



USAID
FROM THE AMERICAN PEOPLE



EVALUATION

Final Report: Impact Evaluation of the Western Cape Emergent Literacy Intervention in South Africa

June 2018

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.

FINAL REPORT: IMPACT EVALUATION OF THE WESTERN CAPE EMERGENT LITERACY INTERVENTION IN SOUTH AFRICA

June 2018

Contract Number: AID-674-I-12-00002, Order Number: AID-674-TO-15-00003

Cover Photo: Early Grade Learners, South Africa. Photo Credit: Nolte Lourens

DISCLAIMER

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

ABSTRACT

From June 2015 – February 2018, Khulisa Management Services (Pty) Ltd conducted an Impact Evaluation of the USAID-funded “Strengthening Teaching of Early Language and Literacy in Grade R” (Stellar) program as integrated into the Grade R Emergent Literacy (ELIT) intervention in the Western Cape Province of South Africa.

The major evaluation questions include:

1. Have Stellar interventions¹ increased the language and literacy achievement of targeted learners over time?
2. To what extent has the train-the-trainer model resulted in more effective and knowledgeable teachers and positive language and literacy outcomes for learners?
3. What is the likelihood of sustainability of organizational and beneficiary gains after Stellar interventions end?

Khulisa employed a quasi-experimental research design for the impact evaluation.

The evaluation found:

- The ELIT intervention, incorporating 8 Stellar story modules and an adapted Letters and Sounds Manual, initially increased the language and literacy achievement of learners over the first half of Grade R but this effect diminished by the end of Grade R and disappeared altogether by the end of Grade 1.
- The cascade training approach may have had an initial effect but, without provincially mandated support post-July 2016, the intervention advantage appears to have been lost.
- Organizational sustainability is beginning to be built but could be improved. For learners to benefit, quality instruction across multiple years is essential.

Keywords: emergent literacy; grade R; training-of-trainers; cluster meetings; block training; impact evaluation; literacy outcomes; sustainability; ongoing teacher support

¹ The evaluation questions were set in the evaluation contract prior to the official incorporation of Stellar into the ELIT intervention. This evaluation measures the effects of the combined Stellar/ELIT intervention.

ACKNOWLEDGEMENTS

Khulisa Management Services (Pty) Ltd would like to thank the United States Agency for International Development (USAID), the ELMA Foundation, and JP Morgan for supporting the development of this report through the funding mechanism which has enabled the evaluation to take place and through their ongoing engagement with the evaluation team. Khulisa specifically thanks USAID for their invaluable contribution and support in the design and delivery of the evaluation, and their input into the content and review of all reports.

Khulisa would also like to thank the Department of Basic Education for their input into the evaluation design and analysis of results, and the Western Cape Education Department for approving the research, readily facilitating access to the schools that are involved in the evaluation, facilitating access to Subject Advisors, tracking missing learners, and for their input and continuing endorsement. Khulisa also thanks Wordworks for their collaboration, support, and input into various evaluation processes.

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

AUTHORS

Dr Helene Aiello (Team Leader)
Katharine Tjasink (Evaluation Manager)
Dr Janeli Kotzé (Statistical Analysis Specialist)
Prof Elizabeth Pretorius (Early Grade Reading Expert)
Dr Louise Holman (Early Grade Assessment Developer)
Jennifer Bisgard (Senior Advisor)
Wade Harker (Project Researcher)
Wendy Magoronga (Analyst)
Nokuthula Mabhena (Data Visualization)

CONTACT DETAILS

Dr Helene Aiello

PO Box 923
Parklands, 2121
Telephone: 011-447-6464
Fax: 011-447-6468

Email: haiello@khulisa.com

Web Address: www.khulisa.com

EXECUTIVE SUMMARY

Khulisa Management Services (Pty) Ltd. (Khulisa) is pleased to submit, to the United States Agency for International Development (USAID), this **Final Report** for an **Impact Evaluation of the “Strengthening Teaching of Early Language and Literacy in Grade R” (Stellar) program as integrated into the Grade R Emergent Literacy (ELIT) intervention** in the Western Cape province of South Africa.

Project Purpose and Background

In July 2015, USAID commissioned Khulisa to conduct an Impact Evaluation of the Stellar program, designed by Wordworks² and implemented in two districts in the Western Cape province of South Africa. Stellar supports Grade R teachers in the development of the knowledge and skills needed to successfully teach oral language and emergent literacy in Grade R classrooms. At the same time, Wordworks, together with the Western Cape Education Department (WCED)³, co-developed the Western Cape ELIT intervention incorporating 8 Stellar story modules and an adapted Letters and Sounds Manual (referred to in this report as the “ELIT intervention” or “ELIT”).

The **purpose of the ELIT intervention** was to strengthen Home Language teaching with appropriate resources to improve language and literacy performance levels in the Foundation Phase (Grade R to 3).

The WCED and Wordworks used a Training-of-Trainers (TOT) cascade training approach to roll out the intervention. Wordworks trained the WCED Subject Advisors who then trained Lead Teachers, who later trained other teachers in the circuits or district-based geographic clusters. Training included teacher cluster meetings, feedback opportunities, peer learning, collaboration, and other opportunities to build on existing knowledge. Dry runs, practical sessions whereby Lead Teachers had an opportunity to practice their newly acquired knowledge and skills, were a key feature of the cascade training approach (Smith & Hugow, 2017).

Initial TOT with Subject Advisors and Lead Teachers began in 2015, and the ELIT intervention was rolled out simultaneously to all Grade R teachers and practitioners – approximately 3000 - at DBE registered sites in the Western Cape in 2016. Wordworks were responsible for training Subject Advisors, providing the Stellar program material and training resources, co-developing ELIT materials with the WCED, and providing support to Subject Advisors and Lead Teachers. The WCED was responsible for selecting, training and providing support to Lead Teachers, managing distribution of materials, training logistics, supporting Lead Teacher training of teachers, occasionally monitoring teacher progress, and providing support and feedback (Smith & Hugow, 2017). The WCED also ensured that the ELIT intervention was approved, communicated to all schools and districts, featured on the provincial General Education and Training (GET) calendar so that it could be implemented in 2016, and that WCED resources were made available as planned.

² Website: <https://www.wordworks.org.za>

³ Website: <https://www.westerncape.gov.za/dept/education>

The ELIT intervention included the provision of Stellar materials: eight existing Stellar Story Packs (classroom materials) and one adapted Letters and Sounds Manual⁴ (teacher manual), in three languages. The intervention also included two purpose-written training manuals for trainers and teachers⁵; the Department of Basic Education (DBE) Grade R Resource KIT; monthly two-hour cluster sessions with teachers; cluster program and facilitators' notes⁶; dry run and cluster session materials⁷; and a five-day block training workshop. The ELIT intervention Theory of Change (TOC) is that through the cascading model of TOT, teachers receive the intervention and this should positively affect the language and literacy abilities of their learners.

This evaluation constitutes an investment by the United States Government to evaluate the effects of the ELIT intervention over the period July 2015 (evaluation award) to February 2018 (final report). The **purpose of this impact evaluation** is to support the South African Government in its efforts to improve Grade R quality (in terms of language and literacy skills) and to raise early grade literacy scores.

Key Evaluation Questions

1. Have Stellar interventions increased the language and literacy achievement of targeted learners over time?
 - 1.1. To what extent have Grade R interventions impacted the language and literacy acquisition and performance of Grade R and Grade 1 learners in comparison to the control group and expected outcomes for these grades?
2. To what extent has the train-the-trainer model resulted in more effective and knowledgeable teachers and positive language and literacy outcomes for learners?
 - 2.1. To what extent has the integrity of the Stellar program been maintained through scale up?
 - 2.2. What factors enhance and/or constrain the scale-up of the Stellar training approach?
3. What is the likelihood of sustainability of organizational and beneficiary gains after Stellar interventions end?

Methods

Given the 2016 simultaneous rollout of the ELIT intervention to all Grade R teachers, Khulisa could not obtain a counterfactual in the same year. Instead, Khulisa proposed a **quasi-experimental research design** using the 2015 cohort of Grade R learners in the classes of teachers who had not yet received the training as the comparison group. The intervention group comprised the 2016 cohort of Grade R learners – one year later - in the classes of the same

⁴ Adapted during the lead up to ELIT.

⁵ One training manual for Subject Advisors and one (resource-focused) manual for Lead Teachers and teachers, both written by Wordworks in collaboration with the WCED.

⁶ For Lead Teachers.

⁷ A variety of materials to support Subject Advisors in monthly preparation workshops with Lead Teachers and to support Lead Teachers in monthly preparation of cluster workshop sessions with teachers.

teachers who had, by that time, received the training. The evaluation team used the same schools to obtain the comparison and intervention cohorts to reduce any bias that may have originated from unobservable school-level variables that remained constant over the evaluation.

The approach for this impact evaluation was **initially designed** using a **difference-in-difference estimator**. However, the signing of the evaluation contract was delayed so that by the time data collection got under way, it was too late to collect a clean baseline and the intervention group had already received the entire training intervention. The estimation approach was subsequently **restructured** to examine the **average score differences** between the groups at three points in time: in the middle of Grade R, after 6 months of training for the intervention group (**Wave 1**); at the end of Grade R (**Wave 2**); and at the end of Grade 1 (**Wave 3**).

Data was collected using several different **tools and instruments**: a Learner Assessment tool; Early Grade Reading Assessment; Classroom Environment Checklist Tool; Teacher Knowledge, Attitudes and Practices Tool; Grade 1 Pacing and Curriculum Coverage Tool; Sustainability Assessment Tool; Key Informant Interview Guidelines; and Cost Reporting Templates.

Key Findings

EQ 1: Have Stellar interventions increased the language and literacy achievement of targeted learners over time?



Assuming no *unobservable confounding factors* may have influenced the outcomes, the results suggest that ELIT interventions had an **initial positive effect** on the learners' language and literacy achievement. However, this effect **decreased with time** – the results show a statistically significant increase in language and literacy achievement in the first half of Grade R, which decreased by the end of Grade R, and disappeared by the end of Grade 1. This suggests that ELIT interventions, incorporating 8 Stellar story modules and an adapted Letters and Sounds Manual, could have given Grade R learners a temporary head start but this learning advantage was not sustained. Correspondingly, an assessment of a small sample of learner writing revealed that Grade 1 teachers did not pace significantly faster through the curriculum with the intervention group.

Sub-question 1.1: To what extent have Grade R interventions impacted the language and literacy acquisition and performance of Grade R and Grade 1 learners in comparison to the control group and expected outcomes for these grades?

Most learners in both the intervention and comparison groups reached the Grade R and Grade 1 criterion levels on the *listening and speaking based on a story*, and *listening and speaking in a conversation*, subtasks of the Learner Assessment. However, in both the comparison and intervention groups, most learners **did not reach the expected grade levels for *phonological awareness*, *phonics letters*, or *emergent reading in Grade R***. Performance on the emergent writing subtask differed between the intervention and comparison groups with intervention group learners performing better and reaching the grade level on most subtask items. By the end of Grade 1, many learners in both the intervention and comparison groups met the expected criterion levels for all these learning outcomes. On the **EGRA items** (Grade 1), learners in the intervention and comparison groups **only reached the WCED benchmarks for *letter sound recognition***. *Word recognition* benchmarks were only met in the Afrikaans comparison group and the isiXhosa intervention group. None of the language groups reached the expected levels on *paragraph reading or comprehension* based on a paragraph.

EQ 2: To what extent has the train-the-trainer model resulted in more effective and knowledgeable teachers and positive language and literacy outcomes for learners?



The cascade training approach appears to have had an **initial effect** on literacy outcomes for learners. The Teacher Knowledge Attitudes and Practices tool shows an **overall improvement** from pre- to post-training and no significant differences from the end of training to the end of the year. However, the fact that the effect on learner language and literacy achievement **decreased** over this period suggests that formal or informal support structures may be necessary to sustain newly acquired knowledge and instructional approaches (Kim, Boyle, Zuilkowski, & Nakamura, 2016).

Sub-question 2.1: To what extent has the integrity of the Stellar program been maintained through scale-up?

The Stellar methodology, detailed to USAID, called for a full year of ongoing teacher support. When Stellar was integrated into the ELIT Intervention, the cluster/support sessions were optional after July 2016. The **integrity** of the Stellar program was thus **not maintained in principle** through scale-up. In terms of the integrity of the ELIT intervention, as expected at the end of 2016, more classrooms showed evidence of the use of Stellar and other ELIT materials in the intervention group. While the materials were observed in almost all classrooms in 2016, many of the ELIT literacy-related **activities were observed in less than half of the classrooms** in 2016 indicating that there may be insufficient uptake to sustain an impact on Grade R learner literacy outcomes. However, **teacher behaviors** before and during the reading exercise **improved overall** from prior to ELIT to after delivery of the intervention, suggesting that some components of the training methodology are beginning to filter through to classroom practices.

Sub-question 2.2: What factors enhance and/or constrain the scale-up of the Stellar training approach?

Key informants noted that, to successfully scale the Stellar training approach through a provincial intervention such as ELIT, **Subject Advisor buy-in** and support is critical. Similarly, they noted that, for Subject Advisors and Lead Teachers, **sufficient training** in adult/peer training approaches is essential. The interviews revealed that trainers and teachers need sufficient training in **phonics**, specifically, to ensure the building blocks are in place for effective literacy instruction. Other factors that affect scale-up, noted by key informants, include **communication** about the intervention, **partnership** in rolling out the intervention and a **training plan** and **resources** for top-up/ refresher training and training new teachers entering the system.

EQ 3: What is the likelihood of sustainability of organizational and beneficiary gains after ELIT ends?



Organizational sustainability capacity is beginning to be built but could be improved through systematic ongoing teacher support. For beneficiary learners, **quality instruction across multiple years is essential**. ELIT has a better chance for sustainability than if it had been rolled out independently because it is incorporated into the operations of the WCED. A number of key informants noted that the WCED should document their **sustainability strategy with a training plan and resources** for refresher courses, top-up training, and training new teachers. The Foundation Phase HOD training on the ELIT program is a step in the right direction and so is the set-up of Grade R teacher communities of practice.

Recommendations

Future impact evaluations should:

- Start with an evaluability assessment.
- Be designed, where possible, in collaboration with policymakers to ensure that the results are fit-for-purpose and use.
- Provide sufficient budget to collect data on important third party variables, particularly those known to affect literacy performance both in South Africa and internationally.
- Require interventions to be implemented in a randomized manner, to allow for the evaluation of impact relative to a robust counterfactual.

Future early grade literacy interventions implemented by provincial governments should:

- Lay out clearly, based on evidence, the minimum dose of instructional time (i.e. hours needed in training over a school year) for effective language and literacy instruction (Kim et al., 2016).
- Articulate what support structures need to be in place to sustain newly achieved learning and pedagogy (Kim et al., 2016), both in the shorter term (within Grade R) and the longer term (carryover effects from Grade R to Grade 1).
- Make space for well-designed impact evaluations, including piloting and refining an implementation model before an impact evaluation is carried out.
- Strategically think about which schools may need which type of interventions most. It is rarely feasible or effective to roll-out a one-size-fits-all intervention to all schools.
- Test whether the funding spent on teacher development is leading to any change in learner performance. Training should be designed so that the effects can be evaluated.
- Ensure strong alignment between the teaching of language in Grade R and in Grade 1. Grade 1 early reading interventions must develop, support, and sustain foundational reading skills.

While more information is needed to better understand policy implications, the evaluation did uncover some implementation lessons:

- The cascade training approach may have had an initial effect but without ongoing provincially mandated support, the intervention advantage appears to have been lost. A recent World Bank report on early grade reading in developing countries indicates that following up initial training with coaching and monitoring compounds the effectiveness of teacher training (Kelly & Graham, 2018).
- The logic of change underlying most professional development programs assumes that there is fidelity to the professional development package in the way that it is administered, rolled out and the contents imparted, demonstrated, and supported in follow-up activities. To document impact, it is very important that the intervention adhere strictly to the program design. In this evaluation, lack of fidelity to the professional development package, implementation modes and support may have negatively affected the change process.

The ELIT intervention appears to have **no lasting impact** beyond Grade R. The evaluation results may suggest that a good Grade R is necessary, but not sufficient, for future language and literacy acquisition.

TABLE OF CONTENTS

Abstract.....	III
Acknowledgements.....	IV
Executive Summary	V
Introduction and Background	1
Evaluation Synopsis of Task, Purpose and Audience	1
Overview of the ELIT Intervention.....	1
Implementation and Evaluation Partners	7
Evaluation Design and Data Collection Methodology	8
Introduction	8
Stellar/ELIT Evaluation Design	8
Assumptions.....	12
Sample Selection	12
Sample Characteristics	14
Data Collection Methodology.....	16
Tools	16
Data Collection.....	19
Data Analysis	19
Reliability.....	20
Constraints and Gaps.....	20
Findings, Conclusions & Recommendations	21
Introduction	21
Findings.....	22
Evaluation Question 1	22
Evaluation Question 2	42
Evaluation Question 3	54
Conclusions, Recommendation, and Lessons	60
Data Quality and Reporting System.....	63
Issues	64
Future Directions.....	65
References	66
Annexes.....	67

ACRONYMS

ANA	Annual National Assessments
AOR	Agreement Officer Representative
CAPS	Curriculum Assessment Policy Statements
CEA	Centre for Evaluation and Assessment
CEMIS	Centralized Educational Management Information System
CET	Classroom Environment Checklist Tool
COR	Contracting Officer Representative
CVC	Consonant, Vowel, Consonant
CVCV	Consonant, Vowel, Consonant, Vowel
DBE	Department of Basic Education
DEC	Development Experience Clearinghouse
DHET	Department of Higher Education and Training
DPME	Department of Planning, Monitoring and Evaluation
ECD	Early Childhood Development
EGRA	Early Grade Reading Assessment
EGRS	Early Grade Reading Study
ELIT	Western Cape Emergent Literacy Intervention
ELMA	ELMA Philanthropies
EQ	Evaluation Question
FET	Further Education and Training
GET	General Education and Training
GPS	Global Positioning System
HO	Head Office
HOA	House of Assembly which formerly served the white (both Afrikaans and English-speaking) population
HOD	Head of Department
HOR	House of Representatives which formerly served the colored (mainly Afrikaans-speaking) population
KAP	Knowledge, Attitudes and Practices
KII	Key Informant Interview

LA/LAT	Learner Assessment
LC	Listening and Speaking in a Conversation
LOE	Level of Effort
LoLT	Language of Learning and Teaching
MO	Missouri
MOU	Memorandum of Understanding
NGO	Non-Governmental Organization
ODK	Open Data Kit
PERSAL	Personal and Salary Administration System
PII	Personally Identifiable Information
PIRLS	Progress in International Reading Literacy Study
RNCS	Revised National Curriculum Statement
SAs	Subject Advisors
SAS	Statistical Analysis Software
SCIP	School Capacity and Innovation Program
SES	Socio-Economic Status
SPSS	Statistical Package for the Social Sciences
Stellar	Strengthening Teaching of Early Language and Literacy in Grade R
TBD	To Be Determined
TOC	Theory of Change
TOCOR	Task Order Contracting Officer Representative
TOT	Training of Trainers
USAID	United States Agency for International Development
USG	United States Government
WC	Western Cape
WCED	Western Cape Education Department
WW	Wordworks

GLOSSARY

Curriculum Assessment Policy Statements (CAPS)	Following a task team review of the Revised National Curriculum Statement (RNCS), the National Curriculum and Assessment Policy Statement (CAPS) was introduced in 2011. CAPS is a policy document that acts as an amendment to the National Curriculum Statement. It was introduced to respond to four key concerns raised in the task team review of the RNCS. Implementation of CAPS was phased in, with the Foundation Phase and Grade 10 scheduled for 2012, the Intermediate Phase and Grade 11 in 2013, and the Senior Phase and Grade 12 in 2014. CAPS replaces the Subject and Learning Area Statements, Learning Program Guidelines, and Subject Assessment Guidelines for the subjects listed in the National Curriculum Statement Grades R - 12.
Criterion-referenced Testing	Criterion-referenced testing measures individual performance against a set of “predetermined criteria or learning standards - i.e., concise, written descriptions of what students are expected to know and be able to do at a specific stage of their education” (Great Schools Partnership, 2014). The purpose of criterion-referenced testing is to determine whether an individual has acquired a specific set of skills, or whether they have learned specific knowledge (e.g. a curriculum). Test-takers who achieve or exceed the required skill level are awarded points. Scoring looks at the number of test-takers who achieve a certain proficiency level, rather than comparing results against a norm.
Grade R	The year prior to formal education in South Africa, equivalent to Kindergarten in the United States. Currently, Grade R is not compulsory but is expected to be required by 2019.
Norm-referenced Testing	Norm-referenced testing refers to the use of standardized tests that are designed to rank test takers and compare them against one another. The results of norm-referenced tests demonstrate the relative performance of test takers against the performance of a “hypothetical average student”. This is determined by comparing individual scores with the scores of other test takers (Hidden curriculum, 2014). Scores are usually reported in terms of a percentage or percentile ranking.

Strengthening Teaching of Early Language and Literacy in Grade R (Stellar)	The Stellar Emergent Literacy Program, designed by Wordworks, supports Grade R teachers in the development of the knowledge and skills needed to successfully teach oral language and emergent literacy in Grade R classrooms. Stellar gives teachers theoretical concepts and practical ideas to build skills with Grade R learners in a “fun, story-based, activity-rich program” (Wordworks, 2016). The program involves year-long attendance of training and reflective workshops where teachers share their experiences and learning with other teachers. The program includes the use of a file of stories, big books, puppets, sequence pictures, and other user-friendly materials. Stellar pre-dates the Western Cape Emergent Literacy intervention (ELIT).
Stellar/ ELIT Evaluation	The Stellar/ELIT Impact Evaluation looks at the 8 original Stellar story modules and an adapted Letters and Sounds Manual, and their use, within the context of the Western Cape Emergent Literacy (ELIT) intervention. Learner Assessment tools for this evaluation are based on the CAPS curriculum and are criterion-referenced. The Teacher Assessment tools are based on Stellar and ELIT. The Classroom Environment Checklist Tool is based on the USAID Early Grade Reading Assessment classroom observation tools and Stellar. While the evaluation does measure the impact of the collective ELIT intervention, it does not specifically measure items related to the Department of Basic Education (DBE) sponsored Grade R Resource Pack, or various other ELIT resources.
Western Cape Emergent Literacy Intervention (ELIT)	The Western Cape Grade R Emergent Literacy (ELIT) intervention is a project implemented in 2015 with training of Subject Advisors and Lead Teachers that reached all Grade R teachers / practitioners at DBE registered sites (approximately 3000) in the Western Cape province of South Africa in 2016. The ELIT intervention included the provision of Stellar materials: eight existing Stellar Story Packs (classroom materials) and one adapted Letters and Sounds Manual ⁸ (teacher manual), in three languages. The intervention also included two purpose-written training manuals for trainers and teachers ⁹ ; the Department of Basic Education (DBE) Grade R Resource KIT; monthly two-hour cluster sessions with teachers; cluster program and facilitators’ notes ¹⁰ ; dry run and cluster session materials ¹¹ ; and a five-day block training workshop. The monthly cluster sessions took place between January and June of 2016, and the 5-day workshop took place in July 2016. ELIT was rolled out by the Western Cape Education Department in partnership with Wordworks.

⁸ Adapted during the lead up to ELIT.

⁹ One training manual for Subject Advisors and one (resource-focused) manual for Lead Teachers and teachers, both written by Wordworks in collaboration with the WCED.

¹⁰ For Lead Teachers.

¹¹ A variety of materials to support Subject Advisors in monthly preparation workshops with Lead Teachers and to support Lead Teachers in monthly preparation of cluster workshop sessions with teachers.

Wordworks Wordworks is a South African non-profit organization that focuses on early language and literacy development in the first eight years of children's lives. Since 2005 Wordworks has worked in under-resourced communities with those adults best positioned to impact on young children's language and literacy development – parents and caregivers, family and community members, home-visitors, early childhood development practitioners and Grade R to Grade 3 teachers. Wordworks developed the "Strengthening Teaching of Early Language and Literacy in Grade R" (Stellar) program

LIST OF FIGURES

Figure 1: ELIT Rollout	3
Figure 2: ELIT Theory of Change	6
Figure 3: Evaluation Design	11
Figure 4: Sample Characteristics	14
Figure 5: Average Wave 1 Score Differences between Intervention and Comparison Groups ..	24
Figure 6: Average Wave 2 Score Differences between Intervention and Comparison Groups ..	25
Figure 7: Average Wave 3 Score Differences between Intervention and Comparison Groups ..	26
Figure 8: Average Score Differences between Intervention and Comparison Groups on EGRA	27
Figure 9: Words Read Correctly per Minute against the WCED Benchmark	28
Figure 10: Paragraph reading: words read correctly per minute against WCED benchmark.....	28
Figure 11: Mean Standardized Index Score for Each Data Collection Wave	29
Figure 12: Distribution of the Index Scores for Each Data Collection Wave.....	30
Figure 13: Teacher Response - Preferred Classroom Context	49
Figure 14: Mean Performance per Sub-Task – Comparison and Intervention Groups.....	72
Figure 15: Correlations between Sub-Tasks in Wave 1 and Wave 2 with Wave 3	73
Figure 16: Main Results per Sub-Task	75
Figure 17: Sensitivity Check - Fieldworker Effects.....	76
Figure 18: Sensitivity Check - Attrition Bias.....	76
Figure 19: Heterogeneous Effects - Learner Gender.....	76
Figure 20: Heterogeneous Effects - Language of Learning and Teaching	77
Figure 21: Heterogeneous Effects - School Quintile.....	77
Figure 22: Enrolment in Public Schools in the Western Cape by Year and Grade, 2007-14.	78
Figure 23: Enrolment in the Western Cape by Year and Grade, 2014 – 2016, SNAP Survey....	78
Figure 24: Average Grade 3 Literacy Score in the Western Cape Systemic Evaluation	79
Figure 25: Performance on Learner Assessment - Expected Criterion Levels (CAPS)	82
Figure 26: Sustainability: Environmental Support.....	83
Figure 27: Sustainability: Funding Stability.....	83
Figure 28: Sustainability: Partnerships	84
Figure 29: Sustainability: Organizational Capacity	84
Figure 30: Sustainability: Program Evaluation	85
Figure 31: Sustainability: Program Adaptation	85
Figure 32: Sustainability: Communications.....	86
Figure 33: Sustainability: Strategic Planning	86
Figure 34: Gr 1 Pacing and Curriculum Coverage: Resources/Materials Used	87
Figure 35: Gr 1 Pacing and Curriculum Coverage: Primary Resource Used	87
Figure 36: Gr 1 Pacing and Curriculum Coverage: Pages Written Work (DBE Workbooks)	88
Figure 37: Gr 1 Pacing and Curriculum Coverage: Pages Written Work (Other Materials)	88
Figure 38: Gr 1 Pacing and Curriculum Coverage: Teacher Feedback (DBE Workbooks)	89
Figure 39: Gr 1 Pacing and Curriculum Coverage: Teacher Feedback (Other Materials)	90
Figure 40: Gr 1 Pacing and Curriculum Coverage: Quality of Feedback (DBE Workbooks)	91
Figure 41: Gr 1 Pacing and Curriculum Coverage: Quality of Feedback (All Other Materials) ...	91
Figure 42: Gr 1 Pacing and Curriculum Coverage: Quality of Writing (DBE Workbooks).....	92
Figure 43: Gr 1 Pacing and Curriculum Coverage: Quality of Writing (All Other Materials).....	92
Figure 44: Gr 1 Pacing and Curriculum Coverage: Creativity of Writing (DBE Workbooks)	93

Figure 45: Gr 1 Pacing and Curriculum Coverage: Creativity of Writing (All Other Materials)	93
Figure 46: Gr 1 Pacing and Curriculum Coverage: Length of Writing (DBE Workbooks)	94
Figure 47: Gr 1 Pacing and Curriculum Coverage: Length of Writing (All Other Materials)	94
Figure 48: Classroom Environment (Teaching Method)	95
Figure 49: Classroom Environment (Activities Seen)	95
Figure 50: Classroom Environment (ELIT Materials Seen).....	96
Figure 51: Classroom Environment (Before Reading)	96
Figure 52: Classroom Environment (During Reading)	97

LIST OF TABLES

Table 1: Attrition Rate	14
Table 2: Difference-in-Differences Results	23
Table 3: Main Results Models	31
Table 4: Lead Teacher Performance on Section 1 of the ELIT Knowledge Questionnaire.....	43
Table 5: Lead Teacher Performance on Section 2 of ELIT Knowledge Questionnaire.....	45
Table 6: Percentage of Teachers with Correct Response on KAP Section 1	45
Table 7: Percentage of Teachers with Correct Response on KAP Section 2	47

LIST OF ANNEXES

Annex 1: Summary Statistics and Results	
Annex 2: Methodology Report Extract	
Annex 3: Data Collection Tools	
Annex 4: Schedules and Project Activities	
Annex 5: Interview Lists and Meetings	
Annex 6: Sources of Information	
Annex 7: Statement of Work	
Annex 8: Statement of Differences	
Annex 9: Evaluation Team	
Annex 10: Learning Brief for the Government of South Africa	

INTRODUCTION AND BACKGROUND

EVALUATION SYNOPSIS OF TASK, PURPOSE AND AUDIENCE

In July 2015, USAID commissioned Khulisa Management Services to conduct an Impact Evaluation¹² of the “Strengthening Teaching of Early Language and Literacy in Grade R” (Stellar) program^{13,14}, funded by USAID¹⁵, designed and implemented by Wordworks in the Western Cape province of South Africa. Stellar supports Grade R teachers in the development of the knowledge and skills needed to successfully teach oral language and emergent literacy in Grade R classrooms. Simultaneously Wordworks, together with the Western Cape Education Department (WCED), co-developed the Emergent Literacy (ELIT) intervention incorporating 8 Stellar story modules and an adapted Letters and Sounds Manual.

This evaluation therefore constitutes an investment by the United States government to evaluate the effects of the **combined Stellar/ELIT intervention** (hereafter referred to as the “ELIT intervention”) over the period July 2015 (evaluation award) to February 2018 (final report).

The **purpose** of the Stellar/ELIT evaluation is to support the South African government in its efforts to improve Grade R quality and to raise early grade literacy scores. The evaluation aims to understand if the ELIT intervention, incorporating 8 Stellar story modules and an adapted Letters and Sounds Manual, is successful in improving the language and literacy skills of Grade R and, ultimately, Grade 1 learners.

The **primary audience** for the evaluation includes USAID and its co-funders - ELMA Philanthropies and JP Morgan - Wordworks, and the WCED. Secondary evaluation audience members include the National Department of Basic Education (DBE). It is expected that the DBE and WCED will **use the results and lessons learned** from the evaluation to inform their approach to improving the quality of Grade R teaching in South Africa, in general, and in the Western Cape Province, in particular. It is also expected that the implementers will use the evaluation results to inform future programming and that USAID and other donors will use the evaluation results to guide future investments in early grade literacy interventions and evaluations in South Africa.

OVERVIEW OF THE ELIT INTERVENTION

The ELIT intervention was a province-wide project for all Grade R educators and practitioners. The purpose was to strengthen the teaching of Home Language (HL) with the use of appropriate

¹² USAID defines an impact evaluation as that which: “require[s] a credible and rigorously defined counterfactual to control for factors other than the intervention that might account for the observed change”.

¹³ Cooperative Agreement No.: APS 674-12-00001, Award Dates: January 2014 through December 2016.

¹⁴ See the Glossary for a description of the Stellar program

¹⁵ USAID funding for implementation (\$940,517.00) and evaluation: (\$ 453,249.13) in partnership with JP Morgan and ELMA Philanthropies: \$159,483.00. Total funding: \$1,553,249.13. The Stellar Program Cooperative Agreement Number: APS 674-12-00001. Implementation dates: January 2014 through December 2016.

resources to improve language and literacy performance levels in the Foundation Phase. The program responds to a need for better quality language and literacy instruction in Grade R in South Africa. As implemented until 2011, the Grade R program had little impact on the educational development of learners in later years (Taylor, 2014).

In 2016, the ELIT intervention was rolled out simultaneously to all Grade R teachers at DBE registered sites (approximately 3000) in the Western Cape province of South Africa. The ELIT intervention included the provision of Stellar materials: eight existing Stellar Story Packs (classroom materials) and one adapted Letters and Sounds Manual¹⁶ (teacher manual), in three languages. The intervention also included two purpose-written training manuals for trainers and teachers¹⁷; the Department of Basic Education (DBE) Grade R Resource KIT; monthly two-hour cluster sessions with teachers; cluster program and facilitators' notes¹⁸; dry run and cluster session materials¹⁹; and a five-day block training workshop.

With funding support through USAID's School Capacity & Innovation Program (SCIP), which seeks to improve literacy in the early grades²⁰, 8 original Stellar story modules and an adapted Letters and Sounds Manual were scaled²¹ through the ELIT intervention. It is important to note that Stellar was not rolled out in its initial form but was integrated into a wider approach to emergent language and literacy development (as discussed above).

The WCED and Wordworks used a Training-of-Trainers (TOT) supported cascade training model to roll out the ELIT intervention. In the first stage, Wordworks trained Head Office (HO) staff, the WCED Early Childhood Development (ECD) and Foundation Phase Subject Advisors. Wordworks also contracted and trained three NGOs in the Western Cape to support WCED Subject Advisors during the roll-out at scale. In the second stage, the WCED Subject Advisors trained Lead Teachers. Finally, in the third stage, the Lead Teachers trained other teachers in circuits or district-based geographic clusters. Training included monthly support/cluster meetings for teachers, opportunities for teacher reflection, peer learning, collaboration, and other opportunities to build on existing knowledge. The actual rollout was timed as follows:

¹⁶ Adapted during the lead up to ELIT.

¹⁷ One training manual for Subject Advisors and one (resource-focused) manual for Lead Teachers and teachers, both written by Wordworks in collaboration with the WCED.

¹⁸ For Lead Teachers.

¹⁹ A variety of materials to support Subject Advisors in monthly preparation workshops with Lead Teachers and to support Lead Teachers in monthly preparation of cluster workshop sessions with teachers.

²⁰ Aligned with USAID's education strategy, SCIP provides funding support to test, refine, and scale up local South African initiatives that support teachers and school management teams to improve their instructional practices.

²¹ Note that ELIT is a scaled intervention in itself. The 8 Stellar story modules, on the other hand, were initially distributed through the Stellar program to a limited number of schools in the Western Cape. These classroom materials have been scaled to the entire province via ELIT.

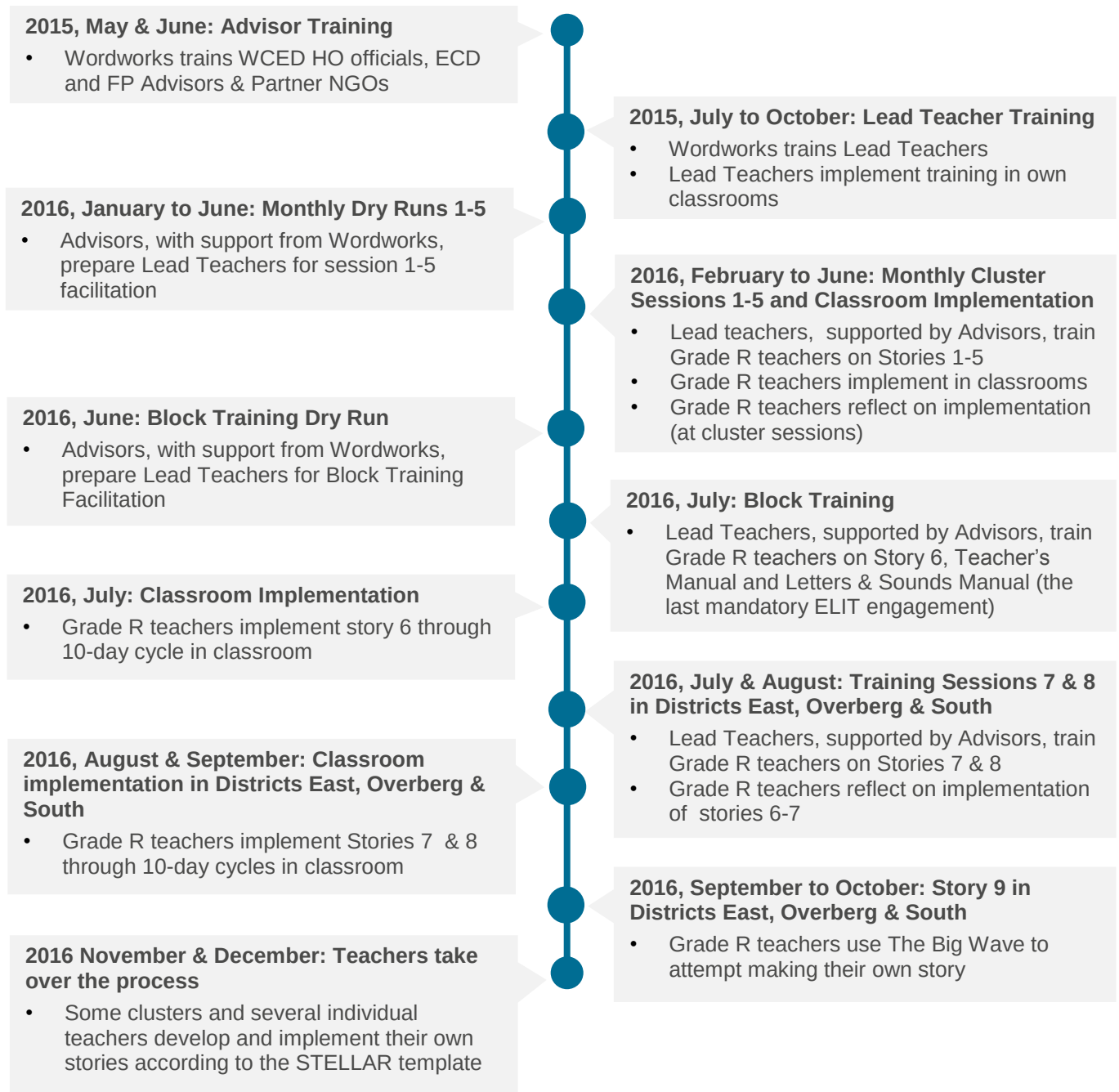


Figure 1: ELIT Rollout²²

²² Source: (Smith & Hugow, 2017)

There are some differences in how the Stellar program was expected to be implemented (i.e. initial design) versus how it was rolled out once it was integrated into the ELIT intervention:

Original Stellar Design	Actual ELIT rollout
<u>Block training</u>: Towards the beginning of 2016 (at the start of the intervention with Grade R teachers), based on a standard Stellar training course.	<u>Block training</u> was conducted half-way through the year and was based on ELIT, incorporating 8 Stellar story modules and adapted Letters and Sounds Manual. The ELIT manuals (one training manual for subject advisors and one resource-focused manual for lead teachers and teachers) were purpose-written by Wordworks in collaboration with the WCED and included CAPS principles and resources.
<u>Cluster meetings</u>: Monthly hands-on training in Stellar material and stories, and reflection on implementation. Eight monthly clusters over a calendar year – February to November excluding July. The purpose was to work with teachers in the second half of the year as children’s literacy develops over the course of the year. Some aspects would not be taught in Terms 1 and 2 that do need to be introduced in Term 3 and 4 – particularly with regard to letters and sounds. Teachers were to be supported in developing their own story modules based on the Stellar method and routine for the remainder of the year. Stellar Letters and Sounds Manual with practical activities for phonics activities during each 2-week story cycle (week 1 focused on whole class activities while week 2 focused on small group activities).	<u>Limited cluster meetings</u>, starting in February and mandated until July. Voluntary for districts to continue²³. It is unclear how many districts continued ELIT cluster meetings. Due to the decentralized nature of the cluster meetings, it is not known exactly what was handed out and in which sequence. Teachers were partially trained on the adapted Stellar Letters and Sounds manual during 3 cluster sessions²⁴ and were more extensively trained on the Manual at the block training in July²⁵. Stellar was adapted to the WCED 10-day cycle; both small group and whole class activities were included in each week and additional small group activities were listed for week 1, and 5 whole class activities for week 2.
<u>Other interventions</u>: No other literacy teaching interventions in Grade R or Grade 1 over the period of the evaluation.	The evaluation team requested that <u>no other language and literacy interventions</u> be rolled out during the evaluation timeframe. There is no evidence to suggest that other system-wide interventions were implemented, but it cannot be confirmed that no schools received any other literacy interventions during the period of the evaluation.

²³ The ELIT rollout plan noted, “Teachers may continue monthly clusters to suit context” and “Optional provincial cluster workshop” in September.

²⁴ Each cluster session included a 25 minute walk-through of the Stellar Letters and Sounds Manual and 20 minutes introducing some of the phonological awareness and letter-sound activities

²⁵ Half a day was dedicated to the Letters and Sounds Manual at the block training in July.

The Theory of Change (TOC) of the ELIT intervention²⁶ is:

IF Subject Advisors are trained in emergent literacy concepts by skilled emergent literacy practitioners (Wordworks) **AND** they receive quality training materials and support, **THEN** they will be sufficiently prepared to train and support Lead Teachers in emergent literacy concepts. **AND IF** Lead Teachers are trained in emergent literacy concepts by Subject Advisors **AND** they receive quality training materials, attend monthly cluster sessions for three months, participate in dry runs (practical sessions whereby trainers modelled the approach to be adopted during the clustered teacher training - Smith & Hugow, 2017), receive support, and attend block training for one week **THEN** they will be sufficiently prepared to train and support Grade R teachers in emergent literacy concepts.

Likewise, **IF** Grade R Teachers receive training from Lead Teachers in emergent literacy concepts **AND** Grade R teachers receive quality training materials, attend monthly support/cluster sessions, and attend block training for one week at the end of the cluster sessions, **THEN** Grade R teachers will introduce emergent literacy concepts in their classrooms **AND** they will change the way they teach emergent language and literacy concepts. **THEN** the quality of Grade R learner reading will improve **AND THEN** Grade R learners will be better prepared for learning in future grades.

The TOC is such that teachers receive the intervention, but that the effects of this should be seen on the language and literacy abilities of their learners. One of the key assumptions of the intervention is that exposing teachers to a series of monthly cluster meetings, followed by a five-day training workshop conducted by Lead Teachers, would be sufficient to build the capacity of Grade R teachers to provide better emergent literacy instruction to Grade R learners. Another key assumption is that the effects of the intervention, if any, on Grade R learners would be retained and built on as the learners moved into Grade 1.

²⁶ See Annex 7 for the Stellar Theory of Change

The Theory of Change (TOC) of the ELIT intervention

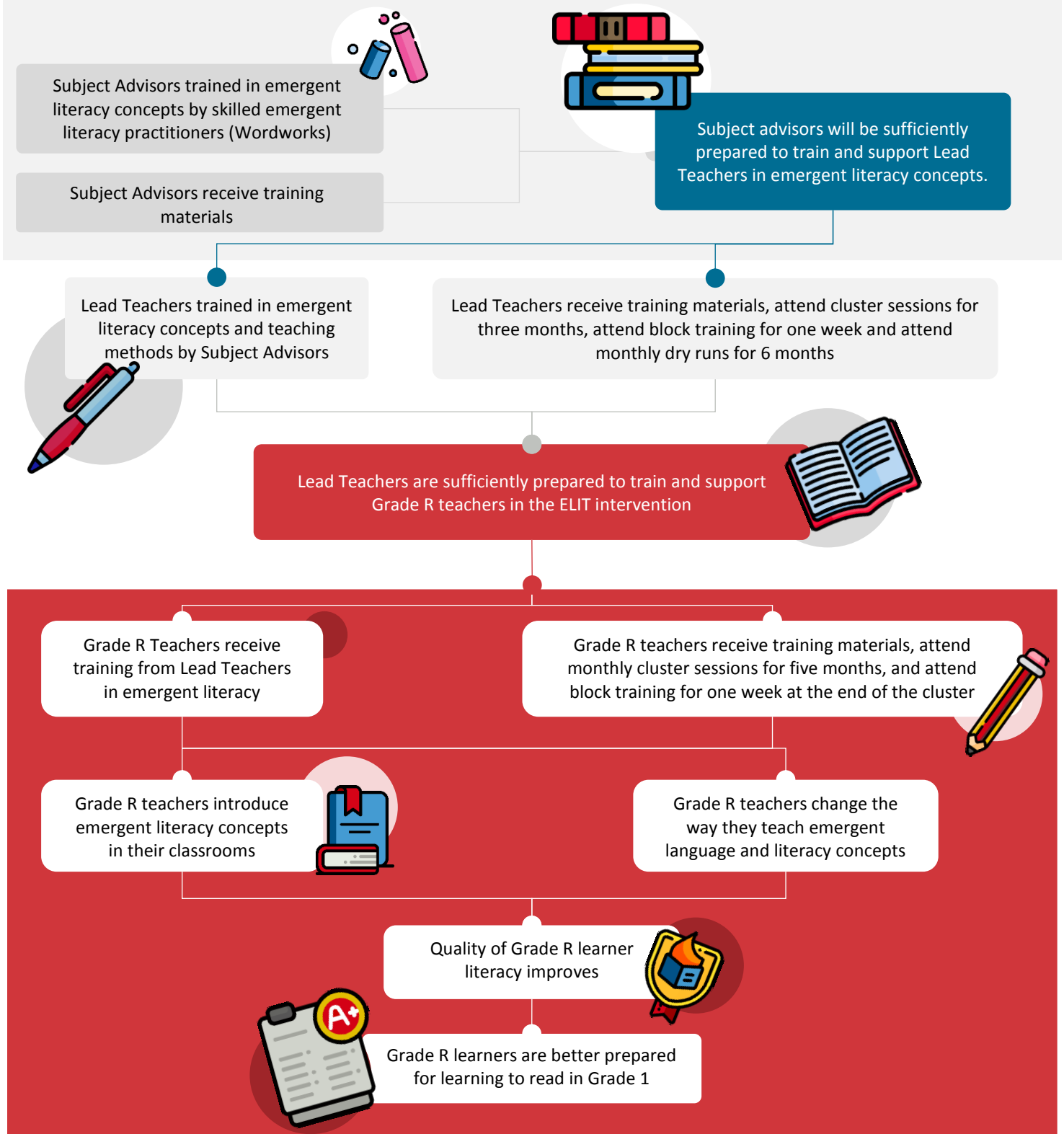


Figure 2: ELIT Theory of Change

IMPLEMENTATION AND EVALUATION PARTNERS

The implementing partners for this evaluation include Wordworks and the Western Cape Education Department. Wordworks was responsible for training the WCED ECD and Foundation Phase Subject Advisors (SAs) and NGO-based support advisors in the ELIT program, providing Stellar program materials and training resources, and providing ongoing teacher support to SAs and Lead Teachers. The WCED Curriculum Directorate's SAs were responsible for selecting, training and providing support to Lead Teachers in the ELIT intervention, managing the distribution of materials and training logistics in each district, supporting Lead Teacher training of teachers, and occasionally monitoring teacher progress and providing support and feedback (Smith & Hugow, 2017).

The evaluation partners include Khulisa, Wordworks, the Western Cape Education Department, ELMA Philanthropies, and USAID. Khulisa led the evaluation with support, engagement, and continual feedback from the evaluation partners.

The Annexures contain more information on the evaluation. These include:

- Annex 1: Summary Statistics and Results
- Annex 2: Methodology Report Extract
- Annex 3: Data Collection Tools
- Annex 4: Schedules and Project Activities
- Annex 5: Interview Lists and Meetings
- Annex 6: Sources of Information
- Annex 7: Statement of Work
- Annex 8: Statement of Differences
- Annex 9: Evaluation Team
- Annex 10: Learning Brief for the Government of South Africa



Photo: Khulisa

EVALUATION DESIGN AND DATA COLLECTION METHODOLOGY

INTRODUCTION

USAID defines impact evaluation as an approach that is based on models of cause and effect, which “...require[s] a credible and rigorously defined counterfactual to control for factors other than the intervention that might account for the observed change” (USAID, 2013). In other words, an impact evaluation needs to account for the fact that changes in outcomes are potentially affected by many factors outside of the intervention being implemented – some of which are observable and some of which are unobservable.

Simple pre-post comparisons of change in outcomes are almost never a credible estimate of the counterfactual because they provide no insight as to what would have happened if the program were not carried out. One can therefore not be sure whether the change observed between the pre- and post-assessment is due to the program, or whether it is change that happened irrespective of the program. All change seen from pre- to post- assessment would have to be due to the effect of the program to be a true estimate of the program’s impact. This assumption simply does not hold most of the time and, for this reason, this kind of comparison is known as a “counterfeit counterfactual” (Gertler, Martinez, Premand, Rawlings, & Vermeersch, 2011).

An impact evaluation first has to establish what would have happened if the program had not been carried out, and then has to establish what additional effect the program had on top of what would have happened anyway – and that the effect was a direct consequence of the program. The most robust estimate of the counterfactual occurs when an intervention is assigned at random to a population. Every individual has an equal probability for selection for intervention and this means that there is a high statistical probability that the intervention and comparison/ control groups will be statistically identical in both their observed characteristics as well as their unobserved characteristics (Gertler et al., 2011). What this does is ensure that any differences between these groups can only be explained by the intervention. An important assumption is that no other (external) factors systematically affect one group and not the other over the time-period of the intervention.

STELLAR/ELIT EVALUATION DESIGN

The ELIT intervention was planned for rollout to all teachers in the province at the same time in 2016 (i.e. not a phased rollout) and so 2015 was the only year that comparison group data could be collected. Given the universal rollout of the ELIT intervention to all teachers across the Western Cape province at the same time, it was impossible for the intervention to be assigned at random to the population of Grade R learners in the Western Cape and so to obtain a credible

counterfactual group of learners during the same year²⁷. The alternative was to randomly select a comparison group of learners in either the Eastern or Northern Cape provinces as these are the only provinces that also offer Afrikaans, English and isiXhosa in Grade R. The evaluation team decided that contextually these provinces were too different to the Western Cape to warrant a credible counterfactual.

The evaluation team explored a number of options (documented in the Methodology Report Extract in Annex 2) before proposing a **quasi-experimental research design** using the 2015 Grade R learners in the classes of teachers who had not yet received training as the comparison group. The intervention group comprised the 2016 cohort of Grade R learners – one year later - in the classes of the same teachers who were in the comparison group in 2015. This design allows for a **counterfactual** that estimates learning that took place in Grade R learners before the teachers received training. Comparing these learning gains with the learning gains after teachers received training allows us to quantify the additional learning as a result of the program. The evaluation team also used the same schools for the comparison and intervention cohorts to reduce any bias in the results that may have originated from unobservable school-level variables that remained constant over the period of the evaluation.

The evaluation team **initially proposed** using a difference-in-difference estimator to estimate the treatment effect, which is defined as “...the difference in average outcome in the intervention group before and after intervention minus the difference in average outcome in the comparison group before and after intervention...” (Albouy, 1994). That is, the learning gains for the 2015 Grade R cohort (the comparison group) would be compared with the learning gains for the 2016 Grade R cohort (the intervention group) in the same schools. However, a number of unexpected changes to the evaluation affected this approach:

Original design	Actual rollout
Evaluation award: Early 2015	The evaluation was awarded in June 2015
<u>Baseline assessment</u> : As close to the beginning of the year as possible, to determine the “raw” state of entry into education, i.e. what learners knew at the beginning of Grade R.	Assessment in August, after half a year of teaching. For the intervention group, this meant that the whole intervention of block training and cluster groups had already occurred.
<u>Mid-term assessment</u> : Near the end of Grade R, after the intervention of block training and with ongoing cluster meetings reinforcing the training. Training via cluster sessions to take place over a full year.	Assessment in October/November. The evaluation did not collect evidence of cluster meeting support. The ELIT roll-out plan shows that there were no provincially mandated cluster sessions after block training in July although the WCED mandated the use of ELIT in the classroom.
<u>End of term assessment</u> : Near the end of Grade 1.	As planned.

²⁷ In our proposal, Khulisa noted “Should random assignment of teachers into intervention and control groups be introduced (or a variant of this approach. For example, random phase-in), Khulisa would be willing to change its evaluation approach from a quasi-experimental to a Randomized Control Trial (RCT)”.

Most estimation strategies for impact evaluations necessitate, or are improved through, the collection of clean baseline data. Impact evaluations also require sufficient time between a baseline and a second measurement for an intervention to yield a measurable impact (Skiles, Hattori, & Curtis, 2014). The evaluation team was awarded the evaluation contract in June 2015, at which point the opportunity for the collection of clean baseline data collection was lost as teaching had already been underway in the schools for the first 6 months of the year.

The earliest data collection could take place was in August 2015 (due to winter school holidays, pre-testing of instruments, and training of fieldworkers). Since the school year ends in early December, the comparison group had to be assessed in November 2015. For the counterfactual to be credible, the intervention group could also only be tested for the same time-period from August 2016 to November 2016. This affected the evaluation team's original assumptions regarding the difference-in-differences approach. The delay decreased the probability that the evaluators would find significant short-term effects due to the short time between the first and second assessments. It also compromised the collection of clean baseline data as the intervention group had already received the whole intervention by August of 2016.

In early August 2015, the evaluation team, evaluation funders, and implementers held an open discussion about the evaluation limitations given these field realities. The decision was taken to pursue the evaluation considering these realities. It was acknowledged that the ability of the evaluators to estimate short-term effects using a difference-in-differences method was compromised but that the design remained strong in its ability to find long-term effects. At this meeting, the WCED noted that teachers would be mandated to attend training until July 2016, but that ongoing teacher support/cluster sessions after July were voluntary.

The evaluation design features were subsequently adjusted and the analysis was restructured to compare the average score differences between the intervention and comparison groups at three points in time – the middle of Grade R (after 6 months of training for the intervention group), the end of Grade R, and the end of Grade 1. To avoid confusion around the term “baseline”²⁸, in this report the three points of data collection in this evaluation are referred to as “wave 1” (mid-Grade R), “wave 2” (end-Grade R), and “wave 3” (end-Grade 1).

The evaluation design is illustrated in Figure 3 below.

²⁸ Because the first point of data collection occurred after the intervention had already taken place in the intervention group, the term “baseline” is misleading and will therefore not be used in this report.

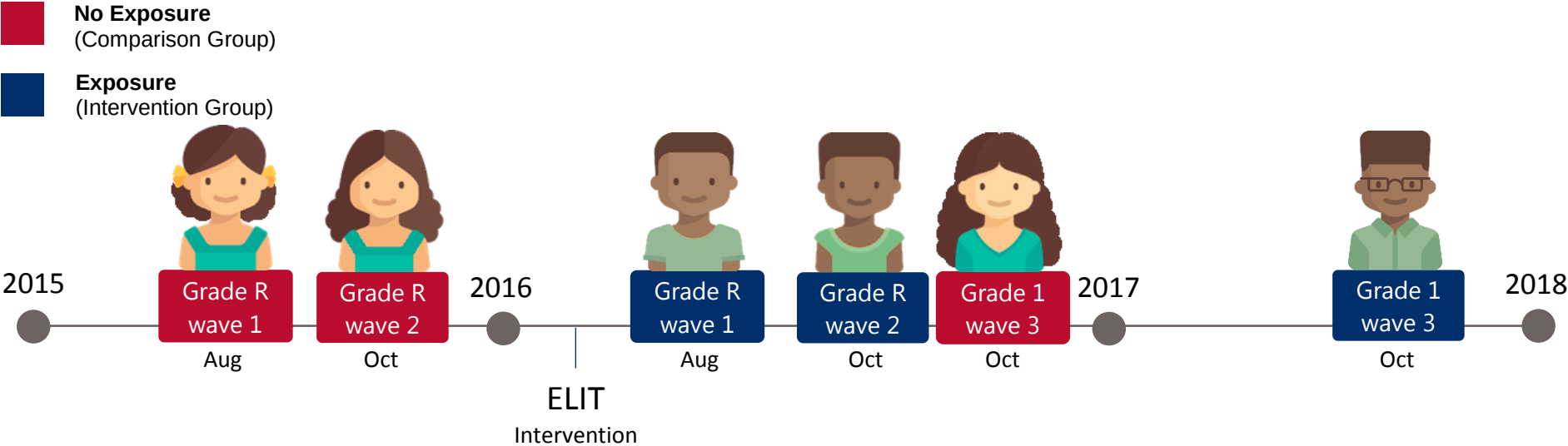


Figure 3: Evaluation Design

Assumptions

Given that it was not possible to obtain clean baseline results, it is not possible to establish whether the intervention and comparison groups were of an equal ability level. The evaluation therefore rests on the major assumption that inherent learner ability levels do not differ significantly from one cohort to the next. A further assumption is that there would be no other provincial-level factors that would affect learning (i.e. no additional school days lost in 2016 vs 2015 due to protests/ strikes, or no changes in policies that would affect all schools in the province).

