

Research Findings

Teacher Benchmark Usage

Pilot Study Summary Report

Asanda Lobelo (University of Cape Town); Nompumelelo Mohohlwane (Department of Basic Education); Lesang Sebaeng (Department of Basic Education); Zamangwe Zwane (Department of Basic Education); Dr Ramashego Mphahlele (UNISA); Prof. Cally Ardington (University of Cape Town)

September 2023



basic education

Basic Education REPUBLIC OF SOUTH AFRICA







Table of Contents

Table o	of Contents	1
Acronyr	/ms	3
Table o	of Figures	4
Table o	of Tables	5
1. Intr	troduction	6
2. Ba	ackground	7
3. Stu	udv Aims and Research Questions	
4 1	Reading assessment practices and efficacy	q
4.0	Evaluation of nilot intervention	10
4.2.		
5. Inte	tervention details	
5.1.	Early Grade Reading Benchmarks	
5.2.	Adapted EGRA Tool	13
5.3.	Frequency of assessments and scheduling options	
5.3	3.1. Home language Lesson (Group Guided Reading)	16
5.4.	Interpretation of assessment results	17
5.5.	Teacher training	
6. Re	esearch design	
6.1.	Sample	19
6.2.	Instruments	
6.2	2.1. Literacy Outcomes	
6.2	2.2. Teacher Interview	
6.2	2.3. HOD/Principal Interview	
6.2	2.4. Qualitative in-depth interviews	21
7. Da	ata Collection	21
7.1.	Fieldwork Team and Training	21

7.	2. Da	ata Collection activities	22
7.	3. Re	alized samples	22
	7.3.1.	Teacher attendance at training	22
	7.3.2.	Teacher and Learner Sample	23
7.	4. Sa	mple Characteristics	24
8.	Results	S	25
8.	1. Re	eading assessment practices and efficacy	25
	8.1.1.	Assessment practice	25
	8.1.2.	Assessment efficacy	29
8.	2. Ev	aluation of pilot intervention	33
	8.2.1.	Intervention compliance and fidelity	34
	8.2.2.	Teacher experience of the intervention	36
	8.2.3.	Understanding of the intervention	37
	8.2.4.	Key challenges with implementation	38
	8.2.5.	Indications of some improvements	39
	8.2.6.	Assessment efficacy	39
	8.2.7.	Estimated impact of pilot intervention	41
9.	Conclu	sion	42
Refe	erences		45

Acronyms

CLPM	Correct Letters per Minute
CWPM	Correct Words per Minute
DBE	Department of Basic Education
EGRA	Early Grade Reading Assessment
HOD	Head of Department
J-PAL	Abdul Latif Jameel Poverty Action Lab
LOLT	Language of Learning and Teaching
ORF	Oral Reading Fluency
PIRLS	Progress in International Reading Literacy Study
RCME	Research Coordination, Monitoring and Evaluation
SALDRU	Southern Africa Labour and Development Research Unit
UCT	University of Cape Town

Table of Figures

Figure 1: Theory of Change	. 11
Figure 2: Letter Sound Knowledge Chart	. 14
Figure 3: isiZulu ORF Passage Teacher Chart	. 14
Figure 4: Class Progress Chart	. 15
Figure 5: Teacher and Class Characteristics	. 25
Figure 6: Distribution of Average EGRA duration	. 27
Figure 7: Teacher Estimate vs Learner Performance at Baseline	. 30
Figure 8: Teacher Estimates of learner benchmark category	. 32
Figure 9: Distribution of proportion of learners correctly classified at baseline	. 33
Figure 10: Teacher rating of intervention relative to EGRA	. 35
Figure 11: Picture of Used Progress Chart	. 36
Figure 12: Picture of assessment letter sound chart given to learners	. 37
Figure 13: CWPM - Teacher Estimate vs Learner Score at Endline	. 40
Figure 14: Spearman correlation coefficient by treatment status	. 41

Table of Tables

Table 1: Usage of Early Grade Reading Benchmarks	7
Table 2: Early Grade Reading Benchmarks for Nguni and Sesotho-Setswana Languages	12
Table 3: Letter Sound Knowledge Benchmark Categories	13
Table 4: Oral Reading Fluency Benchmark Categories	13
Table 5: CAPS Instructional Time by Grade and Activity	17
Table 6: Sampled Schools by DBE EGRA training status	19
Table 7: Timeline of Data Collection Activities	22
Table 8: Training invitation and attendance	23
Table 9: Realised Learner and Teacher Sample by province	24
Table 10: Teacher prior EGRA exposure by Province	26
Table 11: Proportion of teachers administering EGRA by May	26
Table 12: Teacher EGRA Practices	28
Table 13: Teacher usage of EGRA Results	29
Table 14: Correlation between teacher ranking and EGRA ranking	31
Table 15: Take-Up Rates of Intervention	34
Table 16: Treatment teacher categories	35
Table 17: Difference-in-Difference Coefficients	42

1. Introduction

In response to South Africa's 'reading crisis', various ongoing initiatives aim to support early grade reading. These initiatives involve providing reading materials, promoting reading culture, enhancing teacher training and setting language-specific benchmarks for foundational reading skills. The Department of Basic Education (DBE) has led efforts since 2019 to establish such benchmarks for different languages, considering phonological, morphological and orthographic differences between languages. Benchmarks have been established for Nguni, Sesotho-Setswana, Afrikaans and English as a First Additional Language (Mohohlwane, Wills & Ardington, 2022).

The benchmarks have three main purposes outlined in Table 1. Firstly, they establish a national and provincial definition of proficient reading, aiding in target setting and standards monitoring. Secondly, they are valuable at the school level for setting targets, standardizing assessment practices, and pinpointing areas for improvement. Lastly, benchmarks offer classroom-level targets for teachers and learners, helping teachers define successful progress criteria and identify students at risk of not achieving meaningful reading skills by the end of the Foundation Phase. Reading Benchmarks also assist in interpreting assessment results, enhancing teachers' grasp of individual reading proficiency. This understanding is important for implementing curriculum components that group learners by their proficiency levels, and for effective remediation and consolidation when students fall short of curriculum expectations.

Table 1: L	Jsage of	Early Grade	Reading	Benchmarks
------------	----------	-------------	---------	------------

NATIONAL AND PROVINCIAL ADMINISTRATION	SCHOOL	CLASSROOM		
Establishes definition of reading proficiency	Standards and targets that school leaders can aim towards	Standard against which to measure learner skills		
Clearly communicates standards and targets	Standardises assessment practices across and within schools	Identify early on learners at risk of not being able to read		
Monitor progress	Identify the extent of remedial support required	Adapt instructional focus to meet learners' needs		

The study discussed in this report is a collaboration between the DBE, University of Cape Town (UCT), Southern Africa Labour and Development Unit (SALDRU), and Abdul Latif Jameel Poverty Action Lab (J-PAL) Africa, funded by the Zenex Foundation. It builds on the created benchmarks. Its focus is on the classroom level potential of these benchmarks, particularly in clarifying reading curriculum expectations and enhancing formative assessment practices. The study collected and analysed longitudinal data from 197 grade 2 and 3 teachers across 39 schools in four provinces. The goal was to understand existing assessment practices and assess the impact of an intervention that introduced the newly established benchmarks to teachers. This is the summary report, a long form of this report is also available.

