The Tablet Teacher:

Learning Literacy through Technology in Northern Sotho

Submitted in partial fulfilment of the requirements for the degree of

MASTER OF ARTS

DEPARTMENT OF ENGLISH LANGUAGE AND LINGUISTICS

RHODES UNIVERSITY

Submitted by:

Kelly Shiohira

January 2017

ABSTRACT

This study evaluates the efficacy of the Bridges to the Future Initiative – South Africa 2 (BFI) tablet program on early literacy skills, as well as the ways in which learner-operated technology interacts with a traditional South African education system. The BFI is a curriculum-aligned early literacy development intervention implemented through technology in grades 2 and 3 in Northern Sotho[1] first-language schools. A mixed-methods research design was utilized, involving three components: a literacy skills test administered through a time sequence trial design; a curriculum-aligned uptake and retention test using a pre-post design; and a qualitative research component including classroom observation, participant interviews and prompted drawings by learners.

Paired sample t-tests show significantly higher gains during the treatment period in fluency and comprehension, and significantly higher gains in the control period in decoding individual words. It is theorized that this is due to teacher emphasis on emergent literacy. When initial ability is taken into consideration, all ability levels gain more on average during the treatment period in at least one measured skill. Regression analysis determines that time spent on the BFI program is not the most significant determiner of gains in the intervention period. Qualitative analysis supports this finding and suggests that program use cannot replace quality classroom practice in advancing literacy skills.

Learners performed better after a delayed retention period than in an initial uptake test, indicating high rates of retention of knowledge gained through program use and traditional instruction, but inconsistent access to literacy skills gained.

[1] "Northern Sotho" is used as it is the language as named in the constitution. The schools identified themselves as "Sepedi-speaking". Linguistically, Sepedi is referred to as a dialect of Northern Sotho.

ACKNOWLEDGEMENTS

Many thanks to:

USAID, World Vision, and Australian Aid as contributors of the All Children Reading Grant that made BFI possible, and to UNESCO for funding the adaptation and rollout of the tablet program. Especial thanks to Carien Vorster, USAID South Africa and all the education staff involved in both iterations of the BFI program for their support of the development and the evaluation.

My advisors, Kristin van der Merwe and Mark de Vos at Rhodes University, South Africa, for their dedicated guidance and invaluable input.

'Dr. Dan' Wagner, at the University of Pennsylvania, for setting me on the path to South Africa and the Bridges to the Future Initiative.

The Molteno Institute for Language and Literacy, for the wonderful opportunity to work on this project. To Paula Gains, Jenny Katz and Mirna Lawrence, for their advice and input on developing the programme content, and to all the Molteno Institute staff for helping to make the programme and the intervention possible. Especial thanks to Margaret, Jonson, Star, Stanley, Nellie, Rachel, Seipati, Johanna, Thelma and especially Takalani for their assistance and support in development of the programme in African languages and all the sacrificed Saturdays!

Ben Tladi and the Limpopo Board of Education for their support in rolling out the programme, and to the principals and teachers who took the time to learn, deliver the programme and support this research.

Finally, to Thumeka Nakani, Mayli Zapata, Christian Cochon, Amy Lee and Nathan Castello, who supported the intervention in various ways during their internships with the Molteno Institute.

TABLE OF CONTENTS

ABSTRACT i

ACKNOWLEDGEMENTS ii

TABLE OF CONTENTS iii

LIST OF TABLES vii

LIST OF FIGURES viii

LIST OF ACRONYMS x

STRUCTURE AND ORGANIZATION OF THE PAPER xi

CHAPTER 1: CONTEXT AND LITERATURE REVIEW 1

1.1 Background and Rationale for the Intervention 1

1.2 The Acquisition of Literacy 5

1.2.1 Oral Language Development 6

1.2.2 Print Concepts 6

1.2.3 Phonemic Awareness and Phonics 7

1.2.4 Fluency 12

1.2.5 Comprehension 14

1.3 Early Literacy Education Policy in South Africa 15

1.3.1 The Northern Sotho Language and its Pedagogy 15

1.3.2 Curriculum Expectations in the Early Grades 17

1.4 Challenges in Early Literacy Education in Bantu Languages 18

1.5 Technology in Education 19

1.6 The BFI: A Study in Technology and Literacy Acquisition 22

CHAPTER 2: METHODOLOGY 24

2.1 Literature on Mixed-Method Research Design 24

2.2 Research Framework and Design 26

2.2.1 Sample Schools 28

2.3 Implementation Instruments 30

2.3.1 Tablets 30

2.3.2 The Bridges to the Future Initiative Tablet Program 31

2.4 Assessment Tools 35

2.5 Challenges and Limitations to the Study Design 36

CHAPTER 3: QUALITATIVE ANALYSIS 39

3.1 Qualitative Research Design 40

3.1.1 Classroom Observation 41

3.1.2 Teacher Interviews 42

3.1.3 Prompted Drawings 44

3.1.4 Analysis of Data 45

3.2 Presentation of Qualitative Data 45

3.2.1 Overall Observations 46

3.2.2 Classroom Observations: School A 47

3.2.3 Classroom Observations: School B 52

3.2.4 Teacher Interviews: Interactions with Learners during Program Use 55

3.2.5 Learner Drawings 57

3.3 Concluding Remarks on Qualitative Observations and Results 59

CHAPTER 4: THE EARLY GRADE READING ASSESSMENT (EGRA) TEST OF LITERACY SKILL ACQUISITION 61

4.1 Literature and Theory on Literacy Assessment in Schools 63

4.1.1 Approach to Benchmarking Reading Rates in Northern Sotho for Early Grades 65

4.2 Research Design and Methodology 67

4.2.1 Testing Instrument 68

4.2.2 Sample 70

4.2.3 Testing Procedures 71

4.2.4 Treatment and Analysis of Data 73

4.3 Assessment of Literacy Skills Using the Early Grade Reading Assessment 74

4.3.1 Overall Performance Changes 74
4.3.2 Analysis by Age 80
4.3.3 Analysis by Gender 81
4.3.4 Hours of BFI Program Use and improvement in EGRA skills 82
4.3.5 Grapheme-Phoneme Assignment 86
4.3.6 Simple Decoding (Word Reading) 89
4.3.7 Correct Words per Minute 93
4.3.8 Comprehension 99
4.4 Conclusions from the Early Grade Reading Assessment 107
CHAPTER 5: SKILLS ACQUISITION AND RETENTION OF KNOWLEDGE 110
5.1 Literature and Theory on Retention of Knowledge 111
5.2 Methodology for Testing Retention of Knowledge 112
5.2.1 Instrument 113
5.2.2 Sample 115
5.2.3 Testing Procedures 116
5.2.4 Treatment and Analysis of Data 116
5.3 Results and Analysis of the Retention of Knowledge Data 118
5.3.1 Overall Results 118
5.3.2 Manipulation 120
5.3.3 Sentence Mechanics: Grammar and Punctuation 125
5.3.4 Explicit Factual Knowledge 129
5.3.5 Comprehension 132
5.4 Conclusions on Skills and Retention 136
CHAPTER 6: INTEGRATING QUANTITATIVE AND QUALITATIVE DATA 141
6.1 Hours of Program Use 141
6.2 Tablet Integration into the Classroom 144
6.3 Follow-up Activities and Integration into Other Pedagogy 144

6.4 Peer Interactions 146
6.5 Conclusory Remarks on the Integrated Analysis 147
CHAPTER 7: CONCLUSION 148
Question 1: System Placement and Context 149
Question 2: Classroom, School and Grade Factors Contributing to Observed Outcomes 150
Question 3: Impact of the Technology Intervention 151
Question 4: Performance by Age 152
Question 5 and 6: Skills Uptake and Retention 152
Question 7: Correlations between the EGRA and Retention Tests 153
7.1 Other Observations and Directions for Further Research 155
7.1.1 Reading Speed 155
7.1.2 Error 156
7.1.3 Phonemic Strategies of Early Northern Sotho Readers 156
7.1.4 Retention of Knowledge 158
7.2 Final Remarks 159
APPENDIX A: Memorandum of Agreement between the Molteno Institute for Language and Literacy and the Limpopo Provincial Department of Education 160
REFERENCES 170

LIST OF TABLES

Table 1: Comparison of Eye-Tracking Statistics in English and isiZulu, adapted from Land, 2015 12
Table 2: Sample School Learner, Teacher and Class Numbers 29
Table 3: Comparison of Tablet Instruments Used 30
Table 4: Number of Classroom Observations 41
Table 5: Northern Sotho Reading Benchmarks 67
Table 6: English EGRA Comprehension Questions and Classification Under Barrett's Taxonomy 70
Table 7: Final Participant Numbers after Attrition 71
Table 8: Example of EGRA Data Collected 74
Table 9: Attainment Level Descriptions and Formulas 77
Table 10: Change in Attainment Levels of Learners 79
Table 11: Mean and Standard Deviation Results by Grade, School and Test Item 84
Table 12: EGRA Performance by Attainment Level and Task 85
Table 13: Gains in Word Reading by School and Grade, Control and Treatment period 92
Table 14: Retention Test Categories by Grade 114
Table 15: Retention Test Participant Numbers 115
Table 16: Change in Retention Performance by Class and Skill Category 119
Table 17: Overview of Program Use, EGRA scores and Retention Test Scores 120
Table 18: Individual Learner Variance in Manipulation 124
Table 19: Program Use, EGRA and Retention Performance on Manipulation Tasks 125
Table 20: Individual Variance on Sentence Mechanics 127
Table 21: Individual Variance on Factual Knowledge 131
Table 22: Individual Learner Variance in Comprehension 134
Table 23: Program Use, EGRA and Retention Results for Comprehension 135
Table 24: Summary of Quantitative Results and Classroom Observations 141
Table 25: Summary of Quantitative and Qualitative Data by Class 143
Table 26: Performance Ranking by School and Grade 150
Table 27: Summary of Test Performance and Qualitative Data 151

LIST OF FIGURES

Figure 1: Research Timeline 27
Figure 2: BFI Development Process 32
Figure 3: Screenshot of a BFI Audio-Visual Component Frame 33
Figure 4: Screenshot of a BFI Sentence Building Activity 34
Figure 5: Education System Interactions with the BFI 46
Figure 6: Research Timeline 68
Figure 7: Example of Northern Sotho grapheme-phoneme assignment 69
Figure 8: Example of Northern Sotho Familiar Word Reading 69
Figure 9: Boxplot of Overall Learner Performance by Test Period 75
Figure 10: Attainment Level Distribution over Time 78
Figure 11: Average Score on Literacy Skills by Gender 82
Figure 12: Average Change in Score by Gender 82
Figure 13: Performance on Grapheme-Phoneme Relationships over Time 86
Figure 14: Changes in Distribution Patterns, Grapheme-Phoneme Assignment 87
Figure 15: Performance on Grapheme-Phoneme Assignment over time, by School and Grade 88
Figure 16: Changes in Learner Performance, Grapheme-Phoneme Assignment 88
Figure 17: Grapheme-Phoneme Assignment by Attainment Group 89
Figure 18: Simple Decoding over Time 90
Figure 19: Changes in Distribution Patterns, Word Reading 91
Figure 20: Performance on Word Reading over Time, by School and Grade 92
Figure 21: Changes in Word Reading by Attainment Level 93
Figure 22: Performance on Passage Reading (cwpm) over time 95
Figure 23: Changes in Distribution Patterns, Correct Words Per Minute (CWPM) 95
Figure 24: Performance on Reading over Time, by School and Grade 96
Figure 25: Change in Learner Performance, Passage Reading 97
Figure 26: CWPM by Attainment Group 97
Figure 27: Error Rates over Time 98
Figure 28: Error Rates versus Correct Words per Minute 99
Figure 29: Performance on Comprehension over Time 100
Figure 30: Changes in Distribution Patterns, Comprehension 101
Figure 31: Performance on Comprehension over Time, by School and Grade 102
Figure 32: Change in Learner Performance, Comprehension 102
Figure 33: Learner Attainment in Comprehension over Time, by Group 103
Figure 34: Correct Words per Minute versus Comprehension 105
Figure 35: Percent of Learners with Error by Comprehension Score 106
Figure 36: Average Error by Comprehension Score 106
Figure 37: Research Timeline 112
Figure 38: Example of a Phoneme Segmenting Task 115
Figure 39: Example of a Sentence Mechanics Punctuation Task 115
Figure 40: Overall Change in Performance by Skill Category 118
Figure 41: Performance by School, Retention Pre and Post-test 119
Figure 42: Average Percent Scores by Class, Overall Manipulation 121

Figure 43: Average Percent Scores by Class, First Sound Identification 122
Figure 44: Average Percent Scores by Class, Syllable Segmenting 122
Figure 45: Average Percent Scores by Class, Sound Segmenting 122
Figure 46: Average Percent Score by Class, Overall Sentence Mechanics 126
Figure 47: Average Percent Score by Class, Parts of Speech 126
Figure 48: Average Percent Score by Class, Punctuation 126
Figure 49: Average Percent Score by Class, Conjunctions 127
Figure 50: Average Percent Score by Class, Overall Explicit Knowledge 129
Figure 51: Average Percent Score by Class, Story-based Factual Knowledge 130
Figure 52: Average Percent Score by Class, Reinforced Factual Knowledge 130
Figure 53: Average Class Scores, Overall Comprehension 132
Figure 54: Average Class Scores, Direct Questions 133
Figure 55: Average Class Scores, Inference Questions 133
Figure 56: Average Class Scores, Synthesis Questions 133
Figure 57: Interactions between Technology and the South African Education System 149

LIST OF ACRONYMS

ABET	Adult Basic Education and Training
ANA	Annual National Assessments
BFI	Bridges to the Future Initiative
BFISA2	Bridges to the Future Initiative: South Africa 2
CA	Curriculum Advisor
CAPS	Curriculum Assessment Policy Statements
cwpm	correct words per minute
DBE	Department of Basic Education
EDO	Education District Official
EGRA	Early Grade Reading Assessment
ESSS	Education Support and Social Services
FET	Further Education and Training
FP	Foundation Phase
FSI	First Sound Identification
GET	General Education and Training
HOD	Head of Department
ICT	Information Communications Technology
LTSM	Learning and Teaching Support Material
MOOC	Massive Open Online Course
OECD	Organization for Economic Cooperation and Development
PANSALB	Pan-South African Language Board
PC	Personal Computer
PIRLS	Progress in International Reading and Literacy Study
PISA	Programme for International Student Assessment
RCT	Randomized Control Trial
SAEB	Brazilian Basic Education Evaluation System
TST	Time Series Trial
USAID	United States Agency for International Development

STRUCTURE AND ORGANIZATION OF THE PAPER

This paper is divided into seven chapters.

The first chapter provides an overview of the context and literature pertinent to the study, an investigation into literacy skills gained and retained through the use of an early literacy technology program[2] in Northern Sotho, the Bridges to the Future Initiative (BFI).

The second chapter outlines the overall methodological approach to the research, which utilizes a mixed methods design in three parts: a method-independent literacy skills test, a method-dependent skills and retention test and qualitative research into program use and integration into the education system. This chapter discusses research frameworks, instruments used and school selection.

Chapter 3 discusses the findings of the qualitative research, to provide a contextual background for each school and grade under study for reflection against the quantitative outcomes. The chapter covers three qualitative methodologies used – classroom observation, interviews and learner drawings – and discusses the findings from this research.

Chapter 4 presents the methodology, analysis and conclusions from the method-independent literacy skills test.

Chapter 5 presents the methodology, analysis and conclusions from the method-dependent skills and retention test.

Chapter 6 provides an aggregate database summary of the qualitative and quantitative components and explores potential qualitative influences on quantitative outcomes.

Chapter 7 summarizes and presents the conclusions with regard to each research question, and outlines additional observations and directions for further research.

The research included the grade 2 and 3 classes of two schools, which will be referred to as school A and school B. The grade levels will be included as applicable, so that school A, grade 2 will be referred to as A2; school A, grade 3 will be referred to as A3; school B, grade 2 will be referred to as B2; and school B, grade 3 will be referred to as B3.

[2] The BFISA is both computer software and a literacy course. Following South African English conventions, 'program' is used to refer to the computer software, and 'programme' is used to refer to the literacy course and its components.

CHAPTER 1: CONTEXT AND LITERATURE REVIEW

1.1 Background and Rationale for the Intervention

Since the establishment of the new constitution and government in 1996, the South African government has invested heavily in education and education structures as the crux of the country's post-apartheid transformation efforts. By 1999, this theoretical principal had been underwritten not only by the Constitution of the Republic of South Africa[3], but also by the National Education Policy Act, the South African Schools Act, the Education White Papers on Inclusive Education and Early Childhood Education, and the first national curriculum, Curriculum 2005[4]. These documents outline the structures the post-apartheid government envisaged as necessary to fulfil the mandate of free, equal and unbiased education in South Africa, and creating in total an expansive and resource-heavy network of teachers, administrators, governing bodies and resourcing networks.

And yet, despite considerable investment, learner achievement remains low. Although the government itself in 2007 asserted that it had achieved the "most developed and well-resourced system of education and training on the African continent" (Department of Basic Education, 2007), a nearly concurrent review by the OECD of South African education (2008) stated that "South Africa has a high-cost, low-performance education system that does not compare favourably with education systems in other African countries, or in similar developing economies". This reality is further reflected in South Africa's performance on international examinations, such as the 2006 Progress in International Reading and Literacy Study (PIRLS). South Africa presented both grade 4 and grade 5 learners for the 2006 PIRLS examination, which is targeted to grade 4 learners, and with an average score of 302 points on the examination, the South African grade 5 learners were the lowest-performing of 45 participating education systems. Grade 4 learners performed even lower, at an average of 253 points (Howie et al., 2008). In 2011, South Africa opted to participate in prePIRLS, an easier version of the PIRLS examination with shorter texts, simpler vocabulary and grammar and more emphasis on direct comprehension questions, the answers to which can be retrieved verbatim from presented texts (Mullis et al., 2012). On the prePIRLS 2011, South African students averaged 461 points, still below the international centre point of 500 and the lowest of the three participating countries (Land, 2016; Howie et al., 2012).

[3] The Constitution of South Africa can be accessed at: http://www.gov.za/DOCUMENTS/CONSTITUTION/constitution-republic-south-africa-1996-1

[4] School policy documents can be accessed at http://www.education.gov.za/Resources/Policies.aspx

The fact that many learners in South Africa are not achieving functional literacy is clear. However, a number of factors impact the delivery of quality of education in rural communities, which complicates any simple solution. These include poor access roads which complicate service delivery, a prevalence of child-headed households, abuse and/or neglect of learners by parents, declining enrolment due to population migration and competition between local schools, poor or unqualified teachers and administrators, multigrade teaching, poor infrastructure, teacher shortages and high rates of teacher turnover (Shiohira, 2014; Pitsoe, 2013; Department of Basic Education, 2011). National reviews have also found worrying trends regarding retention and repetition of learners in the education sector, contributing to low efficiency of the system (Taylor et al., 2008). A review of the Eastern Cape in 2013 found that less than 15% of cohort members entering the school system in 2000 and 2001 emerged with a matric certificate on time (Shiohira, 2015).

While there is recognition of the complexity of the challenges facing the system, perhaps the most prominently addressed factor influencing educational outcomes in South Africa has to do with teacher capacity and professionalism. Teacher capacity has been determined by research as the most influential school-based factor in sustained learning outcomes (Hanushek et al., 2004; McCaffrey et al., 2003; Bembry et al., 1998). Yet in South Africa teacher quality suffers from under-staffing and poor retention of qualified teachers (Ladbrook, 2009; OECD, 2008; Taylor et al., 2003), in a context in which teaching does not attract highly skilled individuals to begin with (Armstrong, 2009). Additionally, there is significant variation in both quality and content of pre-service teacher training in South African institutions of higher education (JET, forthcoming), and teachers trained by these institutions often find themselves deployed to subjects in which they are not specialists (Van der Berg, 2003), where they engage in active teaching only 6 to 56 percent of scheduled class time (Chisholm et al., 2005). As a result of these realities, interventions aimed at improving educational outcomes in South Africa tend to focus on teacher capacitation through workshops and in-service training and/or provision of teacher support materials such as lesson plans (see, for example, the Gauteng Primary Literacy and Maths Study).

However, ICT presents another possible solution which may be more suited to the complexity of the terrain, namely supplementing teacher instruction with instruction through technology. The appeal of technology is the ability to deliver consistent high quality instruction across contexts, at a cost which decreases as scale is increased due to the initial production cost. And yet, results from educational studies involving technology provide mixed results on efficacy,

and ultimately questions remain regarding what and how much children can learn from educational technology.

The intervention in this study, the Bridges to the Future Initiative (BFI)[5], uses tablet technology as a medium to deliver content training in early literacy directly to learners. The BFI program[6] was developed by the researcher and a team of language experts as part of work done with the Molteno Institute for Language and Literacy. The programme is based on the current South African curriculum for foundation phase home language learners, under the auspices of an All Children Reading grant from USAID, Australia Aid and World Vision. Program implementation offered a unique opportunity to test the uptake of linguistic and curriculum skills taught via technology in foundation phase early literacy learners. The scope of this study investigates impact on a group of Northern Sotho first language[7] speakers in grades 2 and 3 in two schools, and seeks to evaluation the following research questions:

1.) How does structured programme technology such as the BFI interact with the components of South Africa's educational system as outlined in the literature review section?

2.) What observed classroom, school and grade factors likely contributed to gains and retention of literacy skills?

3.) What impact does the Northern Sotho BFI program have in grade 2 and 3 on the following literacy skills?
    a. Letter-sound recognition
    b. Decoding of simple words
    c. Reading speed
    d. Comprehension

[5] BFI was developed and trialled under the auspices of an All Children Reading grant to the Molteno Institute for Language and Literacy in partnership with the International Literacy Institute at the University of Pennsylvania, in a project which ran from October 2013 to June 2015 in 50 schools in the Limpopo Province of South Africa.

[6] The BFI is both a computer software program and a literacy intervention programme. The South African spelling conventions are utilized to differentiate between the software and the content.

[7] 'Northern Sotho first language speakers' is used to describe learners whose mother speaks the language of Northern Sotho. This is distinct from the South African Curriculum Assessment Policy Statements (2011) use of 'home language'. 'Home language' is meant to refer to the first language of learners. However in practice 'home language' refers to the language of teaching and learning in the foundation phase of the school. Schools refer to all learners as 'Sepedi home language learners', regardless of whether the language they speak at home is Northern Sotho. This is in order to avoid creating and finding additional teachers for classes in other languages, as the national policies guarantee learners the right to learn in their 'home language'.

4.) Is there an optimal age for introducing the BFI program? Is it more effective with older or younger learners?

5.) What is the extent of skills uptake from the BFI program, and to what extent are skills gained during the use of the BFI program sustained after use has ended? What types of knowledge, procedural or fact-based, are sustained?

6.) Is there a pattern to retention and non-retention in each task and subtask? In other words, was it possible to predict learners who would and would not retain knowledge?

7.) For tasks included in both tests, does improvement on the Early Grade Reading Assessment (EGRA) correlate in any way to the retention rate?

The following section contextualizes the relevant linguistic, educational and technological literature which informed the design of the research questions.

1.2 The Acquisition of Literacy

The elements of language and literacy development established by research include oral language development, or the development of ability to communicate using adequate vocabulary and understanding of the morphological principles of a language; print concepts, such as awareness of direction of print and sequence of print lines; phonological and phonemic awareness, or the understanding that words are made up of sounds and that sounds can be manipulated to form different words; the alphabetic principle and phonics, or a strong understanding that letters are linked to sounds which can be built up into words; vocabulary; fluency; comprehension; and metalinguistic strategies used to interpret, synthesize, manipulate or otherwise make additional meaning of text (Konza, 2011; Wolf, 2008; Pence & Justice, 2008; Heilman, 2006; Clay, 1991).

While it is arguably apparent that elements such as oral language and print concepts are influential in all languages, the roles of phonemic awareness, phonics, word recognition and comprehension strategies have not been clarified by significant research in the indigenous languages of South Africa, and common literacy practice is based largely upon research performed in industrialized, advanced education systems operating in Germanic or Romance languages. Even among these languages, significant differences have arisen in linguistic research. For example, research has indicated different normal patterns of literacy acquisition for different languages, depending on the opacity of the language (for a review, see Bar-Kochva & Breznitz, 2012). In South Africa, at least three separate studies have suggested that conclusions drawn from English research are not compatible with Bantu languages. An investigation by Taylor et al. (2008) found that English fluency benchmarks were far above the average achievement of Bantu language learners. Diemer et al. (2016) suggest that the established English hierarchy of difficulty for phonological awareness does not match the hierarchy of difficulty in isiXhosa. Finally, Land (2015) reports on an eye tracking study which showed differentials in patterns of eye movement while reading between English and isiZulu readers.

Thus, the present theories of reading, such as the importance of phonemic instruction, commonly applied in the context of developing and linguistically diverse countries like South Africa may actually not be fully appropriate to the linguistic and sociocultural specificities of the country's context (Pretorius & Mokhwesana, 2009; Perfetti, 2003; Pretorius, 2002). Taking this into account, this section seeks to provide an overview of some of the key

components of early literacy as established by the available literature, and to reflect upon the implications of these components in the rural South African context.

1.2.1 Oral Language Development

Literacy skills and capacities are built upon oral language development that begins at birth (Wolf, 2008). Oral language development has been found to impact both vocabulary and the ability to predict words and the meanings of words in sentences (Clay, 2004), and therefore may have a direct bearing on both basic literacy acquisition and higher order reading strategies such as analysis and synthesis. Supporting this theory, research has shown that academic performance in the early grades is linked to measures of oral language exposure, such as the number of words heard in early childhood, and expressive ability such as vocabulary (Pence & Justice, 2008; Hart & Risley, 2003; Anderson & Nagy, 1995; Nagy, 1988). Additionally, oral language provides the foundations of the learning processes, the ability to respond to stimuli and receive and react to feedback on those responses, as well as metalinguistic strategies such as recall, synthesis and problem-solving (Clay, 1991).

Within the South African context, Wilsenach (2015) studied the receptive vocabulary and early literacy skills of Northern Sotho-speaking learners. Her findings supported the assertions of Clay (2004) and Hart and Risley (2003) in showing that receptive vocabulary was a predictor of early literacy, suggesting that oral language development is an important component to literacy acquisition in Northern Sotho as in other languages.

1.2.2 Print Concepts

Print concepts refers to the basic knowledge children must implicitly or explicitly learn in order to read, such as the point at which the text on a page begins; rules about direction, word sequencing and orthography such as letter orientation and use of spaces; and the hierarchy of information – letters form words which form sentences which are linked together to form paragraphs or stories (Clay, 2005). Children with sufficient exposure to reading and books learn most of this information implicitly, either through watching their parents or through engaging in early literacy activities such as bedtime stories (Wolf, 2008), but the extent to which home literacy activities prepare children for academic life varies significantly and may be linked largely to socio-economic status (Brice-Heath, 1982).

South Africans, particularly those who speak indigenous languages, often refer to their language as an oral language or emphasize the strong oral tradition of the language, and some research even indicates that indigenous language speakers do not necessarily value their

language as a written language (Rudwick, 2008; De Klerk, 2000). Partially as a function of poverty, and partially as a function of publisher preference for the measurable and established markets of English and Afrikaans speakers, there are relatively few materials and books available to indigenous language speakers in their home languages, and home literacy and print materials are not emphasized in many rural African[8] homes (Bikitsha & Katz, 2013). As a result, it is reasonable to expect that a large percentage of South African children begin school with relatively little print exposure.

1.2.3 Phonemic Awareness and Phonics

Phonological awareness refers to a broad ability to focus on sounds of speech regardless of meaning and to play on intonation, rhythm, rhyme and individual sounds (Konza, 2011), and has been described as "the ability to recognize, identify, or manipulate any phonological unit within a word, be it phoneme, rime, or syllable" (Zeigler & Goswami, 2005: 4). Phonemic awareness is a subset of phonological awareness defined as "the knowledge or understanding that speech consists of a series of sounds and that individual words can be divided into phonemes" and "the ability to identify and manipulate these sounds" (Heilman, 2006). Research has shown that ability to recognize and manipulate phonological units – through identification of like sounds in a series of words (e.g, *which word starts with a different sound: cat, car, bus?*), addition (e.g., *what is 'ca' with a /t/ at the end?*), deletion (e.g., *what is 'roll' without the /l/?)*, substitution (e.g., *what is 'tip' with a /d/ in place of the /t/?*), and ability to break down and build up words using individual units of sound – is a good predictor of reading ability, with reading ability defined by speed, accuracy and comprehension (Wagner & Torgesen, 1987; Ehri et al., 2001).

Taken from the oral context to print, phonemic awareness is built upon the alphabetic principal and its converse, that words can be broken into their component letters, and that letters represent phonemes. This ability to 'map' phonemes to graphemes and vice versa is called 'decoding' in the literature. Decoding as a strategy enables an early reader to approach a new or difficult word through its component parts, one phoneme at a time (Wolf, 2008). Through the related process of phonological recoding – making meaning of the strings of phonemes present in a word – beginning readers are able to access the full range of their spoken lexicons through print (Zeigler & Goswami, 2005).

[8] The use of 'African' here is in the context of the terms designated by South African law, which designates 'African' as the racial group formerly classified as 'black' under the apartheid system.

While phonological awareness (specifically, understanding of syllables) is present and measurable in children before schooling age, phonemic awareness has been shown to only develop when children are introduced explicitly to literacy instruction. Studies in both English and Italian which asked children to 'tap' the number of syllables and/or phonemes in a word found that few children before schooling age could complete the phoneme version of the task, while children already engaging in reading instruction could complete the exercise. Both studies also found that completion rates for the syllable task were higher than the phoneme task at all ages (Cossu et al., 1998 and Liberman et al., 1974, cited in Zeigler & Goswami, 2005). Within the South African context, Diemer et al. (2016) studied the phonological awareness of early readers of isiXhosa and also found a strong preference for syllabification, concluding that for most tasks syllable manipulation was easier for learners than phoneme manipulation.

Further, there is evidence that instruction plays a reflective role in determining perception of speech. For example, although there are clear syllables in Japanese, and the language is written mostly in accordance with syllable structure, studies of the examples of exceptions to these general rules (loan words that end in a consonant sound, words with long vowels or geminates, for example) have shown that while pre-literate children show mixed preference for syllabic and moraic representation, adults gravitate towards morae (onset-vowel units) rather than syllables (Inagaki, Hatano & Otake, 2000 and Tamaoka & Terao, 2004, cited in Zeigler & Goswami, 2005). This is an example of how acquiring literacy can affect conceptual representations of phonology in a population.

Early Reading and the Dual Route Cascade Model

Phonemic awareness and decoding/recoding are not intrinsically fully matured human abilities, but are developed through instruction (Wolf, 2008), and even phonological awareness improves when explicitly taught. A number of studies and theorists link phonological awareness or improvements in phonological awareness to explicit instruction (Hogan, 2010; Adams, 1990; Perfetti et al., 1987).

However, teaching phonics and phonological awareness is not the only instructional method for literacy acquisition. There is another school of thought that posits early literacy can be taught more effectively using a whole word approach, that learners should engage reading without relying on phonemic decoding but by identifying the word as a whole in the context of reading passages. The whole word approach to reading emphasizes language in context

and was strongly linked to the progressive education movement in the United States, which sought to move away from teacher-led classrooms focused on rote memorization and skills in isolation (Bomengen, 2010; Lemann, 1997). Some proponents, such as Horace Mann, who went so far as to refer to letters as 'bloodless, ghostly apparitions', and Frank Smith, who in his 1986 book *Insult to Intelligence: The Bureaucratic Invasion of Our Classroom* referred to phonics instruction as 'the enemy', argued that reading was as natural as speaking for children and negated any perceived value in phonics instruction as rote memorization bereft of context or meaning. These strong advocates of whole language strategies emphasize what is known as the *look-say* approach as, if not the only, at least the strongly preferred method of instruction (Kim, 2008; Lemann, 1997; Chall, 1967). Other less vehement proponents of whole language approaches have a slightly softer view, including 'opportunistic' or 'intrinsic' phonics instruction as a derivative of whole language approaches, or phonics as one of several strategies taught to early readers to unlock meaning. Still others proposed that both should be taught, it was just a matter of whether words or letters were taught first (Bomengen, 2010; Schantz & Zimmer, 2005; Chall, 1967).

There is ample evidence to indicate that whole language reading theorists are not wrong, and that both strategies are useful in early reading instruction (Adams, 1990). Certainly in opaque languages, like English, there are a significant number of high-frequency words which do not conform to general phonemic patterns: 'a', 'is', 'light', 'my' and so forth, as well as a number of exception-style spelling rules to be remembered as 'sight words' (Ehri, 2005). In addition, early interpretations of psycholinguistic and eye-tracking studies were used to confirm that readers using word reading approaches or 'chunking' did comprehend more (for a summary of studies, see Rayner & Sereno 1994), and it is demonstrably possible to establish literacy based on a non-phonemic reading approach: readers of Chinese do not activate the temporal lobe at all when they read, which indicates that no phonetic assignment or interpretation is occurring during the Chinese literacy event (Wolf, 2008).

However, with the improvement in quality and access of brain imaging studies, new evidence regarding the underlying processes of observed psycholinguistic events became available (Ahlsén, 2006), supporting evidence for the conclusions of Chall (1967) and Adams (1991) that both phonemic and whole-word strategies are integral parts of the reading brain in English. Perhaps most notably, based on neuroimaging Coltheart et al. (2001) proposed a dual-route neurophysiological model of literacy, in which unknown or unfamiliar words present a lateral pattern of neural activation: a slower 'decoding' route that moves from the

occipital lobe through what is believed to be a grapheme-phoneme assignment centre in the temporal lobe. The well-known or 'automatic' words follow a dorsal 'fast track' straight to Broca's area in the frontal lobe, which is associated with language production, without passing through the phonemic assignment centres. The dual-route cascade model provides a model that supports a phonemic approach to early literacy as unfamiliar words must be processed a number of times through the slower decoding route before a reader achieves 'automaticity' and access to the fast-track for those particular words (Wolf, 2008). However, it must be acknowledged that while the phonemic approach may come first for unfamiliar words, whole-word reading, or even whole-phrase reading as an approach is built into the model and is common practice for more fluent readers.

According to Zeigler and Goswami (2005), the developmental progression from large to small units in the 'phonological domain' is well established. However, they note that even from the prerequisite stages of phonological awareness differences can be observed in children learning to read in different languages. For example, children show higher levels of syllable awareness in languages with simple syllable structures (consonant-vowel, vowel-consonant, vowel and consonant-vowel-consonant) than in languages with more complex structures. Zeigler and Goswami further reflect that relatively high levels of phonemic awareness in pre-reading children in Turkish may be the result of the morphology of the language, which demonstrates meaning change by shifts at the phoneme level.

Similarly, a growing body of evidence supports the conclusion that reading in a Bantu language is not the same as reading in English (Diemer et al, 2016; Probert & De Vos, 2016; Land, 2015; Van Rooy & Pretorius, 2013). This is not completely unexpected, as studies show differences in reading patterns between English and European languages with transparent orthographies (Ziegler & Goswami, 2005).

In addition to the level of orthographic transparency, Zeigler and Goswami (2005) assign two other variables which explain differences in literacy acquisition: granularity or grain size of the reading strategies used for phonemic recoding of graphemes, and literacy pedagogy. To provide evidence of the importance of grain size, and its link to orthography, Zeigler and Goswami provide examples of studies which indicate that more skilled readers in less transparent orthographies, such as English, use larger grain size strategies (i.e., basing pronunciation of a non-word such as 'dalk' on a rhyming word 'talk' rather than assigning individual phonemes to each grapheme), while in more transparent orthographies a smaller grain size (usually at the phoneme level in alphabetic scripts) is most efficient. Regarding

pedagogy, Zeigler and Goswami present evidence (for example, Walton & Feyton 2001; National Reading Panel, 2000; Chall, 1967) that using phonemic methodology results in better reading outcomes than whole word strategies previously discussed, and that there seem to be some benefits in some languages to incorporating a large-grain strategy such as onset-rhyme comparisons over strict phoneme-grapheme assignment strategies.

Although no published studies have been done to date examining the patterns of reading in Northern Sotho, some eye-tracking studies have been conducted in isiZulu, which have been compared to English outcomes. Due to the differences in the orthography of the two languages, the results of these studies cannot be directly extrapolated and assumed to apply similarly to Northern Sotho. However, they raise important considerations, establishing that reading patterns in languages such as isiXhosa and isiZulu cannot be equated to reading patterns in English, and providing evidence supporting that reading strategies for these languages predominantly utilize a smaller grain size as proposed by Zeigler and Goswami (2005).

Eye-tracking is a method of research which measures the movement of the reader's eye as it moves across text. The profile of English readers is well-researched, with averages and ranges established for a number of eye-tracking variables, such as saccades, or sudden forward movement in the text, essentially 'skipping' a number of letters or even words; regressions, or backward movements of the eye to previously encountered parts of the text; fixations, or incidences in which the eye pauses on a portion of text longer than others; and re-fixations, in which the eye leaves a word and then returns and fixes on it (Rayner, 2009).

In an eye-tracking reading study of grade 4 isiZulu speakers, Van Rooy and Pretorius (2013) concluded that learners read faster in English than in isiZulu, though with low comprehension in both languages. They also found learners exhibited more and longer fixations and refixations in isiZulu, and in fact found differences in nearly all the measured eye-tracking variables. Interestingly, they also found that difference in reading performance measures did not correlate strongly to difference in eye movements, although they did for learner performance in English.

Land (2015, 2016) performed similar evaluations on adults who classified themselves as fluent readers of isiZulu, and created a comparison of English and isiZulu reader profiles:

Table 1: Comparison of Eye-Tracking Statistics in English and isiZulu, adapted from Land, 2015

	English (all citations in Land, 2015)	**isiZulu** (as reported by Land, 2015)
Average reading speed	300 wpm	105.7 wpm
Fixations per second	5 (Reichle et al., 2003)	3.6
Avg length of fixation	.2 - .25 (Rayner, 2009; Hutzler et al., 2008)	.3
Avg saccade length	7 – 9 characters (Rayner, 2009)	4.05 characters
% of fixations which were regressions	10-15 (Rayner, 2009)	16
Rate of regression	1 per 2 seconds (Paulson, 2005, p. 342; Reichle et al., 2003:348; Rayner, 2009)	1 per 1.89 seconds

Further to the statistical data, Land engaged participants in conversation about the results of eye-tracking. Participants often noted ambiguities leading to regression or fixations. Combining this information with the patterns observed in the quantitative data led Land to conclude that "reading processes that characterise efficient reading in an orthography such as that of English may not be effective in another, such as isiZulu" (2016, pp 151).

1.2.4 Fluency

Early readers rely heavily on the slower phonemic decoding route (Coltheart et al., 2001), which is theorized to compromise comprehension processes due to the limited short-term processing power of the human brain, which research has indicated can only maintain 7 items of information for about 12 seconds (Aadzi, 2008). If those 7 items are made up of individual phonemes, a learner may manage individual words but is unlikely to comprehend the meaning of a full sentence, let alone consider the implications of what is being read or draw in information from additional sources. However, it can be theorized that once the neurological 'fast-track' is accessed, a learner is able to expend less processing power on individual phonemes, and even words. Eye-tracking studies have shown that readers' attention jumps forwards over sections of text, in what are known as saccades. These studies give indications that readers do not actually engage every word in a sentence, but rather 'chunk' words, phrases and even whole sentences into single units of information, which frees up additional processing power for complex operations like comprehension, synthesis and abstraction (Rayner & Sereno, 1994; Rayner, 1975).

A reader who has reached the level of literacy at which they chunk text has moved beyond the emergent stages of literacy to become a fluent reader. Fluency is one demonstration of a reader's mastery of basic phonemic assignment and decoding (Wolf, 2008; LaBerge & Samuels, 1974). Schreiber (1991) identifies three dimensions of fluency: accuracy, automatic processing (or automaticity) and prosody. Accuracy refers to the rate of errors. A fluent reader of English should be able to read a grade-level text with over 90% accuracy (Fountas & Pinnell, 1996). Automaticity refers to the rate of reading, often expressed in words per minute (wpm) (Wolf, 2008). Prosody refers to a reader's ability to enhance text meaning by 'chunking' the text into semantically and syntactically appropriate units while reading, and adding appropriate intonation to match clause boundaries (Kuhn & Stahl, 2003; Schreiber, 1991). Demonstrated ability to read fluently is presumed to be the tangible output of the brain having achieved automaticity (Wolf, 2008; Ahlsén, 2006). However, prosody can be a problematic element to include in evaluating young children's reading ability. Prior research has noted that prosody is poorly-developed in young children (Cutler & Swinney, 1987), and for readers of any age, it is difficult to ascertain the direction and strength of the link between prosody and comprehension. Some research shows prosody to be a poor predictor of comprehension, particularly when reading speed is controlled for (Lopes et al., 2015; Karlin, 1985), and there is evidence that rather than predicting reading ability, prosodic reading is a reflection of high reading capacity (Chafe, 1988).