The evaluation team checked the enrolment rates of learners by grade in the Western Cape over the period 2007 to 2014 to determine whether there were any major fluctuations in learner numbers from year to year. The team also checked the most recently available Grade 3 systemic scores to see if there were any major differences between the evaluation sample and the province as a whole. The results demonstrate the relative stability of the system²⁹.

Recognizing that the cohort identification strategy rests on a relatively strong assumption, various other measures were taken to ensure comparability across the years. The comparison and intervention groups were selected from the same schools to ensure that school and teacher level contextual factors remained constant across the years. Furthermore, the same assessment instruments were used across the data collection points to ensure high correlation between the mid-Grade R, end-Grade R and end-Grade 1 assessments³⁰.

Sample Selection

Given the evaluation design, the comparison and intervention schools are the same schools. Fifty schools were selected in 2015 using a stratified random sampling approach³¹. School quintile³² was chosen as the stratification variable to ensure that the sample would be able to provide information regarding the success of the intervention across all school functionalities. Stratification by school Language of Learning and Teaching (LoLT) would have resulted in over-sampling of highly functional schools, as English schools are overrepresented in quintiles 4 and 5 in the Western Cape. Ten schools per quintile were sampled, with an additional two schools per quintile drawn as potential substitute schools³³. In each school, one class was selected through a process of random sampling whereby the classes were listed, a number between 10 and 15 randomly selected, and the class on which the random number fell was picked. In each classroom, 9 learners were randomly selected using the rule $(1 + (n/9))$, which resulted in 450 learners tested in total.

²⁹ See Figures 27 - 29 in Annex 1

³⁰ Although there were some changes in the instruments *between* the waves, the exact same assessment instruments were used with the comparison and intervention groups *within* each wave.

³¹ Although the intervention was not randomised, random selection of schools and learners was undertaken to ensure that the sample was representative.

³² South African schools are divided into five categories called *quintiles* based on the socio-economic status of the community in which the school is situated. Quintile 1 schools are the poorest, while quintile 5 schools are the least poor. In South Africa, the socio-economic status of the school is correlated with school functionality.

³³ These schools were only assessed in wave 1 if problems arose at the initially sampled schools.

In the final sample, four of the substitute schools were tested in place of the originally sampled schools. Substitution reasons include:

- One selected school was undergoing maintenance and so classes were disrupted over the fieldwork period.
- In three other cases, logistical difficulties prevented the fieldworkers from being able to travel to the schools.

In each round of data collection, the exact same learners were identified and assessed so learners could be tracked longitudinally over the duration of the evaluation. During the August 2016 intervention group assessment (wave 1), one school had discontinued their Grade R classes since the wave 1 data collection for the comparison group in August 2015. This brought the total schools in the sample down to 49. In response, and given the previous year's attrition rate, Khulisa increased the number of learners assessed per school from 9 to 10. This raised the intervention group sample from N = 450 to N = 490 - 40 more learners than the previous year.

Attrition

During the first wave of data collection, test score data was collected from 450 comparison group learners in 50 schools, and 490 intervention group learners in 49³⁴ of the same schools. During the third wave of data collection, the assessments were administered on exactly the same learners that participated in the first and second waves. The first and second waves were in August and November of the same year and therefore it was highly likely that the learners would still have been in the same schools and in the same classes.

Wave 3 was conducted a year later, after learners transitioned to Grade 1. In this transition, it emerged that 25 comparison group learners and 21 intervention learners moved to schools outside of the original sample of schools. The fieldworkers attempted, as far as feasible, to test these learners in their new schools. As expected, given existing trends, six Grade R learners repeated a year and were therefore not assessed in Grade 1. The repeating learners' assessments were excluded in the analysis, as the repeaters in the comparison group would have been exposed to the intervention during their second year in Grade R.

Attrition was very low in the second wave of data collection, largely given the short time duration between the first and second waves. Attrition in the third wave was higher particularly because of Grade R repeaters and learner mobility in Grade 1. Attrition was slightly higher in the comparison group than in the intervention group. Sensitivity analyses were conducted and this did not have a significant effect on the final results (see Figure 23, Annex 1).

³⁴ Only 49 of the comparison schools also served as the intervention schools, as one school discontinued Grade R in 2016.

Table 1: Attrition Rate

	Comparison	Intervention
Wave 1 (mid-Grade R)	450	490
Wave 2 (end-Grade R)	434	481
Wave 3 (end-Grade 1)	397	450
Attrition Wave 2	3.6%	1.8%
Attrition Wave 3	11.8%	8.2%

Figure 16, Annex 1, shows the learner characteristics for the intervention and comparison groups across the three waves of data collection. During wave 1, the sample of learners had no statistically significant differences on any of the observable learner characteristics. Through attrition, the intervention group lost a slightly higher proportion of girls, which means that the intervention group has a slightly higher proportion of boys than the comparison group in the third wave (end-Grade 1). The sample attrition did not result in any further imbalances in terms of observable learner characteristics.

Sample Characteristics

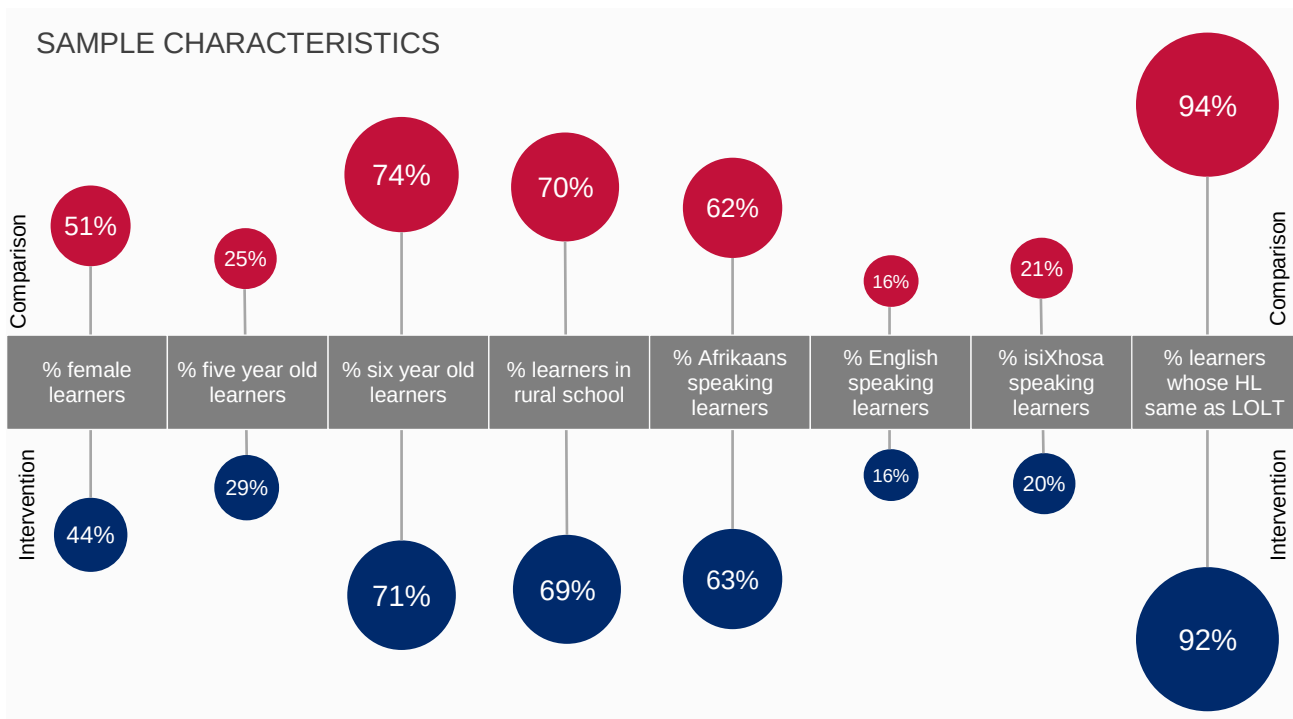


Figure 4: Sample Characteristics

Schools

As noted previously, the sample was specifically selected to ensure the comparison schools in 2015 served as the 2016 intervention schools. This selection ensured that the sample is relatively

balanced on a school level and, mostly, on a teacher level³⁵. Figures 14 and 15 in Annex 1 show the school and teacher characteristics in the sample.

Given that the sample was stratified on school quintile status, the distribution of LoLT, location (urban/rural), and the teacher characteristics were not controlled. Approximately 70% of the schools are urban schools, 64% of the schools have Afrikaans as their LoLT, 18% have English as their LoLT, and a further 18% have isiXhosa as their LoLT.

Although the language distribution is uneven, it is better explained when considering the former department³⁶ under which the schools were administered:

- 58% under the House of Representatives (HOR) administration which formerly served the colored (mainly Afrikaans-speaking) population;
- 10% under the Department of Education and Training (DET) administration which formerly served black isiXhosa speakers;
- 16% under the House of Assembly (HOA) administration which formerly served the white (both English and Afrikaans-speaking) population; and
- 16% are new schools that have been built/established since 1994.

The larger Afrikaans sample is therefore a result of the composition of the Western Cape, as well as the method of stratification used. Most (71%) of the English primary schools in the Western Cape are in the Metro South and Metro Central districts but, as Stellar had previously been implemented in these districts, they were excluded from the sample to avoid potential contamination. Furthermore, 54% of all English primary schools in the Western Cape are quintile 5 schools, which means that the probability of them being included in the sample was reduced due to stratification. Likewise, only 11% of primary schools in the Western Cape have isiXhosa as their language of learning and teaching.

Teachers

As the same schools were tested under intervention and comparison conditions, most Grade R teachers remained the same between 2015 and 2016³⁷. All the teachers in the sample were female. The teachers in the intervention group were slightly younger on average, due to a few new teachers joining the sample. However, the difference in the average age between the two samples is not statistically significant.

The Grade R teachers had on average approximately 10 years' experience, whereas the Grade 1 teachers had approximately 16 years' experience. With regards to qualifications, approximately 80% of the Grade R teachers had a certificate or a diploma and only 8% had a degree or more. This is understandable given the minimum requirements for Grade R practitioners.

³⁵ 41 of the teachers remained the same from wave 1 in 2015 to wave 1 in 2016. The remainder were either substitute teachers (i.e. the teacher was sick on the day of assessment) or new teachers.

³⁶ Before 1994, funding to schools depended on the department under which they were administered.

³⁷ See footnote 18

As learners went to different Grade 1 classes, the Grade 1 teachers varied more substantially. For example, more teachers had a degree in 2017 than in 2016. A third (35%) of teachers who received comparison group learners in 2016 had a degree compared to half (50%) of teachers that received intervention group learners in 2017. Two thirds (64%) of the Grade 1 teachers of comparison group learners only held a certificate or diploma in teaching compared to half the Grade 1 teachers of the intervention group.



Photo: Khulisa

DATA COLLECTION METHODOLOGY

The tools are attached in Annex 3.

Tools



The **Learner Assessment** (LA) instrument was the primary tool designed to measure learner performance in this evaluation. At the time of the evaluation design, there were no known standardized, widely used, and validated in-country Grade R language and literacy assessments available. Language, literacy, and test development consultants developed the tool in line with USAID's instructions in the evaluation Terms of Reference – that it be aligned to the Curriculum Assessment Policy Statements (CAPS) and relevant to Stellar. The tool is therefore criterion-referenced³⁸ as the intervention was carried out within the CAPS framework and is CAPS aligned. As the evaluation covers performance from Grade R into Grade 1, items from both curricula were included in the instrument. Some Grade 2 items were also included to measure “extension” performance.

³⁸ A description of the difference between norm and criterion referenced tests is available at the following hyperlink: <http://www.edpsycinteractive.org/topics/measeval/crmref.html>

The Learner Assessment tool measures all four CAPS Home Language skills:

- Listening and speaking (12 items, 6 each for *listening and speaking in a conversation* and 6 for *listening and speaking based on a story*),
- Phonics (12 items, split into *phonological awareness* and *phonics letters*),
- Emergent *reading* (and viewing - 6 items) and
- Emergent *writing* (6 items).

Items were constructed so that answers could be anchored to as many of the following categories as possible, using a 6-point scale: Not done (0), Partial Grade R level (1), End of Grade R level (2), Partial Grade 1 level (3), End of Grade 1 level (4), Grade 2 level (5). To ensure measurement consistency between the comparison and the intervention group assessments, the same tool was used with both the intervention and comparison groups in each wave. However, to avoid test-retest bias, a few changes were made to the LA tool between the waves³⁹.

 Due to initial ceiling effects in the Grade R assessments, the evaluation team added four subtests from the **Early Grade Reading Assessment**⁴⁰ to the Grade 1 learner assessment. The WCED had already adapted the EGRA tool to be comparable between the three local languages spoken in the province (Afrikaans, English and isiXhosa) and this version was used in the assessment. The four additional EGRA items include the letter sound recognition subtest, word recognition subtest, the paragraph reading subtest and reading comprehension questions based on the paragraph reading subtest.



A **Classroom Environment Checklist Tool** (CET) was used to collect qualitative data on the context of all Grade R classrooms. The CET included items from a USAID Early Grade Reading Assessment (EGRA) evaluation, focusing on instruction approaches, teacher actions, class environment, and pupil engagement in the area of language and literacy, which were adapted to suit the South African Grade R context. The tool also included observation items specific to what was expected of teachers following training in ELIT and looked at whether both Stellar and other ELIT materials and resources filtered through to the Grade R classrooms. The CET covers five broad aspects of classroom practice including pedagogic practices, learning and literacy activities, resources used by teachers, learning and teaching materials visible in the classrooms, planning resources and overall classroom appearance.

The tool was administered during all waves of data collection, in both the comparison and intervention groups⁴¹. In Grade 1 classes, a shortened version of the CET was used focusing specifically on the availability of reading materials and a reading corner, visual stimulation, and overall classroom appearance. Multiple observations of the classroom were made randomly in a

³⁹ The (emergent) *reading* and (emergent) *writing* subtasks were kept the same for all three waves, as these did not draw heavily on memory. For the *phonological awareness* and *phonics letters* subtasks, a different word was constructed in each wave to test the skill (e.g. constructing any CVC word) instead of testing learner memory. In the *listening and speaking based on a story* subtask, two new stories were developed for the end of Grade R (wave 2) and Grade 1 (wave 3) assessments.

⁴⁰ Developed by RTI International, the EGRA is a norm-referenced, standardised, test used to measure learners' progress toward learning to read from Grade 1 onwards.

⁴¹ 50 Grade R classes in 2015 and 49 Grade R classes in 2016

day: when the fieldworker arrived at the school; as the fieldworker fetched each learner from, and returned each learner to, their class (a minimum of 10 short observations); and for a focused observation period of 1 hour during the shared storybook reading session in Grade R (i.e. the “story for instruction” session). Fieldworkers focused their attention on the overall physical appearance of the classroom, literacy-related resources in the classroom, activities engaged in during class teaching, how the Grade R teacher interacted with the class, and the way in which she read a storybook to the learners.



A **Teacher Knowledge, Attitudes and Practices (KAP)** questionnaire was used to collect information on the knowledge, attitudes, and practices of all Grade R and Grade 1 teachers. The tool incorporated some questions from the Wordworks ELIT Knowledge Questionnaire. Data was collected at three points from all Grade R teachers⁴²: before training (in March 2016), after cluster and block training (in August/ September 2016), and after the delivery of the intervention to learners (in November 2016).



A **Pacing and Curriculum Coverage** assessment tool was developed by a South African early grade reading expert in collaboration with USAID. The tool assesses Grade 1 learners’ written work and comprises five main areas, namely, curriculum coverage, learner feedback, quality of writing, creative writing, and length of writing. Learners’ work comprises class workbooks, DBE workbooks, worksheets, homework books, other books (teacher reader, storybook) and additional exercises. Fifteen schools (a third of the sample of schools) were randomly selected for the assessment. First, the schools were separated into quintiles 1 – 3 and quintiles 4 – 5 and the third school on each list was selected. Learners were then randomly selected from a list of learners that were evaluated in Grade R (i.e. the same learners were tracked from Grade R into Grade 1). The fieldworkers were instructed to select every 2nd learner from the list until they had 4 learners. These learners were not necessarily in the same Grade 1 class, which resulted in more teachers participating in the assessment in 2016.

The assessment of learner work stretched from the beginning of Term 3 of the 2016 academic year to the second week of Term 4 of the 2016 academic year. This timing was due to CAPS which specifies that Grade 1 learners will begin complex and creative writing in the third term. Another reason is that the amount of work to be assessed needed to be kept to a reasonable limit to allow each fieldworker half a day to assess the materials.



A **Sustainability Assessment Tool** was adapted to measure the Subject Advisors’ perspective of the sustainability of the ELIT intervention. The tool is a 40-item self-assessment survey adapted from the “Program Sustainability Assessment Tool”, copyright 2012, Washington University; St. Louis, MO. Khulisa obtained written permission from Washington University to adapt the tool and provided the university with a copy of the adaptation. The purpose of the tool was to assess the capacity of the ELIT intervention for sustainability across a range of organizational and contextual factors, identifying sustainability strengths and challenges that could be used to guide sustainability planning.

⁴² 50 Grade R teachers in 2015 and 49 Grade R teachers in 2016

The sustainability assessment was conducted in the districts in which the evaluation was implemented with the Subject Advisors who were responsible for training Lead Teachers in the ELIT evaluation. The assessment was conducted after the delivery of the ELIT intervention in 2016 and repeated one year later in 2017 to gauge whether the project's capacity for sustainability declined or improved over this period. The WCED sent 64 Subject Advisors an email requesting their participation in the online sustainability assessment survey. Of these, 31 Subject Advisors (48%) completed the survey in 2016 and 19 (30%) completed the survey in 2017.



In 2017, the evaluation team also conducted **Key Informant Interviews (KIIs)** with key stakeholders including USAID, Wordworks, ELMA Philanthropies, and the WCED. KIIs were carried out telephonically and in-person over the period July to October 2017. The purpose of these KIIs was to gain the perspective of key stakeholders on the sustainability of the ELIT intervention.

Key Informant Interviews were also carried out with a sample of 36 Grade 1 teachers from 36 sample schools. This was not part of the original design. However, given the differences between the Grade 1 learner results in 2016 and 2017 (discussed in the findings section in this report), the evaluation team felt it was important to investigate what could have caused these differences. In consultation with the evaluation team members and EGRA and literacy evaluation experts outside of the evaluation team, five questions were identified and used to conduct a short follow-up interview with these teachers.

It is important to note that the samples for the CET, Pacing, and Curriculum Coverage assessment, teacher KAP and Sustainability Assessment were small and thus the margin of error is large. All results are interpreted accordingly.

Data Collection

Prior to the first wave of data collection, the Learner Assessment tool was pilot tested at a selection of schools in Gauteng province. The evaluation fieldworker team reviewed the tools and provided their feedback, particularly regarding the translation of the documents into Afrikaans and isiXhosa.

Khulisa engaged the services of a well-established research organization based in the Western Cape, with a range of retired teachers with relevant foundation phase experience to carry out the fieldwork. The fieldworkers comprised first language English, Afrikaans and isiXhosa speakers and the learner assessments were carried out in the LoLT of the school.

Over six rounds of training/refresher training, Khulisa trained 17 fieldworkers on the use of the Open Data Kit (ODK) mobile phone-based data collection system and each of the evaluation tools (see Figure 3 for the evaluation design and points of data collection).

Data Analysis

To ensure that the data collected was explored as far as possible to provide an estimate of the ELIT intervention's effects, the evaluation team conducted an analysis of the differences between the performance of the comparison and intervention groups at each wave of data collection, controlling for learner-level characteristics. Given that the intervention and comparison learners were in the same schools and mostly in the classes with the same teachers, school level and teacher level contextual factors are inherently controlled for in the models.

The evaluation team attempted to strengthen data credibility and validity in this evaluation through two types of triangulation. Data triangulation was used to draw on external research and results from different evaluation tools to contextualize the findings. Investigator triangulation was also used, whereby different evaluators reviewed the same data and provided their interpretations.

Reliability

The overall reliability coefficient (Cronbach's alpha) for all of the items was 0.80 for each of the Learner Assessments in waves 1, 2, and 3. Cronbach's alpha is high enough to indicate good reliability, but not so high that the test items are redundant. This provides a good indication of the internal consistency of the Learner Assessment instrument, signifying that the tool is consistently measuring the underlying construct that it intends to measure - language and reading proficiency.

Figure 19 in Annex 1 shows the correlation between each of the sub-tasks in the wave 1 and wave 2 assessments with the sub-tasks in the wave 3 assessment. Using Cohen's conventions for interpreting effect size (Ellis, 2009), in both the intervention and comparison groups the reading sub-tasks in wave 1 were weakly to moderately correlated with all the sub-tasks in the wave 3 assessment. This was largely driven by the overall weak performance in the reading items in wave 1, as the correlation between the wave 2 reading tasks and the wave 3 items was much higher. The *phonological awareness*, *phonics letters*, and *writing* sub-tasks in both wave 1 and wave 2 had the strongest correlations with the wave 3 sub-tasks.

The same sub-tasks (*phonological awareness*, *phonics letters*, and *writing*) were also moderately correlated with the EGRA items. *Listening and speaking in a conversation*, and *listening and speaking based on a story*, were the least correlated sub-tasks. The correlation in the intervention group was slightly higher than the correlation in the comparison group, specifically related to the latter sub-tasks. A sensitivity check was conducted and the inclusion of these sub-tasks did not significantly affect the results.

Constraints and Gaps

Given that the focus of this evaluation is on the effects of the ELIT project in the Western Cape, the language composition of the sample is a credible representation of primary schools in the province. However, the study will not be able to determine the efficiency of the ELIT program in different languages and the results cannot be generalized to other provinces. On the other hand, stratification by quintile ensures that Socio-Economic Status is adequately represented in the sample, which is one of the largest determining factors of educational performance in South Africa.

As mentioned in the evaluation design section of this report, other constraints included the inability of the evaluation to collect a clean baseline and the inability to conduct a randomized control trial due to the intervention's simultaneous rollout to all Grade R teachers and practitioners in the Western Cape Province.

FINDINGS, CONCLUSIONS & RECOMMENDATIONS

INTRODUCTION

The three overarching questions that are addressed in this evaluation are:

1. Have Stellar interventions increased the language and literacy achievement of targeted learners over time?
 - 1.1. To what extent have Grade R interventions impacted the language and literacy acquisition and performance of Grade R and Grade 1 learners in comparison to the control group and expected outcomes for these grades?
2. To what extent has the train-the-trainer model resulted in more effective and knowledgeable teachers and positive language and literacy outcomes for learners?
 - 2.1. To what extent has the integrity of the Stellar program been maintained through scale up?
 - 2.2. What factors enhance and/or constrain the scale-up of the Stellar training approach?
3. What is the likelihood of sustainability of organizational and beneficiary gains after Stellar interventions end?

FINDINGS

Evaluation Question 1



Have Stellar interventions increased the language and literacy achievement of targeted learners over time?



Assuming no *unobservable confounding factors* may have influenced the outcomes, the results suggest that ELIT interventions had an **initial positive effect** on the learners' language and literacy achievement. However, this effect **decreased with time** – the results show a statistically significant *increase in language and literacy achievement in the first half of Grade R*, which decreased by the end of Grade R, and disappeared by the end of Grade 1. This suggests that ELIT interventions, incorporating 8 Stellar story materials and an adapted Letters and Sounds Manual, could have given Grade R learners a temporary head start but this learning advantage was not sustained over time. Correspondingly, an assessment of a small sample of learner writing revealed that Grade 1 teachers did not pace significantly faster through the curriculum with the intervention group.

The **Learner Assessment** instrument is the primary tool designed to measure learner performance in this evaluation. The instrument measures all four CAPS Home Language skills: listening and speaking (12 items, 6 each for *listening and speaking in a conversation* and 6 for *listening and speaking based on a story*), phonics (12 items, split into *phonological awareness* and *phonics letters*), (emergent) *reading* (and viewing - 6 items) and (emergent) *writing* (6 items).

Due to initial high performance on two items in the Learner Assessment, possible ceiling effects were foreseen and the evaluation team took pre-emptive measures to include four additional items from the **Early Grade Reading Assessment** (EGRA) tool in the Grade 1 assessment: *letter sound recognition*, *word recognition*, *paragraph reading*, and *reading comprehension* based on a paragraph.

To assess whether Grade 1 teachers paced faster through the curriculum, and therefore covered more of the curriculum, a **Pacing and Curriculum Coverage Assessment** was carried out with a small sample (15 of 49) of schools. A Pacing and Curriculum Coverage Tool was developed by a South African early grade reading expert in consultation with USAID. It assessed Grade 1 learners' written work and comprised five areas; quantity of written work, quality of teacher feedback, quality of learner writing, creativity of learner writing, and length of learner writing.

This section is structured as follows:

1. Original Difference-in-Differences Estimation Design
2. Revised Estimation Approach
 - a. Results: Learner Assessment
 - b. Results: Early Grade Reading Assessment
 - c. Results: Combined Index Score
 - d. Results: Regression Results

- e. Results: Pacing and Curriculum Coverage
- 3. Robustness Checks
- 4. Checks for Heterogeneous Effects
- 5. Summary of Findings

1. Original Difference in Differences Estimation Design

As it was not possible to obtain a clean baseline, the wave 1 score contains two sources of information: (1) information on inherent learner ability and (2) information on the intervention effect. For this reason, the original difference-in-differences estimation design was compromised because most of the value added by the intervention was likely effected before the first wave of data collection. Although this situation resulted in positive early gains in the intervention group, it also negatively biased the estimate of the effectiveness of the ELIT intervention as the initial intervention effect (over the first half of the year) was differenced-out in the model. The model therefore is only able to pick up any additional learning after the initial intervention effect. The intervention effect was calculated by adding the coefficients for the intervention variable, the wave variable, and the beta variable (Table 2). Models A2 and B2 include the learner level controls – learner age and learner gender). The results in both models show that at the second wave of data collection (end-Grade R), there was a positive intervention effect (0.237 standard deviations difference in model A1 and 0.241 in model A2). However, by wave 3, the effect was statistically insignificant (-0.102 standard deviations difference in model B1 and -0.096 in model B2).

Table 2: Difference-in-Differences Results

	Wave 2		Wave 3	
	A1	A2	B1	B2
Intervention	0.394*** (0.062)	0.401*** (0.061)	0.394*** (0.064)	0.403*** (0.063)
Learner Age		0.199*** (0.055)		0.188*** (0.055)
Learner Gender = Female		0.194*** (0.053)		0.235*** (0.053)
WAVE	-0.014 (0.037)	-0.015 (0.037)	-0.017 (0.048)	-0.01 (0.048)
BETA	-0.143** (0.051)	-0.145** (0.052)	-0.479*** (0.066)	-0.489*** (0.066)
N	1855	1849	1786	1780

*0.1; **0.05; *** 0.01

2. Revised Estimation Approach

The evaluation team revised the approach to conduct a mean comparison for each round of data collection. This approach allows a direct comparison between the mean wave 1, wave 2 and wave 3 scores to assess the differences in the learning gains made over the full 14-month period. These results are discussed in detail below.

a) **Results: Learner Assessment**

When interpreting these results it is necessary to keep in mind the evaluation design. Due to delays at the start of the project, the comparison group was exposed to 6 months of teaching and the intervention group was exposed to 6 months of teaching - plus the intervention – at the first wave of data collection. It was therefore expected that the wave 1 (mid-Grade R) scores of the intervention and comparison groups would be different.

Figures 5 - 7 below graphically represent the average learner scores for each of the sub-tasks of the Learner Assessment for each wave of data collection. In both the wave 1 and wave 2 assessments, the learners in both the intervention and comparison groups scored relatively poorly on the (emergent) *reading*, *phonological awareness* and *phonics letters* sub-tasks. However, as expected due to an additional year of teaching, large improvements were seen in these sub-task areas by the end of Grade 1 (wave 3).

In wave 1 (Figure 5), there were clear and statistically significant differences between the intervention and comparison groups in the (emergent) *writing* sub-task and in the *listening and speaking based on a story* sub-task.

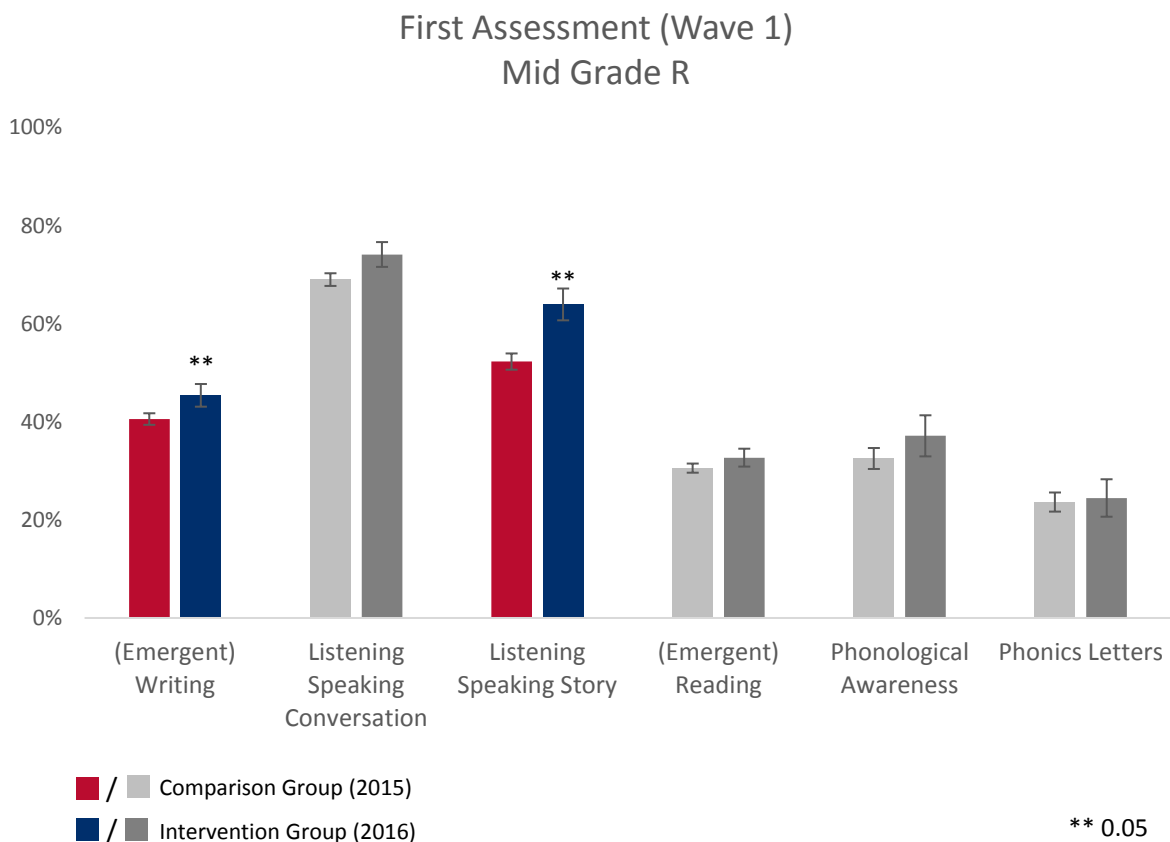


Figure 5: Average Wave 1 Score Differences between Intervention and Comparison Groups

In wave 2 (Figure 6), there was an overall increase in learner scores in both the intervention and comparison groups, as expected due to the effects of Grade R teaching over this time-period.

However, statistically significant differences remained visible in the (emergent) *writing* sub-task at this point.

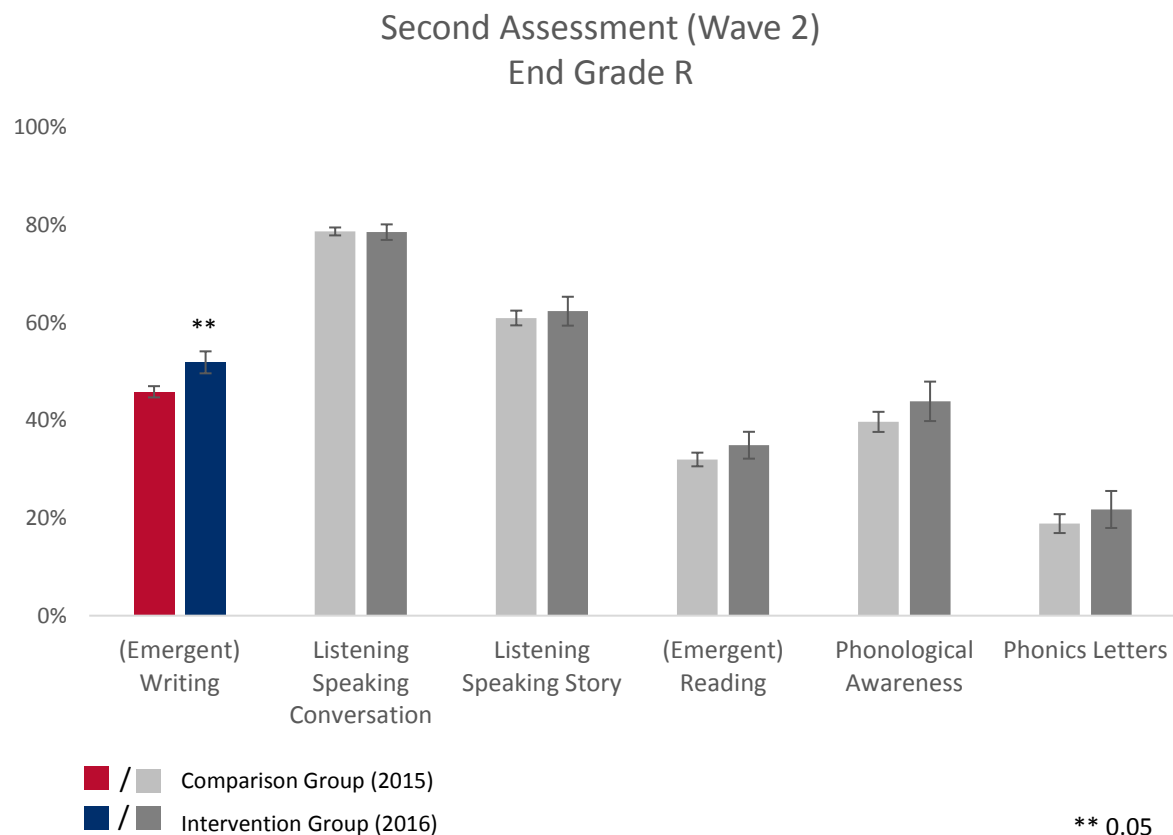


Figure 6: Average Wave 2 Score Differences between Intervention and Comparison Groups

In wave 3 (Figure 7), learners in both the comparison and intervention groups performed very well in all six sub-tasks of the Learner Assessment. However, there were no statistically significant differences visible between these groups, mainly due to ceiling effects in most sub-tasks. Interestingly, both the intervention and comparison groups performed slightly worse in the *listening and speaking based on a story* sub-task at the end of Grade 1 compared to the first and second waves of data collection. It is possible that the story used – Joshua’s goal – was difficult for Grade 1 learners to comprehend⁴³.

⁴³ Due to time constraints, the evaluation team was unable to pilot test the two stories developed for this evaluation – Baby Bird (end Grade R) and Joshua’s Goal (end Grade 1). The stories were reviewed by experts to determine their appropriateness to the South African Grade R context.

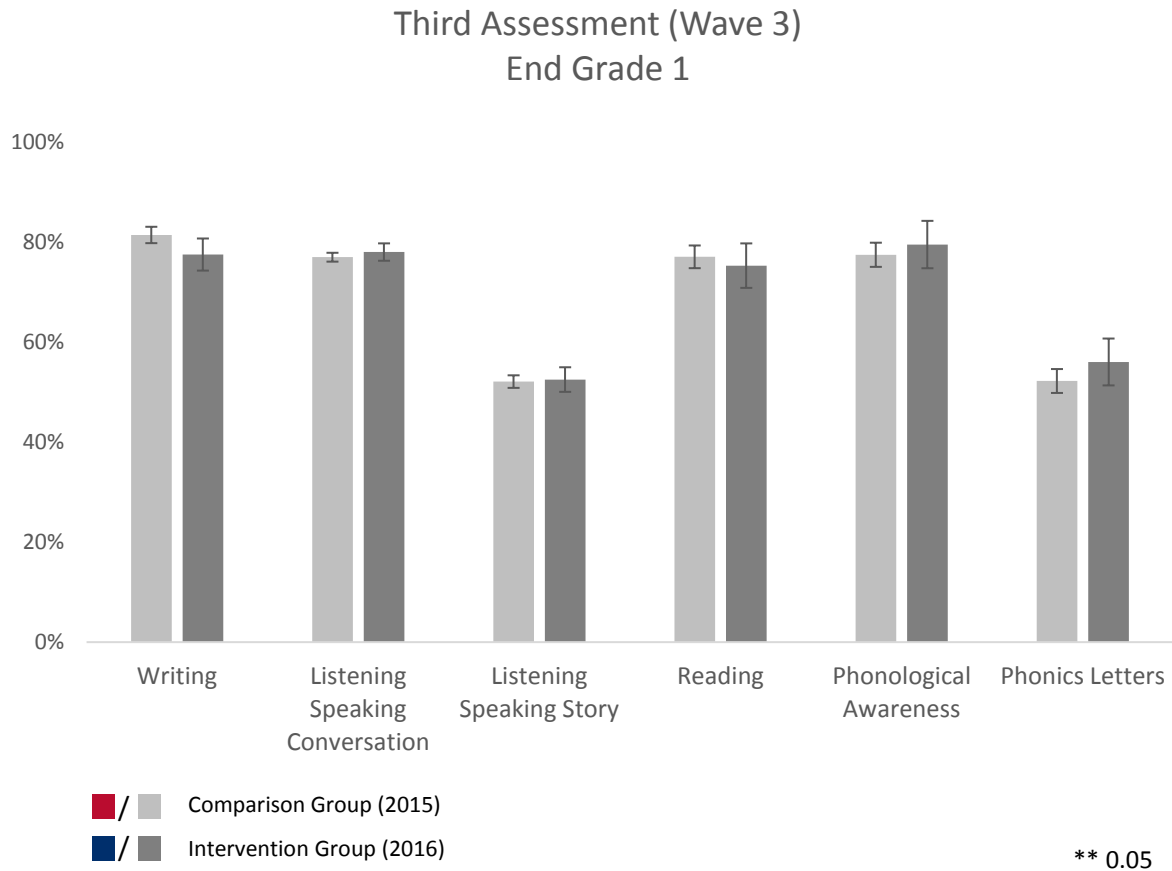


Figure 7: Average Wave 3 Score Differences between Intervention and Comparison Groups

b) Results: EGRA Sub-tasks

In addition to the Learner Assessment, the Early Grade Reading Assessment (EGRA) was administered at the end of Grade 1. Learners in the comparison and intervention groups performed equally well in the *letter sound recognition* subtest of the EGRA. Interestingly, the comparison group outperformed the intervention group in *word recognition* and *paragraph reading*. For the *word recognition* subtest, this difference was statistically significant.

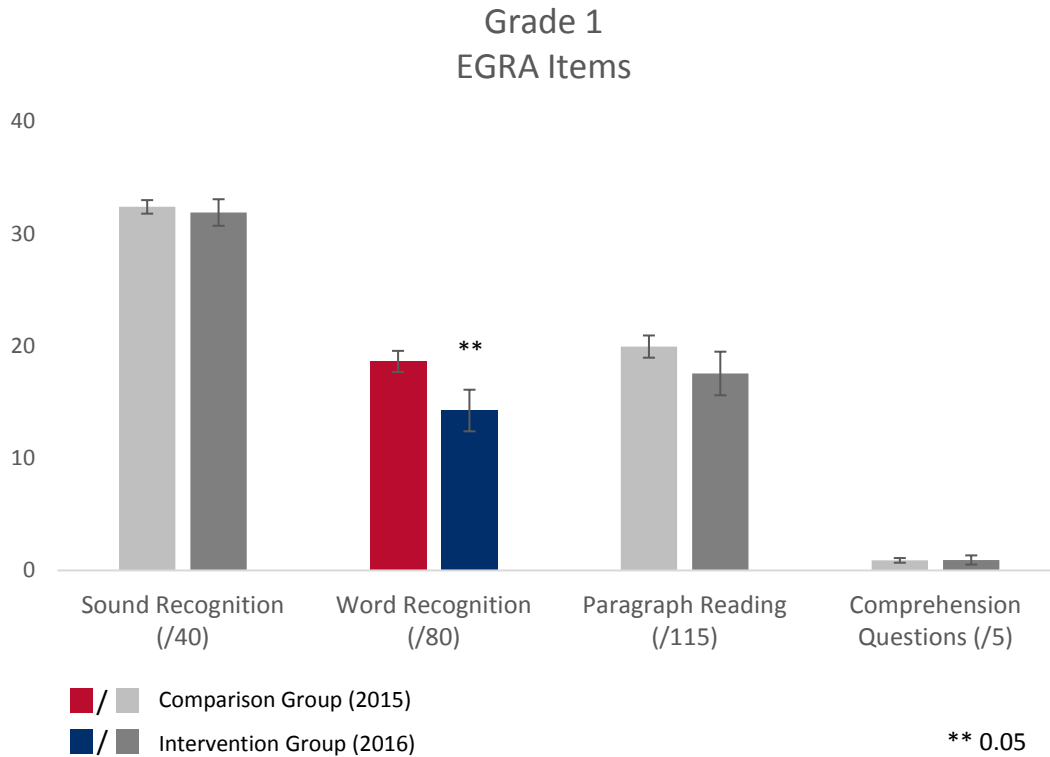


Figure 8: Average Score Differences between Intervention and Comparison Groups on EGRA

Figures 9 and 10 show the differences between the intervention and comparison groups in the two EGRA sub-tasks that measure words read correctly per minute. The *word recognition* sub-task (Figure 9) measured learners' ability to read a list of words within a minute, whereas the *paragraph reading* sub-task (Figure 10) measured learners' ability to read words within the context of a paragraph in a minute. The WCED established their own benchmarks for each language group.

There was a normal distribution in learner performance on the EGRA sub-tasks *word recognition* and *paragraph reading* but poor performance overall. The *comprehension* questions, based on paragraph reading, seemed to have been difficult for the learners to answer with only learners at the 75th percentile managing to answer one question correctly.

With regard to *word recognition*, only the Afrikaans comparison group and the isiXhosa intervention group managed to perform within range of the WCED benchmarks of 20 words per minute, respectively. In wave 1, there were clear and statistically significant differences between the intervention and comparison groups in the (emergent) *writing* sub-task and in the *listening and speaking based on a story* sub-task. Per Figure 9, and contrary to expectation, the Afrikaans intervention group read fewer words per minute than the Afrikaans comparison group (17 words per minute versus 21 words on average per minute). This trend requires further investigation before being interpreted as a real decline. A similar trend was observed in the *paragraph reading* sub-task (Figure 10), with the Afrikaans comparison group learners reading more words per

minute than the intervention group (20 words versus 15 words on average per minute)⁴⁴. Overall, Grade 1 learners are not managing to attain the required levels of reading fluency established by the WCED.

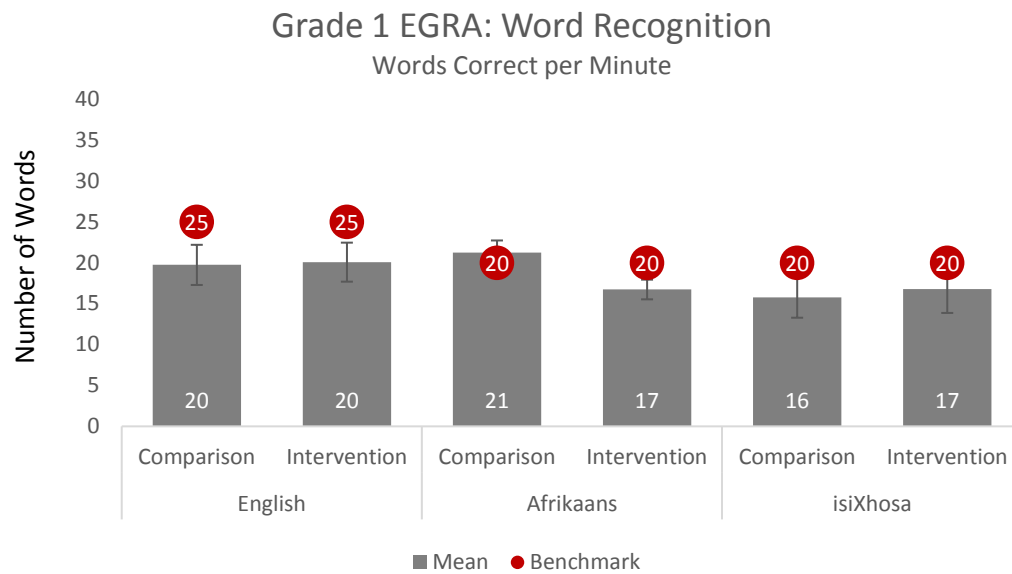


Figure 9: Words Read Correctly per Minute against the WCED Benchmark

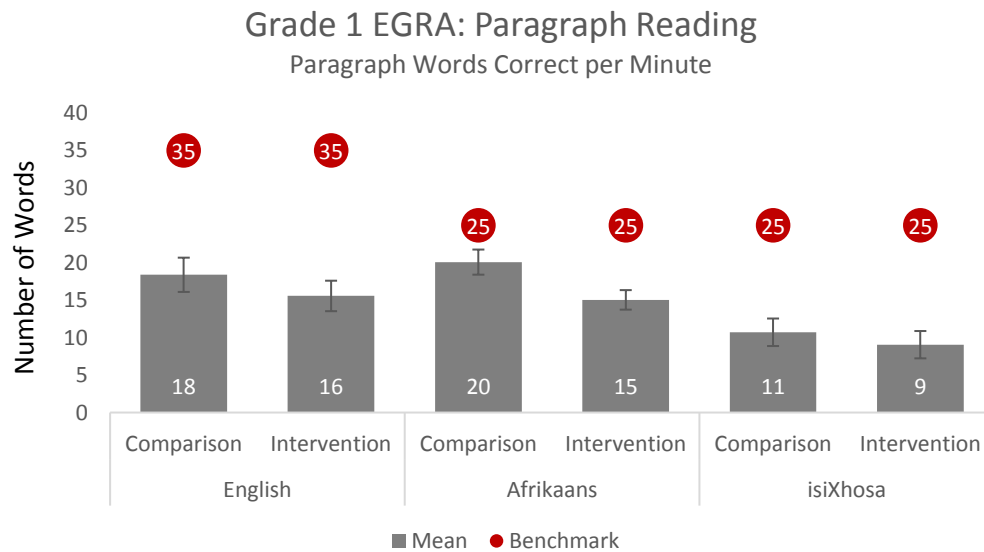


Figure 10: Paragraph reading: words read correctly per minute against WCED benchmark

⁴⁴ Given the stratification of the sample based on school quintile, and the unique demographic composition of the Western Cape, most learners in the sample are Afrikaans speakers.

c) **Overall Results: Combined Index Score**

Using Principal Component Analysis, the learner performance results on all the different sub-tasks - both on the Learner Assessment and the EGRA instruments - were combined to give a single composite index score. The index score was standardized to the mean and standard deviation of the comparison group score. This effectively means that for each wave of data collection, the comparison group mean was set to zero and the intervention group mean is shown relative to the comparison group mean of zero.

Figure 11 below shows the mean index score for the comparison and intervention groups in each wave of data collection. There was a large and statistically significant difference between the performance of the comparison and intervention groups in wave 1 (mid-Grade R, 6 months after ELIT cluster meetings began in the intervention group). This difference decreased in wave 2 (end-Grade R) though remained statistically significant. In wave 3 (end-Grade 1), the difference between the groups was not statistically significant, with the intervention group performing slightly worse overall.

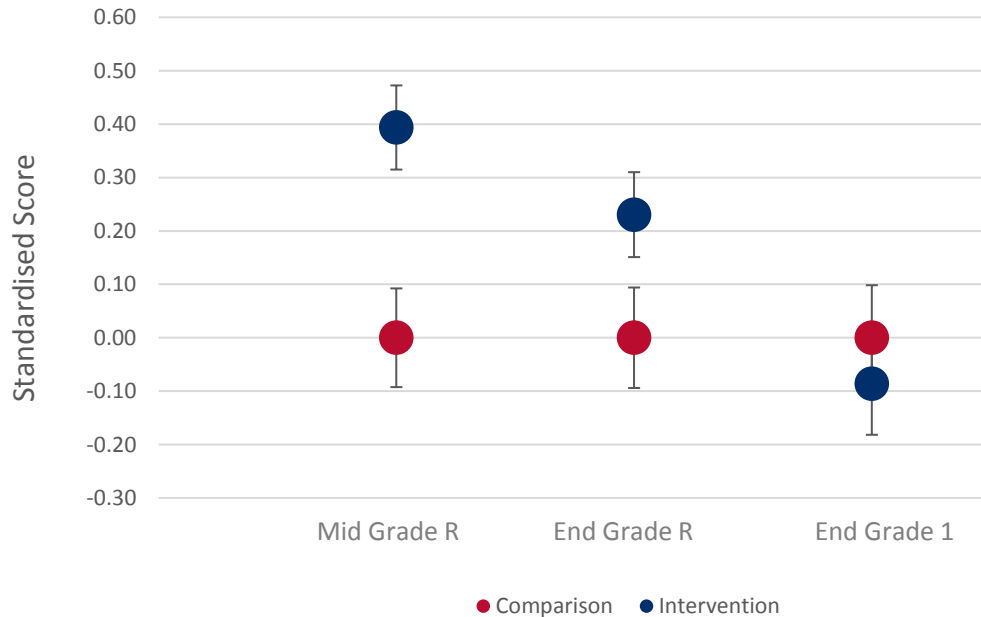


Figure 11: Mean Standardized Index Score for Each Data Collection Wave⁴⁵

Figure 12 below visually represents the difference in the distribution of test scores between the intervention and comparison groups for each wave of data collection. Contrary to Figure 11 above, the index scores used in Figure 12 were not standardized relative to the comparison group, and therefore compare the relative performance of the intervention and comparison groups over all three waves. Figure 12 demonstrates that the intervention group performed better than the comparison group at all levels of learner performance in both wave 1 and wave 2. However, in

⁴⁵ The scores in this Figure are standardised relative to the comparison group mean for each wave of data collection. The error bars represent the confidence interval.

wave 3, the intervention group distribution lies slightly to the left of the comparison group distribution. This indicates that a higher proportion of Grade 1 intervention learners were at the lower end of performance compared to the comparison group learners in Grade 1.

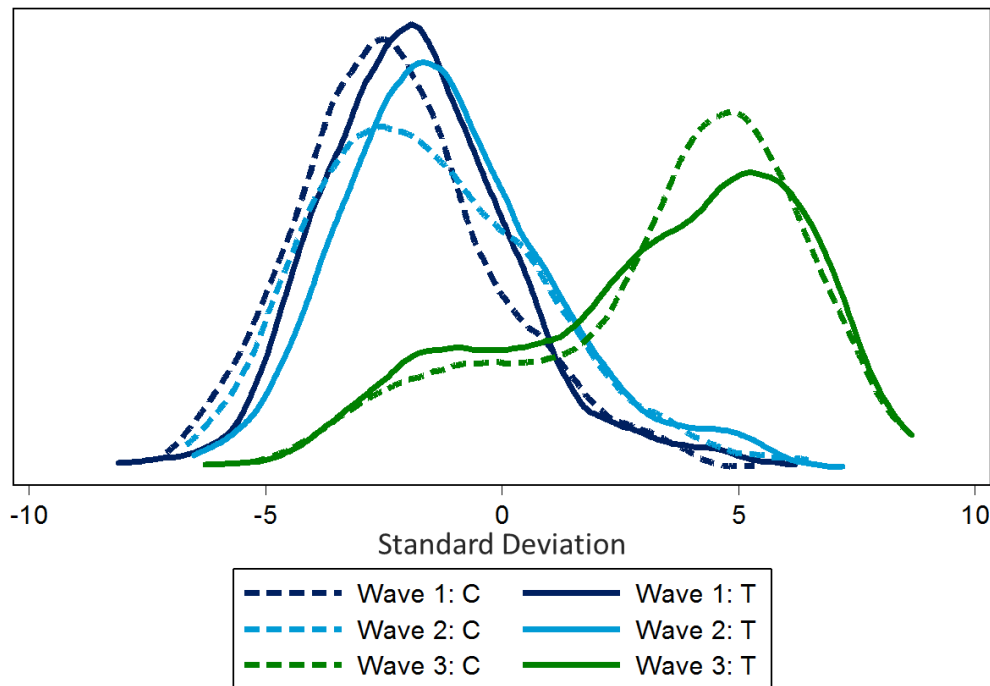


Figure 12: Distribution of the Index Scores for Each Data Collection Wave

d) **Results: Regression Results**

The descriptive statistics in the previous section indicate that the largest difference between the intervention and comparison group was seen in wave 1 (after 6 months of teacher training). This difference in performance diminished by wave 2 (3 months after block training ended), and ceased to exist by wave 3 (a year after the learners were taught by teachers who were trained in the program). This section further explores these results to establish whether the results hold after controlling for learner-level factors.

As noted earlier in this report, the impact evaluation's identification strategy rests on the assumption that the 2015 and 2016 cohorts of Grade R learners are inherently comparable. Any changes between the cohorts that could be confounded with the roll-out of the ELIT program would therefore have had to originate from external factors that changed between 2015 and 2016 on a provincial-, school- or teacher-wide level. Given that the program was rolled out universally across the province, provincial level effects could not be controlled for statistically. However, it is understood that no other system-wide language or literacy programs were rolled out over the period of the evaluation.

Given that the intervention and comparison group learners were in the same schools, school level factors were controlled for inherently. A separate model was run only on the schools where the teachers remained constant across the two years. In this model, inherent teacher level factors were kept constant. Table 3 below shows the mean results for each wave of data collection.

Models A1 and A2 present the results in wave 1; B1 and B2 the results in wave 2; and C1 the results in wave 3. The first model (A1, B1, C1) controls for the school quintile and learner characteristics such as learner gender and age. The second model (A2 and B2) was run only for the schools in which the teacher remained the same in Grade R, thereby keeping the teacher characteristics constant.

Table 3: Main Results Models

	Wave 1		Wave 2		Wave 3
	A1	A2	B1	B2	C1
Intervention	0.405*** (0.099)	0.397*** (0.105)	0.243** (0.095)	0.287*** (0.093)	-0.087 (0.080)
Learner Age	0.168** (0.077)	0.181** (0.080)	0.148** (0.070)	0.175** (0.076)	0.129 (0.084)
Learner = Female	0.212*** (0.059)	0.185*** (0.058)	0.211*** (0.062)	0.192*** (0.059)	0.388*** (0.074)
N	937	823	912	800	778
R-squared	0.125	0.104	0.150	0.135	0.138
Same Teacher	x		x		

*0.1; **0.05; *** 0.01

Reflecting the findings in the descriptive statistics, the difference in performance between the intervention and comparison groups in wave 1 was positive, large, and significant at 0.405 standard deviations. Controlling for classes in which the teacher remained the same in 2015 and 2016 did not result in large changes to the intervention effect. In wave 1 (mid-Grade R) there was already a significant difference between the performance of girls and boys with girls outperforming boys (0.212 standard deviations difference). Older learners (6 years relative to 5 years) also performed significantly better (0.168 standard deviations difference).

Models B1 and B2 present the results in wave 2 (end- Grade R). The first model (B1) only controls for quintile and learner characteristics, whereas the second model (B2) controls for schools in which the Grade R teacher remained the same in 2015 and 2016. In wave 2, the intervention effect was slightly smaller than in wave 1 but it remained positive and significant at 0.243 standard deviations. Once again, controlling for schools in which the teacher remained the same did not change the coefficient. The significant differences seen in wave 1, based on learner age and gender, persisted in wave 2 (0.148 and 0.211 standard deviations difference, respectively).

