USAID
FROM THE AMERICAN PEOPLE

PRE-ANALYSIS PLAN

SUSTAINABILITY IMPACT EVALUATION OF THE EARLY GRADE READING STUDY (EGRS I) AND IMPACT EVALUATION OF THE READING SUPPORT PROJECT (RSP)

September 3 2021

Contract Number: 72067418D00001, Order Number: 72067421F00001

This publication was produced at the request of the United States Agency for International Development. It was prepared independently by Khulisa Management Services, (Pty) Ltd.

PRE- ANALYSIS PLAN

Sustainability Impact Evaluation of the Early Grade Reading Study (EGRS I) and Impact Evaluation of the Reading Support Project (RSP)

[This is a continuation of the Early Grade Reading Study in South Africa, 2016, RCT ID AEARCTR-0001749**]**

3 September, 2021

AUTHORS

Ms. Margaret Roper, Senior Project Manager (Khulisa)
Ms. Benita Williams, Senior Education Research Expert (Benita Williams Evaluation)
Dr. Matthew Jukes and Dr. Jonathan Stern, Research Triangle Institute International (RTI)
Dr. Stephen Taylor and Ms. Nompumelelo Moholwane (National Department of Basic Education)
Ms. Leticia Taimo, Project Manager (Khulisa)
Dr. Gabriella Wills and Dr. Cally Ardington, Research on Socio-Economic Policy (ReSEP)
Ms. Jennifer Bisgard, Project Director (Khulisa)

CONTACT DETAILS

Margaret Roper
26 7th Avenue
Parktown North
Johannesburg, 2196

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

CONTENTS

1. SUMMARY TRIAL INFORMATION 5
GENERAL INFORMATION 5
LOCATION 5
PRIMARY INVESTIGATOR 5
OTHER PRIMARY INVESTIGATORS 5
ADDITIONAL TRIAL INFORMATION 6
ABSTRACT 7
REGISTRATION CITATION 7
2. SPONSORS AND PARTNERS 8
2.1. BACKGROUND 8
2.2. OBJECTIVES OF THE RESEARCH 8
3. EXPERIMENTAL DETAILS 10
3.1. DESCRIPTION OF INTERVENTIONS 10
DESCRIPTION OF THE EGRS I 10
DESCRIPTION OF THE RSP 11
INTERACTIONS BETWEEN EGRS I AND RSP 11
3.2. PRIMARY OUTCOMES 12
3.3. DESCRIPTION OF THE EVALUATION DESIGN EGRS I AND RSP 12
EVALUATION DESIGN EGRS I 12
EVALUATION DESIGN RSP 13
3.1. EXPERIMENT CHARACTERISTICS 15
SAMPLING EGRS I SUSTAINABILITY IMPACT EVALUATION 15
SAMPLING RSP IMPACT EVALUATION 15
SUMMARY OF THE STUDY POPULATION 16
SAMPLE SIZE CALCULATION 16
4. METHODS OVERVIEW 19
4.1. DATA SOURCES AND INSTRUMENTS 19
4.2. ANALYSIS METHODS 19
Main impact analyses 19
Covariates 20
5. HYPOTHESIS UNDER INVESTIGATION 21
5.1. MAIN OUTCOMES 21
5.2. INTERMEDIATE OUTCOMES 22
Fidelity of Implementation 22
Mediators 23

Heterogeneous Impacts 24
5.3. ADDITIONAL CORRECTIONS 25
Assessing the Impact of Attrition 25

LIST OF TABLES

Table 1 EGRS Grade progression..... 9
Table 2 Number of schools in EGRS and RSP treatment groups..... 13
Table 3 Provision of Classroom Libraries by RSP Group 14
Table 4 Learner Sample..... 15
Table 5 Sample size calculations (longitudinal versus cross-sectional) 17

1. SUMMARY TRIAL INFORMATION

GENERAL INFORMATION

Title

Impact Evaluation of the Early Grade Reading Study and the Reading Support Programme

Initial registration Date

3 September 2021

LOCATION

Country

South Africa

Region

North West Province, in the districts of Ngaka Modiri Molema and Dr. Kenneth Kaunda

PRIMARY INVESTIGATOR

Name

Margaret Roper

Affiliation

Khulisa Management Services

OTHER PRIMARY INVESTIGATORS

Name

Benita Williams

Affiliation

Benita Williams Evaluation

Name

Matthew Jukes

Affiliation

Research Triangle Institute International (RTI)

Name

Jonathan Stern

Affiliation

Research Triangle Institute International (RTI)