Although similar benchmarks for reading speed are set for grade 2 and 3 learners in Northern Sotho and English (NEEDU, 2012), both Van Roy and Pretorius (2013) and Land (2015, 2016) noted slower reading speeds in orthographically conjunctive Bantu languages under study than in English for the same participants. While Sesotho is written disjunctively, the same types of discrepancies were noted by Hefer (2013), who observed first-language Sesotho speakers reading subtitles in both English and Sesotho and concluded that Sesotho speakers read subtitles in English faster than subtitles in Sesotho, although they read with greater comprehension in Sesotho. Although intended as a study into the relevance of subtitling in South African media, the results are enticingly familiar when weighed against the studies in isiZulu. Hefer (2013) concluded that a lack of practice in reading Sesotho was the root cause of differentials in reading speed, but the results of other studies in Bantu languages (Land, 2015; Van Rooy & Pretorius, 2013) suggest that lack of practice may not be the primary factor involved. Based on the smaller saccades she observed in isiZulu, Land (2015) theorized that the slower reading rates observed in the language could be due to

engagement with smaller grain sizes, which would be consistent with a more transparent orthography and slower reading speed.

1.2.5 Comprehension

Comprehension is one of the more complex components of literacy acquisition, and has been found to correlate to a number of factors, including socioeconomic status, fluency, vocabulary and IQ (for examples of studies, see Baumann 2005; Jenkins et al., 2003; Kuhn & Stahl, 2003). It is difficult to establish a causal relationship between these factors, which in themselves tend to be interlinked. However, based upon commonly accepted neurolinguistic theories of literacy and working memory, fluency may constitute a key component of comprehension. These theories indicate that the working memory, also referred to as the short-term memory, is one of the key components in both receptive and productive language, as it forms a temporary storage space in which phonemes, words, syllables and sentences can be stored and manipulated. However, the capacity of the working memory is limited. Therefore, a beginning reader who has not yet achieved automaticity, or the ability to automatically link letters and sounds, engages more of the available manipulative space in the simple act of decoding letters into sounds. As automaticity, which would be indicated by aspects of fluency such as reading speed, improves, the brain is able to utilize a "fast track" which eliminates much of the neural engagement previously used in decoding, freeing up working memory for more complex tasks like comprehension (Wolf, 2008; Aadzi, 2008: Baddeley, 2003; LaBerge & Samuels, 1974).

In addition to fluency, metalinguistic strategies also play an important role in comprehension, including the ability to self-monitor comprehension and/or interpretation and apply appropriate problem-solving techniques (Mokhtari & Reichard, 2002). Some research has shown that it is even possible for metalinguistic strategies to compensate for poor phonemic skills. Jackson and Doellinger (2002) found that some students with poor recoding skills showed no significant differences in comprehension from their peers, which may indicate that they developed some metalinguistic coping strategy for comprehension. Research has also demonstrated that it is not necessary for either a first or additional language reader to have prior knowledge of every word presented in a text; in fact, one interpretation of the strong correlation between vocabulary and reading comprehension is that reading comprehension itself is an indicator of the amount of reading done, and reading improves vocabulary – a viable counter to the common interpretation that a more robust vocabulary improves reading comprehension (Baumann, 2005; Anderson & Freebody, 1981).

1.3 Early Literacy Education Policy in South Africa

Internationally, the recommended practices at the intersection of language and content have seen a dramatic shift, with the pre-21st century promotion of a national language of instruction being replaced by advocacy of first language reading instruction, especially in lower grades (Department of Basic Education, 2011b). The Curriculum Assessment Policy Statements (CAPS) curriculum was developed with pedagogy in the foundation phase in the 'home language' of the learners, although in many cases there are learners in each class who do not speak the language of instruction as a first language. The 'first language' education policy is based on research showing stronger literacy skills in both first and second languages for learners taught in their first language (UNESCO, 2008; Williams, 1996).

The current South African curriculum, the Curriculum Assessment Policy Statements (CAPS) is based on best-practice research which incorporates a phonemic approach to literacy in the early grades. The curriculum was rolled out with training to in-service teachers and adopted by Universities for pre-service training in 2012.

The early grade pedagogy suggested by CAPS is strongly influenced by research conducted in English, and advocates a phonemic approach to instruction. In phoneme-based pedagogy in South Africa, combinations of two or more consonant phonemes are referred to as blends, such as 'gl' in 'glad', and orthographically complex phonemes (such as the 'ch' in 'chip') are referred to as digraphs, trigraphs, quadgraphs and so on. In terms of the curriculum, some of these consonant clusters are only introduced in the second or third year of instruction (Department of Basic Education, 2011a), which means that learners in Bantu languages are expected to read before they have completely mastered the phoneme-grapheme relationships in the language. While this is not impossible and could be compared to the later or incidental introduction of "ch", "sh" and "sch" in English, orthographically complex graphemes are relatively more frequent in Bantu languages, and early grade readers do not always consider what clusters have been explicitly taught and which have not, which can complicate the early literacy experiences of South African children (Bikitsha & Katz, 2013).

1.3.1 The Northern Sotho Language and its Pedagogy

Like other Bantu languages in South Africa, Northern Sotho as a language follows a strict consonant-vowel patterning, with a total of 45 phonemes: 7 vowel sounds and 38 consonant sounds. There is only one consonant sound which can end a word, /ŋ/. The language is tonal

and differentiates between aspirated and non-aspirated unvoiced consonants /p/ and /p^h/; /t/ and /t^h/; and /k/ and /k^h/ (Mopida et al., 2010).

The language is written using the 26 letters of the Roman alphabet. The seven vowel sounds are mapped onto the five vowel representations in the Roman alphabet, with two sounds each ascribed to the letters 'i' and 'o', and each of the 39 consonants represented by a letter or unique combination of letters from the 21 consonants of the English alphabet. These comprise both orthographically complex but phonologically simple segments, such as 'hl' and complex segments, which are characterized by "at least two articulator features, with two or more oral tract constrictions occurring simultaneously" (Modipa et al., 2010).

Based on a linguistic analysis of Northern Sotho consonant sounds, Modipa et al. (2010) determined that some complex segments in the Northern Sotho language could be reclassified as simple segments, which would reduce the Northern Sotho phoneme count from 45 to 32. However, this has not been taken up in Northern Sotho pedagogy, and in terms of current pedagogical application a Northern Sotho child must learn 45 phonemes and their corresponding grapheme representations, some of which are orthographically complex (DBE, 2011).

It has been suggested that the Roman alphabet is not the most appropriate vehicle for the written Sesotho word. Besides the consonant clusters often utilized to represent single phonemes, the structure of Northern Sotho would lend itself well to a syllabic script like Japanese hiragana, in which each grapheme represents a consonant-vowel unit rather than a single phoneme (Land, 2015), and in fact many teachers have shown a preference for syllabic teaching despite multiple efforts to engage with early reading through a phonemic approach (Shiohira & Shezi, 2013). The main advantage to moving to a syllabic orthography would be in shortening the words of particularly the Nguni languages such as isiXhosa and isiZulu, which are written conjunctively, meaning that a root word will form a conglomerate with prefixes and suffixes that express or agree to subject, direct and indirect object, tense and so forth. This orthography tends to result in very long words which are especially perplexing to emerging readers.

Northern Sotho, however, is written disjunctively, with the morphological prefixes which could be ascribed individual word units in English written independently from the root word and morphological suffixes, which are joined to the root word. If breaks are assigned based on what constitutes a linguistic unit, the conjunctive orthography is technically correct

(Guthrie, 1970). However, in the disjunctive orthography of Northern Sotho breaks between words have been assigned without consideration for strict word barriers but rather separate prefixes from the root word for the convenience of the reader. The implications for readers are that conjunctive languages such as isiZulu are comprised of fewer, longer words than disjunctive languages, while disjunctive languages such as Northern Sotho are written with breaks in the text which do not always come between words but are often placed within them (Prinsloo, 2009; Louwrens & Poulos, 2006).

In addition to considerations of Northern Sotho orthography as largely – though not completely – transparent and disjunctive, there is a high degree of variation among Bantu languages, with efforts at standardization and lexicography (Prinsloo, 2009). Northern Sotho is no exception, with regional variations in spelling and pronunciation of various words, especially words borrowed from English. For example, during the development of the BFI programme, the English word 'park' was rendered as 'phaka', 'phakha' or 'pakha', sometimes in multiple ways by the same author (Shiohira, 2014). The lack of standard spellings may contribute to difficulties faced by emerging readers as they may be confronted with multiple orthographic representations of the same word.

1.3.2 Curriculum Expectations in the Early Grades

According to CAPS, by the end of grade 1 learners should not only know their letters and the sounds that letters make, but also read words, sentences and short paragraphs of connected sentences. These skills are built upon in grade 2 to include comparing texts with personal experience and reading more difficult passages, as well as the introduction of writing short paragraphs. By the end of grade 3, learners should be reading more difficult texts and explain and support understanding of a text, as well as writing connected paragraphs.

Combrinck et al. (2014) investigated the influence of the introduction of literacy skills on learner performance on an internationally standardized test, the prePIRLS. Combrinck et al. examined the reported grades at which grade 1 level reading strategies were introduced in schools and juxtaposed scores on prePIRLS based on early or late introduction. Skills evaluated included: knowing the letters of the alphabet; letter-sound knowledge; reading words; reading isolated sentences; reading connected text; locating information within a text; and identifying the main idea of a text.

The results showed that while majority of principals indicated that letters, letter-sounds and individual words received emphasis in grade 1 (75 percent) and 56 percent of principals

reported that their schools engaged isolated sentences in grade 1, less than a third of principals reported that grade 1 learners in their schools engaged in reading connected text (31 percent), locating information within a text (23 percent) and identifying the main idea of a text (21 percent) (Combrinck et al., 2014).

1.4 Challenges in Early Literacy Education in Bantu Languages

It is evident from the research by Combrink et al. (2014) that the introduction of basic literacy skills in many rural South African schools is significantly delayed when compared to curriculum targets.

At the same time, there is a ten year record of underperformance in literacy by South African youth when compared to international benchmarks (Howie et al., 2012; Howie et al., 2006). While literacy is an output which has been shown to be related to a number of factors outside the school, such as exposure to oral language (Hart & Risley, 2003), exposure to print in the early years (Brice-Heath, 1982), socio-economic status (Howie et al., 2008) and receptive vocabulary (Wilsenach, 2015; Baumann, 2005), it is also the product of a number of school factors, such as teacher quality (Hanushek et al., 2004) and curriculum coverage (Combrink et al., 2014).

Further, while international best practice cites first language instruction as the most beneficial for literacy achievement in multiple languages (UNESCO, 2008; Williams, 1996), in examining receptive vocabulary's relationship to early literacy skills in foundation phase learners in South Africa, Wilsenach (2015) found that while receptive vocabulary was a predictor of literacy skill, language of instruction was not, leading her to caution that "first language education does not *automatically* guarantee successful attainment of (early) literacy skills". In their examination of South African grade 5 learners taking the PIRLS 2006, Howie et al. (2008) also determined that language of instruction was a primary factor influencing outcomes, regardless of whether learners were attending a school taught in their first language.

These findings call into question the validity of international findings regarding first language instruction within the South African context. Bikitsha and Katz (2013) offer some explanation to these outcomes by noting that efforts to engage learners in first language instruction have been hampered in South Africa by a lack of quality and developmentally appropriate learning materials in Bantu languages.

Further complicating the delivery of curriculum and the creation of materials, Diemer et al. (2016) found evidence that the hierarchy of difficulty in phonological awareness in isiXhosa was not identical to the established hierarchy in English. Additionally, a number of studies have suggested that the grain size of reading in Bantu languages is smaller than in English, and noted differentials in reading patterns (Probert & De Vos, 2016; Land, 2016; Land, 2015; Van Rooy & Pretorius, 2013). There is a need for nascent findings to be supplemented by additional research into literacy skills acquisition in Bantu languages to inform both curriculum expectations and pedagogy. At the moment, the only truly clear conclusion is the lack of literacy outcomes in South African early grade classrooms.

Technology-based educational solutions are increasingly appealing due in part to their potential for scope and scale; their ability to address common challenges in access such as high printing costs; and the fact that they can be customized to specific contexts and yet standardized across systems (Castillo et al., 2015). At the same time, implementing a technology-based solution for literacy may be adding a further dimension of risk to improving literacy skills in languages without an extensive research base from which to form literacy programmes.

1.5 Technology in Education

There is little doubt that education through Information Communications Technology (ICT) is possible and a growing trend, particularly with the rise of distance education, online qualifications platforms and Massive Open Online Courses (MOOCs), as well as a variety of ICT-based education offerings for basic education such as Khan Academy, Mosa Mack and e-classroom. The advent and expansion of ICT into education has created opportunities for new pedagogies, increased or improved access to learners and communities, and established new spaces to exploit in the continued quest for quality education in developing contexts, contexts in which learning outcomes are often negatively impacted by large class sizes, high levels of poverty and unemployment, low community literacy levels and low teacher capacity in both content and pedagogy (Shiohira & Shezi, 2013). Additionally, studies have indicated that children prefer ICT-based interactions to textbook reading or traditional classroom instruction (for a review, see Tolani-Brown et al., 2009).

In fact, technology in various forms has been a part of education for most of the last century. Radio, television, computers, the Internet, mobile phones and tablets have each been expected to revolutionize learning processes and the way education functions in society (UNESCO,

2014). The earliest types of technology-based educational materials, first put into practice in the 1920s, are radio-based (Nwaerondu & Thompson, 1987). Television followed, provoking both great excitement for the potential of television as a method of instruction (see, for example, Anderson et al., 1981), and wary scepticism on the impact of this new medium of consumption on American youth (see, for example, Schramm et al., 1961). Experimentation with educational television continues to the present, through 'edutainment' programmes, education access channels, public broadcasting resources, and educational videos created and marketed specifically for schools. However, the past three decades in particular have seen the rise of new types of ICT which, at least in theory, are capable of tapping into the vast and growing resources of the internet: personal computers, cell phones and tablets.

The early emergence of these devices was accompanied by severe limitations in access, with lines drawn both geographically and economically, due in large part to the high costs of technology in terms of infrastructure, hardware, software, and access subscriptions. By October 2000, there were more than 94 million internet hosts in the world, but 95.6% of these hosts were in higher-income OECD countries, with a mere 4.4% located outside of these areas (OECD, 2001). Even in developed countries, the existence of an "information underclass" was undisputed (NTAI, 1999), with lines generally reflecting existing disparities in gender, ethnicity, income and location (Doczi, 2000; Wolf & MacKinnon, 2002). Thus the "digital divide" existed on two levels: first, between pre-existing divides in national demographics; and second, between nations with different development trajectories.

In OECD countries, the national demographic aspect, or the "first digital divide", has largely been resolved. In the 2012 PISA supplementary ICT survey, 96% of 15 year olds reported having access to a computer at home, and 72% reported that they used a computer, laptop or tablet at school (OECD, 2015). However, with regard to the second tier, developing nations are unable to begin to match the 96% home access benchmark despite making significant gains in potential access – most notably because of the near-ubiquitous use of cellular telephones and their increasing capacity to digitally connect. In the 2013 general survey of South African households, Statistics South Africa found that although 89.5% of households had access to a phone, only 40.9% of households had even one member with access to the internet. Only 10% of households could access the internet at home, a marginal increase from 9.8% in the 2012 survey. And in comparison to 72% of OECD peers reporting access at schools, in 2005 only 22.6% of schools were reported to use ICT for teaching and learning (Isaacs, 2007), a percentage which showed only a modest increase to 28% in 2015 (eNCA,

2015), a modest increase for ten years. Given these statistics, it is undeniable that access to ICT for teaching and learning both within the education system and independently at home remains a challenge for most learners, and substantial further investment would be necessary on the part of the system, teachers and/or learners to attain universal or near-universal access.

However, such an investment may not be warranted, as the reported outcomes of technology-based interventions are mixed. Despite early fervour regarding the potential of ICTs to increase quality and access to education, some large-scale studies show results ranging from negative to mediocre in both developing and developed contexts. The One Laptop Per Child Initiative faced harsh criticism when rolled out in both Peru (Cristia et al., 2012) and America (Kessler, 2012) for failing to produce any sort of educational gain despite enormous expense. Wainer et al. (2008) conducted an analysis of the 2001 Brazilian Basic Education Evaluation System (SAEB), and found computer use had a negative effect on grades, which was only exacerbated in the lower socio-economic status group. A review by Higgins et al. (2012) found lower levels of attainment for technology-based educational interventions than other 'traditional' intervention types. Further research has shown that enjoyment may not be an indication of meaningful engagement, with some researchers noting that the enthusiasm for games-based education is underlined by a lack of evidence showing educational gains (Vogel et al., 2006), and some research shows that "screen time" is not the most effective platform for instruction by demonstrating elements such as the video deficit, or the greater propensity of children to learn from real-life counterparts than from television (Krcmar et al., 2007). Most recently, the OECD (2015) studied the effects of technology use on education attainment in 15 year olds, and found no substantial outcome of ICT investment on learner performance in reading, maths or science. In fact, while schools with moderate technology use performed slightly better than average, schools with high ICT use performed significantly below average.

However, the significant number of studies and meta-analyses with negative results are juxtaposed with other research indicating that technology can be effective in teaching. Early research into technology in education showed mainly positive results (see for example Schacter 1999; Kulik 1994), and a meta-analysis by WestEd (2002) indicated that technology can effectively teach "basic skills" and improve test scores.

Even for studies with negative results, it is difficult to ascertain the cause of low impact with any certainty, as the ICT in the education field is largely exploratory and cannot be said to rest upon large-scale research in developing contexts, conditions exacerbated by the high cost

of ICT hardware and software and continuous advances in the field, often rendering studies obsolete before they are even completed (UNESCO, 2014). Success factors implicated in WestEd's 2002 research included adequate teacher training, reliance on multiple platforms in addition to technology and integration with the curriculum. A dearth of quality language-based content software to apply may also be a key factor resulting in the lack of learning impact (Hinostroza et al., 2012, cited in UNESCO, 2014). The negative results of the 2015 OECD study on technology in education could indicate that subject teaching through technology is a flawed approach; or, as the study points out, it could indicate that in a majority of cases software programs have not been well-integrated with traditional educational approaches.

Ultimately, little is known with any conclusiveness about the links between actual ICT use and educational outcomes, especially in developing countries, and the field is still considered an evolving one (UNESCO, 2014).

1.6 The BFI: A Study in Technology and Literacy Acquisition

As elucidated by the literature review, recent efforts to improve academic understanding of literacy acquisition in Northern Sotho raise important questions, most notably surrounding the use of current best practice methods of phonemic instruction espoused in the South African curriculum. Given the outcomes and insights of prior research, it is possible that internationally recognized methods are mismatched to the structure of Northern Sotho, and further investigation is necessary to determine if these methods are effective in building the core skills of literacy, including phonemic awareness, fluency and comprehension. The Bridges to the Future Initiative was developed in line with the current proposed best practices noted in the 2015 OECD study and the 2014 UNESCO Review of Mobiles in Education, namely explicit links to the national curriculum, provision of immediate and routine feedback and a design which incorporates continuous revision, with the ultimate goal of using established best practices in literacy pedagogy to improve the rate and extent of literacy acquisition of learners in rural South African communities. It is based in the CAPS curriculum, which is linked to the stages of literacy acquisition outlined by phonics instruction proponents, with specific literacy skills by grade. And yet questions remain regarding how effective technology can be in delivering quality literacy education in a South African language.

The current study seeks to elucidate the BFI program's impact on the literacy skills of early literacy learners in Northern Sotho as well as rates of retention in those skills, and in doing so to fill a number of gaps in the research. First, the study will seek to determine whether literacy skills can be learned from technology to the same extent as they can from traditional pedagogy. Second, the study seeks to determine whether skills gained through technology are retained over time. Third, the study seeks to determine what contextual factors within the schooling system influence literacy skills gain of Northern Sotho learners. Finally, the study will investigate evidence on whether Northern Sotho early readers follow the literacy acquisition patterns established in English outlined in the literature review.

Accordingly, the study will seek to answer the following research questions:

1.) How does structured programme technology such as the BFI interact with the components of South Africa's educational system as outlined in the literature review section?

2.) What observed classroom, school and grade factors likely contributed to gains and retention of literacy skills?

3.) What impact does the Northern Sotho BFI program have in grade 2 and 3 on the following literacy skills?
    a. Letter-sound recognition
    b. Decoding of simple words
    c. Reading speed
    d. Comprehension

4.) Is there an optimal age for introducing the BFI program? Is it more effective with older or younger learners?

5.) What is the extent of skills uptake from the BFI program, and to what extent are skills gained during the use of the BFI program sustained after use has ended? What types of knowledge, procedural or fact-based, are sustained?

6.) Is there a pattern to retention and non-retention in each task and subtask? In other words, was it possible to predict learners who would and would not retain knowledge?

7.) For tasks included in both tests, does improvement on the Early Grade Reading Assessment (EGRA) correlate in any way to the retention rate?

CHAPTER 2: METHODOLOGY

In order to respond to the research questions, a mixed-method evaluation was undertaken in two schools which were provided the Bridges to the Future Initiative South Africa 2 (BFI) tablet program for 16 weeks in the second half of the school year. The first part evaluated independent skills of letter-sound knowledge, word reading, oral reading fluency (as measured by correct words per minute) and comprehension. The second part of the evaluation examined delayed retention of information and types of information retained: implicit (or procedural) and explicit (or fact-based). The final part of the evaluation included classroom observation, teacher interviews and learner feedback about the BFI software and programme.

This chapter first undertakes a review of literature on mixed-method research in order to provide a framework and justification for the overall research design. Section 2 of this chapter discusses the overall methodology and research design, including school selection, implementation instruments and timelines. Section 3 briefly outlines the assessment instruments used. Assessment instruments are discussed in detail in their relevant chapters. The final section of this chapter outlines methodological challenges and limitations to be considered in interpreting the conclusions drawn in this research paper.

This study is a complex evaluation in three parts. This chapter only discusses the overall frameworks of the research design, the implementation tools, assessment tools and challenges and limitations. Chapters 3 through 5 introduce the evaluation instruments and component parts of the research in more detail.

2.1 Literature on Mixed-Method Research Design

Experimental research design in education has a history stemming from the turn of the century and involves a control and treatment group in order to maximise internal validity (Cronbach, 1957). The established 'gold standard' experimental design is the randomized control trial, or RCT, in which participants are randomly assigned to either a treatment or control group and evaluated using a difference in difference approach (Ross & Morrison, 2004). While in theory this is the most valid approach, in practice, in real time, the RCT has been called into question in recent years (Sullivan, 2011; Cartwright, 2007). Cartwright (2007) argues that "the claims of randomized controlled trials (RCTs) to be the gold standard rest on the fact that the ideal RCT is a deductive method…this a feature that RCTs share with a variety of other methods, which thus have equal claim to being a gold standard". Sullivan

(2011) elaborates, stating that "the primary advantage of randomization is that it reduces allocation bias…[but] *randomization will not control for other sources of error, which are likely to occur in education studies*" (emphasis in original). Ultimately, with deeper analysis and the emergence of constructivist views, questions have arisen about the validity of experimental design in education, as such a design makes it difficult to control for all variables, or to ascertain the extent to which variables such as the operating system, the environment and the individuals under study influence outcomes (Jonassen et al., 1994).

Further, there are often ethical concerns involved with RCTs, particularly if the treatment is largely perceived to be beneficial. Finally, randomization can hardly be considered a cure for bias in cases in which the sample size is very small and/or constrained by logistical or technical factors.

In response to these considerations, other types of experimental design have emerged in the research space, including repeated measures design, in which individual participants are given multiple treatments, and quasi-experimental approaches such as the pre and post-testing design and the time series design, which involves repeated measures of a group with a treatment introduced between measures (Ross & Morrison, 2004).

The types of research undertaken in this study – namely the Early Grade Reading Assessment (EGRA), benchmarking research, and retention research – are all based in the realm of experimental or quasi-experimental research designs, which emphasize the elimination of researcher bias through evaluation based in the scientific method and, in general, require researchers to distance themselves from participants of the study and evaluation through the use of testing and numerical assignment (Johnson & Onwuegbuzie, 2004; Nagel, 1986).

However, using experimental or quasi-experimental research design, a study limited to these elements is susceptible to accusations of pure positivism and poor external validity by those in favour of constructivism and related concepts such as humanism and relativism, which are evaluated qualitatively through direct interaction with participants. Advocates of qualitative research believe that validity can and should be established through examination of the multiple realities and lived experiences of participants of an intervention or phenomena (Guba, 1990; Smith, 1983).

A growing movement has recognized the value of combining quantitative and qualitative methodologies to improve validity through triangulation, and research increasingly relies on

both paradigms through the application of mixed-methods research designs (Creswell, 2013; Johnson & Onwuegbuzie, 2004; Olsen, 2004; Creswell, 2003; Tashakkori & Teddlie, 1998; Creswell, 1994). According to Creswell (2013), the crux of mixed-methods research is the collection and *integration* of two databases. It is not sufficient to merely collect both types of data; for a study to be considered mixed-methods, the data collected must be aggregated and examined in total in order to draw conclusions.

Creswell and Plano Clark (2011) further outline the reasons to conduct a mixed methods study. They suggest the mixed methods approach can serve to supplement otherwise insufficient data, explain initial results, enhance a study through the use of multiple models and investigate complex research questions. Types of qualitative data collected may include observations, video, surveys or questionnaires, interviews and reflection sessions (see also Patton, 1990), while quantitative data may include aspects such as test results or descriptive information such as age or gender.

A mixed-methods approach is the most appropriate to answer the research questions outlined in this study. While some research questions can be answered using purely quantitative data (e.g.; What impact does the Northern Sotho BFI program have in grade 2 and 3 on literacy skills?), an analysis of data drawn from qualitative research is likely to help explain initial results for these questions (Creswell & Plano-Clark, 2011). Other research questions directly require a more nuanced contextual understanding (e.g.; What observed classroom, school and grade factors likely contributed to uptake and retention of literacy skills?) and can be classified as complex research questions (Creswell & Plano-Clark, 2011).

Further, no single test instrument was sufficient to answer both the literacy skills questions and the retention questions. The literacy skills questions required an assessment which was not linked to a specific pedagogy, due to the inability of the research plan to control for teacher and teaching styles, while the retention test required an assessment specifically linked to the Bridges to the Future Initiative.

Therefore, the research framework ultimately included three elements, two quantitative and one qualitative. Each element is discussed in detail in the three chapters immediately subsequent, while this chapter provides a contextual overview.

2.2 Research Framework and Design

The research framework ultimately included three components.

A qualitative component was undertaken to provide an overview of and provide comparisons of the contexts in which the intervention programme was used. The qualitative analysis included three aspects: classroom observation, teacher interviews and learner drawings. The findings of this analysis are presented in Chapter 3.

The Early Grade Reading Assessment (EGRA) was undertaken using a longitudinal research design in order to address gains in literacy skills. The longitudinal research design was chosen over a randomized control trial due to a small sample size of only two schools and twelve classes. As pointed out by Sullivan (2011), while randomization eliminates allocation bias, it does not affect other types of bias which may be prominent in education studies. Even if a matched pair design were adopted, a number of factors would be difficult to address, such as teacher qualifications, school resourcing, number of learners per class, parental engagement, and so forth. The longitudinal design allowed each class to act as its own control, mitigating the effects of these numerous variables. The research design and results of the EGRA test are discussed in detail in Chapter 4.

A method-dependent test was designed to answer questions pertaining to retention of knowledge, based on the curriculum skills and factual knowledge covered by the technology intervention under study, the Bridges to the Future Initiative - South Africa 2 (BFI). This test was designed to test the knowledge contained within the pedagogy of the technology intervention rather than overall literacy skills, as the research was interested in retention of knowledge pertaining to program use. The test was administered using a quasi-experimental pre and post-test design. The tool, methodology and results are discussed in Chapter 5.

These three component assessments were carried out between February 2015 and January 2016, as shown in Figure 1.

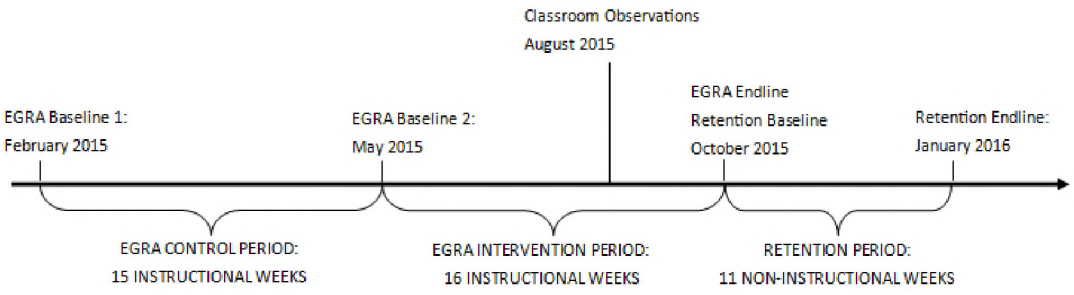


Figure 1: Research Timeline

While ideally the same sample would have been administered all three portions of the evaluation, this was not feasible for three reasons. First, due to the expense and time-consuming nature of the EGRA examination, school requests to limit testing time and the available resources of the research project, only a sample of each school could be administered the EGRA. This sample was drawn from all classes across the school to avoid disrupting class schedules so far as possible. Due to the tight timeline, the retention baseline and EGRA endline had to be administered concurrently, which meant that in any case not all learners taking the EGRA could participate in the retention baseline due to time constraints. A second and related point is that the EGRA itself is taxing to learners and was administered very near the end of term examinations. Asking learners to participate in two taxing examinations on the same day, in the same week as final examinations, could have led to a high degree of error due to exam fatigue. Finally, as EGRA participants were drawn from all classes in the school, the sample for classroom observation could not have included all of the EGRA-tested learners without disrupting class schedules and placing some learners with unfamiliar teachers, which would have impacted the evaluation. Therefore, the samples for the three assessments are different, although drawn from the same two schools.

Ethical clearance for the intervention was provided under the auspices of the Molteno Institute for Language and Literacy with the authorization of the Department of Basic Education and the Limpopo Provincial Board of Education, who signed a Memorandum of Agreement with regard to the programme implementation and accompanying research. In addition, school principals consulted School Governing Bodies and parent meetings before joining the programme. SGBs and parents were given information and a demonstration of the program, and the intended research methods and goals were explained. While ideally parents would have similarly signed consent forms, the principals were concerned by high rates of illiteracy in their communities and opted for oral consent.

2.2.1 Sample Schools

Two schools were included in the study, both in the greater Polokwane area of the Limpopo Province in South Africa. Criteria for selection were: 1) no-fee public ordinary schools under the auspices of the Limpopo Department of Education; 2) a full complement of appointed foundation phase (grade 1-3) teaching staff, foundation phase head of department and principal; 3) permanent infrastructure; 4) using Northern Sotho as a language of instruction in the foundation phase; 5) no prior exposure to technology for learners in the foundation phase; and 6) willing to participate in the study.

Other similarities between the two schools include that both schools use Northern Sotho as the language of teaching and learning up until grade 4, when they switch to English as the language of instruction. Additionally, both schools are classified as high enrolment. However, school B is larger than school A, with nearly twice as many learners enrolled and twice as many teachers per grade. The average class size is similar, however, with classes ranging from 47 to 56 learners per class. Table 2 presents the enrolment and teacher statistics for each school by grade, and the total number of learners and teachers participating in the BFISA2 program in 2015.

Table 2: Sample School Learner, Teacher and Class Numbers

	Grade 2			Grade 3			Total Learners	Total Teachers	Total Classes
	Learners	Teachers	Classes	Learners	Teachers	Classes			
School A	104	2	2	94	2	2	198	4	4
School B	217	4	4	205	4	4	422	8	8
Aggregate	321	6	6	299	6	6	620	12	12

Both schools enrol and teach learners only in Northern Sotho, but have learners who speak other languages at home. It was not always easy to classify learners as "first language speakers", as many learners speak multiple languages at home or have parents who speak two different languages. Ultimately, learners were classified as "first language Northern Sotho speakers" if they responded that they spoke Northern Sotho at home with their families at least most of the time. Only learners meeting this criteria were included in the EGRA and retention portions of the study. Classroom observations were based on full classes so learners of all backgrounds were included.

Differences between the two schools include proximity to a major urban centre, with School B located just 15 minutes outside of Polokwane and therefore classified as "peri-urban" by the Department of Education. School A was located about 1.5 hours from Polokwane and is classified as rural.

Additionally, school B is one of the highest-performing Northern Sotho no-fee schools in the Province in the foundation phase. School A, on the other hand, is moderate-to-low performing in the foundation phase.

The selection of these two schools was intentional. Due to the mix of high and low performance a range of learner abilities was likely to emerge, which would enable analysis of programme effects on learners of varying baseline abilities. A second consideration was to determine to what extent program use enabled a lower-performing school to advance as compared to a higher-performing school.

2.3 Implementation Instruments

2.3.1 Tablets

One school was provided with Nexus 7 tablets, and the other with RCT-800 tablets. The main considerations in choosing these two tablets were ability to run the program, screen size and cost. Both tablets can run the program, although the Nexus 7 has slightly less lag time. A screen size between 7 and 8 inches was chosen as the tablets were then light and small enough for learners under 10 to easily handle. And although the screen resolution and sound quality is better on the Nexus 7, it is still clear on the RCT 800 for about one-third of the cost.

The main differences between the two tablets are outlined in Table 3.

Table 3: Comparison of Tablet Instruments Used

	Nexus 7	**RCT-800**
RAM	2 GB	1 GB
Processor	1.5 GHz	1.3 GHz
Screen Resolution	1920 x 1200	1280 x 800
Screen size	7”	8”
Battery Life (observed)	+/- 8 hrs use	+/- 6 hrs use
Cost	R5,000/tablet	R1,800/tablet

The battery life on both tablets is excellent, with schools reporting that charging was only necessary once a week – a relevant consideration given that South Africa experienced regular power failures in all areas in 2015.

A total of 27 Nexus 7 tablets were procured for use in school B, while 50 RCT 800 tablets were procured for school A. In both schools teachers opted to ‘split’ the tablets between grades, so 13 were allocated to each class in school B and 25 to each class in school A. As the schedules would have permitted the use of all tablets in all classes, it is unclear why this decision was taken by the schools.

2.3.2 The Bridges to the Future Initiative Tablet Program[9]

As discussed in the literature review, the Bridges to the Future Initiative program is an early literacy training programme implemented through ICT. It is based in the phonemic approach to literacy built into interactive software.

The first version of the BFISA program[10] was targeted towards ABET (Adult Basic Education and Training) early literacy learners. After a successful trial with grade 1 to 3 learners, a review and revision of the ABET program and its conceptual framework was undertaken and 40 lessons for learners in grades 1-3 were developed in four languages: English, Northern Sotho, Xitsonga and Tshivenda. As an additional exploratory measure, 10 lessons in each language were also adapted for tablet use and rolled out in seven schools for periods of two weeks to two months. Observations of learners using the tablets were made in both home and school environments, but the tablet program was not evaluated for impact. In October 2014, an additional grant from UNESCO was procured to adapt 10 more PC lessons to the tablet. The complete tablet program therefore has 20 lessons, 10 for grade 2 learners and 10 for grade 3 in each of the four PC program languages.

BFISA2 Tablet Programme Frameworks

There are six aspects to the foundational framework of the BFISA2 tablet programme.

First, content is contextualized to the target audience: rural grade 2 and 3 literacy learners. Animations, stories and content utilize common elements from this environment, including minibus taxis, playing soccer and school gardens[11].

Second, instructional content is provided through the tablet medium, including exercises and, when necessary, revision and instruction visually and/or aurally.

Third, every effort was made to make the program intuitive and easy to use. Necessary elements of the program such as keyboards 'pop up' automatically, and a pointing finger indicates when learners need to take an action such as tapping the 'next' button. Drag and

---

[9] This section on the development and contents of the BFISA2 programme is taken largely from the report on the PC-based BFI project written by Castello, Shiohira, Kochon, Nakani and Zapata (2015) for the United States Agency for International Development (USAID).

[10] As an ICT-based literacy intervention, the BFISA is both computer software and a literacy training course. As per South African English conventions, 'program' is used to refer to the actual software, while 'programme' is used to refer to the course and its contents.

[11] School gardens are one method commonly used in rural South African schools to supplement the school nutrition programme, which provides free meals to learners in government schools in low-income areas.

drop and tapping activities were, after a pilot test, desensitized so learners who were not exactly precise with their fine motor skills were still granted a correct answer.

Fourth, the programme was envisaged to be self-paced, or used in pairs which would pace themselves. Learners can return to previous activities at will, although they cannot skip forward through activities which have not been completed.

Fifth, to empower teachers, the programme is aligned to the Curriculum Assessment Policy Statements released for foundation phase home language and supports classroom teaching. The programme is designed for use in conjunction with class activities to reinforce teacher-based instruction.

Finally, the program provides corrective feedback after incorrect answers. Thus, although the programme is structured for revision, a learner is able to access instructional material as well in areas of challenge or difficulty.

The Development of BFISA2

The approach to the development of BFI-SA2 emphasized independent development in each language. Content development for the Northern Sotho programme was completed in conjunction with African language experts and guidance by the Pan South African Language Board (PANSALB)[12]. The approach to development taken is demonstrated in Figure 2.

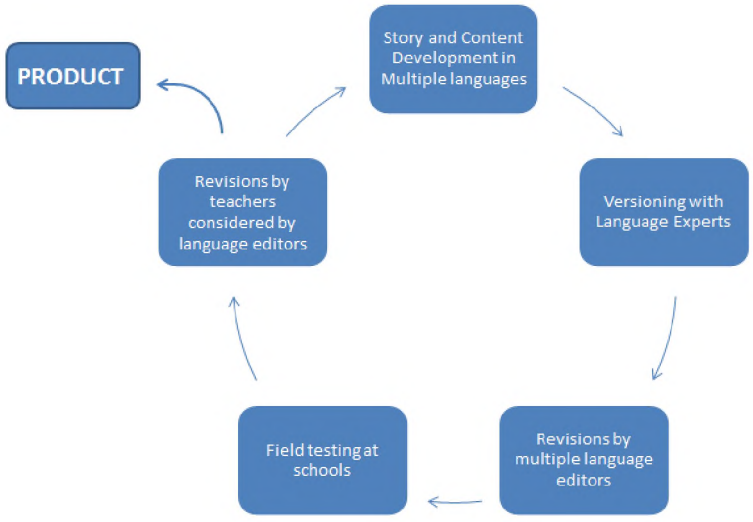


Figure 2: BFI Development Process

Reference: Castello et al, 2015

[12] PANSALB is a South African organization established by Parliament in 1995 to develop South African languages and promote multilingualism in South Africa. http://www.pansalb.org/

In all, each language was reviewed by a minimum of five individuals before content was published. In the process of this, conflicts in acceptable use of the languages commonly emerged, generally influenced by regional dialects of usage of the language. In these cases, PANSALB experts were consulted. In instances when the conflict could not be resolved by PANSALB, a decision was taken by the senior editor in each language.

Components of the BFISA2 Programme

The BFISA2 programme is based on a constructive approach to comprehension. This means that each activity leading up to the comprehension passages introduces key words, sentences, grammatical structures and punctuation that the learner will encounter in the comprehension passage. In this way, supports are provided before the passage is engaged.

The basic structure provides an audio-visual component as background information, which is built upon by independent activities presenting individual words, sentences, morphological items and punctuation, which are combined in the comprehension passage. Thus by the time a learner reaches the comprehension passage and activities, they will have already encountered key elements of that passage which can support their comprehension.

Figure 3: Screenshot of a BFI Audio-Visual Component Frame

Phonemic awareness and phonics

The program contains several exercises which focus on identification of sounds, including phoneme identification and word building through deconstructing and reconstructing key words into phonemes and syllables, identification of words with the same first and last sound as a key word or review key word, and building of words from component graphemes with aural and/or visual prompting. In addition a spelling/typing game and a balloon popping game reinforce word-building from phonemes.

Grammar and punctuation

According to the CAPS home language documents, by the end of grade 3 learners should be able to correctly use punctuation marks including the full stop, exclamation point, question mark and comma. They should also have knowledge of basic parts of speech and the correct use of common conjunctions. The BFI programme introduces all of these elements through decision-based activities, such as clicking the correct part of speech or drag-and-drop the correct punctuation mark. If the learner makes an error, instructive feedback is given explaining the parts of speech or punctuation marks. Learners are then invited to try again.