The final model (C1) presents the results in wave 3 (end-Grade 1), controlling for school quintile and learner characteristics. There was a large dispersion of learners across a range of Grade 1 classes both within the sample schools and in new schools. This led to insufficient overlap of comparison and intervention learners in the same Grade 1 teachers' class to be able to control for schools in which the Grade 1 teacher remained the same. At the end of Grade 1, there were no statistically significant differences in learner scores between the intervention and comparison groups. The significant difference between the performance of girls and boys in Grade R persisted into Grade 1 and increased (0.388 standard deviations difference).

e) **Results: Grade 1 Pacing and Curriculum Coverage**

Part of the evaluation assignment was to assess the extent to which teachers covered the curriculum with learners in Grade 1. The hypothesis underlying this assessment is that Grade 1 teachers would be able to pace through the curriculum faster, and therefore cover more of the curriculum, with learners who were exposed to the ELIT intervention (intervention group) than comparison group learners.

The tool assessed learners' work in the form of learner class workbooks, DBE workbooks, class worksheets, homework books, other books (teacher reader, storybook), and additional exercises.

Fifteen schools were randomly selected for the assessment from the pool of 49 sample schools. Four learners per school were randomly selected from a list of learners that were evaluated in Grade R (the same learners were tracked from Grade R into Grade 1). Data was collected at the end of Grade 1 in 2016 for the comparison group and in 2017 for the intervention group⁴⁶.

The table below shows the quintile, location, and LoLT of the school and whether the teachers remained the same in 2016 and 2017. In 12 of the 15 schools, the teachers remained constant between the two years.

School	Quintile	Location	LoLT	Same Teachers
School 1	1	Rural	Afrikaans	Yes
School 2	1	Rural	Afrikaans	No
School 3	2	Rural	Afrikaans	Yes
School 4	2	Urban	isiXhosa	Yes
School 5	2	Urban	isiXhosa	Yes
School 6	2	Urban	Afrikaans	Yes
School 7	2	Rural	Afrikaans	Yes
School 8	3	Rural	Afrikaans	No
School 9	3	Urban	isiXhosa	Yes
School 10	4	Urban	Afrikaans	Yes
School 11	5	Urban	Afrikaans	Yes
School 12	5	Urban	English	Yes
School 13	5	Rural	English	Yes
School 14	5	Urban	Afrikaans	Yes
School 15	5	Urban	Afrikaans	No

The period of assessment of learner work stretched from the beginning of Term 3 up to the end of the second week of term 4 of the 2016 and 2017 academic years. Following the curriculum, Grade 1 learners only began complex and creative writing in the third term. The amount of work to be assessed also needed to be kept to a reasonable limit to allow each fieldworker half a day to assess the materials.

There were no significant differences in pacing and curriculum coverage from 2016 to 2017 overall. In other words, Grade 1 teachers neither increased the amount of writing nor worked with

⁴⁶ The full Pacing and Curriculum Coverage results are available in Annex 1, Figures 34 to 47.

learners to improve the quality of writing, creativity of writing or length of learner writing. There were two exceptions: there were significant differences in the total number of blank pages in “other materials”, (where more blank pages were evident in 2017) and the quality of teacher feedback in the DBE workbooks (2017 saw a decline in the quality of teacher feedback to learners with more ticks and marks and less evidence of corrective feedback).

Across the years, the assessment found that the DBE workbooks were not the primary learner resource in the majority of Grade 1 classrooms. Creative writing was seen more often in “other materials” than in the DBE workbooks. There was also more evidence of longer sentences in “other materials” than in the DBE workbooks. In terms of the quality of learners’ written work overall, across both years, learners tended to write simple words or shorter and simpler sentences with punctuation although there were at least some examples of longer and more complex sentences.

Three examples demonstrate the findings from the Grade 1 **Pacing and Curriculum Coverage Assessment**. These establish how much work the learners were covering, the quality of feedback from teachers, and the quality and complexity of learners’ work.

In the **best case: School 11** is based in an urban area, is a quintile 5 school, and uses Afrikaans as the LoLT. Their Grade 1 learners wrote more than 50 pages on average during Term 3 and 4, the teachers used mainly corrective feedback (far more than just ticks and marks), and the learners wrote longer and more complex sentences. Finally, learners wrote more than the requisite number of sentences.

In the **medium case: School 5** is a quintile 2 school, based in an urban area, and uses isiXhosa as the LoLT. Learners wrote on average approximately 40 pages each during Term 3 and 4. Teachers mainly provided feedback in the form of ticks and marks with motivational comments. While learners wrote approximately 40 pages, many of the pages contained simple words in 2016. In 2017, the learners wrote at least 2 complete sentences. In 2016, the learners’ writing either was off topic or did not relate to the story, which improved in 2017 to more linked writing.

In the **worst case**: This rural, Afrikaans LoLT, quintile 3 school (**School 8**) deteriorated between 2016 and 2017. At least one teacher left the school. In 2016, the learners completed approximately 30 to 40 pages during Term 3 and 4, but this halved in 2017 to an average of 20 pages. Likewise, teacher feedback deteriorated from motivational feedback in 2016 to minimal ticks and marks in 2017. The learners in both years wrote only simple sentences but, in 2017, their sentences were more often off topic.

3. Robustness Checks

Learner results could have been influenced by three different factors: (1) ceiling effects in the learner assessment in Grade 1, (2) a bias introduced through the administration of the assessment by the fieldworkers, or (3) a bias introduced through learner attrition. In this section, sensitivity checks were conducted to establish whether the abovementioned factors could have had an impact on the overall intervention effect.

The descriptive statistics demonstrated ceiling effects in some of the sub-tasks, particularly in the Grade 1 assessments, which could have masked any differences between the intervention and comparison groups. To evaluate the intervention effect on the sub-tasks of the Learner Assessment, the model was run on each sub-task in the wave 1 and wave 3 assessments. Figure

8 in Annex 1 shows the results. In the *listening and speaking in a conversation*, *listening and speaking based on a story*, *phonological awareness*, (emergent) *writing*, and (emergent) *reading* sub-tasks, there were positive and statistically significant differences between the intervention and comparison groups in wave 1. However, no statistically significant differences were apparent on these or any other Learner Assessment items in wave 3. In the (emergent) *writing* sub-task, the intervention learners out-performed the comparison group in wave 1. However, in wave 3, the opposite was evident with the intervention group learners showing fewer statistically significant learning gains than the comparison group. Likewise, on the *word recognition* and *paragraph reading* sub-tasks of the EGRA, which were included in the wave 3 assessment, the comparison group performed significantly better than the intervention group; recognizing approximately 2.7 words more on average per minute and reading approximately 4 words more on average per minute in a paragraph than the intervention group.

A second form of bias could have been introduced through the fieldworkers. Mostly the same fieldworkers conducted the six rounds of data collection (three waves for each of the intervention and comparison groups) although no specific effort was made to have the same fieldworker, or a different fieldworker, test the same learner in each wave. To assess whether there was any bias in the intervention effect because of the fieldworkers, the model was run for each round of data collection (Figure 22 in Annex 1). Models A1, B1 and C1 show the results for each wave of data collection, whereas models A2, B2, and C2 show the results controlling for any fieldworker effects. It is clear that the fieldworkers did have an effect on learner outcomes (changing the intervention coefficient from 0.405 to 0.591 standard deviations difference in wave 1; 0.243 to -0.091 standard deviations difference in wave 2; and -0.08 to 0.069 standard deviations difference in wave 3). The fieldworker effect was therefore large enough to affect the results in wave 2 but not large enough to contradict the results seen in wave 3.

The attrition rate within the comparison group of learners was slightly higher than for the intervention group of learners, specifically in the third wave of data collection. Reasons for attrition include learners having been held back in Grade R for an additional year, learners having moved schools⁴⁷⁴⁸, or learners having been absent on the day of assessment⁴⁹. Descriptively, attrition did not seem to have a large impact on the overall composition of the sample. However, when considering the predictors associated with attrition – that it is often the younger learners and weaker performing learners that drop off the sample – it was acknowledged that the intervention group could be a slightly weaker group than the comparison group⁵⁰. To control for this bias, a probit model was run to estimate the propensity of each learner in the sample of dropping out of the sample. Using this propensity, inverse probability weights were constructed to re-estimate the model for the wave 3 scores. The weights had an impact on the gender coefficient, but had no significant effect on the intervention coefficient (Figure 23, Annex 1).

⁴⁷ The learners moved to different Grade 1 classes with different teachers.

⁴⁸ These learners could either not be found through the WCED Centralised Educational Management Information System (CEMIS) or they had moved to schools in other districts or provinces.

⁴⁹ The evaluation team made every effort to reassess absent learners in the same week of data collection.

⁵⁰ If the weaker performing learners were more likely to have been part of the intervention group but not the comparison group.

4. Checks for Heterogeneous Effects⁵¹

The ELIT intervention may have influenced various sub-groups differently. Various models were run to investigate whether there was any heterogeneity in schools with a different quintile status, or with a different LoLT, or whether the intervention may have influenced girls and boys differently. Figure 24 in Annex 1 shows the interaction effect of the intervention on girls at each wave of data collection. The coefficient on the interaction effect is positive but not significant, showing that the intervention did not specifically influence girls differently to boys. Figure 25 in Annex 1 shows the effect of the intervention on schools with differing LoLTs compared to schools with an English LoLT. In wave 1, the effect of the intervention on isiXhosa LoLT schools was positive and significant, suggesting that there may have been a positive effect of the intervention in these schools. However, this result was not seen in waves 2 and 3. Overall, there were no consistent interaction effects of the intervention on different LoLTs, showing that the intervention did not have a different influence on different language groups.

Finally, a model was run to evaluate whether schools in quintiles 1 – 3 were influenced differently than schools in quintiles 4 and 5 (Figure 26, Annex 1). In wave 1, the effect of the intervention on quintile 1 - 3 schools was negative and significant, suggesting that the intervention may have had a larger positive effect in quintile 4 – 5 schools. However, there were no interaction effects of the intervention on quintile 1 – 3 schools, showing that the intervention did not have a different influence on different quintile schools.

5. Summary of Findings

Learner performance in the difference-in-differences model, average learner performance on the Learner Assessment and EGRA, and learner performance using a combined index score, taking robustness checks and checks for heterogeneous effects into account, all tell the same story. The ELIT intervention, incorporating 8 Stellar story modules and an adapted Letters and Sounds Manual, appears to have initially increased the language and literacy achievement of targeted learners over the first half of the Grade R year, but this effect was not sustained through to Grade 1. Assuming these results are not caused by extraneous or intrinsic unobservable factors, the results suggest that the ELIT instructional method may be more effective with Grade R learners, giving them a head start, but has little long lasting performance advantage if not sustained.

⁵¹ Heterogeneous effects refer to a case in which an intervention has different effects for different sub-populations based on characteristics such as gender, age, language, etc.

Sub-question 1.1: To what extent have Grade R interventions impacted the language and literacy acquisition and performance of Grade R and Grade 1 learners in comparison to the control group and expected outcomes for these grades?



Most learners in both the intervention and comparison groups reached the Grade R and Grade 1 criterion levels on the *listening and speaking based on a story*, *listening and speaking in a conversation*, and emergent *writing* subtasks of the Learner Assessment.

However, in both the comparison and intervention groups, most learners did not reach the expected grade levels (set in CAPS) for *phonological awareness*, *phonics letters*, or emergent *reading* in Grade R. Performance on the emergent *writing* subtask differed between the intervention and comparison groups with intervention group learners performing better and reaching the grade level on most subtask items.

By the end of Grade 1, many learners in both the intervention and comparison groups met the expected criterion levels on all subtasks of the Learner Assessment.

On the EGRA items, learners in the intervention and comparison groups only reached the WCED benchmarks for *letter sound recognition*. *Word recognition* benchmarks were only met in the Afrikaans comparison group and the isiXhosa intervention group. None of the language groups reached the expected levels on *paragraph reading* or *comprehension* based on a paragraph.

Introduction

Two evaluation tools provided an indication of learner performance against expected Grade levels:

- The Learner Assessment provided an indication of Grade R and Grade 1 performance on *listening and speaking in a conversation*, *listening and speaking based on a story*, (emergent) *reading*, (emergent) *writing*, *phonological awareness* and *phonics letters* against a 6-point scale: not done (0), partial Grade R level (1), end of Grade R level (2), partial Grade 1 level (3), end of Grade 1 level (4), Grade 2 level (5). These “benchmarks” were mapped to the CAPS criteria.
- The EGRA provided an indication of Grade 1 learner performance on *letter sound recognition*, *word recognition*, *paragraph reading*, and reading *comprehension* based on a paragraph, against benchmarks set by the WCED.

The section below is structured as follows:

1. Results: Learner Assessment Grade Levels
2. Results: Early Grade Reading Assessment Benchmarks
3. Summary of Findings

1. Results: Learner Assessment Grade Levels



Listening and Speaking in a Conversation

Being able to answer questions is the basis of both assessment and education. This factor represents the “Speaking” part of “Listening and Speaking” in the CAPS curriculum. Apart from applying conversation skills, dimensions of self-confidence and awareness of other people are involved. Stellar emphasizes building self-confidence with constant praise and encouraging learners to appreciate the efforts of others. Awareness of others is built into group exercises.

The percentage of learners that “reached” or “exceeded” the Grade level (i.e. were competent) on the *listening and speaking in a conversation* items was high in both the intervention and comparison groups across all three waves of data collection. More than 60% of comparison group learners reached this level on 5 of the 6 test items in wave 1, all 6 items in wave 2, and 4 items in Grade 1. Similarly, more than 60% of intervention group learners reached this level on all 6 items in wave 1, 5 in wave 2 and 5 items in wave 3⁵².

Small improvements in answering closed questions and recognizing others were seen in wave 2 in both groups, and these improvements were sustained into Grade 1 (wave 3). In wave 3, 43% of comparison group learners, and 40% of intervention group learners, were able to provide enough detail for a conversation – a Grade 2 skill.



Listening and Speaking based on a Story

This factor represents the “Listening” part of “Listening and Speaking” in the CAPS curriculum and taps into vocabulary, concept formation, and the relationships between spoken words, pictures, and actual life experiences.

The percentage of learners that were competent on the *listening and speaking based on a story* items was high in both the intervention and comparison groups across all three waves of data collection. More than 60% of comparison group learners reached this level on 3 of the 6 test items in wave 1, 5 of 6 items in wave 2, and 4 items in Grade 1. More than 60% of intervention group learners reached this level on all 6 items in wave 1⁵³, 5 in wave 2 and 5 items in wave 3.

Although more than 60% of learners in the comparison and intervention groups reached or exceeded the Grade level at the end of Grade R in terms of “*recalling part of a story / discussing story characters*”, only a fifth of the learners reached this level at the end of Grade 1. This finding is mirrored in learner performance on “remembering the sequence of the story” where more learners were able to tell three or more parts of the story in the correct sequence at the end of Grade R than were able to in Grade 1. These results may indicate that the Grade 1 story – *Joshua’s Goal* - was difficult for learners to follow and remember. The remainder of the learners were able to recall fewer parts of the story and/or not in the correct sequence. The Grade R

⁵² Early on, the evaluation team recognized the ceiling effects in the data and took measures to add items to the Grade 1 assessment to mitigate these effects.

⁵³ As noted in previous reports, the story used for this assessment may have been introduced to the intervention group learners before it was meant to be rolled out in the intervention classes.

intervention group was notably better than the comparison group in drawing a picture of a story in both rounds of data collection (wave 1 and wave 2).



Phonological Awareness

Phonological and phonemic awareness are included under “emergent reading” in CAPS. The Stellar Letters and Sounds Manual (teaching letters and sounds) states that without an understanding of the relationship between sounds and letters, reading, and writing is very difficult for children.

Stellar introduces the following sounds and letters in English, isiXhosa and Afrikaans stories and games:

	ENGLISH	ISIXHOSA	AFRIKAANS
1 The green dress	s	z	s
2 Ali and the paint	b	p	b
3 Spot goes to the shop	m	s	m
4 Baby Bird finds his mother	n	n	n
5 Tortoise and his shell	t	q	t
6 Goldilocks and the three bears	h	k	h
7 The three goats	g	b	g
8 Bushbuck	l	m	l
Additional letters covered in games	r, c, f, d, p	t, c, d, r, l, f, h	r, d, k, v, p

Games repeat all of the letters covered in the stories except the letters in red.

The low numbers of intervention and comparison group learners competent in Grade R *phonological awareness* items suggest that phonological and phonemic awareness need to be carefully taught. Even after prompting, in both the comparison and intervention Grade R groups more than 60% of the learners scored a “0” (i.e. not done or unable) on 2 of the 3 Grade R items in the subtask. Items that learners struggled with the most included joining *sounds to make up words* and identifying an *odd word in sequence*. Learners in both the intervention and comparison groups were able to distinguish *different letter sounds*.

Knowledge of sounds improved in Grade 1. Nearly all (86%) of comparison group learners and 87% of intervention group learners could join *sounds to make up words* (a Grade R task). Two-thirds (63%) of comparison group learners and 60% of intervention group learners could *substitute vowels* in words (a Grade 1 task – primarily CVC words in English and Afrikaans and CVCV words in isiXhosa). However, the ability to identify an *odd word in sequence* (a Grade R task) was problematic with only 48% of comparison group learners and 54% of intervention group learners

reaching competency. Similarly, only 25% of comparison group learners and 28% of intervention group learners were competent in *building words using sounds* (a Grade 1 task).

The general pattern appears to be that little work was done on phonological awareness in either the comparison or intervention group, either up to August or in the second half of Grade R. However, most learners in the intervention and comparison groups were able to distinguish *different letter sounds* across all three waves of data collection.



Phonics Letters

This factor concerns the translation of sounds into letters of the alphabet. Stellar deals only with lower case letters and does not teach alphabet names. Items were based on a letter card worksheet from the Stellar material presenting individual letters, and a similar worksheet with blends of letters and common combinations of letters.

Once again, the results suggest that *phonics letters* have to be carefully taught. Only the basic Grade R item, being able to recognize that *sentences are made up from words*, by clapping on each word, showed an improvement in the intervention group compared with the comparison group in wave 1 (mid-Grade R). Even after prompting, at the end of Grade R in both the comparison and intervention groups, more than 60%⁵⁴ of the learners scored a “0” (i.e. not done or unable) or “1” (mid-Grade R level) on 2 of the 3 Grade R items. Items that learners struggled with the most included *recognizing some consonants* and *pointing to the letter starting a word*.

In Grade 1, the number of learners reaching expected competency levels increased across the board. In both the intervention and comparison groups, more than 60% of learners reached the expected competency levels on 4 of the 6 items. However, learners for the most part remained unable to *recognize two consonants at the start and end of a word* (a Grade 1 task). Again, learners were for the most part unable to reach competency on the Grade 2 item measuring “extension” performance – *recognizing double vowels and 3 consonant strings*.



Reading

Reading is not taught in Grade R, but the process of emergent reading is introduced and learners are exposed to reading activities and books.

In waves 1, 2, and 3 nearly all learners in both groups could *correctly handle a book* (a Grade R item). Similarly, more than 50% of learners in the intervention and comparison groups were able to *predict a storyline from the cover* (a Grade 1 item).

However, in waves 1 and 2, in both the comparison and intervention Grade R groups, more than 60% of the learners scored a “0” (i.e. not done or unable) on 2 of the 3 Grade R items in the subtask. Items that learners struggled with the most included using a *reading voice* and *pointing to words while reading*.

Reading was a major focus in Grade 1. By the end of the year, 84% of the comparison group and 82% of the intervention group were competent in using a *reading voice*. Most (70%) of the comparison group and 69% of the intervention groups *pointed to words while reading*. Three-

⁵⁴ With the exception of the intervention group, wave 2, in which 59% of learners were unable to attempt the task.

quarters (75%) of comparison group learners and 71% of intervention group learners could *recognize sight words*. Finally, over 60% of learners in both groups could *read words correctly* (a Grade 2 task) – showing “extension” performance. This is in contradiction to the EGRA results. However, the Learner Assessment task is not timed and is therefore not identical to the EGRA.



Writing

At the Grade R level, this subtask is more concerned with emergent writing than formal writing. Item responses were rated according to an Evolution of Writing table provided to teachers in training and simplified by Wordworks to 6 steps for the purposes of this evaluation.

Stellar concentrates heavily on writing skills and the formation of letters. Every week, as a letter is introduced, learners reproduce the letter with finger-painting shapes, writing in sand, writing in the air, writing on walls in chalk, play dough, making body-letter shapes and movements. Word ends and chants are taught to learn the direction of letter-writing and templates are provided for coloring in “rainbow” letters. Learners are praised for making letter shapes correctly in their own writing and for inventive spelling. Teachers also model writing by asking for explanations, labelling items in the classroom and showing “grown-up” spelling of what learners have written.

Grade R learners in both the intervention and comparison groups were able to *write their own name* across wave 1 and wave 2. More intervention learners than comparison learners were able to *read their own writing* and more intervention learners were able to *copy a pattern* than their comparison counterparts. However, learners in both groups struggled with “*reading*” their own “*writing*” against expected levels.

In Grade 1, approximately half of the intervention and comparison group learners were able to write something about a story that they had drawn. More than 70% of Grade 1 learners in both the intervention and comparison groups were able to copy a sentence.

Very few learners were able to write a sentence correctly. This is a Grade 2 task, and so not expected, but the numbers show that learners are mostly not doing extension work in writing in Grade 1.

2. Results: Early Grade Reading Assessment Benchmarks

On the EGRA subtest items, neither the comparison nor the intervention groups performed within range of the language benchmarks set by the WCED for Grade 1 on *word recognition* (except for the Afrikaans comparison group and isiXhosa intervention group), or *paragraph reading*. However, both the intervention and comparison groups met the WCED benchmarks for *letter sound recognition*. Reading comprehension was particularly poor, with only learners at the 75th percentile managing to answer one question correctly.

3. Summary of Findings

On the Learner Assessment instrument, most learners in both the intervention and comparison groups reached the criterion levels set for Grade R and Grade 1 on the *listening and speaking based on a story* and *listening and speaking in a conversation* subtests (i.e. more than 60% of learners met or exceeded the expected score on at least 5 out of 6 items in the subtasks),

However, in both the comparison and intervention groups, most learners did not reach the expected grade levels for *phonological awareness*, *phonics letters*, or (emergent) *reading* in Grade R (i.e. on at least 2 out of 3 Grade R items in the subtasks). Performance on the (emergent) *writing* subtask differed between the intervention and comparison groups. In Grade R, more than 60% of the intervention group learners met the expected grade levels on 2 of 3 Grade R items while the comparison group only met the expected grade level on 1 item.

By Grade 1, many learners in both the intervention and comparison groups achieved the expected criterion levels on all subtasks of the Learner Assessment (i.e. more than 60% of learners met or exceeded the expected criterion on at least 4 out of 6 items).

On the EGRA, when measured against the WCED benchmarks, learners in the intervention and comparison groups did not perform well overall. Both groups only reached the WCED benchmarks for *letter sound recognition*. *Word recognition* benchmarks were met in the Afrikaans comparison group and the isiXhosa intervention group. Otherwise, none of the language groups reached the expected levels on *word recognition*, *paragraph reading* (reading words in a paragraph), or *comprehension* based on the paragraph.

The findings show that even though learners in the intervention and comparison groups performed well on the Learner Assessment at the end of Grade 1, learners in both groups are not reaching the Grade 1 WCED benchmarks for *paragraph reading* or *word recognition* on the EGRA. The higher Learner Assessment scores could potentially reflect ceiling effects in the tool.



Photo: Khulisa

Evaluation Question 2



To what extent has the train-the-trainer model resulted in more effective and knowledgeable teachers and positive language and literacy outcomes for learners?



The cascade training approach appears to have had an **initial effect** on literacy outcomes for learners. The teacher Knowledge Attitudes and Practices test shows an **overall improvement** from pre-training to post-training. There were no statistically significant differences from the end of training to the end of the year. However, the fact that the effect on learner language and literacy achievement **decreased** over this period suggests that formal or informal support structures may be necessary to sustain newly acquired knowledge and instructional approaches (Kim, Boyle, Zuilkowski, & Nakamura, 2016)

Introduction

The evaluation team employed a teacher Knowledge, Attitudes and Practices (KAP) survey for this evaluation question and analyzed Wordworks' Lead Teacher Knowledge Questionnaire results for comparison purposes. In Section 1 of the KAP, teachers were required to indicate whether they agreed or disagreed with each of the statements in the section, which measured common understanding and knowledge of key emergent literacy concepts. In Section 2, teachers were required to match a stated learner competency with its area of language and literacy (phonological awareness, concepts about print, letter knowledge, or literate language). In Section 3, teachers were asked to indicate the current practice in their classroom and indicate their preferred classroom context.

Wordworks collected data at three points from Lead Teachers: pre- training, post- training and after cluster sessions. The evaluation team similarly collected data at three points from Grade R teachers: before training (in March 2016), after cluster and block training (in August/September 2016), and after the delivery of the intervention to learners (in November 2016).

The section below is structured as follows:

1. Results: Lead Teacher Knowledge Questionnaire
2. Results: Teacher Knowledge, Attitudes and Practices Survey
3. Summary of Findings

1. Results: Lead Teacher Knowledge Questionnaire

To examine the effects of the ELIT training on trainers, Khulisa re-analyzed the data gathered by Wordworks from the Lead Teachers through their ELIT Knowledge Questionnaire. The analysis was done in the same manner as for the KAP survey administered by the evaluation team.

For Lead Teachers (Table 4), statistically significant improvements were found from pre-training to post-training on all literacy knowledge questions. From the end of block training to the end of Lead Teacher cluster training 7 of the 11 comparable questions answered by Lead Teachers improved even further although, overall, the difference between performance at the end of block training to the end of cluster training was not statistically significant.

Lead Teacher performance on two questions stood out in stark contrast to teacher performance (with differences of 77-80 percent points):

- Question 13 “to encourage a child’s language development the teacher should use simple words to encourage a Grade R learner’s language development”: Only 18.4% of teachers surveyed by Khulisa had retained the information three months after the training, compared to 94.9% of Lead Teachers.
- Question 14 “letters have both names and sounds. It is most important for emergent readers to know the names of letters”: Only 12.2% of teachers answered this correctly three months after the training, compared to 91.9% of Lead Teachers.

When looking at the data more closely, however, these were the two items where Lead Teachers answered correctly in all three instances of taking the test (before the training, immediately after the training and three months later).

Table 4: Lead Teacher Performance on Section 1 of the ELIT Knowledge Questionnaire

Questions in Section 1	Pre-Training	Post-Training *** p value ≤ 0.05	End of Clusters
1. Songs, rhythm and rhyme are fun but not very important for literacy and language learning.	56.5%	62.8%	67.0%
2. Fantasy play helps learners develop advanced forms of language.	45.3%	50.5%	56.6%
3. Children who enter school with limited vocabulary and poor oral language skills will struggle with reading and writing.	32.3%	34.7%	28.6%
6. If children are learning in their second language at school, they should be discouraged from using their mother tongue in the playground and at home.	18.9%	29.8%	31.0%

Questions in Section 1	Pre- Training	Post- Training *** p value ≤ 0.05	End of Clusters
7. Young children should be encouraged to experiment with writing before Grade One.	38.6%	40.0%	54.5%
8. Children should be discouraged from writing words until they can spell them properly.	60.9%	69.5%	70.0%
9. The best way to learn how to form letters in Grade R is using big movements and multisensory activities.	89.8%	93.5%	93.9%
11. An example of a syllable blending question is: Which word would you have if you put these parts of words together: 'to-ma-to'?	84.4%	91.4%	90.9%
13. To encourage a child's language development the teacher should use simple words to encourage a Grade R learner's language development.	90.6%	96.8%	94.9%
14. Letters have both names and sounds. It is most important for emergent readers to know the names of letters.	93.0%	94.6%	91.9%
15. If a Grade R child brings you something he has 'written' with some letters facing the wrong way, it is more important to praise him for trying, than to correct his letter formation.	83.6%	89.2%	90.9%

In Section 2, only five questions were comparable between the Wordworks Knowledge Questionnaire and the KAP survey, as shown in Table 5 below.

Table 5: Lead Teacher Performance on Section 2 of ELIT Knowledge Questionnaire

Questions in Section 2	Pre-Training	Post-Training	End of Clusters
1. Knows the meaning of unusual words.	59.5%	63.5%	56.7%
2. Identifies all the letters of the alphabet.	74.8%	77.1%	77.3%
4. Identifies the first sound in a spoken word.	70.2%	69.8%	73.2%
7. Uses crayons and pencils with ease.	97.7%	97.8%	97.9%
8. Notices detail in pictures.	97.7%	98.9%	95.9%

Overall, more Lead Teachers responded correctly after the training than they did at the beginning of the year. These differences were statistically significant.

2. Results: Teacher Knowledge Attitudes and Practices Survey

The evaluation team administered the KAP survey at three points with teachers in the sample: before training (March 2016); after block training (August/September 2016) and near the end of the Grade R year (November 2016). Statistically significant improvements were found from pre-training (March) to post-training (August/September). While four questions showed further improvements from the end of block training (August/September) to the end of the year (questions 2, 3, 10 and 11) there was no statistically significant difference overall (Table 6).

The fact that learner language and literacy gains declined over the same period suggests that teachers may not have sustained their new knowledge and practices in the classroom. The fact that ELIT literacy-related activities were observed in less than 50% of classrooms by the end of the year (see evaluation question 2.1 below for more details) indicates that this may be the case.

Table 6: Percentage of Teachers with Correct Response on KAP Section 1

Questions in Section 1	Pre-Training	Post-Training *** p value ≤ 0.05	End of Grade R
1. Songs, rhythm and rhyme are fun but not very important for literacy and language learning.	51.3%	66.0%	59.2%
2. Fantasy play helps learners develop advanced forms of language.	53.8%	75.0%	77.6%

Questions in Section 1	Pre- Training	Post- Training	End of Grade R
		*** p value ≤ 0.05	
3. Children who enter school with limited vocabulary and poor oral language skills will struggle with reading and writing.	25.6%	29.2%	34.7%
4. Children only begin to learn to read and write in Grade One.	28.2%	34.0%	24.5%
5. Parents should not be involved in developing their children's literacy and language as they are not trained to be teachers.	61.5%	64.6%	61.2%
6. If children are learning in their second language at school, they should be discouraged from using their mother tongue in the playground and at home.	23.1%	27.1%	27.1%
7. Young children should be encouraged to experiment with writing before Grade One.	48.7%	64.6%	63.3%
8. Children should be discouraged from writing words until they can spell them properly.	28.2%	44.7%	40.8%
9. The best way to learn how to form letters in Grade R is using big movements and multisensory activities.	51.3%	66.7%	65.3%
10. Working with sequence pictures develops children's phonemic awareness.	0.0%	0.0%	6.1%
11. An example of a syllable blending question is: Which word would you have if you put these parts of words together: 'to-ma-to'?	35.9%	42.6%	53.1%
12. Role play is important for language development because children have the opportunity to use the language of a story and make it their own.	74.4%	75.0%	63.3%
13. To encourage a child's language development the teacher should use simple words to encourage a Grade R learner's language development.	0.0%	19.1%	18.4%

Questions in Section 1	Pre- Training	Post- Training *** p value ≤ 0.05	End of Grade R
14. Letters have both names and sounds. It is most important for emergent readers to know the names of letters.	10.3%	15.2%	12.2%
15. If a Grade R child brings you something he has 'written' with some letters facing the wrong way, it is more important to praise him for trying, than to correct his letter formation.	43.6%	58.3%	53.1%

In Section 2 of the KAP survey, teachers were presented with 10 learner competencies, and asked to match each of these competencies with the relevant area of language and literacy (first five questions) or the underlying skill (second five questions). There was only one possible correct response per question based on Wordworks' model answers. Table 7 shows that on 8 of the 10 questions more teachers responded correctly in March 2016, before the teacher cluster sessions and block training, than they did at the end of the year. Overall, there were no statistically significant differences from the end of training to the end of the Grade R year and so this decline is not particularly meaningful.

Table 7: Percentage of Teachers with Correct Response on KAP Section 2

Questions in Section 2	Pre- Training	Post- Training	End of Grade R
1. Knows the meaning of unusual words.	61.1%	54.5%	55.1%
2. Identifies all the letters of the alphabet.	66.7%	60.9%	77.1%
3. Moves their finger from left to right across the page of a book when pretend reading.	66.7%	60.0%	57.1%
4. Identifies the first sound in a spoken word.	68.4%	73.3%	59.2%
5. Able to clap out the syllables in a word.	64.1%	53.5%	58.3%
6. Follows verbal instructions.	89.2%	93.6%	83.7%
7. Uses crayons and pencils with ease.	97.4%	89.6%	87.8%

Questions in Section 2	Pre-Training	Post-Training	End of Grade R
8. Notices detail in pictures.	100.0%	97.9%	100.0%
9. Can recognize letters in the environment.	97.4%	95.6%	95.9%
10. Can distinguish between different sounds at the start of words.	89.7%	86.7%	87.8%

In Section 3, the teachers were required to, for each statement, tick a box that best described the current practice in their classroom and tick a box that indicated their preferred classroom context. The response options were scored as follows:

Existing classroom context:	Preferred classroom context:
3 = Often/regularly 2 = Sometimes/occasionally 1 = Seldom/rarely 0 = Never	3 = No change required 2 = Improvement needed in medium term 1 = Urgent and immediate improvement needed 0 = Don't know

The statements were based on Wordworks' analysis of expected shifts in teacher performance because of the ELIT intervention. If, for example, on the item, "*Storytelling is used to help learners to become familiar with the language of books*", a teacher reported a score of 1 for the existing classroom context, it would be interpreted as a poor practice. In contrast, if the teacher reported a 2 for the preferred classroom context on this item, it would mean that they desire some change (i.e. good attitude), but can wait for it in the medium term.

For existing classroom context, there was no change from before training, to after training, to the end of the school year. Most teachers across all quintiles indicated that they "*often or regularly*" carried out the desired literacy-related practices in their classrooms, both before training, after training, and again at the end of the school year. This could reflect a form of self-presentation bias or it could be attributed to the Dunning Kruger effect⁵⁵.

The Classroom Environment assessment, for example, shows that prior to training, role-play was not observed in any of the classrooms in the sample while, after training, role-play was only observed in 25% of classrooms. This is inconsistent with teacher's rating of "*practice often/regularly*" in relation to the statement, "*learners practice listening and speaking (conversation) skills through role playing and dramatization of stories*". Similarly, in terms of picture card sequencing, this activity was only seen in 10% of classrooms prior to training, and

⁵⁵ The Dunning Kruger effect occurs when people / entities fail to adequately assess their own level of competence, resulting in an over-representation of their skills, abilities, or practices.

42% of classrooms after training. This is also inconsistent with teachers' self-rating in relation to the statement, *"learners are taught to organize their 'thinking' through sequencing of picture cards."*

Regarding preferred classroom context, the results again indicate no substantial change from before to after training, to the end of the school year, with slightly more teachers in quintile 1 and 2 schools indicating that *"improvement is needed in the medium term"*.

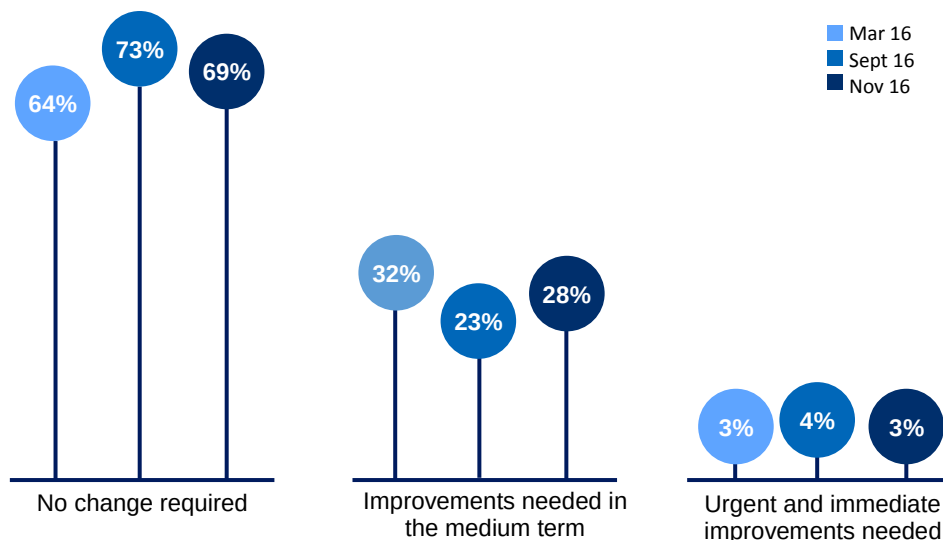


Figure 13: Teacher Response - Preferred Classroom Context

Figure 13 presents a breakdown of the percentage of teachers who required no change; improvement in the medium term; and urgent and immediate improvement – comparing attitudes before and after training.

The findings include:

- Before training, 64% of teachers required no change in their classroom practices while after training, 73% of teachers required no change. In the final administration of the KAP survey (November 2016), 69% of the teachers indicated that they require no change.
- Before training, 32% of teachers required improvement in the medium term while after training only 23% did so. In the final administration of the KAP survey, 28% indicated that they require improvement in the medium term.
- Before training, 3% of teachers required urgent and immediate improvement while, after training, 4% required urgent improvement. In the final administration of the KAP survey, 3% of the teachers indicated that they required immediate improvement, which is the same as reported before training.

3. Summary of Findings

Lead Teacher performance on the ELIT Knowledge Questionnaire suggests that Wordworks' training of Lead Teachers was successful in terms of improving their knowledge of Grade R language and literacy concepts. Overall, the teacher KAP survey revealed improvements in knowledge of language and literacy concepts and pedagogical practices from before to after training. These knowledge gains were, for the most part, maintained over time.

The fact that the learner language and literacy effects declined over the same period suggest that refresher training may be necessary to reinforce learning and motivate teachers to continue using ELIT practices. These findings resonate with a recent USAID landscape report on early grade literacy in developing countries, which notes that formal or informal support structures are necessary to sustain newly acquired knowledge and instructional approaches (Kim, Boyle, Zuilkowski, & Nakamura, 2016).

Sub-question 2.1: To what extent has the integrity of the Stellar program been maintained through scale up?



The Stellar methodology, detailed to USAID, called for a full year of ongoing teacher support. When Stellar was integrated into the ELIT Intervention, the cluster/support sessions were optional after July 2016. The *integrity* of the Stellar program was thus *not maintained in principle* through scale-up.

In terms of the *integrity of the ELIT intervention*, as expected at the end of 2016, statistically significant differences were seen between comparison and intervention group classrooms in the use of Stellar and other ELIT materials and observation of ELIT literacy-related activities with more classrooms showing evidence of these materials and activities in the intervention group. While the materials were observed in almost all classrooms in 2016, many of the Stellar- and ELIT-related activities were observed in *less than half of the classrooms both directly after block training and at the end of the year*, indicating that there may be insufficient uptake to sustain an impact on Grade R learner literacy outcomes.

However, *teacher behaviors* before and during the reading exercise *improved overall* from prior to ELIT to after delivery of the intervention, suggesting that some components of the training methodology are beginning to filter through to classroom practices.

The evaluation team used a Classroom Environment Checklist Tool (CET) to answer this evaluation question. The CET included items from a USAID EGRA evaluation - focusing on instruction approaches, teacher actions, classroom environment, and pupil engagement in the area of language and literacy - which were adapted to suit the South African Grade R context. The tool also included items specific to what was expected of teachers following training in ELIT and looked at whether both Stellar and other ELIT materials and resources filtered through to the Grade R classrooms. The CET covers five broad aspects of classroom practice; pedagogic practices, learning and literacy activities, resources used by teachers, learning and teaching materials visible in the classrooms, and planning resources. Classrooms were also scored overall in terms of classroom appearance (print-richness).

Results: Classroom Environment Checklist Tool⁵⁶

In 2016, the teachers and their classrooms performed adequately to well (i.e. good practices observed in more than 50% of classrooms) in relation to pedagogic practices, use of Grade R classroom resources, emergent literacy materials, planning resources, classroom appearance, teacher's tables and storage facilities, and classroom environment. In general, these activities and resources were seen in more classrooms overall in November 2016 (intervention) compared to November 2015 (comparison).

The largest differences were apparent in the emergent literacy materials observed in the classroom. As expected, due to the rollout of Stellar materials in 2016, many more learning and teaching materials were observed in the 2016 classrooms than the same classrooms in 2015 (Figure 50, Annex 1), and almost all classrooms had evidence of these materials either displayed or the teacher was able to produce them. Stellar materials, specifically, filtered through to the intervention classrooms, with statistically significant improvements in all Stellar materials in 2016 compared to the year before⁵⁷. Stellar story modules (including manuals) were observed in 90% of classrooms in 2016 versus no classrooms in 2015, slightly more often in quintile 4 and 5 schools. Stellar crafts on the story module, and little books made by learners, were seen fairly evenly across all quintiles, in 56% and 71% of classrooms respectively, versus almost no classrooms in 2015⁵⁸. In terms of other ELIT-related materials in the classroom, statistically significant improvements were seen in letter boxes, picture cards (e.g. letter sound cards), and game bags/ activity packs.

Large differences were also seen in literacy-related activities observed in the classroom (Figure 49, Annex 1). In particular, role-play, picture sequencing, letter sound games, and use of little books were observed in statistically significantly more classrooms in 2016. Comparatively, role-play was not observed in any of the comparison classes in 2015, and neither was use of little books. Drawing and painting activities were observed statistically significantly less frequently in the 2016 classrooms indicating that time may have been taken from this activity to engage in other ELIT-related activities.

Even though there were statistically significant differences between the comparison and intervention classrooms in most literacy-related activities, less than 50% of intervention classrooms were implementing these activities at the end of 2016 (with the exception of emergent writing and songs and rhymes) suggesting that, although there was uptake, this was slow to filter through to classrooms.

What did not improve further were the planning resources available, teachers' tables and storage facilities, classroom appearance, and classroom environment as these were already present in many classrooms, or rated as adequate to good, in 2015.

⁵⁶ For a more detailed analysis, please see Interim Evaluation Report 4 (July 2017) on the Development Experience Clearinghouse.

⁵⁷ This difference was expected, given that Stellar materials were not available in the classrooms in 2015. However, the magnitude of the difference indicates the extent to which the materials filtered through to the classroom.

⁵⁸ Little books were seen in one 2015 classroom.

However, teacher behavior before and during the reading exercise improved overall (Figure 51 and 52, Annex 1), with statistically significant improvements in teachers showing the cover of the book and asking questions, reading the title and author, showing learners the pictures on each page, pointing to words and explaining some of them, and asking inferential questions. Although positive differences were also seen in a number of reading-related activities (setting up the reading activity in advance, a routine for changing classroom activities, making predictions about the upcoming story, stimulating awareness of sounds and letters, asking questions to help learners connect the story to their own lives, showing enthusiasm and interest in reading, and asking literal questions), none of these differences were statistically significant.

These results suggest that the train-the-trainer cascade training methodology is beginning to filter through to classroom practices and suggest there is some fidelity to the Stellar model as shown in the use of Stellar materials and resources and in some Stellar-related literacy and reading activities.

Sub-question 2.2: What factors enhance and/or constrain the scale-up of the Stellar training approach?

Key informants noted that to successfully scale the Stellar training approach through a provincial intervention such as ELIT, Subject Advisor buy-in and support is critical. Similarly, they noted that, for Subject Advisors and Lead Teachers, sufficient training in adult/peer training approaches is essential. The interviews revealed that trainers and teachers need sufficient training in phonics, specifically, to ensure the building blocks are in place for effective literacy instruction. Other factors that affect scale-up, noted by key informants, include communication about the intervention, partnership in rolling out the intervention and a training plan and resources for top-up/ refresher training and training new teachers entering the system.

In 2015, Stellar was incorporated into ELIT; a scaled provincial intervention. This evaluation question addresses the factors that affect the scale up of Stellar within the context of any provincial intervention. It draws on the experience of funders and implementers in the ELIT intervention in the Western Cape province of South Africa.

Khulisa conducted Key Informant Interviews (KIIs) with funders and key stakeholders using an interview protocol. The stakeholders represented in the KIIs include USAID, ELMA Philanthropies, Wordworks, and the WCED. Interviews were conducted from July through October 2017 on a one on one basis; either in person or telephonically. The protocol consisted of twelve approved question items covering a range of topics.

Results: Key Informant Interviews

A few key informants noted that ELIT has a good model for scaling and sustaining itself because it is incorporated into WCED operations. The respondents all felt that to sustain something in the system, it has to be embedded, owned, and be part of what officials do. Two respondents noted that the reason why programs are often not sustained is that an outside agency appears, provides

resources and training and then leaves. The respondents noted that a whole system approach was necessary for systemic change. A few respondents noted that funders also often stop funding activities too soon; they should rather provide continued monitoring and support after training.

The existence of program champions was also noted as a factor enhancing scale up and sustainability. Champions need to exist at every level – the implementer, the districts, advisors, and trainers, the ECD Directorate and Lead Teachers. An underlying concern was raised about authority and whether or not the champions have the ability to “run” with the program. Specifically, Subject Advisor buy-in and support was noted as critical to successful scale-up and sustainability of the Stellar training approach within the context of a provincial intervention such as ELIT.

Conflicting priorities was mentioned by most respondents as a factor affecting the scale up of Stellar within the context of the ELIT intervention. For example, the provincial focus on Grade R Maths in late 2016 into 2017 was noted by a few key stakeholders as a risk to the time spent on literacy in the classroom. A few respondents noted that the training burden on district advisors, Lead Teachers, and teachers puts tremendous pressure on the system, including the departmental resources necessary for delivering training. Linked to this, funding was listed by most interview respondents as a factor constraining successful and sustainable scale up.

Other factors affecting scale-up and sustainability, noted by respondents, include communication about the intervention, both with districts and between implementers, partnership in rolling out the intervention and, a training plan for top-up or refresher training and training of new teachers.

Many respondents noted that sufficient training in instructional approaches for adults is essential for Subject Advisors and Lead Teachers. Other opinions raised were that trainers and teachers need sufficient training in phonics, specifically, to ensure the building blocks are in place for effective literacy instruction.

Evaluation Question 3



What is the likelihood of sustainability of organizational and beneficiary gains after Stellar interventions end?



Organizational sustainability capacity is beginning to be built but could be improved through systematic ongoing teacher support. For beneficiary learners, quality instruction across multiple years is essential.

ELIT has a better chance for sustainability than if it had been rolled out independently because it is incorporated into the operations of the WCED. A number of key informants noted that the WCED needs to document their sustainability strategy with a training plan and resources for refresher courses, top-up training, and training new teachers.

The Foundation Phase HOD training on the ELIT program is a step in the right direction and so is the set-up of Grade R teacher communities of practice.

To answer this evaluation question, the evaluation team conducted a *Program Sustainability Assessment* with Subject Advisors that were involved in the evaluation, *Key Informant Interviews* with the program implementers and funders, and a *Cost Element* analysis to gauge the level of effort and other resources necessary to roll out a provincial intervention of this nature.

Subject Advisors were asked to complete a 40-item online self-assessment survey. The survey measured 8 sustainability “domains”: *environmental support, funding stability, partnerships, organizational capacity, program evaluation, program adaptation, communications, and strategic planning*⁵⁹. The survey was administered in 2016 and repeated in 2017.

Khulisa conducted KIIs with funders and key stakeholders using an interview protocol. The stakeholders represented in the KIIs include USAID, ELMA Philanthropies, Wordworks, and the WCED. Interviews were conducted from July through October 2017 on a one on one basis; either in person or telephonically. The protocol consisted of twelve approved question items covering a range of topics.

This section is structured as follows:

1. Results: Sustainability Assessment with Subject Advisors
2. Results: Key Informant Interviews
3. Results: Cost Elements
4. Summary of Findings

⁵⁹ See Annex 3 for the Sustainability Assessment Tool.

1. Results: Sustainability Assessment with Subject Advisors

Per the sustainability assessment findings in Annex 1, Figures 26 to 33, *environmental support* and *program evaluation* both received an overall score of 4.8/7 in 2016 increasing to 4.9/7 (*environmental sustainability*) and 5.4/7 (*program evaluation*) in 2017. *Funding stability* received an overall score of 3.9/7 in 2016 and this score remained the same in 2017. *Partnerships* received the lowest score at 3.1/7 in 2016, improving slightly to 3.3/7 in 2017. While *organizational capacity* received the highest overall score at 5.4/7 in 2016, it increased even further in 2017 to 5.68/7. *Program adaptation* came in at an average score of 4.3/7 in 2016 but went down to 3.9/7 in 2017. *Communications* received a score of 3.7/7 in 2016 and increased slightly to 4.1/7 in 2017. *Strategic planning* received an overall score of 4.2/7 in 2016 compared to 4.4/7 in 2017.

According to the findings, the ELIT intervention's greatest strength lies in the fact that the project has strong champions with the ability to gather or garner resources and has leadership support from within the larger organization (WCED). On the other hand, the project does not yet have strong public support (i.e. support from parents and caregivers).

Overall, the findings suggest that funding stability has room to grow. This could potentially affect the ability of the project to sustain knowledge gains if refresher training for teachers is not available or in the case that materials and resources need to be replaced. Training of new teachers could also be affected.

According to the findings, the ELIT intervention has low support in terms of community engagement (i.e. parents and other stakeholders), involvement, and communication. A recent USAID Landscape Report on early grade literacy in developing countries found that, "...it is important to identify effective ways to involve parents and community members with low literacy levels into literacy programs. Parents must be able to read well enough to coach their children." (Kim et al., 2016, p. 58). Wordworks is in the process of developing a Parent Manual, aimed at teachers, which guides them on how to involve parents at home in the emergent literacy process. Once completed, this resource could contribute to community engagement.

Subject Advisors do feel that, for the most part, the ELIT project has the internal support and resources needed to effectively manage the project and its activities. A particular area of strength is the integration of the project into the operations of the WCED and/or education district offices. Organizational systems seem to be in place to support the project. The ability to articulate the vision of the project to external stakeholders (e.g. parents and caregivers) could be improved. Although the Subject Advisors did not suggest specific ways in which this could be done, parents and caregivers should understand the intended outcome of the program.

Communication of evaluation results has improved from 2016 to 2017. The ELIT intervention seems to have the capacity for quality evaluation, and reports some short term and intermediate outcomes. The Subject Advisors report that there could be better communication of evaluation results to the public and other key stakeholders.

The results indicate that the ability of the ELIT project to take actions that adapt the project to ensure its ongoing effectiveness is good but could be strengthened. As in 2016, in 2017 the capacity to make decisions about which components are ineffective and should not continue was identified as an area that needs specific improvement.

Communication of the project with parents and other stakeholders is another area in which the ELIT project could improve its sustainability capacity. The results indicate that there is a need for

project staff to communicate the importance of the project to the public, in particular. Communication strategies to increase community awareness and demonstrate value to the public could improve buy-in.

The ELIT intervention seems to be using processes to guide the project's direction, goals, and strategies. A particular area of strength, according to the findings, is in the fact that the project's goals are understood by all stakeholders. Once again, the lack of a long-term financial plan was identified as an area needing support. However, there was a marginal improvement from 2016 to 2017 on this item.

The results of the Sustainability Assessment remained largely the same from 2016 to 2017 suggesting that there has been no noteworthy changes in the ELIT intervention's sustainability capacity from the perspective of the Subject Advisors. Overall, there was no significant difference in the average scores from 2016 to 2017. However, the fact that only half the Subject Advisors from 2016 completed the survey in 2017 signifies that these results may not reflect the perspective of the entire Subject Advisor group.



Photo: Khulisa

2. Results: Key Informant Interviews

In relation to Key Informants' understanding of the ELIT intervention, **ownership** emerged as a prominent theme. The interviews revealed some ambiguity surrounding a clear understanding and interpretation of how the Stellar program and ELIT intervention would be integrated and who would "own" it.

Three of the four Key Informants held the view that the Stellar program methodology and practices were adapted to match the ELIT focus in collaboration with the Stellar program developer (Wordworks) and the ELIT implementer (WCED). One stakeholder held the view that Stellar was extended and adapted to fit the WCED daily program. Specifically, the stakeholder noted that,

“...ELIT is the actual intervention but lesson plans for the subjects are an adaptation of the original Stellar material”.

There was no agreement between the respondents on the direction of influence. While the majority of respondents felt that the ELIT intervention was influenced by Stellar, one respondent noted that the ELIT intervention was conceptualized in advance of the Wordworks/WCED partnership. Overall Stellar played an important role in the development of materials to be used by Grade R teachers to help them improve their teaching practices.

Out of ownership, a secondary theme of **role delineation** emerged from the interviews. The ELIT intervention was set up to be an equal partnership between Wordworks and the WCED. USAID funded the project, was on the steering committee, tried to resolve blockages, and kept their finger on the pulse of the intervention, but ultimately had a very limited role. Wordworks was responsible for training the WCED ECD and Foundation Phase Subject Advisors and NGO-based support advisors in the ELIT program, providing Stellar program materials and training resources, and providing ongoing support to SAs and Lead Teachers. The WCED Curriculum Directorate's SAs were responsible for selecting, training, and providing support to Lead Teachers in the ELIT intervention, managing the distribution of materials and training logistics in each district, supporting Lead Teacher training of teachers, and occasionally monitoring teacher progress and providing support and feedback (Smith & Hugow, 2017). Over time, the roles and responsibilities of Wordworks and the WCED became more clarified. WCED protocol had to be followed to avoid confusion and overload on Subject Advisors, teachers and districts.

Negotiation and cooperation was a related theme emerging in the respondent data. While the original plan for the Stellar program, before it was integrated into ELIT, was to have 12 months of support/cluster sessions, the WCED initially did not plan to have cluster meetings. Through negotiation and cooperation between the ELIT partners, the WCED agreed to 6 months of compulsory cluster sessions for Grade R teachers. Although Wordworks initially wanted to hold the block training first, and then the cluster meetings, the order was reversed in line with WCED policy to train during school holidays. This ended up being “a good thing”; by the time the teachers got to the block training, they had an idea of the model and had some experience with it. Some key informants felt that an “ideal model” emerged through the process: clusters, then block training, then further clusters for enhanced classroom practices. However, it is unclear whether the latter set of clusters took place in the majority of the districts because they were not compulsory.

Different respondents **described sustainability differently** in the context of the ELIT intervention. One respondent described it as changed teacher behavior; the fact that two thirds of districts asked for extra training and materials suggested to this respondent that teacher behavior had adjusted. Another described sustainability as continued support and training; the respondent believed that if ELIT “was not available” in 2017 it would not be sustainable. For one respondent, the sustainability of the project was “all about empowering Subject Advisors and Lead Teachers to understand and continue the program once the service providers have left”.

A few respondents noted that ELIT has a good sustainability model because it is **incorporated into WCED operations**. Availability of resources needed to purchase or photocopy files/resources from Wordworks would be necessary. In addition, human resources would be needed to keep the intervention rolling in the districts. One respondent noted that the Stellar file

was a consumable resource. Thus the ELIT intervention, in this respondent's eyes, had to be bigger than the Stellar resources.

The interviews revealed that the WCED is thinking about sustainability strategically. **Top up training** using Heads of Department (HODs) emerged as a strategy to reach new teachers who did not receive ELIT training. **Ongoing communities of practice** were also considered as part of the provincial plan and budget to strengthen classroom practices. Districts were seen as strategic in embracing top up training through an HOD workshop.

A few respondents noted that the Subject Advisors, Lead Teachers and teachers all need **more phonics training** to properly integrate this into their daily practices.

3. Results: Cost Elements

The main elements to the costs involved in scaling an intervention such as ELIT, province-wide, include:

NGO/ USAID Implementing Partner (Wordworks)

a. Personnel

- Executive oversight and governance
- Finance & operations management
- Accounting
- Content & training management
- Logistics & materials co-ordination
- Training and district support
- Monitoring and evaluation
- Communications (including design & translation)

b. Travel and accommodation

c. Equipment

d. Supplies

- Printing of all resources for the provincial roll-out
- Stationery
- Postage/courier of materials

f. Other

- Project meetings
- Administration and overheads

Provincial Government

On the part of the Provincial government, the main elements include:

a. Personnel

- b. Government staff time
- c. Government staff time in training (provincial staff)
- d. Government staff time in training (district staff)
- e. Government staff time in training (teachers)
- f. Executive oversight and governance
- g. Finance and operations management
- h. Accounting
- i. Content and training management
- j. Logistics and materials co-ordination
- k. Training and district support
- l. Monitoring and Evaluation
- m. Communications (including design & translation)
- n. Venue
- o. Materials/ equipment/ supplies
- p. Transportation
- q. Direct monetary contributions
- r. Other

4. Summary of Findings

The fact that the WCED has trained HODs in ELIT concepts, and that it continues to support Grade R teacher communities of practice, suggests that sustainability capacity is beginning to be built in to WCED operations, increasing the possibility of organizational sustainability in the medium to long term.