Name

Stephen Taylor

Affiliation

National Department of Basic Education, South Africa

Name

Nompumelelo Moholwane

Affiliation

National Department of Basic Education, South Africa

ADDITIONAL TRIAL INFORMATION

Status

Ongoing

Start Date

2015 01 05

End date

2022 06 30

Keywords

Education

Additional Keywords

Early Grade Reading, South Africa, Reading Support

ABSTRACT

The South African Department of Basic Education has tested interventions such as scripted lesson plans, teacher training and coaching to improve reading proficiency in learners taught in African home languages by teachers in under resourced schools. The EGRS I was an RCT implemented between 2015 and 2017 which tested three interventions: Treatment 1: Training, scripted lessons, graded readers, Treatment 2: Reading Coaches, scripted lessons, graded readers, Treatment 3: Parental involvement Programme. The EGRS I impact evaluation showed that the coaching treatment resulted in statistically significant improvements in African Home Language reading proficiency. A 2019 sustainability assessment of the EGRS I, indicated that reading proficiency gains were sustained after learners were no longer taught by supported teachers. In response to these encouraging findings, the Reading Support Programme was implemented to test whether a similar intervention can be implemented on a larger scale. The RSP involved all Foundation Phase teachers for two years (The EGRS I targeted one grade's teachers per annum and provided one year's support), and delivered the intervention in both Setswana Home Language as well as English First Additional Language (The EGRS I only provided support in Setswana Home Language). Similar to the EGRS, the RSP was designed to test if there were any reading proficiency gains in schools that benefitted from teacher training and scripted lessons, and there was a differential gain in those schools that also benefitted from literacy coaching support and those that benefitted from a leadership development intervention targeted at School Management Team members.

This 2021 impact evaluation will investigate 1) The impact of EGRS interventions on a cohort of learners tracked for six years from the beginning of Grade 1 (on interventions that ended in either Grade 2 or Grade 3). 2) The impact of RSP interventions (and Classroom Library Program) on a cohort of learners tracked for two years from Grade 1. 3) The impact of RSP interventions (and Classroom Library Program) on learners in Grade 4 in 2021. 4) The impact of both EGRS and RSP on a cohort of learners tracked for two years from Grade 1.

This is a continuation of the Early Grade Reading Study in South Africa, 2016, AEARCTR-0001749.

REGISTRATION CITATION

Roper, M et al. 2021. Sustainability Impact Evaluation of the Early Grade Reading Study (EGRS I) and Impact Evaluation of the Reading Support Project (RSP). South Africa.

2. SPONSORS AND PARTNERS

2.1. BACKGROUND

This Pre-Analysis Plan pertains to research in South Africa for the Data Collection and Analysis for the Early Grade Reading Study (EGRS[1]) and the Reading Support Project (RSP) in 2020.

To address the challenge of children not learning to read for understanding, in 2015 the DBE initiated the Early Grade Reading Study (EGRS) in two districts in the North West province (districts of Ngaka Modiri Molema and Dr. Kenneth Kaunda). The EGRS 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. 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 again in Grade 3 in 2017 – covering the Foundation Phase. The interventions showed significant impacts on learner results when teachers benefitted from training, coaching, and provision of learning materials. 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 on learners' ability to read, continued one year beyond the end of the intervention.

In 2019 and 2020, the DBE requested USAID's support in proceeding with an expansion of the EGRS. The focus of the expansion was to scale up the coaching intervention, which showed the largest impact in the initial evaluations. In response to this request, USAID and the Foundation for Professional Development (FPD) Consortium (comprising FPD, the Molteno Language Institute, Oxford University Press of South Africa, and Voluntary Services Oversees) modified the existing Reading Support Project (RSP) to include selected EGRS components. The RSP has since been implemented in 164 of the original 230 EGRS schools, with a further 50 schools serving as controls.

The DBE and USAID collaborates on this research to contribute to the body of research around early grade reading interventions. They appointed Khulisa Management Services (Khulisa) and their consortium partners Research Triangle Institute International (RTI), Research on Socio-Economic Policy (ReSEP) and Benita Williams Evaluation to carry out research with four objectives.

2.2. OBJECTIVES OF THE RESEARCH

Khulisa will conduct an impact evaluation of the EGRS, to determine if the reading gains made by learners who benefited from the EGRS have been sustained between 2018 and 2021. A previous sustainability impact evaluation of the EGRS I showed that the gains of the EGRS I had been sustained after learners were no longer taught by educators trained in the EGRS. The EGRS I sustainability impact evaluation, planned under USAID PERFORMANCE Task

[1] The original Early Grade Reading Study were replicated in Mpumalanga with a slightly different model. The latter is referred to as EGRS II. In this document we refer the original Early Grade Reading Study as the EGRS or EGRS I.

Order 4, provides the last opportunity to track learners who benefited from the EGRS before they go to secondary schools. With this additional round of data collection, the understanding of the longitudinal benefit of early grade reading interventions can be strengthened. Table 1 shows the EGRS Grade progression from 2018 to 2021.

At the same time, Khulisa will collect data for an impact evaluation of the Reading Support Project to determine if the intervention led to early grade reading gains of learners. The RSP is implemented in the same context as the EGRS, but with a revised design, and on a larger scale. Conducting the RSP impact evaluation in this assignment, may help to draw important lessons about the effect of early grade reading interventions taken to scale.

Table 1 EGRS Grade progression

EGRS progression of Grade sample (2018-2021)		
Grades in 2018	Grades in 2020	Grades in 2021 (during fieldwork)
NA	Grade 2	Grade 3
Grade 1	Grade 3	Grade 4
Grade 3	Grade 6	Grade 7

At the same time as the impact evaluations, two other research objectives will be fulfilled, drawing on the impact evaluation data.

The data collection for the impact evaluations will be expanded to establish letter-sound benchmarks and oral reading fluency thresholds and benchmarks for the early grades in Setswana and EFAL. Since these benchmarks currently do not exist, the research intends to fill an important gap that may benefit other work on early grade reading in future. Benchmarks help to establish the standard or level that every learner should attain at grade specific points to move through a successful language and reading trajectory journey in the primary school years. A separate analysis will be followed for the benchmarking exercise.

