***	5.595**	3.429^t	4.628**	3.943^t	1.441	2.206	1.885
	(2.096)	(1.969)	(2.175)	(1.913)	(2.235)	(2.162)	(2.070)	(2.145)	(2.177)
Setswana Vocabulary	3.561	4.812	5.257^t	4.220^t	4.701^t	5.360**	1.247	4.102	1.610
	(2.541)	(2.937)	(3.109)	(2.259)	(2.764)	(2.680)	(2.679)	(2.806)	(3.085)
Observations	522	321	336	560	371	330	516	294	297

Notes: ***p<0.01; **p<0.05; ^{t}p<0.1. Standard errors are clustered at the school level and reported in parentheses.

Results for the English composite score are presented in Figure 6. There were no effects for any of the interventions on the English composite score. Unlike with Setswana, there was only one significant impact across all subtasks for English, as shown in Annex A (Table 20).

Figure 6: Impact on English composite scores by intervention and subsample
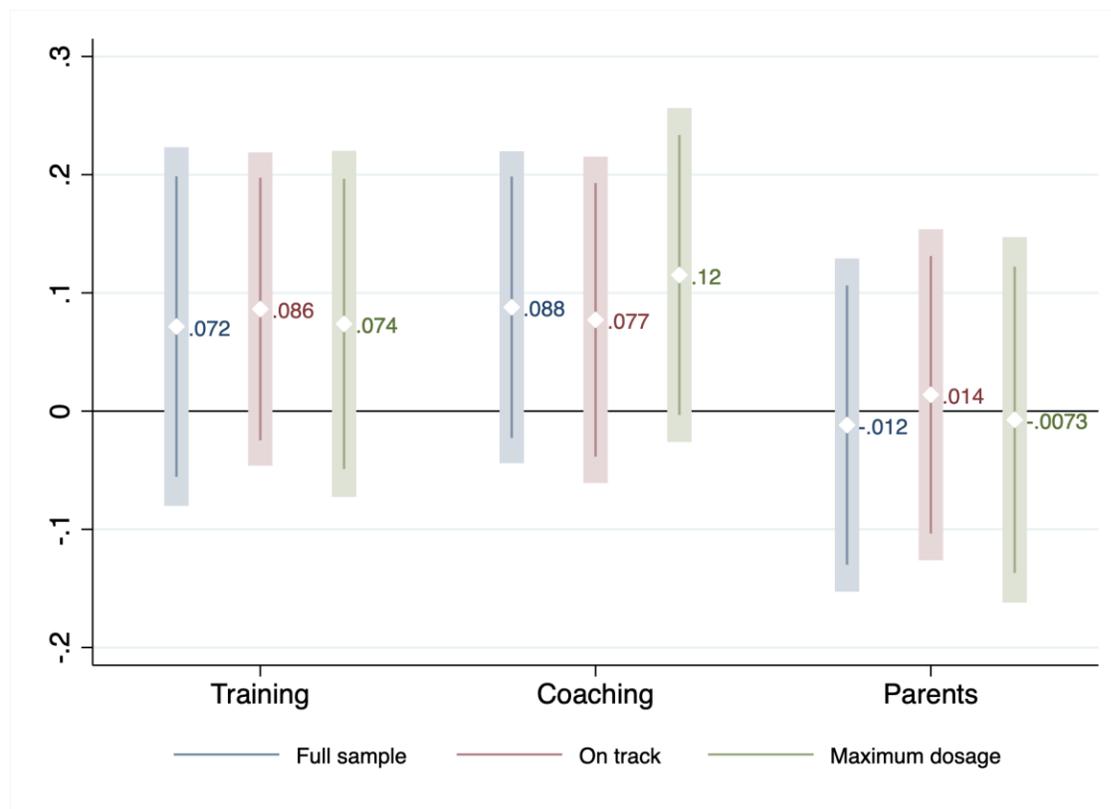
Heterogeneous Impacts

In previous rounds of the EGRS I evaluation, differential effects of the interventions were investigated across a range of factors. We examined these factors again with Wave 5 data to determine whether or not any differential effects persisted. At the learner level, we ran regression models on the Setswana composite score with interaction terms for each intervention group with learner gender, learner age and learner baseline reading scores (see Table 15). These interaction terms (if significant) would signify a differential impact of the intervention for these groups of learners. There are no significant differential effects for gender (i.e., no differential program impact for boys vs girls), nor are there any for age. There is some evidence of a differential effect for the parental involvement intervention based on baseline learner performance. Although the parental involvement intervention was not impactful overall, learners with higher baseline scores may have a sustained benefit from the parental involvement intervention by the time of the Wave 5 data collection. However, since this effect was not found in prior waves, it should be interpreted cautiously. Finally, there was a significant negative interaction squared baseline scores for the training intervention (for both Setswana and English).