Factors that enhance sustainability include proper role delineation, with an understanding of who is responsible for what aspect of the intervention. Negotiation and cooperation is essential with flexibility needed on the part of all partners (including funders) to fit in with provincial mandates. Finally, a common definition of sustainability is needed so that implementers all work towards the same goal.

It is clear that some ongoing costs will be necessary to sustain ongoing teacher support for Grade R teachers.

CONCLUSIONS, RECOMMENDATION, AND LESSONS

Conclusions

As discussed in this report and at every stage with USAID representatives, this impact evaluation was affected by a variety of factors:

- The evaluation contracting process resulted in a late evaluation start. The evaluation should have begun in early 2015 but the contract was only awarded in June 2015. Since the implementation rollout was scheduled to begin in January 2016, the baseline scheduling was compromised with data collected in August and October/November 2015. The testing should have taken place annually in January and November. As anticipated, the delay decreased the probability that the evaluators would see significant short term effects due to the short time between the first point of data collection and the end of the year. Testing in August ruled out the possibility of collecting a clean baseline.
- Because the WCED's standard practice is simultaneous rollout to every district and school, a Randomized Control Trial methodology was ruled out. Instead, it was agreed that a quasi-experimental design would be used, comparing baseline data prior to rollout for Grade R in August and October/November 2015, to learners in the same schools but attending Grade R in 2016 during rollout. The 2015 cohort and the 2016 cohort were tested again in Grade 1. Although there is no reason to believe that the two cohorts were inherently different, without clean baseline data, this is impossible to demonstrate unequivocally.
- Wordworks had designed the methodology and detailed the dosage of the Stellar activity to USAID. However, once the Stellar methodology was integrated into the wider ELIT approach, the fidelity to the original program was compromised in that it was only mandated for 6 months and through the cascade training approach. The General Education and Training Schedule had one "optional" cluster workshop post-block training in September 2016 and stated that "Teachers may continue monthly clusters to suit context".

In summary, the evaluation found that Grade R learners in 2016 (the intervention group) outperformed Grade R learners in 2015 (the comparison group). However, in 2017, there was no significant difference between the intervention group Grade 1 results and the comparison group Grade 1 results in 2016. The intervention effect essentially disappeared.

One hypothesis for this perplexing result is that the intervention spilled over into the comparison group in 2016. In other words, the push to implement with the cluster meetings, materials, and training for the Grade R teachers may have affected Grade 1 teaching practices in 2016 - the year that the ELIT intervention took place. This phenomenon was seen in the USAID/Malawi EGRA Performance Evaluation in 2015 (Orr & Bisgard, 2015).

To explore this possibility, the evaluation team subsequently interviewed Grade 1 teachers in approximately three quarters of the sample schools; two thirds of the teachers indicated that they did not know about the ELIT intervention and a third indicated that while they were aware of the ELIT intervention, they did not know what it entailed. Added to this, given the complex nature of the ELIT rollout, it is unlikely that spill-over would have affected Grade 1 teachers without sufficient training and intervention dosage.

Alternative hypotheses include the theory that third party variables may have affected learner performance in the Western Cape in 2016 and 2017, or that there were inherent differences in the 2015 and 2016 Grade R cohorts.

The evaluation design cannot account for these possibilities because:

- The intervention was not randomized and so there is a possibility of selection bias.
- Despite controlling for school and teacher level characteristics, it is difficult to say unequivocally that the intervention and comparison cohorts are equivalent.
- The WCED were unable to provide information on other initiatives and on the cost of implementing the ELIT intervention.
- Implementation fidelity was variable, so dosage was questionable.
- The tests should have been conducted in January and November to allow learning to take place and improve the evaluation design.

All this being said, given the limitations to the evaluation design, the evaluation cannot say indisputably whether the ELIT intervention was successful or not, either in the short or long term.

Recommendations

Future impact evaluations should:

- Start with an evaluability assessment.
- Be designed, where possible, in collaboration with policymakers to ensure that the results are fit-for-purpose and use.
- Provide sufficient budget to collect data on important third party variables, particularly those known to affect literacy performance both in South Africa and internationally.
- Require interventions to be implemented in a randomized manner, to allow for the evaluation of impact relative to a robust counterfactual.

Future early grade literacy interventions implemented by provincial governments should:

- Lay out clearly, based on evidence, the minimum dose⁶⁰ of instructional time in teacher professional development necessary for effective language and literacy instruction in class (Kim et al., 2016).
- Articulate what support structures need to be in place to sustain newly achieved teacher learning and pedagogy (Kim et al., 2016), both in the shorter term (within Grade R) and the longer term (carryover effects from Grade R to Grade 1).

⁶⁰ Kim et al. (2016, p.11) note that: "Although a rough estimate of 850 to 1000 hours per year has been recommended as the minimum instructional time needed in reading over a school year (Gilles & Jester Quijada, 2008), there is no clear evidence from low-income countries about how much instructional time is necessary or sufficient for successful literacy instruction"

- Make space for well-designed impact evaluations, including piloting and refining an implementation model before an impact evaluation is carried out.
- Strategically think about which schools may need which type of interventions most. It is rarely feasible or effective to roll-out a one-size-fits-all intervention to all schools.
- Test whether the funding spent on teacher development is leading to any change in learner performance. Evaluating teacher training efforts are very important and training should therefore be designed so that the effects can be evaluated.
- Ensure strong alignment between the teaching of language in Grade R and in Grade 1. Grade 1 early reading interventions must develop, support and sustain foundational reading skills. USAID's landscape report found that quality instruction across multiple years is essential; a single year of instructional support is insufficient (Kim et al., 2016).

Lessons

While more information is needed to better understand policy implications, there are implementation lessons:

- The cascade training approach may have had an initial effect but without provincially mandated support past July 2016, the intervention advantage appears to have been lost. A recent World Bank report on early grade reading in developing countries indicates that following initial training with coaching and monitoring is needed to compound the teacher training effectiveness (Kelly & Graham, 2018).
- The logic of change underlying most professional development programs assumes that there is fidelity to the professional development package in the way that it is administered, rolled out and the contents imparted, demonstrated, and supported in follow up activities. To document impact, it is very important that the intervention adhered strictly to the program design. Potentially, lack of fidelity to the professional development package, implementation modes and support may have negatively affected the change process.

If the results are held to be true, the ELIT intervention appears to have no lasting impact beyond Grade R. This is not surprising as neurologically learners need constant reinforcement at this age (Abadzi, 2006).

It may be worthwhile interrogating the underlying assumption that an intervention in Grade R prepares learners for reading in the later grades in spite of the classroom conditions, training, and preparedness of teachers in the later grades and notwithstanding learners' home conditions. The latter point is made in light of USAID's recent Landscape Report on Early Grade Literacy (Kim et al., 2016, p.58), which notes that "...school-based approaches alone have not proved sufficient to reach the goal of widespread reading proficiency among early grade children in many developing countries" and that "...broadening the scope of literacy interventions beyond the school is a necessary step." Examples include routinely incorporating home and community reading components into early grade language and literacy interventions.

Altogether, the evaluation results suggest that a good Grade R is necessary, but not sufficient, for future language and literacy acquisition.



Photo: Khulisa

DATA QUALITY AND REPORTING SYSTEM

Using the ODK mobile-phone-based data collection system played a significant role in ensuring rapid and reliable data collection and more rigorous data quality. The completed forms were automatically uploaded to a cloud-based server when the device was in range of a wireless signal. Khulisa was able to extract and aggregate data on a daily basis and conduct spot checks to ensure that data quality was maintained and that fieldworkers were using the tools correctly. Khulisa was also able to troubleshoot issues in real time to improve data entry, data management, and data quality.

The evaluation team set up a WhatsApp group with the fieldworkers during training which allowed the researchers to readily identify and address data quality issues and, where necessary, deal with fieldworkers either as a group via the WhatsApp application or on an individual basis telephonically to ensure that all data was completed appropriately.

ISSUES

For the most part, the evaluation administrative/logistical issues were minimal. The end of year data collection in 2016 (wave 2 intervention and wave 3 comparison group in Grade 1) was challenging as a large number of research tools had to be administered simultaneously for two learner cohorts. Other technical and administrative issues included:

Learner Tracking

Tracking learners was challenging, as after Grade R 145 learners out of 940 either moved schools (including leaving the province), were deceased, had a long-term illness, or could not be traced.

Within the comparison group, Khulisa could track 57 learners to new schools (23% of the sample) in 2016. Of these 57 learners, 30 moved to other schools within the evaluation sample and 27 learners moved to schools that are not part of the sample. Only 2 of these learners were in the same school, requiring visits to an additional 26 schools outside of the evaluation sample.

In 2017, Khulisa was able to track 23 intervention learners to new schools, of which 2 learners moved to schools within the evaluation sample and 21 learners moved to schools outside the sample. This required visits to an additional 21 schools outside of the evaluation sample.

The Western Cape Education Department assisted the evaluation team with tracking learners. This support was integral to reducing attrition in the sample. However, the schools that the learners were listed as having moved to was sometimes inaccurate. The WCED may therefore find it worthwhile to assess the processes associated with tracking learners in the province.

Pacing and Curriculum Coverage in Grade 1

The Pacing and Curriculum Coverage Tool was designed assuming that all the Grade R classes would be using the most recent DBE workbooks. Each year, the schools are supposed to receive enough Grade R workbooks for each learner. However, it became apparent that there were different DBE workbook versions in the schools. This made it very difficult for the evaluation team to calculate a common denominator for comparison purposes in the analysis.

Planned versus Actual Implementation

It is unclear whether the ELIT intervention was implemented consistently through the cascade model as planned. This makes it difficult to assess the integrity of the intervention. Classroom environment and evidence of materials and practices in the classroom were used as a proxy.

FUTURE DIRECTIONS

The suggested future directions for this evaluation are reflected in the recommendations. In addition, the following practical recommendations are made for future interventions and evaluations based on lessons learned:

- Longitudinal evaluations should consider learner attrition. An attrition rate of 10% of the sample is likely to occur.
- Linked to attrition, over-sampling of schools and learners would likely be more cost effective than tracking learners to new schools and provinces.
- Baseline data should be collected as close to the beginning of the school year as possible. These dates should be negotiated with the relevant department in advance, considering the school calendar.
- Sufficient time should be built in to obtain research permission, develop and/or adapt tools and pilot test the evaluation instruments.
- Translation of instruments is a major issue, particularly for assessing learners in the early grades. Professional translation is necessary but not sufficient. Experienced fieldworkers can provide a valuable service. Given their knowledge and proven ability to assess young learners, they need time to effectively review the instruments and provide feedback.
- Linked to translation, variations in local language/dialect should be taken into account. “Proper” language does not always translate to understandable assessment questions, particularly for young learners.
- The fieldworker quality directly affects the quality of the data collected. It is possible to hire fieldworkers inexpensively, but there may be a trade-off. Hiring experienced fieldworkers who are more expensive, but know how to approach grade R learners, who know how to assess in the early grades, and who are familiar with the schools and school management teams in the province is efficient, effective, and invaluable.
- Stakeholder engagement, at every step of the process, is important. The WCED engaged at a very senior level with the Evaluation Learning Brief (See Annex 10) indicating that they will incorporate the lessons learned into their program. However, evaluation timelines need to accommodate enough time for all stakeholders to engage with the evaluation findings and recommendations before an evaluation is finalized.

REFERENCES

- Abadzi, H. (2006). *Efficient Learning for the Poor: Insights from the Frontier of Cognitive Neuroscience* (Directions in Development No. 36619). Retrieved from <http://documents.worldbank.org/curated/en/438221468134385073/pdf/366190Efficien101OFFICIAL0USE0ONLY1.pdf>
- Albouy, D. (1994). *Program Evaluation and the Difference in Difference Estimator*. *Economics 131-Section Notes*, 1–4.
- Ellis, P. D. (2009). Thresholds for interpreting effect sizes. Retrieved from http://www.polyu.edu.hk/mm/effectsizefaqs/thresholds_for_interpreting_effect_sizes2.html
- Gertler, P. J., Martinez, S., Premand, P., Rawlings, L. B., & Vermeersch, C. M. J. (2011). *Impact Evaluation in Practice*. The International Bank for Reconstruction and Development / The World Bank. <https://doi.org/10.1596/978-0-8213-8541-8>
- Great Schools Partnership. (2014). *Criterion-Referenced Test Definition*. Retrieved February 27, 2018, from <https://www.edglossary.org/criterion-referenced-test/>
- Kelly, S., & Graham, J. (2018). *How Effective Are Early Grade Reading Interventions ? A Review of the Evidence* (Policy Research Working Paper No. 8292).
- Kim, Y.-S. G., Boyle, H. N., Zuilkowski, S. S., & Nakamura, P. (2016). *Landscape Report on Early Grade Literacy*. Washington, D.C.
- Orr, J. K., & Bisgard, J. (2015). *Performance Evaluation of the USAID / Malawi Early Grade Reading Activity (EGRA)*.
- Skiles, M. P., Hattori, A., & Curtis, S. L. (2014). *Impact Evaluations of Large-Scale Public Health Interventions: Experiences from the Field*.
- Smith, A. S., & Hugow, I. (2017). Report on the findings from a qualitative study of the emergent literacy project roll out in the Western Cape, (January), 0–31.
- Taylor, S. (2014). *Strengthening the impact of the Grade R programme*. Retrieved from [http://children.pan.org.za/sites/default/files/publicationdocuments/Impact of introduction of Grade R Policy Brief. DPME.pdf](http://children.pan.org.za/sites/default/files/publicationdocuments/Impact%20of%20introduction%20of%20Grade%20R%20Policy%20Brief.pdf)
- USAID. (2013). *Technical Note: Impact Evaluations*. Retrieved from https://usaidlearninglab.org/sites/default/files/resource/files/ie_technical_note_2013_0903_final_2.pdf
- Wordworks. (2016). *The STELLAR programme: A summary*. Retrieved from <http://www.wordworks.org.za/downloads/STELLAR-programme-summary.pdf>

ANNEXES

- Annex 1: Summary Statistics and Results
- Annex 2: Methodology Report Extract
- Annex 3: Data Collection Tools
- Annex 4: Schedules and Project Activities
- Annex 5: Interview Lists and Meetings
- Annex 6: Sources of Information
- Annex 7: Statement of Work
- Annex 8: Statement of Differences
- Annex 9: Evaluation Team
- Annex 10: Learning Brief for the Government of South Africa

ANNEX 1: SUMMARY STATISTICS AND RESULTS

Learner Assessment / Early Grade Reading Assessment

		Grade R					
		Comparison			Intervention		
		Obs	Mean	Std. Dev.	Obs	Mean	Std. Dev.
Urban:		450	0.70	0.46	490	0.69	0.46
LOLT:	<i>English</i>	450	0.18	0.38	490	0.18	0.39
	<i>Afrikaans</i>	450	0.64	0.48	490	0.63	0.48
	<i>isiXhosa</i>	450	0.18	0.38	490	0.18	0.39
Ex Department:	<i>CED</i>	450	0.16	0.37	490	0.16	0.37
	<i>DET</i>	450	0.10	0.30	490	0.10	0.30
	<i>HOR</i>	450	0.58	0.49	490	0.57	0.50
	<i>WCED</i>	450	0.16	0.37	490	0.16	0.37
Quintile:	<i>1</i>	450	0.18	0.38	490	0.18	0.39
	<i>2</i>	450	0.20	0.40	490	0.20	0.40
	<i>3</i>	450	0.22	0.41	490	0.20	0.40
	<i>4</i>	450	0.20	0.40	490	0.20	0.40
	<i>5</i>	450	0.20	0.40	490	0.20	0.40

Figure 14: School Characteristics

		Grade R						Grade 1					
		Comparison			Intervention			Comparison			Intervention		
		Obs	Mean	Std. Dev.	Obs	Mean	Std. Dev.	Obs	Mean	Std. Dev.	Obs	Mean	Std. Dev.
Female		450	1.00	0.00	490	1.00	0.00	364	1.00	0.00	421	1.00	0.00
Teacher age*		450	41.10	10.74	481	37.95	16.13	362	47.14	11.51	403	45.90	12.13
Teacher experience		450	10.50	9.55	490	9.41	7.13	364	18.52	12.54	394	14.16	11.96
Teacher qualification	<i>Less than matric</i>	450	0.02	0.14	490	0.04	0.20	364	0.01	0.12	421	0.00	0.00
	<i>Matric</i>	450	0.06	0.24	490	0.08	0.27	364	0.00	0.00	421	0.00	0.00
	<i>Certificate / Diploma</i>	450	0.84	0.37	490	0.80	0.40	364	0.64	0.48	421	0.50	0.50
	<i>Degree or More</i>	450	0.08	0.27	490	0.08	0.27	364	0.35	0.48	421	0.50	0.50

* 80% of the comparison group Grade R teachers were in the intervention group - the same classrooms were used in the comparison and intervention conditions. In 20% of classrooms in 2016, the teacher had either been replaced (teacher mobility/retirement) or a substitute teacher was present on the day of testing, which explains the difference in mean age. The Grade 1 teacher age is not useful as there were numerous Grade 1 classes and a number of learners moved to new schools.

Figure 15: Teacher Characteristics

		Comparison									Intervention								
		Wave 1			Wave 2			Wave 3			Wave 1			Wave 2			Wave 3		
		Obs	Mean	Std. Dev.	Obs	Mean	Std. Dev.	Obs	Mean	Std. Dev.	Obs	Mean	Std. Dev.	Obs	Mean	Std. Dev.	Obs	Mean	Std. Dev.
Female:		450	0.51	0.50	434	0.52	0.50	405	0.51	0.50	490	0.47	0.50	481	0.47	0.50	450	0.44	0.50
Age:		449	5.73	0.55	433	5.72	0.55	405	7.37	0.55	488	6.16	0.33	479	6.34	0.33	448	7.34	0.31
Home Language:	<i>Afrikaans</i>	450	0.63	0.48	434	0.63	0.48	405	0.62	0.48	490	0.62	0.49	481	0.64	0.48	418	0.63	0.48
	<i>English</i>	450	0.15	0.36	434	0.15	0.26	405	0.16	0.37	490	0.17	0.37	481	0.15	0.36	418	0.16	0.36
	<i>isiXhosa</i>	450	0.21	0.41	434	0.21	0.40	405	0.21	0.41	490	0.21	0.40	481	0.20	0.40	418	0.20	0.40
	<i>Other</i>	450	0.01	0.05	434	0.01	0.11	405	0.01	0.10	490	0.01	0.08	481	0.02	0.13	418	0.01	0.12
Home Language and LoLT the same:		450	0.94	0.24	434	0.94	0.25	405	0.94	0.24	490	0.95	0.22	481	0.95	0.23	421	0.92	0.27

Figure 16: Learner Characteristics

		p10	p25	p50	p75	p90	min	max
Original Instrument (Percentage Scores)	<i>Writing</i>	0.54	0.65	0.81	0.92	1.00	0.23	1.00
	<i>Listening Speaking Conversation</i>	0.67	0.71	0.79	0.83	0.88	0.21	1.00
	<i>Listening Speaking Story</i>	0.31	0.44	0.50	0.63	0.69	0.06	0.88
	<i>Reading</i>	0.35	0.61	0.86	0.94	0.97	0.00	1.00
	<i>Phonological Awareness</i>	0.39	0.64	0.86	1.00	1.00	0.00	1.00
	<i>Phonics Letters</i>	0.21	0.29	0.57	0.79	0.86	0.00	1.00
EGRA (Count Scores)	<i>Letters sounded - English</i>	22	33	38	40	40	0	40
	<i>Letters sounded - Afrikaans</i>	10	28	38	40	40	0	40
	<i>Letters sounded - isiXhosa</i>	9	21	37	40	40	0	40
	<i>Words sounded - English</i>	3	6	11	31	53	0	80
	<i>Words sounded - Afrikaans</i>	0	3	14	27	43	0	80
	<i>Words sounded - isiXhosa</i>	0	0	11	25	42	0	80
	<i>Words read - English</i>	0	7	12	23	39	0	112
	<i>Words read - Afrikaans</i>	0	3	12	21	46	0	115
	<i>Words read - isiXhosa</i>	0	0	2	16	28	0	52
	<i>Comprehension Questions - English</i>	0	0	1	2	3	0	5
	<i>Comprehension Questions - Afrikaans</i>	0	0	0	1	2	0	4
	<i>Comprehension Questions - isiXhosa</i>	0	0	0	2	4	0	5
Index Score	<i>Combined Score</i>	-4.79	-1.90	0.89	2.41	3.30	-11.44	4.78

Figure 17: Sub-Task Score Distribution of the Intervention Group Wave 3 (Gr 1) Assessment

			Mean		s.e.	
			Comparison	Intervention	Comparison	Intervention
Original Instrument (percentage scores)	Listening & Speaking Conversation	Wave 1	0.69	0.74	0.008	0.007
		Wave 2	0.79	0.79	0.005	0.004
		Wave 3	0.77	0.78	0.004	0.005
	Listening & Speaking Story	Wave 1	0.52	0.64	0.010	0.008
		Wave 2	0.61	0.62	0.009	0.008
		Wave 3	0.52	0.53	0.007	0.006
	Phonological Awareness	Wave 1	0.33	0.37	0.010	0.011
		Wave 2	0.40	0.44	0.012	0.011
		Wave 3	0.77	0.79	0.013	0.012
	Phonics Letters	Wave 1	0.24	0.24	0.011	0.010
		Wave 2	0.19	0.22	0.010	0.010
		Wave 3	0.52	0.56	0.012	0.012
	Reading	Wave 1	0.31	0.33	0.005	0.005
		Wave 2	0.32	0.35	0.007	0.007
		Wave 3	0.77	0.75	0.012	0.012
EGRA (count scores)	Sound Recognition (/40)	English	33.56	35.07	1.04	0.99
		Afrikaans	33.00	31.30	0.69	0.71
		isiXhosa	28.55	29.50	1.81	1.49
	Word Recognition (/80)	English	19.73	20.07	2.18	2.12
		Afrikaans	21.24	16.73	1.31	1.06
		isiXhosa	15.75	16.77	2.20	2.59
	Paragraph Reading (/115)	English	18.39	15.57	2.03	1.80
		Afrikaans	20.09	15.04	1.49	1.15
		isiXhosa	10.72	9.06	1.63	1.62
	Comprehension Questions (/5)	English	1.11	1.05	0.13	0.13
		Afrikaans	0.72	0.55	0.07	0.06
		isiXhosa	1.21	0.94	0.22	0.19

Figure 18: Mean Performance per Sub-Task – Comparison and Intervention Groups

Note: Learner Assessment sub-task scores in this table are converted to percentage scores. The EGRA scores are reported as the number of letter/ words correct.

			Wave 1							Wave 2						
			Index	LC	LS	PA	PL	R	W	Index	LC	LS	PA	PL	R	W
Wave 3 Comparison	Learner Assessment	Index Score	0.46	0.21	0.30	0.41	0.42	0.14	0.33	0.54	0.19	0.37	0.48	0.40	0.42	0.45
		Listening Speaking Conversation	0.29	0.19	0.23	0.24	0.19	0.08	0.23	0.32	0.26	0.27	0.25	0.18	0.20	0.24
		Listening Speaking Story	0.21	0.15	0.25	0.18	0.17	0.04	0.07	0.23	0.13	0.27	0.16	0.10	0.18	0.14
		Phonological Awareness	0.37	0.19	0.26	0.35	0.31	0.10	0.28	0.45	0.18	0.28	0.44	0.34	0.31	0.35
		Phonics Letters	0.35	0.12	0.21	0.38	0.29	0.09	0.30	0.38	0.11	0.17	0.36	0.33	0.26	0.36
		Reading	0.29	0.19	0.19	0.24	0.28	0.06	0.21	0.43	0.18	0.35	0.35	0.28	0.33	0.35
		Writing	0.29	0.18	0.18	0.23	0.30	0.08	0.21	0.35	0.12	0.27	0.29	0.19	0.29	0.29
	EGRA	Sound Recognition	0.34	0.16	0.22	0.28	0.30	0.10	0.31	0.44	0.20	0.27	0.35	0.34	0.31	0.39
		Word Recognition	0.42	0.18	0.26	0.37	0.41	0.18	0.28	0.48	0.16	0.31	0.42	0.36	0.38	0.40
		Paragraph Reading	0.33	0.11	0.22	0.31	0.36	0.12	0.20	0.39	0.08	0.26	0.36	0.30	0.34	0.29
Wave 3 Intervention	Learner Assessment	Index Score	0.57	0.28	0.25	0.45	0.44	0.28	0.51	0.66	0.34	0.37	0.59	0.46	0.45	0.48
		Listening Speaking Conversation	0.44	0.40	0.29	0.30	0.22	0.23	0.31	0.38	0.35	0.24	0.34	0.15	0.30	0.21
		Listening Speaking Story	0.43	0.32	0.29	0.29	0.23	0.26	0.31	0.48	0.34	0.36	0.36	0.21	0.36	0.38
		Phonological Awareness	0.52	0.27	0.23	0.46	0.40	0.25	0.40	0.57	0.27	0.32	0.58	0.40	0.37	0.35
		Phonics Letters	0.45	0.17	0.15	0.45	0.37	0.18	0.40	0.47	0.14	0.21	0.48	0.40	0.33	0.31
		Reading	0.41	0.21	0.21	0.24	0.31	0.22	0.44	0.50	0.34	0.34	0.40	0.31	0.28	0.44
		Writing	0.39	0.19	0.20	0.21	0.34	0.15	0.45	0.54	0.35	0.32	0.43	0.35	0.31	0.49
	EGRA	Sound Recognition	0.40	0.20	0.18	0.28	0.29	0.19	0.42	0.46	0.26	0.26	0.41	0.36	0.25	0.36
		Word Recognition	0.45	0.16	0.14	0.40	0.40	0.23	0.38	0.53	0.18	0.25	0.48	0.44	0.41	0.35
		Paragraph Reading	0.41	0.15	0.15	0.34	0.34	0.23	0.37	0.50	0.19	0.24	0.46	0.38	0.40	0.36

Figure 14: Correlations between Sub-Tasks in Wave 1 and Wave 2 with Wave 3

	Wave 1		Wave 2		Wave 3
	A1	A2	B1	B2	C1
Intervention	0.405*** (0.099)	0.397*** (0.105)	0.243** (0.095)	0.287*** (0.093)	-0.087 (0.080)
Learner Age	0.168** (0.077)	0.181** (0.080)	0.148** (0.070)	0.175** (0.076)	0.129 (0.084)
Learner = Female	0.212*** (0.059)	0.185*** (0.058)	0.211*** (0.062)	0.192*** (0.059)	0.388*** (0.074)
N	937	823	912	800	778
R-squared	0.125	0.104	0.150	0.135	0.138
Same Teacher	x		x		

*0.1; **0.05; *** 0.01

Figure 20: Main Results Models

	LEARNER ASSESSMENT						EGRA			
	Listening Speaking Story	Listening Speaking Conversation	Phonological Awareness	Phonics Letters	Reading	Writing	Word Recognition	Paragraph Reading	Sound Recognition	Comprehension
Intervention	1.874*** (0.328)	1.241*** (0.428)	1.377** (0.602)	0.14 (0.311)	0.689** (0.342)	1.299*** (0.477)				
Learner Age	0.268 (0.255)	0.304 (0.256)	0.549 (0.344)	0.502** (0.188)	0.227 (0.238)	0.889** (0.409)				
Learner = Female	0.251 (0.192)	0.581** (0.242)	1.477*** (0.412)	0.337* (0.196)	0.352 (0.221)	0.836*** (0.221)				
N	937	937	937	937	937	937				
R-squared	0.103	0.073	0.169	0.053	0.04	0.126				
Intervention	0.138 (0.214)	0.224 (0.246)	0.529 (0.551)	0.580* (0.340)	-0.786 (0.553)	-1.084*** (0.355)	-2.696* (1.524)	-3.981** (1.493)	-0.573 (0.852)	-0.153 (0.098)
Learner Age	0.231 (0.152)	0.044 (0.134)	0.424 (0.582)	0.235 (0.320)	0.605 (0.587)	0.299 (0.340)	2.008 (1.391)	3.392** (1.449)	1.317 (1.029)	0.129* (0.075)
Learner = Female	0.21 (0.138)	0.520*** (0.150)	1.825*** (0.495)	0.713** (0.295)	2.810*** (0.596)	1.248*** (0.336)	5.976*** (1.420)	5.503*** (1.318)	4.512*** (0.821)	0.325*** (0.095)
N	779	779	779	779	779	779	779	779	779	778
R-squared	0.123	0.123	0.07	0.095	0.167	0.21	0.079	0.084	0.083	0.079

*0.1; **0.05; *** 0.01

Figure 215: Main Results per Sub-Task

	Wave 1		Wave 2		Wave 3	
	A1	A2	B1	B2	C1	C2
Intervention	0.405*** (0.099)	0.591*** (0.192)	0.243** (0.095)	-0.091 (0.218)	-0.087 (0.080)	0.069 (0.123)
Learner Age	0.168** (0.077)	0.153* (0.084)	0.148** (0.070)	0.116 (0.072)	0.129 (0.084)	0.135 (0.084)
Learner Gender = Female	0.212*** (0.059)	0.204*** (0.056)	0.211*** (0.062)	0.189*** (0.058)	0.388*** (0.074)	0.383*** (0.070)
N	937	937	912	912	778	778
R-squared	0.125	0.277	0.15	0.264	0.138	0.186
Including fieldworker dummies	Yes		Yes		Yes	

*0.1; **0.05; *** 0.01

Figure 22: Sensitivity Check - Fieldworker Effects

	Wave 3	
	(1)	(2)
Intervention	-0.087 (0.080)	-0.085 (0.086)
Learner Age	0.129 (0.084)	0.167* (0.099)
Learner Gender = Female	0.388*** (0.074)	0.315*** (0.091)
N	778	770
R-squared	0.138	0.173
Inverse Probability Weights	Yes	

*0.1; **0.05; *** 0.01

Figure 23: Sensitivity Check - Attrition Bias

	Wave 1		Wave 2		Wave 3	
	A1	A2	B1	B2	C1	C2
Intervention	0.405*** (0.099)	0.333*** (0.117)	0.243** (0.095)	0.215** (0.102)	-0.087 (0.080)	-0.171 (0.103)
Learner Age	0.168** (0.077)	0.171** (0.077)	0.148** (0.070)	0.150** (0.070)	0.129 (0.084)	0.132 (0.083)
Learner Gender = Female	0.212*** (0.059)	0.135 (0.084)	0.211*** (0.062)	0.181** (0.078)	0.388*** (0.074)	0.296** (0.114)
Learner Gender = Female x Intervention		0.148 (0.121)		0.057 (0.097)		0.172 (0.144)
N	937	937	912	912	778	778
R-squared	0.125	0.126	0.15	0.15	0.138	0.14

*0.1; **0.05; *** 0.01

Figure 16: Heterogeneous Effects - Learner Gender

	Wave 1	Wave 2	Wave 3
Intervention	0.479*** (0.160)	0.107 (0.181)	0.121 (0.197)
Learner Age	0.156** (0.074)	0.150** (0.068)	0.108 (0.079)
Learner Gender = Female	0.216*** (0.058)	0.215*** (0.061)	0.392*** (0.072)
LoLT = Afrikaans	0.159 (0.228)	-0.289 (0.190)	0.193 (0.245)
LoLT = isiXhosa	0.564** (0.247)	0.011 (0.221)	-0.016 (0.286)
LoLT = Afrikaans x Intervention	-0.111 (0.210)	0.138 (0.225)	-0.31 (0.219)
LoLT = isiXhosa x Intervention	-0.06 (0.250)	0.167 (0.262)	-0.105 (0.246)
N	937	912	813
R-squared	0.138	0.147	0.128

*0.1; **0.05; *** 0.01

Figure 25: Heterogeneous Effects - Language of Learning and Teaching

	Wave 1	Wave 2	Wave 3
Intervention	0.267** (0.123)	0.191 (0.158)	-0.013 (0.121)
Learner Age	0.186** (0.078)	0.155** (0.068)	0.099 (0.082)
Learner Gender = Female	0.226*** (0.059)	0.230*** (0.065)	0.400*** (0.072)
Quintile 1 - 3	-0.475** (0.193)	-0.354 (0.220)	-0.338* (0.185)
Quintile 1 - 3 x Intervention	0.224 (0.185)	0.075 (0.203)	-0.157 (0.156)
N	937	912	813
R-squared	0.103	0.064	0.088

*0.1; **0.05; *** 0.01

Figure 26: Heterogeneous Effects - School Quintile

CHECK OF ASSUMPTIONS

Source: Van Wyk, Gondwe, & De Villiers, 2017

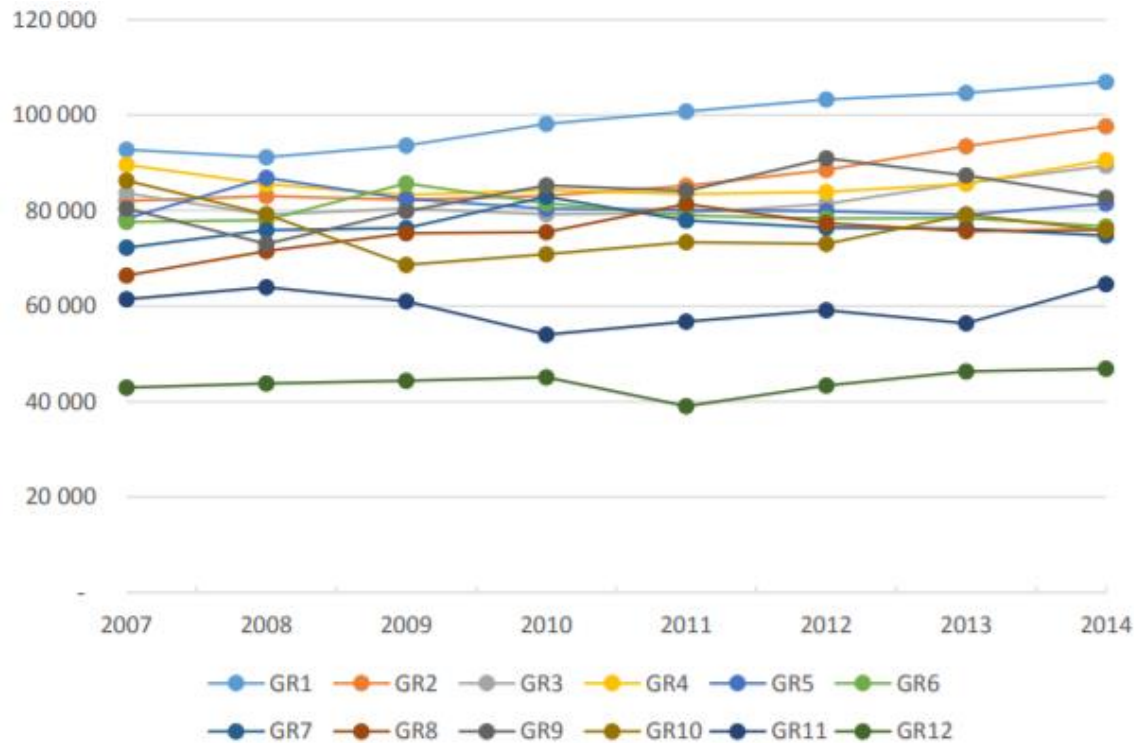


Figure 27: Enrolment in Public Schools in the Western Cape by Year and Grade, 2007-14.

Year	Grade R	Grade 1	% increase in Grade R	% increase in Grade 1
2014	65085	110815		
2015	65569	111856	0.7%	0.9%
2016	68138	110945	3.9%	-0.8%

Figure 17: Enrolment in the Western Cape by Year and Grade, 2014 – 2016, SNAP Survey

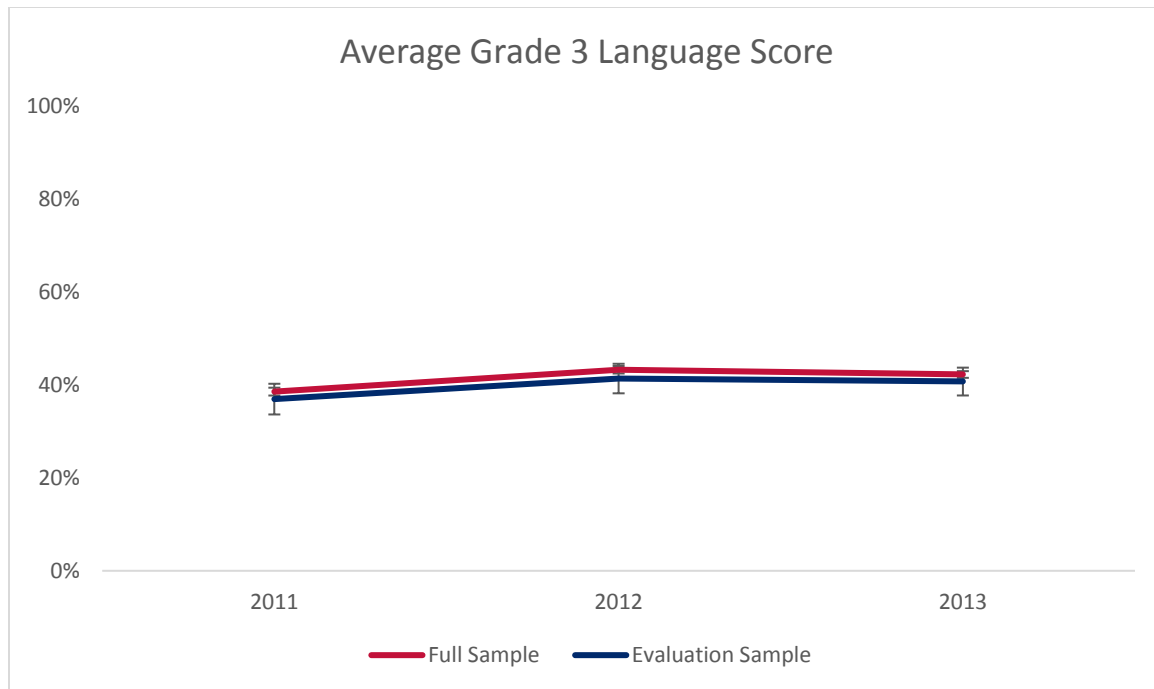


Figure 18: Average Grade 3 Literacy Score in the Western Cape Systemic Evaluation

Performance on Learner Assessment Against Expected Outcomes

					COMPARISON GROUP									INTERVENTION GROUP								
					Wave 1			Wave 2			Wave 3			Wave 1			Wave 2			Wave 3		
					Unable	Below Grade Level	Reached or Exceeded	Unable	Below Grade Level	Reached or Exceeded	Unable	Below Grade Level	Reached or Exceeded	Unable	Below Grade Level	Reached or Exceeded	Unable	Below Grade Level	Reached or Exceeded	Unable	Below Grade Level	Reached or Exceeded
Listening & Speaking Conversation	B1	Gr R	1	Knows name	1%	NA	99%	0%	NA	100%	0%	NA	100%	0%	NA	100%	0%	NA	100%	0%	NA	100%
	B3	Gr R	2	Responds to closed question	4%	8%	87%	1%	5%	94%	2%	1%	97%	13%	5%	82%	4%	4%	93%	3%	1%	96%
	B4	Gr R	2	Complex question and answer (full sentence)	3%	33%	64%	2%	23%	76%	1%	29%	70%	3%	24%	73%	1%	32%	67%	0%	28%	72%
	B5	Gr 1	4	Responds to open question	5%	32%	63%	1%	25%	73%	0%	44%	56%	2%	33%	64%	1%	40%	59%	0%	33%	67%
	B2	Gr 1	4	Knows friends' names (recognises others)	2%	9%	89%	1%	2%	97%	0%	2%	98%	2%	6%	92%	0%	3%	97%	1%	1%	98%
	B6	Gr 2	5	Suggests topics (gives enough detail for a conversation)	47%	NA	53%	6%	26%	68%	2%	55%	43%	23%	NA	77%	2%	24%	74%	2%	58%	40%
Listening & Speaking Story	B10	Gr R	1	Points out objects in a picture	4%	NA	96%	2%	NA	98%	0%	NA	100%	1%	NA	99%	3%	NA	97%	1%	NA	99%
	B9	Gr R	2	Finds answers in a book	6%	28%	66%	4%	14%	82%	6%	8%	86%	5%	18%	77%	7%	15%	78%	8%	6%	86%
	B11	Gr R	2	Recalls part of a story (recognises characters)	14%	28%	58%	19%	11%	70%	59%	24%	18%	12%	14%	74%	19%	14%	67%	47%	32%	21%
	B26	Gr 1	4	Draws a picture to show understanding	16%	36%	48%	11%	18%	71%	0%	6%	94%	5%	18%	78%	1%	14%	85%	1%	5%	94%
	B12	Gr 1	4	Matches caption to picture	4%	15%	81%	18%	21%	61%	8%	7%	85%	3%	5%	92%	24%	15%	60%	8%	13%	80%
	B7	Gr 2	5	Remembers the sequence of a story	43%	NA	57%	16%	47%	38%	8%	63%	29%	21%	NA	79%	7%	41%	52%	6%	58%	36%

	Question #	Grade Level	Expected Score	Item	COMPARISON GROUP									INTERVENTION GROUP								
					Wave 1			Wave 2			Wave 3			Wave 1			Wave 2			Wave 3		
					Unable	Below Grade Level	Reached or Exceeded	Unable	Below Grade Level	Reached or Exceeded	Unable	Below Grade Level	Reached or Exceeded	Unable	Below Grade Level	Reached or Exceeded	Unable	Below Grade Level	Reached or Exceeded	Unable	Below Grade Level	Reached or Exceeded
Phonological Awareness	B14	Gr R	1	Sounds make up words	60%	NA	40%	62%	NA	38%	14%	NA	86%	64%	NA	36%	62%	NA	38%	13%	NA	87%
	B13	Gr R	2	Different letter sound	14%	18%	68%	9%	14%	77%	3%	11%	86%	7%	22%	71%	5%	20%	76%	3%	6%	92%
	B18	Gr R	2	Odd word in sequence	82%	9%	8%	76%	10%	14%	40%	12%	48%	70%	11%	19%	68%	11%	21%	26%	20%	54%
	B17	Gr 1	3	Difference between starting initials of two words	59%	NA	41%	46%	NA	54%	16%	NA	84%	54%	NA	46%	36%	NA	64%	21%	NA	79%
	B15	Gr 1	4	Substitute vowels	84%	6%	10%	79%	14%	6%	29%	8%	63%	86%	7%	7%	83%	10%	7%	32%	7%	61%
	B16	Gr 1	4	Builds words using sounds	79%	16%	5%	76%	15%	10%	48%	27%	25%	79%	16%	6%	76%	17%	7%	37%	35%	28%
Phonics Letters	B19	Gr R	1	Sentences made up of words	24%	NA	76%	15%	NA	85%	1%	NA	99%	12%	NA	88%	4%	NA	96%	2%	NA	98%
	B21	Gr R	2	Recognises some consonants	28%	56%	17%	34%	34%	32%	2%	28%	70%	22%	58%	20%	31%	37%	32%	2%	25%	73%
	B22	Gr R	2	Points to letter starting a word	63%	25%	12%	56%	25%	20%	4%	11%	85%	57%	26%	18%	46%	31%	23%	4%	16%	80%
	B20	Gr 1	4	Letter blends	67%	13%	20%	62%	12%	26%	17%	6%	77%	64%	6%	30%	59%	7%	33%	14%	8%	78%
	B23	Gr 1	4	Recognises 2 consonants at start and end of word	73%	18%	8%	82%	15%	3%	63%	20%	16%	69%	22%	8%	83%	15%	2%	57%	27%	16%
	B24	Gr 2	5	Recognises double vowels & 3 consonant strings	69%	28%	3%	86%	13%	1%	36%	50%	14%	74%	23%	2%	78%	22%	0%	25%	56%	19%
Reading	B29	Gr R	1	"Reading" voice	82%	NA	18%	78%	NA	22%	16%	NA	84%	86%	NA	14%	67%	NA	33%	18%	NA	82%
	B30	Gr R	2	Points to words while reading	85%	10%	4%	87%	8%	5%	16%	14%	70%	90%	6%	4%	78%	15%	7%	18%	13%	69%
	B25	Gr 1	4	Recognises sight words	88%	9%	3%	81%	11%	7%	13%	12%	75%	91%	6%	2%	78%	15%	7%	14%	15%	71%

					COMPARISON GROUP									INTERVENTION GROUP								
					Wave 1			Wave 2			Wave 3			Wave 1			Wave 2			Wave 3		
	Question #	Grade Level	Expected Score	Item	Unable	Below Grade Level	Reached or Exceeded	Unable	Below Grade Level	Reached or Exceeded	Unable	Below Grade Level	Reached or Exceeded	Unable	Below Grade Level	Reached or Exceeded	Unable	Below Grade Level	Reached or Exceeded	Unable	Below Grade Level	Reached or Exceeded
	B31	Gr 2	5	Reads actual words	100%	NA	0%	89%	10%	1%	22%	18%	60%	98%	NA	2%	90%	10%	0%	25%	15%	61%
	B8	Gr R	1	Handles a book correctly	7%	NA	93%	6%	NA	94%	3%	NA	97%	3%	NA	97%	4%	NA	96%	3%	NA	97%
	B28	Gr 1	3	Predicts storyline	20%	NA	80%	16%	NA	84%	3%	NA	97%	14%	NA	86%	9%	NA	91%	6%	NA	94%
Writing	A1	Gr R	2	Writes own name	10%	14%	76%	5%	14%	80%	0%	4%	96%	9%	17%	74%	6%	14%	80%	1%	7%	92%
	A2	Gr 1	4	Write about story	76%	20%	4%	77%	21%	3%	10%	32%	57%	78%	21%	1%	43%	56%	1%	0%	40%	54%
	A3	Gr 1	4	Copy sentence	12%	64%	24%	5%	64%	31%	0%	22%	77%	4%	77%	24%	1%	73%	26%	0%	27%	73%
	A4	Gr 2	5	Write sentence correctly	62%	37%	0%	70%	30%	0%	3%	86%	11%	60%	40%	0%	36%	64%	0%	3%	81%	16%
	A5	Gr R	2	Copy pattern	14%	37%	49%	9%	33%	58%	1%	13%	86%	6%	33%	61%	3%	31%	66%	2%	18%	80%
	B27	Gr R	2	Reads own writing	93%	6%	1%	93%	5%	2%	19%	12%	69%	78%	16%	6%	62%	17%	20%	17%	20%	63%

Figure 30: Performance on Learner Assessment - Expected Criterion Levels (CAPS)

* Table highlights everything 60% and above.

- Red means "unable"
- Orange means "below Grade level"
- Green means "reached or exceeded Grade level"

Sustainability Assessment with Subject Advisors

Q1. Environmental Support: Having a supportive internal and external climate for your programme

(n.s. $p > 0.05$)

	2016	2017
The programme has leadership support from within the larger organization (WCED).	5.29	5.79
The programme has strong champions with the ability to gather/garner resources.	5.58	5.47
Champions exist who strongly support the programme	5.35	5.47
The programme has leadership support from outside of the organization.	4.42	4.63
The programme has strong public support.	3.48	3.37

Figure 31: Sustainability: Environmental Support

Q2. Funding Stability: Establishing a consistent financial base for your programme

(n.s. $p > 0.05$)

	2016	2017
The programme is funded through a variety of sources.	4.1	4.16
The programme implements policies to help ensure sustained funding.	3.87	4.11
The programme exists in a supportive economic climate.	4.03	3.89
The programme has a combination of stable and flexible funding.	3.77	3.84
The programme has sustained funding.	3.65	3.68

Figure 32: Sustainability: Funding Stability

Q3. Partnerships: Cultivating connections between the programme and its stakeholders

(n.s. $p > 0.05$)

	2016	2017
Diverse community organizations are invested in the success of the programme.	3.87	4.16
The programme communicates with community leaders.	2.94	3.68
Community leaders are involved with the programme.	2.97	2.89
Community members are passionately committed to the programme.	3.1	2.89
The community is engaged in the development of programme goals.	2.84	2.89

Figure 33: Sustainability: Partnerships

Q4. Organizational Capacity: Having the internal support and resources needed to effectively manage the programme and its activities

(n.s. $p > 0.05$)

	2016	2017
The programme is well integrated into the operations of the organization (WCED and/or Education District Office).	6.03	6.21
Organizational systems are in place to support the various programme needs.	5.71	5.89
Leadership efficiently manages staff and other resources.	5.13	5.89
Leadership effectively articulates the vision of the programme to external partners.	5.03	5.37
The programme has adequate staff to complete the programme's goals.	5.1	5.05

Figure 34: Sustainability: Organizational Capacity

Q5. Programme Evaluation: Assessing the programme to inform planning and document results

(n.s. $p > 0.05$)

	2016	2017
The programme has the capacity for quality programme evaluation.	5.61	5.89
The programme reports short term and intermediate outcomes.	5.32	5.74
Evaluation results inform programme planning and implementation.	4.77	5.37
The programme provides strong evidence to the public that the programme works.	4.45	5.32
Programme evaluation results are used to demonstrate successes to funders and other key stakeholders.	4.06	4.79

Figure 19: Sustainability: Program Evaluation

Q6. Programme Adaptation: Taking actions that adapt the programme to ensure its ongoing effectiveness

(n.s. $p > 0.05$)

	2016	2017
The programme adapts strategies as needed.	4.61	4.26
The programme proactively adapts to changes in the environment.	4.55	4.21
The programme adapts to new evidence.	4.39	4.16
The programme periodically reviews the evidence base.	4.39	4
The programme makes decisions about which components are ineffective and should not continue.	3.77	3.26

Figure 20: Sustainability: Program Adaptation

Q7. Communications: Strategic communication with stakeholders and the public about the programme

(n.s. $p > 0.05$)

	2016	2017
The programme is marketed in a way that generates interest.	4	4.53
The programme has communication strategies to secure and maintain public support.	3.71	4.37
The programme demonstrates its value to the public.	3.77	4.11
The programme increases community awareness of the issue.	3.65	4
Programme staff communicate the need for the programme to the public.	3.29	3.68

Figure 21: Sustainability: Communications

Q8. Strategic Planning: Using processes that guide the programme's direction, goals, and strategies

(n.s. $p > 0.05$)

	2016	2017
The programme's goals are understood by all stakeholders.	5.03	5.16
The programme clearly outlines roles and responsibilities for all stakeholders.	4.9	5.11
The programme has a sustainability plan.	4.77	4.37
The programme plans for future resource needs.	3.87	4.16
The programme has a long-term financial plan.	2.26	3.26

Figure 22: Sustainability: Strategic Planning

Pacing and Curriculum Coverage Assessment with Grade 1 Teachers

Note: in this section, the cutoff for significance was 0.05.