The beginning of 2020 marked the beginning of the COVID-19 pandemic and significant disruption to schooling. The effect of missed time in school, and the psychosocial impact of COVID-19 will be researched and analyzed in tandem with the data collection and analysis for Early Grade Reading Study (EGRS) and Reading Support Project (RSP).

This pre-analysis plan pertains to

EGRS Impact evaluation: to determine sustained impact, if any, of the original EGRS by administering student assessments of the original EGRS cohort now mostly in Grade 7, analyzing this data and reporting on the findings;

RSP Impact Evaluation: Conduct an impact evaluation of the RSP by administering student assessments in Grades 3 and 4, analyzing the data and reporting on the findings

3. EXPERIMENTAL DETAILS

3.1. DESCRIPTION OF INTERVENTIONS

DESCRIPTION OF THE EGRS I

This EGRS I was set up as a randomized control trial. The EGRS included three different interventions, all aimed at improving early-grade reading in Setswana. 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 again in Grade 3 in 2017.

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. 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. 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.

DESCRIPTION OF THE RSP

The RSP was set up as a randomized control trial, with six focus areas.

1. Quarterly 'just in time' training for teachers (from 263 schools) on the implementations of Curriculum and Assessment Policy Statement (CAPS) and lesson plans (all 263 schools)
2. The provision of Learning and Teaching Support Material (LTSM) packages (all 263 schools).
3. The provision of 14 literacy coaches to 140 schools to offer classroom-based support to Foundation Phase teachers (Coaching in 140 schools)
4. The development of the leadership capacity of principals/deputies and Head of Departments (HODs) to promote a culture of reading in their schools (SMT training in 65 schools)
5. Through the DBE, Classroom libraries are provided to 98 schools

The RSP specifically aims to: improve subject matter knowledge; promote more effective pedagogic practices; improve in-class time management; increase effective use of LTSM; and foster a school environment to support teachers' ability to implement the full curriculum and facilitate successful teaching and learning.

INTERACTIONS BETWEEN EGRS I AND RSP

In order to set up research questions and analyses accurately, it is helpful to understand how the EGRS and RSP interventions may interact and how they compare, in three areas.

1) Teacher Training

Teacher training for EGRS I focused on one language (Setswana) and trained one grade of teachers each year, for one year (Grade 1, then Grade 2, then Grade 3). RSP trained all three groups of teachers at the same time for two years in both Setswana and English. We may expect interactions between the two interventions in that some teachers may have received training in Setswana instructions under both programs.

2) Coaching

Coaches used in the EGRS program were not reemployed under RSP. Thus, we expect interactions to be limited in the coaching intervention of the two programs. It is possible that teachers who received coaching under EGRS may have been more experienced in getting feedback and may have benefited more from the RSP coaching as a result.

3) Learner and Teacher Support Materials

These materials were provided under both programs. In EGRS they were only provided for Setswana while the availability of EFAL materials to teachers was not provided and it is not clear what materials were used. In RSP materials were provided for both Setswana and EFAL. These included structured lesson plans, readers and other supporting LTSM. It is possible that the EGRS teachers were better prepared for LTSM use and further leveraged this when receiving RSP materials for both languages.

3.2. PRIMARY OUTCOMES

The primary outcome of interest is reading proficiency in Setswana home language (HL) and English first additional language (EFAL). For reading proficiency we will measure all the intermediate steps towards comprehension: letter recognition, phonetic awareness, word recognition and reading, vocabulary, reading fluency, and reading with comprehension. These tests are adapted from standard tests that have already been developed for the English First Additional language and Setswana language, such as the Early-Grade-Reading-Assessment (EGRA), but also newly developed and extensively piloted reading tasks. As our main outcome indicator, we will construct an aggregate indicator of learning proficiency, using principal component analysis.

3.3. DESCRIPTION OF THE EVALUATION DESIGN EGRS I AND RSP

EVALUATION DESIGN EGRS I

The EGRS I was designed as a randomized control trial (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 were selected in Grade 1 in 2015 and were tracked into Grade 2. Learners' reading proficiency was tested at the start of Grade 1, at the end of Grade 1, and again at the end of Grade 2.

Treatment Assignment and Random Selection for EGRS I impact evaluation

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)[2] dataset. An additional eight schools were excluded because they took part in the EGRS 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 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

[2] The Annual National Assessments (ANA) are standardised 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.

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.

EVALUATION DESIGN RSP

A sample of 214 schools that were part of the EGRS and part of the RSP 2018 baseline data collection is included in the impact evaluation. This includes all intervention groups. Lastly, 15 schools that were part of the original EGRS sample were not included in the RSP baseline because they had converted to multi-grade schools.

Schools were assigned to RSP treatment groups as follows. The 50 schools that were in the parent group under EGRS were assigned to be the comparison group for RSP, because the parent intervention had no impact. The comparison group received training and LTSM only. In addition, 80 schools that were not part of EGRS received the same (comparison) interventions under RSP. This group is not part of the RSP evaluation design, as they were not included in the baseline evaluation and have not received an intervention to be evaluated. Schools in the three remaining EGRS groups - Training (50 schools), Coaching (50 schools) and Control (80 schools) - received training and coaching under RSP. In addition, half of the schools in each group were allocated randomly to receive a School Management Training (SMT), in addition to coaching. Without schools dropping out, this procedure should have resulted in 90 schools receiving coaching alone, 90 schools receiving coaching and SMT and 50 schools receiving only training and LTSM, received by all other groups. However, approximately 60 schools dropped out of the study, according to the most recent (January 2021) DBE database. Table 2 shows the number of schools remaining in the RSP and EGRS groups.