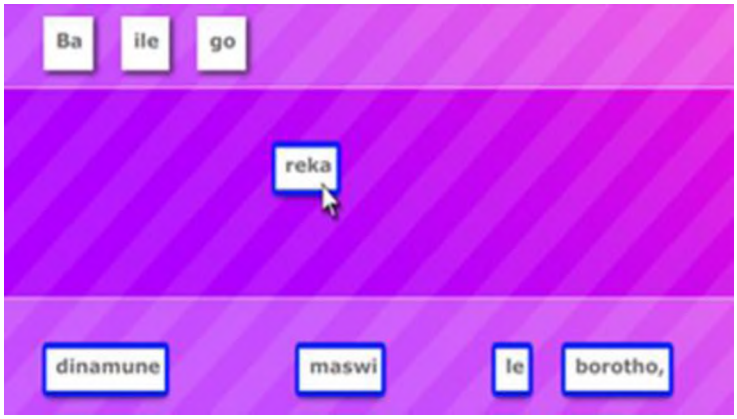
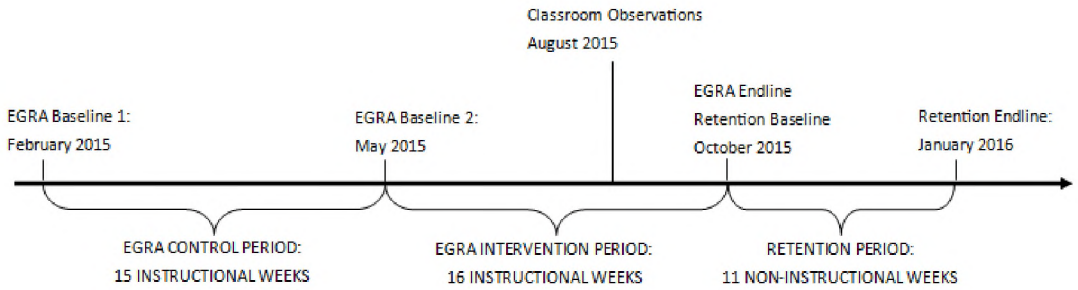


Figure 4: Screenshot of a BFI Sentence Building Activity

Comprehension

Each lesson includes a short story for reading practice and three or four accompanying comprehension questions. Comprehension passages and questions develop in complexity over the programme and are based on the five types of comprehension outlined in Barrett's taxonomy of comprehension skills (Clymer 1968): literal/direct, which refer back to explicit statements in the passage (e.g.; Where did Peter take the books?); organization/reorganization, requiring the learner to classify or clarify the structure of a text through synthesis, summary or description of genre (e.g.; What characteristic do frogs and spiders share?); inference, or questions requiring learners to apply their own knowledge or information from an additional source to material in the text (e.g.; What is the meaning of 'ancient'?); evaluative, or questions asking learners to make judgements about appropriateness or logical reasoning, such as giving reasons for the behavior of characters (e.g.; Why did Peter run away when his friend was hurt?); and appreciative, or questions which ask learners to make judgments about text or occurrences and empathize with characters (e.g.; How did Peter feel when he was playing soccer with his friends?).

The BFI program was implemented over 16 instructional weeks from May to October, after a control period of 15 instructional weeks between February and May. School holidays were omitted from instructional week counts (see Figure 1: Research Timeline, copied below).



2.4 Assessment Tools

While a summary of the assessment tools used for each of the three components is outlined here, each tool will be discussed in detail in their relevant chapters[13].

Qualitative data was collected through classroom observation, teacher interviews and learner drawings. This is discussed in detail in the following chapter, Chapter 3.

To assess the question of gains in the literacy skills under study, the Early Grade Reading Assessment (EGRA) was redesigned in Northern Sotho to cover the relevant skill set of grapheme-phoneme assignment (letter reading), simple word decoding, passage reading and comprehension. This provided knowledge on grapheme knowledge, decoding ability, fluency and comprehension ability. As the test is method-independent, it provided information on literacy skills acquisition without direct links to curriculum components. The EGRA was applied using a longitudinal research framework, a time-series design. The test and its design and application are presented in detail in Chapter 4.

To address the retention test, a paper-based pre and post-test was developed which covered the curriculum skills taught by the BFI programme. This test was specifically aligned to the curriculum and to the programme, and therefore measured the uptake of curriculum skills at the end of the year as well as retention of those skills over 11 non-instructional weeks. The test design and implementation is discussed in detail in Chapter 5.

[13] The work for this thesis was carried out under the auspices of the Molteno Institute for Language and Literacy. Therefore, all tests and products developed remain the intellectual property of the Molteno Institute for Language and Literacy and cannot be included in full in this thesis or as appendices. However, examples of the types of questions and test format will be provided in the relevant chapters.

2.5 Challenges and Limitations to the Study Design

While efforts were made to mitigate as many challenges as possible in the study design and implementation, due to the scope of work, logistics and the nature of research in education settings, there are some limitations which must be noted.

Regarding Sample Size

The sample size of two schools was determined by the number of schools with access to the tablet programme, and was therefore constrained by logistical considerations. The number of learners tested and classrooms observed was also further constrained by logistical and resource considerations. As noted in the methodology section, it is not possible to extrapolate representativeness of the general population of Northern Sotho-speaking South African foundation phase learners in either grade 2 or 3 based upon the results.

Regarding Time Series Research Design

Ross and Morrison (2004) note a dearth of time series designs in current research literature, and attribute it tentatively to the fact that researchers are hesitant to subject participants to so many repeated measures. This may be one factor, but the general acceptability of RCTs as the more internally valid design most likely weighs more heavily. A necessary caveat of any quasi-experimental design is that any difference noted cannot be attributed with certainty to the intervention rather than other unobserved factors (Ross & Morrison, 2004), and education researchers are hesitant to use designs in which their conclusions must necessarily be labelled as tentative, exploratory or indicative.

However, in education practice the time-series design can, if well designed, mitigate many of the variables noted by Jonassen et al. (1994) as problematic in RCTs. In particular, a time series design enables a researcher to work with a relatively small sample, such as the two schools available here, with higher validity than an RCT may offer. In the time series design, each individual participant acts as their own control, which mitigates highly influential variables such as teacher quality, individual capacity, extracurricular training or support and socio-economic status. And while time series studies are subject to validity threats such as maturation of participants, testing and statistical regression (Ross & Morrison, 2004), RCTs when selected from a small whole-school sample would be just as vulnerable to these validity threats (for example, learners in one school may be given a test while learners in another are not, or learners in schools may have different average or modal ages in the same grade, which

would affect maturation), in addition to having a potential selection bias as well. Even if differences in baseline ability are accounted for, this only indicates that the development trajectories of the two populations are predisposed to difference, but not to what extent, therefore throwing doubts upon the results of a difference in difference evaluation.

For these reasons, the time series design was chosen for the EGRA research component. To help mitigate concerns of potential bias, qualitative elements were engaged and triangulated with quantitative results in order to improve validity to the furthest extent possible. However, ultimately results must be weighed against the limitations inherent in the research design.

Regarding Delayed Retention Testing

As program use was discontinued after the first performance test, only direct effects of testing must be taken into account in the current study as a potential source of error. As testing has been shown to improve retention, it is possible that the action of taking the initial retention test itself elevated performance on the delayed retention test. However, while there seems to be universal agreement on the existence of a testing effect, the present study is concerned with two factors: first, the initial performance as a measure of skills uptake, and second, the retention of the performance of those skills over time, however they are acquired. The comparison of performance and retention using control groups with no program interaction is an area for further research, to determine whether retention effects are higher for control groups with traditional instruction or for groups that incorporate the BFI program as an educational tool.

Further, defining implicit and explicit knowledge is imperfect, leading to assumptions about the type of processing used by learners. For example, grammatical knowledge may be accessed through the implicit memory or explicit memory, depending on how automatized procedures are for each learner. With the research design of this study, it is impossible to differentiate within some categories between learners using implicit or explicit memory, which must be considered in interpreting the results of this study.

Regarding Hours of Program Use

The measure of 'time on task' for this study was hours of program use. This was chosen as a measure because it did not cause significant strain on the teachers and was not likely to result in non-compliance in recording use of the program. Additionally, learners were too young and not literate enough to be expected to keep track of their own program use.

However, a challenge emerges in that hours of program use does not translate equally across all learners and classes. Even in the single sessions observed, some learners and groups were witnessed finishing up to four lessons, while other learners or groups were unable to complete even one lesson. Therefore, although some learners may have only used the program for three hours, it is possible they completed the ten lessons for their grade level.

The end result is that it remains a possibility that a more rigorous measure of usage, such as completion percentage, would reveal a different relationship between use and gains.

CHAPTER 3: QUALITATIVE ANALYSIS

A qualitative analysis was undertaken in order to better understand the context in which teachers were working with and using the Bridges to the Future Initiative (BFI) in their classrooms. Qualitative analysis was guided by a specific objective, understanding how the BFI programme was utilized within the context of the South African education system. This section seeks to answer the question:

1) How does structured programme technology such as the BFI interact with the components of South Africa's traditional educational system as outlined in the literature review section?

In addition to being used to investigate this complex question, in Chapter 6 qualitative elements will be integrated with and analysed in conjunction with the results from the EGRA and retention tests in order to explain the quantitative results and add external validity to the study model (Creswell & Plano Clark, 2011). Specifically, this section seeks to address the following research question:

1) What observed classroom, school and grade factors likely contributed to gains and retention of literacy skills?

Three qualitative techniques were utilized: Classroom observation, teacher interviews and prompted drawings from learners. A summary of relevant findings is presented in this chapter.

3.1 Qualitative Research Design

In the field of education, a mixed-methods approach is often employed due not only to the challenges inherent in working across contexts which are highly variable in both operating system, environment and individual contexts (Jonassen et al., 1994), but also due to the structure of education and the existence of a so-called "black box" – the classroom practices and environment in which educational inputs such as resources or training are enhanced, diminished, or otherwise transformed before they can be evaluated as outputs or outcomes (Black & William, 2001). When used in a mixed-methods study, qualitative data can elucidate the teaching and learning practices and environment which contribute to quantitative test results. This study summarizes the findings of qualitative data gathered on observed classroom behaviour as well as teacher and learner perceptions of the program in education.

Data for this portion of the evaluation was collected in August. Prior to intervention rollout, teachers were provided with two days of training on the BFI program itself, which included an overview of the purpose and frameworks of the programme, a walk-through of a lesson to explain the activities and technical expectations of learners, the expected hours of use and suggested strategies for linking to the curriculum. However, teachers were not given explicit instructions on how to use the tablet program with their learners, as part of the goal of the research was to evaluate how teachers make use of technology.

Qualitative research was undertaken by two researchers. Both researchers were trained in qualitative research methodology by the Graduate School of Education at the University of Pennsylvania, and both had prior experience and collaborated on published qualitative research in a variety of countries and educational contexts (including Japan, America, Peru and South Africa). Techniques previously employed by the researchers included interviews, surveys, discourse analysis, and analysis of pictographic representations (drawings) by vulnerable populations.

Data was gathered through classroom observation, teacher interviews and learner drawings. Two classrooms per school were observed by two researchers for between 30 minutes and 1 hour as they worked with the tablets. The classrooms were selected based on a convenience sample of which classes were scheduled to use the tablets on the day of the observation visit. Two classrooms were observed working in English, and two in Northern Sotho. The total number of learners observed was 150, as outlined by Table 4: **Number of Classroom**

Observations. 97% of learners present in classroom observations submitted drawing of the program (45 out of 50), and a total of three class teachers and one principal were interviewed.

Table 4: Number of Classroom Observations

	NORTHERN SOTHO		ENGLISH		TOTAL	
	Classes	Learners	Classes	Learners	Classes	Learners
GRADE 2	1	24	2	101	3	125
GRADE 3	1	25	0	--	1	25
TOTAL	2	59	2	101	4	150

Data was analysed through general inductive analysis. General inductive analysis is a systematic procedure appropriate when the analysis is guided by objectives and aims to summarize various and extensive raw data, used to connect the summary findings of various data to the objectives of the research (Thomas, 2003). Inductive analysis was therefore the most appropriate qualitative approach to this study, as a variety of data was collected (recordings, drawings and field notes) through different means (interviews, prompted drawings and classroom observation) with a specific research objective (to determine how the BFI interacts with the components of the South African education system).

3.1.1 Classroom Observation

Classroom observation was loosely structured with two researchers recording detailed field notes independently in each classroom. In total four classrooms were observed: a grade 2 and a grade 3 classroom in school A and two grade 2 classrooms in school B. The grade 3 of school B was undergoing testing on the day of the visit, and also had not yet used the program so was omitted from the qualitative analysis.

The researchers positioned themselves at opposite sides of each classroom and sat with notebooks. Notes were taken by both researchers on observed teacher and learner interactions with each other and with the technology during regular class time. Periodically, researchers would circumnavigate the room to better identify learner actions with the tablets. Researchers did not engage learners, but did respond when engaged by the learners.

Notes from the two researchers were typed and aggregated for each classroom observed at the end of each day. The compiled data was coded following procedures introduced by Creswell (2003), which were reviewed and added to after each submission of new data. Coding was

completed independently by each of the two researchers and codes and emerging themes noted by both researchers were included in a final framework for analysis.

3.1.2 Teacher Interviews

Interviewees were drawn from school staff currently involved in the BFI program. All interviews were reviewed and notes taken in the initial interviews were added to so that all utterances by participants were represented. The researcher analysed the coded data in accordance with the guidelines established by Creswell (2003), grouping codes to create themes and categories which present the participants' collective experience of the interactions between learners, teachers and technology in the classroom. This was a continuous process throughout the interviews, and previously coded data was returned to after each interview in order to apply newly perceived codes where applicable until a point of saturation was reached (Creswell, 2003).

The resulting aggregate data was analysed for common themes. Quotations were drawn from the recordings which were representative of participant views on key themes identified. This write-up was then sent to the participants, who were given the opportunity to make corrections, request exclusion of information, or provide additional information.

Confidentiality, permission to record and use and treatment of data concerns were discussed with participants before interviews were conducted. Interviews were open interviews which explored the relationships between the teacher(s) and the BFI programme; the teacher and learners while using the programme; and the programme and the larger school system and system(s) of education including the district office, the curriculum, assessment and relationships with other schools and provinces. Ultimately, a teacher from grade 2 and 3 as well as the principal were interviewed in school A, while a grade 2 teacher and a focus group of all grade 2 and 3 teachers including the head of department (HOD) were interviewed from school B.

Teachers were not able to leave their classrooms, so interviews were conducted while learners were drawing their impressions of the programme[14]. This did lead to some distractions in the interview process as teachers paused to reprimand learners or answer questions. Interviews were paused for such occurrences.

[14] Research on the BFI was conducted with the knowledge and permission of the Limpopo Provincial Board of Education.

The principal of school B was not available on the day of observations, and in fact was rarely at the school, but the principal of school A was interviewed regarding the implementation of the programme in her school. The interview took place in the make-shift principal's office, which was actually the defunct computer room of the school (all computers had ceased functioning by 2013 when the school was first surveyed).

Interviews were tape-recorded with the permission of the participants, and detailed notes of the participants' responses as they were replying to questions were taken. Interviews for teachers lasted between 20 and 30 minutes and covered the following topics:

- *Confidence with tablets and tablet or smartphone ownership*
- *Logistics of tablet use;*
- *Preparation for tablet lessons;*
- *Perceived strengths and weaknesses of the software and programme;*
- *Perceived strengths and weaknesses of learners;*
- *Perceived changes in learner behaviour and aptitude;*
- *Follow-up to program use;*
- *Any additional comments.*

The interview with the principal of School A lasted 15 minutes and covered the following topics:

- *Logistics of tablet use in the school*;
- *Effect of tablet use on school management*;
- *Observed changes in teacher and learner behaviour*;
- *Parent reactions to the tablet program in the school*;
- *District official involvement in the program and the school as a whole*.

In all interviews, the target questions outlined above were augmented by additional questions or areas specific to each participant's experiences.

3.1.3 Prompted Drawings

While qualitative data pertaining to adults is often gained through interview processes or surveys, qualitative research incorporating young children can be particularly problematic. Research on and about children are intellectual spaces historically dominated by adults talking about children, the recognition of which led to widespread acknowledgement of the need for 'children's voices' to enter such research (Eldén, 2012). One current method of engaging young children to fill the need for children's voices is employed in this study -- drawing. As noted by Eldén (2012), drawings offer "a different way of revealing experiences and perspectives while at the same time democratically involving children as 'producers of knowledge'". Drawings are always in some way elicited, but can be either directly or indirectly prompted (Duncan, 2013). A direct prompt includes a topic or question posed to children for the drawing (see for example Harrison et al., 2007; Alerby, 2000), while drawings may be indirectly prompted by a researcher by the provision of materials or an invitation to draw something. As a research methodology, drawing is a participatory methodology of particular use when linguistic competence of the participants is questionable, or there is a language barrier between researchers and participants. Further advantages are that this method mitigates the ethical concerns of working with children, as it "discourages the sense of a hierarchy between researcher and subject" (Literat, 2013, p. 88).

Yet, regardless of intention or medium, in order for communication to take place two individuals must interact: the speaker and the listener (Komulainen, 2007), and ultimately the nature of drawing analysis is hierarchical, with the researcher in a positon of power, interpreting the work of learners (Eldén, 2012). While some researchers attempt to mitigate the interpretive aspects and thus the hierarchical nature of drawing evaluations by engaging children in dialogue while or about their drawings (Literat, 2013), this may not always be possible due to linguistic or resource constraints.

Further, analysis of drawings is not a straightforward process. First, to date no specific framework has been developed and adopted for the analysis of drawings (Rose, 2007), and although proposals have been made (see for example Duncan, 2013) there is still no established protocol specifically for the education space.

A total of 145 learners submitted drawings. Learners were given blank, white A4 paper, coloured pencils and the direct prompt, "Draw yourself using the BFI tablet program." The prompt was given in Northern Sotho. Learners were allowed 30 minutes for drawing and

colouring. Drawings were done independently by each learner, although some learners drew themes or elements from neighbours, demonstrating peer influence in the study (Duncan, 2013).

To mitigate challenges in interpretation raised by Eldén (2012) and Literat (2013), as a proxy for discussion, learners were invited to write, if they wished to, in Northern Sotho or English about their drawing and/or their experiences with the tablet. The learner's writings were transcribed and translated, and drawings were analysed with both the written text and drawing media.

As no applicable framework for evaluation of learner drawings could be found, the researcher created one using five criteria specifically related to the relevant research objectives using a deductive approach (Thomas, 2003): engagement with the topic; mood and impressions conveyed by the drawing; placement of learner(s), teacher(s) and technology; engagement with specific parts of the program; and self-placement, including learner relationship to the tablet, other individuals and the school. Each drawing was evaluated according to these criteria, and emerging themes were determined based on patterns in the drawings across learners. Themes were categorized by grade and class in order to compare similarities and differences across sub-groups.

3.1.4 Analysis of Data

Data collected from teacher interviews, classroom observations and learner were analysed individually as described in the previous sections. Codes and themes emerging as well as the research objectives were then aggregated into a complete framework for analysis, which was used to review all data included in the qualitative analysis in an inductive approach (Thomas, 2003). The framework ultimately included 1) interactions between participants and technology; 2) interactions between types of participants (teachers, learners, school administrators); 3) technology within the systems and structures of education (curriculum requirements, classroom practice and accountability departments); 4) attitudes and perceptions towards the technology.

3.2 Presentation of Qualitative Data

"Now we are in the world of technology…" – study participant

The qualitative data collected was structured to answer the question: How does structured programme technology such as the BFI interact with the components of South Africa's

traditional educational system as outlined in the literature review section? The question was investigated through three perspectives: the observations of researchers, teacher perceptions as elucidated by interviews, and learner perspectives as gleaned from participatory drawings.

Figure 5 summarizes the relationships elucidated by the qualitative data. Solid lines indicate strong or particularly influential relationships, while dotted lines express infrequent or secondary relationships experienced by some participants.

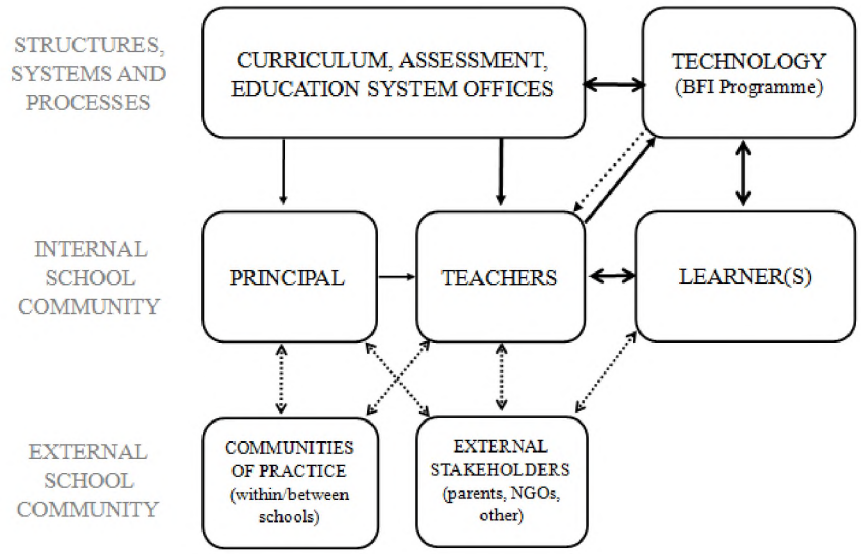


Figure 5: Education System Interactions with the BFI

Classroom-implemented technology was found to influence all levels of the system, from changing the way the curriculum was interpreted and delivered to the expectations of principals, teachers and learners and the way the school interacted with the community and with other schools. As the focus of this research is primarily linguistic, only a broad overview of the systemic and external factors elucidated will be provided, with an increased focus on classroom delivery and interactions, in order to determine factors which may have influenced quantitative outcomes discussed in subsequent chapters.

3.2.1 Overall Observations

In general, the grade 2 classes observed had more difficulty in handling the tablet than the grade 3 class. In terms of gross motor skills the grade 2 learners seemed much less capable than the grade 3 class. In one class observed, the most interesting part of the whole lesson was watching the learners try to put away the tablets. With the headphones, charger cords, tablet and box, they were basically trying to work out a 3D puzzle. It took some groups a long time to manage and they all had to try several times. Another teacher indicated that a difficult

aspect of program use was "switching the tablets off and on…Because they are not used to the tablet they take a long time. Others will say ours is not on, or not be on the same page."

Learners also had difficulty with fine motor skills. Unlocking the tablets was the first challenge for learners, and many were unable to follow the pattern (the letter N) the teacher drew on the board. In drag-and-drop activities, they would often miss the box the word/punctuation mark should have been placed in, resulting in an "incorrect" feedback.

In terms of tablet use, grade 2 learners knew how to turn tablets on but many had to be supported individually to perform basic functions such as opening the program or adjusting the volume. In all grade 2 classes, pairs or groups were observed to be "stuck" due to small technicalities, such as clicking off the keyboard and not knowing how to get back to it. This was added to the more expected challenges of learners accidentally exiting the program and having to restart the lesson. One grade 2 teacher expressed that most learners were completely unfamiliar with the technology before the programme began:

> *Tablets were something totally new to them, and they were excited just to touch the tablet. It was difficult for them to concentrate…to give them that opportunity to use it was something very huge for them. They talked about it for the whole day. They even imitated what they heard from the tablet outside during break.*

3.2.2 Classroom Observations: School A

School A was given a total of 50 tablets for use by grade 2 and 3 learners. The school opted to split the tablets between the grade 2 and 3 classes, meaning that each grade had a total of 25 tablets to use. Strategies for tablet use were devised by grade and were dissimilar. The school does have a lockable room which used to house computers, but had opted to take the tablets to classes. The principal asserted that "taking tablets to classes takes less time than taking the learners to a computer room".

In her interview, the principal indicated that she had observed tablet lessons herself and was impressed by "the concentration…if a pin dropped you could hear it because everyone was looking at the tablets".

School A, Grade 2 (A2): Large group work

The classes for grade 2 are arranged with learners in groups, with space to move around between desks and learners. There are 50 learners in the class observed.

The class was scheduled to use the program weekly, however they did not use the program every week due to frequent teacher absence. The school policy was not to use the tablets without a teacher present. Learners generally use the program for 1.5 hrs, from 9:00 to 10:30. The school has a nutrition break at 10:30, and after learners return the teachers may give learners additional tasks.

The learners generally use the program in pairs or groups, depending on how many learners are present and how many tablets are charged. Although they have the means to charge all tablets after every use, it seems there is some inconsistency in performing this task. At the time of observation, grade 2 had only used the program in Northern Sotho and had completed two lessons several times, as reported by the teacher. For the observation the learners began from lesson 1 in Northern Sotho.

The learners became very excited when the teacher announced 'tablet time', and when the teacher asked for participants most learners raised their hands. Only 24 were allowed to remain. The teacher explained:

> *Even those that are dull are able to use the tablets and they speak, unlike in a normal class. It gets interesting to each and every one. They laugh, they enjoy…There's more life than a normal class. The most interesting thing is that the tablets are talking to them like a teacher.*

The teacher explained that the tablets did not have enough charge for more than one lesson, and that they had not been charged the night before[15]. The end result was that although originally the learners were set up with one tablet per learner, half of the batteries ran out and the learners began working in groups as learners were reassigned by the teacher to the nearest working tablet. The largest group was 4, and some learners continued using the tablet

[15] The tablets used in the class have an average battery life of 6 hours.

individually. Overall there were nine groups with nine working tablets. When open programs were checked, it was found that a majority of the tablets had not been shut down properly and were still running more than ten different programs, including multiple copies of the BFI program, camera programs, games (fruit slash) and the calculator. The teacher was very familiar with tablets and said she owned a tablet. However, she preferred to perform most functions herself rather than instructing learners on how to do them. Perhaps as a result, learners in the class did not seem to be able to perform basic functions such as closing programs or adjusting the volume on the devices.

The teacher gave directions to the learners, but most did not wait for her instructions and opened the program as quickly as possible. Some learners opened to different lessons (one opened lesson 8) and two learners opened the program in English. Some learners were already halfway through the lesson when they began because the program had been left open. The learners did not use headphones and the atmosphere in the classroom was one of, at first, disorganized chaos, which gradually improved somewhat as learners improved their focus and began the activities.

In the beginning there was a lot of conversation and learners required assistance with basic functions: they would accidentally close the program and open a new copy of it; the sound was too loud; the learners were all talking to each other and shouting across the tables. Outside, learners who were on break came to the windows of the classroom and knocked and tried to peek in on the lesson until the teacher chased them away. The researchers were also pulled into assisting and closed the open programs so that the power would not die on the remaining machines. At other points a researcher stepped in to remedy or prevent learners from opening multiple copies of the program.

As the lesson continued learners became more comfortable with the tablets and the set up. The program noise continued and the learners constantly spoke to each other and the teacher as well as the researchers, but concentration increased for all learners. As the teacher asked for volunteers, it is possible that the high level of engagement was due to self-selection bias, or the tendency of those who volunteer to be more likely to fully engage. Celebration was common after completion of tasks by a pair or group, and even the researchers were prompted for high-fives by excited learners.

All learners focused on the program the whole lesson, but it was obvious from the other programs that were open that this is not always the case.

The teacher rotated between groups of learners performing the following functions: giving hints to learners about correct answers; turning the volume on the devices down or up; giving physical assistance (i.e., taking learners hands and moving them when learners had trouble with drag-and-drop or fine motor skills); encouraging learners after a wrong answer; celebrating with groups who achieved difficult tasks; and helping learners to power down and put away the machines. The teacher confirmed that the majority of what she does during a tablet lesson is checking on learners. She said learners would accidentally change the program to English, and that "some are just silly" and would open games or other applications. She saw her primary role for learners during a tablet lesson as an enforcer, ensuring learners stayed on task and to provide troubleshooting for them if necessary.

This class had the highest levels of peer collaboration, with learner-learner interaction constant in all groups. Even learners using tablets individually collaborated, with the faster learners being called in to help by slower learners. Learners were seen celebrating with high-fives, fist pumps and exchanged smiles, especially after completing the balloon popping game[16], both within and between groups.

Although the teacher did not assign one, in most groups a "pointer" was established, who controlled the actual action on the screen (drag and drop, click, etc.). The pointer was not in any observed case the strongest member of the group – the answers were often provided by other members and executed by the pointer. In many cases, especially at the beginning of the lesson, even when the pointer did not perform adequately (e.g., she chose the wrong answer or was slow to answer), the others would point and encourage her rather than taking over. However, as the difficulty of tasks increased learners began arguing to control the pointer position, especially in the larger groups and especially when the pointer hesitated or made an error. This did not seem to affect collaboration negatively, and at no point was a pair seen disengaging from each other. Spats were quickly replaced by the status quo and celebration after completion of the difficult task.

School A, Grade 3 (A3): Independent Tablet Work

School A has two grade 3 classes with between 45 and 50 learners. The grade 3 classes were arranged in groups of between four and eight learners, which provided ample space to move around the classroom.

[16] See section 2.3.2 for details on the program activities.

Each class was scheduled to use the tablet once per week. The teacher explained that when they used the tablets they did so in four groups, since the grade had only 25 tablets and the teachers had elected for the learners to use the tablets individually. As a result, during tablet time one class has 70 learners and the other has 20-25. After the learners completed the lesson for the day, they moved into the other classroom. Each group used the tablets for 30 minutes according to the schedule. However, the 24 grade 3 learners observed used the tablets for a total of 44 minutes. Learners were familiar with all activities and overall had little difficulty in completing the exercises. They were able to follow the directions in the program and learned from the corrections offered when they made errors. They were able to type the words in the typing game without observed challenges. About 40% of learners completed lesson 2 and moved on to lesson 3. One learner also worked through part of lesson 4.

When the teacher announced tablet time, some learners appeared excited and all put away their books promptly. The tablets were brought in in a box and distributed swiftly and silently, and the teacher began instructions to open the tablet box, turn on the device and choose the correct lesson for the day. Learners at that point put on headphones and worked through the lesson individually.

Learners were obviously familiar with the tablets and were able to perform all functions, including opening the tablets, turning them on, identifying and opening the BFI program, click, drag-and-drop, and typing. One learner was observed controlling the sound on her tablet, and another learner that accidentally closed the program was able to re-open it without assistance. The learners also knew how to care for them; a majority of the tablets still had covers and were kept in them. This is especially notable as the teacher of this class was one of the least familiar with tablets; she had obviously taken pains to not only educate herself but also her learners on their use and care.

The teacher kept the class silent throughout. At two points in the lesson individual learner's sound became audible and the teacher quickly instructed them on how to correct this. At another point, one learner began speaking and the teacher asked quickly, "Who's talking?" The learner fell silent. The net effect was that although the external environment was very noisy, with learners in other classrooms shouting and playing, inside the tablet classroom

everything was completely silent. The teacher shut the door about halfway through the lesson which reduced some of the external noise.

The teacher spent the lesson walking around, observing learners. She supported some mainly through encouragement and congratulations. For example, one learner was having difficulty on the balloon popping exercise, and she first told him "Ok" and then, when he failed, laughed in a friendly way and said, "Be fast!" She congratulated him and others upon completion of some exercises and when they turned in their tablets.

Learners were seldom observed interacting with each other. On two occasions, finished learners were observed trying to assist other learners who pushed the would-be helpers' hands away. It was clear that for these learners completion was a personal mission and they did not appreciate unsolicited assistance. On two occasions learners were seen gesturing to each other to relay information, in one case a correct answer, and in the other to stop a classmate from entering the answer the gesturer had just incorrectly entered. These were the only instances of collaboration observed.

Other learners left the classroom when they finished lesson 2. The lesson ended when the last learners completed lesson 2, in about 44 minutes.

3.2.3 Classroom Observations: School B

School B was given 27 Nexus 7 tablets for use during the intervention period. The school opted to use the tablets twice per week for grade two learners for the first half of the intervention period, with the intention of switching or adding grade 3 in the second half. The teachers indicated that grade 3 was heavily engaged in Annual National Assessment (ANA) preparation even from June, and that the program had taken a back seat to this, although they said a shortage of tablets also played a role in the decision.

As a result, only grade 2 classes were observed in the school. One grade 2 educator was interviewed about the program, and what was meant to be the HOD interview transformed into a focus group and workshop as the HOD asked that educators all be gathered to discuss the program. As a result, discussion was teacher-led, as the interview protocol developed for the survey had a high potential for bias if conducted in a focus group, due to the revealing and individual nature of the questions, such as asking about teacher performance.

Challenges raised by the focus group were all technical and due to user error, ranging from two 'broken' tablets that had not been charged to teachers asserting the program had 'frozen'

when they had not completed an activity and were therefore not prompted to move to the next activity. The group worked through a lesson together in a reiteration of the initial training, with the researcher explaining how to resolve challenges raised by the educators and the challenges teachers indicated were faced by learners. They also worked through the basic settings on the device and swiping between multiple home screens.

In terms of pedagogy, the teachers all favoured a teacher-led approach (described in detail below). Some teachers were clearly uncomfortable with the noise generated by independent or group work, specifically, and were concerned about classroom management.

School B, Grade 2 (B2): Teacher Led Tablet Work

School B is very large, with four grade 2 classes of between 50 and 60 learners. There were two classrooms observed, one laid out in rows and the other with six clusters of desks positioned for groups of seven to nine learners. Like school A, the school decided to split the tablets between two classes two days a week, which meant each class had either 13 or 14 tablets to use two days a week and learners formed groups of four or five during program use. A general session with the program lasts about 1 hour and fifteen minutes. The learners used the program once a week in English and once a week in Northern Sotho. The lesson observed in both classes was English lesson 5, indicating learners are about halfway through the program.

On the day of observation all 27 were allocated to a single class of learners, allowing for pair work with the odd group of three learners. Classes were noisy, and the learners were obviously distracted by the presence of the researchers. The teachers began by asking all learners to turn the sound off on their tablets, although some did not and were not corrected. The teacher stood in front of the class with one of the tablets with the volume at maximum.

Both teachers skipped the audio-visual component, which provides background context, and began the lesson at the sentences activity. For each activity, the teachers listened to the tablet instructions and repeated them to the class, asking for the answer. For some activities, the teachers added repetitions or components. For example, for the sentence building one teacher asked the learners to read the sentence several times, both call-and-repeat and as a group, and then they moved one word at a time as a group into its correct place in the sentence.

In school B, the teacher-centred approach meant that teachers spent the majority of the lesson talking, with instruction and direction coming from the front of the class and learners expected to pay attention and respond on demand. As a result, the teacher's interactions with learners primarily took three forms: soliciting answers from an individual; soliciting answers from the class; and providing corrective feedback, mostly in the form of discipline. One classroom exchange is outlined below:

Teacher: "Which words are the same?"

(Some learners raise their hands, and the teacher points to one of them)

Learner: "Clock and clean."

Teacher (to class): "Is she correct?"

Learners (chorus): "Yes! Let's move!"

Teacher: "No, we are not moving until we get direction."

As demonstrated by the dialogue above, when prompted for an answer by the program the teacher would ask the class for volunteers, at which point in each class a core group of clever learners would raise their hands. If the learner selected to answer was correct, the teacher would ask, "Is she correct?" and, at the chorused *yes*, would indicate the correct answer on her tablet and say, "Let's see. Press here." At the congratulatory message, the teacher would say, "Let's move!" and the learners would press the cue for the next activity. If the learner was incorrect she would choose another learner to respond or give instruction herself. For example, after only a few hands were raised in response to a solicitation for an answer to a grammar activity, the teacher of one class explained what a 'verb' is in Northern Sotho, which raised a few more hands. The teacher read the sentence again, emphasizing the verb words and providing a kicking motion with the word "kick". When the teacher was satisfied with the number of learners willing to respond, she called on one of them to give the correct answer. As a result of this style of interaction, the corrective feedback provided by the programme was never used.

The teachers had established "group leaders" who controlled the tablet. The group leader was always the most capable learner in the group, and one teacher explained that they did not switch off as "the others cannot read and write and are unable to answer the questions". Remaining learners were meant to be relegated to observer status.

However, within that structure collaboration between learners still happened frequently, though under threat of scolding. Some pairs turned to help other pairs when advice was solicited, and the class in general became disruptive between activities as groups spoke to each other. A few pairs of learners moved ahead of the class, leaning in close to hear directions and whispering to each other.

At the typing activities and the balloon popping game, the learners exhibited obvious excitement and were allowed to work independently. There was chaos during these activities, with a lot of cross-group collaboration and learners speaking to each other excitedly and celebrating after completing the activity.

3.2.4 Teacher Interviews: Interactions with Learners during Program Use

> *[The best thing about the program is] that they will know how to use the tablets and they will understand technology. They already know how to open it, how to search for the lesson they want. They will not be empty; they will have the basics.*
>
> – Programme participant

All teachers interviewed indicated that at first the tablets had introduced a measure of chaos to their classrooms. One teacher indicated that it was challenging "to control them [learners], because they were too excited" and another indicating that learners would "fight over who holds it…that is what causes noise". Teachers in school B indicated ongoing challenges, but teachers of school A noted that the situation had calmed in their classrooms, with the grade 2 teacher expressing that "it is simpler once everyone is familiar with the tablet…they are able to listen and share the tablet at the same time" and the grade 3 teacher commenting that now learners were easily managed during tablet lessons as "they just go with the program and it is very easy".

One teacher noted that learners were able to engage more with the program than in her regular classes:

> *The better way of learning for them is practical. They learn to be fast readers. Even spelling they are improving, and building sentences. I have a difficulty in getting them to build sentences even in Sepedi [Northern Sotho] during class, but I can see they are able to do it in the program. With the tablets they are written there, jangled around so they can practice…they like it more than being taught in class, because it's learning whilst playing. It's sort of a game.*

Of the four teachers observed, only one owned a tablet and considered herself fully competent with the device. Other teachers had mostly mastered basic functions such as turning a tablet on and off, but were unable to support learners in more complex functions. For example, in school B due to the method of tablet lesson delivery there were long periods of inactivity with the tablets, and learners' tablets would 'sleep' automatically. Teachers could not access the settings and did not know how to correct this, and in fact assumed the tablets were 'broken'. The principal interviewed noted that her school did not have such a difficult time adopting the tablet program in the classes because many learners were familiar with touch-screen phones. The teachers, however, posed more of a challenge:

> *Educators weren't that computer literate. Maybe…maybe if…If maybe something can happen and they are given more, there should be an intense training for educators. Because if they know everything then there won't be a problem that when they have this problem they run to the office, [and] I run to Google…*

Despite obvious challenges in using the device and the program, only the teacher of class A3 indicated that she did any preparation work for a tablet lesson, expressing that to prepare she went through each lesson before the learners attempt it to ensure she was able to work the device and the lesson. She said she had learned and was continuing to learn with her class despite being unfamiliar with tablets. She said she learns something new every day.

Other teachers responded that they did no preparation, with the two teachers from school A indicating that they did not have access to the tablets except during class time, as they were locked in the office after school. The teacher of class A2 also indicated that she does not particularly prepare for the tablet lessons: "I can't say I have something that I've written down as preparation. For the day I just tell myself that I want to achieve this. For example, all the learners to finish lesson 1 activity 1 today. Something like that."

All three teachers interviewed indicated that they followed the BFI program sessions with writing activities and/or workbook exercises. However, the difficulty of these activities varied. The teacher of A2 indicated they go straight to class workbooks with DBE exercises, while the teacher of A3 indicated that as a follow-up to program activities she has learners practice re-telling the story from the audio-visual component and/or the comprehension passage, and write the words that they learned on the chalkboard. She commented that the learners can usually write 10 or 15 words. In school B, the grade two learners of both classes

followed up with sentence writing and workbook or other writing activities. One teacher explained how she linked the tablet activities to her regular classwork:

> *If you do like today, we have done verbs, pronouns, punctuation, and then we will write them. Last week we were busy with nouns. After the lesson I write the sentences on the board and they have to underline the nouns. Today we were going to do verbs. After each and every lesson from that tablet we do writing.*

Teachers in both schools and the principal interviewed noted the alignment of the programme to CAPS, and all interviewed expressed that they believed it covered the CAPS curriculum for the grades concerned. However, teachers felt their learners struggled with the integrated approach to skills practice incorporated in the program. One teacher in the focus group commented to general agreement:

> *Sometimes it is heavier for the learners because they do more activities at the same time. In the DBE workbooks there is a day for punctuation, a day for pronouns. In the program you have to do everything at the same time. But it's already the same as the DBE books…You can see here we are doing sounds, and here we are doing sounds. We have pronouns, and here are pronouns...*

3.2.5 Learner Drawings

Drawings by learners were used to elicit a general understanding of how learners perceived the program and their participation in the program, as a proxy for interviews. Learners were provided with crayons and paper to draw with, and were given between 30 and 45 minutes to complete the instruction: Draw how you feel about the tablets and the BFI program. Instructions were given and learners' questions answered in Northern Sotho, as English competency among the foundation phase learners was low. In total, 145 learners submitted drawings, all of which were included in the analysis. Drawings were analysed under five criteria: engagement with the topic; mood and impressions conveyed by the drawing; placement of learner(s), teacher(s) and technology; engagement with specific parts of the program; and self-placement. The results are summarized in this subsection.