2. Background

Accurate formative assessment of students' academic abilities is important for effective teaching. This accurate assessment, combined with understanding curriculum expectations, helps teachers measure progress compared to curriculum requirements and identify struggling students who need extra support. Additionally, it is crucial for implementing strategies that group students by ability and provide targeted instruction.

However, South African schools face challenges in implementing effective formative assessment due to factors like insufficient teacher training, limited resources, and time constraints in completing individual assessments (Kanjee & Bhana, 2022). Large class sizes and teacher-led teaching methods focused on whole-group activities limit individualized assessment and feedback

opportunities (Hoadley, 2018). Research indicates that teachers often lack a clear understanding of their students' learning levels. Studies have shown a weak connection between grade advancement and actual learning ability, a low correlation between continuous assessment scores and external exams, and significant discrepancies between learner performance in independent assessments and teacher estimations thereof (Lam et al., 2011; Van der Berg & Shepherd, 2015; Ardington & Meiring, 2020).

The Department of Basic Education (DBE) introduced Early Grade Reading Assessments (EGRA) to enhance formative assessments practices in reading. Developed in 2006, the EGRA is available in across languages, cost-effective and user-friendly for teachers or fieldworkers. EGRA evaluates foundational reading skills such as the alphabetic principle, phonemic awareness, word recognition, fluency, and comprehension (Mohohlwane, Wills & Ardington, 2022). Since 2007, EGRA has been adapted into all 11 official languages with teacher training of EGRA led by subject advisors across primary schools (Department of Basic Education, 2010). While its implementation remains limited, EGRA acts as a classroom resource for teachers with minimal data comparison across schools within provinces.

At the time of EGRAs development, Early Grade Reading benchmarks in African languages were not yet available. The DBE made use of expert knowledge to form reading benchmarks for those EGRA assessments. The newly established reading benchmarks improve upon the benchmarks that were used in the past in that they are evidence-based, they incorporate expert knowledge, and they allow for language specific reading score interpretations.

The administration of the Early Grade Reading Assessment (EGRA) has posed challenges for teachers, especially those with large classes. Tasks involving mental math adjustments, handling multiple documents, and deciphering lengthy instructions that are required by EGRA are demanding for teachers. In classrooms with many students, individually assessing each learner consumes significant time, hindering frequent use of the EGRA as a formative assessment tool. Revisions to the EGRA that alleviate cognitive demands on teachers and reduce assessment time hold the potential to enhance its usability for tracking reading proficiency advancement in the Foundation Phase.

In recent years, there has been growing interest in differentiated and targeted instruction in many developing countries to address poor learning outcomes. Many learners struggle to meet curriculum requirements, falling behind as teachers progress without considering individual understanding. The Teaching at the Right Level (TaRL) approach, pioneered by the Indian NGO

Pratham, targets instruction to learners' actual knowledge rather than their grade level as specified by the curriculum. These programs typically group students based on their learning level for part of the school day or provide supplementary support after regular classes and have demonstrated effectiveness in enhancing learner performance, particularly for those at risk (Banerjee et al., 2007; Bassi et al., 2020). Such strategies require accurate assessment of learners' abilities.

3. Study Aims and Research Questions

The purpose of this study is to generate insights into how the newly established benchmarks could be productively used in South African classrooms. To this end, the study has the following two key aims:

- To inform our understanding of teachers' use of the EGRA tool and general assessment practices with respect to reading in the Foundation Phase.
- To implement and evaluate a pilot programme to introduce newly established African language benchmarks into Foundation Phase classrooms.

These aims are further discussed in the section below:

4.1. Reading assessment practices and efficacy

The first objective of this research is to enhance our understanding of how teachers make use of the EGRA tool, the obstacles they encounter and their assessment methods. This aim is exploratory due to the scarcity of existing research on teacher assessment practices in South Africa, especially regarding the implementation of the EGRA tool.

The research questions with respect to this aim are as follows:

- 1. What is teachers' orientation to the EGRA tool? Have they received training on the EGRA? Are they using the tool?
- 2. Where teachers are using the EGRA, how are they implementing it? How long does it take? How do they manage the class? What are the challenges?
- 3. How do they use the results from EGRA? Is there a feedback loop between assessment results and pedagogic and instructional practice?

4. Do teachers know the reading level of the learners in their classroom? Does this differ by whether they have received DBE EGRA training or have implemented EGRA?

In responding to these questions, we hope to contribute to knowledge of what is currently happening in classrooms with respect to assessment of early grade reading skills.

4.2. Evaluation of pilot intervention

The second aim of the research is to implement and evaluate a pilot programme that seeks to improve the efficacy of formative assessment practices through the use of the newly established language-specific reading benchmarks. To this end, our research questions are:

1. What is the effect of providing teacher training on reading benchmarks on the ability of teachers to know the learning levels of their learners?

2. Does easy-to-administer assessment address the cognitive burden that teachers face in conducting and interpreting assessment?

The next section describes the intervention in detail.

5. Intervention details

Figure 1 shows the theory of change of a typical programme using targeted and differentiated instruction to improve reading outcomes, drawing explicit attention to the links in the causal chain that focus on teachers' knowledge of their learners' abilities (highlighted in green). In order to achieve the desired outcome of improved learner performance, teachers need to know how to address gaps, but they cannot begin to do so if they don't know what those gaps are. While not sufficient alone, teachers' ability to accurately judge their learners' level is a necessary link in the causal chain. The intervention focuses on strengthening that link.



The intervention discussed in this report aims to improve the efficacy of formative assessment through:

- Introducing teachers to the newly established early grade reading benchmarks for the foundation phase LoLT in their school.
- Facilitating the meaningful interpretation of the EGRA results by linking them to the benchmarks. We do this by showing teachers how to classify each learner's reading proficiency according to the benchmark categories using an EGRA-like assessment.
- Providing teachers with high level guidance on the appropriate use of the benchmark categories in the classroom over and above the assessment component
- Simplifying the administration of EGRA, reducing the cognitive burden and the time taken to conduct the assessment with each learner. The assessment is trimmed into two subtasks for the intervention, namely Letter Sound recognition and Oral Reading Fluency (ORF) and we remove the need for additional calculations to determine fluency. We also

provide classroom charts that give teachers a readily accessible visual summary of individual learner reading proficiency.

Through this intervention, teachers could conduct more frequent assessments and enhance their interpretation of assessment outcomes. This would increase opportunities for teachers to interact with their students' reading levels, providing a clear measure for defining reading proficiency. Consequently, teachers will enhance their understanding of learners' reading abilities. This improved awareness will enable teachers to address learning gaps through tailored instruction, potentially necessitating extra training or support for effective implementation, as indicated in the theory of change.

The components of the intervention are discussed in the section below.

5.1. Early Grade Reading Benchmarks

The newly established Early Grade Reading Benchmarks for the African languages used in the study are shown in Table 2 below.