Figure 23: Gr 1 Pacing and Curriculum Coverage: Resources/Materials Used

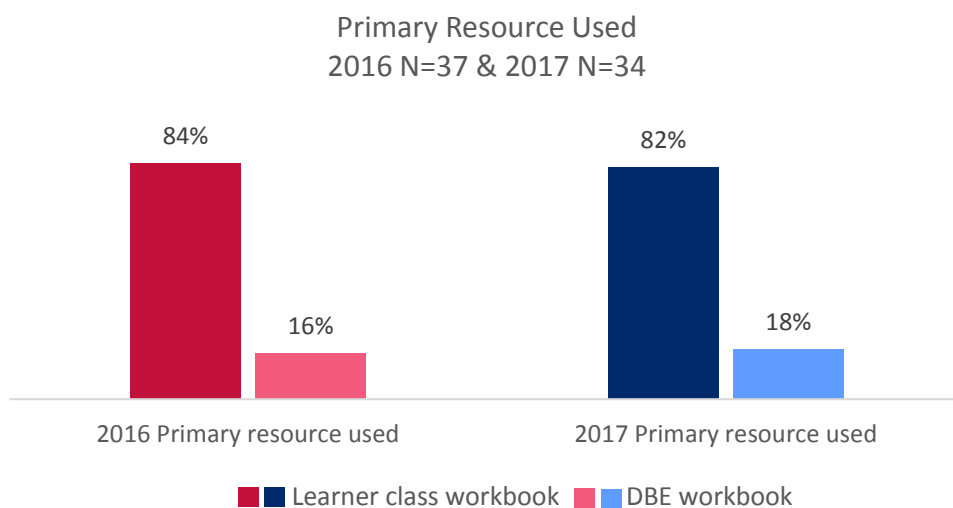


Figure 40: Gr 1 Pacing and Curriculum Coverage: Primary Resource Used

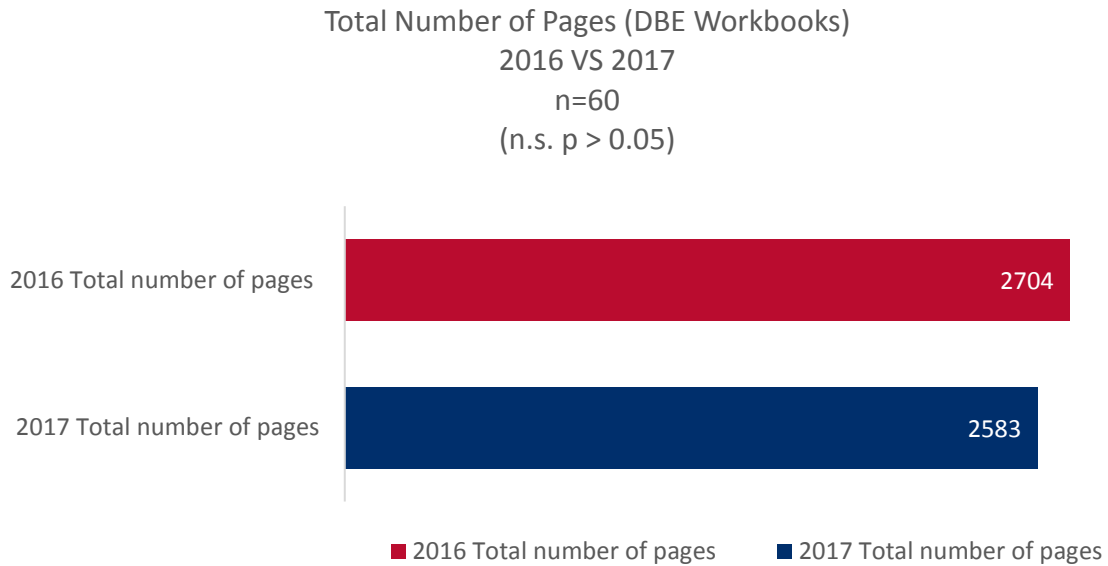


Figure 41: Gr 1 Pacing and Curriculum Coverage: Pages Written Work (DBE Workbooks)

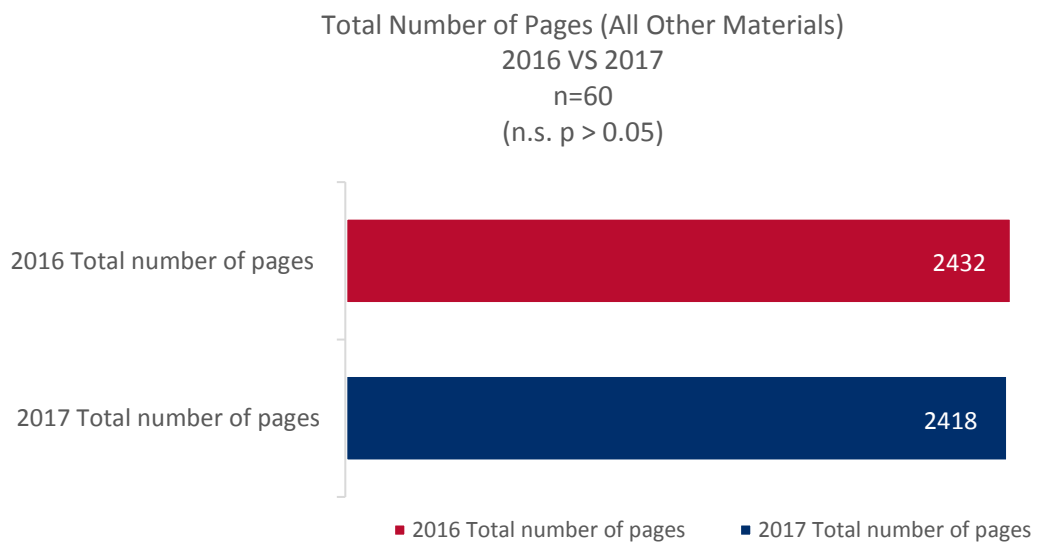


Figure 42: Gr 1 Pacing and Curriculum Coverage: Pages Written Work (Other Materials)

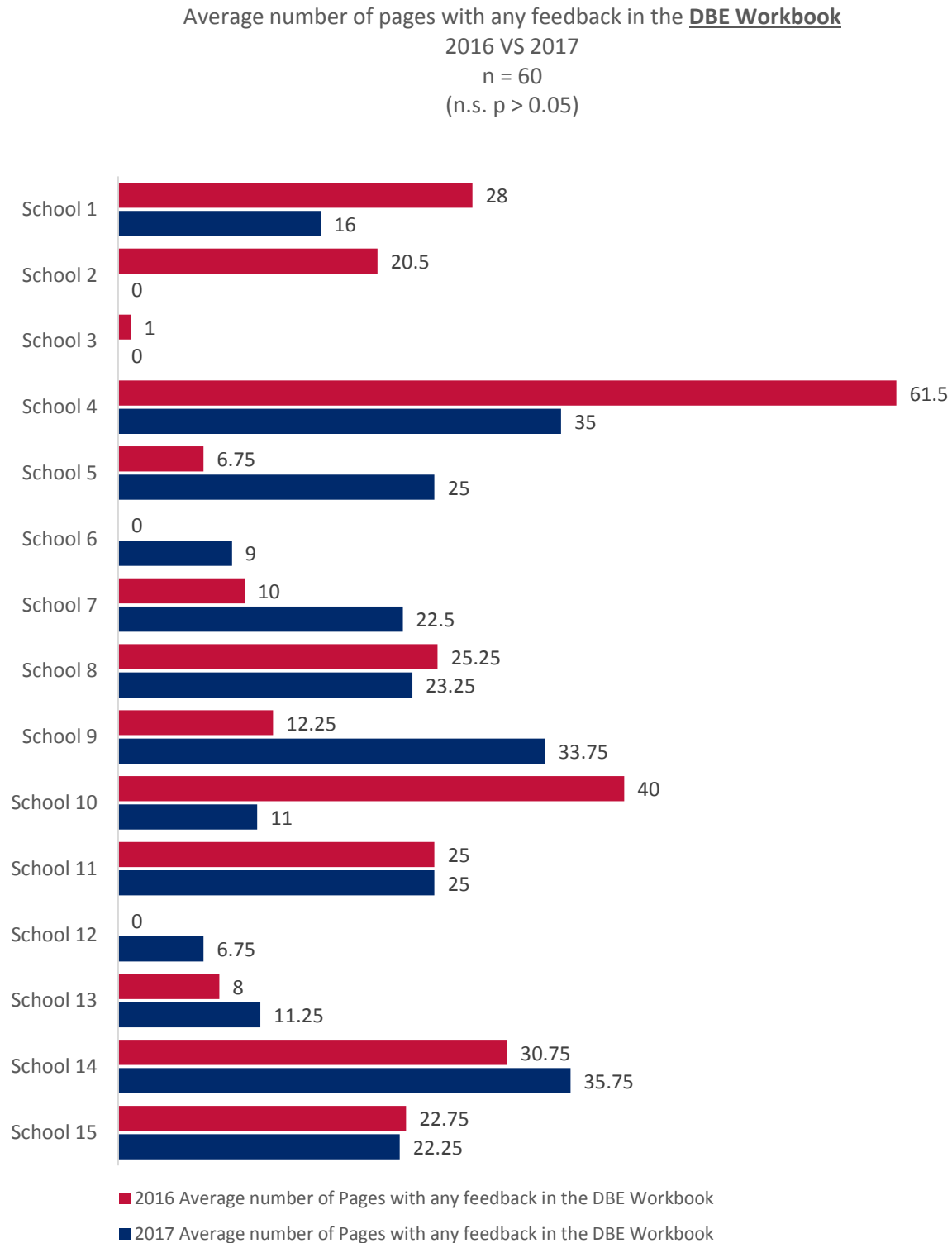


Figure 43: Gr 1 Pacing and Curriculum Coverage: Teacher Feedback (DBE Workbooks)

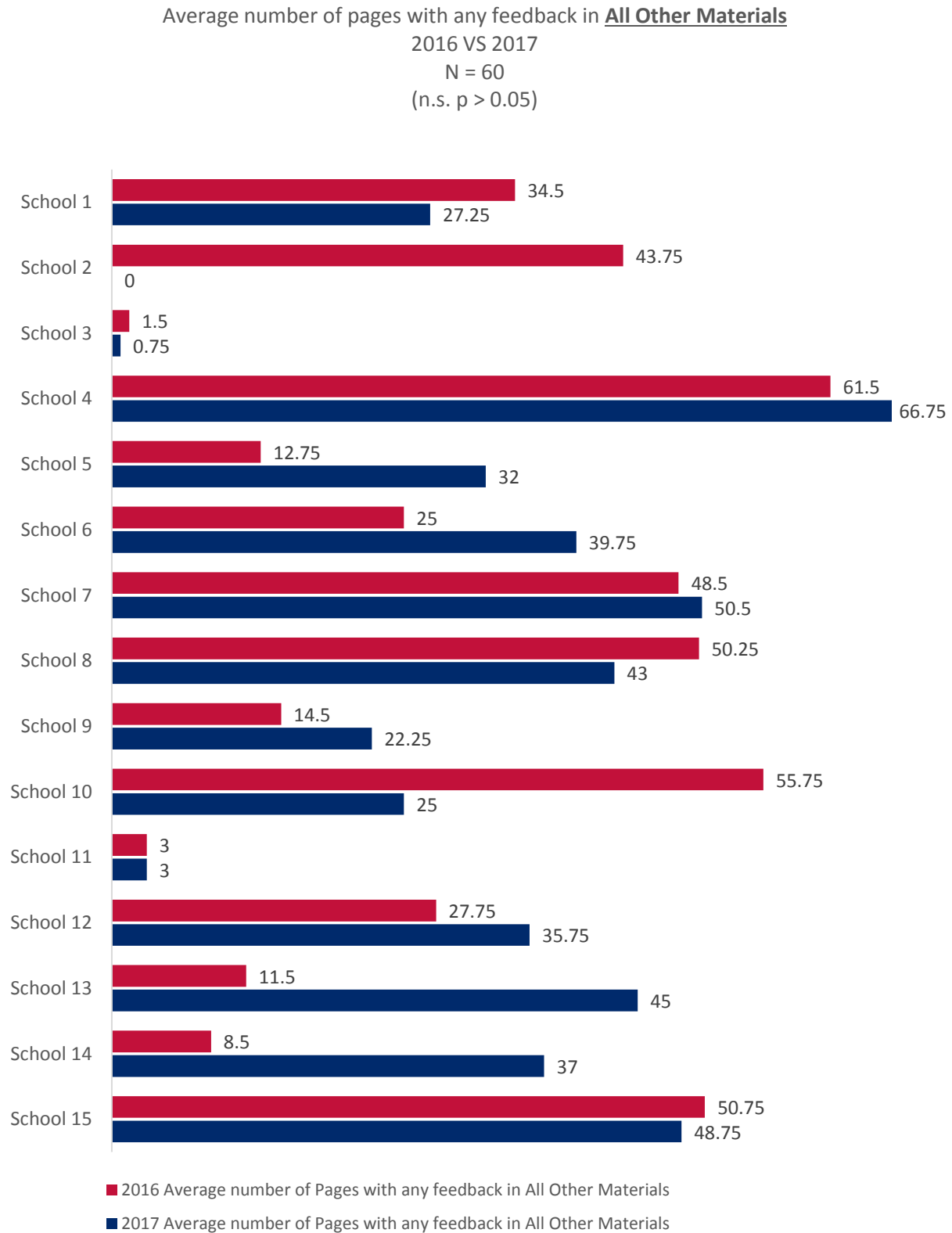


Figure 24: Gr 1 Pacing and Curriculum Coverage: Teacher Feedback (Other Materials)

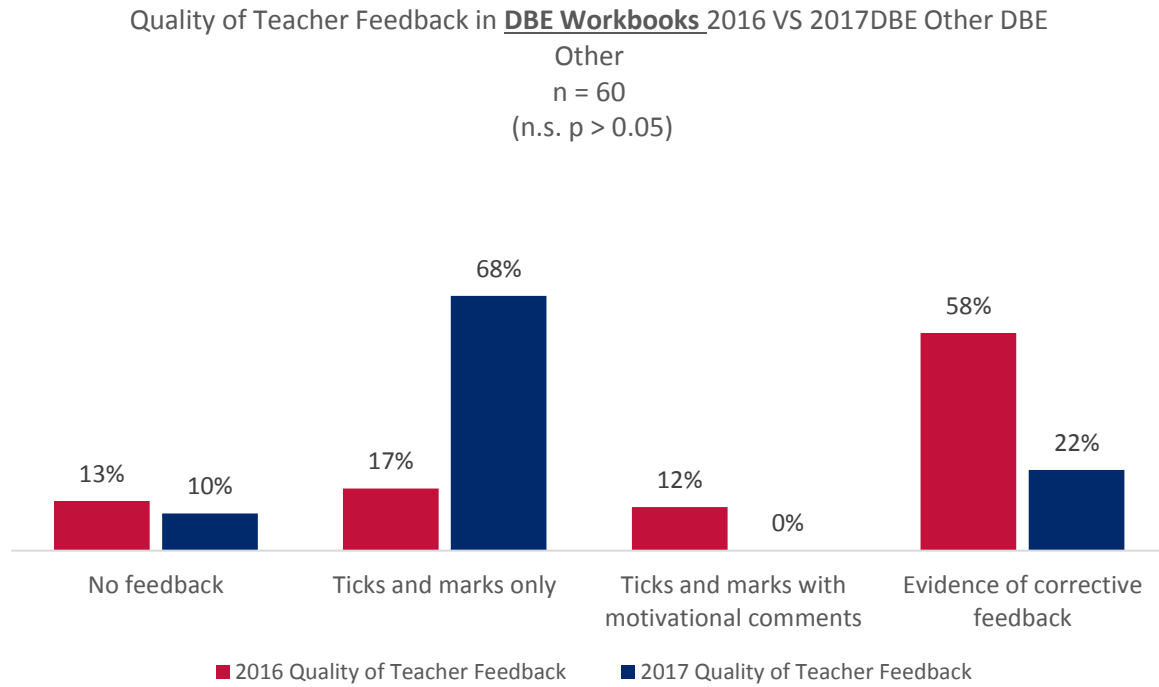


Figure 25: Gr 1 Pacing and Curriculum Coverage: Quality of Feedback (DBE Workbooks)

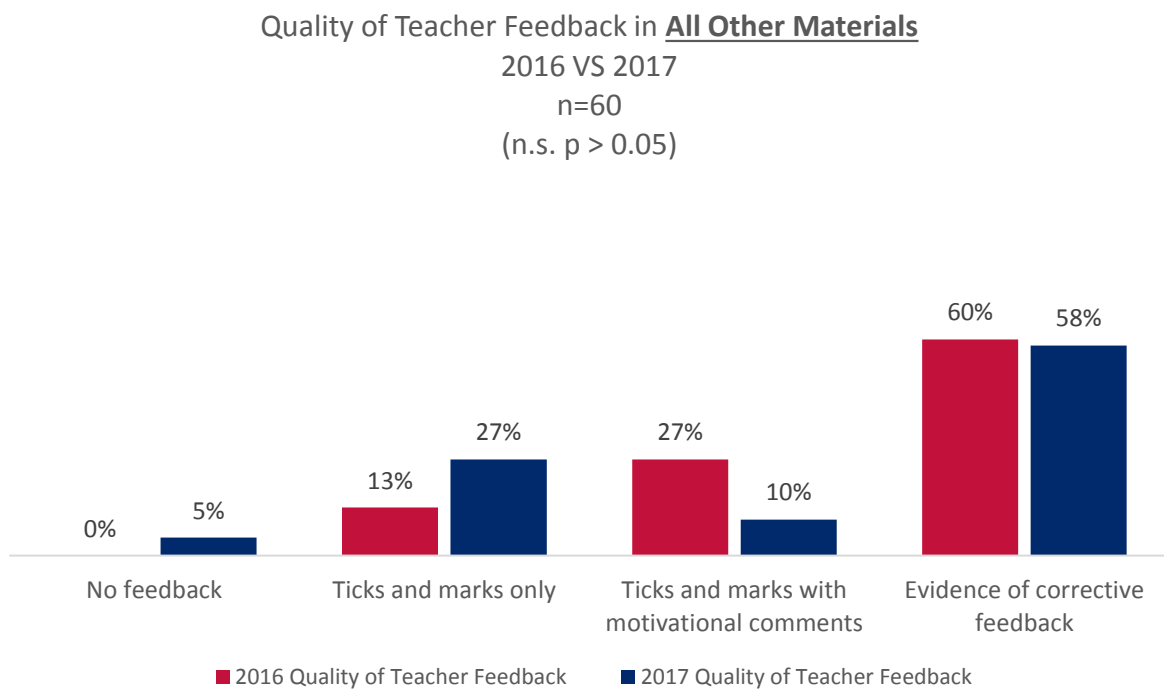


Figure 26: Gr 1 Pacing and Curriculum Coverage: Quality of Feedback (All Other Materials)

Quality of Writing in **DBE Workbooks**
 2016 VS 2017
 n=60
 (n.s. $p > 0.05$)

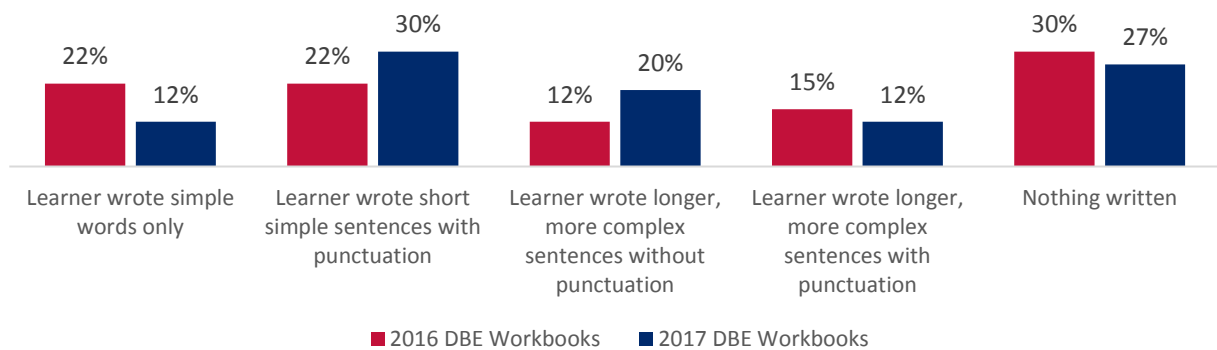


Figure 27: Gr 1 Pacing and Curriculum Coverage: Quality of Writing (DBE Workbooks)

Quality of Writing in **All Other Materials**
 2016 VS 2017
 n=60
 (n.s. $p > 0.05$)

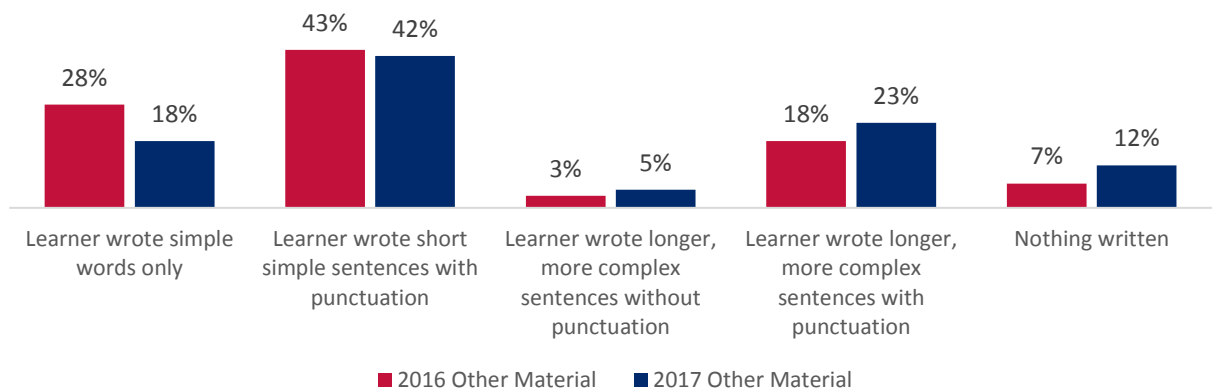


Figure 28: Gr 1 Pacing and Curriculum Coverage: Quality of Writing (All Other Materials)

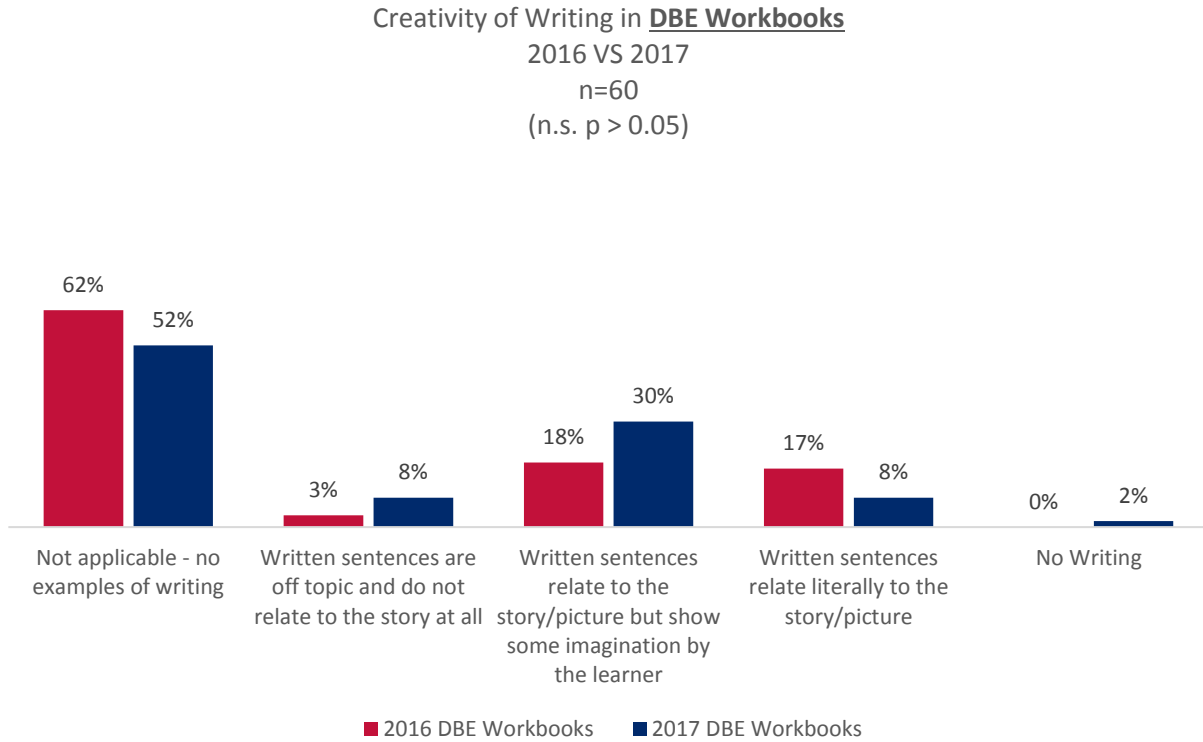


Figure 29: Gr 1 Pacing and Curriculum Coverage: Creativity of Writing (DBE Workbooks)

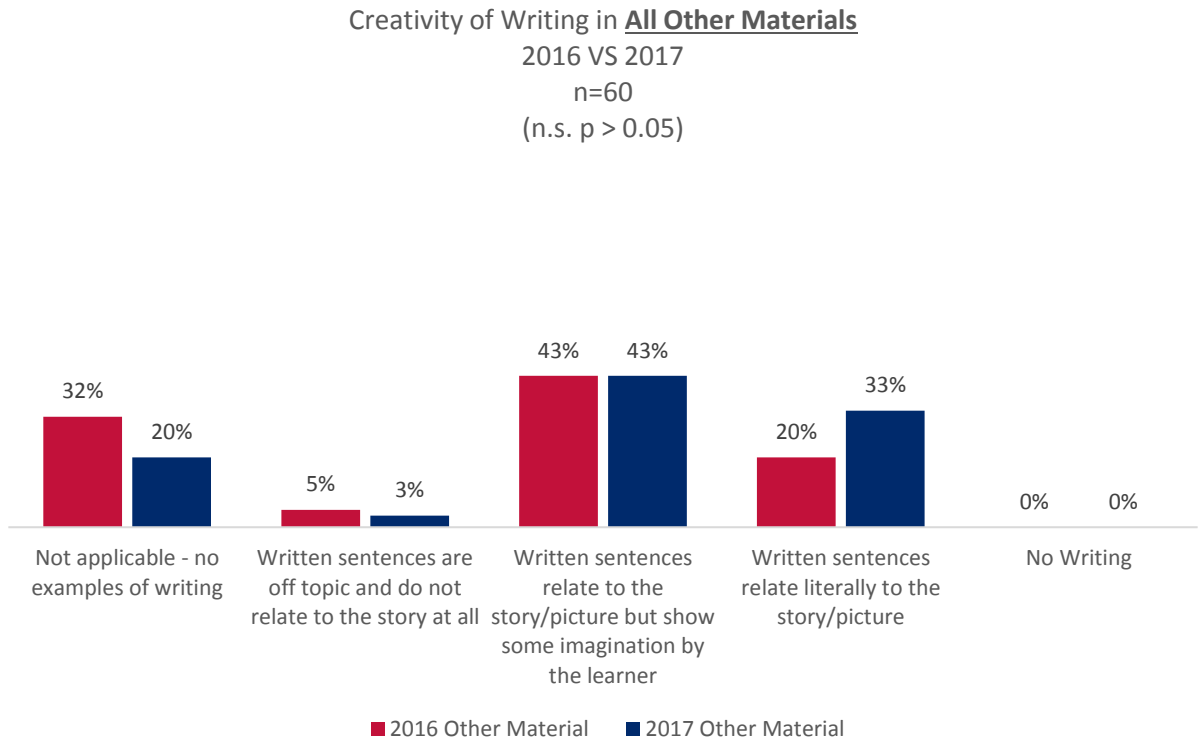


Figure 30: Gr 1 Pacing and Curriculum Coverage: Creativity of Writing (All Other Materials)

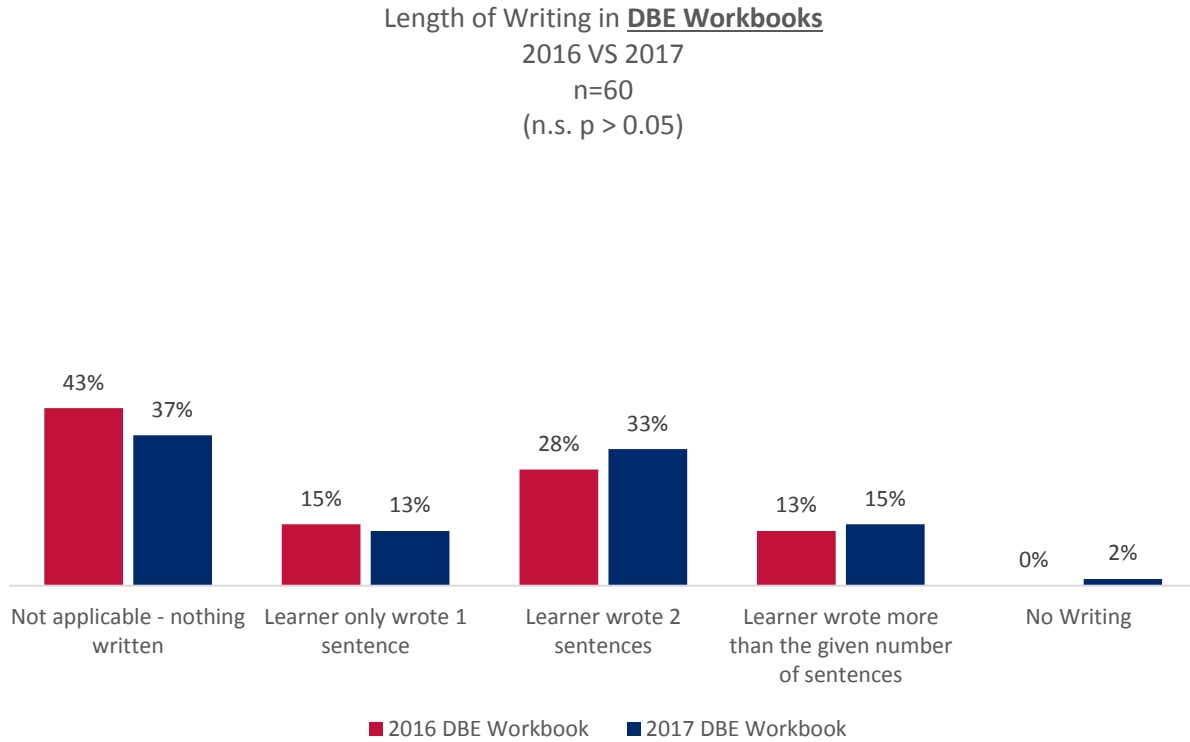


Figure 31: Gr 1 Pacing and Curriculum Coverage: Length of Writing (DBE Workbooks)

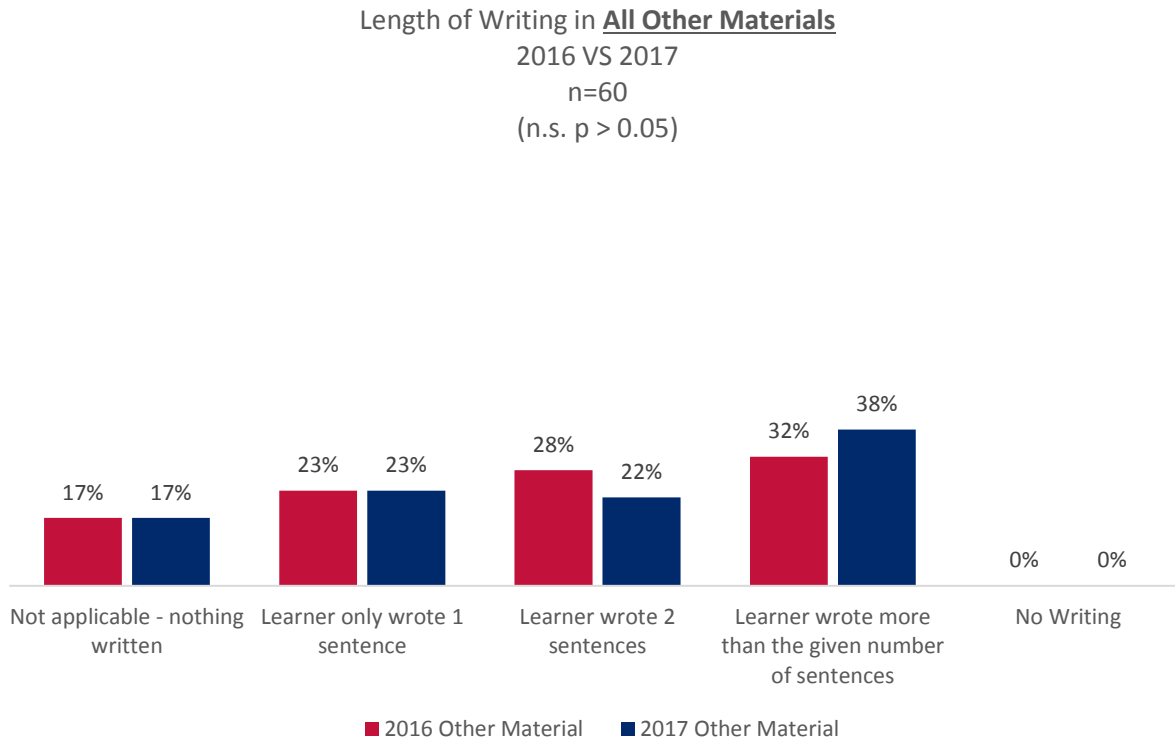


Figure 52: Gr 1 Pacing and Curriculum Coverage: Length of Writing (All Other Materials)

Classroom Environment Assessment with Grade R Teachers

Note: in this section, the cutoff for significance was 0.05.

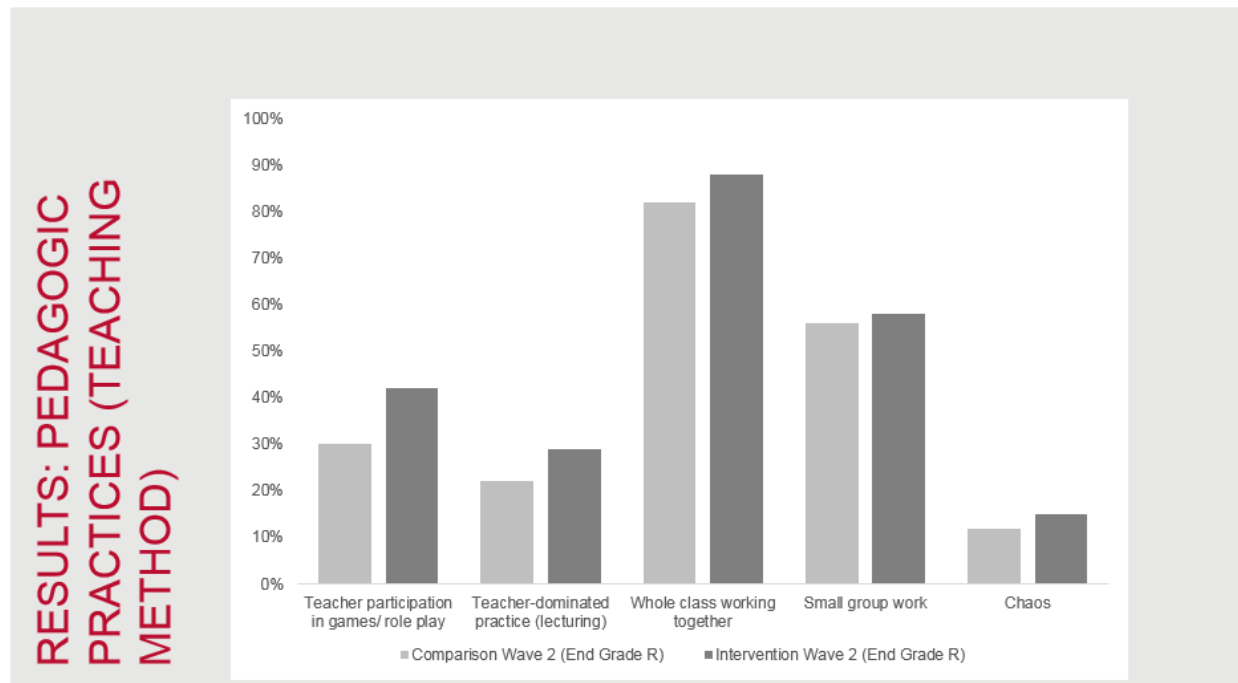


Figure 32: Classroom Environment (Teaching Method)

P < 0.05; *P < 0.01

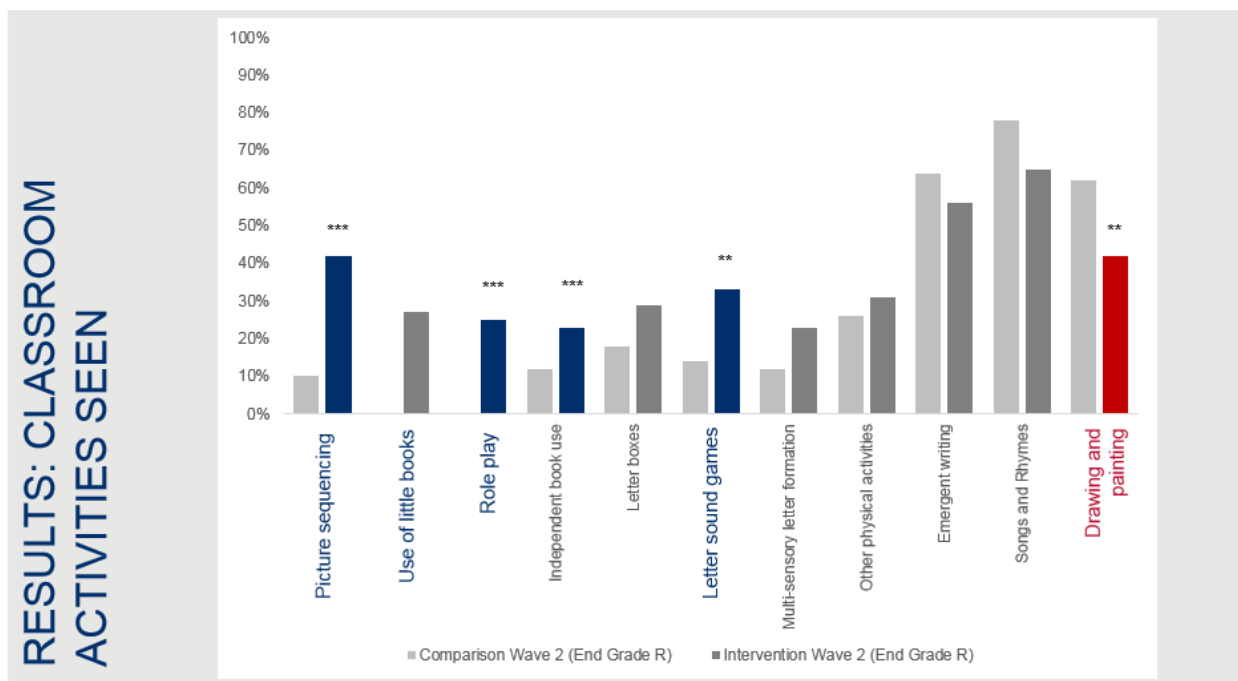


Figure 33: Classroom Environment (Activities Seen)

P < 0.05; *P < 0.01

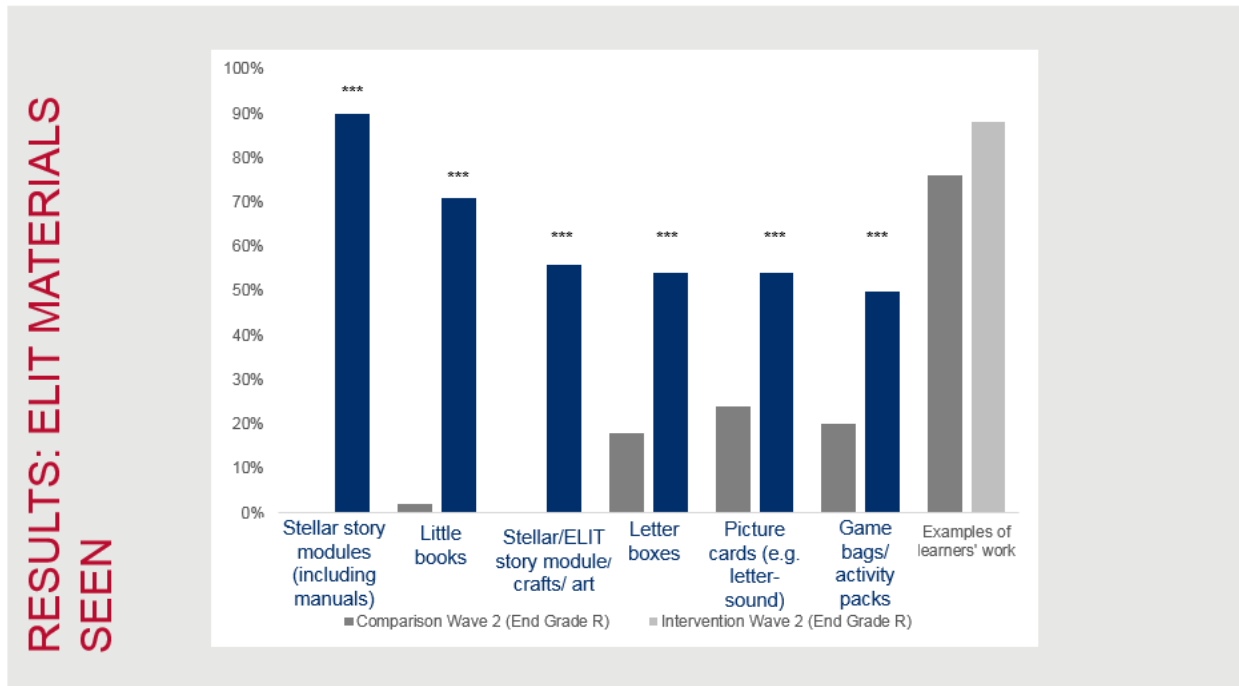


Figure 34: Classroom Environment (ELIT Materials Seen)

P < 0.05; *P < 0.01

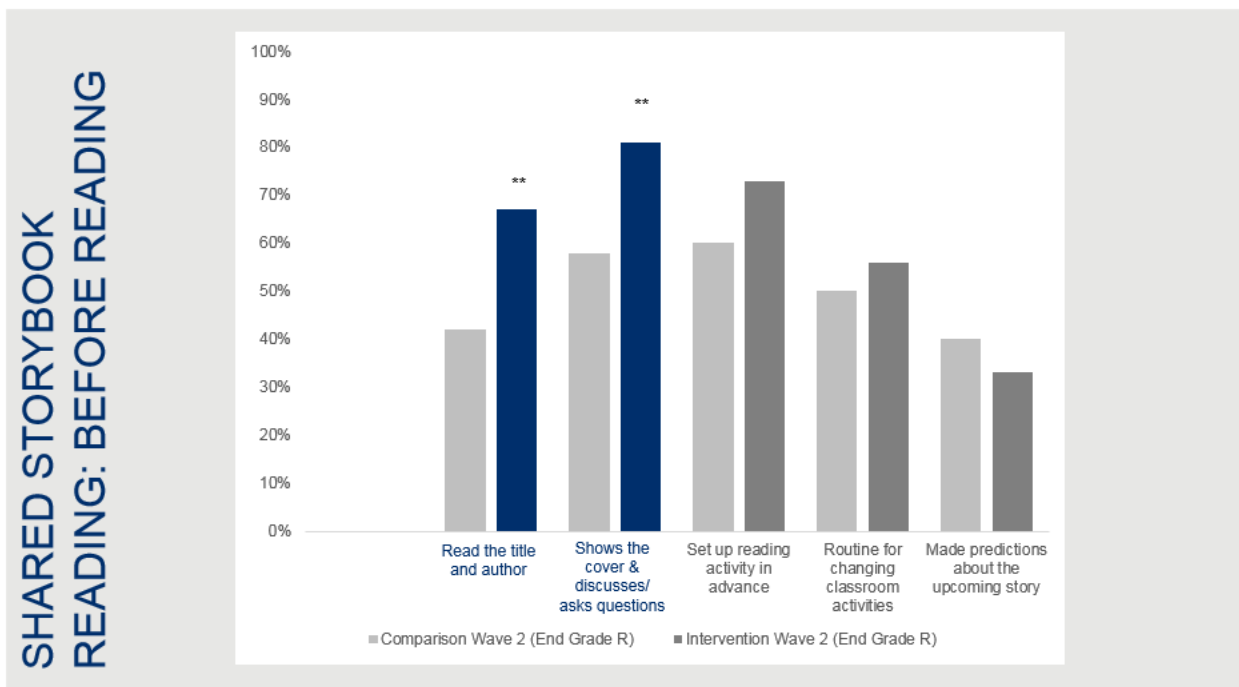


Figure 35: Classroom Environment (Before Reading)

P < 0.05; *P < 0.01

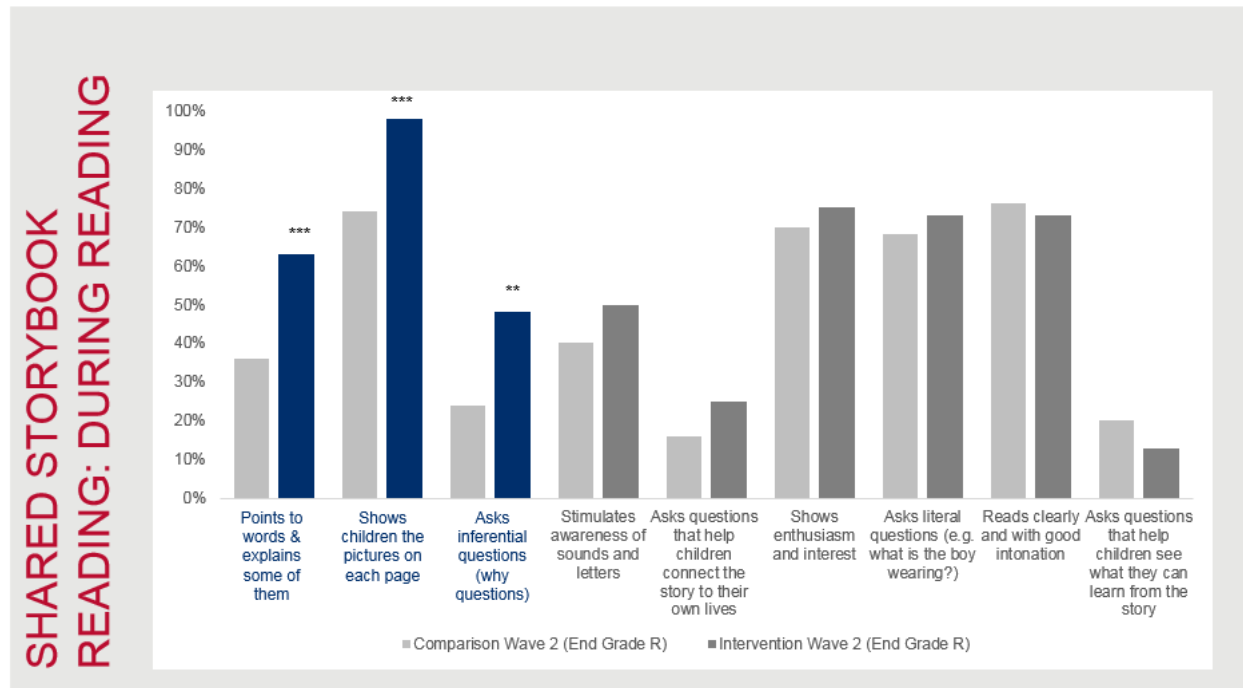


Figure 36: Classroom Environment (During Reading)

P < 0.05; *P < 0.01

ANNEX 2: METHODOLOGY REPORT EXTRACT

Note: Annex 2 is an extract from Khulisa's Methodology and Study Protocol submitted to USAID in July 2015.

Overview

The study will be conducted in the Western Cape Province of South Africa. The **settings** within which the research will take place include schools (Grade R classrooms), districts (teacher training venues), and province (provincial offices). The **population** of interest includes teachers of Grade R, trainers of Grade R teachers, and Grade R and Grade 1 learners in the Western Cape Province. It is expected that the study findings will be directly **applicable** to the program funders, implementers, and other stakeholders including USAID Southern Africa, Wordworks, the ELMA Foundation, JP Morgan, the WCED, and the Department of Basic Education (DBE). The information generated through this evaluation will add to the evidence base for policy decisions about the most effective way to improve early language and literacy in South Africa. Khulisa received **authorization** from the WCED in July 2015 to conduct the evaluation.

Evaluation Design

Since the award of the evaluation has been delayed, Khulisa's assumption that there would be sufficient time for learning to take place (on which the sample sizes were estimated) has been affected. Khulisa considered a number of **alternative approaches**, given the limitations to the original design. Following all considerations, Khulisa believes that it is outside of the evaluation team's scope to request a change to the roll-out of the Stellar program and therefore will continue with the original design as planned.

The 2015 cohort of teachers and learners is the only chance for the evaluation team to capture a comparison group in line with the contractual requirement to collect "comparison group data on language and literacy skills from a representative sample of the current 2015 Grade R learners", given the current roll-out plans for the STELLAR program as documented in the evaluation Work Plan⁶¹. Based on the current rollout plan, and in line with the research design as proposed, the experiment design is as follows:

Group	Cohort	When Tested
Comparison Group	2015 Grade R Cohort	<i>In Grade R:</i> Start of 3 rd Term 2015 End of year 2015* <i>In Grade 1:</i> End of year 2016*
Intervention group (learners in the same teacher's class who was tested in 2015)	2016 Grade R Cohort	<i>In Grade R:</i> Start of 3 rd Term 2016 End of year 2016* <i>In Grade 1:</i> End of year 2017*

⁶¹ Khulisa assumes that the roll-out of the Stellar program will not change to accommodate an alternate research design, given the WCED's requirement that all teachers in the Western Cape be trained at the same point in time.*It is important for the design that sufficient time passes between the baseline and end-of-year points of data collection to allow for learning to take place. Khulisa assumes that the WCED will be willing to lift their research restriction in the fourth quarter of the year to allow for this.

One of the things that must be noted in regard to continuing with this research design is that the evaluation is going to get intermediate results from which we may not get strong effects. In other words, the short-term effects of the intervention may not be as robust as initially expected, due to the changes that have occurred since the evaluation's inception. The research design will, on the other hand, be able to answer questions around long-term language and literacy gains.

Study Population

For the Grade R learner study, the population from which our sample will be drawn includes all Grade R learners at primary schools that have Grade R classes in the Western Cape. Grade R learners are officially required to be four years of age, turning five by 30 June in the year of their admission. The Actuarial Society of South Africa (ASSA)'s mid-year estimates for the number of 5 year olds in the Western Cape for 2015 are 96,303⁶². From the 2013 General Household Survey, it is evident that 55% of 4 – 6 year olds in the Western Cape are of Coloured ethnicity, 33% are of Black ethnicity, 12% are of White ethnicity, and 0.4% are of Indian or Asian ethnicity. Furthermore, there is not a large gender difference, with 51% being boys and 49% being girls. Approximately 46% of learners come from families that receive a child support grant, and on average children in the bottom income quintile come from homes with an average monthly income of R1, 120. For the teacher observation study, all Grade R teachers in schools selected for the study will be sampled.

Schools will be eligible for inclusion in the study under the following conditions:

- a. Are within a 2-hour driving time from where fieldworkers are located, so that fieldworkers are able to reach the school by 8am.
- b. Are primary schools with a Grade R class; Early Childhood Development (ECD) centres will be excluded due to the fact that it would be difficult to track their learners to Grade 1.
- c. Have more than 20 learners and less than 120 learners, to ensure that class size is not a confounding factor for learner performance.
- d. Have not previously been exposed to Stellar or any other Wordworks program, to reduce systematic differences between selected schools based on prior exposure to treatment.

Grade R learners will be eligible for inclusion in the study under the following conditions:

- a. Are no older than 6 years of age, to ensure that age is not a confounding factor for learner performance.
- b. Have not repeated Grade R, to ensure that both age and previous experience/exposure to Grade R are not confounding factors for learner performance.
- c. Do not have special needs (for example, are not hearing or sight impaired) due to the nature of the test design, which relies on visual and verbal cues.
- d. At the time of selection, the learner is not known to be moving to a new school (according to the Grade R teacher), or to be repeating Grade R, the following year. This is important to minimize the effect of attrition.

⁶² Actuarial Society of South Africa, 2008 demographic model

Teachers will be eligible for inclusion in the study under the following conditions:

- a. Are not Lead Teachers that have been trained in the Stellar Emergent Literacy Approach in 2015, to ensure that prior exposure to treatment and special status (Lead Teachers may potentially be more experienced, better trained and/or more motivated teachers) are not confounding factors for learner performance.
- b. Have not moved from another school where they previously received the Stellar intervention, to ensure that prior exposure to treatment is not a confounding factor for learner performance.

Sampling

The schools for this evaluation have been randomly selected using systematic sampling. The units of analyses are (1) Grade R learners and (2) Grade R teachers. The data used for the sampling frame was the 2013 Masterlist for Quarter 4, as this is the latest Masterlist data available, which includes SNAP survey data on Grade R learner enrolment. Using the 2014 Annual National Assessment (ANA) results, a combined dataset was constructed in order to have a measure of performance in the sampling frame.

Grade R learner enrolment for 2014 was used to determine whether a school has Grade R classes, and the sampling frame was limited to only include schools with more than 20 learners and less than 120 Grade R learners. Furthermore, the sampling frame excluded the Metro South district and the Metro Central district because they have been receiving the STELLAR intervention previously. The West Coast district was also excluded from the sampling frame due to the restriction placed on the driving time for fieldworkers.⁶³ The sample was stratified based on Quintile, and therefore all schools without a Quintile classification in the data had to be excluded (only two schools). Subsequently, we were left with 446 schools from the remaining districts⁶⁴ from which we could draw our sample.

A random sample of 10 schools per quintile was selected for assessment, with an additional 2 schools drawn as potential substitute school⁶⁵. Since the schools will act as their own control in 2015 (when Grade R's are not exposed to treatment) and will be compared to their own performance in 2016 (when Grade R's have been exposed to treatment), it was not necessary to sample 100 schools as originally proposed. Rather, the same 50 schools would be visited twice, in 2015 and 2016, to compare the performance of the learners and teachers prior to, and after, treatment. This would give us a total sample of 900 learners (50 schools, 9 learners per school, measured in consecutive years). The rationale behind selecting this approach is that it reduces inter-school variability and therefore improves the quality of the counterfactual.

The sample size for the treatment and comparison schools was selected based on power calculations done in Stata using the `clustersampsi` command, and using the Optimal Design Software. The selected sample size was based on an initial assumption about the amount of learning that would likely take place over various points of data collection (based on international and local literature).

⁶³ The Eden and Karoo district was included as one of the fieldworkers live in Oudsthoorn.

⁶⁴ Metro North, Metro East, Cape Winelands, Overberg, Eden and Central Karoo,

⁶⁵ These schools will be assessed if any problems arise at the initially sampled schools.

Within each school, students will be randomly selected using systematic sampling and will be tested by the fieldworkers individually. To minimize the disruption to the school day, the number of children tested will be restricted to only 9 students. The fieldworkers will start at the 3rd name on the class list and then select every (N/9)th student. In cases where a class list is not available, the fieldworker will utilize an alternate random sampling approach. All Grade R teachers in the treatment and comparison schools will be observed/assessed.

Learner Assessment Instrument

Learner scores will be evaluated using a **Grade R and Grade 1 Learner Assessment Tool** (LAT), which is a composite, built-up test that incorporates content from Grade R through to Grade 2 to allow for comparison over time, and demonstrable improvement across the Grades. The tool is aligned with the Curriculum Assessment Policy Statement (CAPS) and follow CAPS recommendations regarding assessment in the Grade R context.

The parameters for the LAT include the following:

1. One-on-one assessment.
2. Time limit of 20 minutes per learner assessed.
3. The instrument to be available in English, Afrikaans and isiXhosa.
4. Grade R and Grade 1 learners will be assessed on the same instrument so that improvement could be measured from Grade to Grade.
5. Each learner will be assessed three times: mid-Grade R, end Grade R and end Grade 1
6. The study aims at measuring Home language learning at the end of Grade 1, so criteria have had to be drawn from CAPS.
7. Some learners could be expected to be ahead of others at the end of Grade 1, so provision has had to be made to measure extension work as well (i.e. including some Grade 2 items).

Steps in developing the instrument include the following:

1. CAPS Home language criteria for Grades R, 1 and 2 were analysed and tabulated by school term.
2. Duplicate criteria were eliminated so that learning of a particular outcome is shown only when it first appears. For example, if “Knows the whole alphabet” occurs for Grade 1 terms 3 and 4, it is included only for Grade 1 Term 3.
3. Criteria were arranged to be continuous so that increased learning in the same area can be tracked through the Grades. For example, knowing 10 single letters in term 1, expanded to 21 letters in term 2, all 26 letters in term 3, and two-letter combinations in term 4.
4. Criteria with an expected increase in learning across time were extracted to be included in the instrument. This meant that, if a criterion occurred only in one term of one Grade, it would not be assessed.
5. Items were constructed so that answers could be anchored to as many of the following categories as possible, using a 6-point scale: Not done (0), Partial Grade R level (1), End of Grade R level (2), Partial Grade 1 level (3), End of Grade 1 level (4), Grade 2 level (5).
6. Anchoring was described for each item, For example, for the item “Write your name” the response would be rated as 0 if not done, 1 if the writing didn’t look like a name, 2 if it looked like a name, 3 if it started with a Capital letter etc. Where learning in a particular area only started in Grade 1, the anchoring only started at score 3.
7. Provision was made in the anchoring for “prompted” answers, where the tester has to further explain a question before a learner can answer it.

8. Stellar big books modules 7 and 8 were used as reading material for the first test (mid-Grade R). These modules will only be given to teachers in the last school term and are thus unfamiliar to learners.
9. After compiling and internally testing the English instrument, translations into Afrikaans and isiXhosa were carried out.
10. Two additional versions of the test were produced for end of Grade R and end of Grade 1. This involved changing the examples in the phonics sections and using extra reading material supplied by Wordworks that had not been included in the eight Stellar modules.

The final instrument is described as follows:

1. The instrument measures all four CAPS (a) Home Language skills: Listening and speaking (12 items), (b) Phonics (12 items), (c) Reading and viewing (6 items) and (d) Handwriting and writing (6 items)
2. Each CAPS skill was measured with 6 items drawn from the CAPS criteria: 3 at the Grade R level, 2 at the Grade 1 level and 1 at the Grade 2 level. This doubles up where there are 12 items.
3. Demographic information per teacher, classroom, and school are collected before the test begins.

The LAT is administered in two parts. In Part 1, the group of 9 randomly selected Grade R learners work in a group, listen to a story, and carry out writing tasks. In Part 2, each learner is tested separately, one-on-one, and the responses to each item are recorded according to the anchoring described above.

Data Handling and Collection

Khulisa will use a team of Grade R/ early grade assessors/ retired teachers to capture the data for this evaluation. These fieldworkers will use a tried and tested, open source, information and communication technology tool - the Open Data Kit (ODK) - in combination with smart phones to collect information directly from the field, where applicable. The ODK system allows the evaluation team to build forms for collection of qualitative or quantitative information, collect the data, send it to a cloud-based server when the device is in range of wireless signal, and aggregate, and extract the data from the server. The system improves data entry, data management, and data quality.

The fieldworkers will be trained for three days by Khulisa staff on the use of the ODK system and each of the evaluation tools. Since the ODK system has a GPS tracking system, Khulisa will be able to track where and when data is collected. The evaluation team will also create a WhatsApp group to allow fieldworkers to interact with the evaluation team and ask questions during field work. The ODK system will send the data directly to a private cloud-based server, hosted at Khulisa. This will ensure confidentiality during transmission, use, and storage of data. At the end of the study, Khulisa will provide USAID with data, technical materials, and other relevant materials produced in the execution of this evaluation. Once USAID has received the data from Khulisa, and has signed off on it, the data will be purged from the server and deleted.

Grade R learners who are selected for this evaluation will be tracked using a unique identification number. Khulisa will generate a unique number for each learner who is evaluated in Grade R, and use this number as a unique identifier for the specific learner throughout the evaluation. Each teacher will also be given a unique identifier which will be linked to the learners in his/ her class. This will enable the evaluation team to link specific teacher characteristics to changes in learner

outcomes. As the evaluation is being conducted in primary schools with Grade R and Grade 1 classes, it is expected that many of these learners will continue on to Grade 1 in the same school. However, it is possible that attrition could occur due to migration of learners to other areas or other schools. Khulisa will collect contact details for all Grade R learners to see to what extent it will be possible to locate the learners who have migrated to other schools.

ANNEX 3: DATA COLLECTION TOOLS

Note: Only the English tools are included in this Annex for reference purposes. Please see the Development Experience Clearinghouse at <https://dec.usaid.gov/dec/home/Default.aspx> for the full suite of Afrikaans and isiXhosa tools.

LEARNER ASSESSMENT: WAVE 1

START TIME: _____ DATE: _____ FIELDWORKER: _____

DISTRICT: _____ SCHOOL: _____

LEARNER NAME: _____

CEMIS NUMBER: _____

PART A – IN THE GROUP

Exercise 1: Shared reading of a Big Book

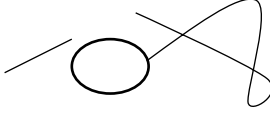
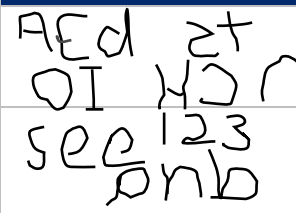
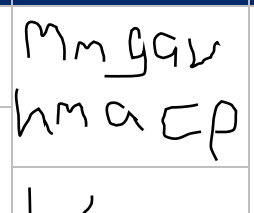
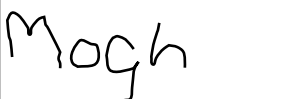
I am going to read you a story. This is a story about “Bushbuck’s visitor”.

Read WW Big Book 7 “Bushbuck’s visitor” to the group. Show them the pictures and point to the words as you read. You may dramatize the story with different voices and point to relevant pictures. DO NOT explain or expand the story, or repeat any part of the story or use words from another language.