Table 2 Number of schools in EGRS and RSP treatment groups

	EGRS Group					Total	
RSP Group;	Control	Training	Coaching	Parent	Not included	Overall	In the Evaluation
Coaching	24	22	18	0	6	**70**	**64**
Coaching and SMT	31	12	20	0	7	**70**	**63**
LTSM only	0	0	0	43	80	**123**	**43**
Not included	25	16	12	7	0	**60**	**0**
Total	**80**	**50**	**50**	**50**	**93**	**323**	**170**

These numbers are provisional. We recommend visiting or contacting all 214 schools from the RSP baseline to confirm whether or not they received RSP interventions.

As shown in Table 3, the 263[3] schools receiving RSP interventions are as follows:

- **Group 0:** 80 schools not included in EGRS (LTSM and Teacher Training only; not part of the RSP evaluation design)
- **Group 1:** 43 schools included in EGRS parent intervention (now receiving LTSM and Teacher Training only)
- **Group 2:** 140 schools receiving Coaching, LTSM and Teacher Training schools (of which 127 were in EGRS and part of the 2018 RSP baseline)
- **Group 3** (a *subset* of Group 2): 70 schools also receiving an SMT program (of which 63 were in EGRS and part of the 2018 RSP baseline)

Within the main experimental groups, 100 schools were randomly selected to take part in the Classroom Library Project. As shown in Table 3, 14 of these libraries were assigned to schools that were not part of EGRS and therefore not part of the current evaluation design.

Table 3 Provision of Classroom Libraries by RSP Group

	Included in EGRS			**Not Included in EGRS**		
	Classroom Libraries			**Classroom Libraries**		
RSP Group	Yes	No	Total	Yes	No	Overall Total
Coaching	32	32	**64**	0	6	**70**
Coaching & SMT	32	31	**63**	3	4	**70**
LTSM only	22	21	**43**	1	69	**123**
Total	**86**	**76**	**170**	**14**	**79**	**263**

[3] 263 school participated in the RSP. The sample for this evaluation is 214 of these schools. However the final number may change upon verification of the school participation during this evaluation.

3.1. EXPERIMENT CHARACTERISTICS

SAMPLING EGRS I SUSTAINABILITY IMPACT EVALUATION

Since the start of EGRS, one school has closed leaving 229 of the original 230 schools. These 229 schools form the current sample for data collection. Data will be collected on all learners for whom data was collected in Grade 1 and who are 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. We expect this number to be reduced to approximately 12 learners in 2021 (see below). Although learners are expected to be in Grade 6 this year, all learners from the original sample will be tested regardless of Grade (as long as they can be found within the school).

SAMPLING RSP IMPACT EVALUATION

The RSP sample consists of 214 schools of the 229 schools in the EGRS sample because 15 of the original schools had moved to multi-grade teaching, which rendered grade-specific EGRS lesson plans inappropriate.

The Grade 3 sample for this study will be based on the 2018 RSP baseline, which consisted of 20 Grade 1 learners randomly selected in each of the 214 schools. The sample will consist of all learners for whom baseline data are available and who are present at the school on the day of assessment. Based on previous attrition rates we anticipate around 16 learners per school to be available in this sample with around 13 learners in Grade 3 and approximately three learners in lower grades.

Table 4 Learner Sample

Program	Data collection in 2021 cohort	Previous data collection wave (2018)	Average number of traceable learners per school [4]	Number of schools	Total learners
EGRS	Grade 6 (some in Grade 4 & 5)[5]	Grade 4	12	229	2,748
RSP	Grade 3 (some in Grade 1 & 2)[6]	Grade 1	16	214	3,424
RSP	Grade 2[7]	Not applicable (New sample)	16	214	3,424
					7,715

[4] This is our assumptions for the number of learners that may be traceable in 2021. The section on sample size calculations indicate what the anticipated MDES is with this sample size. In the original RFTOP, the prescribed sample sizes were: Grade 6: 15 learners, Grade 2: 10 learners, Grade 3: 10 learners. The summary sample table indicates our estimate for the number of learners that can be tested within the 2 days that are allocated for fieldwork in each school.

[5] This is the cohort of learners that were tested in Grade 1 in 2015, were part of Wave 4 data collection in 2015, is mostly in Grade 6 in 2020, and is mostly in Grade 7 in 2021.

[6] This is the cohort of learners that were tested in Grade 1 in 2018, and will mostly be in Grade 3 in 2020, and in Grade 4 in 2021

[7] This is the cohort of learners that were in Grade 2 in 2020, and will be mostly in Grade 3 in 2021

SUMMARY OF THE STUDY POPULATION

The study population is in the North-West province of South Africa, in the districts of Dr. Kenneth Kaunda and Ngaka Modiri Molema.