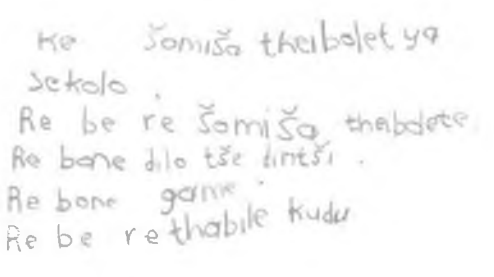
In terms of mood and impression, responses from learners were overwhelmingly positive, indicated by bright colours and smiling versions of the learners seated at a desk with a tablet, often accompanied by hearts and/or stars. There were two instances of learners who expressed negativity through drawing, colouring themselves in dark, heavy pencil despite the crayons available (see the example to the right).

One other learner drew her mother and sister rather than something related to the programme. The other 142 responses conveyed a general sense of satisfaction, enjoyment and engagement.

Although learners were not specifically required to write their impressions or feedback, many took the opportunity offered to try and express themselves. A few learners even tried to write in English. Many of these learners referred to the tablet as "ephone" (the phone).

Other learners wrote in Northern Sotho. Some learners in all classes wrote sentences or words from the lesson or about the lesson they had just completed. This was especially prevalent in class 3A, in which almost all learners wrote something about one of the characters, a girl called Mosibudi.

Grade 2 learners were more likely to express some emotion about the tablet (for example, one learner wrote "I love ephone"), explain what they drew (such as "Ke be ke kgotla gebelet Rebe tsennytša." – *I am pressing (touching the screen of) the tablet.*), or try to explain their relationship to the tablet (such as "Sekolong go bose. Re šomiša tablet. Re phela re šomiša tablet. Go bose sekolong. Ba re hlokomela sekolong sa rena. Re thabile kudu ge re kwišiša." – *At school it's nice. We are using tablets. We always use them. At school it's nice. They are taking care of us. We are very happy because we understand.* Another example is a learner who wrote: "Ke šomiša theibolet ya sekolo. Re be re šomiša theibolete. Re bone dilo tše dintši. Re bone game. Re be re thabile kudu." – *I use the school tablet. We were using a tablet. We have seen a lot. We see games. We are very happy.*)



3.3 Concluding Remarks on Qualitative Observations and Results

There were several findings which are pertinent in interpreting the quantitative data.

First, learners displayed similar high rates of positive engagement with the topic in every class, and learner perceptions remained consistent regardless of pedagogical style. This means that while the type of pedagogy may have affected learning outcomes, it is unlikely to be because learners were less interested or engaged due to a particular style of pedagogy.

Second, learners displayed immediate retention of programme knowledge through writing words, names and events from the story as well as drawing program activities. This means that engagement with the program did result in immediate changes in knowledge, as learners were able to recall facts from the tablet narrative and the activities undertaken. This pertains more to explicit knowledge, as implicit knowledge was not on display in the drawings.

Third, despite similarities in learner perceptions, the pedagogical styles and the ways in which teachers interacted with the tablet, as well as the ways they expected learners to interact with the tablet, may have influenced both outcomes of the intervention and learners' willingness to experiment and learn from the device itself.

In both classes in school A, the role of the teacher shifted from instruction to facilitation and support. Teacher interactions with the tablets in school A took three forms: observation of learners using the program; elicited support; and unsolicited support which seemed to come in response to perceived difficulty or slowness in following instructions. For both types of support, teachers in both schools performed functions for learners having difficulties rather than using oral instruction or correction.

In school B, teachers chose rather to adapt the tablet to the typical role of a teaching and learning aid in the classroom, similar to how a teacher might use a poster or another teaching prop. Teacher interactions with the device were constant, while learners interacted with the device mainly by mimicking teacher actions after an answer had been derived and they were given instruction to do so. In other words, the teachers retained the traditional role as the purveyor of knowledge from the front and centre of the class rather than shifting to facilitators as was seen in school A.

Finally, there were differentials in the type of follow-up activities performed in class, which could be a factor affecting retention. Class A2 followed up with only regular classwork from the DBE workbooks, which cover the same skills as can be found in the programme but are

not linked to the BFI story. Class A3 followed programme use with retelling and writing activities linked to the story, including words and exercises. Class B2 had the most rigorous follow-up activities, in which learners wrote sentences from the program content and did linked grammar exercises.

These differentials will be explored in Chapter 6 alongside the quantitative results.

CHAPTER 4: THE EARLY GRADE READING ASSESSMENT (EGRA) TEST OF LITERACY SKILL ACQUISITION

In order to determine changes in literacy skills achievement, a time series design was used with three testing periods. Gains between the first two baselines were used as a control to determine intervention impact, or gains during the intervention period. This was chosen above a randomized control trial due to the small number of schools involved in the study and the corresponding high influence of uncontrolled variables such as school schedule, teacher competence and pedagogy, rates of teacher and learner absenteeism, support of district and provincial school officials, and so forth (for a review of relevant literature on the relative benefits of randomized control trials and quasi-experimental methodologies in the education sector, see Chapter 2).

In order to evaluate literacy skill achievement, a method-independent tool had to be appropriated or developed to apply in each testing period. The tool had to be method-independent because by definition the control and intervention used different pedagogies, and therefore a method or pedagogy dependent test which tested the specific skills in the BFI program may have shown greater gains which were meaningless, akin to asking learners if they had learned something, teaching the item, and then asking again if they had learned it. While the retention part of the study was interested in specific skills uptake, questions around literacy acquisition were interested in overall progression achieved, and therefore required an independent rather than derivative test.

Ultimately, the Early Grade Reading Assessment (EGRA) was chosen to fulfil this need, as it could be adapted to Northern Sotho based on internationally-recognized and utilised test development guidelines, and offered an opportunity to test skills in relation to each other using a single test (RTI International, 2009). Testing multiple literacy skills in relation to each other enables an analysis which plots learner performance and changes in learner performance along a development continuum from phonemic awareness through decoding to fluency and comprehension using a single tool.

This chapter first provides a discussion of the EGRA tool through pertinent literature and frameworks. This section also outlines the theoretical framework underpinning the methodology used to administer the EGRA. The second section of the chapter discusses the methodology used for data collection and analysis. The third section of the chapter presents

the results of the EGRA analysis. The final section discusses the outcomes in relation to the following research questions:

1.) What impact does BFI program use have on the following literacy skills in Northern Sotho-speaking grade 2 and 3 learners:
    a. Letter-sound recognition
    b. Simple word decoding
    c. Reading speed
    d. Comprehension
2.) Is there an optimal age for introducing the BFI program? Is it more effective with older or younger learners?

4.1 Literature and Theory on Literacy Assessment in Schools

The Early Grade Reading Assessment (EGRA) was developed through a collaborative effort of international education and literacy professionals working under the auspices of USAID, the World Bank and RTI International, and has been used in over 200 languages to date (https://www.eddataglobal.org/reading/index.cfm), including the impact evaluation of both the PC and tablet versions of the BFI technology program addressed in this report.

The Early Grade Reading Assessment (EGRA) examination was chosen for this study to evaluate the functional literacy skill of grade 2 and 3 learners in four areas: phoneme-grapheme assignment, simple word decoding, reading speed as a measure of fluency and comprehension. The EGRA was chosen because it fulfilled the following criteria:

- It is an internationally recognized method-independent tool, meaning outcomes can be compared across different methods of pedagogy or even different curriculums as the results reflect literacy outcomes;
- The EGRA examination is highly adaptable, and could be manipulated easily to answer the research questions involved using a single tool;
- There is a published EGRA Toolkit, developed by international early literacy experts including South African experts. Following the established guidelines in test creation ensured the Northern Sotho tool ultimately used 'best practice' assessment knowledge and was of an international quality standard;
- The EGRA undertakes to evaluate distinct identified phases of literacy acquisition: letter name knowledge, phonemic awareness, letter-sound knowledge, familiar word reading, unfamiliar word reading, oral reading fluency with comprehension, listening comprehension and dictation. Therefore learners can be mapped at levels of literacy achievement which are relative to each other; and
- The EGRA is a rigorous assessment undertaken individually, which eliminated some important types of biases which can emerge in classroom testing due to copying or unintended peer support. The results of each learner belong to that learner alone (RTI International, 2009).

The particular Northern Sotho version of the EGRA examination used by this evaluation was developed by the author of this study in collaboration with language editors at the Molteno Institute for Language and Literacy. The test was developed based on EGRA toolkit guidelines and was matched to a late grade 1 or early grade 2 level. The test was produced

and field tested by the Molteno Institute for Language and Literacy in 2013 in Northern Sotho. The test was then revised and used to evaluate early grade reading skill in 28 Northern Sotho schools in 2013 and 2014, and showed high internal validity (Castillo et al., 2015). This particular Northern Sotho EGRA was therefore selected, as it was a field-tested tool in the Northern Sotho language.

In order to target the skills of particular interest and shorten the evaluation time, three components of the EGRA were included: letter-sound knowledge, familiar word reading and passage reading.

First, letter-sound knowledge was included. While a case could be made that letter name knowledge is the first stage of literacy acquisition in the formal school curriculum in South Africa, in Northern Sotho the names of the letters are the same as the sounds they make (as opposed to, for example, English, in which a letter may be called 'k' but make the sound /k/). Therefore, the two sections of letter name and letter-sound knowledge are synonymous in Northern Sotho.

Second, familiar word reading was included to test decoding. In the first iteration of the test, used for the USAID Bridges to the Future Initiative Project, unfamiliar word reading was also included, as several studies have indicated the usefulness of nonsense or unfamiliar word reading as measures of phonemic decoding skill and for identifying literacy deficits in early readers (for a review, see Farrell et al., *The Value of Nonsense Words*, available at http://www.readsters.com/wordpress/wp-content/uploads/2010/12/ValueOfNonsenseWords.pdf). Cardenas (2009) further suggests that phonemic instruction is more effective when reading lists which include non-words are used to reinforce phonemic concepts.

However, the evaluation team encountered objections from the teachers involved, including threats of union action to block project access to schools. A secondary concern was raised in the nature of the South African curriculum, which favours additive bilingualism, usually in home language and English, two languages with differing phonological 'rules' to apply to unfamiliar words. In the pilot, when faced with nonwords learners presented expected pronunciations in English more often than Northern Sotho, perhaps because of an expectation or repeated experience that unfamiliar words were English words. Finally, a large number of learners were unable to perform the task at all as they seemed to not understand the directive. Ultimately, due to these concerns unfamiliar word reading was omitted and familiar word reading was adopted as the measure of simple word decoding.

Thirdly, passage reading and comprehension were included. For the purposes of analysis, oral reading fluency (correct words per minute) and comprehension were treated as two distinct components, bringing the total count of measured variables to four.

The EGRA toolkit calls for learners to read only one minute of text, and then to be asked questions based on the portion of the reading passage they were able to complete. The modified version of the EGRA test used in this study deviated from this procedure. Learners were timed for one minute, and the total number of words correct recorded. However, they were allowed to finish the passage in their own time, which is a deviation from standard EGRA procedure. The reason for this was to mitigate a potential analysis challenge in which correct words per minute (cwpm) automatically correlated to comprehension due to non-exposure to parts of the passage. There are no established reading benchmarks in Northern Sotho for the grades under consideration, and little evidence for a reading speed which indicates automaticity in Northern Sotho, so the researcher felt it would be a mistake to make assumptions with regard to reading speed and comprehension as the traditional administration of the EGRA implicitly does.

The lack of established reading benchmarks is a challenge which could not be mitigated by the testing tool. Cwpm as a measure of fluency would have best been adopted against standard norms, such as those available in English. The following section discusses the available literature on benchmarking of reading rates, and outlines the strategy adopted to mitigate this significant challenge to the research.

4.1.1 Approach to Benchmarking Reading Rates in Northern Sotho for Early Grades

Reading benchmarks for learners of English in America were developed as early as 1992, and revised by Hasbrouck and Tindal (2006) to include a larger and more representative demographic. According to the outcomes of their investigation, readers at the 50th percentile in grade 2 could be expected to read 72 correct words per minute, and readers at the 50th percentile of grade 3 could be expected to read 92 correct words per minute by mid-year. Benchmarks are therefore descriptive in nature, derived from the actual performance of a sample and applied with a degree of representation assumed.

In 2012, NEEDU (National Education Evaluation and Development Unit) investigated the state of literacy in grade 2 in South Africa, in part through the administration of the EGRA test to 641 grade 2 learners in all languages, attempting to use a similar strategy to benchmark reading rates in Northern Sotho. However, due to the low performance of South African

learners, descriptive benchmarks were unreasonably low, and the unit advised benchmarks based instead on a combination of their evaluation and the American benchmarks. Ultimately, the NEEDU report suggested 50th percentile benchmarks of 70 correct words per minute for grade 2 and 95 correct words per minute for grade 3. However, the results of the NEEDU study indicate that in South Africa, 72% of *the top performing learners* in grade 2 did not meet the 50th percentile benchmark (NEEDU, 2012).

Given the caveat that only 28% of the best-performing Northern Sotho-speaking learners could reach the 70 correct words per minute benchmark, the benchmark itself is called into question as it merely echoes the American benchmark, without consideration of factors such as orthography of the languages, average socioeconomic status of the learners, school resourcing, adherence to school nutrition programmes, or other variables specific to the context and the language which may influence descriptive benchmarks. Further, the results of the 2012 NEEDU study were not disaggregated by language or community grouping, which is problematic given the comparative complexity of the orthographies under study (NEEDU, 2012) and the high degree of variability in other factors influencing educational outcomes between communities and schools.

A later reading trajectory study by Pretorius and Currin (2010) measured gains by Northern Sotho-speaking grade 7 learners in both Northern Sotho and English during a library support intervention. Learners gained 37 words per minute over two years in English, and 26 words per minute in Northern Sotho. The average reading rate of learners in English was 143 words per minute, which, according to the benchmarks set by Hasbrouck and Tindal (2006), is on par with expectations for grade 7. In Northern Sotho learners averaged 119 words per minute, which, although this cannot be placed against a grade benchmark, should be sufficient to indicate automaticity in decoding and the access of 'fast-track' neurological processing of the material (Coltheart et al., 2001).

However, the accomplishment of grade parity is somewhat offset by the accompanying comprehension gains: a mere 18.3% in English and 8.7% in Northern Sotho. At the close of the study, comprehension performance averaged at 47.8% in English and only 38.7% in Northern Sotho – the learners' home language. It is significant that comprehension rates were less than 50%, although reading speed reached acceptable measures to indicate automaticity. This indicates that in South Africa, speed does not necessarily indicate engagement in comprehension or analysis, in home language or in first additional language, and that there

are likely other factors that determine when and if a learner can or does engage in comprehension processes.

As purely descriptive benchmarks set the benchmark near 0, and evidence of grade-parity fluency and comprehension was lacking, ultimately the question remained: where should reading benchmarks be set for South Africa as a whole, and in particular for rural learners of indigenous languages in home language? Specific to this study, how could fluency or degrees of fluency be described in the absence of rigorously validated benchmarks?

Given the lack of extensive research external to the project, benchmarks had to be created by the EGRA data collected. A total of 158 learner tests conducted in Northern Sotho were used to establish the descriptive reading benchmarks in Table 5.

Table 5: Northern Sotho Reading Benchmarks

Percentile	Average Reading Speed	Minimum reading speed	Maximum reading speed
below 20th	0 cwpm	0 cwpm	0 cwpm
20th – 49th	15.3 cwpm	0 cwpm	39 cwpm
50th – 75th	51.9 cwpm	40 cwpm	65 cwpm
above 75th	82.5 cwpm	65 cwpm	132 cwpm

The 75th percentile benchmark fell in a group of learners with the same cwpm score (65 cwpm). Therefore, this score is technically split between the two groups encompassing and above the 75th percentile. The same is true for the 20th percentile; as 22 percent of the study population scored 0 or below, the 20th percentile falls amidst 0 scores.

Based on the descriptive data curve, the EGRA data was used to establish a 50th percentile reading benchmark at 40 correct words per minute, and a 75th percentile reading benchmark at 65 correct words per minute. These benchmarks were important in establishing changes in the population, and in grouping learners according to ability.

4.2 Research Design and Methodology

Once the EGRA had been established as the method-independent test which would be used to evaluate literacy skills gain, and concerns emerging about relative measurement using benchmarks were placated so far as possible, the EGRA was implemented three times throughout the year in a longitudinal design (see Chapter 2 for details) as outlined in Figure 6.

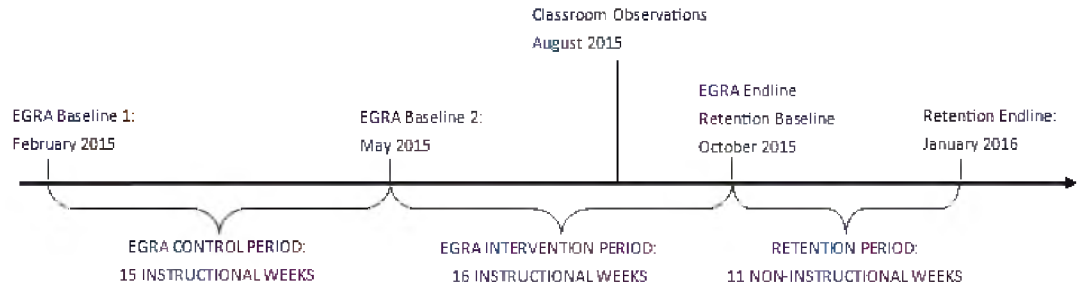


Figure 6: Research Timeline

In a longitudinal design, the participant sample acts as its own control group, which eliminates the influence of many additional variables which may affect results, management capacity and style, teacher capacity, infrastructure, additional resourcing and school location as well as individual variables such as aptitude, attitude towards schooling, existence or non-existence of extra-curricular support structures, and so forth. Therefore, a longitudinal design is complementary to the small sample that results from a labour-intensive evaluation tool such as the EGRA and the small scope of the sample schools.

In accordance with this design, EGRA testing was implemented in February, May and October 2015. Using the standard 2015 South African school calendar for Limpopo, school was in session for 15 weeks between the first baseline testing and the second testing of learners, and for 16 weeks between the second testing and the endline testing of learners. Although efforts were made to make the duration equivalent, ultimately due to competing school responsibilities the endline testing had to be delayed for a week.

The EGRA test was administered by three experienced EGRA administrators, trained in both the testing procedure and in interacting with young learners by the Molteno Institute for Language and Literacy. All administrators were first language Northern Sotho speakers who were also fluent in spoken English. All test administration was overseen by the researcher. The span between the first two tests acted as the control period and the span between the second and third test encompassed the intervention period.

4.2.1 Testing Instrument

As discussed in section 1 of this chapter, the EGRA examination was comprised of four elements, one in response to each literacy skill under evaluation.

The first part of the test was based on letter-sound knowledge and required learners to assign phonemes to graphemes. One observation of the Northwest Needs Analysis undertaken by

Shiohira & Shezi (2013) was that some teachers had a poor understanding of orthographically complex phonemes in their home languages, and were unsure of the appropriate pedagogy. Therefore, to prevent learner confusion and avoid the possible encroachment of pedagogical influence on testing outcomes, only single-letter phonemes were presented. While a case could be made for their inclusion, a secondary consideration was the intention that the test be pitched to late grade 1 or early grade 2 literacy levels. Figure 7 shows an example of two lines of the Northern Sotho letter reading activity. In total, 110 test items were presented, with all capital and lower case representations of single letters included.

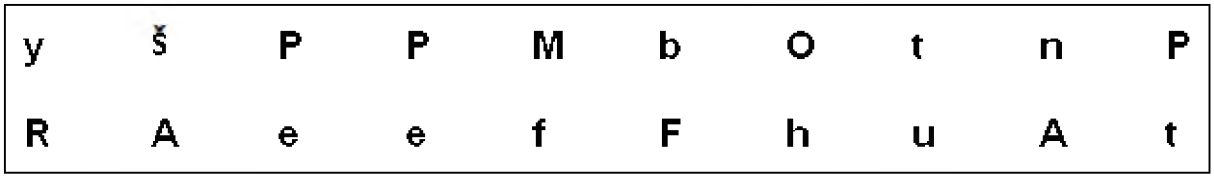


Figure 7: Example of Northern Sotho grapheme-phoneme assignment

The second test item was simple word decoding, which was accessed through the familiar word reading EGRA task. Lacking a reliable word frequency list in Northern Sotho, the English Dolche list 500 were referenced to determine high frequency words. Language experts and community teachers were consulted to remove words which were not relevant or likely to be encountered by learners in their everyday life, as the intention was to measure reading of familiar words. In total, 50 words were presented with a maximum of 6 phonemes, with orthographically complex phonemes included predominantly in the second half of the test item. Words included were limited to the following structures: V, CV, CVV, CVCV, VCVCV, CVVCV, and CVCVCV. Figure 8 provides an example of the type and presentation of words.

Figure 8: Example of Northern Sotho Familiar Word Reading

Familiar word reading was followed by passage reading. The passage consisted of 66 words and had a type-token ratio of 1 to 2.5, which is higher than the English EGRA ratio of 1 to 1.7 but is a product of the repeated noun and pronoun morphological structure of Northern Sotho, which inflates the number of tokens and depresses the number of types.

Language experts and teachers were consulted with reference to both grade level expectations and relevance of the content of the reading passage before it was finalized.

Comprehension questions were based on the passage reading, and were structured to mirror the presentation of questions in the English EGRA, which is based on Barrett's taxonomy. Some English EGRA questions are provided in Table 6.

Table 6: English EGRA Comprehension Questions and Classification Under Barrett's Taxonomy

Type of Question	Examples from EGRA tests
Direct	Where did the snake bite Sam? Who is this story about? What does Sam have?
Evaluative	Why did Sam run away? Why do you think Rosie wrote a letter?
Inference/Synthesis	What made the family happy again? Who do you think she sent the letter to?
Appreciative	What do you think happened to the snake? Does the story have a happy ending?

The Northern Sotho EGRA test included two direct questions, one evaluative, one synthesis and one appreciative question.

4.2.2 Sample

The Early Grade Reading Assessment is a one-on-one assessment which takes between 10 and 20 minutes per learner. As a result, the sample was limited by the number of trained researchers available to administer the test, and the amount of time schools were willing to devote to testing. With the personnel available and accounting for travel time, it was possible to comfortably administer roughly 40 tests per language in a school day.

School A allowed for multiple days of testing, and so for the EGRA analysis a random sample of 10% was taken. The same strategy was applied in school B, but with a 5% sample. This lower percentage was due to the significantly higher number of learners, and the fact that the school requested testing be completed within one day.

Sampling for the EGRA was done using the class lists for each grade. The total number of learners to be tested was determined by percentage, and the n^{th} learner was chosen based on the results. For example, in the case of 200 learners and a 10% sample, 20 learners total would be chosen; therefore, every tenth learner was selected. If a learner selected did not meet the first-language criteria or was absent on the day of testing, the learner directly below the selected learner on the class list was substituted. For the subsequent test period in May, the list of learners tested in February was produced and the same learners were tested. In case of absence, the learner directly below the absent learner on the class list was substituted, so

long as they met the Northern Sotho first-language criteria. In October, an aggregate list of all learners tested in February and May was produced and all available learners from the list were tested.

The attrition rate for the year was 21% for learners tested in Northern Sotho and 25% for learners tested in English. The overall attrition rate for school B was 35%, while school A had an attrition rate of 10%. This led to a disproportionately low number of learners from school B being included in the final sample.

The full set of participant data was used to determine reading benchmarks and for correlations between aspects of the evaluation (e.g., the correlation between reading speed and comprehension, or error and comprehension). However, because of the relatively small sample, representation of class, school or grade could not be reliably determined. Therefore, the analysis is focused on the progression of the cohort of learners who sat in all three EGRA testing periods through the year, and specifically compares changes in performance in the control and intervention periods. In this way, the sample acts as its own control and issues of representativeness are mitigated with respect to the research questions.

Table 7: Final Participant Numbers after Attrition

	Grade 2	Grade 3	TOTAL
School A	14	14	28
School B	7	10	17
TOTAL	21	24	45

4.2.3 Testing Procedures

During a pilot period of the tool in 2013, it was found that learners shared information after being released from the examination room, as some learners who were unable to read the passage were able to answer comprehension questions. To mitigate this, all learners selected in a grade were assembled in a waiting area and called one by one to the testing area. After the test was administered, they were escorted back to their classrooms to avoid conversation with those still waiting to be tested. A teacher from the school was engaged to monitor learners waiting their turns, both to keep order and to ensure there was no conversation between those who had completed the test and those waiting.

Another challenge observed in piloting was in explaining the test to the learners. Although the research team included Northern Sotho speakers, the young learners were excited by the newness of the research team and this seemed to compromise their comprehension. As a

result, one of the class teachers was recruited to explain to the entire learner sample the testing procedure before testing began, as the teacher and her way of speaking were familiar to the learners. The explanations were repeated in the examination room for each individual learner by the research team before each activity, and no problems were encountered with learners who did not understand what was expected of them.

In the testing areas, test administrators worked independently, one at a time with the learners. Administrators began by introducing themselves and asking learners, "How are you?" as per standard class greeting procedure to put them more at ease. Learners were also asked, "How old are you?" and "Are you ready?" in Northern Sotho.

Before the letter and word reading activities, learners were given two 'practice' examples in addition to the repeated instructions. Each of these activities was timed at one minute using stopwatches, or the stopwatch function of a cell phone, and learners were stopped after the minute had elapsed or if they made six consecutive errors in letter or word reading. The total number of correct letters or words read as well as the number of errors was recorded for each learner.

For the comprehension passage, learners were timed for one minute and the number of correct words read and errors was recorded. However, in a deviation from standard EGRA practice, learners were allowed to finish the passage in their own time and not stopped after one minute, as discussed in the literature and test review in this chapter.

All learners were allowed to attempt letter reading, word reading and comprehension. However, a number of learners were unable to read any words in the passage; these learners were not given the comprehension section and were assigned a 0 score.

In the comprehension section, learners were asked and answered five questions orally. Questions were repeated twice if necessary. Some learners pointed to words in the passage in response to the questions; these learners were asked to say their answer. If they were unable to respond orally, answers were counted as incorrect. All questions had a limited set of acceptable answers. However, for the evaluative question, which asked whether the story had a happy ending and why, the set was expanded after the 2013 trial to include both 'yes' and 'no' answers so long as the explanation was sufficiently logical. (The set answers were originally specified as 'no', because in the story a girl breaks a window accidentally and her father is very angry. However, a number of learners answered 'yes' unexpectedly, deciding with child-logic that the story was happy because the girl was able to escape from her father

without being beaten.) All other questions had a limited set of answers which did not change from the pilot testing to administration.

4.2.4 Treatment and Analysis of Data

After testing, data was collected by the researcher and data entry into an MS Excel spreadsheet was completed on each day of testing. Overall, 2% of collected data was compromised due to inconsistent recording (e.g.; scores higher than the maximum or mismatched languages, meaning the learner was mistakenly tested in different languages during different testing periods due to administrator error) or illegibility; this data was discarded. All data was entered twice and checked for discrepancies, which when discovered were resolved through consultation with the hard-copy data. Dated hard copies were retained and used to perform a 10% random check for accuracy in data entry before the aggregate analysis commenced (the resulting error rate was 0%). They were retained through the end of the analysis period in case of discrepancies arising or the necessity of data checks. However, as hard copies included learner names they were not disseminated beyond the researcher and were destroyed at the completion of the EGRA analysis.

Learner names were used to keep records of learners tested as the assistance of the school was necessary in ascertaining that the correct learners were called, but names were removed in the final dataset to protect learner identities.

In addition, teachers provided a record of hours of program use in their individual classrooms, which was entered as a variable into the dataset and used to analyse gains against hours of program use.

Finally, name, school, grader, class, grade, gender and age in October were collected as descriptive data in order to further explore the dataset and quantitative performance.

4.3 Assessment of Literacy Skills Using the Early Grade Reading Assessment

The analysis is divided into five sections. The first section provides an overview of performance changes over time. The subsequent four sections deal with the skills examined by the EGRA: phoneme-grapheme assignment, individual word reading, correct words per minute (cwpm) and comprehension. Each of these four sections begins with an overview of how the task was conducted, followed by an analysis including comparison of performance in the control and treatment periods, changes in distribution patterns, disaggregated results by school and grade and analysis by attainment level. For the cwpm and comprehension tasks, additional analysis of errors and the effect of reading speed on comprehension are presented.

4.3.1 Overall Performance Changes

In order to measure change, it was necessary to design a formula which would measure the achievement on all tasks in aggregate, while taking error into account. Table 8 is an example of the quantitative data gathered from a single learner over the course of the evaluation using the EGRA tool. The table also shows the maximum possible score and the average score for each test area. Scores were tallied by subtracting 25% of the error from the total correct, following the procedure from the American Scholastic Achievement Test for five answer questions[17]. The smallest fraction was selected for use in this evaluation as the responses were open and therefore infinite incorrect answers were possible.

Table 8: Example of EGRA Data Collected

	Grapheme-phoneme assignment			Simple word decoding			Passage reading (fluency)				Comprehension	TOTAL SCORE
Date	Correct	Error	Score	Correct	Error	Score	CWPM	Error	Passage Time	Score		
Feb	18	5	16.75	3	7	1.25	0	6	60	-1.5	0	16.5
May	38	2	37.5	12	8	10	12	9	60	9.75	0	57.25
Oct	46	14	42.5	12	19	7.25	30	12	60	27	4	116.75
MAX SCORE			110			50				132	5	342
AVG SCORE			39			20				40	2.4	124

[17] See https://collegereadiness.college board.org/sat-subject-tests/scores/how-tests-are-scored for details.

Comprehension was problematic, as there were only five questions. Therefore, if the raw score was taken as in other items, due to the discrepancy in maximum score comprehension ability counted for very little in a learner's overall capacity. Conversely, if the percent was used, each correct answer counted for 20 points, and given the average performance on the other items and the fact that realistically some maximums could not be reached within the minute timeline, this also seemed extravagant. Further, answering one question correctly meant a score of 0 if error was accounted for, which led to a lack of differentiation between learners with no comprehension ability and those who demonstrated a basic ability. Ultimately, a formula was devised by the researcher based on a maximum of 50 points for comprehension:

$$\begin{aligned} &\big(phonemes\ correct - (.25 \times phoneme\ error)\big) \\ &+ \big(words\ correct - (.25 \times words\ error)\big) \\ &+ \big(cwpm - (.25 \times passage\ error)\big) \\ &+ (.5 \times comprehension\ percent) \\ &= overall\ score \end{aligned}$$

A total of 45 learners were tested in all three testing sessions, 21 learners in grade 2 and 24 learners in grade 3. A boxplot of the participants overall scores, as calculated using the devised formula in each of the three testing sessions, shows the pattern of skills improvement for learners over the year (see Figure 9).

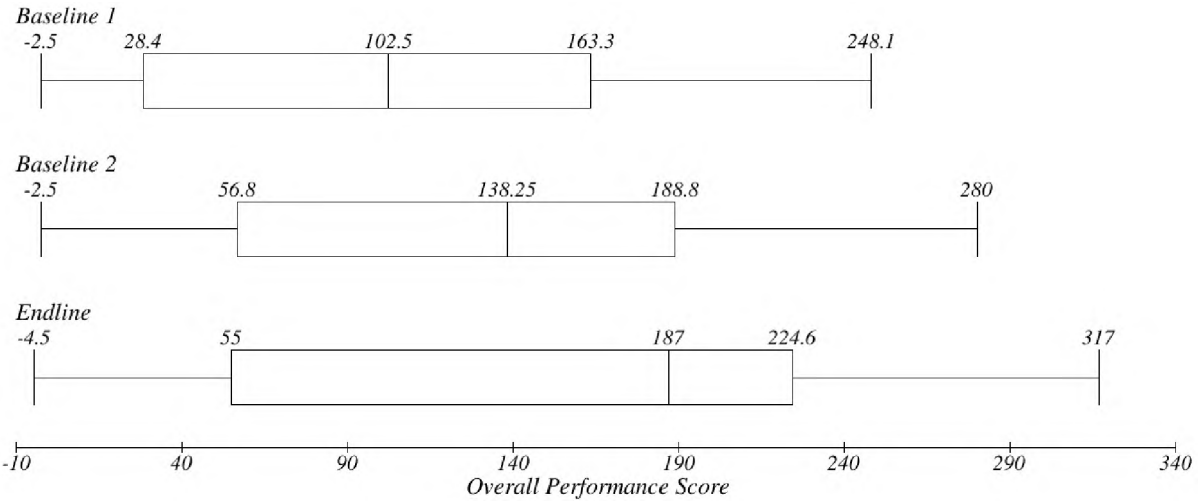


Figure 9: Boxplot of Overall Learner Performance by Test Period

The progression is characterized by gains in the top half of the sample's performance and no or faltering momentum in the lowest performing quarter. The end result of this dichotomous change in performance is that while the interquartile range of the overall achievement scores remains relatively constant in the control period, 134.9 points in the first baseline and 132 in

the second baseline, an elongated interquartile range of 169.6 points develops in the endline testing period. Likewise, the overall range of the data increases over time: 250.6 points at the beginning of the year; 282.5 at the midline; and 321.5 at the end of the year. These expanding ranges indicate that although overall on average gains were made by the learners, at least 25 percent of learners ceased to progresses in literacy at some point in the year.

The median overall score of learners increased in the control period by 35.75 points, and increased by 48.75 points during the treatment period. The third quintile score increased by 25.5 points in the control period, and by 35.8 points in the treatment period, while the maximum score increased by 31.9 points in the control period and 37 points during the treatment period. Therefore, a pattern emerges in which learners performing at or above the median showed gains throughout the year and increased their performance more during the treatment period than in the control period.

When the minimum score and the first quintile are considered, however, the opposite trend appears. The minimum score remained steady during the control period, and dropped by two points in the treatment period, while the first quintile doubled in value in the control period, between the two baselines, and then lost 1.8 points in the treatment period, or from the second baseline to the endline. This may indicate that learners in the lowest 25 percent of the sample were below an ability threshold necessary to gain from the program. As class time shifted from traditional teaching to program use, program use may have even negatively impacted their performance.

Performance Changes by Attainment Level

In her 2008 book *Proust and the Squid, The Story and Science of the Reading Brain*, Wolf outlined five stages of reading: the emerging pre-reader; the novice reader; the decoding reader; the fluent comprehending reader; and the expert reader. The emerging pre-reader stage included the introduction to and understanding of sounds and words as well as early print concepts such as images. The novice reader learned grapheme to phoneme correspondence rules and engage in early decoding. The decoding reader begins to use 'sight chunks' such as pre and suffixes and forms an emotional attachment to print as they can engage words, sentences and stories. Fluent comprehending readers and expert readers are able to extract meaning from texts of various difficulty levels.

An approach which took initial level of ability into consideration was necessary for this intervention analysis, in order to determine the effects of the technology literacy programme on individual linguistic skills. While an overall analysis can provide general answers, it was useful to consider learners in groups as Wolf did, to investigate the gains in various skills were made by learners at differing levels of attainment.

However, the stages as laid out by Wolf could not be used exactly as defined. First, the study did not deal with print concepts or oral language development. Secondly, the skills specifically under study were at times included in a single grouping, such as phoneme-grapheme assignment knowledge and single word decoding, both achievements of the novice reader in Wolf's grouping. Third, decoding to expert readership must be considered along a continuum with no clear pre-determinable breaks in the language under study.

Therefore, the researcher used the framework provided by Wolf as a basis for an attainment level framework for this study. As no information was gathered on pre-literate print knowledge or oral language, the emerging pre-reader stage was replaced by a "pre-reader" group, which was unable to demonstrate ability on any EGRA literacy skill task. The novice reader stage was subdivided based on acquisition of two skills under study: phoneme-grapheme assignment knowledge and early decoding. Lacking external definitions of "decoding", "fluent" or "expert" fluency for Northern Sotho readers in the grade levels under study, or indications of reading speed necessary to attain fluency, rather than sub-divide the remaining readers into three groups, the 50^{th} and 75^{th} percentile benchmarks established by the data collected were used to divide learners into groups. The formula and parameters used to group learners into these five attainment levels is described in Table 9.

Table 9: Attainment Level Descriptions and Formulas

ATTAINMENT LEVEL	Description	Formula Used	Percent of Study Population
1 – Pre-readers	Unable to assign phonemes to graphemes or read words	$(letters\ correct - (.25 \times letters\ error)) < 10$ AND $(words\ correct - (.25 \times words\ error)) < 5$	9%
2 – Phonemically Aware	Able to assign phonemes to graphemes, but unable to read	$(letters\ correct - (.25 \times letters\ error)) \geq 10$ AND $(words\ correct - (.25 \times words\ error)) < 5$	13%

	words		
3 – Beginning Readers	Able to read words, but unable to reach the 50th percentile average	$5 \leq (words\ correct - (.25 \times words\ error)) < 40$	23%
4 – Developing Readers	Able to reach the 50th percentile average benchmark, but not the 75th percentile average	$40 \leq (words\ correct - (.25 \times words\ error)) < 65$	24%
5 – Fluent Readers	Reading at or above the 75th percentile	$65 \leq (words\ correct - (.25 \times words\ error))$	30%

The purpose of grouping readers according to ability levels is first to assist in determining change in ability distribution patterns in the sample under study, and second to help determine if and how learners of different baseline abilities responded differently in the control and treatment periods. Different responses would provide evidence of a base and/or a maximum capacity threshold at which the Bridges to the Future Initiative is effective.

Change in attainment level distribution

In terms of distribution patterns, the population began with a nearly normal distribution, with a slight skew to the right (see Figure 10).

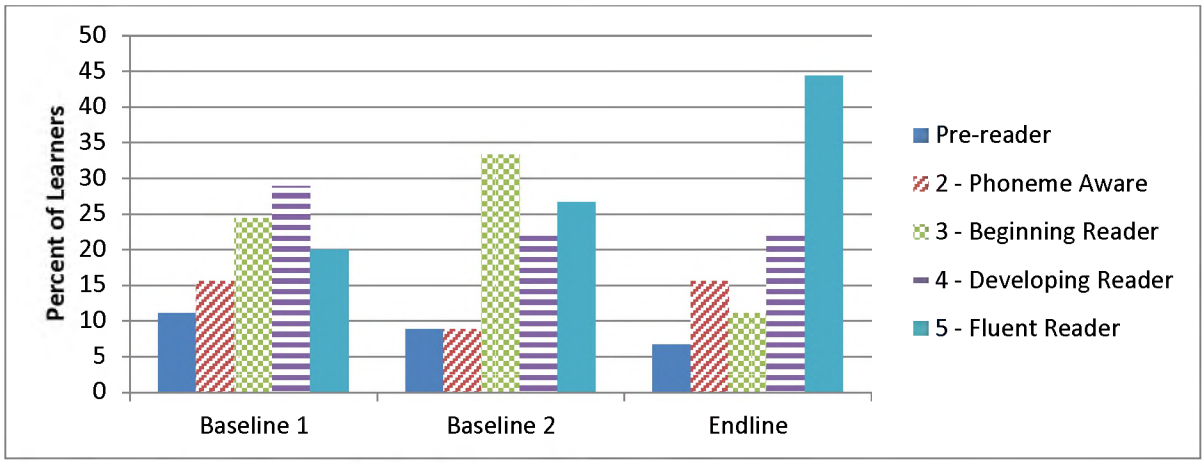


Figure 10: Attainment Level Distribution over Time

Considering the test was pitched at the late grade 1 or early grade 2 level, one would expect a majority of grade 3 learners to perform at the developing or fluent reader stage (attainment levels 4 and 5), and for a majority of grade 2 learners to at least manage some portion of the text. However, 51 percent of learners performed below the developing reader level in the first baseline, and 27 percent could not read even individual words.

By the second baseline, the percentage of learners unable to read words (level 1, basic readers, and level 2, phonemically aware) had decreased to 18 percent, and a bulge developed in the beginning reader level (3) as learners gained basic word decoding ability. Some further shifts were seen between developing and fluent readers as reading speed overall increased.

By the endline, a majority of learners had shifted out of the beginning reader level, resulting in a highly skewed distribution which favoured developing and especially fluent readers.

The results of the examination of beginning readers lent itself to a corollary investigation: How accurate was the assumption of a straight-line progression from basic readers to fluent readers on the part of individual learners? Table 10 answers this question by plotting the movement from pre to post-test scores from the intervention and control periods. Learners in each attainment level on the first baseline were tracked to see which attainment level they achieved on the second baseline, and learners at each attainment level on the second baseline were tracked to see which attainment level they achieved on the endline. For example, of the five learners in attainment level 1 (pre-readers) on the first baseline, in the second baseline three remained in the pre-reader level, one progressed to phonemically-aware, and one progressed to basic decoding. The shaded diagonal represents no change in attainment level between tests.

Table 10: Change in Attainment Levels of Learners

	Baseline 2 Level					
Baseline 1 Level	Attainment Level	1	2	3	4	5
	1	3	1	1		
	2	1	1	5		
	3		2	7	1	1
	4			2	7	4
	5				2	7

	Endline Level					
Baseline 2 Level	Attainment Level	1	2	3	4	5
	1	1	3			
	2		2		1	1
	3	2*	2	5	3	3
	4				5	5
	5				1	11

Overall, in the control period (baseline 1 to baseline 2) 56 percent of learners remained within the same attainment level between the two baselines, while 29 percent advanced to a subsequent level (85 percent of these learners advancing to the attainment level immediately subsequent) and 16 percent regressed, in all cases to the attainment level immediately below that achieved in the first baseline.