Exercise 2: Writing. *Hand out the activity sheet. Make sure everyone has the blank side of the paper in front of them. It is important that learners know WHAT to do, so instructions can be repeated. Wait for everyone to finish an instruction before giving the next instruction.*

No	Assessing	Question to be asked	0	1	2	3	4	5
A1	Writes own name. W, R	Write your name at the top of the paper.	Does not look like a name or not done.	A name. No capital letter. Not all letters present.	Capital letter at start of name. All letters present.	Name and Surname.		
A2.	Emergent writing (write about a picture). W, R	Draw a picture of the part of the story you liked best. <i>(Allow about 4 minutes)</i> Now write something about the story under the line. <i>(Allow about 4 minutes)</i>	Rate writing as below. <i>Drawing will be rated later.</i>					
A3	Copy a sentence. W, 1	Turn the page over. There is a sentence that says, "The Frog Hops Out". Copy this on the line.	Not done or unable.	Scribbles.	Does not look like a copy.	Looks like a copy.	Copy, with correct spaces.	Exact copy with punctuation.
A4	Write a sentence correctly. W, 1	On the next line, write "The cat saw a butterfly".	Not done or unable.	Scribbles.	Not a sentence.	Correct but incomplete. 4 words.	Complete. 5 words correct spaces.	Correct with correct punctuation.
A5	Copy a pattern. W, R	Now copy the drawing in the block on the left into the block on the right.	Blank or wrong figures.	Approximate triangle and circle or one correct.	Closed correct figures within square.	Closed figures in the right places.	Closed figures touch correctly.	Exact copy.

Rating examples for questions A2 to A4

0	1	2	3	4	5
					Mom gave him a cup.
					
Blank or picture only.	Random scribbles or scribble writing.	Symbols representing letters, random letters, or letter strings. No beginning sounds.	Groups of letters and/or numbers, more than one beginning sound.	(Early) inventive spelling.	Proper writing.

PART B - ONE ON ONE

No	Assessing	Question to be asked	0	1	2	3	4	5
1	Knows name. L, R	What is your name?	No answer.	After asked twice. Gives name.	Unprompted. Gives name.	Gives name and Surname.		
2	Knows friends' names. L, 1	What are your friends' names?	No answer.			After asked twice. Gives one name.	Unprompted. Gives one name.	Unprompted. Gives more than one name.
3	Respond to closed question. L, R	Your teacher is _____. Is that right?	Not done or Yes/No.	After asked twice. Yes/No.	Unprompted Yes/No.	Unprompted Yes/No plus extra.		
4	Complex question and answer. L, R	What do you do at school?	Not done or unable.	Prompted or one word answer.	Unprompted. Full sentence.	Unprompted Long explanation.		
5	Respond to open question. L, 1	How do you get to school? (Prompt: By taxi?)	Not done or Yes/No.		Prompted or one word answer.	Unprompted. Full sentence.	Unprompted Extra.	
6	Suggests topics. L, 2	Is there anything else you would like to tell me about your family? (Unprompted only)	Not done or unable.					Unprompted. Tells family details.
7	Remembers sequence of story. L, 2	Show learner the <u>cover</u> of the book "Bushbuck's visitor" Can you remember what happened in this story?	Not done or unable.					Tells 3 or more parts of story in correct sequence.
8	Handles book correctly. R, R	(Hand learner the book "Bushbuck's visitor" upside down and back to front. Observe how the learner handles the book).	Needs help to position book.	Knows some things, but some are wrong.	Unprompted Turns book and opens correctly.			
9	Looks to book for answers. L, R	Show me whom Bushbuck called.	Not done or unable.	Prompted. Shows lion or elephant.	Unprompted. Shows lion or elephant.	Unprompted. Shows both Lion and elephant.		
10	Points out objects in a picture. L, R	Show me where the door is.	Not done or unable.	Prompted.	Unprompted. Shows one door.	Unprompted. Shows more than one.		

No	Assessing	Question to be asked	0	1	2	3	4	5
11	Discusses characters. L, R	What did the frog say when he was in the house?	Not done or unable.			<i>Prompted.</i> One of: Do not open the door or I will eat you.	<i>Unprompted</i> One of: Do not open the door or I will eat you.	<i>Unprompted</i> Do not open the door or I will eat you.
12	Matches caption to picture. L, 1	Show me the picture where the frog came out of the house.	Not done or unable.			<i>Prompted</i> Shows correct page.	<i>Unprompted</i> Shows correct page.	<i>Unprompted</i> Explains picture.
13	Different letter sound. P, R	Say the letters with me; c, c, g, c. <i>(Repeat two times in total)</i>	Not done or unable.	Says all "c".	Says "c" and "g".	<i>Unprompted.</i> Repeats whole letter string.		
14	Sounds make up words. P, R	What word does this make? c – a – t <i>(Show combining letters with hands. Allow time to answer and then give answer "cat" as an example).</i> What word does this make? d – o – g	Not done or unable.	<i>Prompted</i> "dog".	<i>Unprompted.</i> "dog".	<i>Unprompted.</i> Both correct: "cat, dog".		
15	Substitute vowels in words. P, 1	Say the word "hat". Now say it with "o" instead of "a". <i>(Allow time to answer and then give answer "hot" as an example).</i> Say the word "pen". Now say it with "a" instead of "e".	Not done or unable.	<i>Prompted</i> "Pan".	<i>Unprompted</i> "Pan"	<i>Unprompted.</i> Both correct: "hot, pan".		
16	Builds words using sounds learnt. P, 1	Tell me some more words with "at" at the end. <i>(Prompt: We had "Cat" and "Hat". Give me more words with "at".)</i>	Not done or unable.			<i>Prompted or</i> only one.	<i>Unprompted.</i> More than one e.g. "sat, bat".	<i>Unprompted.</i> Volunteers many words.
17	Difference between starting initials of two words. P, 1	Ffff. "Fire" starts with Fff. What other word starts with Fff? <i>(Prompt: Fan, fall start with Fff. Give me another word that starts with Fff.)</i>	Not done or Fire, fan and fall repeated.	<i>After prompt.</i> A new word.	<i>Unprompted.</i> One new word.	<i>Unprompted.</i> More than one new word.		
18	Odd word in sequence. P, R	Say these words with me: Rose, Cap, Round. <i>(Emphasise R).</i> Which word does not start the same as the others? <i>(Allow time to</i>	Not done or unable.	<i>After example and further prompt.</i> "Desk".	<i>Unprompted.</i> "Desk".	<i>Unprompted.</i> Both correct: "Cap, Desk".		

No	Assessing	Question to be asked	0	1	2	3	4	5
		answer and then give answer “cap” as an example.) Test, Desk, Tank. (Emphasise T). Which word does not start the same as the others?						
19	Sentences made up of words. P, R	Say this sentence with me. “She likes the cat.” (Show to clap on every word.) Now clap on every word in this sentence. “He ran to the shop.”	Not done or unable.	After example and further prompt. 5 claps.	Unprompted. 5 claps.	Unprompted. 5 claps for both.		
20	Builds up words with blends. P, 2	Say this with me: Gr – a – nd. What word does that make? (Allow time to answer, then give answer “grand” as an example.) Say this with me: Bl – i – nk”. What word does that make?	Not done or unable.			After example and further prompt. “Blink”.	Unprompted “Blink”.	Unprompted. Both correct “Grand, Blink”.
21	Recognises some consonants. P, R	(On worksheet 1) Show me “m” for “mother”. Show me “d” for “duck”. (Point to “s”). What is this?	Not done or unable.	Prompted or one correct.	Unprompted. All correct “m, d, s”.			
22	Points to letter starting a word. P, R	(On worksheet 1) Say this word with me: “Bird”. Show me what this word starts with. (Allow time to answer and then show answer “b” as an example.) Say this name with me: “Peter”. Show me what this word starts with.	Not done or unable.	Prompted or one correct.	Unprompted. All correct “b, p”.			
23	Recognises 2 consonants at start and end of word. P, 1	(On worksheet 2) Say this word with me: “Shut”. Show me what this word starts with. Say this word with me: “Wish”. Show me what this word ends with.	Not done or unable.			Prompted or one correct.	Unprompted. Two correct “sh, sh”.	Unprompted. Points out they are the same.
24	Recognises double vowels and 3	(On worksheet 2)	Not done or unable.			Prompted or one correct.	Prompted or two correct.	Unprompted. All correct

No	Assessing	Question to be asked	0	1	2	3	4	5
	consonant strings at word start and end. P, 2	Show me “oo” as in “moon” Show me “ing” as in “running” (Point to “spr”). What is this?						‘oo, ing, spr’.
25	Recognises sight words. R, 1	(On worksheet 2 – last line) Can you read these words?	Not done or unable.			Reads one or two of the words.	Reads three or four slowly.	Reads four quickly “The, is, and of”.

Show learner his/her own writing on worksheet 1. Hand over the page

No	Assessing	Question to be asked	0	1	2	3	4	5
26	Draws picture to show understanding. L, 1	What is this picture? (Related to story)	Not done.	Prompted or unrelated picture.	Unprompted Related picture.	Extra e.g. 2 pictures.		
27	Reads own writing. W, R	Please read me what you wrote. (From story under picture drawn, or last sentence before copying exercise.)	Not done.	Prompted “Reads”.	Unprompted “Reads”.	Unprompted “Reads” correct story or sentence.		

Show learner both sides of the Story worksheet, then the picture side.

No	Assessing	Question to be asked	0	1	2	3	4	5
28	Predicts storyline from cover. R, 1	(Show cover of the book) What is this story called?	Not done or unable.			Prompted “Cat and Dog”.	Unprompted “Cat and dog”.	Unprompted Reads “The cat and dog”.
29	“Reading” voice. R, R	(Turn over.) Please read me the story.	Not done or unable.	Tells story in normal voice	Unprompted. Tells Unrelated story in reading voice	Some actual words		
30	Points to words while reading. R, R	(From previous. Ask during exercise). Show me where you are reading.	Not done or unable. Points to picture.	Prompted. Points to words.	Unprompted. Points to words.	Unprompted. Points to exact word.		
31	Reads actual words. R, 2	(From whole of previous)	Not done or unable.					All correct

LEARNER ASSESSMENT: WAVE 2

LEARNER NAME: _____

EDUCATION ID: _____ GRADE: _____

GENDER: _____ DATE OF BIRTH: _____ HOME LANGUAGE: _____

TEACHER: _____

FIELDWORKER: _____ DATE: _____

PART A – IN THE GROUP

Exercise 1: Shared reading of a Big Book

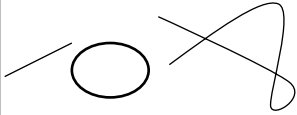
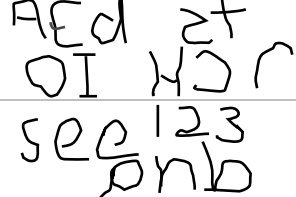
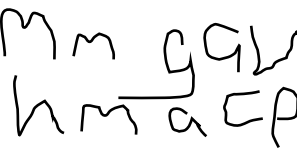
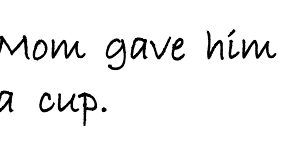
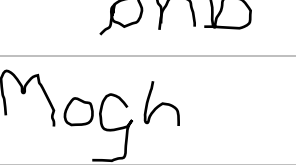
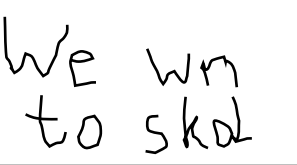
I am going to read you a story. This is a story about “The baby birds”.

Read “The baby birds” to the group. Show them the pictures and point to the words as you read. You may dramatize the story with different voices and point to relevant pictures. DO NOT explain or expand the story, or repeat any part of the story or use words from another language.

Exercise 2: Writing. *Hand out the activity sheet. Make sure everyone has the blank side of the paper in front of them. It is important that learners know WHAT to do, so instructions can be repeated. Wait for everyone to finish an instruction before giving the next instruction.*

No	Assessing	Question to be asked	0	1	2	3	4	5
A1	Writes own name. W, R	Write your name at the top of the paper.	Does not look like a name or not done.	A name. No capital letter. Not all letters present.	Capital letter at start of name. All letters present.	Name and Surname.		
A2.	Emergent writing (write about a picture). W, R	Draw a picture of the part of the story you liked best. <i>(Allow about 4 minutes)</i> Now write something about the story under the line. <i>(Allow about 4 minutes)</i>	Rate writing as below. <i>Drawing will be rated later.</i>					
A3	Copy a sentence. W, 1	Turn the page over. There is a sentence that says, "The frog jumps out". Copy this on the line.	Not done or unable.	Scribbles.	Does not look like a copy.	Looks like a copy.	Copy, with correct spaces.	Exact copy with punctuation.
A4	Write a sentence correctly. W, 2	On the next line, write, "The cat saw a butterfly".	Not done or unable.	Scribbles.	Not a sentence.	Correct but incomplete. 4 words.	Complete. 5 words correct spaces.	Correct with correct punctuation.
A5	Copy a pattern. W, R	Now copy the drawing in the block on the left into the block on the right.	Blank or wrong figures.	Approximate triangle and circle or one correct.	Closed correct figures within square.	Closed figures in the right places.	Closed figures touch correctly.	Exact copy.

Rating examples for questions A2 to A4

0	1	2	3	4	5
					
					
Blank or picture only	Random scribbles or scribble writing	Symbols representing letters, random letters or letter strings, No beginning sounds.	Groups of letters and/or numbers, more than one beginning sound	(Early) inventive spelling	Proper writing

PART B - ONE ON ONE

No	Assessing	Question to be asked	0	1	2	3	4	5
17	Difference between starting initials of two words. P, 1	Ssss. “Snake” starts with Sss. What other word starts with Sss? (<i>Sit, sand, start with Sss Give me another word that starts with Sss</i>)	<i>Not done or Sit, stand, repeated.</i>			<i>After prompt. A new word.</i>	<i>Unprompted. One new word.</i>	<i>Unprompted. More than one new word.</i>
18	Odd word in sequence. P, R	Say these words with me: Wait, Hat, Wind. (<i>Emphasise W</i>). Which word does not start the same as the others? (<i>Allow time to answer and then give answer “hat” as an example.</i>) Cap, Gum, Cost. (<i>Emphasise C</i>). Which word does not start the same as the others?	<i>Not done or unable.</i>	<i>After example and further prompt. “Gum”.</i>	<i>Unprompted. “Gum”.</i>	<i>Unprompted. Both correct: “Hat, Gum”.</i>		
19	Sentences made up of words. P, R	Say this sentence with me. “He wants the red sweet.” (<i>Show to clap on every word.</i>) Now clap on every word in this sentence. “She climbed up the tree.”	<i>Not done or unable.</i>	<i>After example and further prompt. 5 claps.</i>	<i>Unprompted. 5 claps.</i>	<i>Unprompted. 5 claps for both.</i>		
20	Builds up words with blends. P, 1	Say this with me: Br – a – nd. What word does that make? (<i>Allow time to answer and then give answer “brand” as an example.</i>) Say this with me: Tr – u – nk”. What word does that make?	<i>Not done or unable.</i>			<i>After example and further prompt. “Trunk”.</i>	<i>Unprompted “Trunk”</i>	<i>Unprompted. Both correct “Brand, Trunk”</i>

No	Assessing	Question to be asked	0	1	2	3	4	5
21	Recognises some consonants. P, R	(On worksheet 1) Show me “t” for “table”. Show me “r” for “rose”. (Point to “m”). What is this?	<i>Not done or unable.</i>	<i>Prompted or one correct.</i>	<i>Unprompted.</i> All correct “t, r, m”.			
22	Points to letter starting a word. P, R	(On worksheet 1) Say this word with me: “Dance”. Show me what this word starts with. <i>(Allow time to answer and then show answer “d” as an example.)</i> Say this name with me: “Nadia”. Show me what this word starts with.	<i>Not done or unable.</i>	<i>Prompted or one correct.</i>	<i>Unprompted.</i> All correct “d, n”.			
23	Recognises 2 consonants at start and end of word. P, 1	(On worksheet 2) Say this word with me: “This”. Show me what this word starts with. Say this word with me: “Cloth”. Show me what this word ends with.	<i>Not done or unable.</i>			<i>Prompted or one correct.</i>	<i>Unprompted.</i> Two correct “th, th”.	<i>Unprompted.</i> Points out they are the same.
24	Recognises double vowels and 3 consonant strings at word start and end. P, 2	(On worksheet 2) Show me “ee” as in “bee” Show me “ier” as in “funnier” <i>(Now, point to “str”).</i> What is this?	<i>Not done or unable.</i>			<i>Prompted or one correct.</i>	<i>Prompted or two correct.</i>	<i>Unprompted.</i> All correct ‘ ee, ier, str”.
25	Recognises sight words. R, 1	(On worksheet 2 – last line) Can you read these words?	<i>Not done or unable</i>			Reads one or two of the words.	Reads three or four slowly.	Reads four quickly “And, is, the of”.

Show learner his/her own writing on worksheet 1. Hand over the page

No	Assessing	Question to be asked	0	1	2	3	4	5
26	Draws picture to show understanding. L, 1	What is this picture? (Related to story)	Not done.			Prompted or unrelated picture.	Unprompted Related picture.	Extra e.g. 2 pictures.
27	Reads own writing. W, R	Please read me what you wrote. (From story under picture drawn, or last sentence before copying exercise.)	Not done.	Prompted "Reads".	Unprompted "Reads".	Unprompted "Reads" correct story or sentence.		

Show learner both sides of the Story worksheet, then the picture side.

No	Assessing	Question to be asked	0	1	2	3	4	5
28	Predicts storyline from cover. R, 1	(Show cover of the book) What is this story called?	Not done or unable.			Prompted "Cat and Dog".	Unprompted "Cat and dog".	Unprompted Reads "The cat and dog."
29	"Reading" voice. R, R	(Turn the page over.) Please read me the story.	Not done or unable.	Tells story in normal voice.	Unprompted. Tells Unrelated story in Reading voice.	Some actual words.		
30	Points to words while reading. R, R	(From previous. Ask during exercise) Show me where you are reading.	Not done or unable. Points to picture.	Prompted. Points to words.	Unprompted. Points to words.	Unprompted. Points to exact word.		
31	Reads actual words. R, 2	(From whole of previous)	Not done or unable.			Prompted Reads some words correctly.	Unprompted Reads some words correctly.	All correct.

LEARNER ASSESSMENT: WAVE 3

LEARNER NAME: _____

EDUCATION ID: _____ GRADE: _____

GENDER: _____ DATE OF BIRTH: _____ HOME LANGUAGE: _____

TEACHER: _____

FIELDWORKER: _____ DATE: _____

PART A – IN THE GROUP

Exercise 1: Shared reading of a Big Book

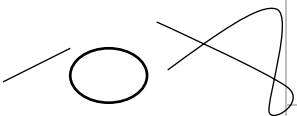
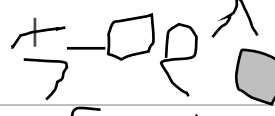
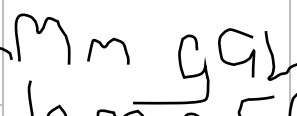
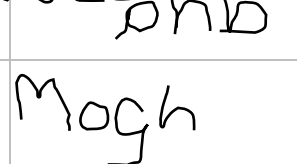
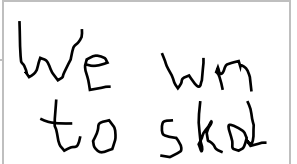
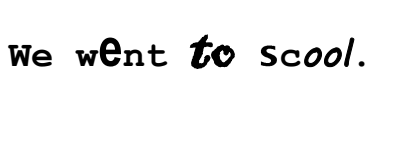
I am going to read you a story. This is a story about “Joshua’s goal.”

Read “Joshua’s goal” to the group. Show them the pictures and point to the words as you read. You may dramatise the story with different voices and point to relevant pictures. DO NOT explain or expand the story, or repeat any part of the story or use words from another language.

Exercise 2: Writing. *Hand out the activity sheet. Make sure everyone has the blank side of the paper in front of them. It is important that learners know WHAT to do, so instructions can be repeated. Wait for everyone to finish an instruction before giving the next instruction.*

No	Assessing	Question to be asked	0	1	2	3	4	5
A1	Writes own name. W, R	Write your name at the top of the paper.	Does not look like a name or not done.	A name. No capital letter. Not all letters present.	Capital letter at start of name. All letters present.	Name and Surname.		
A2.	Emergent writing (write about a picture). W, R	Draw a picture of the part of the story you liked best. (Allow about 4 minutes) Now write something about the story under the line. (Allow about 4 minutes)	Rate writing as below. Drawing will be rated later.					
A3	Copy a sentence. W, 1	Turn the page over. There is a sentence that says, "The frog jumps out". Copy this on the line.	Not done or unable.	Scribbles.	Does not look like a copy.	Looks like a copy.	Copy, with correct spaces.	Exact copy with punctuation.
A4	Write a sentence correctly. W,2	On the next line, write "The cat saw a butterfly".	Not done or unable.	Scribbles.	Not a sentence.	Correct but incomplete 4 words.	Complete. 5 words correct spaces.	Correct with correct punctuation.
A5	Copy a pattern. W, R	Now copy the drawing in the block on the left into the block on the right.	Blank or wrong figures.	Approximate triangle and circle or one correct.	Closed correct figures within square.	Closed figures in the right places.	Closed figures touch correctly.	Exact copy

Rating examples for questions A2 to A4

0	1	2	3	4	5
					
					
Blank or picture only	Random scribbles or scribble writing	Symbols representing letters, random letters or letter strings, No beginning sounds.	Groups of letters and/or numbers, more than one beginning sound	(Early) inventive spelling	Proper writing

No	Assessing	Question to be asked	0	1	2	3	4	5
1	Knows name. L, R	What is your name?	No answer.	After asked twice. Gives name.	Unprompted. Gives name.	Gives name and Surname.		
2	Knows friends' names. L, 1	What are your friends' names	No answer.			After asked twice. Gives one name.	Unprompted. Gives one name.	Unprompted. Gives more than one name.
3	Respond to closed question. L, R	Your teacher is _____. Is that right?	Not done or Yes/No.	After asked twice. Yes/No.	Unprompted Yes/No.	Unprompted Yes/No plus extra.		
4	Complex question and answer. L, R	What do you do at school?	Not done or unable.	Prompted or one word answer.	Unprompted. Full sentence.	Unprompted Long explanation.		
5	Respond to open question. L, 1	How do you get to school? <i>Prompt: By taxi?</i>	Not done or Yes/No.			Prompted or one word answer.	Unprompted. Full sentence.	Unprompted Extra.
6	Suggests topics. L, 2	Tell me about the place where you live?	Not done or unable.			Prompted. Says only one thing.	Unprompted. Says only one thing.	Unprompted. Gives details enough for a conversation.
7	Remembers sequence of story. L,2	Show learner the <u>cover</u> of the book "Joshua's goal" Can you remember what happened in this story? Please tell me.	Not done or unable.			Prompted. Less than 3 parts or not in sequence.	Unprompted. Less than 3 parts or not in sequence.	Tells 3 or more parts of story in correct sequence.
8	Handles book correctly. R, R	<i>(Hand learner the book "Joshua's goal" upside down and back to front. Observe how the learner handles the book).</i>	Needs help to position book.	Knows some things, but some are wrong.	Unprompted Turns book and opens correctly.			
9	Finds answers in a book. L, R	Show me the page where the goal was scored.	Not done or unable.	Prompted. Shows picture.	Unprompted. Shows picture.	Unprompted. Shows the goal being scored.		
10	Points out objects in a picture. L, R	In the book, show me the soccer ball.	Not done or unable.	Prompted.	Unprompted. Shows one picture.	Unprompted. Shows more than one picture.		

No	Assessing	Question to be asked	0	1	2	3	4	5
11	Discusses characters. L, R	What did Andrew say when he saw his brother's soccer was getting better?	Not done or unable.	<i>Prompted.</i> One day you will be a good soccer player or You must practise to be the best you can.	<i>Unprompted</i> One day you will be a good soccer player or You must practise to be the best you can.	<i>Unprompted</i> Reply both One day you will be a good soccer player. You must practise to be the best you can.		
12	Matches caption to picture. L, 1	Which picture matches the words, The crowd shouted, "Goal, goal"?	Not done or unable.			<i>Prompted</i> Shows correct page.	<i>Unprompted</i> Shows correct page.	<i>Unprompted</i> Explains picture.
13	Different letter sound. P, R	Say the letters with me. b, b, p, b. (Repeat two times in total)	Not done or unable.	Says all "b".	Says "b" and "p".	<i>Unprompted.</i> Repeats whole letter string.		
14	Sounds make up words. P, R	What word does this make? n – a – p (Show combining letters with hands. Allow time to answer and then give answer "nap" as an example.) What word does this make? b – u – n	Not done or unable.	<i>Prompted</i> "Bun".	<i>Unprompted</i> "Bun".	<i>Unprompted.</i> Both correct: "nap, bun".		
15	Substitute vowels in words. P, 1	Say the word "net". Now say it with "o" instead of "e". (Allow time to answer and then give answer "not" as an example.) Say the word "lap". Now say it with "i" instead of "a".	Not done or unable.			<i>Prompted</i> "not".	<i>Unprompted</i> "not".	<i>Unprompted.</i> Both correct: "not, lip".
16	Builds words using sounds learnt. P, 1	Tell me some more words with "ap" at the end. (Prompt: We had "lap" and "nap". Give me more words with "ap".)	Not done or unable.			<i>Prompted</i> or only one.	<i>Unprompted.</i> More than one e.g. "cap, tap, sap".	<i>Unprompted.</i> Volunteers many words gap, map, rap, yap, (ap).

No	Assessing	Question to be asked	0	1	2	3	4	5
17	Difference between starting initials of two words. P, 1	G for “Garden” starts with G. What other word starts with G? (<i>Get, gap, start with G Give me another word that starts with G</i>)	<i>Not done or unable.</i>			<i>After prompt.</i> A new word.	<i>Unprompted.</i> One new word.	<i>Unprompted.</i> More than one new word.
18	Odd word in sequence. P, R	Say these words with me: Pen, Bank, Past. (<i>Emphasise P</i>). Which word does not start the same as the others? (<i>Allow time to answer and then give answer “bank” as an example.</i>) Dark, Test, Doll. (<i>Emphasise D</i>). Which word does not start the same as the others?	<i>Not done or unable.</i>	<i>After example and further prompt.</i> “Test”.	<i>Unprompted.</i> “Test”.	<i>Unprompted.</i> Both correct: “Bank, Test”.		
19	Sentences made up of words. P, R	Say this sentence with me. “We all sing in class.” (<i>Show to clap on every word.</i>) Now clap on every word in this sentence. “I want to go now.”	<i>Not done or unable.</i>	<i>After example and further prompt.</i> 5 claps.	<i>Unprompted.</i> 5 claps.	<i>Unprompted.</i> 5 claps for both.		
20	Builds up words with blends. P, 1	Say this with me: Cr – a – sh. What word does that make? (<i>Allow time to answer and then give answer “crash” as an example.</i>) Say this with me: Bl – u – nt”. What word does that make?	<i>Not done or unable.</i>			<i>After example and further prompt.</i> “Blunt”.	<i>Unprompted</i> “Blunt”	<i>Unprompted.</i> Both correct “Crash, “Blunt”.
21	Recognises some consonants. P, R	(On worksheet 1) Show me “c” for “cat”. Show me “b” for “bus”. (Point to “h”). What is this?	<i>Not done or unable.</i>	<i>Prompted or one correct.</i>	<i>Unprompted.</i> All correct “c,b,h”.			
22	Points to letter starting a word. P, R	(On worksheet 1) Say this word with me: “Lap”. Show me what this word starts with. (<i>Allow time to answer and then show answer “l” as an example.</i>) Say this word with me: “Fire”. Show me what this word starts with.	<i>Not done or unable.</i>	<i>Prompted or one correct.</i>	<i>Unprompted.</i> All correct “l, f”.			

No	Assessing	Question to be asked	0	1	2	3	4	5
23	Recognises 2 consonants at start and end of word. P, 1	(On worksheet 2) Say this word with me: “Chat”. Show me what this word starts with. Say this word with me: “Lunch”. Show me what this word ends with.	Not done or unable.			Prompted or one correct.	Unprompted. Two correct “ch, ch”.	Unprompted. Points out they are the same.
24	Recognises double vowels and 3 consonant strings at word start and end. P, 2	(On worksheet 2) Show me “ay” as in “say” Show me “tch” as in “catch” (Point to “ing”). What is this?	Not done or unable.			Prompted or one correct.	Prompted or two correct.	Unprompted. All correct “ay, tch, ing”.
25	Recognises sight words. R, 1	(On worksheet 2 – last line) Can you read these words?	Not done or unable.			Reads one or two of the words.	Reads three or four slowly.	Reads four quickly “of, and, is, the”

Show learner his/her own writing on worksheet 1. Hand over the page

No	Assessing	Question to be asked	0	1	2	3	4	5
26	Draws picture to show understanding. L, 1	What is this picture? (Related to story)	Not done.			Prompted or unrelated picture.	Unprompted Related picture.	Extra e.g. 2 pictures.
27	Reads own writing. W, R	Please read me what you wrote. (From story under picture drawn, or last sentence before copying exercise.)	Not done.	Prompted “Reads”.	Unprompted “Reads”.	Unprompted “Reads” correct story or sentence.		

Show learner both sides of the Story worksheet, then the picture side.

No	Assessing	Question to be asked	0	1	2	3	4	5
28	Predicts storyline from cover. R, 1	(Show cover of the book) What is this story called?	Not done or unable.			Prompted “Cat and Dog”.	Unprompted “Cat and dog”.	Unprompted Reads “The cat and dog.”
29	“Reading” voice. R, R	(Turn the page over.) Please read me the story.	Not done or unable.	Tells story in normal voice.	Unprompted. Tells Unrelated story in Reading voice.	Some actual words.		

No	Assessing	Question to be asked	0	1	2	3	4	5
30	Points to words while reading. R, R	(From previous. Ask during exercise) Show me where you are reading.	Not done or unable. Points to picture.	Prompted. Points to words.	Unprompted. Points to words.	Unprompted. Points to exact word.		
31	Reads actual words. R, 2	(From whole of previous)	Not done or unable.			Prompted Reads some words correctly.	Unprompted Reads some words correctly.	All correct.

TEACHER INTERVIEW TOOL: WAVE 1

Record your current location (Not mandatory)

Date

Choose Fieldworker (Select only one)

- ☐ Fieldworker 1
- ☐ Fieldworker 2
- ☐ Fieldworker 3
- ☐ Fieldworker 4
- ☐ Fieldworker 5
- ☐ Fieldworker 6
- ☐ Fieldworker 7
- ☐ Fieldworker 8
- ☐ Fieldworker 9
- ☐ Fieldworker 10

Please select your District

Please type in the name of the District (Free text)

Please select the school

Please type in the name of the school

LEARNER INFORMATION (HEADING)

Learner 1 name (Free text)

Learner 1 CEMIS number (Numerical free text)

Learner 1 Grade (Select only one)

- ☐ *Grade R*
- ☐ *Grade 1*

Learner 1 Gender (Select only one)

- ☐ *Male*
- ☐ *Female*

Learner 1 Date of Birth (formatted)

Learner 1 Main language used at home (Select only one)

Please specify (Free text)

Repeat the above questions for 10 learners

TEACHER INFORMATION (HEADING)

Teacher name (Free text)

Teacher PERSAL number (Free numerical text)

Teacher date of birth (format)

Teacher Gender (Select only one)

☐ *Male*

☐ *Female*

Language of Learning and Teaching (Select only one)

☐ *English*

☐ *Afrikaans*

☐ *Xhosa*

Other, please specify (Free text)

Which language/s does the teacher use in the classroom? (Select all that apply)

☐ *English*

☐ *Afrikaans*

☐ *Xhosa*

Other, please specify (Free text)

How many years of experience does the teacher have teaching Grade R?

____ Years

____ Months

Have you heard about the STELLAR program? (Select only one)

- ☐ Yes
- ☐ No

If yes, have you used and/or are you using it in your classroom? (Select only one)

- ☐ Yes
- ☐ No

Have you heard about the Grade R Emergent Literacy Intervention (ELIT)? (Select only one)

- ☐ Yes
- ☐ No

If yes, have you used and/or are you using it in your classroom? (Select only one)

- ☐ Yes
- ☐ No

Which of the STELLAR/ELIT stories have you read to your class? (Select all that apply)

- ☐ The green dress
- ☐ Ali and the paint
- ☐ Spot goes to the shop
- ☐ Baby bird finds his mother
- ☐ Goldilocks and the three bears
- ☐ Tortoise and his shell
- ☐ Bushbuck's visitor

When did you read the first one?

DATE – MONTH & YEAR (format) _____

When did you last read one?

DATE DROPDOWN – MONTH & YEAR (format) _____

Can you show me examples of activities that you have done with your learners, related to these stories? For example, little books, something produced by a child, puppets (select all that apply)

- ☐ Little books
- ☐ Something produced by a child (a relevant drawing, writing, craftwork)
- ☐ Puppets punched out of the templates.

If other, please specify

Which letter-sound game did your learners like best? (Select only one)

- ☐ Post Box game
- ☐ Letter-sound bingo
- ☐ Letter Snake game
- ☐ Racing car game

☐ Letter pirate game.

Which letter-sound game did your learners find most useful or taught the children the most? (Select only one)

- ☐ Post Box game
- ☐ Letter-sound bingo
- ☐ Letter Snake game
- ☐ Racing car game
- ☐ Letter pirate game.

Can you show me examples of activities that you have done with your learners, related to the letter sound games? For example, letter templates, picture cards. (Select all that apply)

- ☐ Letter formation templates copied and cut out
- ☐ Picture cards

If other, please specify

How many STELLAR cluster meetings (full sessions) has the teacher attended in the last 12 months?

_____ Cluster meetings

Did you attend any cluster sessions after the block training in July?

- ☐ Yes
- ☐ No

If yes, how many? _____

Which of these resources does the teacher use? (Select all that apply)

- ☐ STELLAR materials
- ☐ CAPS materials
- ☐ Grade R resource kit
- ☐ Self-sourced materials
- ☐ Other

If other, please specify

Highest teacher qualification (Select the highest)

- ☐ Less than Matric
- ☐ Matric
- ☐ Certificate / Diploma
- ☐ Degree or higher

Where did the teacher obtain this qualification? (Select the highest one)

- ☐ No qualification. Experience only
- ☐ School

- ☐ *FET College*
- ☐ *Teacher training college*
- ☐ *University*
- ☐ *Other*

If other, please specify

Did the teacher obtain this qualification in the last year? (Select only one)

- ☐ *Yes*
- ☐ *No*

If yes to the above, please specify (Free text)

Does the teacher read for enjoyment at home? (Any print materials)

- ☐ *Yes*
- ☐ *No*

On average, how many hours does the teacher read for enjoyment per week?

Hours

Has the teacher benefitted from any outside intervention, resources or support in the last 12 months, other than the STELLAR/Emergent Literacy intervention?

- ☐ *Yes*
- ☐ *No*

If yes to the above, specify (Free text)

PICTURES (HEADING)

ODK prompts four pictures.

Instructions:

- Please use your phone to take four pictures of the Grade R classroom;
- Try to stand in the centre of the class and get a picture of each of the four walls;
- Ensure that your picture captures the entire surface of the wall to ensure that every poster or picture on the wall can be seen in the photo; and
- Please ensure that the picture quality of each photo is **VERY GOOD** before clicking save.

NOTE: High quality pictures will enable the team to do an accurate and effective analysis of the classroom environment.

TEACHER INTERVIEW TOOL: WAVE 2

Record your current location (Not mandatory)

Date

Choose Fieldworker (Select only one)

- ☐ Fieldworker 1
- ☐ Fieldworker 2
- ☐ Fieldworker 3
- ☐ Fieldworker 4
- ☐ Fieldworker 5
- ☐ Fieldworker 6
- ☐ Fieldworker 7
- ☐ Fieldworker 8
- ☐ Fieldworker 9
- ☐ Fieldworker 10

Please select your District

Please type in the name of the District (Free text)

Please select the school

Please type in the name of the school

PICTURES (HEADING)

ODK prompts four pictures.

Instructions:

- Please use your phone to take four pictures of the Grade R classroom;
- Try to stand in the centre of the class and get a picture of each of the four walls;
- Ensure that your picture captures the entire surface of the wall to ensure that every poster or picture on the wall can be seen in the photo;
- Please ensure that the picture quality of each photo is **VERY GOOD** before clicking save.

NOTE: High quality pictures will enable the team to do an accurate and effective analysis of the classroom environment.

LEARNER INFORMATION (HEADING)

Learner 1 name (Free text)

Learner 1 CEMIS number (Numerical free text)

Learner 1 Grade (Select only one)

☐ *Grade R*

☐ *Grade 1*

Learner 1 Gender (Select only one)

☐ *Male*

☐ *Female*

Learner 1 Date of Birth (formatted)

Learner 1 Main language used at home (Select only one)

Please specify (Free text)

Repeat the above questions for 10 learners

TEACHER INFORMATION (HEADING)

Teacher name (Free text)

Teacher PERSAL number (Free numerical text)

Teacher date of birth (format)

Teacher Gender (Select only one)

☐ *Male*

☐ *Female*

Language of Learning and Teaching (Select only one)

☐ *English*

☐ *Afrikaans*

☐ *Xhosa*

Other, please specify (Free text)

Which language/s does the teacher use in the classroom? (Select all that apply)

- ☐ *English*
☐ *Afrikaans*
☐ *Xhosa*

Other, please specify (Free text)

How many years of experience does the teacher have teaching Grade R?

____ Years
____ Months

Has the teacher heard about the Grade R Emergent Literacy Intervention (ELIT)? (Select only one)

- ☐ *Yes*
☐ *No*

If yes, has the teacher used and/or are using the ELIT approach in their classroom? (Select only one)

- ☐ *Yes*
☐ *No*

How many ELIT cluster meetings has the teacher attended in the last 12 months? (Full sessions)

____ Cluster meetings

On a scale of 1 to 7, with 1 being to a little extent and 7 being to a great extent, did the teacher feel the ELIT cluster training was worth their time?

Rating: _____

On a scale of 1 to 7, with 1 being to a little extent and 7 being to a very great extent, did the teacher put any learning from the ELIT cluster meetings to use in their classroom?

Rating: _____

On a scale of 1 to 7, with 1 being very low and 7 being very high, how does the teacher rate the quality of the ELIT cluster training?

Rating: _____

What has been the teacher's major learning, discovery or insight as a result of the cluster training? (Free text)

If there were anything that the teacher could have changed about the cluster training, what would that be? (Free text)

Which of these resources does the teacher use? (Select all that apply)

- ☐ *STELLAR materials*
- ☐ *CAPS materials*
- ☐ *Grade R resource kit*
- ☐ *Self-sourced materials*
- ☐ *Other*

If other, please specify

Highest teacher qualification. (Select the highest)

- ☐ *Less than Matric*
- ☐ *Matric*
- ☐ *Certificate / Diploma*
- ☐ *Degree or higher*

Where did the teacher obtain this qualification? (Select the highest one)

- ☐ *No qualification. Experience only*
- ☐ *School*
- ☐ *FET College*
- ☐ *Teacher training college*
- ☐ *University*
- ☐ *Other*

If other, please specify

Did the teacher obtain this qualification in the last year? (Select only one)

- ☐ *Yes*
- ☐ *No*

If yes to the above, specify (Free text)

Has the teacher benefitted from any outside intervention, resources or support in the last 12 months, other than the STELLAR/Emergent Literacy intervention?

- ☐ *Yes*
- ☐ *No*

If yes to the above, specify (Free text)

TEACHER INTERVIEW TOOL: WAVE 3

Record your current location (Not mandatory)

Date

Choose Fieldworker (Select only one)

- ☐ Fieldworker 1
- ☐ Fieldworker 2
- ☐ Fieldworker 3
- ☐ Fieldworker 4
- ☐ Fieldworker 5
- ☐ Fieldworker 6
- ☐ Fieldworker 7
- ☐ Fieldworker 8
- ☐ Fieldworker 9
- ☐ Fieldworker 10

Please select your District

Please type in the name of the District (Free text)

Please select the school

Please type in the name of the school (Free text)

PICTURES (HEADING)

ODK prompts four pictures.

Instructions:

- Please use your phone to take four pictures of each Grade 1 classroom from which the learners are drawn;
- Try to stand in the centre of the class and get a picture of each of the four walls;
- Ensure that your picture captures the entire surface of the wall to ensure that every poster or picture on the wall can be seen in the photo;
- Please ensure that the picture quality of each photo is **VERY GOOD** before clicking save.

NOTE: High quality pictures will enable the team to do an accurate and effective analysis of the classroom environment.

LEARNER INFORMATION (HEADING)

Learner 1 name (Free text)

Learner 1 CEMIS number (Numerical free text)

Learner 1 Grade (Select only one)

☐ *Grade R*

☐ *Grade 1*

Learner 1 Gender (Select only one)

☐ *Male*

☐ *Female*

Learner 1 Date of Birth (formatted)

Learner 1 Main language used at home (Select only one)

Please specify (Free text)

Repeat the above questions for all 9 learners

TEACHER INFORMATION (HEADING)

Teacher name (Free text)

Teacher PERSAL number (Free numerical text)

Teacher date of birth (Format)

Teacher Gender (Select only one)

☐ *Male*

☐ *Female*

Language of Learning and Teaching (Select only one)

☐ *English*

☐ *Afrikaans*

☐ *Xhosa*

Other, please specify (Free text)

Which language/s does the teacher use in the classroom? (Select all that apply)

- ☐ *English*
- ☐ *Afrikaans*
- ☐ *Xhosa*

Other, please specify (Free text)

How many years of experience does the teacher have teaching Grade 1?

____ Years
____ Months

Which of these resources does the teacher use in the classroom? (Select all that apply)

- ☐ *CAPS materials*
- ☐ *Self-sourced materials*
- ☐ *Other*

If other, please specify (Free Text)

Highest teacher qualification. (Select the highest)

- ☐ *Less than Matric*
- ☐ *Matric*
- ☐ *Certificate / Diploma*
- ☐ *Degree or higher*

If other, please specify (Free Text)

Does the teacher do group guided reading with the class? (Select one)

- ☐ *Yes*
- ☐ *No*

If yes, what do the other learners do while the teacher does group guided reading? (Select all that apply)

- ☐ *Work in the DBE workbooks*
- ☐ *Writing*
- ☐ *Paired reading*
- ☐ *Independent reading*

If other, please specify (Free Text)

CLASSROOM ENVIRONMENT OBSERVATION TOOL: WAVE 1

THIS SECTION IS COMPLETED ON PAPER AND THE DATA ENTERED ON SMARTPHONE LATER

Choose Fieldworker (Select only one)

- ☐ Fieldworker 1
- ☐ Fieldworker 2
- ☐ Fieldworker 3
- ☐ Fieldworker 4
- ☐ Fieldworker 5
- ☐ Fieldworker 6
- ☐ Fieldworker 7
- ☐ Fieldworker 8
- ☐ Fieldworker 9
- ☐ Fieldworker 10

Please select your District

Please type in the name of the District (Free text)

Please select the school

Please type in the name of the school

Date

INSTRUCTIONS

This Grade R Classroom Environment (Observation) Tool is intended for use in Grade R classes, and requires that information be collected both before and during the class reading lesson. There are two sections:

Section A: Classroom and Resources

Section B: Activities

Section A: This section should be completed throughout the day, when the fieldworker has an opportunity to observe the class (either between learner assessments, or during the reading hour in which the fieldworker will observe the class).

Section B: The final section of this tool is primarily intended for use during the scheduled one-hour reading period. However, some of the information will also need to be gathered throughout the day.

SECTION A: CLASSROOM AND RESOURCES

Which of the following did you see taking place? (Select all that apply)

- ☐ Chaos and lack of discipline
- ☐ Teacher “lecturing”
- ☐ Teacher participating in games/role-plays
- ☐ Whole class working together
- ☐ Small groups with directed activities

Which of the following did you see taking place? (Select all that apply)

- ☐ Children looking at books without the teacher
- ☐ Songs/rhymes
- ☐ Role-plays
- ☐ Picture sequencing
- ☐ Work with letterboxes
- ☐ Letter-sound games
- ☐ Use of (hand-made) little books
- ☐ Painting and drawing
- ☐ Writing
- ☐ Multi-sensory letter formation
- ☐ Physical activities not involving any of the above (e.g. dancing, miming, hand/body movements)

Which of these Emergent Literacy materials, if any, do you observe in the classroom? (Please select all that apply)

- ☐ STELLAR story module(s) including manuals, puppets, big book and sequence pictures
- ☐ Game bags and activity packs
- ☐ Letterboxes
- ☐ Picture cards (e.g. CVC words, letter-sound cards)
- ☐ Art and construction from the last STELLAR story module
- ☐ Examples of learners’ work
- ☐ STELLAR little books made by children

Which of the following planning resources are in the classroom? (Select all that apply)

- ☐ Class register
- ☐ School calendar marked with holidays, public holidays and special days
- ☐ Teacher’s term plans
- ☐ File containing information about learners

Teacher’s table. (Select all that apply)

- ☐ Teacher’s table tidy with a tablecloth and chair
- ☐ Lock up cupboard
- ☐ Cupboard tidy, with resources stored and labelled
- ☐ Other places to store print resources (e.g. shelves, small tables)

The reading corner. (Select only one)

- ☐ No reading corner
- ☐ Very basic with limited books, limited seating arrangements, resources or creativity
- ☐ Adequate space, resources and books, although limited creativity and could be improved
- ☐ Adequate space and comfort, resources and books, and creatively designed

Children's books. (Select only one)

- ☐ None observed
- ☐ Minimal, tatty, only a few or not well displayed/stored
- ☐ Adequately displayed/stored; shows effort although could be improved
- ☐ Well-preserved, well displayed/stored, colourful, interesting

The theme table. (Select only one)

- ☐ No theme table
- ☐ Obviously not changed regularly; few objects on display
- ☐ Very basic with limited resources and creativity; combination of some objects and some print resources displayed
- ☐ Adequate space and resources; creatively arranged interesting collection of objects, posters, books, labels)

Alphabet chart/friezes, other charts and posters. (Select only one)

- ☐ Nothing
- ☐ Minimal, tatty, inappropriate to height and age of learners
- ☐ Adequate and appropriate, although could be improved
- ☐ Well-preserved, colourful, interesting layout

Labels around the classroom. (Select only one)

- ☐ Nothing
- ☐ Minimal, tatty, inappropriate to height and age of learners
- ☐ Adequate and appropriate, although could be improved
- ☐ Well-preserved, colourful, interesting layout

Word wall. (Select only one)

- ☐ None
- ☐ Minimal, tatty, inappropriate to age of learners
- ☐ Adequate and appropriate, although could be improved
- ☐ Well-preserved, colourful, interesting layout

Games for letter recognition and vocabulary building. (Select only one)

- ☐ None
- ☐ Minimal, only a few; not well displayed/stored; inappropriate to age of learners
- ☐ Adequate and appropriate; shows effort, although could be improved
- ☐ Well-preserved, colourful, interesting; clearly stored and labelled

Toys and puppets. (Select only one)

- ☐ None
- ☐ Minimal, only a few; not well displayed/stored
- ☐ Adequate and appropriate; shows effort, although could be improved
- ☐ Well-preserved, colourful, interesting; clearly stored and labelled

Equipment for multi-sensory letter formation (e.g. Sand trays, Play dough, White board and markers (commercial or home-made, with plastic sleeves). (Select only one)

- ☐ None
- ☐ Minimal, only a few; not well displayed/stored
- ☐ Adequate and appropriate; shows effort, although could be improved
- ☐ Well-preserved, colourful, interesting; clearly stored and labelled

Writing and drawing instruments (e.g. Paints, brushes, crayons, pencils). (Select only one)

- ☐ None
- ☐ Minimal, only a few; not well displayed/stored
- ☐ Adequate and appropriate; shows effort, although could be improved
- ☐ Well-preserved, colourful, interesting; clearly stored and labelled

Pictures cards “for sequencing”. (Select only one)

- ☐ None
- ☐ Minimal, only a few; not well displayed/stored
- ☐ Adequate and appropriate; shows effort, although could be improved
- ☐ Well-preserved, colourful, interesting; clearly stored and labelled

Overall, how would you rate the amount of visual stimulation in the classroom? (e.g. how print rich is the classroom?)

- ☐ No visual stimulation
- ☐ Some visual stimulation, but limited and not always appropriate
- ☐ Adequate amount of visual stimulation, mostly appropriate
- ☐ A lot of visual stimulation, mostly appropriate

Classroom appearance. (Select all that apply)

- ☐ Brightly painted, clean and tidy
- ☐ Learners' small tables and chairs are the right size and in good condition
- ☐ Learners' chairs have chair bags
- ☐ Learners' tables have their names and alphabet strips

SECTION B: ACTIVITIES

During Reading Exercise

Number of boys present: _____

Number of girls present: _____

Is there sufficient space for the learners in the classroom? (Select only one)

- ☐ The class is over-crowded
- ☐ The class is crowded
- ☐ There is sufficient space
- ☐ The class is roomy
- ☐ The class is outside

How comfortable are learners in regard to the teacher? (Select only one)

- ☐ Learners appear afraid or uncomfortable with the teacher
- ☐ A few learners respond well and feel comfortable with the teacher and initiate interaction
- ☐ Some learners (but not all) respond well and feel comfortable with the teacher and initiate interaction
- ☐ Learners respond well and feel comfortable with the teacher and initiate interaction with him/her

How engaged are the learners? (Select only one)

- ☐ All learners appear bored and uninvolved

- ☐ A few learners are engaged and interested
- ☐ The majority of learners are engaged and interested
- ☐ All learners are engaged and interested

Before reading, the teacher: (Select all that apply)

- ☐ Is organised, has everything set up in advance
- ☐ Has a routine for changing from the previous activity
- ☐ Calms and quiets the class
- ☐ Shows the cover and discusses/asks questions about the pictures
- ☐ Reads the title and author
- ☐ Asks what the story could be about

During reading, the teacher: (Select all that apply)

- ☐ Shows enthusiasm and interest
- ☐ Reads clearly and with good intonation
- ☐ Shows the children the pictures on each page
- ☐ Points to words and explains some of them
- ☐ Stimulates awareness of sounds and letters
- ☐ Asks literal questions (*e.g. what is the boy wearing?*)
- ☐ Asks inferential questions (*why questions, e.g. Do you think the girl is frightened? Why do you think so?*)
- ☐ Asks questions that help children connect the story to their own lives (*e.g. what would you have done?*)
- ☐ Asks questions that help children see what they can learn from the story

CLASSROOM ENVIRONMENT OBSERVATION TOOL: WAVE 2

THIS SECTION IS COMPLETED ON PAPER AND THE DATA ENTERED ON SMARTPHONE LATER

Choose Fieldworker (Select only one)

- ☐ Fieldworker 1
- ☐ Fieldworker 2
- ☐ Fieldworker 3
- ☐ Fieldworker 4
- ☐ Fieldworker 5
- ☐ Fieldworker 6
- ☐ Fieldworker 7
- ☐ Fieldworker 8
- ☐ Fieldworker 9
- ☐ Fieldworker 10

Please select your District

Please type in the name of the District (Free text)

Please select the school

Please type in the name of the school

Date

INSTRUCTIONS

This Grade R Classroom Environment (Observation) Tool is intended for use in Grade R classes, and requires that information be collected both before and during the class reading lesson. There are two sections:

Section A: Classroom and Resources

Section B: Activities

Section A: This section should be completed throughout the day, when the fieldworker has an opportunity to observe the class (either between learner assessments, or during the reading hour in which the fieldworker will observe the class).

Section B: The final section of this tool is primarily intended for use during the scheduled one-hour reading period. However, some of the information will also need to be gathered throughout the day.

SECTION A: CLASSROOM AND RESOURCES

Which of the following did you see taking place? (Select all that apply)

- ☐ Chaos and lack of discipline
- ☐ Teacher “lecturing”
- ☐ Teacher participating in games/role-plays
- ☐ Whole class working together
- ☐ Small groups with directed activities

Which of the following did you see taking place? (Select all that apply)

- ☐ Children looking at books without the teacher
- ☐ Songs/rhymes
- ☐ Role-plays
- ☐ Picture sequencing
- ☐ Work with letterboxes
- ☐ Letter-sound games
- ☐ Use of (hand-made) little books
- ☐ Painting and drawing
- ☐ Writing
- ☐ Multi-sensory letter formation
- ☐ Physical activities not involving any of the above (e.g. dancing, miming, hand/body movements)

Which of these Emergent Literacy materials, if any, do you observe in the classroom? (Please select all that apply)

- ☐ STELLAR story module(s) including manuals, puppets, big book and sequence pictures
- ☐ Game bags and activity packs
- ☐ Letterboxes
- ☐ Picture cards (e.g. CVC words, letter-sound cards)
- ☐ Art and construction from the last STELLAR story module
- ☐ Examples of learners’ work
- ☐ STELLAR little books made by children

Which of the following planning resources are in the classroom? (Select all that apply)

- ☐ Class register
- ☐ School calendar marked with holidays, public holidays and special days
- ☐ Teacher’s term plans
- ☐ File containing information about learners

Teacher’s table. (Select all that apply)

- ☐ Teacher’s table tidy with a tablecloth and chair
- ☐ Lock up cupboard
- ☐ Cupboard tidy, with resources stored and labelled
- ☐ Other places to store print resources (e.g. shelves, small tables)

The reading corner. (Select only one)

- ☐ No reading corner
- ☐ Very basic with limited books, limited seating arrangements, resources or creativity
- ☐ Adequate space, resources and books, although limited creativity and could be improved
- ☐ Adequate space and comfort, resources and books, and creatively designed

Children's books. (Select only one)

- ☐ None observed
- ☐ Minimal, tatty, only a few or not well displayed/stored
- ☐ Adequately displayed/stored; shows effort although could be improved
- ☐ Well-preserved, well displayed/stored, colourful, interesting

The theme table. (Select only one)

- ☐ No theme table
- ☐ Obviously not changed regularly; few objects on display
- ☐ Very basic with limited resources and creativity; combination of some objects and some print resources displayed
- ☐ Adequate space and resources; creatively arranged interesting collection of objects, posters, books, labels)

Alphabet chart/friezes, other charts and posters. (Select only one)

- ☐ Nothing
- ☐ Minimal, tatty, inappropriate to height and age of learners
- ☐ Adequate and appropriate, although could be improved
- ☐ Well-preserved, colourful, interesting layout

Labels around the classroom. (Select only one)

- ☐ Nothing
- ☐ Minimal, tatty, inappropriate to height and age of learners
- ☐ Adequate and appropriate, although could be improved
- ☐ Well-preserved, colourful, interesting layout

Word wall. (Select only one)

- ☐ None
- ☐ Minimal, tatty, inappropriate to age of learners
- ☐ Adequate and appropriate, although could be improved
- ☐ Well-preserved, colourful, interesting layout

Games for letter recognition and vocabulary building. (Select only one)

- ☐ None
- ☐ Minimal, only a few; not well displayed/stored; inappropriate to age of learners
- ☐ Adequate and appropriate; shows effort, although could be improved
- ☐ Well-preserved, colourful, interesting; clearly stored and labelled

Toys and puppets (select only one):

- ☐ None
- ☐ Minimal, only a few; not well displayed/stored
- ☐ Adequate and appropriate; shows effort, although could be improved
- ☐ Well-preserved, colourful, interesting; clearly stored and labelled

Equipment for multi-sensory letter formation (e.g. Sand trays, Play dough, White board and markers (commercial or home-made, with plastic sleeves). (Select only one)

- ☐ None
- ☐ Minimal, only a few; not well displayed/stored
- ☐ Adequate and appropriate; shows effort, although could be improved
- ☐ Well-preserved, colourful, interesting; clearly stored and labelled

Writing and drawing instruments (e.g. Paints, brushes, crayons, pencils). (Select only one)

- ☐ None
- ☐ Minimal, only a few; not well displayed/stored
- ☐ Adequate and appropriate; shows effort, although could be improved
- ☐ Well-preserved, colourful, interesting; clearly stored and labelled

Pictures cards “for sequencing”. (Select only one)

- ☐ None
- ☐ Minimal, only a few; not well displayed/stored
- ☐ Adequate and appropriate; shows effort, although could be improved
- ☐ Well-preserved, colourful, interesting; clearly stored and labelled

Overall, how would you rate the amount of visual stimulation in the classroom? (e.g. how print rich is the classroom?)