The data collection will be conducted in the two districts. The settings within which the research will take place are exclusively Quintile 1 – 3 schools,[8] which use Setswana as the language of learning and teaching in the Foundation Phase (Grade 1-3 classrooms).

Data will be collected across a sample of 229 treatment and control schools.

The population of interest includes:

- Sample of learners identified from the 229 schools
- Parents of learners being assessed in the 229 schools
- Teachers in Grades 3 and Grade 7 in the participating schools in the North-West province who were part of the EGRS I and RSP implementation
- Principals and SMT members in participating schools
- North West Provincial officials and District officials in Kenneth Kaunda and Ngaka Modiri Molema districts.

SAMPLE SIZE CALCULATION

Sample size calculations are conducted to inform design decisions. Below we calculate the number of learners required to achieve a reasonable minimum detectable effect size (MDES) of approximately 0.21 SDs (the MDES for the originally-designed EGRS comparison between treatment and control). For the RSP Grade 3 and EGRS Wave 6 evaluations, the sample size is determined by the number of learners who were assessed at baseline and who are present on the day of assessment. We based the calculations below on the sample sizes estimated in and using a longitudinal design (i.e., controlling for baseline values). The RSP Grade 2 evaluation uses a cross-sectional design (i.e., there is no baseline) and therefore relies only on a comparison across groups being made at a single time point.

We have also conducted sample size calculations (and associated MDES) for two evaluation approaches for both EGRS and RSP. In both cases, we estimate the MDES for the original, longitudinal design and also include an estimate of the MDES for a cross-sectional design that includes a 'top-up' of students. It is clear from the results in Table 5 that the marginal benefit of the cross-sectional top-up is minimal and we therefore recommend using the longitudinal design. Our recommended samples sizes are therefore ~12 students in EGRS Grade 6; ~16 students in Grade 2 and ~16 students in Grade 3 for RSP. Additional assumptions and explanations are provided in the section below Table 5.

[8] In South Africa, public schools are grouped into quintiles, which describe the wealth of the school, and therefore how much government funding they are entitled to. Quintile 5 schools are the wealthiest schools in the country, while Quintile 1 schools are the poorest. Quintile 1 – 3 schools do not charge school fees, and a large proportion of these schools participate in the National School Nutrition program where learners receive a meal at the school. These schools are legally not allowed to charge school fees, and are usually the worst performing in the system.

Table 5 Sample size calculations (longitudinal versus cross-sectional)

Assumption	EGRS Longitudinal	EGRS Cross-Sectional	RSP Grade 3 Longitudinal	RSP Grade 3 Cross-Sectional	RSP Grade 2 Cross-Sectional
Zα =	1.6450	1.6450	1.6450	1.6450	1.6450
Zβ =	0.8416	0.8416	0.8416	0.8416	0.8416
Group 1 Total Schools	80	80	70	70	70
Group 2 Total Schools	50	50	43	43	43
Group 1 Students/School	12	20	16	20	16
Group 2 Students/School	12	20	16	20	16
Group 1 Students	960	1600	1120	1400	1120
Group 2 Students	600	1000	688	860	688
ICC	0.2	0.2	0.2	0.2	0.2
Design Effect	3.2	4.8	4	4.8	4
Corr	0.7	n/a	0.7	n/a	n/a
Minimum Detectable Effect Size (MDES)	**0.18**	**0.22**	**0.19**	**0.237**	**0.241**

RSP Impact Grade 2 (2021)[9]

Based on the equation and assumptions below, we recommend sampling 16 learners per school for a cross-sectional analysis in order to detect an effect size of 0.24 or larger.

The estimate is based on the following parameters:
$\boldsymbol{\mu_1}$ is the control's mean of the Grade 2 learners' scores.
$\boldsymbol{\mu_2}$ is the treatment's mean of the Grade 2 learners' scores.
A difference of 5 cwpm[10] ($\boldsymbol{\mu_1 - \mu_2 = 5}$) is equivalent to an effect size of 0.2, with a standard deviation of 25.
$\boldsymbol{\sigma_1}$ is the standard deviation of the control's scores. 25
$\boldsymbol{\sigma_2}$ is the standard deviation of the treatment scores. 25
$\boldsymbol{1 - \beta}$ is the power set at 0.8
$\boldsymbol{\alpha}$ is the alpha 0.05
$\boldsymbol{\rho}$ is the intraclass correlation. 0.2
$\boldsymbol{m}$ is the average cluster size (average number of learners sampled within each school). [to be solved for]

[9] This is the cohort of learners that were in Grade 2 in 2020, and will be mostly in Grade 3 in 2021
[10] correct words per minute

RSP Impact Grade 1 (2018) – Grade 3 (Grade 4 or lower in 2021)[11]

Given the current sample size of schools and assuming an attrition rate of 17 percent, we would expect an average of 16 learners who were assessed in Grade 1 in 2018 to be assessed in 2021. Given the following assumptions below, we expect the minimal detectable effect size approximately 0.18 with a longitudinal analysis. This translates to an approximate gain of five correct words per minute in the measure of reading fluency of the intervention groups over the comparison group.