However, when the treatment period is considered a somewhat different pattern emerges. While 53 percent of learners remained within the same attainment level, 36 percent progress and 11 percent regressed, 31 percent of learners progressing skipped at least one attainment level entirely, indicating exceptional individual variance (compared to 15 percent in the control period). Additionally, 40 percent of learners regressing 'lost' previously held phonemic assignment ability (a total of 4 percent of the population). As indicated by the distribution charts, the beginning reader category emptied by the endline and a clear distinction emerged between those who could read and those who could not, with 27 percent of the study population remaining in the latter category.

In what seemed initially a surprising result, some learners 'lost' word decoding ability and regressed to the phonemically aware or even basic reader levels between testing periods (see Table 10). Of learners who scored within the beginning reader category in the second baseline, by endline 27 percent regressed, 33 percent remained within the category, and 40 percent progressed. However, all regressing learners were at the lower end of the emerging reader scale, as none were able to read more than 10 words with errors accounted for, indicating that the differing attainment level results for those regressing one level could be due to incidental variance that overlapped the category demarcation lines.

The two cases of learners who regressed more than one attainment level on the endline, indicated by the asterisk, are worth further study. One learner moved from basic level to beginning reader and then back to basic, which could be due to context factors such as illness on the day of testing, or it could indicate a genuine loss of ability. The other learner showed an abnormal pattern of development; in the first baseline she was able to assign phonemes to graphemes but not read words, in the second baseline she could read words but not assign phonemes, and in the endline she was unable to perform either task. This learner constituted the sole example in the entire dataset of a learner who at any point read more words than they were able to assign phonemes. The reasons are unclear without further study, but could be due to anything from disinterest to a progressive developmental disorder.

4.3.2 Analysis by Age

One of the research questions asked whether age affected the efficacy of the program. The age of learners participating ranged from 7 to 11 years, with learners in grade 2 spanning from 7 to 10 years of age and learners in grade 3 spanning from 8 to 11 years of age. A regression analysis at the alpha .05 level across both grades with age as the independent

variable shows that age is not significantly correlated to overall performance on any of the EGRA tasks.

Regression: Age x Variable	DEPENDENT VARIABLE	Overall Performance	Comprehension	CWPM	Individual Words	Letters
	multiple r	.2642	.2463	.2504	.2054	.2700
	p	.0795	.1028	.0971	.1759	.0728

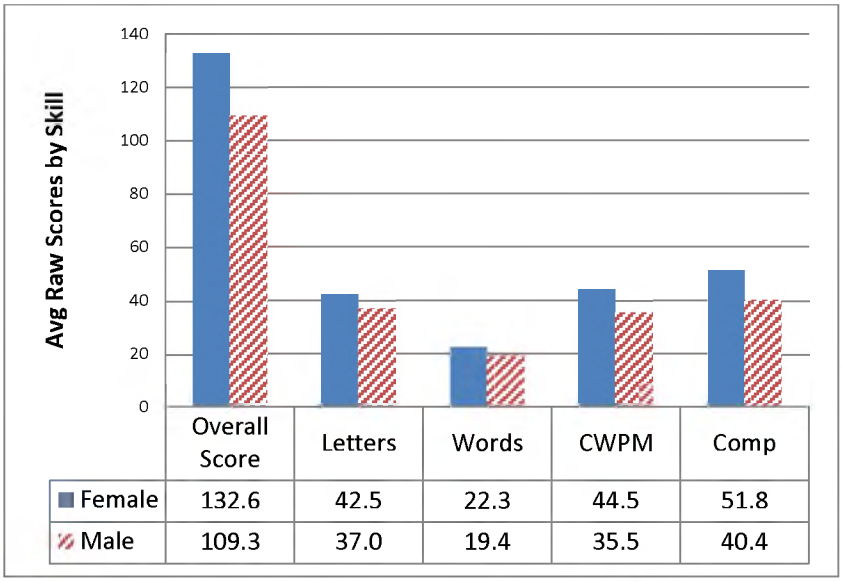
When age is considered as the independent variable in a regression equation with change in performance during the treatment period considered as the dependent variable, correlations are even less significant.

Regression: Age x Variable	DEPENDENT VARIABLE	Change Overall Performance	Change Comprehension	Change CWPM	Change Individual Words	Change Letters
	multiple r	.1112	.0061	.1199	.1457	.1770
	p	.4670	.9681	.4329	.3397	.2449

The results indicate that age is not a significant predictor of either performance or improvement during the treatment period. Therefore, in answer to the research question regarding ideal age of engagement, as far as this study can determine there is no ideal age at which to begin interacting with technology within the age range included in the study, 7 to 11 year old Northern Sotho learners.

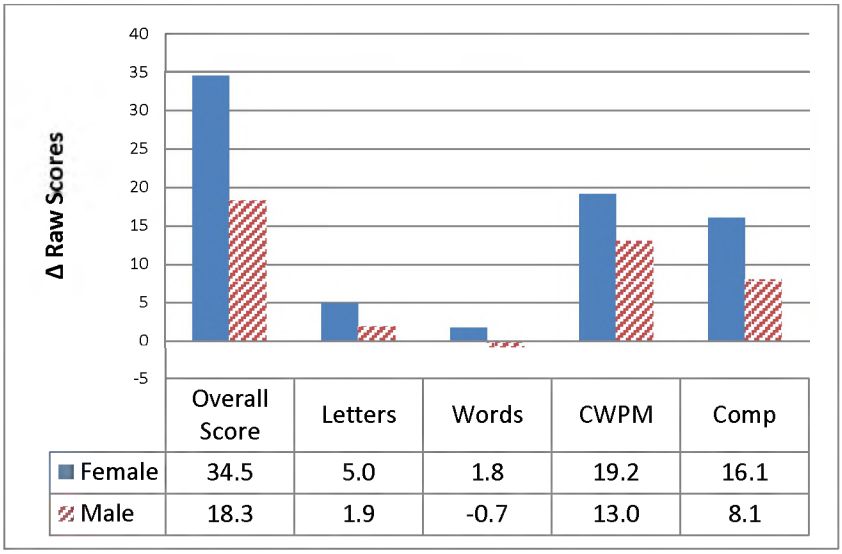
4.3.3 Analysis by Gender

Girls made up 63 percent of the population, and outperformed boys on every category of the EGRA assessment. Girls also gained more during the treatment period than boys. However, gender was not a significant predictor of either performance ($r = .1843$; $p = .2257$) or change in performance during the treatment period ($r = .2333$; $p = .1230$). Figure 11 and Figure 12 show the average scores by gender and the change in scores by gender during the treatment period, respectively.



	Overall Score	Letters	Words	CWPM	Comp
Female	132.6	42.5	22.3	44.5	51.8
Male	109.3	37.0	19.4	35.5	40.4

Figure 11: Average Score on Literacy Skills by Gender[18]



	Overall Score	Letters	Words	CWPM	Comp
Female	34.5	5.0	1.8	19.2	16.1
Male	18.3	1.9	-0.7	13.0	8.1

Figure 12: Average Change in Score by Gender

4.3.4 Hours of BFI Program Use and improvement in EGRA skills

Overall, hours of program use showed a significant but weak negative correlation with changes in performance during the treatment period (r = -.1215; p = .0122). The significance disappeared as the results were considered by school, with an insignificant negative correlation in school A (r = -.129; p = .2927), and an insignificant positive correlation in school B (r = .4877; p = .8788). When the same analysis was performed by grade, a negative correlation appeared for grade 2 which approached significance (r = .2594; p = .0681), while a positive correlation for grade 3 did not approach significance (r = -.0588; p = .1471).

[18] No correlations were significant

As discussed in in section 2.2.1, school A overall had lower performance than school B, although school A had lower learner attrition and retained a higher percentage of the sample over time. School A also had higher hours of program use in both grades, which influenced the outcome of the overall correlation. Additionally, there was a high degree of variance in learner performance in both schools in both grades. Finally, as only two schools and 12 classes were included in the study, there are a small number of data points available. These factors contribute to a conclusion that, although the overall correlation is significant, individual school factors and a low number of data points make that finding questionable.

The lack of correlation between program hours and overall performance does not necessarily indicate that the program is ineffective; rather, it indicates that factors beyond hours of program use influence its effectiveness, an idea which is also supported by the high degree of variance in performance changes.

To provide evidence for or against program effectiveness in improving each literacy skill under study, a comparison of gains during the control and treatment periods was performed for each skill tested by the EGRA examination. The mean results of these investigations and a comparison of gains between the control and treatment periods is provided in Table 11. Standard deviations, which indicate the variance in performance of each group, are provided in brackets. Table 12 shows the change in performance by attainment level, with standard deviations in brackets. Significance was established using a paired sample t-test to compare the means of changes in scores between the control and treatment periods. The results summarized in these tables will be presented in detail in each subsequent section, and examined in conjunction with results from the retention tests and qualitative data in the discussion section.

Table 11: Mean and Standard Deviation Results by Grade, School and Test Item[19]

Avg Scores by school and grade

	GRADE 2						GRADE 3					
	SCHOOL A avg use: 23.6 hrs			SCHOOL B avg use: 7.7 hrs			SCHOOL A avg use: 16.7 hrs			SCHOOL B avg use: 2.8 hrs		
	Baseline1	Baseline2	Endline	Baseline1	Baseline2	Endline	Baseline1	Baseline2	Endline	Baseline1	Baseline2	Endline
grapheme-phoneme assignment	14.1 (13.6)	23.3 (19.6)	24.3 (19.3)	34.4 (7.3)	46.3 (15.4)	51.8 (25.5)	38.8 (9.0)	48.1 (13.5)	48.8 (18.5)	40.0 (13.0)	55.1 (15.4)	67.3 (20.4)
individual word reading	5.5 (10.7)	7.0 (9.1)	7.9 (14.5)	14.1 (10.6)	27.0 (12.2)	28.4 (13.3)	21.1 (12.8)	25.0 (14.2)	25.1 (13.9)	25.6 (12.9)	35.1 (12.6)	36.8 (12.4)
correct words per minute	7.9 (19.5)	6.3 (18.2)	12.7 (28.4)	26.9 (26.1)	31.7 (23.7)	63.2 (21.5)	37.3 (25.5)	48.5 (32.1)	71.7 (30.0)	61.1 (29.9)	65.5 (28.6)	76.9 (31.1)
comprehension	7.1 (20.8)	8.6 (21.0)	18.6 (36.6)	25.7 (41.0)	51.4 (18.1)	74.3 (25.6)	51.4 (36.8)	62.9 (30.1)	80 (23.9)	60.0 (33.5)	76.0 (32.0)	80.0 (29.7)

Change in avgerage scores

	GRADE 2				GRADE 3			
	SCHOOL A		SCHOOL B		SCHOOL A		SCHOOL B	
	change control	change treatment	change control	change treatment	change control	change treatment	change control	change treatment
grapheme-phoneme assignment	9.2 (12.9)	0.9 (14.0)	11.9 (9.8)	5.4 (20.5)	9.3 (17.0)	0.7 (21.7)	15.2 (16.4)	12.2 (23.6)
individual word reading	1.5 (5.0)	0.9 (11.2)	12.9 (8.4)	1.4 (5.0)	3.4 (8.5)	0.0 (7.8)	9.5 (5.4)	1.7 (4.7)
correct words per minute	-1.5 (4.5)	6.4 (17.8)	4.8 (22.0)	31.5 (25.9)	11.3 (20.0)	23.2 (23.3)	4.5 (14.2)	11.3 (13.9)
comprehension	1.4 (11.9)	10.0 (27.0)	25.7 (31.6)	22.9 (27.1)	11.4 (18.1)	17.1 (26.0)	16.0 (17.4)	4.0 (47.2)

* $p < .05$ ** $p < .01$ (note there are no significant differences in the means for any group on any test area)
Shaded items show larger gains in the treatment period than the control period

Table 12: EGRA Performance by Attainment Level and Task[20]

Grapheme-Phoneme Assignment

Attainment Level	baseline1	baseline2	endline	CHANGE CONTROL	CHANGE TREATMENT
1	4.3 (4.1)	2.1 (1.2)	10.1 (6.7)	-2.3 (3.1)	8.0* (5.8)
2	21.6 (13.9)	42.0 (16.9)	48.1 (21.3)	20.4* (5.8)	6.1 (10.6)
3	28.5 (13.3)	38.3 (17.8)	36.9 (25.5)	9.8 (14.0)	-1.4 (23.9)
4	37.9 (15.7)	57.6 (13.2)	57.3 (15.1)	19.7* (11.3)	-0.4 (15.5)
5	39.2 (8.8)	45.8 (14.1)	58.3 (23.5)	6.5 (17.2)	12.5 (21.7)

Correct Words Per Minute

Attainment Level	baseline1	baseline2	endline	CHANGE CONTROL	CHANGE TREATMENT
1	-1.5 (0.0)	-1.5 (0.0)	-1.5 (0.0)	0.0 (0.0)	0.0 (0.0)
2	1.3 (4.1)	-0.2 (0.9)	29.3 (30.7)	-1.5 (3.5)	29.5 (30.5)
3	7.6 (14.6)	11.8 (13.7)	33.5 (30.7)	4.2 (21.3)	21.7 (26.2)
4	49.6 (13.0)	49.7 (7.0)	69.0 (19.4)	0.1 (11.1)	19.3*(18.6)
5	69.1 (16.4)	82.4 (17.4)	91.7 (24.1)	13.3 (15.8)	9.4 (13.5)

* $p < .05$ ** $p < .01$

Shaded items show larger gains in the treatment period than the control period

Individual Word Reading

Attainment Level	baseline1	baseline2	endline	CHANGE CONTROL	CHANGE TREATMENT
1	-0.5 (1.0)	-1.1 (0.6)	-0.6 (1.1)	-0.6 (1.4)	0.5 (1.3)
2	1.8 (3.1)	1.4 (2.5)	12.4 (15.9)	-0.4 (3.4)	11.1 (15.5)
3	8.3 (9.8)	14.3 (7.6)	12.7 (12.1)	6.0** (8.4)	-1.7 (6.2)
4	22.1 (10.2)	26.5 (5.5)	28.1 (9.7)	4.4 (7.3)	1.6 (8.4)
5	31.3 (7.7)	42.3 (7.8)	42.5 (6.6)	11.0 (7.1)	0.2 (4.8)

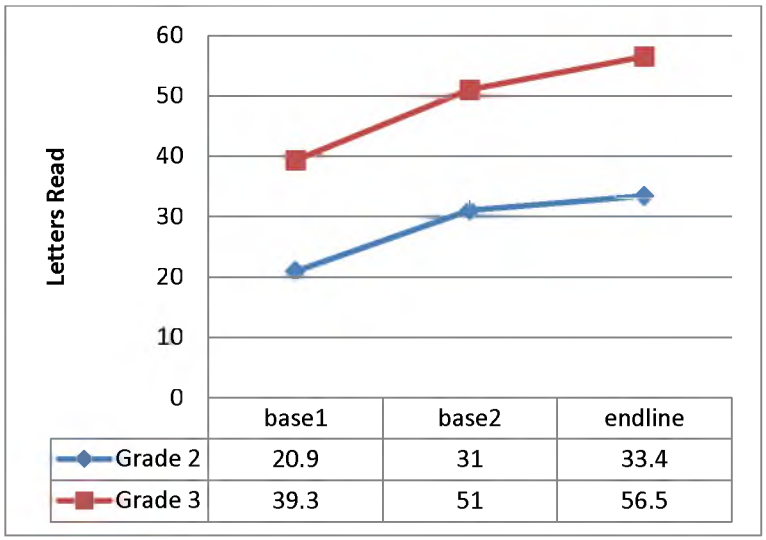
Comprehension

Attainment Level	baseline1	baseline2	endline	CHANGE CONTROL	CHANGE TREATMENT
1	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)
2	0.0 (0.0)	0.0 (0.0)	45.0 (45.6)	0.0 (0.0)	45.0 (45.6)
3	2.7 (10.0)	25.3 (23.6)	41.3 (35.4)	22.7 (24.1)	16.0 (28.5)
4	56.0 (30.0)	60.0 (8.9)	90.0 (16.1)	4.0 (19.6)	30.0* (16.1)
5	83.3 (18.0)	95.0 (8.7)	83.3 (28.1)	11.7 (17.2)	-11.7 (30.0)

4.3.5 Grapheme-Phoneme Assignment

The first section of the EGRA test examined grapheme-phoneme assignment, the ability to assign phonemes to graphemes or "read letters". Participants were given instructions to make the sounds of the letters presented on a page. (It must be noted that in Northern Sotho the letters are called by their sounds so there is no difference between the name of the letter and its sound.)

In both grade 2 and 3, gains in the control period surpassed gains during the treatment period, as shown in Figure 13. However, a paired sample t-test comparing mean gains in the control and treatment periods approached but did not achieve significance ($t(44) = 1.6$; $p = .0622$).



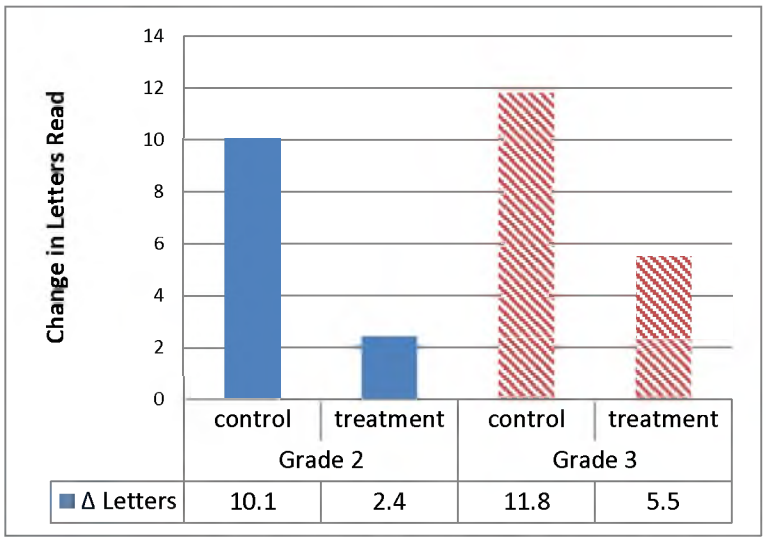


Figure 13: Performance on Grapheme-Phoneme Relationships over Time

The range of scores on this item was 61.5 in baseline 1 (30.7 mean), 78 in baseline 2 (41.7 mean) and 112.5 in the endline (45.7 mean), indicating that by the second baseline more than

50 percent of learners were reading letters at a speed of 1.5 seconds or faster per letter, which indicates a degree of automaticity had been reached before the treatment period began.

An analysis of the distribution of learner scores at each testing period (Figure 14) also shows a gradual flattening of an initial skew to the left, with a bimodal distribution at the endline.

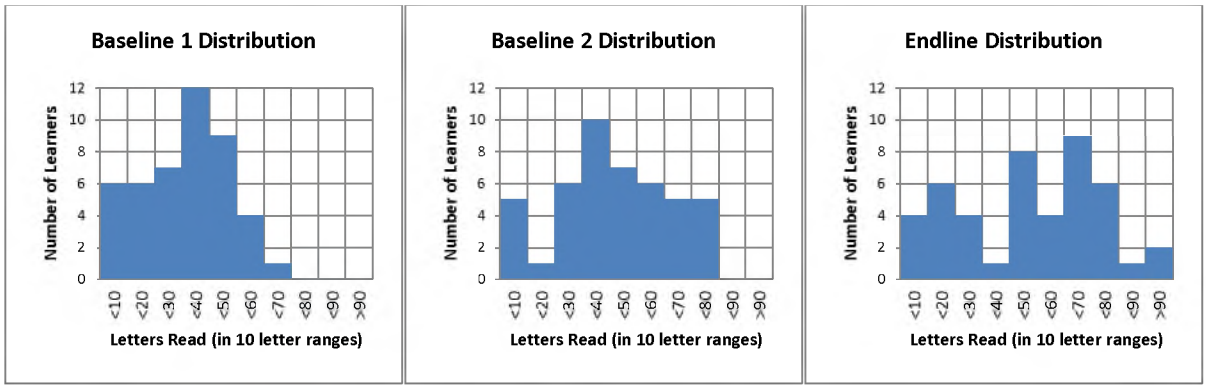
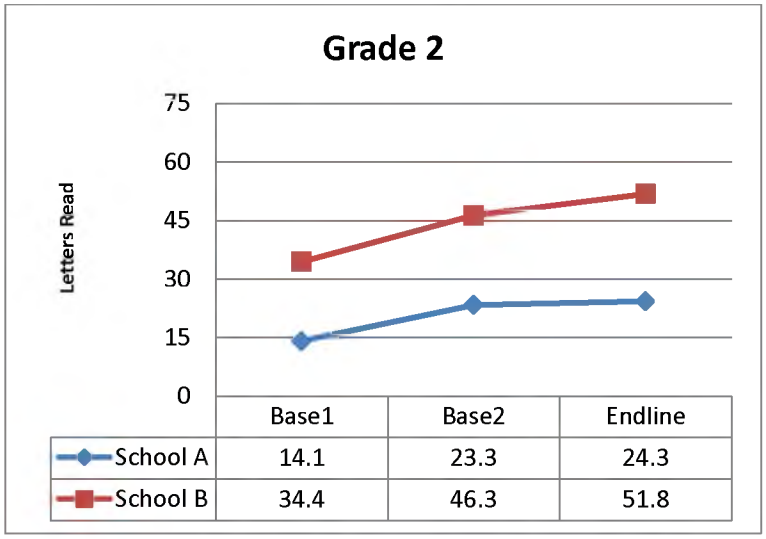


Figure 14: Changes in Distribution Patterns, Grapheme-Phoneme Assignment

An analysis by school and grade shows that the bimodal break is not between schools or between grades, but between grade 2 of the lower-performing school and all other groups, as can be seen in Figure 15. In all three testing periods, the grade 2 learners of the higher performing school approach the averages of the grade 3 learners in the lower performing school, while the grade 2 of the lower performing school lagged an average of between 15 and 20 points behind.



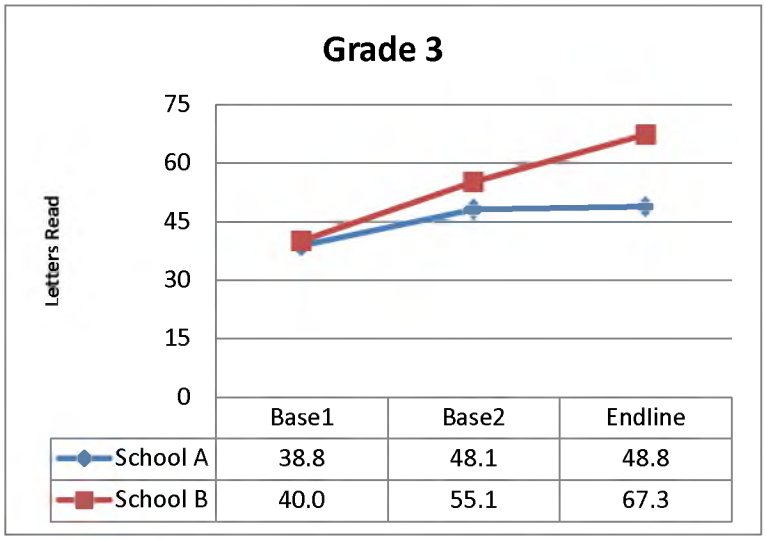


Figure 15: Performance on Grapheme-Phoneme Assignment over time, by School and Grade

The changes in distribution scores indicate continuous improvement in grapheme-phoneme assignment, though some sizable regressions during the treatment period affected the overall mean scores. The cause of regressions is not determinable. However, 71 percent of learners scoring the top 15 percent in the control period had negative gains in the treatment period, which may indicate that boredom with the task played a factor in negative gains. Only one learner had a negative change in both periods, indicating that 98 percent of learners improved in phonemic assignment automaticity in the year, as demonstrated by Figure 16.

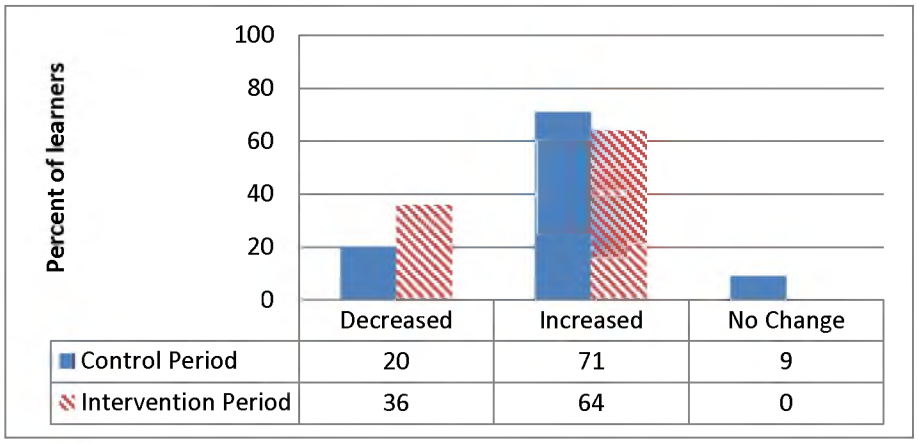
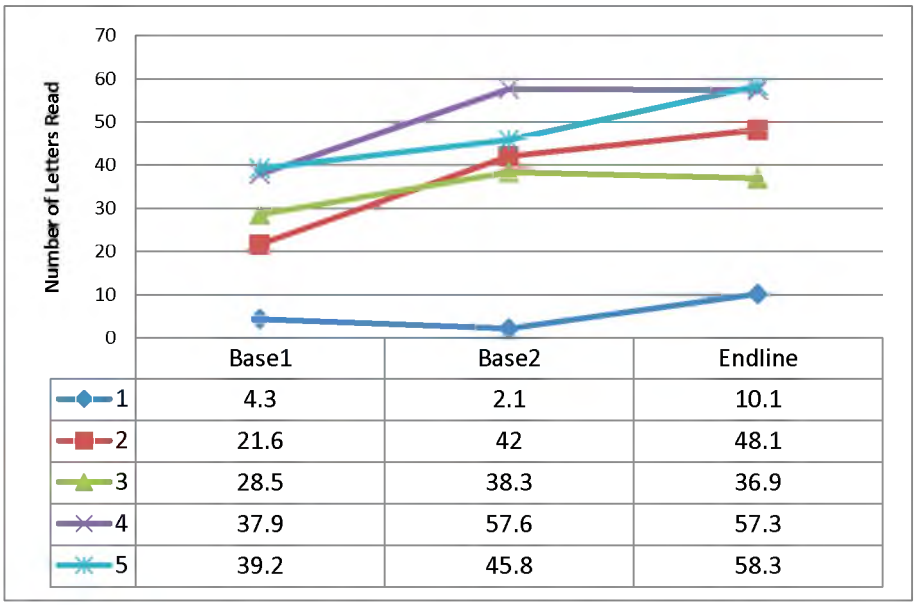


Figure 16: Changes in Learner Performance, Grapheme-Phoneme Assignment

Analysis by Attainment Level

Figure 17 shows that most attainment groups made higher gains in the control period. The only exceptions were the fluent reader attainment level, which made steady gains throughout the year, and learners in the pre-reader level, who showed negative gains in the control period but more than quadrupled their average score in the treatment period. When the null hypothesis that gains in the control and intervention periods would be equivalent was tested

at the alpha = .05 level by a paired sample t-test, the result showed significantly larger gains in the intervention period for pre-reading learners ($p < .05$).



	Base1	Base2	Endline
1	4.3	2.1	10.1
2	21.6	42	48.1
3	28.5	38.3	36.9
4	37.9	57.6	57.3
5	39.2	45.8	58.3

Figure 17: Grapheme-Phoneme Assignment by Attainment Group

Therefore, while the overall analysis shows that learners did not improve performance in phoneme-grapheme assignment during the intervention period, learners in the pre-reader attainment level, or those who were unable to assign at least 10 phonemes to 10 graphemes in one minute and were unable to read words, did improve their performance more in the intervention period. This may indicate that for some learners, specifically remedial learners operating below grade level, the technology-based BFI instruction was more effective than over one year of traditional literacy instruction.

In answer to research question 1, part a, the conclusion of the analysis shows that while program use overall does not improve grapheme-phoneme assignment ability, for very low-performing learners program use is more effective than a typical teaching environment.

4.3.6 Simple Decoding (Word Reading)

The second section of the test required participants to decode and read individual words. Participants were instructed to read as many words as possible in a minute after being shown two examples. Correct words were recorded. Learners were stopped after six consecutive errors.

Gains during the control period outstripped gains during the treatment period in individual word reading (see Figure 18). Learners made average gains of 5.9 cwpm in the control period, compared to an average gain of 0.8 cwpm in the treatment period. A paired sample t-test to

compare the mean gains in control and treatment periods shows significantly higher improvement in the control period (t(44) = 2.7; p = .0051). If ceiling effects are removed from the sample[21], the control mean gain drops to 4.3 words, while the treatment mean increases to 1.1 words; a paired sample t-test shows that the difference in mean gains is still very near significance (t(38) = 1.7; p = .0507).

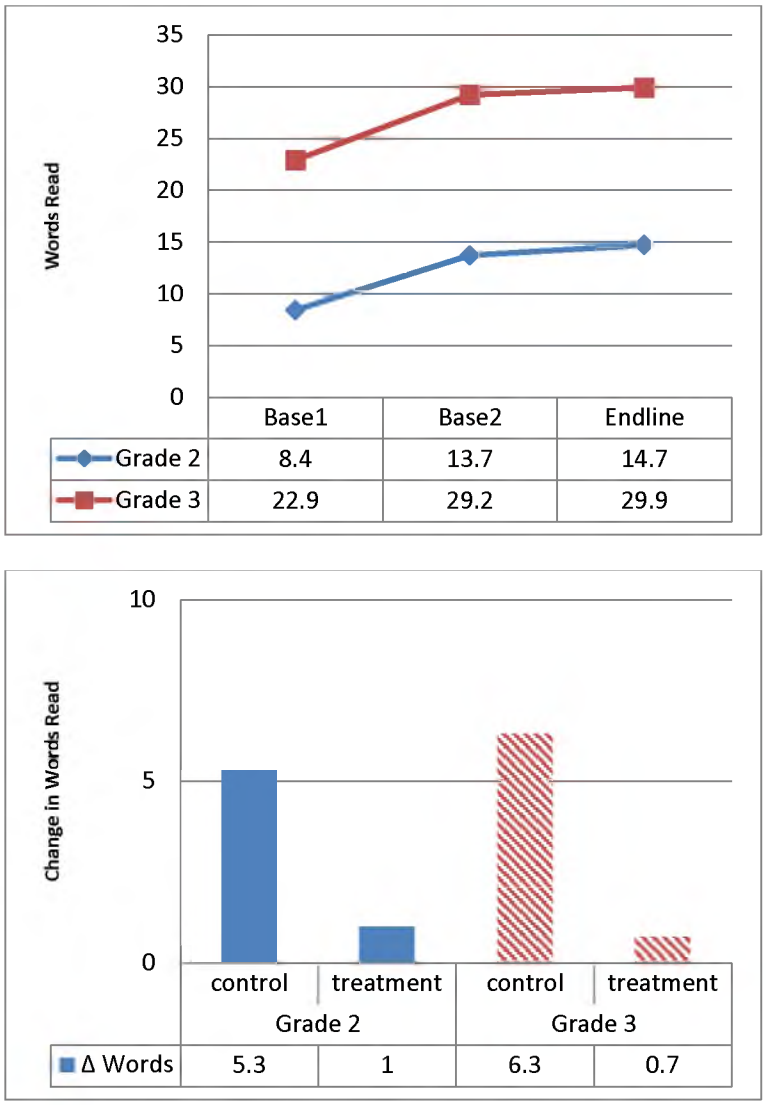


Figure 18: Simple Decoding over Time

When the distribution of learners over the three testing periods is analysed (see Figure 19), a bimodal distribution emerges at the beginning of the year, with a third of the study population reading less than five individual words per minute. The second baseline shows an abnormal distribution, with spikes at the low and high end of the spectrum and a somewhat normal distribution curve in the centre. In the endline a bimodal distribution similar to the initial

[21] Ceiling effects for word reading are defined as learners who could not read more words in the treatment period than the control period due to high performance at the end of the control period.

pattern but shifted to the right emerges. The number of learners reading less than five words dropped in the control period from 15 to 9, but increased from 9 to 11 in the treatment period.

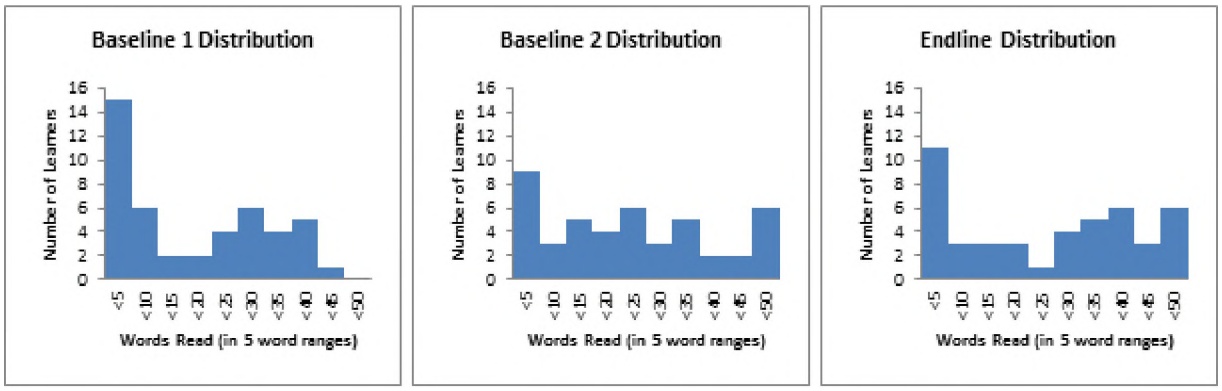
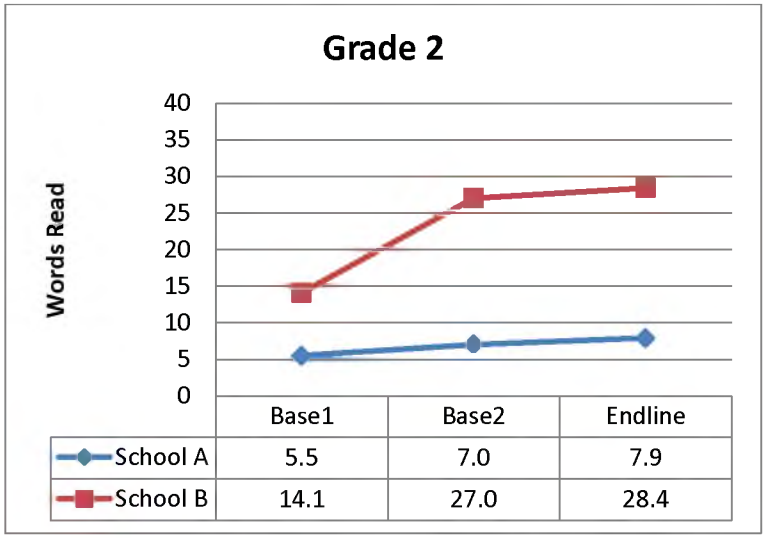


Figure 19: Changes in Distribution Patterns, Word Reading

Initially, the bimodal nature of the distribution on this task was due to grade level, with grade 2 learners averaging below 15 cwpm and grade 3 learners averaging above 20 cwpm. By the second baseline, the same split emerged as found in the grapheme-phoneme assignment task, with grade 2 in the lower performing school significantly behind all other grades. This deficit remained through the endline. However, the bimodal distribution emerging at the endline actually separates the higher performing school's grade 3 learners from the other three classes, all of which averaged below 30 cwpm. This is demonstrated by Figure 20, which shows longitudinal performance by school and grade.



	Base1	Base2	Endline
School A	5.5	7.0	7.9
School B	14.1	27.0	28.4

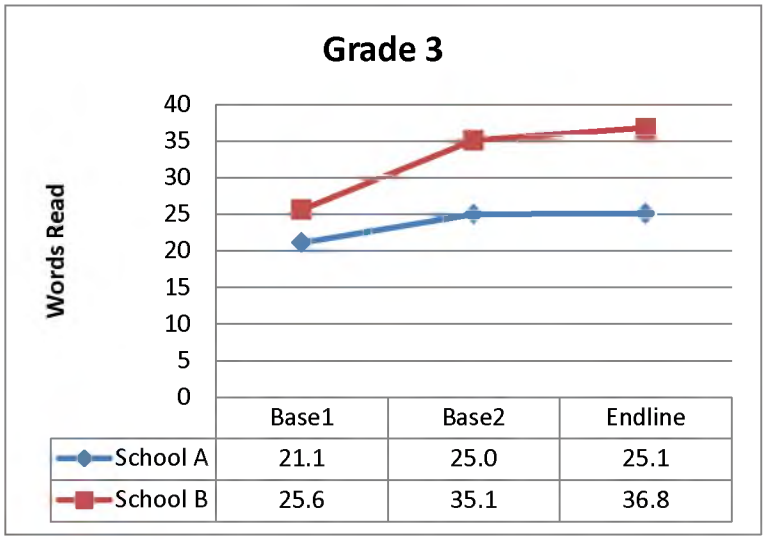


Figure 20: Performance on Word Reading over Time, by School and Grade

Individual word reading was one of the most variable skills, with 18 percent of learners dropping in performance in the control period and 36 percent dropping in the treatment period. The majority of gains were made in the control period in this exercise.

Low results in the treatment period warranted further investigation, so the dataset broken by school and grade was further analysed to determine if other factors might have played a role in the substantial amount of regression. However, although overall school B showed higher gains than school A, for both schools in both grades the outcome was similar: gains in average performance in the control period and smaller or no average gains in the treatment period. The near-exception was grade 2 in school A, which had small gains even in the control period; however this group also showed lesser gains in the treatment period. Table 13 provides evidence indicating that overall, traditional teaching may have been stronger than the BFI program in teaching learners to read individual words.

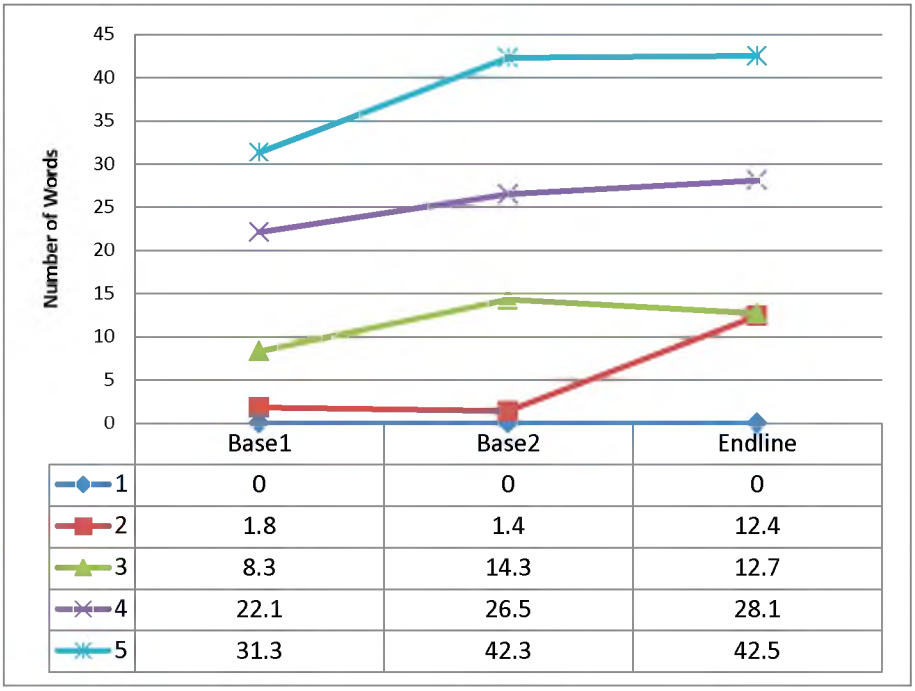
Table 13: Gains in Word Reading by School and Grade, Control and Treatment period

		Control Period	Treatment period
School A	Grade 2	1.5	0.9
	Grade 3	3.9	0
School B	Grade 2	12.9	1.4
	Grade 3	9.5	1.7

Analysis by Attainment Level

When learner performance is analysed by attainment level (see Figure 21), the only groups to make higher gains in word reading in the treatment period than the control period were

learners who began the intervention phonemically aware (attainment level 2). Basic readers were not able to achieve word reading in the year, and gains in this task for beginning, developing and fluent readers were negligible in the treatment period after gains in the control period. However, phonemically aware learners, a group that had mostly stagnated in this ability in the first half of the year, were able to triple the number of words read in the treatment period. Nevertheless, this result did not approach significance ($p = .1565$). Therefore, it may be concluded that the skill of simple decoding in Northern Sotho was not significantly improved by use of the literacy through technology program under study.