Table 15: Impact on Setswana composite scores by intervention group for learner characteristics

VARIABLES	(1) Learner: Female	(2) Learner: Age	(3) Learner: Baseline Score	(4) Learner: Baseline Score Squared
Training	0.0817	0.332	0.0282	0.0688
	(0.0843)	(0.512)	(0.0732)	(0.0770)
Coaching	0.149[t]	0.827	0.137**	0.134[*]
	(0.0884)	(0.635)	(0.0689)	(0.0719)
Parents	0.0180	0.635	0.0189	0.0185
	(0.0853)	(0.564)	(0.0718)	(0.0760)
Training x group	-0.0952	-0.0450	0.00600	0.0774
	(0.0951)	(0.0789)	(0.0730)	(0.0985)
Coaching x group	-0.0299	-0.108	-0.00709	0.0534
	(0.116)	(0.0989)	(0.0568)	(0.106)
Parents x group	0.0148	-0.0934	0.110**	0.136
	(0.112)	(0.0867)	(0.0527)	(0.0940)
Training x group squared				-0.0659**
				(0.0318)
Coaching x group squared				-0.0125
				(0.0269)
Parents x group squared				-0.00171
				(0.0232)
Observations	2,353	2,345	2,353	2,353
R-squared	0.193	0.196	0.195	0.203

Notes: ***$p<0.01$; **$p<0.05$; [t]$p<0.1$. Standard errors are clustered at the school level and reported in parentheses.

There were no significant learner-level differential effects on the English composite score. However, as in prior rounds, there was some evidence that weaker baseline learners may have benefited more from the training intervention (see additional results in Annex A). Additional differential impacts were investigated based on school-level factors. However, no statistically significant effects were found on either the Setswana or English composite scores for any of the factors examined: urban vs rural, school location (district) or the condition of the school (based on the school functionality index).

CONCLUSIONS

This sustainability impact evaluation of EGRS I has provided a rare opportunity to track the progress of learners (longitudinally) over nearly seven years and to examine the long-term impacts of an early grade reading intervention, three to four years after the intervention has been completed.

The results show evidence of a sustained impact from the coaching intervention on several Setswana (home language) outcomes. Although there was a significant average impact for all learners in the coaching intervention on only one of the Setswana oral reading passages, the most consistent impacts were for learners who received the maximum dosage of the EGRS I intervention in Grades 1 to 3 (i.e., those who progressed, as intended, through the three years of the intervention from 2015 to 2018). Significant impacts were found for these learners on four of the seven Setswana reading outcomes (with marginally significant impacts on the rest). This provides evidence that results from an early grade reading intervention can be sustained into upper grades but also points to the importance of dosage. Interestingly, the coaching intervention also showed a statistically significant impact on retention, with a smaller proportion of attriting learners than the control group (38% versus 45%).

While EFAL results followed a similar pattern to Setswana HL results (with the largest magnitude, positive impacts for the coaching intervention), there were ultimately no statistically significant long-term intervention crossover effects on EFAL. Additionally, heterogeneous impacts were limited, with some evidence that learners with different baseline scores (from Grade 1) may have been impacted differently by the interventions.

Overall, the results are very promising for the EGRS I coaching intervention, particularly in light of the fact that few other research studies have shown such a sustained, long-term impact of an early grade reading program.

ANNEX A: ADDITIONAL TABLES

Table 16: Regression estimates for balance on sample of learners in schools that were dropped from Wave 5

	Age	Female	Baseline Composite Score
Training	-0.283	0.0875	-0.0254
	(0.268)	(0.110)	(0.265)
Coaching	-0.300	0.1000	0.280***
	(0.265)	(0.0714)	(0.0721)
Parents	-0.0115	0.0750	0.464^t
	(0.363)	(0.0470)	(0.248)
Observations	280	280	280
R-squared	0.040	0.004	0.111
Control mean	6.746	0.350	-0.659
Std Dev	0.773	0.483	0.195

Notes: ***p<0.01; **p<0.05; ^{t}p<0.1. Standard errors are clustered at the school level and reported in parentheses.

Table 17: OLS Regression estimates for impact of interventions on Setswana reading outcomes without controls

VARIABLES	(1) Setswana Composite	(2) Setswana Text 1 Reading	(3) Setswana Text 2 Reading	(4) Setswana Text 1 Comp	(5) Setswana Text 2 Comp	(6) Setswana Written Comp	(7) Setswana Vocabulary
Training	-0.0500	-0.917	0.245	-3.069	-1.210	0.349	0.782
	(0.0805)	(2.345)	(3.077)	(2.231)	(2.480)	(2.232)	(2.714)
Coaching	0.118[t]	2.372	5.091[t]	2.601	3.550[t]	2.860	4.271[t]
	(0.0701)	(1.986)	(2.825)	(2.203)	(2.132)	(2.063)	(2.561)
Parents	-0.0891	-3.746[t]	-3.918	-1.134	-1.313	-1.125	-1.632
	(0.0793)	(2.242)	(2.901)	(2.507)	(2.531)	(2.342)	(2.752)
Observations	2,354	2,357	2,355	2,357	2,356	2,260	2,264
R-squared	0.015	0.011	0.011	0.020	0.015	0.015	0.017
P-value	0.0454	0.165	0.136	0.0192	0.0682	0.299	0.187
Control mean	0.006	56.86	75.61	29.67	41.17	49.24	61.95

Notes: ***$p<0.01$; **$p<0.05$; [t]$p<0.1$. Standard errors are clustered at the school level and reported in parentheses. P-value represents statistical significance for the difference between training and coaching interventions.