$\boldsymbol{\mu_1}$ is the control's mean gain in learners' scores from Grade 1 to Grade 3.
$\boldsymbol{\mu_2}$ is the treatment's mean gain in learners' scores from Grade 1 to Grade 3.
Assuming a difference of 5 cwpm, $(\boldsymbol{\mu_1 - \mu_2 = 5})$
$\boldsymbol{\sigma_1}$ is the standard deviation of the difference in comparison group scores. 22
$\boldsymbol{\sigma_2}$ is the standard deviation of the difference in treatment group scores. 22
$\boldsymbol{1 - \beta}$ is the power 0.8
α is the alpha 0.05
$\boldsymbol{\rho}$ is the intraclass correlation. 0.2
$\boldsymbol{m}$ is the average cluster size (average number of learners sampled within each school). 15
r is the correlation coefficients between the baseline (2018) and endline (2021) = 0.7

EGRS 1 Impact Grade 1 (2015) – Grade 4 (2018) – Grade 6 (Grade 7 or lower in 2021)[12]

The current sample size of schools and assuming an attrition rate of 40 percent, we would expect on average 12 learners (who were assessed in Grade 1 in 2015) per school will be assessed in 2021. Given the following assumptions below, we expect the minimal detectable effect size would be approximately 0.18 for this longitudinal analysis.

$\boldsymbol{\mu_1}$ is the control's mean gain in learners' scores from Grade 1 to Grade 6.
$\boldsymbol{\mu_2}$ is the treatment's mean gain in learners' scores from Grade 1 to Grade 6. $\boldsymbol{\mu_1 - \mu_2}$ to be solved for. A difference of 8 cwpm with a standard deviation of 32 would give an effect size of 0.25**.**
$\boldsymbol{\sigma_1}$ is the standard deviation of the difference in control's scores. 32
$\boldsymbol{\sigma_2}$ is the standard deviation of the difference in treatment scores. 32
$\boldsymbol{1 - \beta}$ is the power 0.8
α is the alpha 0.05
$\boldsymbol{\rho}$ is the intraclass correlation. 0.2
$\boldsymbol{m}$ is the average cluster size (average number of learners sampled within each school). 12
r is the correlation coefficients between the baseline (2015) and endline (2021). 0.7
(This applies to the longitudinal design only)

[11] This is the cohort of learners that were tested in Grade 1 in 2018, and will mostly be in Grade 3 in 2020, and in Grade 4 in 2021
[12] This is the cohort of learners that were tested in Grade 1 in 2015, were part of Wave 4 data collection in 2015, is mostly in Grade 6 in 2020, and is mostly in Grade 7 in 2021.

4. METHODS OVERVIEW

4.1. DATA SOURCES AND INSTRUMENTS

Analyses will make use of three data sources.

1) Data collected as part of the current evaluation study. These include:
    a. Principal Questionnaire
    b. Grade 3 Teacher questionnaire
    c. Grade 7 Teacher questionnaire
    d. Head of Department Questionnaire
    e. Classroom Observation and Learner Workbook Assessment
    f. School functionality assessment
    g. Home/parent questionnaire
    h. Learner assessment for English First Additional Language
    i. Learner assessment for Home Language (Setswana)

2) The COVID-19 research component of Task Order 4 will investigate how schooling and curriculum, learner and teaching performance were affected by COVID-19.
3) Data collected as part of the RSP design and implementation evaluation conducted under USAID PERFORMANCE Task Order 3.

4.2. ANALYSIS METHODS

Main impact analyses

The first analyses will address the key research question for each study – the impact of interventions on reading outcomes. Three sets of impact analyses will be conducted:

(1) The impact of EGRS interventions on the cohort tracked for six years from the beginning of Grade 1 (on interventions that ended in either Grade 2 or Grade 3)

(2) The impact of RSP interventions (and Classroom Library Program) on the cohort tracked for two years from Grade 1.

(3) The impact of RSP interventions (and Classroom Library Program) on learners in Grade 4 in 2021.

(4) The impact of both EGRS and RSP on cohort tracked for two years from Grade 1

Our main estimating equation will be:

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

Where y_{isb1} is the is the outcome indicator of interest (such as the score on an EGRA sub-test) for pupil 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 baseline controls (such as age, gender etc.), and ε_{isb1} is the error term clustered at the school level. To estimate the respective impacts of the three interventions, we restrict the sample to the control schools and the schools from the relevant treatment group.

Covariates

For the analysis of EGRS I impact and RSP impact in Grade 3 we will conduct longitudinal analyses which control for baseline measures of reading skills. This includes a baseline control for each domain of reading proficiency collected at baseline: vocabulary, letter recognition, working memory, phonological awareness, word recognition, words read, and sentence comprehension. In a second analysis, we will control for individual and community level characteristics (e.g., SES, measures of school functionality) which are highly correlated with y_{isb1} or were imbalanced at baseline and which are exogenous to treatment effects (i.e., cannot plausibly have been impacted by the interventions). The aim of the second analysis is to increase statistical power and account for any incidental differences that may exist between treatment groups. To assess the impact of RSP in Grade 2 we will include only school- and community-level baseline covariates; no individual-level covariates were collected at baseline for this sample. Depending on decisions made on the basis of sample size calculations, we may conduct similar analyses – i.e., without individual baseline-covariates – for EGRS in order to increase the sample size by including learners who were not assessed at baseline.

An important additional covariate for the evaluation of the RSP interventions is EGRS group to which the school had previously been assigned. It is possible that there is an interaction between the treatment status of the school under EGRS and RSP. For example, teacher training may be more effective under RSP if the same school had been allocated to receive teacher training in EGRS – perhaps because the same teachers were included in the training, or because of a change in school culture. Additional analyses of RSP impact will include binary variables indicating membership of EGRS treatment group.