- ☐ No visual stimulation
- ☐ Some visual stimulation, but limited and not always appropriate
- ☐ Adequate amount of visual stimulation, mostly appropriate
- ☐ A lot of visual stimulation, mostly appropriate

Classroom appearance. (Select all that apply)

- ☐ Brightly painted, clean and tidy
- ☐ Learners' small tables and chairs are the right size and in good condition
- ☐ Learners' chairs have chair bags
- ☐ Learners' tables have their names and alphabet strips

SECTION B: ACTIVITIES

During Reading Exercise

Number of boys present: _____

Number of girls present: _____

Is there sufficient space for the learners in the classroom? (Select only one)

- ☐ The class is over-crowded
- ☐ The class is crowded
- ☐ There is sufficient space
- ☐ The class is roomy
- ☐ The class is outside

How comfortable are learners in regard to the teacher? (Select only one)

- ☐ Learners appear afraid or uncomfortable with the teacher
- ☐ A few learners respond well and feel comfortable with the teacher and initiate interaction
- ☐ Some learners (but not all) respond well and feel comfortable with the teacher and initiate interaction
- ☐ Learners respond well and feel comfortable with the teacher and initiate interaction with him/her

How engaged are the learners? (Select only one)

- ☐ All learners appear bored and uninvolved
- ☐ A few learners are engaged and interested
- ☐ The majority of learners are engaged and interested
- ☐ All learners are engaged and interested

Before reading, the teacher: (Select all that apply)

- ☐ Is organised, has everything set up in advance
- ☐ Has a routine for changing from the previous activity
- ☐ Calms and quiets the class
- ☐ Shows the cover and discusses/asks questions about the pictures
- ☐ Reads the title and author
- ☐ Asks what the story could be about

During reading, the teacher: (Select all that apply)

- ☐ Shows enthusiasm and interest
- ☐ Reads clearly and with good intonation
- ☐ Shows the children the pictures on each page
- ☐ Points to words and explains some of them
- ☐ Stimulates awareness of sounds and letters
- ☐ Asks literal questions (*e.g. what is the boy wearing?*)
- ☐ Asks inferential questions (*why questions, e.g. Do you think the girl is frightened? Why do you think so?*)
- ☐ Asks questions that help children connect the story to their own lives (*e.g. what would you have done?*)
- ☐ Asks questions that help children see what they can learn from the story

CLASSROOM ENVIRONMENT OBSERVATION TOOL: WAVE 3

THIS SECTION IS COMPLETED ON PAPER AND THE DATA ENTERED ON SMARTPHONE LATER

Choose Fieldworker (Select only one)

- ☐ Fieldworker 1
- ☐ Fieldworker 2
- ☐ Fieldworker 3
- ☐ Fieldworker 4
- ☐ Fieldworker 5
- ☐ Fieldworker 6
- ☐ Fieldworker 7
- ☐ Fieldworker 8
- ☐ Fieldworker 9
- ☐ Fieldworker 10

Please select the district

Please type in the name of the district (Free text)

Please select the school

Please type in the name of the school (Free text)

Date (Free text)

INSTRUCTIONS

This Grade 1 Classroom Environment (Observation) Tool is intended for use in all Grade 1 classes in which the tracked learners are located, and requires that information be collected at any point in the school day (for example, during the lunch break). The fieldworker will NOT observe classroom activities but rather rate the classroom on appearance alone.

The reading corner (select only one):

- ☐ No reading corner
- ☐ Very basic with limited books, limited seating arrangements, resources or creativity
- ☐ Adequate space, resources and books, although limited creativity and could be improved
- ☐ Adequate space and comfort, resources and books, and creatively designed

Children's books (select only one):

- ☐ None observed
- ☐ Minimal, tatty, only a few or not well displayed/stored
- ☐ Adequately displayed/stored; shows effort although could be improved
- ☐ Well-preserved, well displayed/stored, colourful, interesting

Alphabet chart/friezes; other charts and posters (select only one):

- ☐ Nothing
- ☐ Minimal, tatty, inappropriate to height and age of learners
- ☐ Adequate and appropriate, although could be improved
- ☐ Well-preserved, colourful, interesting layout

Overall, how would you rate the amount of visual stimulation in the classroom (e.g. how print rich is the classroom?)

- ☐ No visual stimulation
- ☐ Some visual stimulation, but limited and not always appropriate
- ☐ Adequate amount of visual stimulation, mostly appropriate
- ☐ A lot of visual stimulation, mostly appropriate

Classroom appearance (select all that apply):

- ☐ Brightly painted, clean and tidy
- ☐ Learners' small tables and chairs are the right size and in good condition
- ☐ Learners' chairs have chair bags
- ☐ Learners' tables have their names and alphabet strips

GRADE R TEACHER KNOWLEDGE, ATTITUDES AND PRACTICES SURVEY

The following questions consider your knowledge, attitudes, and practices related to teaching in Grade R. Your answers will help us to better organize and understand Grade R education in the Province, and will be used for research purposes only. Your name will not be used in any reports and your responses will be kept confidential and anonymous. Thank you for your participation.

SCHOOL NAME (PLEASE PRINT):

TEACHER NAME (PLEASE PRINT):

PLEASE MAKE A MARK (*) TO THE LEFT OF THE QUESTION IF A QUESTION IS UNCLEAR.

SECTION 1

For each statement given below, please indicate whether you agree or disagree.

No	Statement	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
1	Songs, rhythm and rhyme are fun but not very important for literacy and language learning.					
2	Fantasy play helps learners develop advanced forms of language.					
3	Children who enter school with limited vocabulary and poor oral language skills will struggle with reading and writing.					
4	Children only begin to learn to read and write in Grade One.					
5	Parents should not be involved in developing their children's literacy and language as they are not trained to be teachers.					
6	If children are learning in their second language at school, they should be discouraged from using their mother tongue in the playground and at home.					

No	Statement	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
7	Young children should be encouraged to experiment with writing before Grade One.					
8	Children should be discouraged from writing words until they can spell them properly.					
9	The best way to learn how to form letters in Grade R is using big movements and multisensory activities.					
10	Working with sequence pictures develops children's phonemic awareness					
11	An example of a syllable blending question is: Which word would you have if you put these parts of words together: 'to-ma-to'?					
12	Role-play is important for language development because children have the opportunity to use the language of a story and make it their own.					
13	To encourage a child's language development the teacher should use simple words.					
14	Letters have both names and sounds. It is most important for emergent readers to know the names of letters.					
15	If a Grade R child brings you something, he has 'written' with some letters facing the wrong way, it is more important to praise him for trying, than to correct his letter formation.					

SECTION 2

Tick the appropriate column to match the learner competency with your understanding of language and literacy.

	Learner competency:	AREA OF LANGUAGE AND LITERACY			
		Phonological Awareness	Print Concepts	Letter knowledge	Literate language
1.	Knows the meaning of unusual words				
2.	Identifies all the letters of the alphabet				
3.	Moves their finger from left to right across the page of a book when pretend reading				
4.	Identifies the first sound in a spoken word				
5.	Able to clap out the syllables in a word				

Tick the appropriate column to match the learner competency with your understanding of the underlying skill.

	Learner competency:	UNDERLYING SKILL		
		Auditory processing	Visual processing	Fine motor skills
6.	Follows verbal instructions			
7.	Uses crayons and pencils with ease			
8.	Notices detail in pictures			
9.	Can recognise letters in the environment			
10.	Can distinguish between different sounds at the start of words			

SECTION 3

The statements in this tool are used to assess the current as well as the preferred classroom context. For each statement below, tick the box to the left indicating the **current practice** in your classroom (please be accurate – this information will be used anonymously to inform the improvement of Grade R teacher training programmes), and tick the box to the right of the statement indicating how much change is needed in your **preferred** classroom context. For example:

Existing Classroom Context				STATEMENT	Preferred Classroom Context			
How is the situation at the moment (current practice in your Grade-R classroom)?					How much change is needed in your classroom?			
Often/ regularly	Sometimes/ occasionally	Seldom/ rarely	Never		No change required	Improvement needed in med term	Urgent and immediate improvement necessary	Don't know
Section A: Principles								
☐	☐	☐	☐	Storytelling is used to help learners to become familiar with the language of books.	☐	☐	☐	☐

Existing Classroom Context				STATEMENT	Preferred Classroom Context			
How is the situation at the moment (current practice in your Grade R classroom)?					How much change is needed in your Grade R classroom?			
Often/ regularly	Sometimes/ occasionally	Seldom/ rarely	Never		No change required	Improvement needed in medium term	Urgent and immediate improvement necessary	Don't know
PART A								
☐	☐	☐	☐	1. Storytelling is used to help learners to become familiar with the language of books.	☐	☐	☐	☐
☐	☐	☐	☐	2. Learners are given many opportunities to engage with the same stories and build up story vocabulary.	☐	☐	☐	☐
☐	☐	☐	☐	3. Drawing, mark-making, and "pretend" reading and writing are encouraged.	☐	☐	☐	☐
☐	☐	☐	☐	4. Learners are given opportunities to try "have-a-go" writing.	☐	☐	☐	☐
☐	☐	☐	☐	5. Teachers interact with learners around language and literacy at any time of the day, not just during lessons.	☐	☐	☐	☐
☐	☐	☐	☐	6. Nurturing, developmental, teacher-learner relationships are a top priority.	☐	☐	☐	☐
☐	☐	☐	☐	7. Ideas and resources are shared with parents to strengthen what is being done in the Grade R class.	☐	☐	☐	☐
PART B								
☐	☐	☐	☐	8. Teachers use a learning framework and teaching program to guide Grade R classroom activities.	☐	☐	☐	☐
☐	☐	☐	☐	9. Teachers integrate literacy materials from different sources (e.g. CAPS, Grade R Resource Kit, other).	☐	☐	☐	☐
☐	☐	☐	☐	10. Discipline is balanced with teacher-learner relationships and keeping learner attention.	☐	☐	☐	☐
☐	☐	☐	☐	11. Learners practice listening and speaking (conversation) skills through role-playing and dramatization of stories.	☐	☐	☐	☐

Existing Classroom Context				STATEMENT	Preferred Classroom Context			
How is the situation at the moment (current practice in your Grade R classroom)?					How much change is needed in your Grade R classroom?			
Often/ regularly	Sometimes/ occasionally	Seldom/ rarely	Never		No change required	Improvement needed in medium term	Urgent and immediate improvement necessary	Don't know
☐	☐	☐	☐	12. Learners are taught the sounds and form of letters and to build up words from sounds and sentences from words.	☐	☐	☐	☐
☐	☐	☐	☐	13. Learners are taught the construction of language by clapping to syllables, blending words and repetition.	☐	☐	☐	☐
☐	☐	☐	☐	14. Learners are encouraged to read.	☐	☐	☐	☐
☐	☐	☐	☐	15. Learners are encouraged to use writing instruments and to draw.	☐	☐	☐	☐
☐	☐	☐	☐	16. Art and construction, and movement to rhymes, rhythms and songs are encouraged.	☐	☐	☐	☐
☐	☐	☐	☐	17. Teachers organize and store literacy material and resources.	☐	☐	☐	☐
☐	☐	☐	☐	18. Learners are taught to organize their "thinking" through sequencing of picture cards.	☐	☐	☐	☐
☐	☐	☐	☐	19. Picture cards are used during story-telling.	☐	☐	☐	☐
☐	☐	☐	☐	20. Teachers read stories and texts aloud.	☐	☐	☐	☐
☐	☐	☐	☐	21. Learner vocabulary is built through repetition, use of labels and reinforcement.	☐	☐	☐	☐
☐	☐	☐	☐	22. Songs and rhymes are taught in the classroom.	☐	☐	☐	☐

GRADE 1 TEACHER KNOWLEDGE, ATTITUDES AND PRACTICES SURVEY

The following questions consider your knowledge and concepts related to teaching in Grade R. Your answers will help us to better organize and understand Grade R education in the Province, and will be used for research purposes only. Your name will not be used in any reports and your responses will be kept confidential and anonymous. Thank you for your participation.

SCHOOL NAME (PLEASE PRINT):

TEACHER NAME (PLEASE PRINT):

PLEASE MAKE A MARK (*) TO THE LEFT OF THE QUESTION IF A QUESTION IS UNCLEAR.

SECTION 1

For each statement given below, please indicate whether you agree or disagree.

No	Statement	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
1	Songs, rhythm and rhyme are fun but not very important for literacy and language learning.					
2	Fantasy play helps learners develop advanced forms of language.					
3	Children who enter school with limited vocabulary and poor oral language skills will struggle with reading and writing.					
4	Children only begin to learn to read and write in Grade One.					
5	Parents should not be involved in developing their children's literacy and language, as they are not trained to be teachers.					
6	If children are learning in their second language at school, they should be discouraged from using their mother tongue in the playground and at home.					
7	Young children should be encouraged to experiment with writing before Grade One.					
8	Children should be discouraged from writing words until they can spell them properly.					

No	Statement	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
9	The best way to learn how to form letters in Grade R is using big movements and multisensory activities.					
10	Working with sequence pictures develops children's phonemic awareness					
11	An example of a syllable blending question is: Which word would you have if you put these parts of words together: 'to-ma-to'?					
12	Role-play is important for language development because children have the opportunity to use the language of a story and make it their own.					
13	To encourage a child's language development the teacher should use simple words.					
14	Letters have both names and sounds. It is most important for emergent readers to know the names of letters.					
15	If a Grade R child brings you something, he has 'written' with some letters facing the wrong way, it is more important to praise him for trying, than to correct his letter formation.					

SECTION 2

Tick the appropriate column to match the learner competency with your understanding of language and literacy.

	Learner competency:	AREA OF LANGUAGE AND LITERACY			
		Phonological Awareness	Print Concepts	Letter knowledge	Literate language
1.	Knows the meaning of unusual words				
2.	Identifies all the letters of the alphabet				
3.	Moves their finger from left to right across the page of a book when pretend reading				
4.	Identifies the first sound in a spoken word				
5.	Able to clap out the syllables in a word				

Tick the appropriate column to match the learner competency with your understanding of the underlying skill.

		UNDERLYING SKILL		
	Learner competency:	<i>Auditory processing</i>	<i>Visual processing</i>	<i>Fine motor skills</i>
6.	Follows verbal instructions			
7.	Uses crayons and pencils with ease			
8.	Notices detail in pictures			
9.	Can recognise letters in the environment			
10.	Can distinguish between different sounds at the start of words			

PACING AND CURRICULUM COVERAGE TOOL

This pacing and curriculum coverage tool assesses Grade 1 learner writing and comprises five main assessment areas namely:

1. Curriculum coverage
2. Learner feedback
3. Quality of writing
4. Creative writing
5. Length of writing

INSTRUCTIONS AND INITIAL QUESTIONS

On the day of the evaluation:

- Randomly select 4 Grade 1 learners from the list you have been provided. In order to randomly select the 4 Grade 1 learners, select every **SECOND LEARNER** from the list of Grade 1 learners. This is essentially the same random selection method you used in August. However, now, you are required to randomly select the 4 learners from the list which holds the previously assessed learner.
- When selecting the learners randomly, you might find that the learners you have selected might be in a separate class.
- When you collect the Grade 1 learners for the learner assessment group activity, please ask each teacher to set aside **ALL LEARNER RESOURCES INCLUDING THE DBE WORKBOOKS** used by each of these learners.
- After reading the instructions, mark the specific pages on which you will assess creativity, and length of writing, using section dividers.

DATE _____

FIELDWORKER NAME _____

SCHOOL NAME _____

Learner number	Learner Name	Teacher Name
Learner 1		
Learner 2		
Learner 3		
Learner 4		

Teacher number	Teacher Name
Teacher 1	
Teacher 2	
Teacher 3	
Teacher 4	

Materials used by the teacher. (Mark ALL that apply – please ask the teacher)

Teacher	DBE Workbook	Learner class workbook	Worksheets	Additional exercises	Homework books	Other, specify
Teacher 1						
Teacher 2						
Teacher 3						
Teacher 4						

What is the primary resource used by the teacher? (Mark ONE – please ask the teacher)

Teacher	DBE Workbook	Learner class workbook	Worksheets	Additional exercises	Homework books	Other, specify
Teacher 1						
Teacher 2						
Teacher 3						
Teacher 4						

If worksheets are available, where are they kept? (Mark ONE)

Teacher	In a file/folder	Disorganised	Not applicable
Teacher 1			
Teacher 2			
Teacher 3			
Teacher 4			

PART 1: CURRICULUM COVERAGE

DBE workbooks

Note: In the DBE workbooks, some pages do not require written input from the learner (e.g. let's speak or let's read). Some pages require that the learner complete written exercises. Only count the pages on which the learner is expected to complete written exercises for Term 3 & up to end of the 2nd week Term 4.

Learner number	Total number of Pages	Total number of blank pages	Total number of pages with partially completed written exercises (less than 50%)	Total number of pages with partially completed written exercises (50% and above)	Total number of pages with fully completed exercises
Learner 1					
Learner 2					
Learner 3					
Learner 4					

All Other materials (e.g. learner class workbook, worksheet portfolio etc.)

Note: Now count the ACTUAL number of pages blank/ partially completed/fully completed. Please ensure that you review all available written work other than the DBE workbook for the assessment.

Learner number		Total number of Pages	Number of blank pages (Exercises not at all attempted)	Number of pages with partially completed written exercises (less than 50%)	Number of pages with partially completed written exercises (50% and above)	Number of pages with fully completed exercises
Learner 1	Actual					
Learner 2	Actual					
Learner 3	Actual					
Learner 4	Actual					

PART 2: TEACHER FEEDBACK

DBE workbooks

Note: In the DBE workbooks, assess teacher feedback in terms of learner written work, from Term 3 & up to end of the 2nd week Term 4.

Count the total number pages that show evidence of ANY teacher feedback.

Learner Number	Number of pages with evidence of ANY teacher feedback
Learner 1	
Learner 2	
Learner 3	
Learner 4	

Now rate (mark with an x) the overall quality (holistic appraisal) of teacher feedback. (Select the score that represents the most advanced form of teacher feedback evident)

Learner number	0 – No feedback	1 – Ticks and marks only	2 – Ticks and marks with motivational comments, e.g. “well done”	3 - Evidence of corrective feedback, e.g. marked incorrect and provided / gave example of the correction.
Learner 1				
Learner 2				
Learner 3				
Learner 4				

All other materials

Note: In the other materials, assess teacher feedback in all available written work.

Count the total number pages that show evidence of ANY teacher feedback.

Learner Number	Number of pages with evidence of ANY teacher feedback
Learner 1	
Learner 2	
Learner 3	
Learner 4	

Now rate (mark with an x) the overall quality of teacher feedback. (Select the score that represents the most advanced form of teacher feedback evident)

Learner number	0 – No feedback	1 – Ticks and marks only	2 – Ticks and marks with motivational comments, e.g. “well done”	3 - Evidence of corrective feedback, e.g. marked incorrect and provided / gave example of the correction.
Learner 1				
Learner 2				
Learner 3				
Learner 4				

PART 3: QUALITY OF WRITING

DBE workbooks

Assess three exercises where learners attempted to write 2 – 3 sentences about a **given picture/story**. Assess ANY three exercises in term 3 and up to the second week in term 4 where learners attempted to write 2 – 3 sentences.

Now rate the overall quality of learner writing about a given picture/story in these exercises. (Select the score that represents the most advanced quality of learner writing evident)

Learner name	1 – Learner wrote simple words only	2 – Learner wrote SHORT simple sentences WITH punctuation	3 - Learner wrote LONGER, MORE COMPLEX sentences WITHOUT punctuation	4 - Learner wrote LONGER, MORE COMPLEX sentences WITH punctuation
Learner 1				
Learner 2				
Learner 3				
Learner 4				

Assess three exercises where learners attempted to write 2 – 3 sentences about a **self-drawn story**. Assess ANY three exercises in term 3 and up to the second week in term 4 where learners attempted to write 2 – 3 sentences.

Now rate the overall quality of learner writing relating to a self-drawn picture about a given story in these exercises. (Select the score that represents the most advanced quality of learner writing evident)

Learner name	1 – Learner wrote simple words only	2 – Learner wrote SHORT simple sentences WITH punctuation	3 - Learner wrote LONGER, MORE COMPLEX sentences WITHOUT punctuation	4 - Learner wrote LONGER, MORE COMPLEX sentences WITH punctuation
Learner 1				
Learner 2				
Learner 3				
Learner 4				

All other materials

Assess ANY three exercises in term 3 and up to the second week in term 4 where learners attempted to write 2 – 3 sentences.

Rate the overall quality of the learners' writing. (Select the score that represents the most advanced form of learner writing available in the materials)

Learner name	1 – Learner wrote simple words only	2 – Learner wrote SHORT simple sentences WITH punctuation	3 - Learner wrote LONGER, MORE COMPLEX sentences WITHOUT punctuation	4 - Learner wrote LONGER, MORE COMPLEX sentences WITH punctuation
Learner 1				
Learner 2				
Learner 3				
Learner 4				

PART 4: CREATIVITY OF WRITING

DBE workbooks

Assess three exercises where learners attempted to write 2 – 3 sentences about a self-drawn story. Assess ANY three exercises in term 3 and up to the second week in term 4 where learners attempted to write 2 – 3 sentences.

Rate the overall creativity of learner writing about a self-drawn story: (Select one)

Learner number	1 – Not applicable – no examples of writing.	2 – Written sentences are off topic and do not relate to the given picture/story at all	3 – Written sentences relate to given story/picture but show some imagination (are created/made up) by the learner.	4 – Written sentences relate literally to the given story/picture in the DBE workbook.
Learner 1				
Learner 2				
Learner 3				
Learner 4				

All other materials

Find three example of creative writing in term 3 and up to second week of term 4 where learners have attempted to write 2-3 sentences about a story/topic.

Now rate the overall creativity of the learners' writing. (Select the score that represents the most advanced form of learner writing available in the materials)

Learner name	1 – Not applicable – no examples of writing.	2 – Written sentences are off topic and do not relate to the given story at all	3 – Written sentences relate to given story but show some imagination (are created/made up) by the learner.	4 – Written sentences relate literally to the given story
Learner 1				
Learner 2				
Learner 3				
Learner 4				

PART 5: LENGTH OF WRITING

DBE workbooks

Assess three exercises where learners attempted to write 2 sentences about a picture. Assess ANY three exercises in term 3 and up to the second week in term 4 where learners attempted to write 2 – 3 sentences.

Now rate the overall length of the sentences written by the learner: (Select one)

Learner name	1 – Not applicable – nothing written	2 – Learner only wrote 1 sentence	3 – Learner wrote 2 sentences	4 – Learner wrote more than the given number of sentences
Learner 1				
Learner 2				
Learner 3				
Learner 4				

All other materials

Select any three exercises done in term 3 and up to the second week of term 4 where learners attempted to write 1 or more sentences about a given picture/story/topic.

Now rate the overall length of the sentences written by the learner (select one):

Learner name	1 – Not applicable – nothing written	2 – Learner only wrote 1 sentence	3 – Learner wrote 2 sentences	4 – Learner wrote more than the given number of sentences
Learner 1				
Learner 2				
Learner 3				
Learner 4				

ELIT Programme Sustainability Online Survey

What is program sustainability capacity?

This tool defines program sustainability capacity as the ability to maintain programming and its benefits over time. The tool assumes that the following domains influence a project or program's capacity for sustainability:

- Environmental support: Having a supportive internal and external climate for the program.
- Funding sustainability: Establishing a consistent financial base for the program.
- Partnerships: Cultivating connections between the program and its stakeholders.
- Organizational capacity: Having the internal support and resources needed to effectively manage the program.
- Program evaluation: Assessing the program to inform planning and document results.
- Program adaptation: Taking actions that adapt the program to ensure its ongoing effectiveness.
- Communications: Strategic communication with stakeholders and the public about the program.
- Strategic planning: Using processes that guide the program's directions, goals, and strategies.

Why is program sustainability capacity important?

Programs at all levels and settings struggle with their sustainability capacity. Unfortunately, when programs end, often hard-won improvements can dissolve. To maintain these benefits to society, stakeholders must understand all of the factors that contribute to program sustainability. With knowledge of these critical factors, stakeholders can build program capacity for sustainability and position their efforts for long-term success.

What is the purpose of this tool?

This tool will enable the program implementers and evaluators to assess the current capacity of the Western Cape Education Department Emergent Literacy Project (ELIT) for sustainability across a range of specific organizational and contextual factors. Your responses will help identify sustainability strengths and challenges, which can be used to guide sustainability planning for the ELIT project.

Licensing

This assessment tool is adapted from: the Program Sustainability Assessment Tool, copyright 2012, Washington University, St. Louis, MO. All rights reserved. If you would like more information about the original framework or Program Sustainability Assessment Tool, visit <http://www.sustaintool.org>.

Creative Commons Attribution-Non-commercial-ShareAlike 3.0 Unported

Helpful definitions

This tool has been designed for use with a wide variety of program, both large and small, across different settings. Given this flexibility, it is important to understand how "program", "organization", and "community" are defined before starting this assessment.

Below are a few definitions of terms that are frequently used throughout the tool.

- Program refers to the set of formal organized activities that you want to sustain over time. In this case, "program" refers to the ELIT project for teaching language and literacy in Grade R that aims to support Grade R learners in becoming self-assured listeners, speakers and emergent writers and readers.
- Organization encompasses all the parent organizations or agencies in which the program is housed. In this case, "organization" refers to the Western Cape Education Department, in general, and the Education District Office in which you work, in particular.
- Community refers to the stakeholders who may benefit from or who may guide the program. This could include parents of learners that are receiving the intervention, other local residents, organizational leaders, decision-makers, local Non-Profit Organizations, etc. Community does not refer to a specific town or neighbourhood.

In the following questions, you will rate the ELIT program across a range of specific factors that affect its sustainability. Please respond to as many items as possible. If you truly feel you are not able to answer an item, you may select "NA." For each statement, select the rating that best indicates the extent to which the ELIT program has or does the following things.

1. ENVIRONMENTAL SUPPORT: HAVING A SUPPORTIVE INTERNAL AND EXTERNAL CLIMATE FOR YOUR PROGRAMME

Statement	Too little or no extent 1	2	3	4	5	6	To a very great extent 7	Not able to answer N/A
Champions exist who strongly support the programme								
The programme has strong champions with the ability to gather/garner resources.								
The programme has leadership support from within the larger organization (WCED).								
The programme has leadership support from outside of the organization.								
The programme has strong public support.								

Comments, if any:

2. FUNDING STABILITY: ESTABLISHING A CONSISTENT FINANCIAL BASE FOR YOUR PROGRAMME

Statement	Too little or no extent 1	2	3	4	5	6	To a very great extent 7	Not able to answer N/A
Diverse community organizations are invested in the success of the programme.								
The programme communicates with community leaders.								
Community leaders are involved with the programme.								
Community members are passionately committed to the programme.								
The community is engaged in the development of programme goals.								

Comments, if any:

3. PARTNERSHIPS: CULTIVATING CONNECTIONS BETWEEN THE PROGRAMME AND ITS STAKEHOLDERS

Statement	Too little or no extent 1	2	3	4	5	6	To a very great extent 7	Not able to answer N/A
The programme exists in a supportive economic climate.								
The programme implements policies to help ensure sustained funding.								
The programme is funded through a variety of sources.								
The programme has a combination of stable and flexible funding.								
The programme has sustained funding.								

Comments, if any:

4. ORGANIZATIONAL CAPACITY: HAVING THE INTERNAL SUPPORT AND RESOURCES NEEDED TO EFFECTIVELY MANAGE THE PROGRAMME AND ITS ACTIVITIES

Statement	Too little or no extent 1	2	3	4	5	6	To a very great extent 7	Not able to answer N/A
The programme is well integrated into the operations of the organization (WCED and/or Education District Office).								
Organizational systems are in place to support the various programme needs.								
Leadership effectively articulates the vision of the programme to external partners.								
Leadership efficiently manages staff and other resources.								
The programme has adequate staff to complete the programme's goals.								

Comments, if any:

5. PROGRAMME EVALUATION: ASSESSING THE PROGRAMME TO INFORM PLANNING AND DOCUMENT RESULTS

Statement	Too little or no extent 1	2	3	4	5	6	To a very great extent 7	Not able to answer N/A
The programme has the capacity for quality programme evaluation.								
The programme reports short term and intermediate outcomes.								
Evaluation results inform programme planning and implementation.								
Programme evaluation results are used to demonstrate successes to funders and other key stakeholders.								
The programme provides strong evidence to the public that the programme works.								

Comments, if any:

6. PROGRAMME ADAPTATION: TAKING ACTIONS THAT ADAPT THE PROGRAMME TO ENSURE ITS ONGOING EFFECTIVENESS

Statement	Too little or no extent 1	2	3	4	5	6	To a very great extent 7	Not able to answer N/A
The programme periodically reviews the evidence base.								
The programme adapts strategies as needed.								
The programme adapts to new evidence.								
The programme proactively adapts to changes in the environment.								
The programme makes decisions about which components are ineffective and should not continue.								

Comments, if any:

7. COMMUNICATIONS: STRATEGIC COMMUNICATION WITH STAKEHOLDERS AND THE PUBLIC ABOUT THE PROGRAMME

Statement	Too little or no extent 1	2	3	4	5	6	To a very great extent 7	Not able to answer N/A
The programme has communication strategies to secure and maintain public support.								
Programme staff communicate the need for the programme to the public.								
The programme is marketed in a way that generates interest.								
The programme increases community awareness of the issue.								
The programme demonstrates its value to the public.								

Comments, if any:

8. STRATEGIC PLANNING: USING PROCESSES THAT GUIDE THE PROGRAM'S DIRECTION, GOALS, AND STRATEGIES

Statement	Too little or no extent 1	2	3	4	5	6	To a very great extent 7	Not able to answer N/A
The programme plans for future resource needs.								
The programme has a long-term financial plan.								
The programme has a sustainability plan.								
The programs goals are understood by all stakeholders.								
The programme clearly outlines roles and responsibilities for all stakeholders.								

Comments, if any:

ELIT Programme Sustainability Key Informant Interview Guide

The purpose of the interview is to deepen our understanding of the inherent sustainability of the ELIT and Stellar projects. The semi-structured survey will be used to interview ELIT/Stellar project managers as well as project champions who played an intimate role in the implementation of the ELIT/Stellar project. The Key Informant Interviews are used to narrate the perceptions, experience and the lessons learned by project managers and champions. The interviews also allow the respondents to reflect on their contributions and whether their contributions have been implemented in a sustainable manner. As such, we would like to ensure the confidentiality of all respondents to garner honest opinions.

Suggested question	Comments
What is your understanding of the ELIT and Stellar projects? <i>(Prompts: How are they similar, how are they different, how/where are they integrated?)</i>	
How are you, or were you, involved in the ELIT and/or Stellar Programme?	
How would you describe program sustainability? (In the context of the ELIT and Stellar programs and associated activities) a) What has worked well that you think has improved the ELIT and Stellar programs potential for sustainability? b) Do you have any concerns specifically around the short and/or long-term sustainability of this initiative? (Factors that could hinder sustainability)	
Do you think the program champions have the necessary skills to manage program activities effectively (During and after)? Prompt: Stellar / ELIT	
Are there any funding mechanisms in place to ensure the ongoing sustainability of the ELIT and Stellar program? (Such as funds for training of new teachers, circulation of new materials, etc.) If so, please would you explain how these would work? If not, what do you think will happen without these funds?" <i>Prompt: What mechanism was used during the life cycle, has that changed / how has it changed, what is being used now?</i>	
To what extent did the ELIT and Stellar program involve the community?	

Suggested question	Comments
Would you say the ELIT and Stellar program are well integrated into the WCED? a) Provincial and district level b) Are there champions responsible for managing it? c) If staff leave, would new staff be able to take over ownership?	
The training for the ELIT and Stellar program ended last year. Since then, what activities have been ongoing? <i>Prompt: Teacher training, district level, provincial level</i>	
Do you think teachers have been sufficiently trained, and are there any plans for future training? a) What is/was the plan for training new teachers entering the system?	
Have teachers/schools been provided with sufficient resources? (Such as learning materials, resources, etc.) a) Will additional resources be needed in the future? b) Is there a plan to provide additional resources? <i>(Prompt: Frequency of provision? New teachers. Teachers who have not yet received resources. Was there EVER a plan in place?)</i>	
What do you think are the main challenges/threats to the sustainability of the ELIT and Stellar programs?	
Will teachers continue to use Stellar resources? <i>(Prompt: District / Provincial level, teacher-dependant)</i>	
What would be required to make the ELIT and Stellar programs sustainable? a) If they think it is sustainable, ask what would be required in the future b) If they do not think it's sustainable, ask what is needed	

ELIT Cost Templates

WORDWORKS COST TEMPLATE (COMPLETED)

ESTIMATED COST TO WORDWORKS OF ELIT ROLL OUT 2015/6 (excluding development of materials) Description of cost	Yr 1 Total	Yr 2 Total	Total
a. Personnel	1 494 064	2 615 698	4 109 762
Executive oversight and governance	220 000	195 173	415 173
Finance & Operations Management	147 517	157 843	305 361
Accounting	103 377	110 613	213 990
Content & Training Management	406 065	434 490	840 555
Logistics & Materials Co-ordination	260 000	262 899	522 899
Training and District Support	120 000	590 000	710 000
M & E	143 000	337 000	480 000
Communications (including design & translation)	146 000	157 680	303 680
b. Travel and accommodation	29 000	137 000	166 000
c. Equipment	18 400	20 770	39 170
d. Supplies	1 422 817	55 435	1 478 252
Printing	1 250 000	637 000	1 887 000
Stationery	16 000	27 000	43 000
Postage/Courier of materials	6 155	24 000	30 155
f. Other	333 959	373 136	707 094
Project meetings, including out of province	60 000	26 000	86 000
Administration and overheads	255 000	355 000	610 000
TOTAL	3 359 069	3 157 716	6 516 785

WCED COST TEMPLATE (NOT COMPLETED)

Category	Number of staff involved in the project	% of full time employment	Number of months dedicated to the project	Approximate/ average monthly salary	Brief description of the contribution
Government staff time					

Category	Number of staff involved in the project	Number of person hours (total)	Approximate/ average monthly salary	Brief description of the contribution
Government staff time in training (provincial staff)				
Government staff time in training (district staff)				
Government staff time in training (teachers)				

Category	Value in local currency		Brief description of the contribution
Venue			
Materials/ equipment/ supplies			
Transportation			
Direct monetary contributions			
Other (please specify here)			

ANNEX 4: SCHEDULES AND PROJECT ACTIVITIES

Project Activities	Due Date
Work Plan Finalisation	
Review STELLAR documents	05 June 2015
Hold inception meeting with USAID	12 June 2015
Hold inception meeting with Wordworks	26 June 2015
Develop and submit work plan	03 July 2015
Methodology Plan and Study Protocol	
Select sample	10 July 2015
Locate fieldworkers	10 July 2015
Design assessment tools	17 July 2015
Paper-based learner assessment	17 July 2015
Verbal learner assessment	17 July 2015
Classroom observation assessment	17 July 2015
Teacher assessment tool	17 July 2015
Trainer assessment tool	17 July 2015
Teacher KII guide	17 July 2015
District official KII guide	17 July 2015
Provincial official KII guide	17 July 2015
National official KII guide	17 July 2015
Wordworks KII guide	17 July 2015
Donor KII guide	17 July 2015
Sustainability Assessment	17 July 2015
Design database and develop and test ODK forms	17 July 2015
Develop and submit methodology plan and study protocol	17 July 2015
Client, Wordworks, and WCED review of tools and methodology plan/study protocol	31 July 2015
Pilot assessment tools	31 July 2015
Refine assessment tools	31 July 2015
Translation of tools into isiXhosa and Afrikaans	07 August 2015
Provincial access	31 July 2015
Interim Evaluation Report 1	
Collect and analyse Lead Teacher training data for districts that attend block training from 13 - 17 July	24 July 2015
Secure school visits	31 July 2015

Project Activities	Due Date
Prepare for fieldwork, print fieldworker materials for training, and gain consent from parents/guardians of learners	31 July 2015
Train fieldworkers	07 August 2015
Collect baseline control learner & educator data	21 August 2015
Clean and screen data	28 August 2015
Analyse data	11 September 2015
Develop and submit draft Interim Evaluation Report 1	11 September 2015
Debriefing with USAID and other stakeholders	18 September 2015
Client review	25 September 2015
Develop and submit final Interim Evaluation Report 1	16 October 2015
Conduct Donor KII	25 September 2015
Interim Evaluation Report 2	
Secure school visits	30 October 2015
Prepare for fieldwork	30 October 2015
Train fieldworkers	13 November 2015
Collect end of year (mid-term) control learner and educator data	27 November 2015
Clean and screen data	4 December 2015
Analyse data	18 December 2015
Develop and submit draft Interim Evaluation Report 2	18 December 2015
2016	
Interim Evaluation Report 2	
Client review of Draft Interim Evaluation Report 2	15 January 2016
Debriefing with Client on Draft Interim Evaluation Report 2	29 January 2016
Develop and submit Final Interim Evaluation Report 2	05 February 2016
Debriefing with stakeholders	12 February 2016
Via WCED Lead Teachers, remotely administer training pre- and post- tests to regular teachers at the first cluster meeting	12 February 2016
Submission of data (Round 2)	19 February 2016
Resubmit Final Interim Evaluation Report 2 incorporating stakeholder comments	26 February 2016
Administer Sustainability Assessment (KIIs)	11 March 2016
Introduce Cost Effectiveness Assessment	11 March 2016

Project Activities	Due Date
Secure letters to schools for Baseline Intervention group (from WCED)	06 May 2016
Interim Evaluation Report 3	
Secure school visits	29 July 2016
Prepare for fieldwork	29 July 2016
Train fieldworkers	05 August 2016
Collect baseline treatment learner & educator data	02 September 2016
Clean and screen data	16 September 2016
Analyse data	16 September 2016
Develop and submit Draft Interim Evaluation Report 3	30 September 2016
Client review	07 October 2016
Submit datasets (Round 3)	14 October 2016
Debriefing with USAID	14 October 2016
Develop and submit Interim Evaluation Report 3 (second draft)	28 October 2016
Debriefing with stakeholders	04 November 2016
Resubmit Final Interim Evaluation Report 3 incorporating stakeholder comments	04 November 2016
Analyse data	02 December 2016
Develop and submit draft Interim Evaluation Report 4	09 December 2016
2017	
Interim Evaluation Report 4	
Submission of annual work plan	16 January 2017
Cost analysis with WCED, continued	30 January 2017
Cost analysis with Wordworks/USAID	30 January 2017
Develop and submit Draft Interim Evaluation Report 4	30 January 2017
Client review	06 February 2017
Debriefing with USAID and Elma Foundation (10th February)	06 February 2017
Submit Revised Interim Evaluation Report 4 to USAID	13 February 2017
Debriefing with other stakeholders (26th April)	24 April 2017
Submit Final Interim Evaluation Report 4 incorporating stakeholder comments	15 May 2017
Submission of datasets (Round 4)	20 February 2017

Project Activities	Due Date
Interim Evaluation Report 5	
Secure letters to schools for Endterm Intervention group (from WCED)	28 August 2017
Secure letter from WCED to subject advisors asking for their participation in the online sustainability assessment	14 August 2017
Repeat online sustainability assessment with subject advisors	25 September 2017
Develop sustainability assessment Key Informant Interview guides	11 September 2017
Team planning meeting	04 September 2017
Conduct sustainability assessment Key Informant Interviews with USAID, WCED and Wordworks	25 September 2017
Secure school visits	25 September 2017
Prepare for fieldwork	25 September 2017
Train fieldworkers	09 October 2017
Collect endterm treatment learner & educator data	20 November 2017
Clean and screen data	11 December 2017
2018	
Final Evaluation Report	
Completed processed data from Janeli by 8 January	08 January 2018
Finalised graphics for the report from Nokuthula by 17 January	15 January 2018
Submission of Final Evaluation Report Draft by 19 January	15 January 2018
Client review and feedback by 8 February	05 February 2018
Debriefing with USAID and Elma Foundation 9 February	05 February 2018
Revision of Final Evaluation Report Draft by 14 February	12 February 2018
Submission of Summary Report Draft by 14 February	12 February 2018
Stakeholder review and feedback by 22 February	19 February 2018
Debriefing with Stakeholders 23 February	19 February 2018
Finalisation of report and submission of Final Evaluation Report and Cleaned Datasets	29 June 2018

ANNEX 5: INTERVIEW LISTS AND MEETINGS

Sustainability Key Informant Interviews

1. USAID representative
2. ELMA Philanthropies representative
3. Wordworks representatives
4. Western Cape Education Department representative

Key Meetings

1. Kick-off Meeting USAID_10 June 2015
2. Meeting USAID and WW_22 June 2015
3. Evaluation meeting_30 June 2015
4. Meeting USAID WCED and WW_4 August 2015
5. Project Internal Team Meeting_21 Sept 2015
6. Debriefing with USAID and other stakeholders: Draft Interim Evaluation Report 1_October 2015
7. Meeting between USAID, Khulisa, and the Elma Foundation: Debriefing Interim Evaluation Report 1_21 October 2015
8. Fieldwork Preparation Meeting_03 November 2015
9. Teleconference with USAID_23 November 2015
10. Stakeholder meeting with USAID, WW and WCED_2 December 2015
11. USAID Midterm Debriefing: Interim Evaluation Report 2_29 January 2016
12. Evaluation Meeting with USAID_15 June 2016
13. Debriefing with USAID and other stakeholders: Draft Interim Evaluation Report 3 _1 November 2016
14. Stakeholder debriefing_09 November 2016
15. Evaluation Internal Team Meeting_19 January 2017
16. Debriefing with USAID: Draft Interim Report 4 _10 February 2017
17. Teleconference with USAID_27 March 2017
18. Debriefing with stakeholders: Draft Interim Report 4 _24 May 2017
19. Project planning meeting_4 September 2017
20. Evaluation Project Management meeting_30 November 2017
21. Evaluation Project Management meeting_18 January 2018
22. Debriefing with USAID and other stakeholders: Draft Interim Evaluation Report 4_February 2018
23. Final Evaluation Report Debriefing meeting_7 February 2018
24. Final debriefing follow-up meeting_13 February 2018
25. Final Report Debriefing with Stakeholders_21 February 2018

ANNEX 6: SOURCES OF INFORMATION

Key Informants

1. USAID representative
2. ELMA Philanthropies representative
3. Wordworks representatives
4. Western Cape Education Department representative
5. Sample Grade R and Grade 1 Teachers
6. ORT SA Cape representative

Research Tools

1. Learner Assessment
2. Arrival Instrument
3. Classroom Environment Checklist Tool
4. Teacher Knowledge, Attitudes and Practices Survey Tool
5. Pacing and Curriculum Coverage Tool
6. ELIT Programme Sustainability Online Survey
7. ELIT Programme Sustainability Key Informant Interview Guide

ANNEX 7: STATEMENT OF WORK

Evaluation Scope of Work

SCHOOL CAPACITY & INNOVATION PROGRAM (SCIP)

Strengthening Teaching of Early Language and Literacy: STELLAR Program

PROJECT TO BE EVALUATED

Project name: *Strengthening Teaching of Early Language and Literacy in Grade R: STELLAR Program*

Cooperative Agreement No.: APS 674-12-00001

Project Dates: January 2014 through December 2016

Evaluation Dates: March 2015 through February 2018

Implementing Organization: Wordworks

1. OBJECTIVE

USAID/Southern Africa's School Capacity & Innovation Program (SCIP) seeks to improve literacy in the early grades. The aim of this solicitation is to conduct a longitudinal study on the SCIP award "STELLAR" to evaluate the effectiveness of Grade R literacy interventions and follow the impact of these interventions through Grade 1. Specifically, USAID seeks to understand if the scale up of the STELLAR model to all teachers and learners in Grade R in the Western Cape Province, through the Western Cape Emergent Literacy Project, is successful in improving the language and literacy skills of Grade R and 1 students, and if positive impacts for beneficiaries (students, teachers, and trainers) are evident upon delivery of the intervention as well as over time. If this is the case, this will be a welcome addition to the evidence base for policy decisions moving forward about the most effective way to improve early language and literacy in South Africa.

This impact evaluation will primarily focus on 1) collecting baseline data on language and literacy skills on a comparison group of current 2015 Grade R learners at the end of the school year and endline data in 2016, at the end of Grade 1; 2) to collect baseline, interim and endline data from lead teachers who will serve as trainers and teachers; and 3) baseline and endline data on language and literacy skills a treatment cohort of 2016 Grade R learners and a cohort of 2017 Grade 1 learners and assessing their language and literacy skills at the end of the 2017 school year. The exact timing for the follow up will be determined in consultation with USAID and the implementing partner. Data from students will be collected at intervals sufficient to evaluate language and literacy achievement over time; for example, at the start of Grade R, end of Grade R, and end of Grade 1. Teachers of Grade R should also be observed during the year to evaluate fidelity to the STELLAR model. The evaluation is meant to be an evaluation of program impact, utilizing robust methodology and appropriate mix of approaches to assess the impact of the STELLAR intervention model on teacher and trainer effectiveness and on student achievement from baseline through Grade 1 of primary school. The results should also evaluate the effectiveness of the training-of-trainer methodology from both a cost and fidelity perspective, with

particular emphasis on STELLAR methods to address teacher need for ideas on how to build early literacy and language.

2. PURPOSE AND USE

The evaluation's purpose is to assess: (1) the effects of the STELLAR intervention on Grade R and Grade 1 achievement in language and literacy; and (2) the effectiveness of the STELLAR intervention on Grade

R teacher knowledge and teaching practice; and (3) the effectiveness of the STELLAR training-of-trainers model in replicating impacts for (1) and (2). Baseline data collected as part of the evaluation may be used as formative data for the STELLAR program so Wordworks can better tailor their interventions to beneficiary needs during subsequent implementation. The evaluation will also provide recommendations for taking successful literacy interventions to scale in South Africa. The primary audience of the evaluation is USAID/Southern Africa, the ELMA Foundation, JP Morgan, Wordworks and the Western Cape Education Department (WCED). A secondary audience of the evaluation is the Government of South Africa's (GoSA) national Department of Basic Education. The approved final report will be released for the use of other stakeholders, such as other USG agencies, bilateral and multilateral donors, other NGOs, etc., and is to be published on the Development Experience Clearinghouse (DEC).

3. BACKGROUND ON PROGRAMMING

USAID/Southern Africa's Education Office has two main priorities:

Priority 1: To support the GoSA Department of Basic Education's (DBE) *Action Plan to 2014: Towards the Realization of Schooling 2025* by working to improve the quality of teaching and learning with a special focus on primary grade reading.

Priority 2: To advance Goal 1 of USAID's *Education Strategy 2011 -2015* to improve the reading skills for 100 million children in primary grades, worldwide, by 2015.

The USAID/SA basic education program focuses on interventions to improve the reading outcomes of primary grade learners. USAID's basic education investments contribute to GoSA's strategic framework for education and support both education quality and sector reform initiatives led by the GoSA. USAID/SA basic education programming is also aligned to Goal 1 of the Education Strategy, working to improve the literacy skills of SA children in the primary grades through support to innovative and targeted reading interventions in classrooms and through improving teacher effectiveness in the primary grades. See Annex C for the USAID/SA Country Development Cooperation Strategy Results Framework.

School Capacity & Innovation Program (SCIP)

In response to the priorities of the South African Department of Basic Education (DBE), USAID, in partnership with the ELMA Foundation and JP Morgan, developed the School Capacity & Innovation Program (SCIP). SCIP is designed to improve learning outcomes as measured by improvements in primary grade reading. Interventions include building teacher effectiveness through in-service training and strengthening classroom and school management through targeted capacity development. The first three local organizations selected for award, the Human Sciences Research Council, Mindset Teach, and siyaJabula siyaKhula, completed the first of three years of planned implementation in FY 13. This first year culminated in a Learning Forum

where key education stakeholders and partners gathered to present results, exchange ideas and share resources to support continued literacy achievement in the early grades. All awards under SCIP introduce innovative primary grade literacy focused interventions with sustainability plans for scaling up these interventions in partnership with the DBE.

The SCIP mechanism is well-placed to first recognize innovative ideas for educational interventions and to then provide the necessary resources to take those successful, innovative interventions to scale. Apart from the three awards mentioned above, the Wordworks STELLAR intervention is one of two recently made awards. SCIP's overarching goal is to stimulate improvements in primary grade reading outcomes across South Africa.

Strengthening Teaching of Early Language and Literacy in Grade R: STELLAR Program

The *Strengthening Teaching of Early Language and Literacy: STELLAR Program* has been implemented by Wordworks in two education districts in the Western Cape via direct, six-month long training delivery for in-service Grade R teachers. With USAID support, STELLAR will in 2015 and 2016 expand both geographically and in scale via a "training-of-trainers" or "cascade" approach to program delivery. For the first year of implementation (2014), delivery was via direct training where Wordworks personnel work with groups of Grade R teachers through monthly workshops over a period of six months. For implementation years two and three (2015 and 2016) a training-of-trainer methodology will be employed in partnership with the WCED, and the STELLAR model will be taken to scale in 2015 in the whole province as part of the Western Cape Emergent Literacy Project.

The intended goal of STELLAR is to positively impact the language and literacy achievement of early grade learners and to provide essential language and literacy skills to children from disadvantaged communities in South Africa. Illustrative activities are included in Annex A: Overview of the STELLAR Program Content.

In order for STELLAR to achieve its overarching goals of significantly improving the language and literacy achievement of Grade R students and successfully bringing the STELLAR implementation model to scale, the theory of change (Annex B) points to the necessary accomplishment of the following results:

Result 1: Develop and distribute innovative, high quality, user-friendly teaching and learning materials for the teaching of language and literacy to Grade R learners.

Select indicators:

- Number of textbooks and other teaching and learning materials (TLM) provided with United States Government (USG) assistance
- Custom indicator(s) or tool, TBD, to demonstrate effectiveness and quality of TLM

Result 2: Through training over time, provide Grade R teachers and trainers with the knowledge, skills and resources to teach language and early literacy effectively.

Select indicators:

- *Total number of trainers who successfully completed training*
- Number of teachers/educators/teaching assistants who successfully completed in-service training or received intensive coaching or mentoring with USG support Proportion of teachers who (a) demonstrate improved knowledge and understanding of early language

and literacy teaching and learning, and (b) reflect their new knowledge in improved classroom practices

- Proportion of trainers who demonstrate improved knowledge and understanding of early language and literacy teaching and learning
- Proportion of learners whose language and literacy scores have improved significantly by the end of Grade R (above the expected improvement for this time period)
- The average improvement in learners' language and literacy scores by the end of Grade R and Grade 1
- Number of learners receiving reading interventions at the primary level

Result 3: Identify key partners and 'train the trainer' in order that innovations reach as many schools and learners as possible.

Select indicators:

- Number of textbooks and other teaching and learning materials (TLM) provided with USG assistance
- Number of Memoranda Of Understanding established between Wordworks and Provincial DoEs, Departments of Social Development (DSDs), including district offices, Registered Training Organizations (RTOs), Early Childhood Development (ECD) providers, or Further Education and Training institutions (FETs)
- Number of administrators and officials successfully trained with USG support

Result 4: Collect, analyse and disseminate data that demonstrates the impact of these interventions in order to build an evidence base and ensure that programs are responsive to emerging learning.

Select Indicators:

- Custom indicator(s) or tool, TBD, to measure or reflect the number of and/or quality of forums held to disseminate data, press events or news articles published, academic papers published, guidance issued, etc.
- Custom indicator(s) such as: Percent change in perception of teachers participating in the program who report feeling mentored/successful/supported/integrated into a larger teaching community
- Custom indicator(s) such as: Percent change in responsiveness of programs to emergent learning, as measured by developed index tool/observation method/other

To achieve these results, STELLAR focuses on strengthening pedagogy and equipping teaching professionals with essential teaching and learning resources. Program activities will support the South African government in its efforts to improve the quality of Grade R and in achieving its ambitious targets to raise early grade literacy scores.

4. SITES AND REACH

STELLAR's initial training sites in 2014 were in 68 schools in the Metro South and Metro Central Education Districts in the Western Cape. In 2015 and 2016 STELLAR will be expanding to all other districts in the Western Cape as part of the Western Cape Grade R Emergent Literacy Project. The implementation plan is detailed below:

Block training-of-trainers for approximately 70 District Officials by Wordworks trainers for a minimum of 20 hours: 20-23 April 2015.

Block training-of-trainers for approximately 150 Lead Teachers by Wordworks trainers for a minimum of 20 hours: 13-17 July 2015. Previously trained District Officials support Wordworks trainers at these sessions.

Lead Teachers implement the STELLAR method in their own classrooms: 3rd term of 2015. Lead Teachers receive about four mentoring and support sessions facilitated by Wordworks trainers: 3rd term of 2015

STELLAR training for all 3000 teachers in the Western Cape presented by the Lead Teachers with support from Wordworks trainers and District Officials: October school holidays 2015.

First cluster meeting between Lead Teachers and their respective groups of Grade R teachers, for planning purposes for 2016: 4th term of 2015. Telephonic and email support provided to Lead Teachers by Wordworks trainers and District Officials.

All Grade R teachers implement the STELLAR approach in their classrooms: Jan – Dec 2016.

8 Monthly cluster meetings are held for grade R teachers in the course of the year, facilitated by the Lead Teachers: 2016. Telephonic and email support provided to Lead Teachers by Wordworks trainers and District Officials.

5. DATA POTENTIALLY TO BE COLLECTED VIA THIS EVALUATION

As the intervention goes to scale in the province, all teachers and learners will receive the intervention and there will no longer be a comparison group available for comparison. However, the majority of Grade R's of 2015 will not receive the intervention and therefore they can be used as a comparison group. Other groups that will be evaluated include the District Officials and Lead Teachers, both of whom will receive the training-of-trainers sessions, the teachers who will receive the STELLAR training, and the learners who will be taught according to the STELLAR model and materials.

We expect to see at least the following data resulting from the evaluation:

1. Comparison group data on language and literacy skills from a representative sample of the current 2015 Grade R learners at the end of the school year and from a representative sample of learners at the end of Grade 1 in 2016. The sample should exclude learners taught by teachers who have been previously trained by Wordworks.
2. Comparison group data on pacing and curriculum coverage from a small sample of classes at the end of Grade 1 in 2016. This could include a workbook analysis and/or other methodologies.
3. Baseline data on subject knowledge and pedagogical content knowledge from a sample of Lead Teachers before they receive training.
4. Baseline data on subject knowledge and pedagogical content knowledge, as well as teaching practices, from a sample of Grade R teachers before they receive training
5. Baseline data on language and literacy skills from a representative sample of 2016 Grade R learners at the beginning of 2016.
6. Interim data on subject knowledge and pedagogical content knowledge from Lead Teachers after they have received training and again after they have delivered training.
7. Interim data on subject knowledge and pedagogical content knowledge from teachers after they have received training and again at the end of the year of school years in 2016

and 2017, as well as data collected at intervals throughout the year on their teaching practice.

8. Data on pacing and curriculum coverage at the end of Grade 1 in 2017.
9. Data on the degree of fidelity to the content of the STELLAR materials as the train-the-trainer model is rolled out.
10. Data on the effectiveness of the cluster meetings where teachers receive refresher training, mentoring, and support.
11. Data on the cost effectiveness of the training of trainer approach.

USAID seeks data on learner language and literacy outcomes with a confidence level of at least 95% and a confidence interval of $\pm 5\%$. Data on pacing and curriculum, teachers and lead trainers does not need a confidence level as high as the data on learners. Sample sizes for these groups should balance the expense of gathering the data with the value of the data for answering the evaluation questions.

6. EVALUATION QUESTIONS

The evaluation's purpose is to assess: (1) the effects of the STELLAR intervention on Grade R and Grade 1 achievement in language and literacy; and (2) the effectiveness of the STELLAR intervention on Grade R teacher knowledge and teaching practice; and (3) the effectiveness of the STELLAR training-of-trainers model in replicating impacts for (1) and (2).

The evaluation questions to be answered are:

1. Have STELLAR interventions increased the language and literacy achievement of targeted learners over time? (50% LOE) (*impact* evaluation design desired)

Key Sub-question:

- To what extent have Grade R interventions impacted the language and literacy acquisition and performance of Grade R and Grade 1 learners in comparison to the comparison group and expected outcomes for these grades?
2. To what extent has the training-of-trainers model approach resulted in (1) more effective and knowledgeable teachers and (2) positive language and literacy outcomes for learners? (40% LOE)

Key Sub-questions:

- To what extent has the integrity of the STELLAR program been maintained through scale-up?
 - What factors enhance and/or constrain the scale-up of the STELLAR training approach?
3. What is the likelihood of sustainability of organizational and beneficiary gains after STELLAR interventions end? (10% LOE) (*performance* evaluation design acceptable)

7. EVALUATION DESIGN AND METHODOLOGY

USAID is interested in rigorous evaluations of both the process and impact of the STELLAR program and is open to offeror suggestions on methods to evaluate. The proposed methodology must include both quantitative and qualitative methods.