	Base1	Base2	Endline
1	0	0	0
2	1.8	1.4	12.4
3	8.3	14.3	12.7
4	22.1	26.5	28.1
5	31.3	42.3	42.5

Figure 21: Changes in Word Reading by Attainment Level

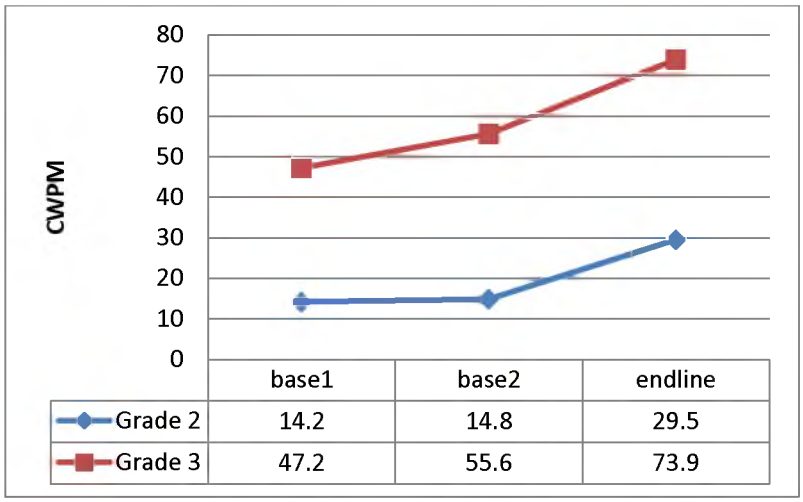
Close evaluation of the data shows dramatic gains for half of the learners in the phonemically-aware attainment level during the treatment period, from phonemically-aware to developing or fluent level, while the remaining half showed no change in attainment level. As all learners in the phonemically-aware group were from the same school, and 75 percent in the same class, it is unclear what factors contributed to advancement.

4.3.7 Correct Words per Minute

The third element of the EGRA test required participants to read a story aloud independently. The story included a total of 66 words and was written on a late first-early second grade

reading level[22]. The total number of words read correctly within a minute as well as the number of errors within the same minute was recorded by trained Northern Sotho first language speakers. An error was counted if a learner grossly mispronounced, substituted, omitted or added a word into the passage without self-correcting. Learners were stopped after six consecutive errors. However, in a deviation from a typical EGRA administration, learners were not stopped after the minute had elapsed, but were allowed to complete the passage in their own time (words and errors read after the minute had elapsed were not recorded). This was done to guard against reading speed correlating to comprehension only because slower readers did not access parts of the passage relating to some comprehension questions.

Learners in both grades gained substantially in passage reading during the treatment period as compared to the control period, as shown by Figure 22. Overall, average reading speed increased by 4.8 cwpm in the control period, and 16.6 cwpm in the treatment period. A paired sample t-test found significantly higher gains in the treatment period ($t(44) = -2.5$; $p = .0084$).



	base1	base2	endline
Grade 2	14.2	14.8	29.5
Grade 3	47.2	55.6	73.9

[22] Currently there are no reliable readability tests for Northern Sotho. The reading level was determined by teachers and curriculum specialists at the Molteno Institute for Language and Literacy.

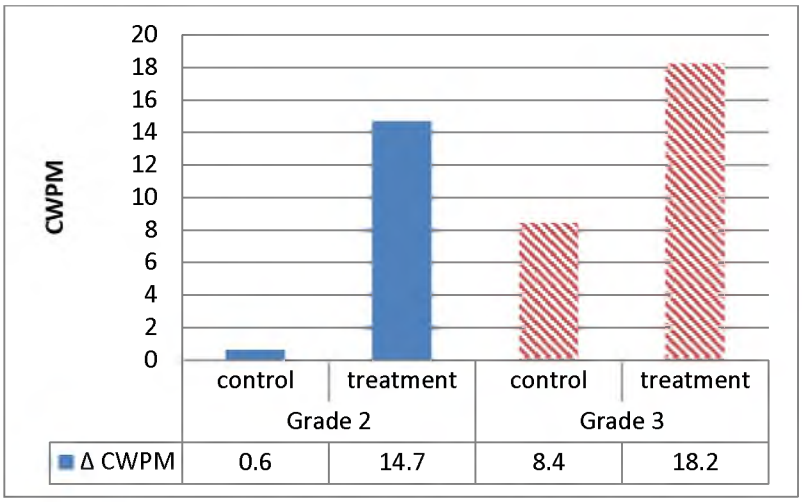


Figure 22: Performance on Passage Reading (cwpm) over time

This indicates that the BFI program may be effective in improving reading speed. However, the average performance at the end of grade 2 did not approach the opening performance of grade 3 learners. This is likely due to high repetition rates in early grades in South African schools, as only those learners at the higher end of performance will advance to grade 3.

When distribution patterns are analysed (see Figure 23), it emerges that a bimodal distribution exists in the first baseline, with over 40 percent of learners unable to read 10 cwpm at one end, and a second group with a slight skew to the right with a mean of 58 cwpm after error is accounted for. By the second baseline, the curve had flattened, but still with more than 35 percent of learners unable to read 10 cwpm, with the remainder of learners achieving a mean of 59 cwpm. By the endline, the percentage of learners who could not read at least 20 cwpm had dropped to 24 percent, with the remainder averaging 71 cwpm.

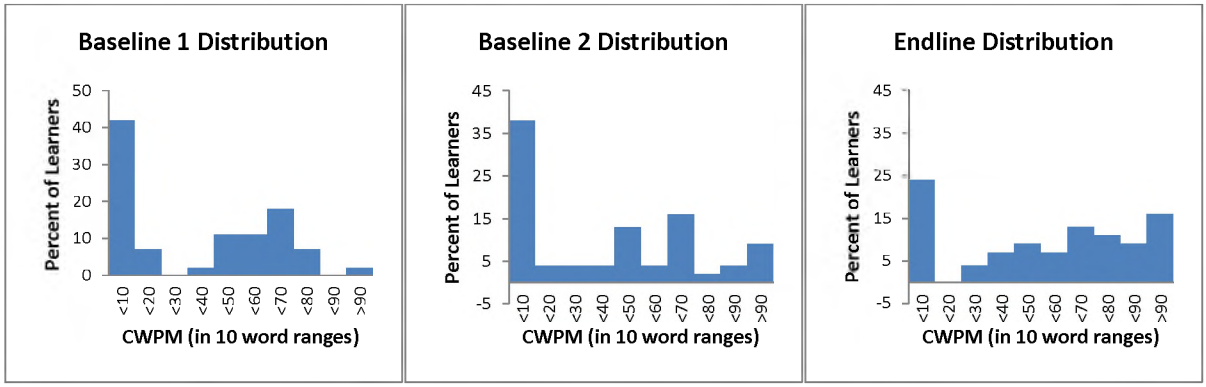
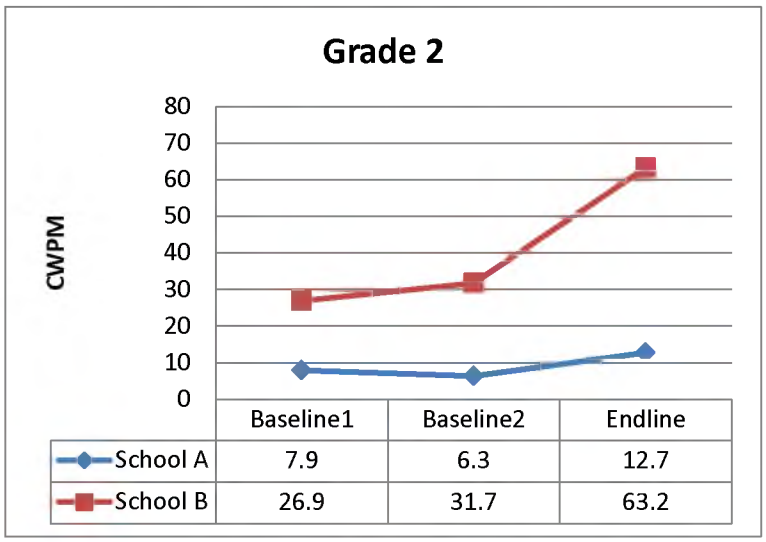


Figure 23: Changes in Distribution Patterns, Correct Words Per Minute (CWPM)

The bimodal distribution evident in the initial baseline was largely attributable to grade. Grade 2 learners averaged below 25 cwpm and grade 3 learners in both schools averaged above 35 cwpm. Discrepancies between grades remained throughout the year. On average at

endline, grade 3 learners performed at 70 cwpm. The grade 2 learners of the higher-performing school performed at just under 50 cwpm. The grade 2 learners of the lower performing school made the smallest gains, ending with an average passage reading speed of less than 15 cwpm (see Figure 24).



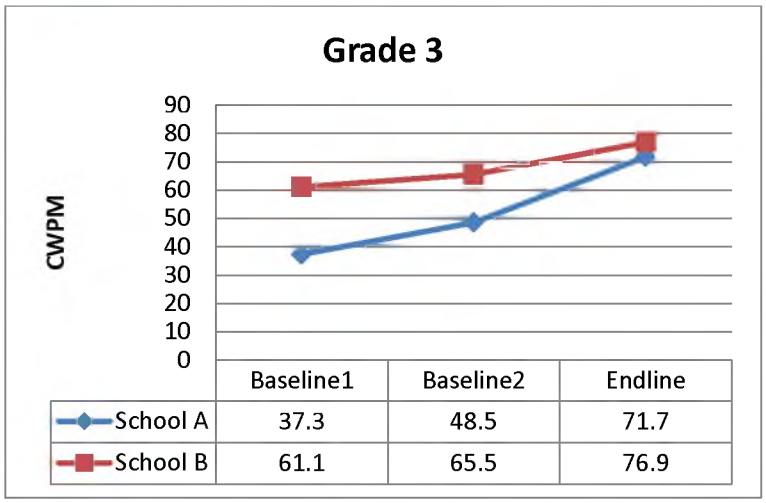
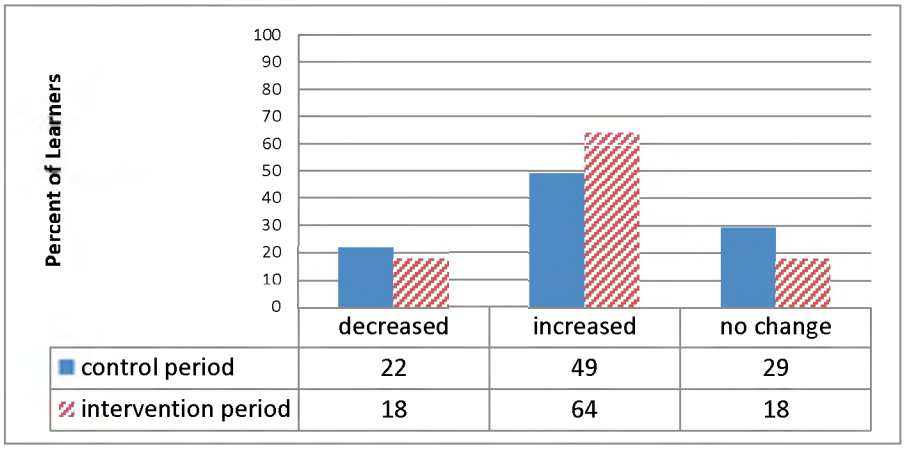


Figure 24: Performance on Reading over Time, by School and Grade

As shown in Figure 25, 64 percent of learners were able to improve their passage reading speed in the treatment period, compared to 49 percent in the control period. If only the 64 percent of learners with gains are considered, reading speed increased by an average of 27 cwpm in the treatment period. If only the 49 percent of learners with gains in the control period are considered, passage reading speeds increased by 17 cwpm. Therefore, more learners were able to make higher average gains in the treatment period.

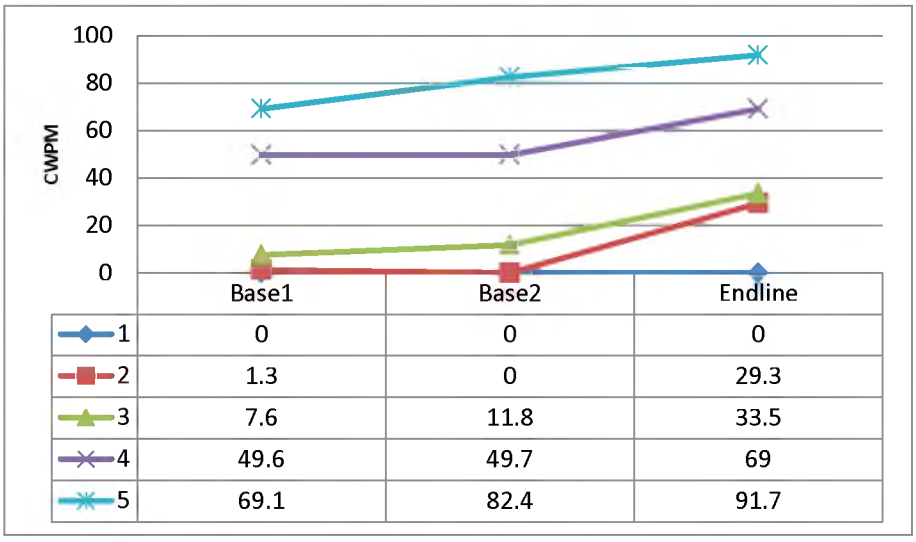


	decreased	increased	no change
control period	22	49	29
intervention period	18	64	18

Figure 25: Change in Learner Performance, Passage Reading

Analysis by Attainment Group

Learners at the lower end of the spectrum in level 1 remained unable to crack the literacy code and read the passage, and learners at the high end of the spectrum in level 5 showed consistent gains through the year. However, learners from the phonemically aware, beginning reading and developing reader groups gained substantially more during the treatment period than the control period. During the treatment period the phonemically aware group gained an average of 29.3 cwpm, approaching the beginning reader endline average of 33.5 cwpm, despite gains of 21.7 cwpm by beginning readers in the treatment period. However, paired sample t-tests showed that only the gains in attainment level 4 were significant (p =.0134). This is due to high variance among small sub-populations for the other two groups with gains. Figure 26 demonstrates changes in cwpm by attainment group.



	Base1	Base2	Endline
1	0	0	0
2	1.3	0	29.3
3	7.6	11.8	33.5
4	49.6	49.7	69
5	69.1	82.4	91.7

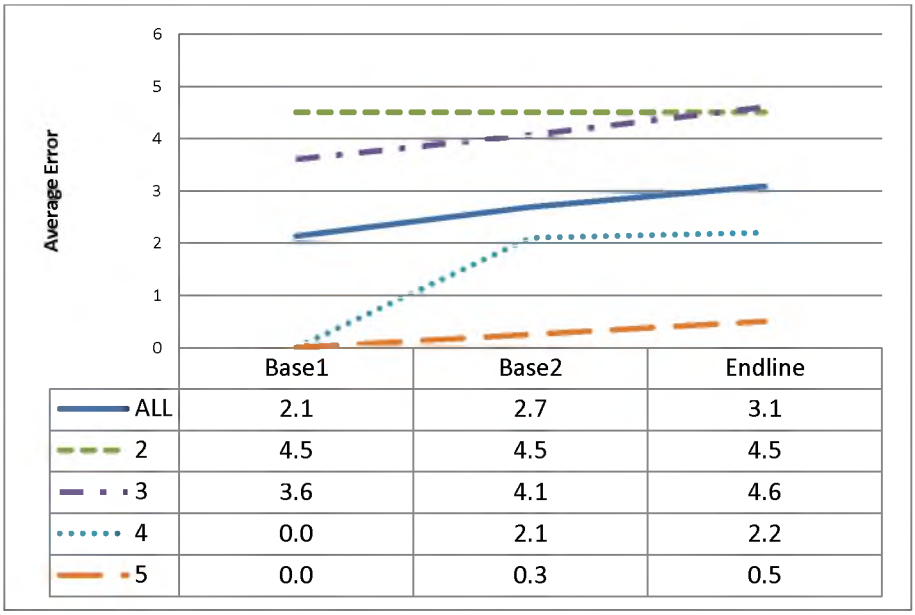
Figure 26: CWPM by Attainment Group

Nevertheless, gains were substantial in the treatment period, and significant for the overall test population (p = .0084). It is therefore the finding of this paper that use of the technology

program did correlate significantly to increased gains in fluency as measured by correct words per minute on the part of the study population. However, further research with larger sample sizes would be necessary to establish the attainment levels with significant gains.

Analysis of Error and Correct Words per Minute

Figure 27 demonstrates the error rates over time by attainment level. Overall, error increased slightly in the control period and again in the treatment period, with the largest increase by developing readers in the control period. The expectation of higher error in lower attainment groups was met, although all groups increased in error. Learners with a zero score in passage reading were omitted from regression calculations, as a high number of 0-0 scores would inaccurately strengthen correlations.



	Base1	Base2	Endline
ALL	2.1	2.7	3.1
2	4.5	4.5	4.5
3	3.6	4.1	4.6
4	0.0	2.1	2.2
5	0.0	0.3	0.5

Figure 27: Error Rates over Time

Figure 28 shows a significant negative correlation occurs between cwpm and passage error ($r = -.3858$; $p < .0005$). The correlation becomes tighter at higher reading speeds, with variance all but disappearing after 80 cwpm, indicating that learners reading at 80 or more cwpm are unlikely to make errors on the level of text presented. On the lower end of the spectrum, more variance can be seen, with higher average numbers of errors for learners reading between 20 and 40 cwpm.

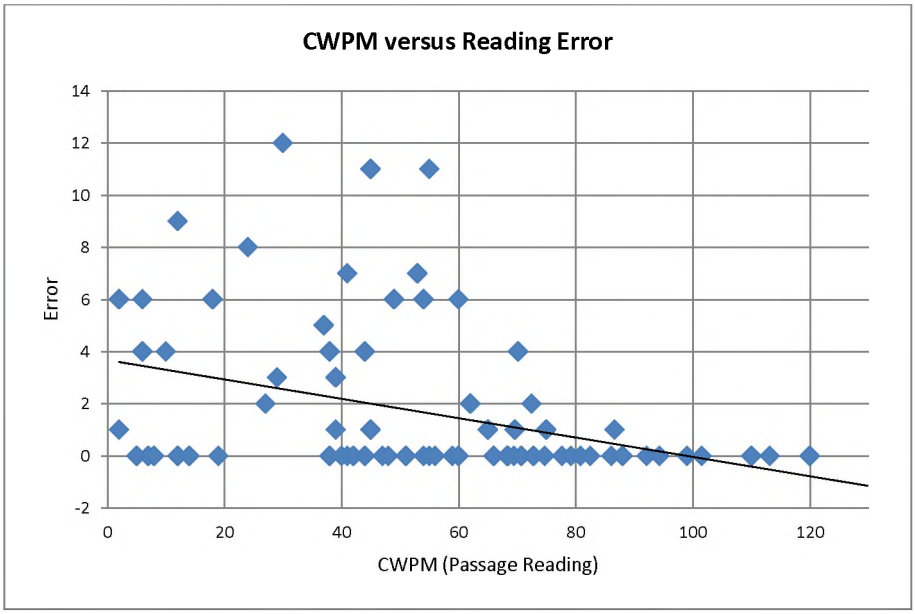


Figure 28: Error Rates versus Correct Words per Minute

While not specifically linked to a research question, findings on relationships between error rates, reading rates and comprehension provide an interesting comparison point to prior research in English. This is discussed in Chapter 7.

4.3.8 Comprehension

The final exercise the EGRA test administered was a comprehension test based on the passage reading. A total of five questions were asked and answered orally by learners.

Overall, learners gained slightly more in comprehension ability during the treatment period, with both grades improving by an average of 11.6 percent in the control period and by between 12.9 percent in the treatment period. When the means of improvement in control and improvement in treatment period were compared using a paired-sample t-test, the results were insignificant ($t(44) = 0.2$; $p = .4199$). However, if ceiling effects are removed[23], average gains in the control period drop to 6.7 percent, while increasing to 22.4 percent in the treatment period, and a paired sample t-test establishes the significance of higher gains in the treatment period ($t(32) = 2.7$; $p = .0058$). Therefore, it can be determined that learners who were not successful in making gains in routine instruction did make gains when technology was added to their instruction.

As shown by Figure 29, the highest comprehension gains were in grade 2 during the treatment period.

[23] Ceiling effects are defined as learners who could not gain more percentage points in the treatment period than control period due to high performance at the end of the control period.

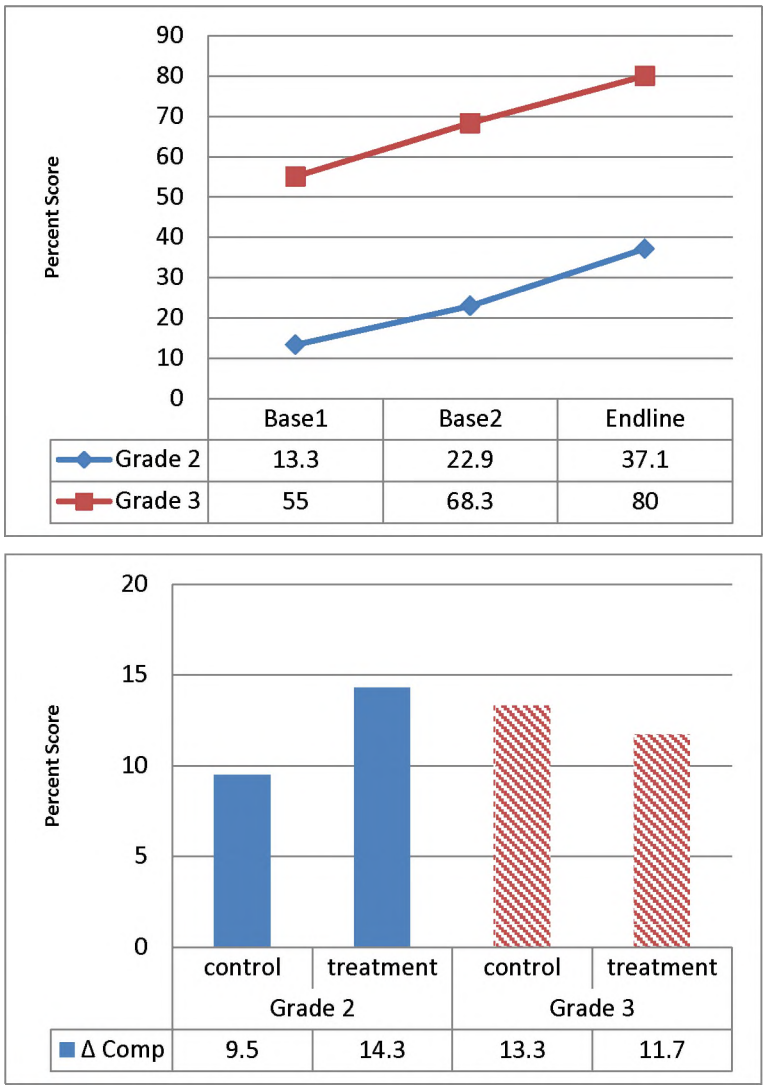


Figure 29: Performance on Comprehension over Time

Changes in distribution patterns, shown in Figure 30, reveal that in the first baseline over 50 percent of learners achieved a 0 score, creating a bimodal distribution pattern with one mean at 0 percent and a second mean at 73 percent. The relatively high value of the second mean indicates that learners were binary on comprehension ability: either they answered no questions or they answered a majority of questions correctly. The percentage of learners with a 0 score dropped to 31 percent by the second baseline, and decreased by a further 4 percent at the endline to 27 percent. The second baseline saw a slight drop in the non-zero mean to 68 percent, which is likely due to lower scores by learners moving out of the 0 score range during the control period. In the endline, a full separation occurs between learners unable to answer any questions and the rest of the learners, who achieved a mean score of 82 percent.

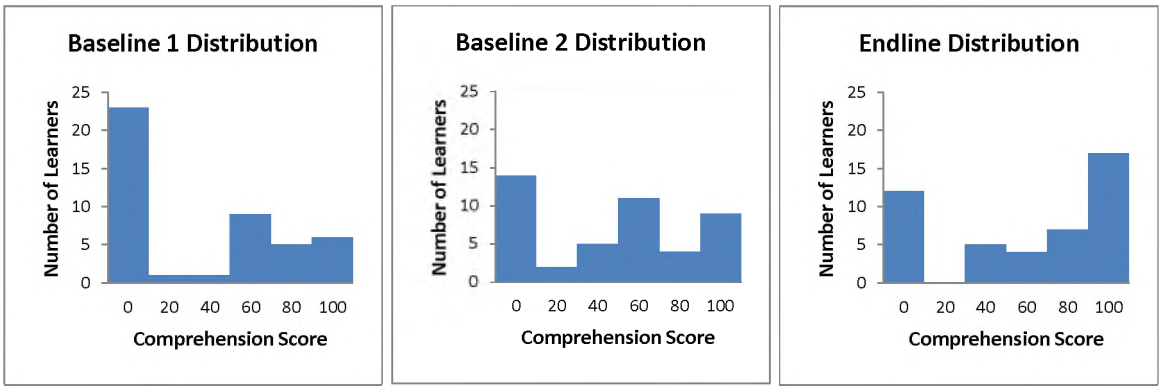
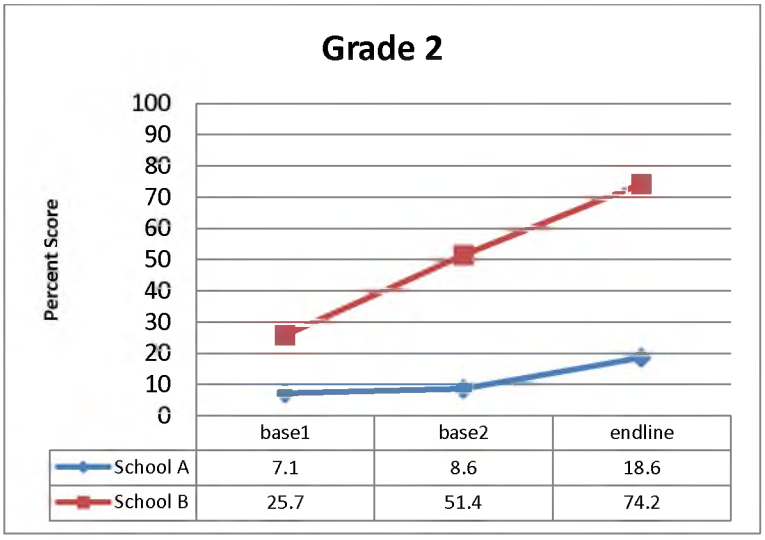


Figure 30: Changes in Distribution Patterns, Comprehension

Similarly to the correct words per minute analysis, there were clear differentials between the means of the grade 2 and grade 3 learners in both schools (see Figure 31). In the endline, 73 percent of all zero scores rested in the lower performing school's grade 2 classes. In the grade 3 classes, a similar pattern to the correct words per minute emerged, with the lower performing school learners attaining the same average score as the higher performing school by the end of the treatment period, with gains in both periods but higher gains in the treatment period.



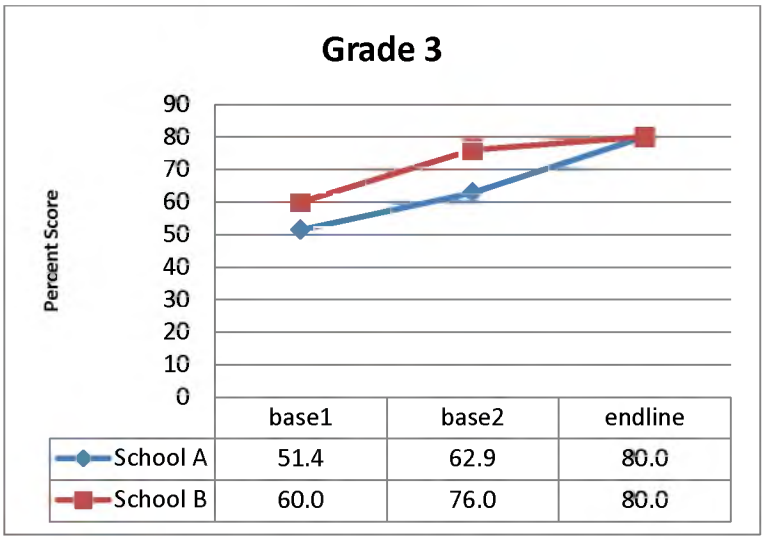


Figure 31: Performance on Comprehension over Time, by School and Grade

As shown in Figure 32, the percent of learners whose scores increased, decreased and remained the same was similar in both the control and treatment periods, with just over a third of the study population increasing in score. 40 percent of learners showed no improvement in comprehension ability in either period. Of these, 24 percent scored a 0 in all three testing periods, and 11 percent scored a 100 percent in all three testing periods, and of the remaining 5 percent (representing 2 learners), one began with 100 percent and dropped to 80 percent, and one remained at a score of 60 percent for all three testing periods.

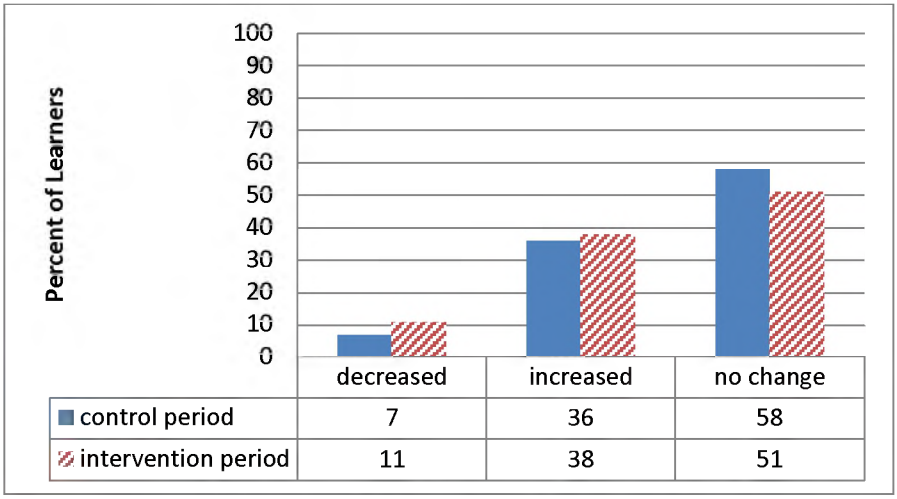
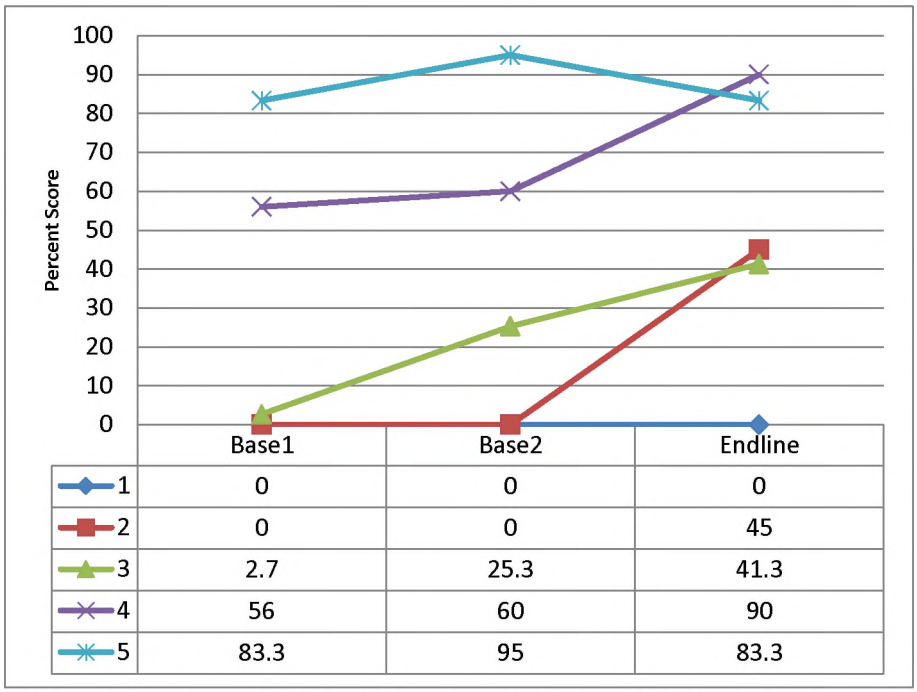


Figure 32: Change in Learner Performance, Comprehension

Analysis by Attainment Level

Figure 33 shows changes in score by attainment group. The highest gains emerge in the phonemically aware group (level 2), followed by the developing reader group (4) and the beginning reader group (3). The basic level readers were unable to read the passage or answer

any questions by endline, and the fluent readers improved at the midline but dropped at endline.



	Base1	Base2	Endline
1	0	0	0
2	0	0	45
3	2.7	25.3	41.3
4	56	60	90
5	83.3	95	83.3

Figure 33: Learner Attainment in Comprehension over Time, by Group

Learners in the fluent reader group performed best in the second baseline, with scores in both the first baseline and the endline 11.7 percentage points lower than the high of 95 percent achieved mid-year. When this was examined learner-by-learner, it was found that while 75 percent of learners in the fluent reader group either remained the same or gained in comprehension, 25 percent of learners regressed, one of them an outlier who dropped from 100 percent to 0 percent between the second baseline and the endline. The other two learners answered one or two fewer questions correctly in the endline than in the second baseline. Due to the open-ended nature of the comprehension questions, this result is unlikely to be due to guessing error and may be due to individual factors on the day of testing, such as boredom with the test or lack of engagement for other reasons. It may also indicate that early readers experience various regressions as well as progressions in reading ability. Perhaps, for example, they may gain a level of comprehension in which they second-guess correct responses, or a level of fluency in which they read automatically but without as much comprehension. The data collected in this study was not sufficient to draw conclusions in this regard, but it is an interesting direction for further study as 25 percent is not an insignificant percentage.

Learners in the basic level were unable to read the passage, so were also unable to answer comprehension questions. Learners in level 2 made all of their year gains during the treatment period, moving from 0 percent correct in the second baseline to 45 percent correct in September, concurrent with impressive increases in reading speed. Developing readers also made a majority of comprehension gains during the treatment period, improving by 30 percent compared to a 4 percent gain in the control period. While the beginning readers made gains in both the control and treatment period, gains were slightly less in the treatment period.

Given the ceiling effect on the results, it is difficult to ascertain if and/or to what extent high performing learners benefitted from the intervention. The EGRA was pitched at late grade 1 or early grade 2 level, and some of the learners were high achievers even on the first baseline, with more joining this group by the second baseline. This left little room for improvement between the second baseline and the endline. A more difficult test would need to be administered to these learners to determine if and to what extent they were able to improve their comprehension skill through program use.

For the remaining learners, the technology-based intervention yielded significantly better results than traditional classroom teaching ($p = .0058$). It is therefore the finding of this study that comprehension skill can be advanced by the use of the Bridges to the Future Initiative program.

Reading Speed and Comprehension in Northern Sotho

When cwpm are plotted against comprehension scores with cwpm 0-scores omitted, a strong and significant correlation occurs ($r = .7563$; $p < .0005$). The relationship is shown in Figure 34. It can be inferred from this that reading speed is a good predictor of comprehension performance in Northern Sotho. This echoes the assertions of prior research which has found that reading fluency is a good predictor of comprehension (Stanovich, 1991).

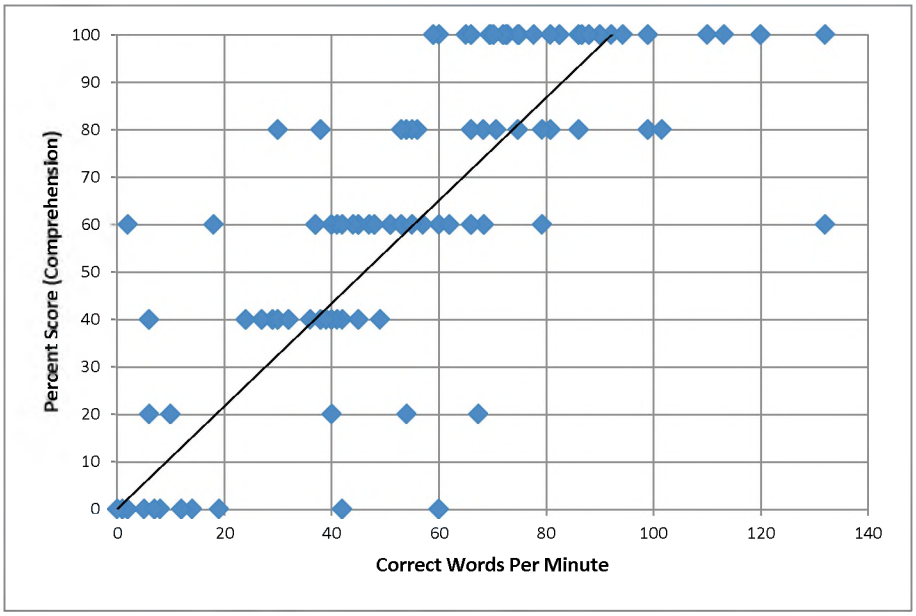


Figure 34: Correct Words per Minute versus Comprehension

The data also begins to point to a reading threshold in Northern Sotho between 20 and 24 cwpm: If learners are able to reach this threshold, they are most likely to be reading with comprehension. There are two instances of data points with higher reading speeds and 0 scores on comprehension; one was a learner who scored 80 percent and 100 percent on comprehension in the prior testing periods, and the other scored a 60 percent and 100 percent respectively in the second two testing periods. It is unclear what caused these inconsistencies, but they make up only 2 percent of data points with reading speeds over 24 cwpm and do not follow the general trend.

Analysis of Error and Comprehension

The analysis of error against comprehension reveals that that high rates of error do not correlate strongly with drops in comprehension scores, though the correlation is significant ($r = -.2817$; $p < .005$). Figures 35 and 36 show that the percent of participants with error and average error decrease sharply for learners attaining at least 60 percent comprehension, but that error rates are relatively high even among learners who scored well in comprehension. In fact, 25 percent of participants who scored above 50 percent had errors in passage reading, and errors did not necessarily affect comprehension negatively, with some participants at error rates of over 6 percent able to answer more than half of comprehension questions correctly.

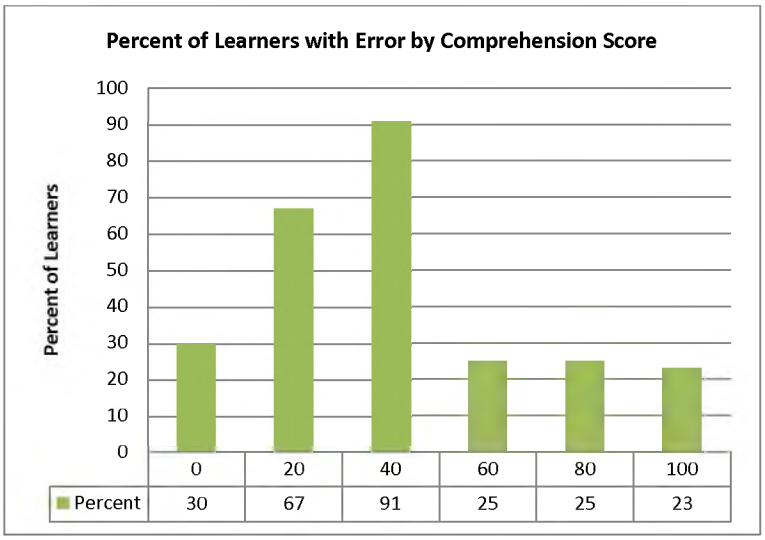


Figure 35: Percent of Learners with Error by Comprehension Score

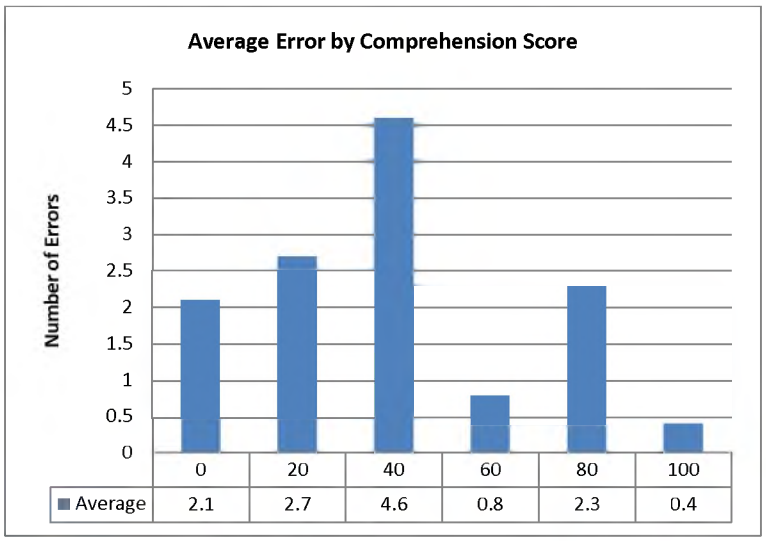


Figure 36: Average Error by Comprehension Score

4.4 Conclusions from the Early Grade Reading Assessment

This chapter set out to answer two of the research questions:

1.) What impact does BFI program use have on the following literacy skills in Northern Sotho-speaking grade 2 and 3 learners?
    a. Letter-sound recognition
    b. Simple word decoding
    c. Reading speed
    d. Comprehension

2.) Is there an optimal age for introducing the BFI program? Is it more effective with older or younger learners?