A final additional analysis will be conducted controlling for the impact of the COVID-19 pandemic on learner outcomes. The impact of COVID-19 is not exogenous to the treatment effect. That is, the response to the pandemic may have differed by treatment group. For example, teachers who improved their skills under RSP may have been more motivated to continue instruction during the pandemic. For these reasons, the effect of COVID-19 should be conducted after the effect of all exogenous covariates (described above) have been considered. Previous EGRS data collection wave data on Grade 2 will be compared with current Grade 2 data to account for.

For the EGRS evaluation, we will construct a longitudinal model to examine predictors of growth in reading scores since the last evaluation. Predictor variables will include the EGRS intervention to which children were exposed, SES as a proxy for the support provided to their learning by their families and their experience of the COVID-19 pandemic.

5. HYPOTHESIS UNDER INVESTIGATION

This section outlines the main and intermediate outcomes we want to measure, as well as heterogeneous treatment effects of interest. In brackets we indicate the source of the data as well as question number if it is based on a survey. If an outcome is based on multiple indicators, it will be aggregated using principal component analysis.

[*SPQ = School Principal Questionnaire; TRQ = Grade 3 and Grade 7 Teacher Questionnaire; HODQ = Foundation Phase Head of Department Questionnaire; PHQ = Parent Home Questionnaire; CLW = Classroom Observation and Workbook Assessment; LAL = Learner Asset list; LWQ = Learner Wellbeing Questionnaire].*

5.1. MAIN OUTCOMES

Our main research questions are:

EGRS

What is the impact of (i) three years of Coaching + TT + LTSM (ii) three years of TT + LTSM and (iii) two years of Parental Involvement on reading outcomes in Setswana and EFAL, compared to a comparison group receiving no intervention, nearly four years after the completion of all interventions.

RSP

What is the impact of (i) SMT + Coaching + TT + LTSM and (ii) Coaching + TT + LTSM compared to a comparison group receiving TT + LTSM alone on:

- Reading outcomes in Grade 2 and 3 in Setswana and EFAL?
- Classroom processes, teacher behavior, school functionality?

Within the main experimental groups, what is the impact of the Classroom Library Program on Reading outcomes in Grades 2 and 3 in Setswana and EFAL?

What is the relationship between fidelity of implementation (i.e. attendance at teacher and SMT training, participation in individual and group coaching) and the improvement in reading in the three groups?

What are the school-level factors mediating the effect of each intervention on learning outcomes?

How do treatment effects differ between subgroups based on, for example, pupil gender, teacher experience, school location and household profile in parent/guardian education, employment, literacy practices.?

What is the impact of:

i) 5 years of coaching (EGRS 3 years + RSP 2 years) + 5 years of TT + 5 years of LTSM

ii) 2 years of RSP coaching + 5 years of TT + 5 years of LTSM

iii) 2 years of parental involvement + 2 years of RSP coaching +2 years of RSP TT+ 2 years of RSP LTSM

The skills to be assessed include

Grade 3 learners:

- **Home language:** Rapid object naming, letter sound recognition, complex consonants & diacritics recognition, word reading, oral reading fluency and comprehension and listening comprehension.
- **EFAL:** Decodable word reading, sight word reading, oral reading fluency and comprehension.

Grade 4 learners:

- **Home language:** Rapid object naming, letter sound recognition, complex consonants & diacritics recognition, word reading, oral reading fluency and comprehension, written text comprehension.
- **EFAL:** Decodable word reading, oral reading fluency and comprehension, written text comprehension.

Grade 7 learners:

- **Home language:** Oral reading fluency and comprehension, written text comprehension, vocabulary.
- **EFAL:** Oral reading fluency and comprehension, written text comprehension, vocabulary.

The instruments include tasks that were adapted from standard early grade reading assessments in Setswana and English firs additional language, as well as newly developed and extensively piloted reading tasks.

As our main outcome indicator, we will construct an aggregate indicator of HL and EFAL learning proficiency, using principal component analysis.

5.2. INTERMEDIATE OUTCOMES

For the evaluation of RSP we will assess the role of intermediate outcomes in achieving impact in two categories – fidelity of implementation and mediators.

Fidelity of Implementation

The fidelity of implementation analysis will aim to understand whether schools implementing the interventions to a higher standard achieved greater improvements in reading outcomes. To do this, we will identify candidate variables from the contextual tools during the data collection pilot, in addition to data already collected from the RSP implementation to indicate fidelity of

implementation (e.g., teacher use of national workbook) and assess the relationship between each of these variables and reading outcomes. It is not possible to identify a causal relationship between fidelity of implementation and reading outcomes because of confounding factors. To control for observed confounding factors, we will include baseline characteristics of schools as covariates in these analyses.

Data available from the RSP implementation evaluation include: number of teachers per school attending training and receiving coaching and whether schools received project materials.

Mediators

The mediator analysis aims to identify the mechanism for the impact of interventions on outcomes. Mediators are typically school level processes (e.g., teacher corrects learner work) that are enhanced by the intervention and which lead to improved reading outcomes. Mediator variables that meet some or all of the following criterial will be considered.