Language Group	Grade 1	Grade 2	Grade 3
	By the end of the year, le	earners should be able to:	
Nguni	Sound out <u>40 letters</u> correctly in one minute	Correctly read <u>20</u> words in a passage	Correctly read <u>35</u> words in a passage
Sesotho-Setswana		Correctly read <u>40</u> <u>words</u> in a passage	Correctly read <u>60</u> words in a passage

 Table 2: Early Grade Reading Benchmarks for Nguni and Sesotho-Setswana Languages

The benchmarks were introduced to intervention teachers using color-coded categories outlined in Table 3 and Table 4. These categories divide the benchmarks into five proficiency levels: nonreader, struggling reader, emerging reader, proficient reader, and, in Grade 2, fluent reader. This categorization is designed since the benchmarks represent end-of-year standards, thus not suitable for tracking in-year progress. The categories help teachers monitor whether their students are on track to achieve the end-of-year target as the terms progress, while also indicating how many words they need to acquire to meet the benchmark. This naming system effectively describes students' progress relative to expectations. The proficient reader category represents the minimum benchmark for year-end achievement, and in Grade 2, the 'fluent reader' category denotes a student's reading proficiency at the level of the next grade.

Category	Grade 2 and 3
Non-Reader	0
Struggling Reader	1-25
Emerging Reader	26-39
Proficient Reader	40 and above

Table 3: Letter Sound Knowledge Benchmark Categories

Table 4: Oral Reading Fluency Benchmark Categories

Category	Grade 2	Grade 3		Grade 3		Category	Grade 2	Grade 3
Non-Reader	0	0		Non-Reader	0	0		
Struggling Reader	1-9	1-19		Struggling Reader	1-19	1-39		
Emerging Reader	10-19	20-34		Emerging Reader	20-39	40-59		
Proficient Reader	20-34	35 and above		Proficient Reader	40-59	60 and above		
Fluent Reader	35 and above			Fluent Reader	60 and above			

5.2. Adapted EGRA Tool

For the intervention, the adapted EGRA tool focused on two skills aligned with the benchmarks: Letter Sound Knowledge and Oral Reading Fluency. This reduced assessment time per learner from 8-10 minutes to 5 minutes, making it feasible to assess a class of 45 students in less than 4 hours. The teacher assessment charts were color-coded to match the benchmark levels in Table 3 and Table 4. Teachers received a handbook explaining the colour codes, along with two identical Letter Sound Recognition and ORF assessment charts. The teacher version of the assessment chart was colour coded denoting the reading categories as seen in Figure 2 and Figure 3 while the version for use by learners was not colour coded. Teachers also got a wipeable learner progress chart (Figure 4) with matching color-coded stickers to track progress.

During the assessments, learners are instructed to sound out letters or read words on the charts for 60 seconds per task. The teacher times the learner and marks any incorrect items. After 60 seconds, the teacher marks the final attempted item and adjusts backward for each error, determining the learner's benchmark category based on the colour associated with the final position on the chart.

Chart 1 – LETTER SOUND RECOGNITION									
Izibonelo:	b	К	f						
m	T	н	g	S	А	Z	W	Р	е
L	k	т	D	a	d	с	0	n	ĩ
В	Μ	U	j	к	u	G	q	Е	f
м	Х	S	Ν	b	Y	v	Q	r	t
D	u	А	t	Р	о	h	е	а	t
у	н	В	F	U	J	V	n	С	R

Figure 2: Letter Sound Knowledge Chart

Figure 3: isiZulu ORF Passage Teacher Chart

Chart 2 – PASSAGE READING

UJabu unenja encane.	3
Ngelinye ilanga uJabu nenja yakhe	8
baya kuyodlala enkandla.	11
Inja yabona unogwaja yase yawujaha.	16
Inja yalahleka.	18
<mark>UJabu</mark> wayimemeza kodwa yangabuya.	22
UJabu wakhala wase ephindela ekhaya.	27
Kodwa kwathi ngaphambi kokuhlwa inja yabuya.	33
<mark>UJabu</mark> wajabula ukubona umngani wakhe.	38

In this intervention, the adapted EGRA assessment has been simplified. Instead of calculating precise fluency in correct letters or words per minute, learners are now assigned to benchmark categories. Teachers no longer need to count attempted and incorrect items or perform calculations. In cases where learners finish the passage before 60 seconds, traditional EGRA methods would involve adjusting the number of correct items based on remaining time.

Another enhancement is the learner progress chart (Figure 4). Following each assessment, teachers place a colour sticker next to the learner's name and task on the chart displayed in the classroom, replacing the need to record numbers in a booklet. This enables quick visual tracking of each learner's progress. By the end of each assessment period, teachers can use the chart's information to tailor instruction to suit individual needs. Over time, the chart reflects learners' progress towards meeting end-of-year benchmarks.

ZENEX					(5		bas Depa Bast RCP4	sic education short statuten usus of south AFRICA		0	J -	PA												
Class Teacher:												~	_	_			1								
Grade: Names of Learners	Table Sounds	Faile in the	Latter Sounds	Parage for day	Teller Sounds	Parage for day	Latter Scurren	Passage man deal	Names of Learners	(Facult for first 2	Parage Reading	Tarter Sounds	Parage for day	Test County and an an and an									
1		-	-	-	-	+	-	-	26		>	-		-											
2			-	-	-	-	-	-	27		_			-			$ \rangle$								
3			-	-	-	+	-		28			-	+	+		\vdash		٩							
4		-	-	+	-	+	+	-	29		_	-	+	-	-	\vdash		Assess	ment 1	Assess	ment 2	Assess	ment 3	Assess	ment 4
5		-	-	+	+	+	⊢	+	30			+	+	+	-	\vdash	-	Begi	n June	Begin	August	End Sep	tember	End O	ctober
6 7		-	-	-	+	+	+	-	31		_	+	+	+	-				2		2		p p		<u>به</u>
/			-	+	+	+	+	+	32		-	+	+	+	-	\vdash		pu	adir	1 a	l iji	- pu	adir	p	adir
o o	-		+	+	+	+	⊢	+	33			+	+	+	+	\vdash	-	Sou	Se .	Se	8e	Sou	8e	Sou	Se.
10	-		H	+	+	+	⊢	+	36		-	+	+	+	+	\vdash		ter	age	E.	88	ter	36	ter	386
n			-	+	+	+	⊢	-	36		-	+	+	+	-			Let	ass	Let	ass	Lett	ass	Let	ass
12	-			+	+	+	t	+	37			+	+	+	-	\vdash	1		•		•		•		•
13				+	+	+	t	+	38		-	+	+		\vdash	\vdash	1								
14				+	\square	\mathbf{T}	t	+	39			+		+	\vdash		1								
15							t		40			+					1								
16						\top	F		41								1								
17							F		42								1								
18							Γ		43								1								
19							T		44								1								
20									45																
21									46																
22									47																
23									48																
24									49																
25									50																

Figure 4: Class Progress Chart

5.3. Frequency of assessments and scheduling options

Teachers were instructed to assess students four times between June and October 2023 on Letter Sound Knowledge and Oral Reading Fluency using the provided intervention materials. The subsections below indicate the possible times outlined in the national curriculum, often referred to as the Curriculum and Assessment Policy Statement (CAPS), where teachers were advised to administer the assessment. Teachers used the same charts for Letter Sound Knowledge and Oral Reading Fluency for all four assessments. Teachers were advised not to use the charts for any other purpose to prevent students from memorizing them, which could distort their true reading ability and instead reflect their memorization skills.