These questions were analysed through the use of the EGRA test in Northern Sotho. As a method-independent examination, meaning that results are meaningful regardless of the type of pedagogy learners engage in (RTI, 2011), the EGRA test results are not solely dependent upon explicit knowledge, but measure only outcomes, or implicit skills gained. Therefore, the EGRA examination was ideal to compare traditional teaching methods to teaching supplemented by the BFI program, as it was possible to compare outcomes across teaching methodologies which included and did not include the technology component.

As a time series design was used, gains between the first two baselines were used as a control to determine intervention impact. Therefore, conclusions drawn are based on gains made in the first half of the school year against gains made in the second half of the school year.

In answer to the first research question, when gains are taken in aggregate across the sample population, in both grades 2 and 3 larger gains were recorded in the control period for both grapheme-phoneme assignment and simple word decoding for both grades and in both schools. Therefore, it can be concluded that traditional pedagogy is more effective than the combination of pedagogy and technology at imparting the necessary skills for phoneme-grapheme coding and decoding.

However, it is important to note two additional observations.

First, ceiling effects may have played a role in the lack of gains in the second half of the year. In other words, once learners achieve a certain benchmark ability in these two tasks, they are unlikely to improve further due to the limits of the number of test items presented and limits on physical speech production. These two items were timed for only one minute, but times

were not recorded for learners who finished in less than one minute, which means ceiling effects cannot be reliably estimated by the data collected.

Second, for both skills there was a subset of learners who improved more in the intervention period. For phoneme-grapheme assignment, it was learners who began the intervention period at the pre-reader level, with no ability to assign phonemes to graphemes. For decoding, it was phonemically aware learners, or learners who began the intervention period with the ability to assign phonemes to graphemes, but who were unable to read words.

Therefore, although overall the sample population exhibited better gains in the control period, when learners who were not able to achieve a particular literacy skill through traditional pedagogy are considered, it was found that they were able to gain some ability in that skill during the period in which technology was introduced to their classroom pedagogy.

However, this was not true for all learners. Roughly half of learners in each category were able to improve during the intervention period, with the remaining half still displaying no gains. Further research is necessary to determine what factors affect the ability of remedial learners to learn from the program and to concretely prove that it is the introduction of technology which influences their increased ability in these literacy skills.

In terms of fluency, significantly higher gains were made during the intervention period. All grades and classes performed better during the intervention period than the control period on this literacy skill, and overall the sample population showed four times more improvement in the intervention period than the control period, improving by an average of 16 correct words per minute in the intervention period. This implies that fluency can be improved through the use of a technology-based literacy program.

In the comprehension section, it is more useful to consider the results with ceiling effects removed, as the test was not difficult enough to properly gauge improvement in high performing learners in this area. For remaining learners, average intervention gains were three times higher than gains in the control period. This significant improvement is linked to improvements in fluency, as there was a high correlation between fluency and comprehension in this study. Therefore, it can be concluded that both fluency and comprehension were improved by the inclusion of the technology program in the classroom.

With regard to the second question, there was no correlation observed between age and performance on any of the literacy skills. The largest gains from program use were observed

in grade 3 in school A, and grade 2 in school B. Additionally, age was not a significant predictor of either performance or improvement during the intervention period. This implies that there is no threshold age for technology to be effective within the age range under study. Therefore, it cannot be concluded that the program is more effective with any particular age or age range within this sample population.

However, it is possible that there is still an age threshold at which technology becomes an effective teaching tool below the youngest members of the sample, who finished the year at seven years old. Prior research on learning from television by Krcmar et al. (2007) indicates dramatic improvement in ability to learn from television in children between three and five years old; it may be that such a threshold would lie in a similar age range.

CHAPTER 5: SKILLS ACQUISITION AND RETENTION OF KNOWLEDGE

The retention test served two purposes. The first was to determine how much and what types of information are gained and retained by learners after concluding use of the program.

The second was to supplement the findings of the EGRA examination. As noted, one of the strengths of the EGRA is that it is not dependent on instruction and can be used with various types of pedagogy. However, at the same time, the BFISA2 programme was designed with a specific set of curricula in consideration, namely the Curriculum Assessment Policy Statements (DBE, 2011). And while the skills tested by the EGRA are certainly points of consideration, the skill set the BFISA2 programme is intended to convey to learners expands beyond stages of literacy.

Therefore, while the EGRA test provided a method-independent analysis of how learners gained literacy skills mapped against a theoretical trajectory, it was necessary to examine concrete tasks and skills covered in the programme in order to determine whether the knowledge gained through the program was retained over time.

A retention test was designed in order to fulfil this need and to answer the following research questions:

1.) To what extent is knowledge gained by the conclusion of the BFI program sustained after use has ended?

2.) What types of knowledge, procedural or fact-based, are sustained?

3.) Is there a pattern to retention and non-retention in each task and subtask? In other words, was it possible to predict learners who would and would not retain knowledge?

4.) For tasks included in both tests, does improvement on the EGRA correlate in any way to the retention rate?

Section one of this chapter outlines the relevant research on and regarding retention testing that informed the test instrument development and methodology designed. Section two details the methodology of the test administration. Section 3 presents the results, and finally section 4 discusses the results in relation to the literature and the research questions.

5.1 Literature and Theory on Retention of Knowledge

A retention interval is defined as the amount of time between initial learning of material and the recall stage, or application of a cognitive test. According to Haynie (2007), "delayed retention represents the important and significant learning" achieved through instruction. Therefore, it is important for the intervention impact assessment to establish to what extent knowledge is sustained as well as gained. The extent of the delay in retention test varies depending on field of study, research questions and other factors, but for the purposes of education a delay of at least three weeks is recommended (Haynie, 2007).

Prior research has shown that test taking itself is a form of learning which significantly enhances performance on subsequent tests, above the effects of additional studying or review, a phenomenon known as the testing effect (Toppino & Cohen, 2009; Haynie, 2007; Carpenter et al., 2007; Carpenter & DeLosh, 2005; Wheeler et al., 2003 Runquist, 1983). The mechanisms through which Roediger and Karpicke (2006) suggest that the effects of testing include both mediated and direct effects – testing may act indirectly to improve performance by encouraging further study (mediated effects) and/or may itself contribute to the knowledge base and recall of facts (direct effects). Therefore, delayed retention test achievement of learners who take an initial test may be elevated from the achievement of groups given only a delayed retention test. This is demonstrated by Toppino and Cohen (2009), who found higher retention in the 'study' condition as opposed to the 'test' condition after a 2 minute retention interval, but higher retention in the 'test' condition after a 48 hour interval. In the study condition, participants recalled just over 60% of Swahili words after 2 minutes and just over 30% after 48 hours; in the test condition participants recalled just over 50% in both the two minute and 48 hour retention periods, with slightly higher performance after 48 hours. However, studies over longer periods of time, such as the retention test by Carpenter et al. (2007) showed up to a 63% decay of knowledge between testing at five minute and 42 day delays.

The psychological processes and mechanisms behind the retention effect are not fully understood (Toppino & Cohen, 2009). Possible explanations include that testing improves the process of retrieval of information (Wheeler et al., 2003); or that matching testing conditions with study conditions enhances transfer-appropriate processing, or the relationship between encoding and retrieval (Toppino & Cohen, 2009). Bjork and Bjork (1992) suggested that the extent of learning, which they referred to as storage strength, could slow changes in accessibility or retrieval strength, the assumption being that testing improves storage strength,

and therefore improves retrieval on delayed tests. Context is an additional factor in recall, as a lack of appropriate context or differing contexts can lead to declining performance in secondary text sessions, leading to depressed test performance (Averell & Heathcote, 2009).

Testing can incorporate different types of memory, as well. Averell and Heathcote (2009) experimented with implicit (nondeclarative) and explicit (declarative) memory tasks. Implicit memory tasks are defined as tasks which do not require conscious thought, and include categories such as procedures implemented by rote and knowledge of language. Explicit memory includes both semantic memory, or memory of facts, and episodic memory, or long-term memory of specific events. Through testing stem-cued recall as an example of an explicit memory task and stem completion as an example of an implicit memory task, Averell and Heathcote found that between one day and one week performance on both tasks declined by a similar amount, with little change on test performance with delays between one week and one month.

5.2 Methodology for Testing Retention of Knowledge

The uptake and retention of knowledge was tested through a quasi-experimental pre and post-test design using a paper-based retention test based on the skills and information conveyed through the BFISA2 tablet program, implemented from the conclusion of program use to the beginning of the subsequent school year, as outlined in Figure 37. The retention test was written and given in Northern Sotho and tested four categories of skills: phoneme manipulation, grammar and sentence structure, explicit factual knowledge and comprehension.

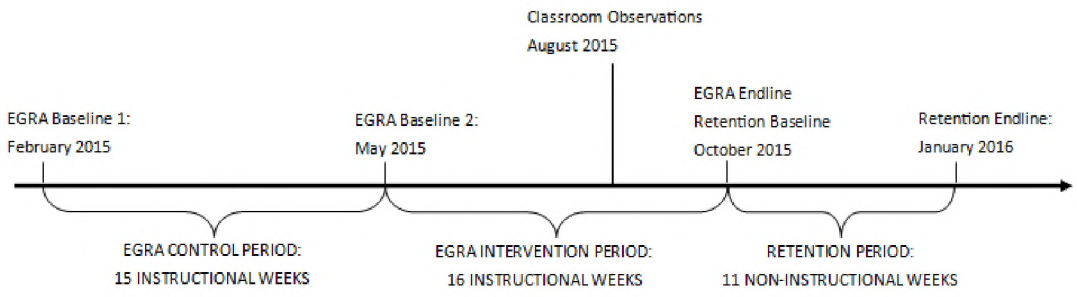


Figure 37: Research Timeline

The results of the retention baseline and endline tests were analysed against observed classroom characteristics and EGRA scores to attempt to answer the research questions outlined for the section:

1.) To what extent is knowledge gained by the conclusion of the BFI program sustained after use has ended?

2.) What types of knowledge, procedural or fact-based, are sustained?

3.) For tasks included in both tests, does improvement on the EGRA correlate in any way to the retention rate?

5.2.1 Instrument

In alignment with the research questions the test sought to answer, the purpose of the retention test was to test the uptake of the specific curriculum-aligned literacy skills taught by the Bridges to the Future Initiative program, as well as to measure the retention of these skills over time. The test further sought to differentiate between implicit and explicit tasks, or tasks based in skills and tasks based in factual knowledge. To fulfil the requirements, a paper-based testing tool was developed by the researcher with the assistance of Northern Sotho language editors from the Molteno Institute for Language and Literacy. Separate tests were developed for grade 2 and 3 as the content covered differed in some respects between grades to include items stipulated by the Curriculum Assessment Policy Statements (DBE, 2011). As a result, unlike the EGRA examination, the difficulty of the two tests varies and the findings are not directly comparable by grade.

The retention test was drawn from the BFISA2 program and includes all skills covered in the two grades. These can be grouped into four categories: phoneme/word manipulation, sentence mechanics and grammar, factual knowledge and comprehension.

Phoneme/word manipulation forms the bulk of the programme. Almost half of all activities in the program deal with introduction of phonics, phoneme identification, segmenting of words into syllables and sounds, and building words from phonemes. Activities included first sound identification, in which learners were asked to circle words beginning with given letters, syllable segmenting and phoneme segmenting. The instructions mirror the structure and syntax of the BFISA2 program activity on breaking words into phonemes.

Sentence mechanics and grammar account for another 40 percent of activities in the program, with learners expected identify correct punctuation marks, insert commas into the correct

places in a sentence, identify simple parts of speech, build sentences and choose the correct conjunction for a sentence. Learners were expected to duplicate these tasks on the paper examination.

Comprehension forms only 10 percent of activities, as there is one comprehension passage and three or four accompanying questions per story. However, in terms of time taken the share of this component is significantly higher, and the three sections (manipulation, mechanics and comprehension) take roughly equal time for the average learner.

Concurrent to these activities, the program engages learners on a variety of topics and introduces characters, events and cross-curricular information. For example, learners read about photosynthesis, cultural customs in Ethiopia and dangerous animals in Africa through the adventures of the BFI family and their friends. This type of information is considered as fact-based rather than procedural or skills-based, and is the sole category of completely explicit knowledge in the retention test.

In order to determine what skills and types of skills were adopted by program users and what skills and types of skills were retained over time, the retention test was designed to include both procedural and factual information, and to cover all of the skills present in the program in each of the areas.

An outline of the skills tested in each grade is provided in Table 14.

Table 14: Retention Test Categories by Grade

	Grade 2	Grade 3
Word Manipulation	First-sound identification; syllable segmenting; sound segmenting	First-sound identification; syllable segmenting; sound segmenting
Grammar and Punctuation	Identification of parts of speech; end sentence punctuation	Identification of parts of speech; end and mid-sentence punctuation; conjunctions
Fact-based knowledge	Questions on storyline from grade 2 lessons	Questions on storyline from grade 3 lessons
Comprehension	Direct, Inference and Synthesis questions based on grade 2 text adapted from the VulaBula [24] reading programme	Direct, Inference and Synthesis questions based on grade 3 text adapted from the VulaBula reading programme

[24] VulaBula is a trademarked reading programme from the Molteno Institute for Language and Literacy. Test schools had no prior exposure to the content.

With the exception of the comprehension passage, test items were taken directly from the BFI program to ensure that skills were of equivalent grade level to the material presented by the program and that learners had prior exposure to the content. Figure 38 and Figure 39 are examples of the types of tasks given on the retention tests[25].

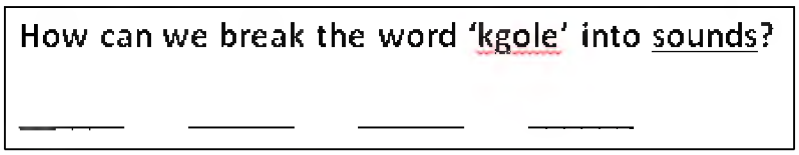


Figure 38: Example of a Phoneme Segmenting Task

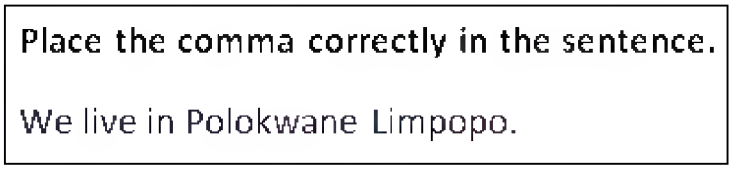


Figure 39: Example of a Sentence Mechanics Punctuation Task

5.2.2 Sample

Two classes per school were selected to undertake the retention test, one in each grade. The selection of classes was based on random selection of one class from the set of classes. Both schools named classes after letters of the alphabet ('Class A', 'Class B', etc.). As per the result of the random draw class b was tested in each school and grade after ascertaining with the Heads of Department and principals in both schools that learners were randomly assigned to classes and not grouped by ability.

The same learners were tested in October and January. Because the school year had rolled over, learners had to be taken from various classrooms and grades, as some learners were repeating and others had progressed, and classes had been redistributed. Attrition played a significant role as learners were absent, had dropped out or switched schools, particularly in the higher-performing school. The final sample included 109 learners as indicated in Table 15.

Table 15: Retention Test Participant Numbers

	Grade 2	Grade 3	TOTAL
School A	31	32	63
School B	25	21	46
TOTAL	56	53	109

[25] Please note the retention tests are the intellectual property of the Molteno Institute for Language and Literacy. Activities are displayed as examples of the type of activities given in the retention tests; they are not actual test questions. Actual test questions and directions were written in Northern Sotho.

5.2.3 Testing Procedures

In October, concurrent to the final EGRA examination and the termination of BFISA2 program use, the sample of learners described above was given the baseline retention test. For most learners, this was their first experience with an independent test, as the practice in foundation phase is for teachers to read the entirety of the test to learners, including comprehension passages, questions and answers, which clearly would not yield any meaningful information as to independent learner reading ability. As a compromise, teachers were allowed to read the directions to the learners, and were allowed to read the questions but not the answers to the learners. They were also allowed to read the comprehension passage once aloud to the class before learners looked at the comprehension questions. In grade 2, they were also allowed to read the comprehension questions but not the answers.

A second retention test using identical procedures was administered in January 2016 to test the amount and types of knowledge learners were able to recall over time. The second test was originally scheduled in November, but upon arriving at the schools it was found that learners had already departed for summer holidays, although the school calendar did not end for another three weeks. As a result, the second test was rescheduled for January 2016 and administered in the second week of classes as schools requested the first week for administrative purposes. For the post-test, a list of learners taking the pre-test was produced. Analysis included only learners participating in both tests.

For both the pre-test and post-test, a classroom was set up to administer the test. Learners were called into the testing area two at a time and spaced to mitigate copying, with two learners per long desk. Learners in each class were provided with one hour to complete the examination; however most learners finished within 45 minutes. After completing the test, learners submitted their papers to the examiner and exited the testing room.

5.2.4 Treatment and Analysis of Data

Test information included learner name, school and grade. This data was used for matching purposes, but in the final dataset learner names were replaced with numbers to protect learner identities. Teachers were not privy to individual learner results, as poor performance on the test could have resulted in negative consequences for learners. As hard copy data contained learner names, it was retained through the end of the analysis period and destroyed at the conclusion of the evaluation.

Marking and data entry for both tests was done by the researcher using an examination answer key. Data was entered twice and checked for discrepancies, which were resolved using hard copies. A 10% sample was then randomly selected and checked for error rate, which was 0%.

Evaluation and analysis was done using MS Excel, hard copy data, and the program R.

5.3 Results and Analysis of the Retention of Knowledge Data

This section presents the results of the skills test and delayed retention test for each of the four areas addressed by the BFI program: phoneme manipulation, sentence mechanics, explicit factual knowledge and comprehension. The presentation begins with an overall analysis, followed by analysis of each of the four skill categories with subtasks included. Results for each skill category are presented as change in average percent score.

However, an analysis of averages does not display movement in the scores of individual learners. Therefore, the data was also analysed by individual to determine the percentage of learners who progressed, regressed and retained on each subtask. The results are disaggregated by grade and school to allow for analysis and discussion in conjunction with EGRA performance and classroom observations.

5.3.1 Overall Results

In terms of skills uptake, the program did not achieve 50% overall averages in any category in the test of skills uptake, in other words the retention pre-test. Even after significantly higher scores on the second retention test ($t(108) = -2.5$; $p = .0069$), learner averages were still below 50% on procedural knowledge and showed a minor decrease in factual knowledge. Figure 40 shows the average change by skill category. A paired sample t-test to compare mean performance on the skills upake and retention tests showed significance in manipulation ($t(108) = -3.1$; $p = .0013$) and comprehension ($t(108) = -2.0$; $p = .0248$), and approached significance in grammar ($t(108) = -1.5$; $p = .0666$).

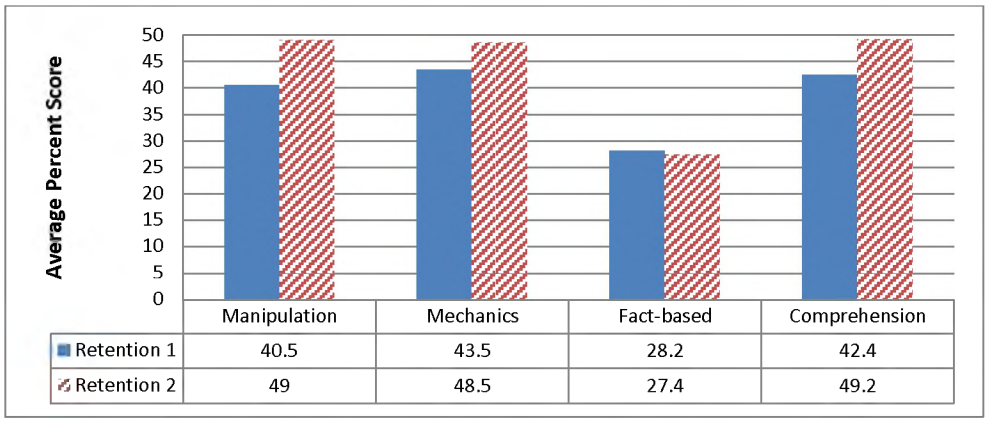


Figure 40: Overall Change in Performance by Skill Category

Figure 41 shows that although both schools improved overall between the tests, the higher performing school B improved more in every category except factual knowledge. School B made overall gains of more than 12 percent in every other category between the two testing

periods, compared to gains of between 0 and 5 percent in school A. Factual knowledge scores were lowest on both tests, and also had the only regression between tests. It is also the only category in which school A outperformed school B on both tests.

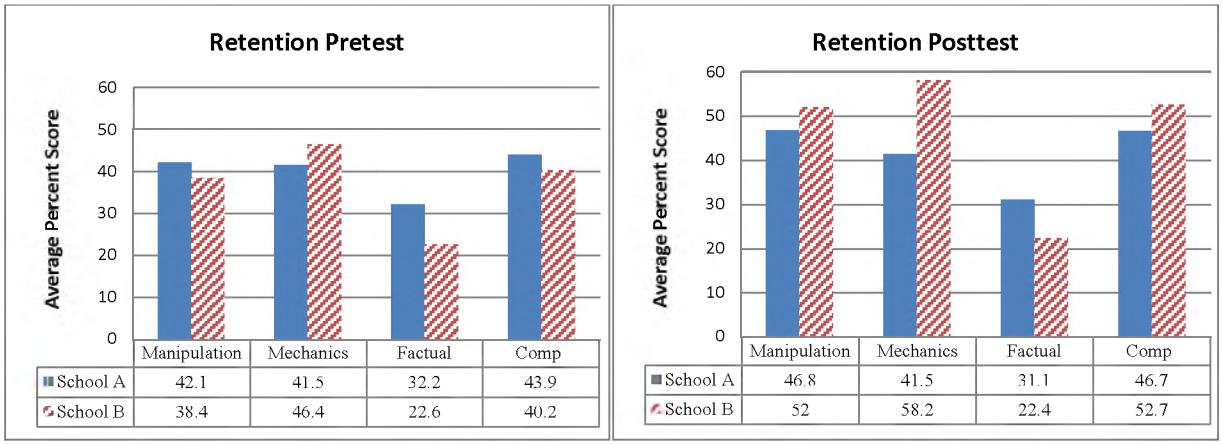


Figure 41: Performance by School, Retention Pre and Post-test

When change in performance by class is considered (see Table 16), B3 showed improvement between the pre and post-test in phoneme manipulation, factual knowledge and comprehension, ultimately performing above 40 percent in all categories except factual knowledge. Class B2, as well, improved in three out of the four categories, the exception being factual knowledge which decreased by 7.6 percent.

School A had drops in performance in two categories in each grade. Grade 2 performance decreased by 5.4 percent and 5.9 percent respectively in manipulation and mechanics, while grade 3 decreased by 3.4 percent in factual knowledge and 7 percent in comprehension.

Three out of four classes showed gains of over 10 percent in comprehension. The exception was A3, which was the highest performer on the pre-test at 50 percent. Grade 3 learners in both schools improved on manipulation of phonemes by more than 10 percent, with class B2 also improving by 10 percent. For mechanics, skills increased by more than 25 percent in class B2, with moderate positive changes in A3 and moderate negative changes in A2 and B3.

Table 16: Change in Retention Performance by Class and Skill Category

		Manipulation	Mechanics	Factual	Comprehension
GRADE 2	School A	-5.4	-5.9	1.3	12.9
	School B	10	28.0	-7.6	12.0
GRADE 3	School A	14.5	5.8	-3.4	-7.0
	School B	17.9	-7.5	8.6	13.1

When retention test results are presented by individual class, average scores on the retention pre-test were 40 percent or less for all classes tested, with a range of 8.6 percent (see Table 17). The class with the highest performance on the EGRA (B3) was the worst-performing on the retention pre-test, and conversely, the class with the lowest performance on the EGRA (A2) achieved on par with the other classes. Excluding B3, the range of retention pre-test scores was only 2.7 percent, although the range of the EGRA scores was 35.7 percent. Overall average scores on the retention post-test were higher than on the pre-test, with the exception of A2, which lost 1.1 percent. These factors will be examined in the discussion section after a more complete examination of performance in each category by class.

Table 17: Overview of Program Use, EGRA scores and Retention Test Scores

		Hours of BFI Program Use	Final EGRA Test Score (average %)	Retention Pre-test Score (average %)	Retention Post-test Score (average %)	Change in Retention
GRADE 2	School A	25	14.7	37.3	36.1	-1.1
	School B	10	49.0	40.0	51.3	11.3
GRADE 3	School A	18	50.4	38.1	40.6	2.4
	School B	3	60.0	31.4	36.6	5.2

5.3.2 Manipulation

The retention test dealt with three components under manipulation: first sound identification (FSI), syllable segmenting and sound segmenting. FSI required learners to select pictures with a given first sound; syllable segmenting required learners to parse words into syllables; and sound segmenting required the learners the parse words into sounds. Figure 42 highlights performance in the manipulation concept area and its subtasks.

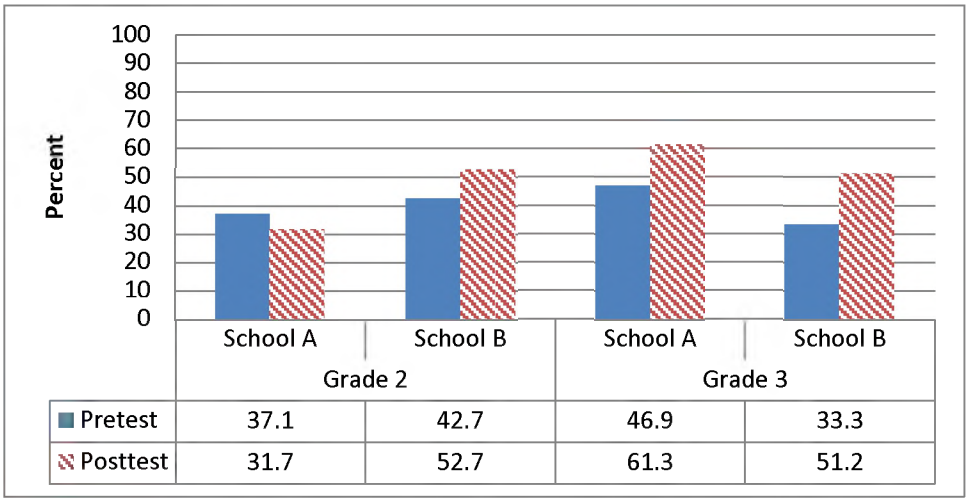


Figure 42: Average Percent Scores by Class, Overall Manipulation

When examined by average percent achieved, grade 2 learners in both schools performed best on both retention tests in FSI as compared to other subtasks in the category. Class A2 showed a decrease of 12.9 percent in average performance from pre to post-test, while class B2 showed an increase of 21.5 percent. Averages were higher in A2 on the pre-test, and in B2 on the post-test, but were above 50 percent for both classes on both tests. Syllabification average scores were below 50 percent in both schools on the pre-test, with B2 performing 9.3 percent better than A2. A2 dropped from pre to post-test by 9.3 percent, while B2 improved by 10 percent. Sound segmenting was the most difficult subtask in this category for learners, with A2 performing below 5 percent and B2 performing below 15 percent on the pre-test. Both schools showed increases of between 5 and 10 percent on the post-test, but no learner in either school on either test was able to perform at 100 percent on this task. The most common error in grade 2 was in breaking the digraph "kg", which represents one sound in the language, into individual letters.

Figures 43 to 45 highlight performance on individual subtasks. Performance on FSI was lower in grade 3 than in grade 2, with both schools at 50 percent or below and B3 initially performing 17.2 percent better. However, average scores in both schools rose between 23 and 25 percentage points in the post-test to achieve average percent scores of above 55 percent.

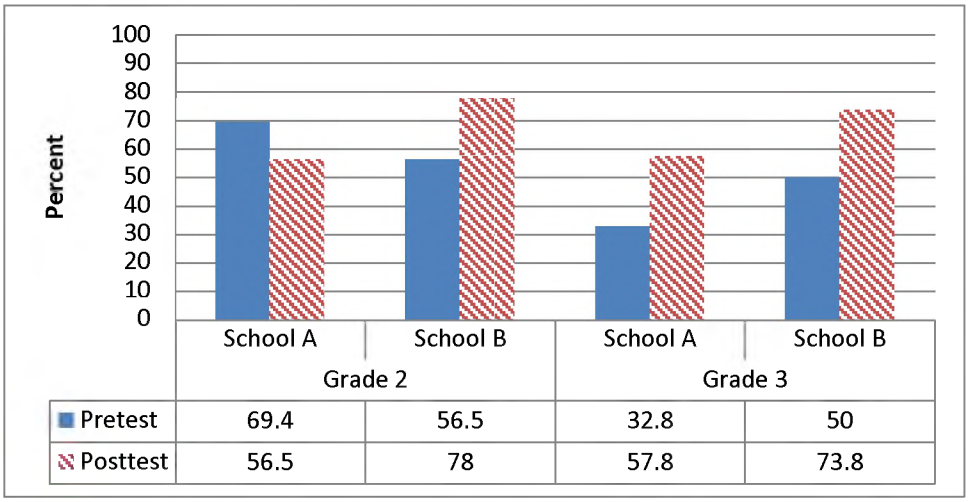


Figure 43: Average Percent Scores by Class, First Sound Identification

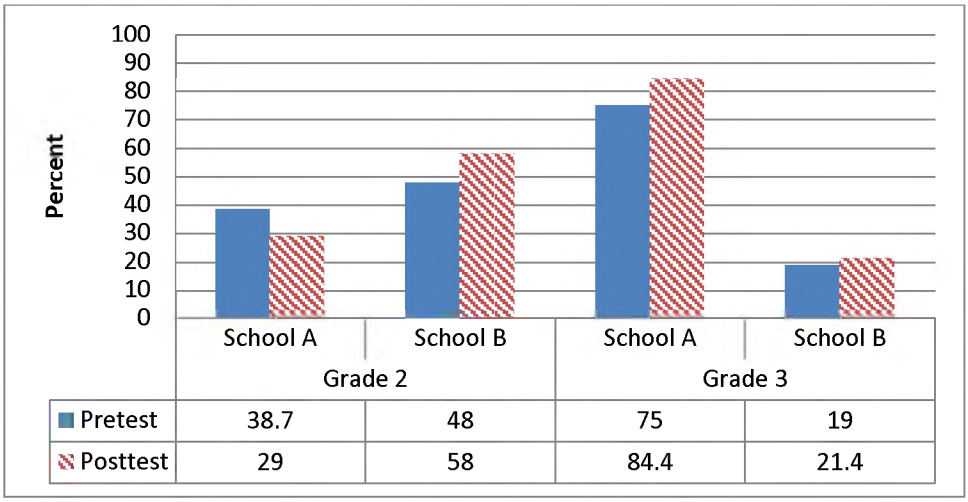


Figure 44: Average Percent Scores by Class, Syllable Segmenting

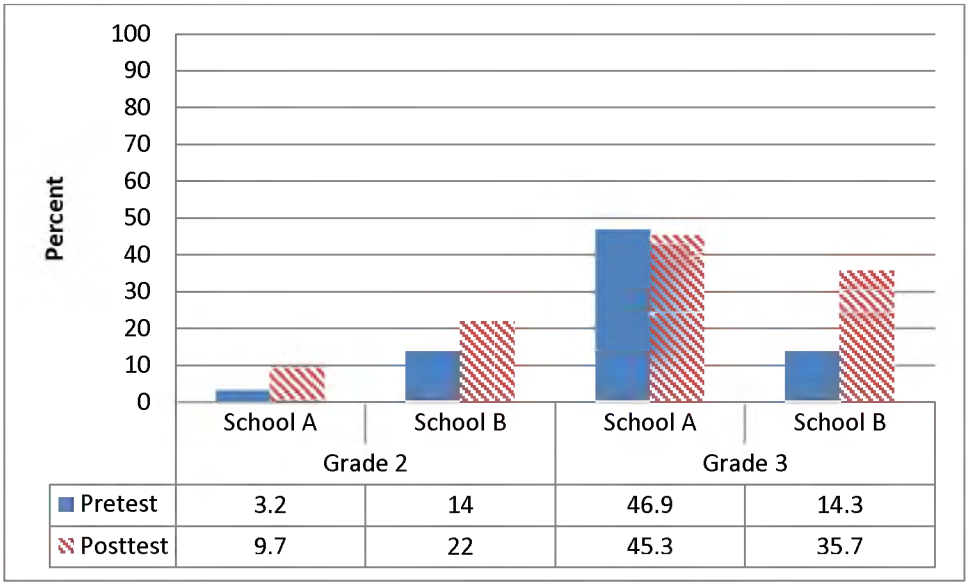


Figure 45: Average Percent Scores by Class, Sound Segmenting

Learners in grade 3 in both schools made a common error in first sound identification when asked to circle items beginning with the letter /s/ and one of the pictures was of a pair of shoes. The word for 'shoe' in Northern Sotho does begin with /s/, however if there are two shoes the plural prefix 'di-' is added and therefore the selection is incorrect. It is unclear whether they were confused by the English digraph 'sh' or failed to notice the plural, but just over 50 percent of learners in grade 3 circled this answer incorrectly in October. Error on this item was lower (almost 0) on the second retention test, which contributed to gains.

A3 performed better than B3 in both syllable and sound segmenting, with average scores 56 percent and 32.6 percent better respectively on the pre-test. Syllabification scores increased in both schools in the post-test, and while for sound segmenting A3 had a 1.6 percent drop and B3 more than doubled its score, A3 still outperformed B3 in both subtasks in the post-test.

When change in individual learner performance is evaluated (see Table 18), overall regression rates ranged from 7.9 percent to 20.4 percent. However, in the case of class B3, the relatively low rate of regression is reflective of low pre-test achievement on two of the three tasks (syllable and phoneme segmenting). This is also the case with low regression rates in phoneme segmenting in grade 2. If these are omitted from the average score, regression rates for manipulation increase to 29.1 percent in A2 and 20 percent in B3.

Progression rates for manipulation are particularly high in school B, averaging 34.7 percent in grade 2 and 27 percent in grade 3. School A had an average progression rate of 29.2 percent in grade 3 and 12.8 percent in grade 2. The relatively low progression rates and high regression rates in A2 is consistent through all categories.

Phoneme segmenting was the most challenging subtask in three classes, with high rates of low performance in A2 and B3. B3 also showed difficulty in syllable segmenting, which was shared to a lesser extent with A2. Classes A3 and B2 showed comparative achievement on these two subtasks, with rates of regression at 20 percent or below.

Table 18: Individual Learner Variance in Manipulation

GRADE 2	SCHOOL A				SCHOOL B			
	retained, <50%	regressed	progress-ed	retained, 50%+	retained, <50%	regressed	progress-ed	retained, 50%+
First Sound Identification	0.0	32.3	12.9	54.8	0.0	20.0	40.0	40.0
Syllable Segmenting	41.9	25.8	9.7	22.6	16.0	20.0	40.0	24.0
Phoneme Segmenting	77.4	3.2	16.1	3.2	48.0	8.0	24.0	20.0
AVERAGE (all subskills)	39.8	20.4	12.9	26.9	21.3	16.0	34.7	28.0

GRADE 3	SCHOOL A				SCHOOL B			
	retained, <50%	regressed	progress-ed	retained, 50%+	retained, <50%	regressed	progress-ed	retained, 50%+
First Sound Identification	9.4	6.3	53.1	31.3	0.0	9.5	47.6	42.9
Syllable Segmenting	3.1	12.5	21.9	62.5	71.4	9.5	9.5	9.5
Phoneme Segmenting	40.6	15.6	12.5	31.3	61.9	4.8	23.8	9.5
AVERAGE (all subskills)	17.7	11.5	29.2	41.7	44.4	7.9	27.0	20.6

Table 19 shows no discernible pattern in regression or progression based on hours of program use. The class with the highest use (A2) presented the highest regression and lowest progression, while A3 had the second-lowest regression and second-highest progression. B3 had the lowest regression, but due to the comparatively low baseline of 33.3% it is difficult to attribute much weight to the data, as regression from 0 is not possible.

An inconsistent pattern also emerges when performance and gains on the EGRA are compared to the retention pre-test, regression and progression. A2 performed better on the retention test than the EGRA, while B3 performed better on the EGRA than the retention test. The two classes in the middle performed 2 to 3 percentage points better on the EGRA.

Table 19: Program Use, EGRA and Retention Performance on Manipulation Tasks

		Hours of BFI Use	Phoneme Assignment Gains (EGRA)	EGRA average performance at endline	Pre-test (Avg % correct)	% of Learners Regressed	% of Learners Progressed
Grade 2	School A	25	10.1	24.3	37.1	20.4	12.9
	School B	10	17.3	51.8	42.7	16.0	34.7
Grade 3	School A	18	10.0	48.8	46.9	11.5	29.2
	School B	3	27.4	67.3	33.3	7.9	27

5.3.3 Sentence Mechanics: Grammar and Punctuation

The mechanics section in the two grades included slightly different sets of subskills. Both grades were asked to identify parts of speech and punctuate the end of sentences; however grade 3 learners encountered the additional skills of mid-sentence punctuation and conjunctions. Therefore, for grade 2 this section encompassed two subskills (identification of parts of speech and punctuation) while grade 3 encompassed three (identification of parts of speech, punctuation and conjunctions).

As shown in Figure 46 to 49, grade 2 showed similar trends in mechanics to the manipulation section, with A2 decreasing between the pre-test and post-test on both subskills, while conversely B2 improved in both subskills on the post-test, in this case dramatically, with gains of 31 percent in identifying parts of speech and 22 percent in punctuation.

In grade 3, overall A3 showed an increase of 5.8 percent in mechanics, while school B showed a decrease in performance of 7.5 percent. A3 improved in identification of parts of speech and conjunctions, although performance on punctuation saw a decrease of 18.7 percent. B3 performed below its relatively higher pre-test scores in all subskills: 8.3 percent on parts of speech; 2.4 percent on punctuation; and 14.3 percent on conjunctions.

The result was that on the pre-test, B3 outperformed A3, but on the retention test, A3 outperformed B3 in all subskills except punctuation.