(1) The intervention has a significant impact on the mediator

(2) There is a significant relationship between the mediator and reading outcomes

(3) The impact of the treatment variables on HL and EFAL reading proficiency is attenuated when the mediator is included in the regression equation.

Below are potential mediating variables (these will be determined during the final data analysis).

Teachers

Did the EGRS I and RSP treatment arms which involves training and coaching change **teaching practice and effort**?

1. Provide regular individualized assessment [*TRQ – 3.7 and 3.10, CLW Section 2 document review)]*
2. Require student to read out loud [*TRQ – 3.7 and 3.10)*]
3. Stream by ability [*TRQ – 3.5*]
4. Assess reading ability [*TRQ – 3.7 and 3.10*]
5. The completion of more writing exercises [*TRQ – 3.7 and 3.10, CLW Section 2 document review*]
6. The assessment of more writing exercises [*CLW Section 2 document review*]
7. Use of the national workbook [*CLW Section 2 document review*]
8. Increase the availability of home language text in classrooms [CLW Section 3]

Parents

Did EGRS I treatment arm lead to a change in **parent attitudes and behavior**?

1. Happy with child's reading progress [PHQ -- 27]
2. Read to child [PHQ – 19]
3. Support child doing homework [PHQ – 30 and 31]

Pupils

1. Pupil attendance [PRQ – 25]
2. Pupil do homework on days not at school [PHQ – 29]

Heterogeneous Impacts

At the pupil level, we could expect two opposing heterogeneous treatment effects of the EGRS and RSP interventions, based on baseline pupil reading proficiency. The scripted lesson plans require streaming by ability *within* the same classroom and provides opportunity for individualized attention and could benefit children who have otherwise been left behind. However, at the same time the scripted lesson plans are aligned to the national curriculum, which prescribes an ambitious pace in the South African context. The worst-performing pupils might benefit less if the teachers who follow the scripted lesson plans now progress at too fast a pace. Furthermore, boys/girls might benefit more/less from the individualized attention. Finally, the emphasis on individualized attention and tracking means that pupils might benefit more from the scripted lesson plans when the class size is large.

For the RSP evaluation, we expect that the success of the interventions depend on teacher motivation, prior levels of effort, and ability. The scripted lesson plans will only be applied by teachers who have a sufficient level of intrinsic motivation. Related, teachers who have a higher burden of lesson preparation are most likely to switch to scripted lesson plans, because they have most to gain from scripted lesson. Furthermore, teachers may need a sufficient baseline level of reading proficiency in order to effectively apply the scripted lesson plans. On the other hand, the scripted lesson plans might be too restrictive for exceptional teachers who are effective at adjusting their instruction to the needs of the classroom and ability of the pupils.

We will therefore examine the following possible interactions:

Learner level

1. Learner gender [learner information form]
2. Learner age [learner information form]
3. Learner attendance of grade R [PHQ 15]
4. Learner attendance of other ECD [PHQ 18]
5. Learner baseline performance [pupil baseline test]
6. Learner attendance [PHQ 25, LWQ 2]
7. Learner speaks English at home [LWQ B1
8. Learning when not at school [PHQ 29 ,30; LWQ 2.2]
9. Learner reported amenities and possessions [LAL 1]
10. Learner reported books in home [LAL 2]

Teacher level

1. Teacher characteristics:
    a. Education level [TRQ 2.5]
    b. Training in Foundation Phase [TRQ 2.8]
    c. Years of experience [TRQ 2.3]
2. Teacher Attendance [TRQ 4.1]

3. The extent to which teachers feel supported and recognized in their work [TRQ 5.2l]
4. The sufficiency of the classroom LTSM [TRQ 3.1, 3.2, 3.3., 3.4]

School-level

1. Socio-economic background of school [administrative data]
2. School resources [SFO Section 2]
3. School urban / rural [administrative data]
4. School district [administrative data]
5. Pupil-teacher ratio [administrative data, SPQ section 3]
6. School days lost [SFO 5.1, SPQ 7.5, 7.7, 7.8, 7.11)]
7. Teacher absenteeism [SPQ 5.1]

Parent level

1. Identity of the parent/guardian: [PHQ 8]
   E.g., single mother/grandmother; child-headed household.
2. Education level [PHQ 11,12]
3. Parent employment [PHQ 13]
4. Household amenities and possessions [PHQ14]
5. Learner use of English at home [PHQ 21, 22]

5.3. ADDITIONAL CORRECTIONS

Assessing the Impact of Attrition

Attrition threatens validity if it differs between treatment groups. For example, if struggling learners in the intervention group are less likely to drop out of school than the control group, there may be a greater representation of weaker learners in the control group in the final sample. We will assess attrition in two stages. In the first stage, we will assess whether learners included in the final sample differ in baseline characteristics from those who were included at baseline. If such a difference is found (as is likely), it will imply that the final sample of learners is not representative of the initial sample. The implications of such a finding should be discussed, but they do not undermine the validity of the impact evaluation. In the second stage, we will look for evidence of differential attrition – a potential threat to validity of the impact evaluation. We will assess whether the three intervention groups differ in the number of learners not present in the final sample, or if the baseline characteristics of the final sample differs among groups. If there is evidence of differential attrition we will consider conducting analyses with imputation to assess the effect of differential attrition on treatment outcomes.