5.3.1. Home language Lesson (Group Guided Reading)

According to CAPS instructional time for Grade 2 and 3, a maximum of 8 hours and a minimum of 7 hours are allocated for home language per week.

Table 5 shows suggestions on how teachers can break down their minimum time into different components. In the minimum 7 hours per week of Home Language instruction, 30 minutes per day (2.5 hours a week) is set aside for Group Guided Reading (GGR). To implement our assessment, teachers could use two consecutive weeks of GGR to assess learners.

Table 5: CAPS Instructional Time by Grade and Activity

	GRADE 2		GRADE 3			
HOME	Total per week		Total per week			
LANGUAGE						
Listening & speaking	15 minutes per day for 3 days	45 minutes	15 minutes per day for 3 days	45 minutes		
Reading & phonics	Phonics : 15 minutes per day for 5 days (1 hour)	4 hours 30 minutes	Phonics : 15 minutes per day for 4 days (1 hour)	4 hours 30 minutes		
	Shared Reading: 20 minutes per day for 3 days (1 hour)		Shared Reading: 20 minutes per day for 3 days (1 hour)			
	Group Reading: 30 minutes per day (2 groups each for 15 minutes for 5 days (2 hours 30 minutes)		Group Reading: 30 minutes per day (2 groups each for 15 minutes for 5 days (2 hours 30 minutes)			
Handwriting	15 minutes per day for 3 days	45 minutes	15 minutes per day for 3 days	45 minutes		
Writing	15 minutes per day for 4 days	1 hour	20 minutes per day for 3 days	1 hour		
	Total per week	7 hours	Total per week	7 hours		

5.4. Interpretation of assessment results

The approach to communicating benchmarks was deliberately simple. Instead of conveying multiple figures that represent different aspects of expected learner performance for various tasks and intervals, only one figure per grade was communicated. This number serves as the year-end target for all learners, as shown in Table 2. During periodic assessments, teachers assess each learner's progress in relation to the benchmark for that grade. While this communication resembles the EGRA approach, further breakdown into sub-categories (Table 3 & Table 4) aids in grouping learners, understanding each category's implications, and tracking transitions between categories throughout the year.

The intervention's second intermediary outcome is that teachers adjust their instruction based on students' current needs. Teachers were firstly encouraged to rely on their expertise to address skill gaps in learners. Additionally, they were directed to the National Framework for the Teaching

of Reading in African Languages in the Foundation Phase. This resource outlines the skills necessary for fluent reading and comprehension, gap identification, and remediation strategies. Another resource that was made known to the intervention teachers is an online teacher development platform¹ that offers courses to enhance pedagogical practices, including various aspects of reading instruction.

5.5. Teacher training

The full-day training program was comprised of two parts. The first part aimed to familiarize teachers with the study, while the second focused on implementing assessments. The teachers were explicitly informed about the research's purpose, which was to enhance their ability to measure their learners' reading levels and understand their learning needs. This knowledge would allow them to refine their teaching methods to better address their learners' requirements. Teachers were then introduced to benchmarks pertinent to their languages and grades. They received a Teacher Handbook as a resource detailing the intervention's purpose and rationale.

The second part of the training provided teachers with practical skills for assessing their students' reading levels in alignment with the benchmarks. Specifically, they learned how to assess Letter Sound Knowledge and Oral Reading Fluency. This training included paired and group practice among the teachers.

¹ www.tpd-dbe.org

6. Research design

The study combines quantitative longitudinal data at the teacher and learner level with qualitative data collected through in-depth interviews and materials review. Details of the sample, instruments and methodology are outlined below.

6.1. Sample

The sampling process involved a database of schools already involved in the DBE's EGRA program within the provinces of Mpumalanga, Eastern Cape, Limpopo, and North West. The goal was to enhance existing learner assessment practices in these schools rather than introducing early grade reading assessment for the first time. Districts were selected based on the number of schools and proximity to major cities within each province.

The desired sample comprised 60 schools, with 15 from each province. These were primary schools of Quintiles 1-3, each with at least three grade 3 and three grade 2 classes. After the initial selection, random sampling identified 10 schools in each province, and the remaining 5 schools served as reserves in case some of the initially chosen schools couldn't participate.

Upon contacting the schools, it was found that North West and Limpopo schools lacked the necessary Grade 2 and 3 classes. Consequently, the sampling was expanded to include schools not part of the EGRA project. Non-EGRA schools were randomly selected within each Language of Learning and Teaching (LoLT) category from the Education Management Information Systems (EMIS) database. Table 6 below shows the sample of schools by DBE EGRA training status.

Table 6: Sampled Schools by DBE EGRA training status

Province	EGRA	Non-EGRA
Eastern Cape	10	0
Mpumalanga	9	
North West	0	10
Limpopo	0	10

In each school, three teachers from both grades 2 and 3 were chosen for the study. For each selected teacher, 10 students were randomly sampled from their class using the lottery method. The intended sample size was 240 teachers and 2400 students.

For the qualitative part of the study, 4 schools in Eastern Cape and Limpopo provinces were selected based on the highest and lowest average EGRA scores at the baseline. In these schools, all grade 2 and 3 teachers were eligible for interviews.

6.2. Instruments

6.2.1. Literacy Outcomes

Reading outcomes were evaluated using specific tasks from an adapted Early Grade Reading Assessment (EGRA) tool. These tasks were designed for African languages by academics and education professionals and have been utilized in previous education impact evaluations. The EGRA's widespread usage and understanding make it advantageous for cross-program comparisons. To maintain consistency with the intervention and allow for comparability, the assessment included the Letter Sound Recognition and Oral Reading Fluency tasks. Additionally, comprehension questions tied to the Oral Reading Fluency task were included for informational purposes, although this task wasn't part of the intervention materials.

6.2.2. Teacher Interview

We conducted a short interview with the six randomly selected grade 2 and 3 teachers from each school to probe literacy teaching and assessment practices and attitudes. We collected basic demographic information and data on the teacher's qualifications and experience. Teachers were asked to rank the reading proficiency of the learners for whom we conducted EGRA assessments. Teachers were provided with levelled reading passages and asked to indicate what proportion of their class is reading independently at each level and to estimate how far each of the sampled learners read up to in one minute.

6.2.3. HOD/Principal Interview

Short interviews were conducted with either the Foundation Phase Head of the Department or the school principal depending on who was available to collect information around school resources.

6.2.4. Qualitative in-depth interviews

Teacher interviews were conducted with available teachers in the selected schools. Control teachers were asked questions on their assessment practice around the EGRA as well as their general teaching practice with a focus on differentiated instruction. In addition to questions on their assessment and teaching practice, treatment teachers were asked about their understanding and implementation of the intervention in their classrooms.