Table 18: Logistic regression estimates for impact of interventions on Setswana zero scores without controls

VARIABLES	(1) Setswana Text 1 Reading	(2) Setswana Text 2 Reading	(3) Setswana Text 1 Comprehension	(4) Setswana Text 2 Comprehension	(5) Setswana Written Comprehension	(6) Setswana Vocabulary
Training	1.219	1.050	1.040	0.926	1.189	0.769
	(0.329)	(0.276)	(0.176)	(0.194)	(0.271)	(0.217)
Coaching	1.073	0.707	0.795	0.730	0.892	0.719
	(0.320)	(0.227)	(0.128)	(0.154)	(0.231)	(0.178)
Parents	1.972***	1.494	1.458**	1.235	0.791	0.966
	(0.487)	(0.373)	(0.228)	(0.235)	(0.217)	(0.270)
Observations	2,357	2,355	2,357	2,356	2,260	2,264
Control mean	0.0559	0.0681	0.272	0.166	0.0579	0.100

Notes: ***p<0.01; **p<0.05; ᵗp<0.1. Estimates are odds ratios. Standard errors are clustered at the school level and reported in parentheses.

Table 19: OLS Regression estimates for impact of interventions on English reading outcomes without controls

VARIABLES	(1) English Composite	(2) English Text Reading	(3) English Text Comprehension	(4) English Written Comprehension	(5) English Vocabulary (A)
Training	-0.0496	-0.997	-2.250	-2.591	-1.770
	(0.0862)	(3.692)	(2.731)	(2.314)	(2.038)
Coaching	0.0624	2.141	2.198	2.179	-0.486
	(0.0707)	(3.078)	(2.334)	(2.098)	(1.902)
Parents	-0.137	-5.526	-4.214	-2.538	-4.727**
	(0.0827)	(3.509)	(2.715)	(2.178)	(1.904)
Observations	2,353	2,353	2,353	2,214	2,213
R-squared	0.024	0.013	0.035	0.022	0.018
P-value	0.199	0.395	0.119	0.0517	0.559
Control mean	0.0336	77.94	52.19	40.28	25.66

Notes: ***p<0.01; **p<0.05; ᵗp<0.1. Standard errors are clustered at the school level and reported in parentheses. P-value represents statistical significance for the difference between training and coaching interventions.

Table 20: Logistic regression estimates for impact of interventions on English zero scores without controls

VARIABLES	(1) English Text Reading	(2) English Text Comprehension	(3) English Written Comprehension	(4) English Vocabulary (A)
Training	1.202	1.102	1.154	1.256
	(0.289)	(0.227)	(0.277)	(0.234)
Coaching	0.896	0.847	0.850	1.443^t
	(0.225)	(0.184)	(0.217)	(0.289)
Parents	1.546^t	1.404^t	0.892	1.375^t
	(0.366)	(0.285)	(0.232)	(0.255)
Observations	2,353	2,353	2,214	2,213
Control mean	0.0973	0.125	0.0760	0.183

Notes: ***p<0.01; **p<0.05; ^{t}p<0.1. Estimates are odds ratios. Standard errors are clustered at the school level and reported in parentheses.

Table 21: Impact on English reading outcomes by intervention and subsample

	Training			Coaching			Parental Involvement		
	Full	On Track	Max Dose	Full	On Track	Max Dose	Full	On Track	Max Dose
English Composite	0.0715	0.0863	0.0738	0.0878	0.0772	0.115	-0.0118	0.0138	-0.00735
	(0.0769)	(0.0672)	(0.0742)	(0.0669)	(0.0700)	(0.0716)	(0.0715)	(0.0710)	(0.0784)
English Text 1 ORF	3.431	4.290	4.675	3.334	2.707	5.613[t]	-0.536	0.440	0.388
	(3.298)	(2.890)	(3.198)	(2.967)	(3.083)	(3.353)	(3.200)	(3.168)	(3.517)
English Text 1 Comp	1.821	2.093	1.066	2.867	2.680	2.866	-0.324	0.512	-0.717
	(2.503)	(2.453)	(2.573)	(2.223)	(2.460)	(2.325)	(2.301)	(2.446)	(2.610)
English Written Comp	1.169	2.561	2.340	3.293*	3.199	1.133	0.883	1.330	-0.301
	(2.067)	(2.372)	(2.450)	(1.885)	(2.406)	(2.167)	(1.843)	(2.265)	(2.259)
English Vocabulary (A)	1.547	1.959	1.453	0.932	1.385	-0.293	-2.046	-2.914	-4.740**
	(1.810)	(2.380)	(2.236)	(1.762)	(2.453)	(2.436)	(1.564)	(2.290)	(2.148)
Observations	522	321	336	560	371	330	516	294	297

Notes: ***$p<0.01$; **$p<0.05$; [t]$p<0.1$. Standard errors are clustered at the school level and reported in parentheses.