For each of the evaluation questions listed in section 6, offerors should propose illustrative indicators, data sources and collection methods, sampling and selection criteria and data analysis methods.

Additionally, the offeror must discuss data disaggregation by sex, and should by other categories in the evaluation, as well as take any disaggregates and localized considerations that relate specifically to the project site context into account to the fullest extent possible.

The evaluators will have access to routine program data collected by Wordworks and its partners, and will be expected to work in collaboration with Wordworks on the design and utilization of data collection tools and procedures so as to minimize disruption of classroom and instructional time for teachers and learners. The selected evaluator will be expected to collect primary data in addition to that collected by Wordworks and their partners wherever possible in order to get the most objective evaluation possible.

Upon award, but before fieldwork is conducted, the evaluator will submit a detailed evaluation design, methodology and implementation plan for review and approval by USAID.

Before data collection, the contractor, in coordination with USAID and Wordworks, will finalize the data analysis methods as part of the methodology plan.

As a lesson learned from previous evaluations, the evaluation needs to be carried out in a participatory fashion, forming a team that, in various places and times, includes a range of USAID staff, GoSA staff, Wordworks staff, and other stakeholders. This stakeholder engagement should be reflected in the evaluation plan.

Potential Limitations to Evaluation Methods

The STELLAR program is already being implemented Offerors will address potential limitations to their proposed evaluation methodology and suggest ways to mitigate any identified weaknesses or potential weaknesses in the proposed approach.

8. TEAM COMPOSITION

The contractor will propose a core evaluation team, with a preference for local South African hires. The team proposed must have the appropriate combination of knowledge, skills, experience and abilities necessary to successfully carry out the evaluation.

The evaluation team should have:

- a. Extensive evaluation experience in South Africa.
- b. Experience undertaking impact evaluations using experimental and quasi-experimental methods.
- c. Experience in the social sciences, particularly in early grade language and literacy education and assessments and evaluations involving children.
- d. A comprehensive understanding of training and professional development, including knowledge of the public sector institutions and NGOs working in education.
- e. Experience working with donor-funded development programs.
- f. Experience in qualitative and quantitative data collection methods

9. DELIVERABLES AND TIMETABLE

The offer should outline their timeframe for each deliverable, with particular care to maintaining cohesiveness across the full number of years of data collection that relates to the longitudinal component of the evaluation.

- a. **Team planning meeting(s)**
At least one team planning meeting will be conducted in advance of any data collection activities. This meeting will allow USAID to present the evaluation team with the purpose, expectations, and agenda of the assignment as well as provide any feedback on the evaluation design and methodology submitted as part of the proposal.
- b. **Work plan**
Before data collection is authorized, the team will prepare a detailed work plan that will include the methodologies to be used in the evaluations. The work plan will be submitted to the Contracting Officer Representative (COR) at USAID for approval after the project is awarded.
- c. **Methodology plan & study protocol developed in coordination with Wordworks**
Written methodology plans (e.g. evaluation design, key questions, methods, tools, data collection instruments, and data analysis plan, operational work plan) will be prepared by the evaluation team and approved by USAID prior to international travel. The written design of the evaluations will be shared with stakeholders before being finalized. After USAID approval, the study protocol ***must be submitted to Wordworks and consensus must be reached on the approach.***
- d. **Interim briefings, including status reports**
The team leader will provide weekly status reports to USAID on work plan implementation. The evaluation team will also conduct at least one interim briefing per quarter with USAID to review the progress and methodology of the evaluation.
- e. **Debriefing with USAID**
The team will present the major findings of the evaluations to USAID after submission of the draft reports.
- f. **Debriefings with other stakeholders**
The team will independently present the major findings of the evaluations to the USAID partner (as appropriate and as defined by USAID) and to the DBE through PowerPoint presentations. The debriefing will include a discussion of achievements, activities, and recommendations. The evaluation team will consider partner comments and revise the draft reports accordingly, as appropriate.
- g. **Interim data reports**
Draft reports of the findings and recommendations after each round of data collection should be submitted to the USAID COR. The written reports should describe the baseline data collected, interim data collected, and any findings, conclusions, or recommendations. The reports will be shared with the project partner, Wordworks. There will be a minimum of three interim data reports, submitted annually at the end of the fourth quarter of the USAID fiscal year, on September 30th.
- h. **Draft evaluation report**
Draft reports of the findings and recommendations should be submitted to the USAID Task Order Contracting Officers Representative (TO COR). One hard copy and one electronic copy, in Microsoft Word, will be provided to USAID.
- i. **Final evaluation report**
The USAID has 14 working days to review the draft report and provide written comments to the contractor on the draft report. The contractor will have 14 working days to incorporate these comments into a final report. Contractor must provide both an electronic version, in Microsoft Word, and 5 hard copies of the final report to USAID. The contractor will be responsible for releasing the approved report as a public document on the USAID Development Experience Clearinghouse (DEC) (<http://dec.usaid.gov>) after the TO COR provides formal written approval.
- j. **Data sets**
Provide data, technical materials, and other information produced in the execution of USAID funded activities

The Awardee will provide USAID with data, technical materials, and other relevant materials produced in the execution of this Award. This includes pedagogical materials and other technical inputs developed to support early grade reading outcomes and other Award objectives, as well as data and information needed for reporting under the relevant foreign assistance objectives, areas and elements.

Pedagogical Materials and Technical Inputs

The Awardee is required to provide pedagogical materials and other technical inputs developed to support early grade reading outcomes and other Award objectives. Examples of technical inputs to be provided to USAID include scripted lesson plans, supplementary readers, assessment instruments, observation tools, training guides, workshop reports, radio programs, assessment tools, sampling frames, photographs, videos, and other recordings. The Awardee will transmit technical materials to the relevant TOCOR (if applicable) and/or AOR/COR, and submit them to the USAID Development Experience Clearinghouse (<https://dec.usaid.gov/>).

Data for Reporting Under Foreign Assistance Objectives

The Awardee is required to provide datasets and codebooks that include data on student learning outcomes and information needed to estimate the number of unique pupils benefiting from program activities over the life of the program. Implementing partners may be responsible, in collaboration with USAID, for obtaining country level memoranda of understanding that allow for the sharing of the datasets and other data with USAID, as well as public access to the data through the partner organization, where possible.

Within 90 days of the completion of data collection, the Awardee will transmit requested data to USAID. The transmittal shall be according to the following specifications:

- Datasets should be complete, clean, and final, and include any derived or secondary variables used to calculate indicator values provided in assessment reports.
- Datasets must be cleansed of Personally Identifiable Information (PII) prior to transmittal to USAID. PII includes any information that could be used to identify an individual student, teacher, or administrator for whom data have been collected.
- Datasets will include all variables included in the initial data collection, with the exception of any data that must be edited or cleaned to protect the privacy and anonymity of students, teachers, or administrators represented in the data.
- If variables are edited or removed in order to protect the privacy and anonymity of research subjects, steps should be taken to ensure that sufficient information is retained to allow analyses that require grouping students by school, or track schools/students across datasets if appropriate.
- Data must be transmitted along with relevant supporting materials and instruments. This includes questionnaires and other instruments, codebook, data dictionary, information on sample design, setup and weights, assessment reports, performance management plans or other materials that describe the structure of the assessment and/or program, and any other information a researcher may need when working with the data.
- Learning Assessment data can be transmitted in formats including Stata, SPSS, SAS, R, or an open and machine readable format. Supporting documents can be transmitted in MS Office or an open and machine readable format.
- Awardee will provide information on the number of pupils benefiting from the program, disaggregated by sex and grade for each year that the program is active.

- Datasets will be delivered through email, addressed to the relevant COR/AOR. The Awardee may also be directed by USAID to submit data and related documents to a third party site (e.g. <https://sartdatacollection.org>).
- All prerequisites to providing the complete, cleaned datasets must be completed by the implementing partner prior to the provision of the dataset to USAID, such as review and approval by Missions and host country governments, as appropriate."

10. REPORTING FORMAT AND CRITERIA

Five separate reports will be produced for USAID: 1) Three interim reports incorporating baseline and interim data collected; 2) an end-of-evaluation report, once endline data has been collected, with impact-evaluation findings; and 3) a brief (<10 pages) summary report on the evaluation findings.

10.1 Format

The report format should be restricted to Microsoft Office products and 12-point type font should be used throughout the body of the report, with page margins 1 inch top/bottom and left/right. The end- of-evaluation report should not exceed 35 pages, excluding references and annexes.

Interim Reports

The interim report should include:

1. Baseline assessment data and/or interim assessment data on student literacy levels and teacher and trainer content knowledge, including knowledge of the Wordworks package of training materials.
2. Baseline and interim survey data on treatment cohorts and counterfactual/comparison groups.
3. Qualitative and/or quantitative interview data with program staff, school officials, and other stakeholders on the quality of project implementation, what things are working, what needs to be changed, etc.
4. Summary of relevant findings and conclusions drawn from baseline and interim data collected as it relates to the state of the learning environment and all beneficiaries before, during and in the absence of STELLAR program implementation.

The first interim report should include: (1) baseline data collected and analyzed; (2) any needed revisions in the key questions to be addressed in the impact evaluation phase.

The baseline phase will review the development hypothesis, the identified project goals and results, and the key indicators to determine whether the development hypotheses relates to the achievement of expected results as articulated in the program description. If not, why not?

Final Evaluation Report

The final evaluation report produced by the successful offeror shall: (1) Use rigorous evidence to answer the evaluation questions presented in this task order; (2) Identify best practices and lessons learned; (3) Make recommendations for any future literacy interventions and in particular, recommendations on scaling up literacy interventions.

The evaluation report should include the following sections:

Initial baseline and interim data reports

1. Executive Summary: summarizes project purpose and background, key evaluation questions, methods, findings, and recommendations. (1-2 pgs.);
2. Table of Contents (1 pg.);
3. Introduction and Background: purpose, audience, and synopsis of task, brief overview of STELLAR program, including program strategy and activities implemented in response to the problem, brief description of Wordworks and any relevant STELLAR partners (1-2 pgs);
4. Evaluation Design and Data Collection Methodology: describes evaluation methods, including constraints and gaps (1-2 pg.);
5. Findings/Conclusions/Recommendations: for each objective area; and also include data quality and reporting system that should present verification of spot checks, issues, and outcome (8-10 pgs.);
6. Issues: provide a list of key technical and/or administrative, if any (1–2 pgs.);
7. Future Directions: to inform any necessary changes to the design of the final evaluation questions (1-2 pgs. if needed);
8. References (including bibliographical documentation, meetings, interviews and focus group discussions);
9. Annexes: annexes that document the evaluation tools, schedules, interview lists, tables, all sources of information, the evaluation statement of work, statements of differences—should be succinct, pertinent and readable.

Final Evaluation Report

1. Executive Summary: summarizes project purpose and background, key evaluation questions, methods, findings, and recommendations. (3-5 pgs.);
2. Table of Contents (1 pg.);
3. Introduction and Background: purpose, audience, and synopsis of task, brief overview of STELLAR program, USAID program strategy and activities implemented in response to the problem, brief description of implementing partner (1-2 pages);
4. Evaluation Design and Data Collection Methodology: describes evaluation methods, including constraints and gaps (1-2 pg.);
5. Findings/Conclusions/Recommendations: for each evaluation question answered by objective area; and also include data quality and reporting system that should present verification of spot checks, issues, and outcome (17–20 pgs.);
6. Issues: provide a list of key technical and/or administrative, if any (1–2 pgs.);
7. Future Directions - to inform the design of any further projects (2-3 pgs.);
8. References (including bibliographical documentation, meetings, interviews and focus group discussions);
9. Annexes: annexes that document the evaluation tools, schedules, interview lists, tables, all sources of information, the evaluation statement of work, statements of differences—should be succinct, pertinent and readable.

Summary Report

Format is flexible, but should briefly summarize the key findings, in particular the impact, or lack of impact, identified.

This is meant to be a policy brief; very briefly summarizing the evaluation questions, the methods used, the key findings, and the programmatic implications.

10.2 Final Report Criteria

The evaluation report should represent a thoughtful, well-researched, and well organized effort to objectively evaluate what worked in the project, what did not, and why.

Evaluation reports shall address all evaluation questions included in this Scope of Work.

The evaluation report should include the scope of work (as written in the signed contract) as an annex. All modifications to the scope of work, whether in technical requirements, evaluation questions, evaluation team composition, methodology, or timeline need to be agreed upon in writing by the Contracting Officer's Representatives.

Evaluation methodology shall be explained in detail and all tools used in conducting the evaluation such as questionnaires, checklists, and discussion guides will be included in an Annex in the final report.

Evaluation findings will assess outcomes and impacts on males and females.

Limitations to the evaluation shall be disclosed in the report, with particular attention to the limitations associated with the evaluation methodology (selection bias, recall bias, unobservable differences between comparator groups, etc.).

Evaluation findings should be presented as analyzed facts, evidence, and data and not based on anecdotes, hearsay, or the compilation of people's opinions. Findings should be specific, concise, and supported by strong quantitative or qualitative evidence.

Sources of information need to be properly identified and listed in an annex.

Recommendations should be action-oriented, practical, and specific, with defined responsibility for the action.

11. RELATIONSHIPS, SCHEDULING AND LOGISTICS

The Contracting Officer will formally appoint a Contracting Officer's Representative (COR) and an Alternate Contracting Officer's Representative in a formal letter to the contractor. The Contracting Officer and the COR are the only official representatives of USAID for this contract and are the only ones authorized to provide technical direction to the contractor throughout the evaluation. The contractor is expected to work together with the COR to implement the scope of work.

USAID will provide overall direction to the evaluation team, identify key documents, and assist in facilitating a work plan. The evaluation team will be responsible for arranging meetings with key stakeholders and will be required to advise USAID prior to each of those meetings. The evaluation team is also responsible for arranging transportation in and around South Africa. The evaluation team will be responsible for procuring its own office space, computers, Internet access, printing, and photocopying. Evaluation team members will be required to make their own payments. USAID personnel will be made available to the team for consultations regarding sources and technical issues, before and during the evaluation process.

USAID will send letters and emails of introduction informing key DBE staff and other high-level partners of the nature, timing, and scope of the evaluation and of the evaluation team members.

Annex A: Overview of the STELLAR Program Content

	Activities	Research evidence
Telling stories using picture sequences and cut-outs	Teachers use the cut-outs and picture sequences to tell stories to a class. Children role play and retell stories and engage in small group follow-up activities. Songs and rhymes for each story help to reinforce new language. Goals: Children become familiar with story structure – beginning, middle, end Concept development e.g. colors, size, concept of place (on, under, in, behind, through, next to, over) New vocabulary is learnt in a meaningful context Children are prepared for ‘book language’ Small group activities help to consolidate new vocabulary and build language skills	Children’s experiences with oral language in their early years provide critical foundations for later reading success (Burns, Griffin, & Snow, 1999). Research has shown a relationship between listening to and interacting with stories, and subsequent literacy competence. Wells (1987) proposed that hearing stories is the most crucial of all the interactions that contribute towards children becoming literate. General language development and the ability to understand more complex grammar and story structures are positively affected by consistent story telling in the early years. Learning the meanings of words and developing an understanding of concepts means that children will be able to understand what they read. Having opportunities to use new words and phrases and explain concepts, gives children the expressive language skills necessary for expressing their ideas in writing.
Drawing and emergent writing	Teachers use the story-telling experience as a platform for drawing and writing. Children draw part of the story, and ‘write’ about the story through scribbles or letter-like forms. Teachers can also write for children. Goals: Children realize that drawing and writing are different and that writing is meaningful Children start to represent sounds in words (invented spelling)	Children do important “brain work” when they draw, think and notice. Through drawing children make the big “jump” from noticing real or concrete objects (3 dimensional) to being able to represent these objects on paper (2-dimensional). Realizing that real things can be represented by a drawing is a big step for a young child. After this they realize that real things can be represented by words and sounds, and that drawing is different to writing.

	Activities	Research evidence
Shared storybook reading	Teachers read storybooks with rich language. Goals: Encourage a love of books Develop language through discussion before, during and after the story Develop concepts about print	Shared storybook reading at a young age has been shown to be linked to language and literacy outcomes at school. It is not only story reading that is vital, but the talk that accompanies book reading. The impact of shared reading is not necessarily seen on early word reading skills, but rather on language skills which support reading comprehension. Through books, children encounter new vocabulary and more complex sentence structures, syntax and descriptions (Adams, 1990). Books can lead to discussions about past experience or predictions about what will happen. Inferences can also be made about things that are not in the text or illustrations.
Making little books	Children read and make their own little library of books based on the stories that have been told in class. Goals: Children learn concepts about print (reading from left to right; concept of a word) Children pretend 'read' and have an experience of being readers	Through the little books children are engaged in developing language by retelling stories they have heard. They learn important concepts about print such as reading from left to right.

	Activities	Research evidence
Creating opportunities for emergent writing	Teachers are given ideas for including writing in play activities e.g. writing a letter, writing a shopping list. Goals: Teachers realize the value of emergent writing as part of children's play. Children use writing for meaningful purposes and learn to understand why we read and write. Children experiment with writing by forming scribbles, letter-like forms, and random strings of letters.	Writing begins with children experimenting with scribbles, letter-like forms, and random strings of letters. Children realize that written language represents ideas and meaning and they will often "read" their printed messages. Scribbles and pretend writing are the first steps towards increasingly accurate representations of sounds in invented spelling. As children think about how to represent the sounds of words through their writing, they are building skills that will be useful for reading as well (Dyson, 1982; Teale & Sulzby, 1986). It is vital that five- to six- year old children are given the time and space to engage with written language – using writing in play and experimenting with print as they start to figure out the way the system works. Without this foundation, many children spend their first few months in Grade One learning letters and remembering words by sight and stories by rote, but not understanding how the letters they are learning are connected to words and books. They also become passive recipients of knowledge from the teacher, rather than beginning school with the tools to be active participants in the process of learning to read and write.

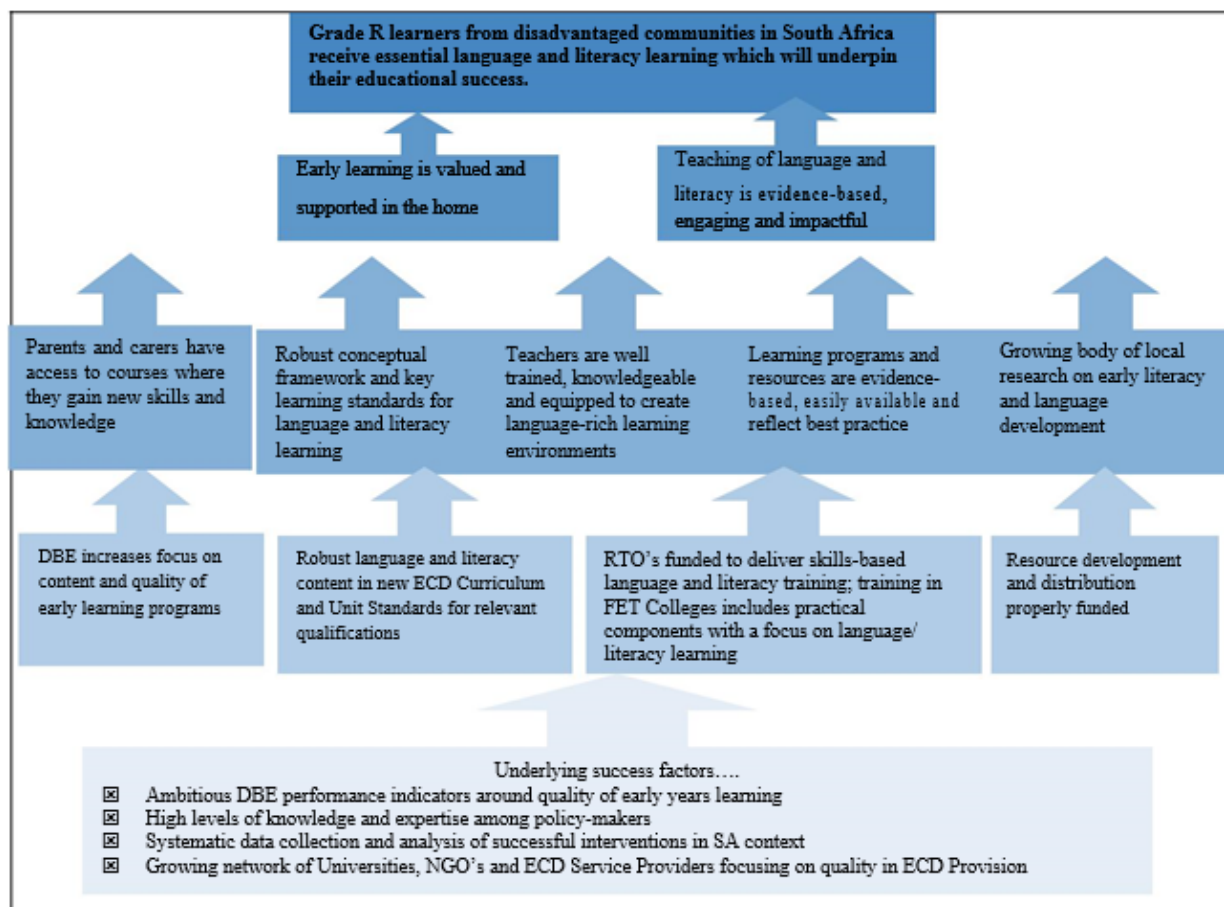
	Activities	Research evidence
Developing letter-sound knowledge	Children play games to develop awareness of beginning sounds and letter-sound knowledge. Goals: Children learn to hear beginning sounds in words and relate these to letters. Children have opportunities to write letters using paintbrushes, sticks in the sand, whiteboard markers Children start to observe letters in their environment and use letter cues to 'read' words	One of the key challenges facing beginner readers is discovering the alphabetic principle - that letters represent sounds in words, even though the representation is only approximate. Letter-sound knowledge plays a role from the earliest stages of literacy development (Ehri, 1997; 1998; Dixon, Stuart and Masterson, 2002; Stuart, Masterson and Dixon, 2000). Studies have shown that children from high poverty contexts generally begin school with less well-developed letter knowledge and phonological awareness than their middle-class peers and that these differences contribute to differences in early word reading skills (Stuart, 1990; Bowey, 1995). When preschool children from disadvantaged contexts participate in targeted interventions, their letter knowledge and phonological awareness improves, and effects are seen in their word reading and spelling skills. Research in the South African context has shown the importance of early letter-sound knowledge and phonological awareness (O'Carroll, 2006; Nadler-Nir, 1997).
Teaching listening skills and phonological awareness	Teachers use songs and games to develop listening skills and awareness of sounds in words. Goals: Children learn to think of words in terms of sounds, not just what they mean. Children begin to hear beginning and end sounds in words and break words into syllables/sounds.	Juel (1988) found that 'children who became poor readers entered first grade with little phonemic awareness poor entering phonemic awareness appeared to contribute to a very slow start in learning spelling- sound correspondences by the end of fourth grade the poor decoders had still not achieved the level of decoding that the average to good readers had achieved by the beginning of second grade' (Juel, 1988, p. 14).

Annex B: STELLAR Program Hypothesis and Theory of Change

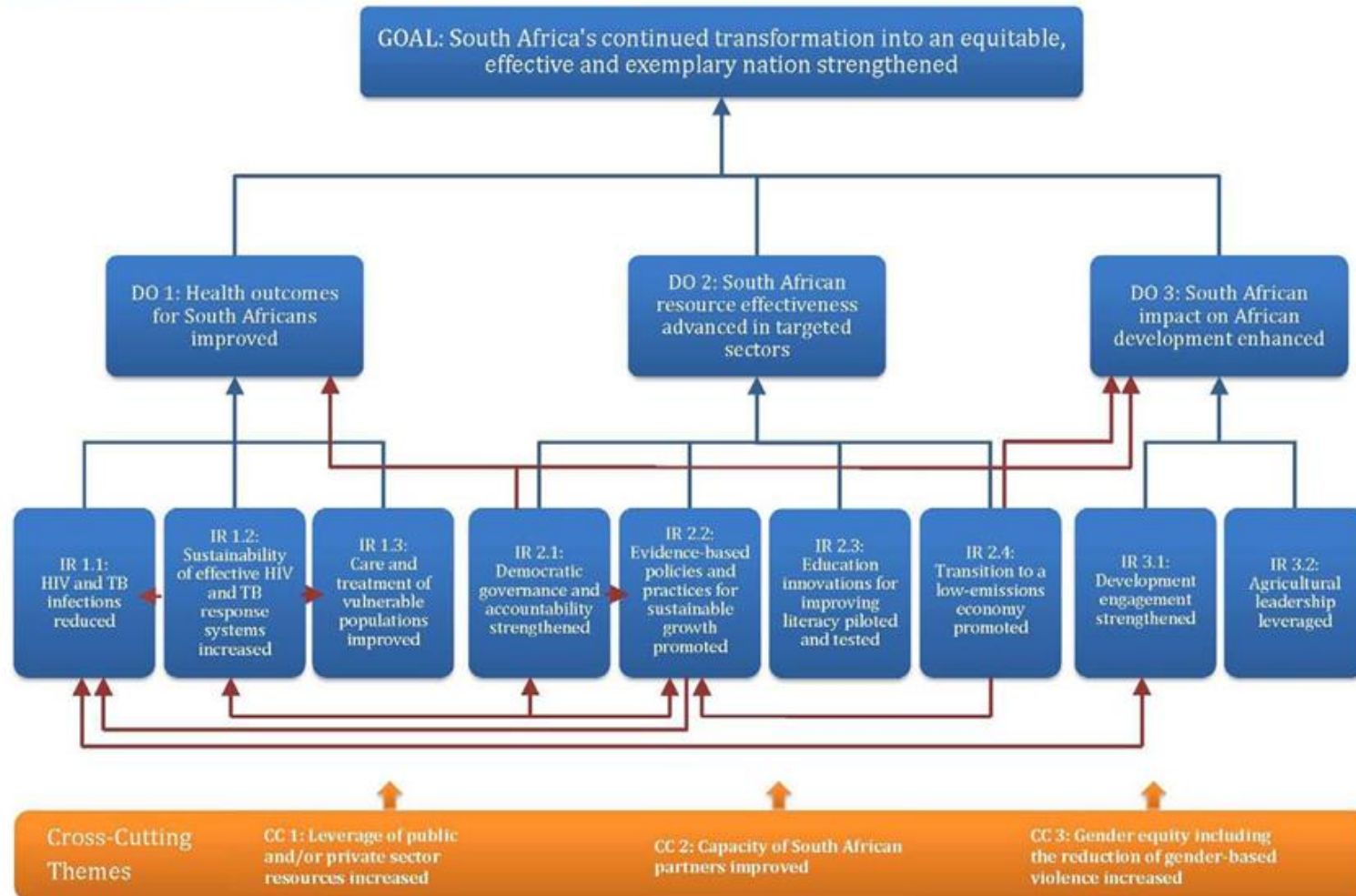
The STELLAR Program seeks to address the following problem statement:

Grade R Children from disadvantaged communities in South Africa miss out on the essential language and literacy learning which would otherwise underpin their educational success.

The diagram below provides the theory of change that underpins Wordworks' programs.



Annex C. USAID/SA Bilateral Country Development Cooperation Strategy – Public Version
Goals and Development Objectives: USAID/Southern Africa



ANNEX 8: STATEMENT OF DIFFERENCES

Statement from Wordworks

16 May 2018

We offer the following comments in response to the evaluation report:

1. Development of evaluation tools

We acknowledge the fact that Grade R is an under-researched area in the South African context, and that there is a lack of evaluation and assessment tools, specifically focussed on children's early language and literacy competencies at this level, that have been developed, validated and normed. We hope that the tools developed for this study will inform further test development for future evaluation studies. While we understand that the tools were designed to be aligned with the current national curriculum, we would caution that there are limitations to this approach because some aspects of the Grade R curriculum for literacy and language are not clearly articulated, and are not necessarily aligned with the latest research. It seems important that this is taken into account when assessment tools are developed, refined or selected for future studies.

2. Design of study and delayed start of the evaluation

In we consider the impact of the intervention on teacher practice and learner outcomes in Grade R, the report describes positive outcomes for both learners and teachers, and there are also positive findings on the Classroom Environment Tool. The report indicates clear change in the correct direction but does not point to statistically significant change for children benefiting from ELIT-trained teachers. Lack of statistical significance does not imply that positive change didn't happen, but instead that we can't say with certainty that the positive change is as a result of the intervention. The uncertainty extends in both directions, so the real change could, in fact, be even larger. The report notes that statistical significance might well have been found if the delayed starting data of the evaluation had not interfered with the design.

We have over the course of the evaluation articulated our concerns that the pre- and post-tests for learners were only three months apart, and that they took place in the last months of the year. This was not part of the original design, but resulted from delays in signing the contract for the evaluation. The question is whether the research should have been abandoned for this reason, or another approach found. The delay meant that this was not a pre- and post-test, but instead a midline (or even later) and post-test. Relying on the last months of the intervention also does not take into account that improvements effected through this type of programme would probably result in bigger "leaps" in performance nearer the beginning than towards the end of the period. The results would also have been affected by the fact that schooling activities and energy generally tail off quite substantially from November of each year, if not earlier.

3. Shifting parameters

Wordworks acknowledges the challenges of conducting an evaluation of a multi-stakeholder project. We are aware that as implementation proceeded, it became difficult to ascertain what was in fact being evaluated as the model and timeframes for the project shifted during implementation. The original model included training and cluster meetings over the course of a calendar year and this was the basis for assumptions about learner progress. When the project was truncated by the Western Cape Education Department to a 6 month training cycle, perhaps it warranted a review of assumptions and predicted outcomes. Given that the literature on teacher

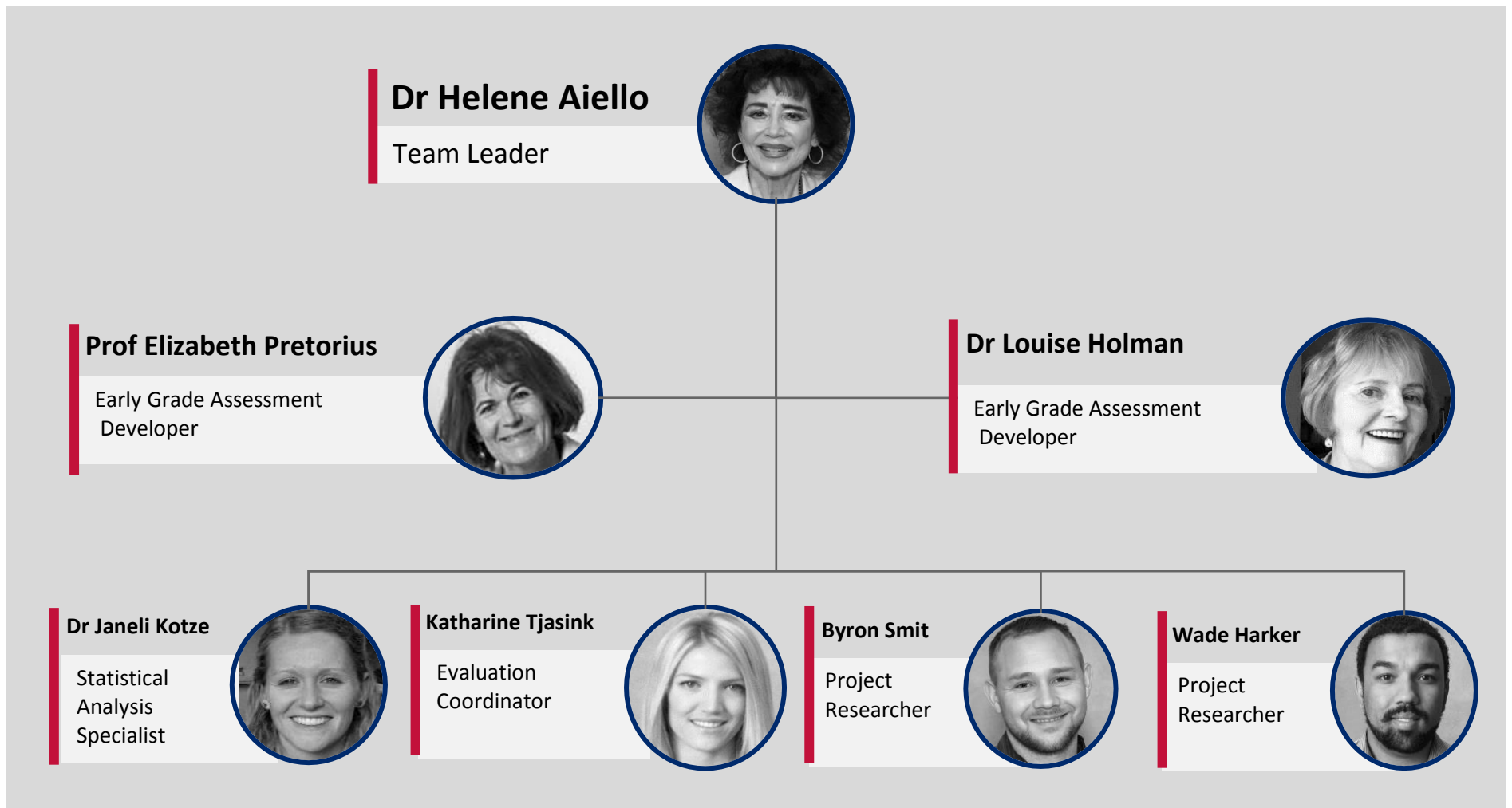
development anticipates that training and support is needed over two years to effect change in teaching practices and learner outcomes, it was perhaps ill-advised to anticipate improvements in learner outcomes at the end of Grade One given that the teacher training and support was only conducted over 6 months. It may have been more realistic to assess whether teachers carried over their learning to a second cohort of learners, for example. A review of evaluation plans might well be necessary where project plans change and therefore the theory of change no longer holds.

4. Published of evaluation results in the context of points 2 and 3 above

Depending on how the results of the evaluation are disseminated there is potential reputational risk and a potential impact on resourcing of a similar implementation. Where technicalities of evaluation design and statistical power are not understood, the conclusions drawn risk undermining the participant's lived experience of change and the motivation and optimism of role players. We have appreciated the opportunity we have had to engage with the results during the course of the evaluation, however, we are uncertain as to how the findings will be interpreted without a considered and carefully managed engagement process.

We would like to thank USAID and Khulisa Management Services for the detailed and thorough report and the opportunities to engage with and learn from the evaluation team. The evaluation outcomes have contributed to our materials revision process and a stronger programme going forward. The learning from the evaluation is also being used to inform project planning and future training models.

ANNEX 9: EVALUATION TEAM



ANNEX 10: LEARNING BRIEF FOR THE GOVERNMENT OF SOUTH AFRICA



USAID
FROM THE AMERICAN PEOPLE

LEARNING BRIEF: IMPACT EVALUATION OF THE WESTERN CAPE EMERGENT LITERACY INTERVENTION

Contract Number: AID-674-I-12-00002, Order Number: AID-674-TO-15-00003



February 2018

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

LEARNING BRIEF: IMPACT EVALUATION OF THE WESTERN CAPE EMERGENT LITERACY INTERVENTION

CONTACT DETAILS

Dr Helene Aiello

PO Box 923

Parklands, 2121

Telephone: 011-447-6464

Fax: 011-447-6468

Email: haiello@khulisa.com

Web Address: www.khulisa.com

INTRODUCTION

In 2015, the United States Agency for International Development (USAID) funded an Emergent Literacy intervention in Grade R (ELIT) targeting Grade R teachers at 3000 Department of Basic Education (DBE) registered sites in the Western Cape. Simultaneously, USAID commissioned an impact evaluation of ELIT. The evaluation looked at the language and literacy performance of Grade R learners and followed the same learners into Grade 1 to assess whether the intervention had a lasting impact.

The impact evaluation also looked at the effects of the ELIT intervention on teachers and at the implications of the findings for the provincial education system as a whole.

This learning brief presents the results of the ELIT evaluation.

EVALUATION RESULTS

The ELIT intervention appears to have had an initial positive effect on the emergent language and literacy of Grade R learners. However, this effect decreased by the end of Grade R, and disappeared by the end of Grade 1.

Although learners made consistent progress from Grade R into Grade 1, The Grade 1 results in both the comparison and intervention groups show that early reading performance is generally very low in the Western Cape. Grade 1 learners in the province are not yet mastering the foundational decoding skills that learners need for reading; phonemic awareness, letter-sound knowledge, and oral reading fluency.

The ELIT intervention, as currently implemented, appears to have had **no lasting impact** beyond Grade R.

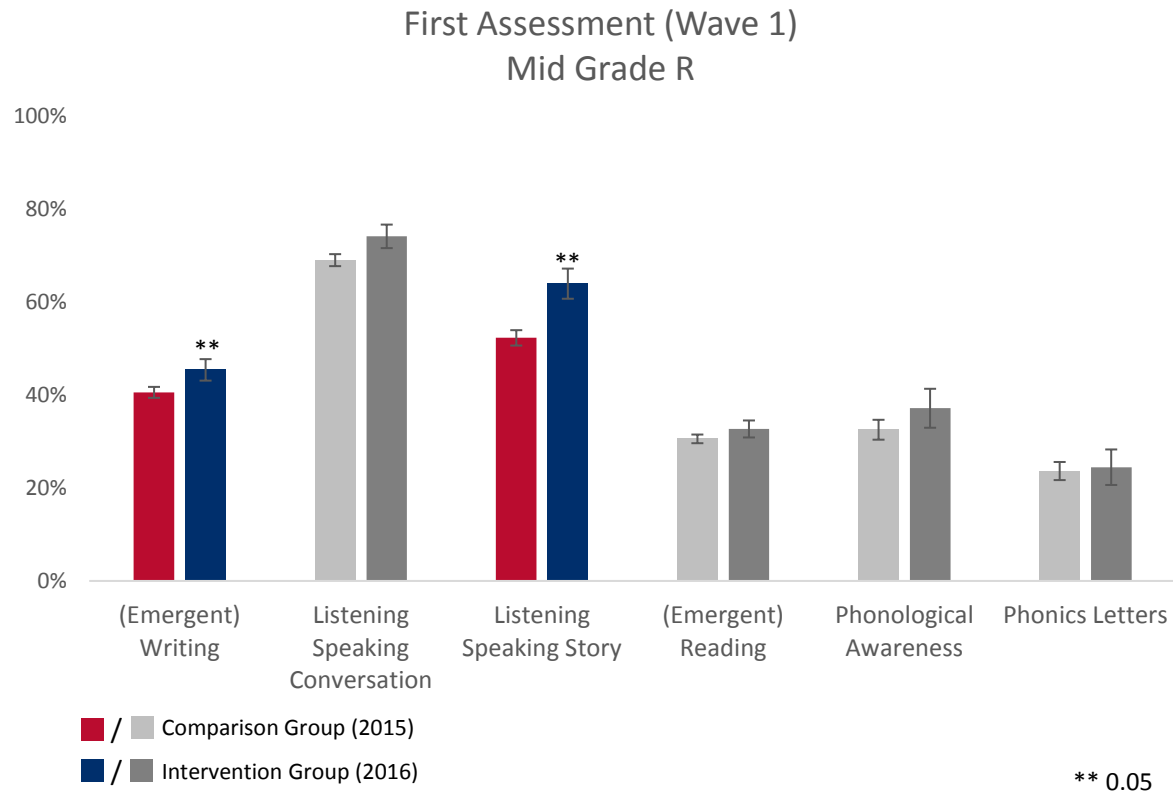
Results: Learner Assessment

This evaluation assessed learner language and literacy performance at three points;

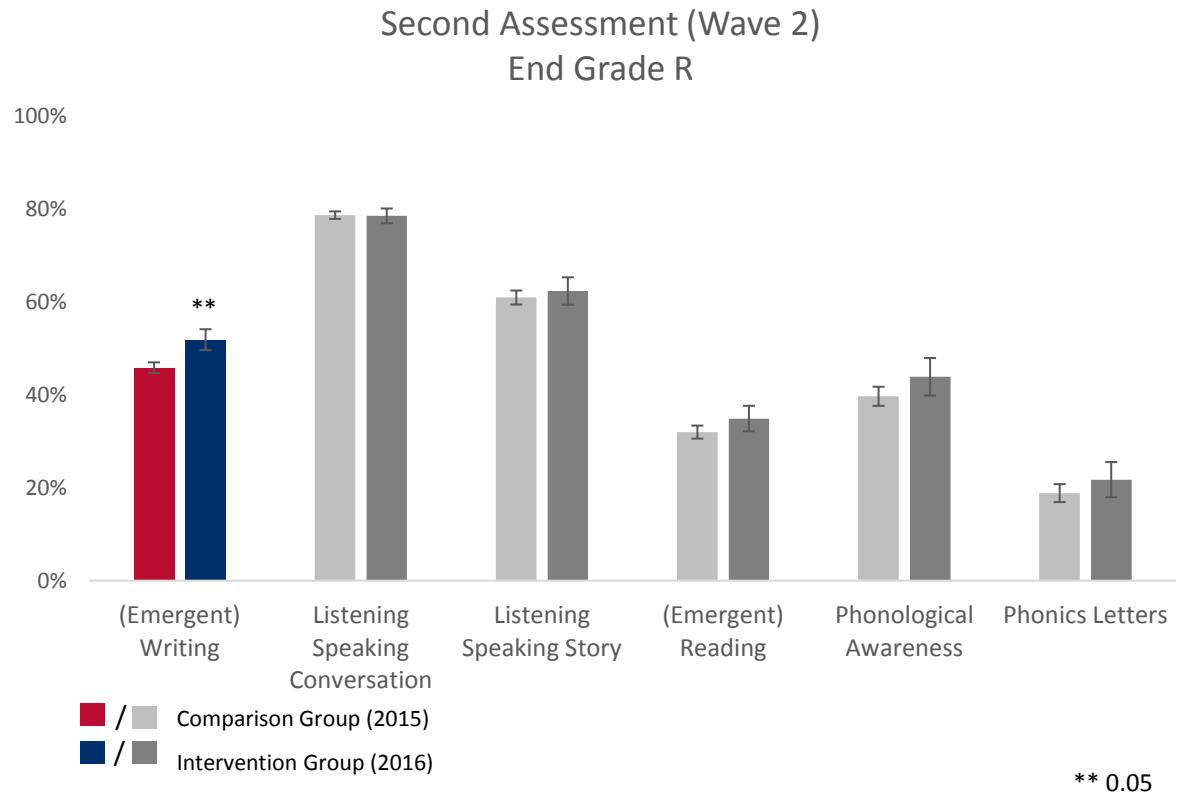
- First assessment at the midpoint of Grade R
- Second assessment at the end of Grade R
- Third assessment at the end of Grade 1

The Learner Assessment Tool was based on the Curriculum Assessment Policy Statements (CAPS) and ELIT.

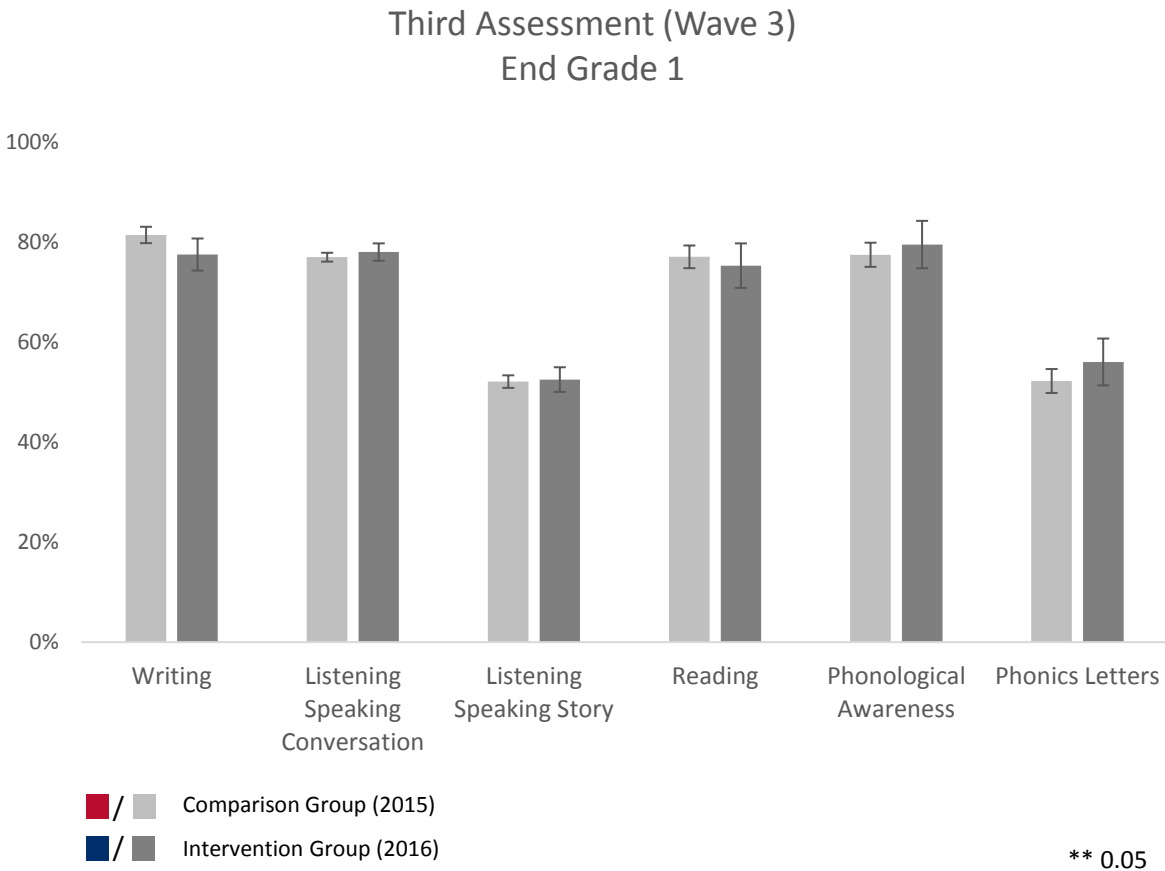
The first assessment found that the intervention group did significantly better than the comparison group in the *emergent writing* and *listening and speaking based on a story* subtasks of the Learner Assessment. Learners in both groups scored relatively poorly on the *reading*, *phonological awareness* and *phonics letters* subtasks.



At the end of Grade R, learner scores in the intervention and comparison groups improved overall with the intervention group still significantly better in the *emergent writing* sub-task. Learner performance in *emergent reading*, *phonological awareness* and *phonics letters* was still relatively weak in both groups.

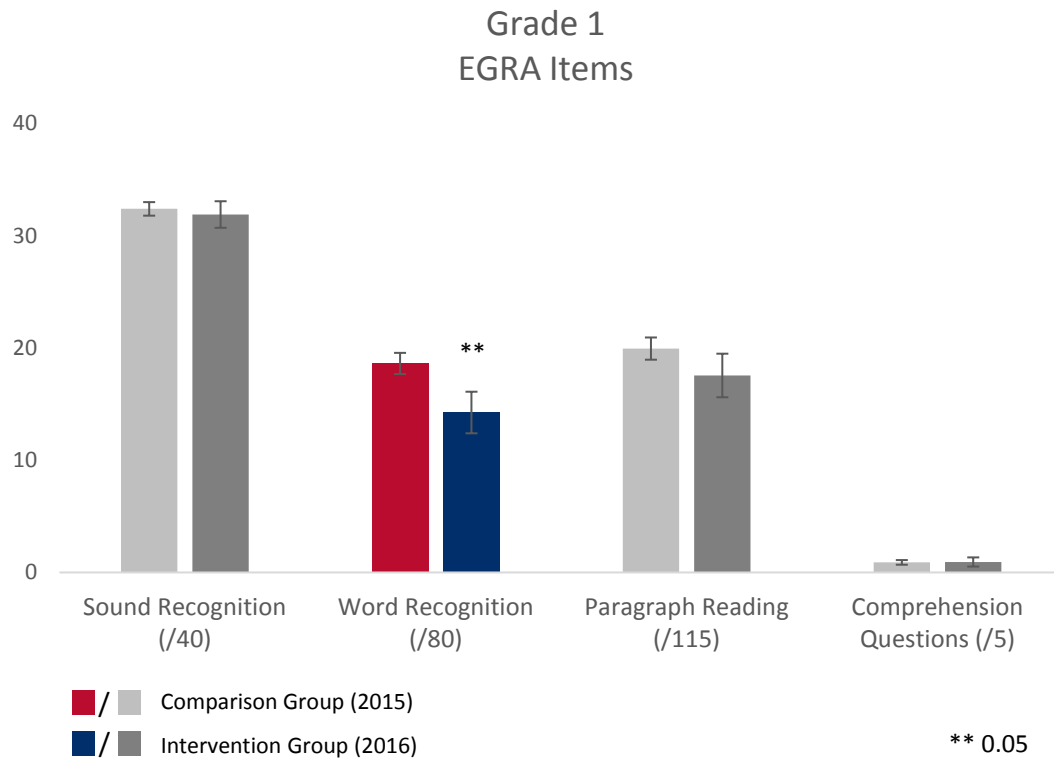


Large improvements were seen in all subtasks by the end of Grade 1. However, there were no statistically significant differences between the intervention and comparison groups at this point.



Results: EGRA Sub-Tasks

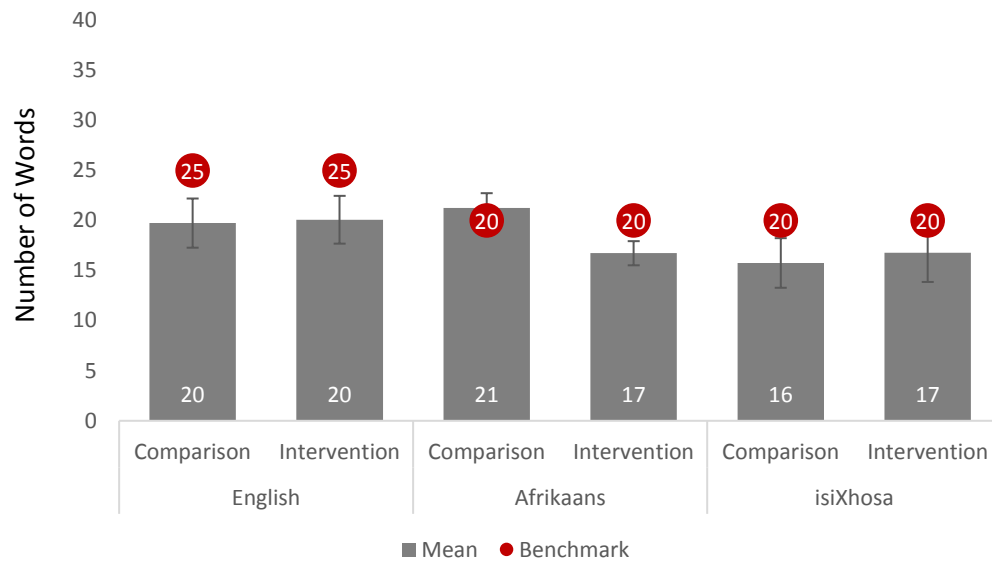
In addition to the Learner Assessment Tool, the Early Grade Reading Assessment (EGRA) was administered at the end of Grade 1. Learners in the comparison and intervention groups performed equally well in the *letter sound recognition* subtest of the EGRA. Interestingly, the comparison group outperformed the intervention group in *word recognition* and *paragraph reading*. For the *word recognition* subtest, this difference is statistically significant.



However, other than the Afrikaans comparison group and the isiXhosa intervention group, learners in both the intervention and comparison groups did not reach the Western Cape Education Department (WCED) benchmarks in *word recognition* or *paragraph reading*.

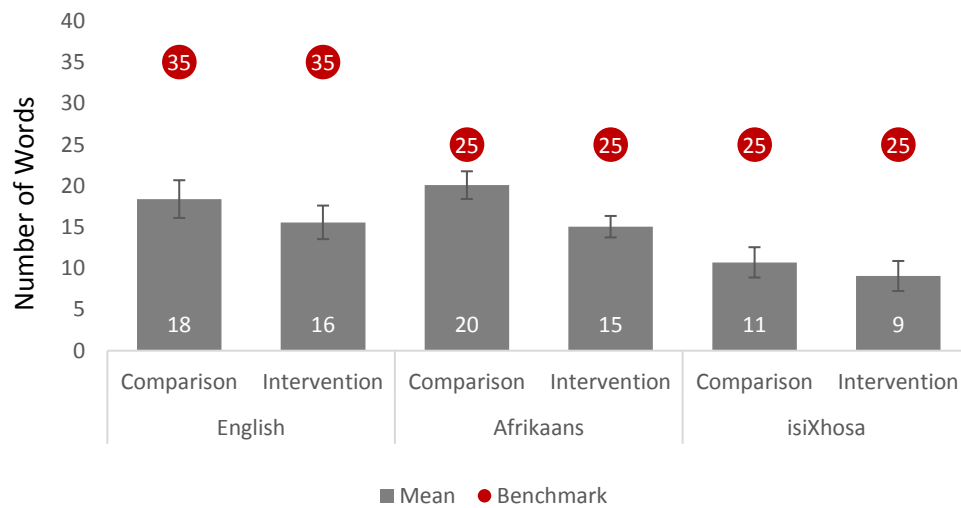
Grade 1 EGRA: Word Recognition

Words Correct per Minute



Grade 1 EGRA: Paragraph Reading

Paragraph Words Correct per Minute



Only learners at the 75th percentile could answer one *paragraph comprehension* question correctly.

LESSONS AND IMPLICATIONS

International research consistently finds that quality teaching/ learning in Grade R can make a difference to early reading success once children start school formally, but the results from the ELIT intervention do not reflect this. This raises questions about implementation of the intervention, exposure to pre-reading skills in the years preceding Grade R and pedagogies and classroom environments following Grade R.

- Implementation: It appears that the 6-month intervention of monthly cluster meetings and a five-day block training that the Grade R teachers received was enough to initially improve Grade R scores but not enough to sustain those improvements. More ongoing support may have been needed to sustain changes in classroom practice that were starting to emerge. USAID's landscape report on early grade literacy found that changing teacher behaviour takes time and necessitates sustained efforts (Cochran-Smith & Zeichner, 2005; cited in Kim, Boyle, Zuilkowski, & Nakamura, 2016). South African research finds that intensive and sustained coaching can improve learner results in reading (EGRS Endline Report 2017, DBE).
- Reading performance: The Grade 1 results (in comparison and intervention groups equally) show that early reading performance is generally very low in the Western Cape. Girls outperform boys in emergent reading skills in Grade R, and their advantage doubles in Grade 1.
- Alignment between Grade R and later grades: There must be strong alignment between the teaching of language in Grade R and in Grade 1. Grade 1 early reading interventions must develop, support and sustain foundational reading skills. USAID's landscape report found that quality instruction across multiple years is essential; a single year of instructional support is insufficient (Kim et al., 2016).

While more information is needed to better understand policy implications, the evaluation did uncover some implementation lessons:

- The cascade training approach may have had an initial effect but without provincially mandated support post July 2016, the intervention advantage appears to have been lost. A recent World Bank report on early grade reading in developing countries indicates that following up initial training with coaching and monitoring is needed to compound the effectiveness of teacher training (Kelly & Graham, 2018).
- The logic of change underlying most professional development programs assumes that there is fidelity to the professional development package in the way that it is administered, rolled out and the contents imparted, demonstrated, and supported in follow up activities. To document impact, it is very important that the intervention adhered strictly to the program design. Potentially, lack of fidelity to the professional development package, implementation modes and support may have negatively affected the change process.

The following are the **main recommendations** emanating from this evaluation:

Future impact evaluations:

- Should rigorously evaluate the impact of a program before it is implemented at scale. To facilitate rigorous impact evaluation studies implementation plans need to allow for the random selection of intervention and comparison groups.
- Should test whether the funding spent on teacher development is leading to any change in learner performance. Evaluating teacher training efforts are very important and training should therefore happen in a way that allow for the effects to be evaluated.

Future early grade literacy interventions:

- Should be tested before being taken to scale.
- Should mandate a specific dosage of training and ongoing support that must be sustained.
- Should articulate how ongoing teacher training and support will become part of the system so that it will continue to reach new and experienced teachers.
- Should be strategically tailored to the needs of schools. It is not always feasible to roll-out a one-size-fits-all intervention to all schools in a province.

Overall, the evaluation results suggest that a good Grade R is necessary but may not be sufficient for future language and literacy acquisition.