7. Data Collection

7.1. Fieldwork Team and Training

Quantitative fieldworkers were recruited for data collection, and their tasks included conducting data collection in assigned schools, administering Early Grade Reading Assessments (EGRAs) for Grade 2 and 3, and conducting interviews with educators about teaching and assessment practices. For the qualitative work, six fieldworkers were recruited with post-graduate qualifications and proficiency in isiXhosa or Sepedi, the Home Language spoken in the schools where they worked, as well as English. Their responsibilities included administering a qualitative classroom observation tool, photographing of completed learner charts, and interviewing teachers.

Fieldworker training consisted of three days for baseline data collection and an additional three days for endline data collection. The research team responsible for the study led the training, held in Pretoria and East London. A separate training for qualitative data collection was conducted over three days by an experienced senior qualitative researcher. This training covered various aspects including questionnaire administration, recording open-ended questions and interviews, utilizing recording and transcribing software, expectations, as well as paired practice and simulations.

7.2. Data Collection activities

Data collection followed the timeline shown in Table 7. Baseline data was gathered from mid-May 2022 to early June 2022 across provinces. Teacher Intervention training was then conducted immediately after the baseline data collection activities were concluded which effectively gave teachers 6 months to implement and use the intervention. After 5 months of intervention, qualitative data was collected to compare pedagogical approaches between treatment and control teachers. Ultimately, endline quantitative data was collected from late October to early November 2022.

Survey Type	Instruments	Dates
Baseline (Quantitative)	Teacher Questionnaire; EGRA Learner Assessment; HoD/Principal Questionnaire	May/June 2022
Midline (Qualitative)	Semi-structured Teacher Interview; Classroom Observations	September/October 2022
Endline (Quantitative)	Teacher Questionnaire; EGRA Learner Assessment; HoD/Principal Questionnaire	October/November 2022

Table 7: Timeline of Data Collection Activities

7.3. Realized samples

7.3.1. Teacher attendance at training

A total of 228 teachers were interviewed during the baseline. After baseline data collection, random selection of teachers within each grade and school was done for training invitations. However, not all invited teachers attended due to some being replaced by non-invited teachers, often at the discretion of school principals. About 86% of invited teachers attended the training, with the lowest attendance in Mpumalanga province. Interestingly, 16% of non-invited teachers in

the sample still attended the training. Out of 40 schools, one school did not send any teachers to the training. Table 8 combines attendance register data from training with the invitation lists to show the number and proportion of those invited who actually intended. In the analysis that follow, treatment refers to training attendance, whether invited or not.

rable 0. fraining invitation and attendance

	Invited to training	Not invited
Province	% Attending Training	% Attending Training
Eastern Cape	87%	22%
Limpopo	85%	16%
Mpumalanga	83%	0%
North-West	88%	25%
Total	86%	16%

7.3.2. Teacher and Learner Sample

The analytical sample was restricted to the learners of the 202 teachers who were observed at baseline and endline.

shows the realized sample of teachers and learners. The sample of teachers is comprised of teachers that were interviewed at baseline and endline, categorized by province. Due to teacher absence during fieldwork, 202 out of 228 baseline-interviewed teachers were re-interviewed at endline. Attrition rates are similar for treatment and control (12% vs 10%). The analysis focuses on the 202 teachers observed at both points, with five teachers from the non-training school excluded for treatment-control balance. The analysis involves 197 teachers, examining baseline characteristic balance between treatment groups.

During the baseline fieldwork, various issues arose that led to deviations from the original sampling plan. Consequently, some cases had less than 10 sampled learners per teacher, making the baseline sample size 1,998. At the endline, the same learners were re-evaluated, with

replacements for unavailable ones. The analytical sample was restricted to the learners of the 202 teachers who were observed at baseline and endline.

Province	Teachers	Learners
Eastern Cape	49	451
Mpumalanga	52	472
North West	48	512
Limpopo	53	472
Total	202	1907

Table 9: Realised Learner and Teacher Sample by province

7.4. Sample Characteristics

Figure 5 presents the average traits of the 197 teachers interviewed both at baseline and endline, categorized for the full sample, as well as treatment and control groups. The average teacher has 17 years of experience, 14 of which are in foundation phase teaching. About 34% hold a bachelor's degree, and 66% are formally trained for foundation phase teaching. Class size varies from 27 to 70, with a mean of 46. No statistically significant differences were found between intervention (treatment) and business-as-usual (control) teachers. This indicates balanced characteristics, even with attrition, at baseline.





8.1. Reading assessment practices and efficacy

This section draws on baseline teacher interviews to generate insights into formative assessment practices in relation to learning to read with a particular focus on the use of EGRA. Thereafter, we consider the efficacy of these assessment practices by examining the correspondence between teacher ratings of learners' ability with results from EGRAs independently administered by our field teams.

8.1.1. Assessment practice

Although all 10 schools in the Eastern Cape and 9 out of 10 schools in Mpumalanga were included in the DBE EGRA initiative, only 67% and 38% of teachers in these provinces, respectively,

confirmed their exposure to EGRA through training or assessment (Table 10). Surprisingly, 54% and 70% of teachers in North West and Limpopo, respectively, reported exposure to EGRA despite their schools not being part of the DBE EGRA rollout.

Variable	North West	Limpopo	Mpumalanga	Eastern Cape	Overall
EGRA Trained	39%	65%	36%	54%	48%
EGRA experienced – No Training	15%	5%	2%	13%	9%
No EGRA exposure	46%	30%	62%	33%	43%
No. of Teachers	59	54	58	54	225

Table 10: Teacher prior EGRA exposure by Province

Baseline data were collected in May which corresponds to roughly the middle of the second school term. According to the DBE EGRA programme, Grade 2 and 3 teachers are required to have completed at least one assessment by this point. Of the 128 teachers who have had exposure to EGRA, 74% had completed at least 1 assessment round by halfway through term 2 (Table 11).

Table 11: Proportion of teachers administering EGRA by May

Province	Completed EGRA by May
Eastern Cape	76%
North-West	69%
Mpumalanga	38%
Limpopo	74%
Overall	74%
No. of Teachers	128

Of the 95 teachers who completed at least one assessment at Baseline Figure 6 below shows the distribution of the average length of an EGRA with a single learner. Approximately 35% of teachers require up to 5 minutes per learner, while 33% allocate 5 to 10 minutes. The mean duration is 8.74 minutes, skewed by a few teachers who take up to 60 minutes per learner, making the median of 5 minutes a more representative summary. With an average class size of about 46,

using a median EGRA duration of 5 minutes per learner implies roughly 4 contact hours to assess the whole class.



Figure 6: Distribution of Average EGRA duration

However, the CAPS curriculum allocates 7 hours per week for Home Language instruction. To administer one EGRA round for Home Language, teachers would need to forfeit nearly an entire week of instruction time. Some teachers manage 60 to 70 students, taking 5 to 6 hours to assess the entire class. Factoring in transition time between learners, these teachers must use instructional time allocated to other subjects or spend over a week for assessment.

Mitigating "lost" instructional time involves effective classroom management during assessments, such as assigning individual tasks while assessing a single learner. Strategies like these, outlined in Table 13, can sustain learning time even without whole-group instruction. Most teachers (62%) report engage learners with DBE workbooks or worksheets. Additionally, 20% report setting up reading corners, and 13% have students complete writing tasks. Although most teachers maintain learner engagement, 5% do not, resulting in lost structured learning time during a home language assessment week.

Table 12: Teacher EGRA Practices

Variable	Mean
Average EGRA Time per Learner (Minutes)	8,74
Classroom management during assessment:	
Workbooks/worksheets	62%
Reading	20%
Free time	5%
Writing	13%
Avenues for easier administration:	
Decrease Time Burden	33%
More Training	23%
More Resources	21%
Other (None of the above)	35%
No. of Teachers	95

Teachers were asked about ways to simplify EGRA administration and their utilization of assessment results. Multiple responses were allowed. Regarding EGRA administration, 33% of teachers suggested reducing assessment time as it currently takes a week of home language instruction to conduct one assessment round. Given that Grade 2 and 3 teachers must perform a minimum of 3 assessment rounds each year, this time constraint, combined with curriculum pressures and struggling students, contributes to 26% of exposed teachers not yet conducting an assessment by halfway through the second term.

Additionally, 23% and 21% of teachers respectively indicated that increased training and resources would aid EGRA administration. Notably, this response comes from teachers who have been exposed to EGRA and completed an assessment by mid-term. Another 35% had alternative views for simplifying EGRA administration, as listed under "Other" in Table 12. Many of these teachers suggested extending reading time per learner and using simpler assessment passages.

These responses may reflect a lack of understanding of the EGRA tool and its purpose, possibly highlighting training issues.

Concerning EGRA use, a significant portion of teachers employ it for identifying students needing remediation or reinforcement (Table 13). Among the 13% not choosing the provided options, several mentioned using the results to monitor learner progress.

Table 13: Teacher usage of EGRA Results

Variable	Mean
Reports to DBE	13%
Create Group Guided Reading Groups	23%
Identify Learners for Remediation/Consolidation	44%
Differentiate Instruction	11%
Set Lesson Plans	17%
Classroom Management (exec GGR)	11%
Assign Materials to Learners	13%
Report to School Management	9%
Other (None of the above)	3%
Num. of Teachers	95

In summary, we find that many teachers have no exposure to EGRA, and this does not align with DBE roll out. While it is possible that the teachers are new, it is more likely that it reflects absenteeism or that training was too light touch. This is supported by the number of teachers reporting the need for further training.

8.1.2. Assessment efficacy

This section examines teachers' assessment practices and their knowledge of learners' reading proficiency levels, comparing their estimates to the results from the EGRA assessments

administered by the field teams at baseline. Teachers were asked to estimate how far each of the 10 assessed learners could read within a minute using the same passage as in the learner assessment. Figure 7 illustrates the comparison in Nguni LoLT schools between teachers' estimates and the measured correct words read per minute for each learner. The corresponding plot for Sesotho-Setswana LoLT schools paints a similar picture. Points on the equality line indicate accurate teacher estimation. However, most points lie above this line, showing teachers generally overestimated their learners' performance.

Around 71% of learners, whose teachers had conducted an EGRA by May or had prior EGRA experience, and 68% of learners whose teachers lacked prior EGRA experience, had their performance estimated. Three teacher groups are identified in the figures: those who administered EGRA in that year (light blue), those who had EGRA exposure but didn't administer it that year (grey), and those with no EGRA exposure (navy).

Interestingly, even teachers with EGRA training tend to overestimate reading proficiency across the achievement spectrum. This overestimation is consistent, and teachers struggle to accurately estimate reading proficiency at various achievement levels.



Figure 7: Teacher Estimate vs Learner Performance at Baseline

Table 14 provides summary measures for the association between teachers' estimates and their learners' performance. Instead of assuming a specific distribution, non-parametric rank correlation measures are used: Spearman's correlation coefficient and Kendall's rank correlation coefficient. The expectation is that teachers who understand their learners' learning levels will exhibit higher correlations compared to those with less understanding. The average Spearman's correlation coefficient is 0.64, and the average Kendall-Tau correlation coefficient is 0.55, indicating a moderate positive association for all three teacher groups. Even though teachers with no EGRA exposure have the lowest correlation coefficients, the difference is minimal.

Table 14: Correlation between teacher ranking and EGRA ranking

	EGRA by May	EGRA exposed	No exposure	Overall
Spearman	0.65	0.65	0.62	0.64
Kendall-Tau	0.56	0.56	0.54	0.55

Learners were categorized into benchmark levels based on Grade 2 thresholds. A comparison is made with teachers' estimated benchmark levels derived from their estimates of each learner's correct words per minute (CWPM), as illustrated in Figure 8. For learners classified as Non-Readers, only 18% of teachers correctly classify them, while approximately 66% classify them as Struggling Readers, 8% as Emerging Readers, 3% as Proficient, and 5% as Fluent. Among learners classified as Proficient, 1% are classified as Non-Readers, 8% as Struggling, 12% as Emerging, and 57% as Fluent by teachers.

The Proficient benchmark ranges from 40 to 58 CWPM for Sotho-Setswana Languages and 20 to 34 for Nguni. Even within this wide range, teachers could correctly classify Proficient Readers. However, only 22% of learners were correctly classified, suggesting that teachers frequently misjudge their learners' levels, even with a considerable margin of error provided by benchmark thresholds. In summary, teachers not only overestimate their learners' performance, but the extent of misestimating is substantial.



Figure 8: Teacher Estimates of learner benchmark category

For each teacher, the proportion of their 10 assessed learners accurately classified based on Grade 2 reading benchmark thresholds is computed. The distribution of these proportions is depicted in Figure 9. On average, teachers correctly classified 37% of their learners, equating to 4 out of 10 sampled learners. This aligns with existing literature, indicating that in settings with large and diverse classes emphasizing communal teaching methods, personalized assessment is limited and teacher knowledge of individual learners' reading proficiency is deficient.

Although one might anticipate that teachers, after exposure to and implementation of EGRA for learner assessment, would improve in estimating their students' reading levels, our analysis doesn't show significant differences in accuracy between teachers with EGRA exposure and those without. This implies that prior to the intervention introduction, the process of teachers grasping the assessment tool, using it to gauge learner performance, and gaining understanding of their students' levels was ineffective.





8.2. Evaluation of pilot intervention

We begin with an examination of intervention compliance and fidelity and identify two groups of treatment teachers, namely those who appear to have implemented the programme (high use) and those with little or no evidence of take up (low use). We then draw on the quantitative and qualitative teacher interview data to describe teachers' experience of the intervention materials, assessment practices, knowledge of learners' reading levels and home language pedagogy. We examine whether there are differences between the three groups of high use, low use and control teachers. Finally, we estimate the impact of the intervention on our main outcome, teachers' knowledge of their learners' reading proficiency.

8.2.1. Intervention compliance and fidelity

The evaluation of development programs often considers program compliance, which pertains to participants being granted access to an intervention but possibly not implementing or using it. In this study, participants, particularly teachers, might not only fail to use the intervention, but they might also misuse it or use it inadequately. Prior research by Brodie et al. (2002) highlighted how teachers in under-resourced schools can adopt program aspects superficially without grasping their essence.

Table 15 illustrates the intervention's uptake rates. Out of 127 teachers receiving the intervention, 86 confirmed attending training. However, 29 teachers claimed non-attendance despite contrary training register records, while 10 reported attending unrelated assessment training. Those affirming training attendance, 75 teachers indicated using intervention materials, and four teachers avoided the question. Among the seven who didn't use the materials, some cited non-receipt or incompleteness despite universal distribution.

Among users, 68% followed the desired frequency for progress chart use, involving 3 to 4 assessment rounds.

		Frequency	Proportion		
Training attendance					
	Confirmed attendance	86	44%		
	Attended other training	37	19%		
	Attended no training	74	38%		
	Total	197			
Use of materials for those confirming attendance					
	Yes	75	87%		
	No	7	8%		
	Refused	4	5%		
	Total	86			
Progress chart use rating for those using materials					
	No use	4	5%		
	Low use	20	27%		
	High use	21	68%		
	Total	75			

Table 15: Take-Up Rates of Intervention

Note: Sample excludes the school where no teachers attended training

In the analyses that follow, we distinguish between treatment teachers for whom we have evidence of reasonable uptake (high use) and the other treatment teachers. The sample sizes and proportions are shown in Table 16.

Table 16: Treatment teacher categories

	Frequency	Proportion
Treatment – high use	51	26%
Treatment – other	80	41%
Control	66	33%
Total	197	

Note: Sample excludes the school where no teachers attended training

Treatment teachers were asked to rate the difficulty of administration and interpretation of the intervention assessment relative to the traditional EGRA (Figure 10).

Figure 10: Teacher rating of intervention relative to EGRA



Administration Difficulty

8.2.2. Teacher experience of the intervention

Qualitative interviews provided valuable insights into how teachers perceive and engage with the intervention. Ten treatment teachers were interviewed, but one Eastern Cape teacher admitted not implementing the intervention due to busyness, favoring the National Education Collaboration Trust (NECT) program instead. She also mentioned not receiving all intervention materials. Among the nine other teachers, all acknowledged intervention implementation, though none completed all six prescribed activities.

Interviewers took photos of learner progress charts and noted their display and usage levels. Figure 11 depicts the chart for one of the teachers interviewed (Teacher T2).



Figure 11: Picture of Used Progress Chart

8.2.3. Understanding of the intervention

Teachers' implementation of the intervention largely indicates a misunderstanding of its role as a formative assessment tool to enhance teaching practices. Instead, they focused on utilizing the letter sound chart and passage chart by directly teaching them to students. Some teachers even distributed copies of the charts for home practice. For instance, one of the teachers (LS1IG3T6) pasted the letter sound chart to student desks (Figure 12). Another teacher proposed sharing audio recordings of the letter sounds for classroom practice. During the interview one of the teachers said *"I made a copy of the booklet they gave us on one page. Then I gave them to put them in the file. I told them I'm not going to tell you the day, but I will call you, you are allowed to read this thing. This paper at home. I will call you one day. You will come to me. Then you will read. I will give you the sticker."* Some teachers solely used the intervention materials for instruction and practice, while others combined teaching with individual assessment.

Figure 12: Picture of assessment letter sound chart given to learners



8.2.4. Key challenges with implementation

The main obstacle cited for finding the intervention difficult to implement was due to time. Time was mentioned in three contexts. First, teachers spoke about workload and competing priorities. Below are some of the responses from the qualitative interviews with the teachers.

"... lessons that needs to take place, there are meetings that we need to go to. So, time and the amount of the workload that is there in a day." (T1).

"These learners are not always ahead, so I told myself that if I focus on this program, because in the end, I also have goals in my class. So, this program would have disturbed me in my goals."

"….it was a challenge for us to or maybe let me say for me to be able to do that the assessment with all the learners in the in the given timeframe" (T1).

In large, overcrowded classes, an additional challenge arises in identifying suitable spaces for learner assessment. Teachers encountered issues conducting the intervention within the classroom due to the presence of other students, who could overhear the assessment process. This situation led to concerns like memorization of passages, as stated by Teacher T7 who noted, *"By the time we get to the seventh learner, they already memorized the passage."* Similarly, Teacher T9 highlighted that struggling readers often memorize passages already read, hindering accurate assessments.

Some teachers responded by adapting the intervention, incorporating alternative resources. For instance, one teacher favoured the use of DBE workbooks due to their story variety, which offers a more diverse approach.

The main feedback from teachers on how to improve the intervention surround resources. One teacher made a plea for more reading passages, "That is why I said, let maybe next time let it not be only one paragraph or one paragraph for a quarter, one paragraph for a quarter, one paragraph for the third quarter and the second paragraph, one paragraph for the second time and that so that they can be more improvement and so that learners can be good, good readers.

8.2.5. Indications of some improvements

Despite all the challenges that teachers faced with conducting the assessments, majority of the teachers reported positive changes as a result of the intervention. A couple of the teacher's responses can be seen below.

"..... I realised that some of the learners are unable to complete the letter sounds" (T2)

"It has changed in this way; since I've realized which one are able to read even if it's a little bit" (T3)

"...learners now know that they know how to read..." (T4)

"There is a change because now it is no longer difficult to assess learners; it has become easier now" (T8).

"It makes it easier now because you begin the assessment knowing where the learner is, their abilities" (T8).

Most of the teachers interviewed reported that there were changes in the way they taught since implementing the intervention. The main strategy for this was grouping the learners according to their reading abilities.

8.2.6. Assessment efficacy

We go back to examining the accuracy of teacher evaluations of learners reading proficiency, now comparing different groups based on treatment status. For intervention-trained teachers, they are divided into those using materials as intended and those with low use. The examination also focuses on changes between baseline and endline.

In Figure 13, teacher estimates of learners' correct words per minute read (CWPM) at endline are plotted against the CWPM measured by EGRA. The findings reveal notable disparities between teacher estimates and actual ORF, indicating a clear tendency for teachers to overestimate performance. Overall, teacher estimates are higher than measured values for 75% of learners. Figure 13 represents the Nguni languages as with Figure 7 however the Sesotho-Setswana language group paints a similar picture.

The figure distinguishes among control teachers, low use treatment teachers, and high use treatment teachers. However, apparent differences between these three groups are not immediately evident.



Figure 13: CWPM - Teacher Estimate vs Learner Score at Endline

Figure 14 plots the average spearman correlation coefficient by treatment status. At baseline, control teachers displayed a strong positive association with an average correlation around 0.7. By endline, treatment teachers who effectively used the intervention achieved the highest average correlation of approximately 0.7, experiencing the most substantial improvement in correlation. In contrast, control teachers experienced a decline to a moderate positive association, while treatment low use teachers maintained a relatively stable correlation level.





8.2.7. Estimated impact of pilot intervention

Although the treatment and control groups were found to be very similar on baseline characteristics, unobserved systematic differences between the groups could introduce selection bias into our estimates of the impact of the intervention. To address this and the potential teachers becoming more familiar with their students' learning levels due to increased interaction over time, a panel difference-in-difference with fixed effects model was used to estimate the intervention's effect. This approach compares the average change in outcomes of treatment teachers with control teachers, separating the impact of the intervention from the "getting to know your learners" effect. The results can be seen in Table 17.

Table 17: Difference-in-Difference Coefficients

	(1)	(2)	(3)
Variables	Percent Correct	Spearman	Kendall
Endline	0.0258	-0.0507	-0.0600
	(0.0314)	(0.0444)	(0.0389)
Treatment x Endline	0.0360	0.0969*	0.0995**
	(0.0384)	(0.0542)	(0.0475)
Constant	0.343***	0.629***	0.546***
	(0.0131)	(0.0185)	(0.0162)
Observations	358	355	355
R-squared	0.051	0.022	0.028
Number of Teachers	197	197	197

Note: Models include individual fixed effects. *, **, *** indicate significance at the 10-percent, 5-percent and 1-percent significance level respectively. Standard errors in parentheses

In terms of the "percent correctly classified" measure, none of the coefficients are statistically significant. When using the Kendall rank correlation coefficient, the correlation between control teacher estimates and learner performance decreases by 0.06 points between baseline and endline, although this change is not statistically significant. In contrast, treatment teachers exhibit an improvement of 0.1 points in their correlation, which is statistically significant at the 5% level.

9. Conclusion

In this report, the findings from the Teacher Benchmark Usage pilot study are presented. Data was collected from Quintile 1-3 schools in urban areas, involving learner assessments, teacher interviews, and head of department (HoD) interviews. The schools were chosen based on certain criteria, and the study focused on Grade 2 and Grade 3 classes. We included both schools that received the DBE EGRA training and those that didn't. The intervention aimed to improve teacher

knowledge of learners' reading abilities through a simplified reading benchmark-aligned assessment tool.

Key findings of the study include the following:

- **Study Design and Groups**: Teachers who attended training and received the intervention materials (treatment group) were compared with other teachers in the same grade at the same school (control group). The characteristics of treatment and control teachers were statistically similar, indicating that the groups were comparable before the intervention.
- Intervention Benefits: The report suggests that continuing to provide simplified tools to teachers at scale is beneficial. The intervention, which focused on letter sound recognition and Oral Reading Fluency assessment, was found to be easier to administer than EGRA by 95% of teachers in the high-use category. This type of intervention could be a valuable addition to the official DBE EGRA toolkit.
- Impact on Teacher Knowledge: The intervention led to an increase in teacher knowledge of learner levels of 0.1 correlation points. This increase in knowledge was significant relative to control teachers.
- Gaps in Teacher Understanding of Formative Assessment: The report highlights important lessons about teacher knowledge of diagnostic and individualized assessment. Teachers often lack an understanding of the purpose and use of individualized assessments such as EGRA. The report emphasizes the need to account for these gaps as individualized assessments are designed and implemented.
- Challenges with Assessment Practice: The study identifies challenges with assessment practice, including the lack of adequately graded resources for teaching and assessment. Teachers' use of assessments for practice purposes and the need for secure assessment texts are discussed.
- **Resource Constraints and Classroom Size**: Resource constraints and large class sizes pose challenges to individualized assessment. The time required for assessment in large classes can be substantial and needs to be factored into teaching planning.

The report suggests the need for resources that support learners' practice and secure assessment texts. Multiple assessments might be necessary to maintain assessment integrity. The study also addresses the importance of providing teachers with effective teaching strategies for areas like letter sounds.

In conclusion, the report emphasizes the value of simplified assessment tools for teachers and highlights the need to address pedagogical gaps in implementing individualized assessments. The study's findings provide insights into teacher knowledge, assessment practices, and challenges within the context of the intervention, contributing to the broader understanding of effective teaching and learning strategies.

References

- Ardington, C. & Meiring, T., (2020) "Impact Evaluation of Funda Wande Coaching Intervention Midline Findings" SALDRU, University of Cape Town, Cape Town.
- Ardington, C., Wills, G., Pretorius, E., Mohohlwane, N. & Menendez, A. 2021. Benchmarking oral reading fluency in the early grades in Nguni languages. *International Journal of Educational Development*, *84*, p.102433. Banerjee, A.V., Cole, S., Duflo, E. and Linden, L., 2007. Remedying education: Evidence from two randomized experiments in India. *The Quarterly Journal of Economics*, *122*(3), pp.1235-1264.
- Bassi, M., Meghir, C. & Reynoso, A., 2020. Education quality and teaching practices. *The Economic Journal*, *130*(631), pp.1937-1965.
- Brodie, K., Lelliott, A. & Davis, H. 2002. Forms and substance in learner-centered teaching: Teachers' takeup from an in-service programme in South Africa. *Teaching and teacher Education*, *18*(5), pp.541-559.
- Djaker, S., Ganimian, A.J. & Sabarwal, S. 2022. Primary-and middle-school teachers in South Asia overestimate the performance of their students. [Unpublished Manuscript]
- Hoadley, U. 2018. Pedagogy in poverty: Lessons from twenty years of curriculum reform in South Africa. Abingdon/New York: Routledge.
- Howie, S.J., Combrinck, C., Tshele, M., Roux, K., McLeod Palane, N. & Mokoena, G. 2017. *PIRLS 2016: South African highlights report.* Centre for Evaluation and Assessment (CEA).
- Kanjee, A., 2020. Exploring primary school teachers' use of formative assessment across fee and no-fee schools. *South African Journal of Childhood Education*, *10*(1), pp.1-13.
- Kanjee, A. & Bhana, J., 2022. Lessons learned and evidence of impact: Formative assessment in an integrated reading and mathematics intervention. In *Early Grade Reading and Mathematics Interventions in South Africa.* N. Spaull & E. Pretorius, Eds. Cape Town: Oxford University Press Southern Africa.
- Menendez, A. & Ardington, C. 2018. Impact Evaluation of USAID/South Africa Story Powered School Program—Baseline. Technical Report.
- Mkhwanazi, H.N., Joubert, I., Phatudi, N.C. and Fraser, W.J., 2014. Teachers' Use of Formative Assessment for the Teaching of Reading Comprehension in Grade 3. *Mediterranean Journal of Social Sciences*, *5*(7), p.468.89

- Mohohlwane, N., Wills, G. & Ardington, C. 2022. A review of recent efforts to benchmark early grade reading skills in South African languages. In *Early Grade Reading in South Africa*. N. Spaull & E. Pretorius, Eds. Cape Town: Oxford University Press Southern Africa.
- Taylor, S., Cilliers, J., Prinsloo, C., Fleisch, B. & Reddy, V. 2017. The Early Grade Reading Study: Impact evaluation after two years of interventions. *EGRS Evaluation Report*.