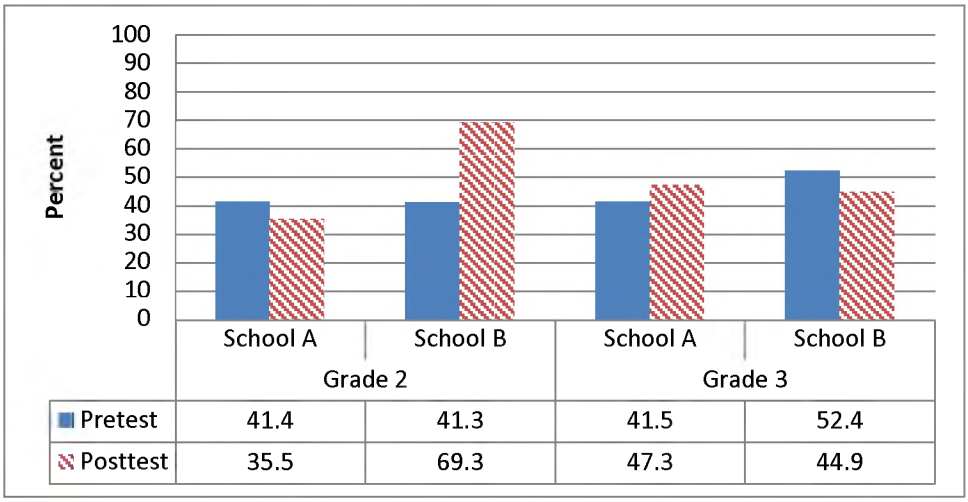


Figure 46: Average Percent Score by Class, Overall Sentence Mechanics

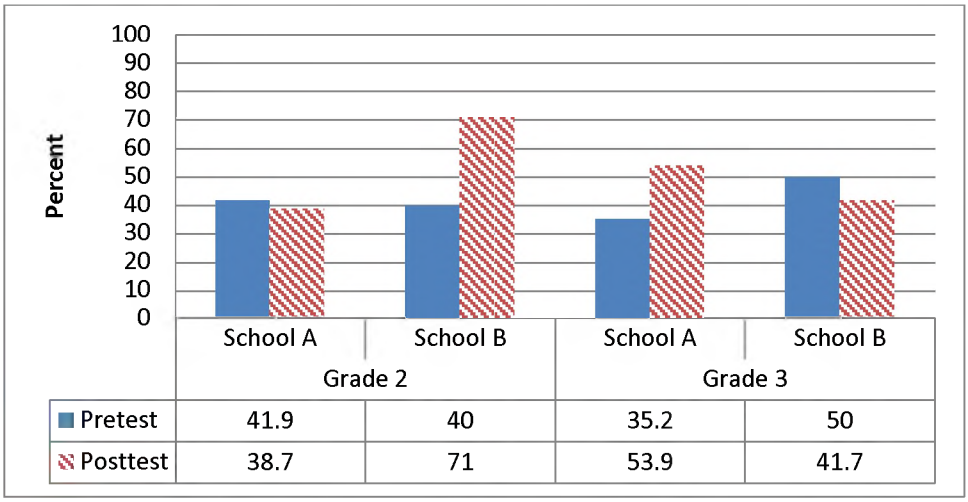


Figure 47: Average Percent Score by Class, Parts of Speech

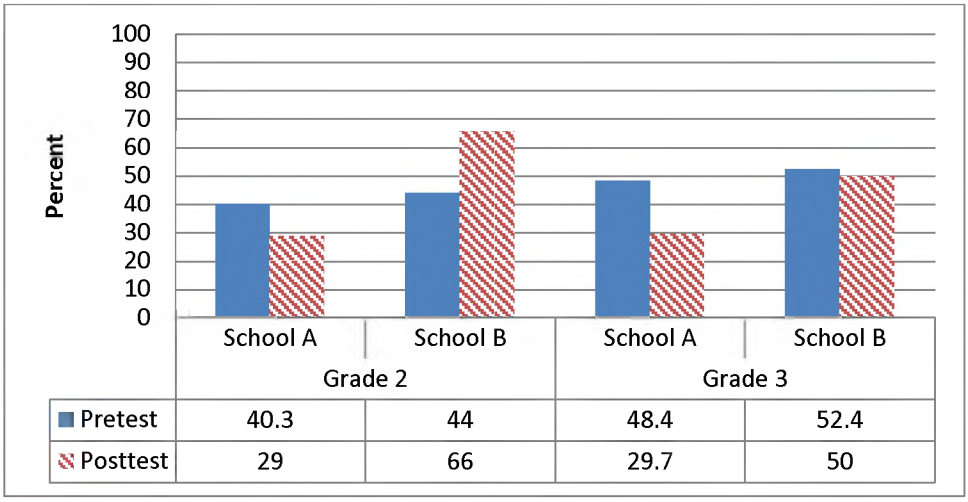


Figure 48: Average Percent Score by Class, Punctuation

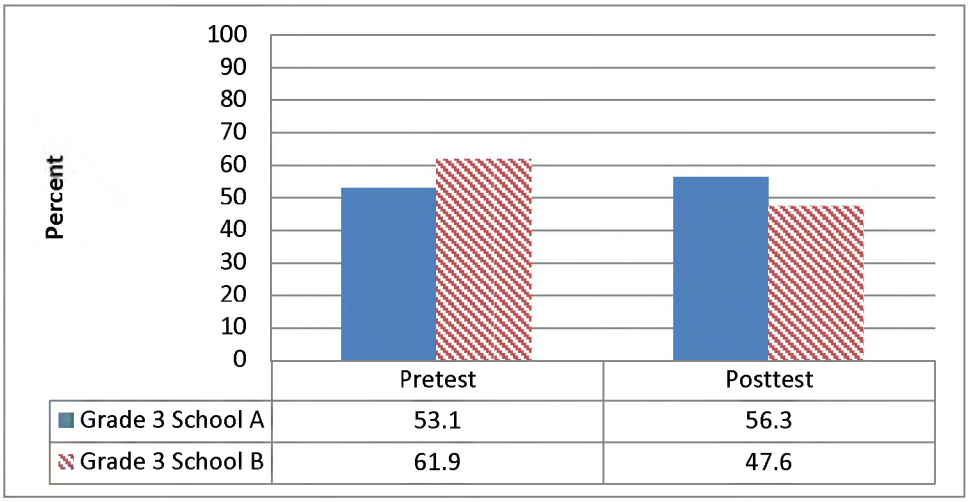


Figure 49: Average Percent Score by Class, Conjunctions

Table 20 shows progression and regression for the two subskills of parts of speech and punctuation.

Table 20: Individual Variance on Sentence Mechanics

GRADE 2	SCHOOL A				SCHOOL B			
	retained, <50%	regressed	progress-ed	retained, 50%+	retained, <50%	regressed	progress-ed	retained, 50%+
Adjective	16.1	38.7	12.9	32.3	12.0	20.0	36.0	32.0
Noun	6.5	29.0	29.0	35.5	4.0	12.0	48.0	36.0
Verb	80.6	3.2	12.9	3.2	32.0	4.0	64.0	0.0
AVERAGE (parts of speech)	34.4	23.7	18.3	23.7	16.0	12.0	49.3	22.7
Full stop	45.2	16.1	19.4	19.4	32.0	8.0	28.0	32.0
Question mark	45.2	35.5	9.7	9.7	20.0	8.0	32.0	40.0
AVERAGE (punctuation)	45.2	25.8	14.5	14.5	26.0	8.0	30.0	36.0
AVERAGE (all subskills)	38.7	24.5	16.8	20.0	20.0	10.4	41.6	28.0

GRADE 3	SCHOOL A				SCHOOL B			
	retained, <50%	regressed	progress-ed	retained, 50%+	retained, <50%	regressed	progress-ed	retained, 50%+
Adjective	31.3	3.1	21.9	43.8	38.1	28.6	9.5	23.8
Noun	31.3	12.5	43.8	12.5	28.6	33.3	23.8	14.3
Verb	31.3	18.8	21.9	28.1	23.8	23.8	23.8	28.6
Conjunction	21.9	21.9	25.0	31.3	23.8	28.6	14.3	33.3
AVERAGE (parts of speech)	28.9	14.1	28.1	28.9	28.6	28.6	17.9	25.0
End punctuation	37.5	40.6	6.3	15.6	28.6	38.1	14.3	19.0
Comma	50.0	12.5	9.4	28.1	23.8	9.5	28.6	38.1
AVERAGE (punctuation)	43.8	26.6	7.8	21.9	26.2	23.8	21.4	28.6
AVERAGE (all subtasks)	33.9	18.2	21.4	26.6	27.8	27.0	19.0	26.2

In the parts of speech subtasks, the grade 2 classes of the two schools performed on par in the pre-test. However, B2 improved dramatically on the post-test, showing rates of progression of 20 percent or more on all subskills, accompanied by regression rates at 12 percent or lower. A2 showed a converse pattern of high regression and low progression, with regression over 20 percent on all subskills and progression rates between 14 and 20 percent. Overall, punctuation followed the same pattern as parts of speech for grade 2, with similar pre-test scores in both schools but high rates of progression and low rates of regression in B2, and high rates of regression and low rates of progression in A2.

B3 had a moderate progression rate of 23.8 percent on noun identification, while A3 achieved a 43.8 percent progression rate. Regression rates overall were higher in B3 than A3, while progression rates were higher in A3 than B3. Conjunctions showed similar patterns of high regression and similar retention in both schools, but a higher progression rate in A3.

In punctuation, grade 3 learners in both schools showed high regression and low progression rates on end of sentence punctuation. Particularly, learners were challenged by the exclamation mark. Comma use showed low to moderate rates of regression in both schools, but with a higher progression in B3. Ultimately, although the two schools began within 4

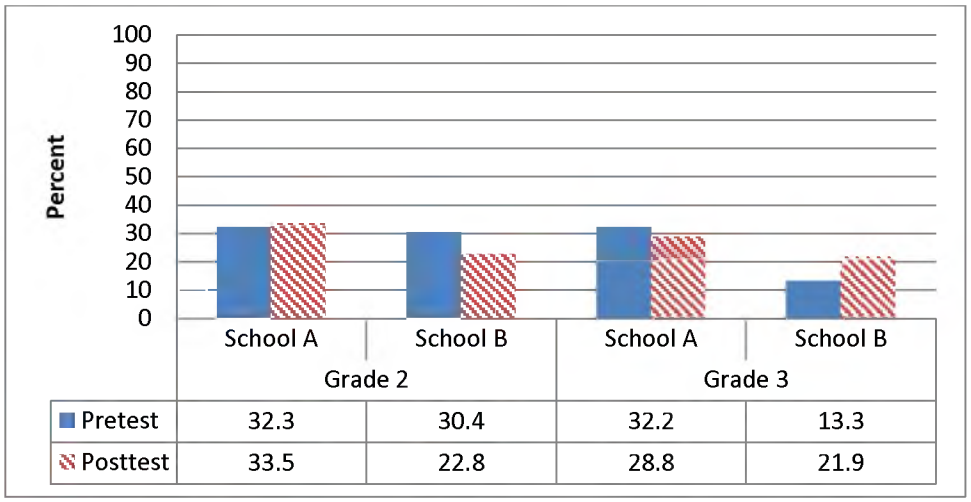
percentage points of each other, due to the high progression rate on comma usage, on the endline B3 outperformed A3. Both schools had high degrees of regression on this item.

The most difficult question in this section for grade 2 learners in both schools was verb identification, with only 5 percent of learners in the grade 2 sample answering the question correctly on the pre-test. Class A2 also struggled with question marks, particularly on the post-test. Interestingly, for the parts of speech subskill grade 3 learners in both schools struggled in identifying nouns, which grade 2 found to be the easiest questions.

5.3.4 Explicit Factual Knowledge

Performance in the category of factual knowledge was the lowest on both tests, indicating that uptake and retention of facts introduced in the program is not as strong as uptake and retention of skills-based categories. Factual knowledge was divided into two subskills: story-based knowledge, defined as program-specific components such as the names of characters or storyline events, and reinforced knowledge, or facts introduced in the program relating to real-world applications, e.g., information about cultural practices.

As shown in Figure 50, grade 2 performance in both schools on story-based knowledge was lower than reinforced knowledge, though Grade 3 was inconsistent in this respect, with B3 performing better on reinforced knowledge on the post-test only, and A3 performing better on reinforced knowledge on the pre-test only.



	School A	School B	School A	School B
	Grade 2		Grade 3	
Pretest	32.3	30.4	32.2	13.3
Posttest	33.5	22.8	28.8	21.9

Figure 50: Average Percent Score by Class, Overall Explicit Knowledge

For both tests and both schools in grade 2, the subtask with the lowest performance was story-based knowledge. Overall, performance on explicit factual knowledge in A2 rose slightly between the two tests, with higher gains in reinforced knowledge than story-based

knowledge. In class B2, performance on reinforced facts was higher than in class A2 initially but fell 16 percent on the post-test, contributing to an overall loss in the category.

Grade 3 showed different trends. A3 showed gains in their ability to correctly respond to program storyline questions, while in the reinforced question category scores dropped by 34 percent. B3 showed a loss on story-based questions, but made remarkable gains in reinforced questions, improving 650 percent from the average pre-test score.

Figure 51 shows the average performance on explicit story-based knowledge, while Figure 52 shows the performance on reinforced knowledge, or facts pertaining to the real world environment introduced by the program.

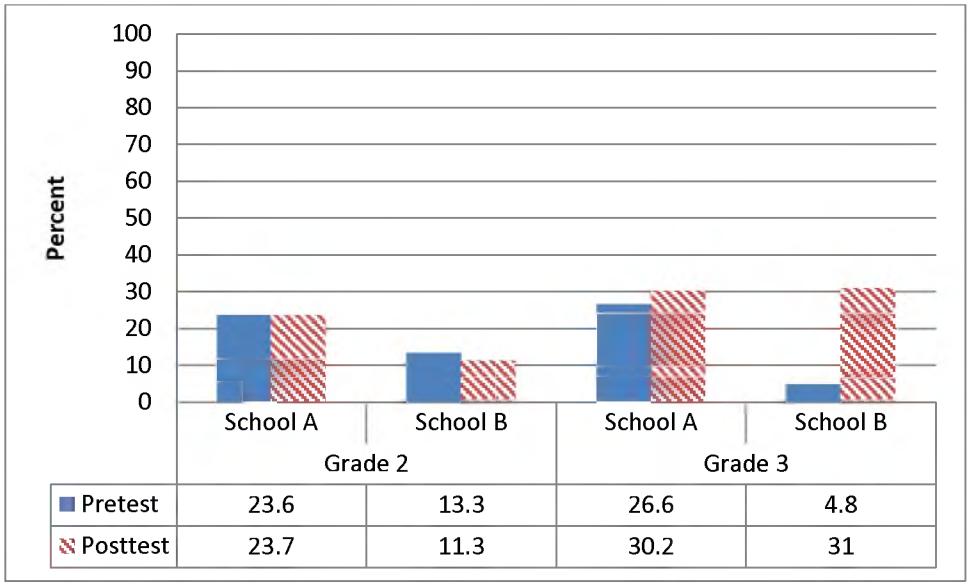


Figure 51: Average Percent Score by Class, Story-based Factual Knowledge

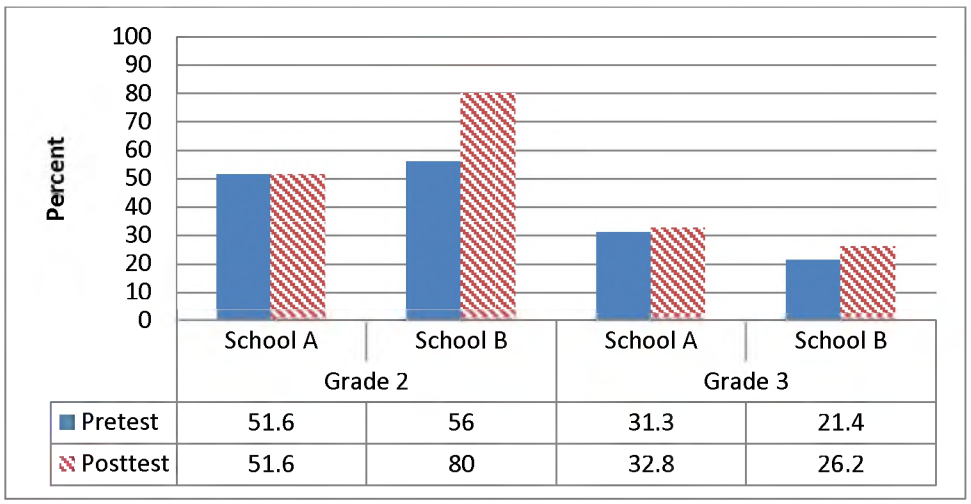


Figure 52: Average Percent Score by Class, Reinforced Factual Knowledge

Although the correct response rates for retention of factual knowledge appear relatively stable in Figure 51 and Figure 52, the analysis of individual variance presented in Table 21 shows a lot of movement, particularly on story-based factual knowledge. In both schools in both grades, progression rates were within 25 percent of regression rates on story-based knowledge, which led to the appearance of stability when overall scores were examined. However, regression and progression rates indicate a high degree of variability in this subskill, indicating that learners who are able to access story-based factual knowledge seem to do so inconsistently. In fact, only 1.8 percent of learners were able to retain a score of 50 percent or better on story-based factual knowledge.

Reinforced items had relatively higher retained scores in three classes, with the exception of B3 which had a very high progression rate. Regression rates were inconsistent, with rates over 40 percent in B2 and A3, and a rate below 20 percent in A2. The low regression rate in B3 cannot be considered as the average pre-test score for the school was below 5 percent.

Progression rates on reinforced facts ranged from 16 to 52 percent, and were higher than regression in A2 and B3, but lower than regression in A3 and B2. Given such inconsistency it is difficult to ascertain what could have caused gains and losses in this area, though the retained percentage echoes program use to some degree.

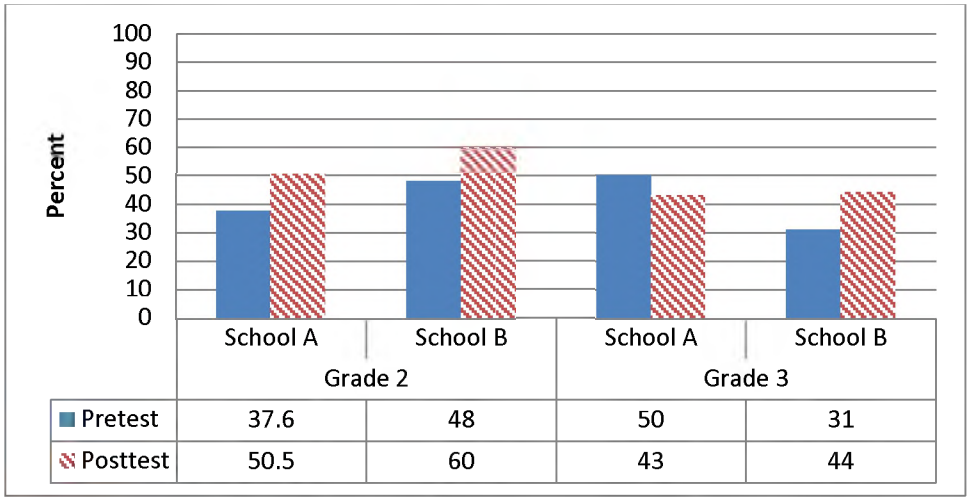
Table 21: Individual Variance on Factual Knowledge

GRADE 2	SCHOOL A				SCHOOL B			
	retained, <50%	regressed	progress-ed	retained, 50%+	retained, <50%	regressed	progress-ed	retained, 50%+
Story-based	22.6	41.9	32.3	3.2	56.0	24.0	20.0	0.0
Reinforced	22.6	19.4	25.8	32.3	12.0	44.0	20.0	24.0
AVERAGE (all subskills)	22.6	30.6	29.0	17.7	34.0	34.0	20.0	12.0

GRADE 3	SCHOOL A				SCHOOL B			
	retained, <50%	regressed	progress-ed	retained, 50%+	retained, <50%	regressed	progress-ed	retained, 50%+
Story-based	25.0	31.3	37.5	6.3	33.3	33.3	33.3	0.0
Reinforced	15.6	43.8	15.6	25.0	38.1	4.8	52.4	4.8
AVERAGE (all subskills)	20.3	37.5	26.6	15.6	35.7	19.0	42.9	2.4

5.3.5 Comprehension

Learners in grade 2 were given two direct questions (one easy and one more difficult) and one inference question, while grade 3 learners answered direct, inference and synthesis questions. Results are shown in Figure 53.



	School A	School B	School A	School B
	Grade 2		Grade 3	
Pretest	37.6	48	50	31
Posttest	50.5	60	43	44

Figure 53: Average Class Scores, Overall Comprehension

Overall, B2 performed better on the comprehension section than A2, which is not unexpected given the comparative reading levels of the two schools in grade 2 on the EGRA. However, both schools made gains over the retention period, although in different subtasks. A2 gained 19.4 percent on direct questions (Figure 54), while performance on the inference question remained static (Figure 55). B2 made gains in the both the direct and inference categories, but gained more in the inference category.

Grade 3 showed opposite patterns in the two schools, with overall averages dropping in A3 on direct and synthesis questions, while performance improved in all three subtasks in B3. In both schools initial and final results showed the lowest achievement on the inference task. Variability in the inference task performance was also lowest, with a 1.5 percent gain in A3 and a 4.6 percent drop in B3. As shown in Figure 56, similar trends were seen in synthesis, with A3 decreasing 12.5 percent from the pre-test to the post-test and B3 nearly doubling its pre-test score on the post-test.

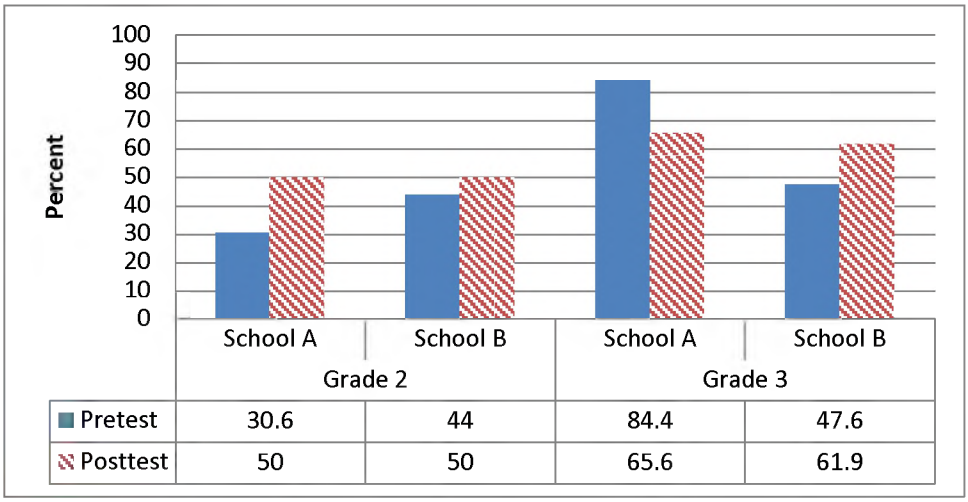


Figure 54: Average Class Scores, Direct Questions

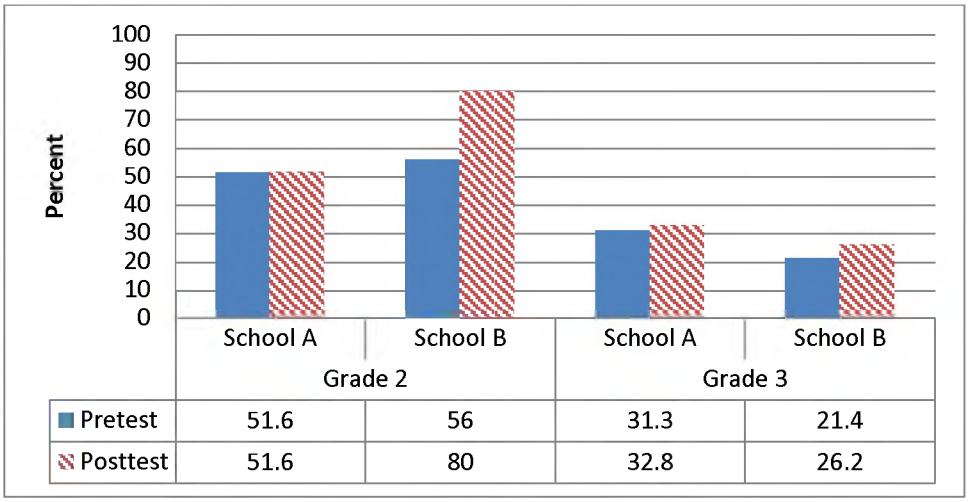


Figure 55: Average Class Scores, Inference Questions

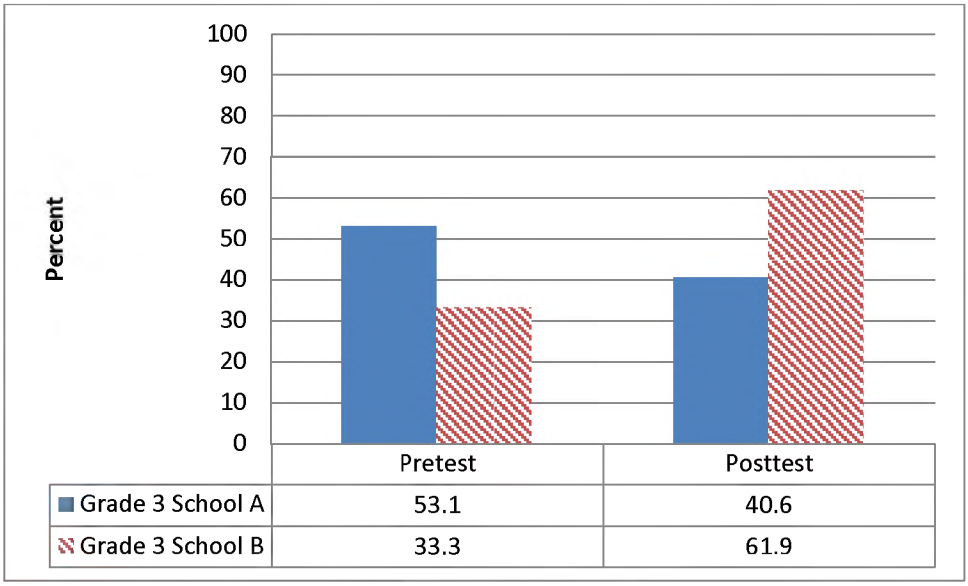


Figure 56: Average Class Scores, Synthesis Questions

Table 22 displays the relative regression rates of the four classes, and shows a high degree of variance, with the lowest regression in B2 at 10.7 percent, and the highest in A3 at 27.9 percent. Regression rates were higher in school A than in school B, and highest in A3. Progression rates were above 20 percent for all classes.

Table 22: Individual Learner Variance in Comprehension

GRADE 2	SCHOOL A				SCHOOL B			
	retained, <50%	regressed	progress-ed	retained, 50%+	retained, <50%	regressed	progress-ed	retained, 50%+
DIRECT	25.8	9.7	35.5	29.0	32.0	16.0	16.0	36.0
DIRECT 2	51.6	12.9	25.8	9.7	40.0	12.0	24.0	24.0
INFERENCE	22.6	25.8	25.8	25.8	16.0	4.0	28.0	52.0
Average (all)	**33.3**	**16.1**	**29.0**	**21.5**	**29.3**	**10.7**	**22.7**	**37.3**

GRADE 3	SCHOOL A				SCHOOL B			
	retained, <50%	regressed	progress-ed	retained, 50%+	retained, <50%	regressed	progress-ed	retained, 50%+
DIRECT	0.0	34.4	15.6	50.0	23.8	14.3	28.6	33.3
SYNTHESIS	31.3	28.1	15.6	25.0	33.3	4.8	33.3	28.6
INFERENCE	37.5	18.8	37.5	6.3	57.1	14.3	23.8	4.8
INFERENCE2	50.0	28.1	12.5	9.4	61.9	14.3	14.3	9.5
Average (all)	**29.7**	**27.3**	**20.3**	**22.7**	**44.0**	**11.9**	**25.0**	**19.0**

Both grade 2 classes had the most difficulty with the second direct question: "What did the tree promise the woodcutter?" This is likely because the wording provided in the answer did not exactly match the wording in the passage. However, grade 3 learners had more difficulty with inference questions. Learners in both classes struggled with an inference question which asked them to define 'venomous' based on a passage about frogs and spiders which used but did not define the word. For all questions, A3 initially responded better than B3, but high regression rates and low progression rates in A3 combined with low regression and high progression rates in B3 led to near parity on the post-test in most questions.

Table 23 provides a comparison of BFI program use and comprehension gains on the EGRA, which gives some indication that higher gains translate to higher losses, as three out of the four classes display regression within 5 percent of EGRA gains. However, these are offset by

the aforementioned higher progression gains in three of the four classes. Hours of program use shows an inconsistent pattern when weighed against both regression and progression, and shows a weak positive correlation with pre-test scores (r = .3173). With only four observation points, significance cannot really be established with any reliability.

For grade 2, the EGRA measured text based comprehension while the retention test measured aural comprehension, which contributes to the higher score on the retention pre-test than the EGRA endline in A2. However, B2 shows a lower comprehension score on the pre-test than the EGRA, which cannot be easily explained. Both grade 2 schools improved by 12-13 percent on the post-test, which may be attributed to repeat exposure.

A3 was the only class to exhibit high regression rates on comprehension, although all exhibited high progression rates. This may have to do with the pace of learning through the year.

Table 23: Program Use, EGRA and Retention Results for Comprehension

		Hours of BFI Use	Gains, Comp. (EGRA)	EGRA avg performance at endline	Pre-test (Avg % correct)	% Learners Regressed	% Learners Progressed
Grade 2	School A	25	11.4	18.6	37.6	16.1	29.0
	School B	10	48.6	74.3	48	10.7	22.7
Grade 3	School A	18	28.5	80.0	50	27.3	20.3
	School B	3	20.0	80.0	30.9	11.9	25.0

5.4 Conclusions on Skills and Retention

This chapter set out to discuss the results when a retention pre and post-test were designed and used to answer the following research question:

1.) To what extent is knowledge gained by the conclusion of the BFI program sustained after use has ended?

2.) What types of knowledge, procedural or fact-based, are sustained?

3.) Is there a pattern to retention and non-retention in each task and subtask? In other words, was it possible to predict learners who would and would not retain knowledge?

4.) For tasks included in both tests, does improvement on the EGRA correlate in any way to the retention rate?

Before discussing the results, it is important to note that performance was low on the retention tests. Overall average performance on the first retention test, which measured the uptake of specific skills taught in the BFI program and the South African curriculum, was below 45 percent on all categories: manipulation, mechanics, factual knowledge and comprehension. Even on the post-test, none achieved 50 percent.

This is not measured against a baseline and therefore cannot be taken to conclude that the BFI program was ineffective in improving literacy; on the contrary, results from the EGRA indicate that the program is more effective than traditional pedagogy in improving some literacy skills of early readers, particularly reading speed and comprehension. However, it does indicate that the knowledge of this sample of learners at the end of grade 2 and 3 is below the expectations of the national Department of Basic Education, which expects at least 70 percent of learners to perform at 50 percent or better on the Annual National Assessments (ANA), which test the same knowledge as the retention test. ANA examinations were unfortunately not held in 2015 for political reasons, so a comparison between the retention test and ANA performance could not be made.

With regard to the first question, to what extent knowledge gained by the conclusion of the program was sustained, the results suggested higher overall performance rates on the post-test as compared to the immediate skills uptake pre-test. While some past literature has found that practical knowledge decreases over intervals of weeks or months (Carpenter et al., 2007; Haynie, 2007), *the anticipated drop in retention scores between tests did not occur, and was in fact reversed in most cases.* In aggregate, scores increased in the retention period in three of the four classes, and dropped by just over one percentage point in the final class (A2).

This is remarkable because learners did not even attend school for nine of the eleven weeks of the retention period, and so were subject to very little further formal instruction. These results are in clear defiance of both conventional wisdom and prior research, which advise researchers avoid testing at the beginning of the year due to the 'January drop' – a phenomenon in which learners return to school after a long absence having forgotten a notable portion of what they learned the previous year.

Possible reasons for this include the repeated introduction of content and skills and improved familiarity with the test and testing environment. With regard to the testing environment, as noted by Averell and Heathcote (2009), context plays an important role in retention, and repeated encounters with the context of testing through end of year exams may have improved recall.

Regarding repetition, information which is encountered on multiple occasions may be more likely to be retained, and for longer intervals, than information encountered only once, as repeated exposure can increase the 'storage strength' and slow changes in accessibility or retrieval as described by Bjork and Bjork (1992). A further explanation which is supported by literature is that test repetition played some role in performance gains, through the direct effects of testing in improving retrieval of information or the enhancement of transfer-appropriate processing (Toppino & Cohen, 2009; Roediger & Karpicke, 2006; Wheeler et al., 2003).

However, the conclusion that repeated exposure and/or repeated testing are primary contributors is challenged by *high rates of progression and regression*. Thus, with regard to question 3, whether retention was predictable or patterned, the findings of this study suggest that the concept of 'retention' should be interrogated more deeply when group averages are considered as they were in this study, as there were equal and sometimes greater numbers of learners 'remembering' what they had 'forgotten' than learners 'forgetting' what they had 'remembered'. Rather than true 'retention', increases in average scores on the retention test indicated higher retrieval rates, including from learners who initially scored incorrectly on the uptake test. This means that information was accessed periodically rather than consistently by participants, which cannot be explained by multiple choice error due to the open nature of test questions.

Ultimately, the findings of this report with regard to retention and its predictability support the assertion of Toppino and Cohen (2009) that the mechanisms behind retention effects are

not truly understood, and it was not possible to form concrete conclusions about the characteristics of learners or questions which would lead to higher retention rates, or what caused learners to gain, lose or retain knowledge between the pre and post-tests.

With regard to the second question, whether explicit or implicit knowledge had better retention, it is the finding of this research that *uptake and retention scores are lower on explicit knowledge than on implicit knowledge, and lowest on programme-specific explicit knowledge.* For all grades and classes, performance on explicit factual knowledge was lower on the pre-test and post-test than any other skill category, and in all classes performance on story-based knowledge was lower than explicit factual knowledge with real-world applications, which contradicts findings by Averell and Heathcote (2009) in which performance on explicit knowledge was higher than on implicit knowledge.

Additionally, regression and progression rates indicate a high degree of variability in explicit story-based knowledge, meaning that learners who are able to access story-based factual knowledge seem to do so inconsistently. This was also true of explicit factual knowledge in some cases, with class B3 improving in reinforced factual knowledge by over 650% between the retention pre-test and the post-test. As it is unlikely large numbers of learners spontaneously encountered the facts included in the post-test (e.g.; the number of languages spoken in Africa), it is likely this also indicates that reinforced factual knowledge is also more likely to be sporadically retrieved.

Additionally, B3 outperformed A3 on the post-test in reinforced knowledge, indicating that higher exposure did not translate to higher sustained gains in explicit factual knowledge when that knowledge can be applied to a real world context. This may suggest that direct repetition is not the only route to long-term memory storage of factual information. It was suggested by Baddeley (2003) that neurolinguistic processing through the short-term memory includes a 'visio-spacial sketchpad' to manipulate inputs from sensory organs, long and short-term memory; it could be that the functionality of such a structure extends to long-term memory creation and storage as well.

1.) For tasks included in both tests, does improvement on the EGRA correlate in any way to the retention rate?

With regard to the final question, on whether there was a correlation between EGRA performance and performance on the retention test, there were three areas which were possible to analyse: overall performance; phoneme manipulation; and comprehension.

When EGRA performance is compared to retention test performance in all three areas, some anomalies occur. Particularly, *weak results on the EGRA did not translate to weak results on the retention test.* In particular, the high performance of A2 on the retention test as compared to the EGRA and the high performance of B3 on the EGRA as compared to the retention test require consideration.

EGRA performance indicates that the majority of learners in class A2 could not read the grade 1 passage even in October, while the majority of B3 learners were not only able to read the passage but also able to answer a majority of the comprehension questions. Considering their limitations, A2 learners were able to complete a remarkable percentage of the tasks set to them on the retention test, while B3 learners had lacklustre results on retention tests skills given the high literacy rates in the class. While it is possible that sampling error in the EGRA component played a role, with a 10 to 20 percent sample it is unlikely to alone account for the discrepancy, especially as teachers made no indication that the selection of learners was abnormally biased towards low or high performers.

There are likely two factors to be considered in both cases: The relative hours of program use, and the relative difficulty of the two examinations.

While there are too few data points to declare a concrete relationship between hours of program use and performance on the retention test, there are indications of a link. Class A2 had the highest number of hours of program use during the intervention period, which translated to a high exposure to and repetition of the tasks in the program, while B3 had the lowest hours of program use at only 3 hours.

The first impact of low hours of program use can be seen in the factual knowledge category, which accounted for 23 percent of the grade 3 retention test and 20 percent of the grade 2 test. Preliminarily, B3 performed at 13.3 percent on this category, while all other classes performed between 30 and 32.5 percent. This contributed to overall low retention pre-test scores in class B3.

The second impact of hours of program use has to do with the amount of exposure to the types of tasks presented in the retention test. While the programme only covered skills which are included in the national curriculum, the way in which they were presented and the difficulty level of the tasks may have been higher than learners routinely receive through traditional instruction. For example, the fact that both grades in school A outperformed school B in the syllabification and sound segmenting sections may be reflective of the

difference in time spent with the program, which focused on breaking words into syllables and sounds rather than syllables and letters as is often done in traditional class instruction. Pedagogical difficulty in the programme may also have played a role in comprehension, where a literacy needs analysis of teachers in foundation phase showed a heavy emphasis on direct questions and little attention to inference, synthesis, evaluation and appreciation (Shiohira & Shezi, 2013). All of these categories of comprehension questions are included liberally in the BFI program.

In addition to difficulty and presentation, repetition must also be considered. While class B3 engaged in reading and sentence mechanics through traditional instruction, they may not have been given sufficient exposure to the skills and accompanying instruction of the BFI program to firmly establish the concepts included. Thus, higher exposure to the program may have translated to better performance for A2, and poorer performance for B3.

Concerning the second factor of relative difficulty, if the EGRA and the grade 2 retention test are compared, the EGRA is the more demanding as it is a one-on-one test which leaves no room whatsoever for external support, while the paper-based test was mostly read to the learners as per normal foundation phase practice (see the methodology sections of Chapter 4 and 5 for further details). In other words, learners were provided with literacy support on the paper-based test which was not available on the EGRA.

However, if the EGRA is compared to the grade 3 retention test, the EGRA is the easier examination. First, the EGRA was written with a difficultly pitched at late grade 1 or early grade 2, while the retention test provided a grade-level passage and comprehension questions. It is likely that class B3 was not able to achieve parity with A3 because of A3's increased exposure to not only grade level passages but also different types of comprehension questions through program use.

CHAPTER 6: INTEGRATING QUANTITATIVE AND QUALITATIVE DATA

The purpose of this chapter is to explore the ways in which the qualitative and quantitative data interact for the three classrooms observed, grades 2 and 3 in school A and grade 2 in school B, in order to answer the following research questions:

1.) What observed classroom, school and grade factors likely contributed to uptake and retention of literacy skills?

2.) What is the relationship, if any, between hours of program use and skills uptake and/or retention?

Chapter 3 outlined the qualitative data collected through classroom observation, teacher interviews and learner drawings. It was noted that while learner engagement seemed high and enthusiastic in all schools and classes, there were significant differentials in outcomes as measured by the quantitative EGRA and retention test analyses. A summary of outcomes by class is outlined in Table 24.

Table 24: Summary of Quantitative Results and Classroom Observations

Class	Frequency of tablet use	Final EGRA Score	Change in EGRA score	Final Retention Score	Change in Retention Test	Tablet Lesson and Language Observed	Number of learners per tablet
A2	1.47 / wk	54.1	13	36.1	-1.1	Lesson 1, Northern Sotho	1 to 4
A3	1.04 / wk	185.5	32.5	40.6	2.4	Lesson 5, Northern Sotho	1
B2	0.48 / wk	180.5	49.8	51.3	11.3	Lesson 2, English	4 to 5; pairs observed
B3	0.18 / wk	220.9	27.2	36.6	5.2	NONE	

6.1 Hours of Program Use

As displayed by Table 24, the highest performing class on the literacy skills was B3, the class with the lowest program use. However, the classes with the largest gains were B2 and A3. These two classes were also the highest performing on the retention test. A2 had the highest hours of program use and the lowest outcomes on all tests, as well as the smallest gains between test periods.

As the summary in the chart suggests, hours of program use overall showed no meaningful correlation to change in performance on EGRA skills. The lack of correlation between program hours and overall performance indicates that factors beyond hours of program use influence its effectiveness. In other words, differentials in the average performance of classes may be due more to how the technology was used than to the presence of technology. Consideration of results in light of the qualitative data yields some insight in this regard.

In addition to hours of program use, three potential areas of influence were noted in the qualitative research: 1) The use of the tablets within the classroom; 2) The types of follow-up activities completed by classes; and 3) types and frequency of peer engagements were considered as factors which may have influenced outcomes.

In this chapter, these three aspects of engagement will be discussed relative to the quantitative outcomes achieved. A summary of gains during and after the treatment period and the observed classroom behaviours is presented in Table 25[26].

[26] Note that B3 was not observed, for reasons outlined in Chapter 3.

Table 25: Summary of Quantitative and Qualitative Data by Class

		CLASS A2				CLASS A3				CLASS B2			
		Final Score: EGRA	**Change in EGRA score**	**Final Score: Retention**	**Change in Retention Test**	**Final Score: EGRA**	**Change in EGRA score**	**Final Score: Retention**	**Change in Retention Test**	**Final Score: EGRA**	**Change in EGRA score**	**Final Score: Retention**	**Change in Retention Test**
		54.1	**13**	**36.1**	**-1.1**	**185.5**	**32.5**	**40.6**	**2.4**	**180.5**	**49.8**	**51.3**	**11.3**
QUALITATIVE OBSERVATIONS	Class Style	Teacher facilitating student-led groups as needed				Independent learning; minimal teacher interaction with learners				Teacher-led classes; learners respond to teacher prompts			
	Class Work	As follow-up, learners did exercises in the DBE workbook but did not explicitly link these to programme activities.				As follow-up, the teacher has learner retell the stories from the program and write word lists from the program on the board.				As follow-up, learners write sentences and performed linked activities from the DBE workbooks.			
	Teacher Engagement	Observation of learners using the program; elicited support; and unsolicited support which seemed to come in response to perceived difficulty or slowness in following instructions.				Observation of learners using the program; elicited support; and unsolicited support which seemed to come in response to perceived difficulty or slowness in following instructions				Adapt the tablet to the typical role of a teaching and learning aid in the classroom, similar to how a teacher might use a poster or another teaching prop.			
	Peer Engagements	Constant interaction and peer support, open celebration at completing tasks. Learners self-assigned a group leader; occasional leading by other learners				Minimal interaction; tablets were individual work. Suggestions from one learner to another were rebuffed by recipients.				Collaboration within and between groups, although learners were scolded for this			
	Technology-System	Used the tablets 1.47 times per week on average during the treatment period, for 1.5 hours per session. Indicated minimal preparation or planning for BFI lessons.				Classes used the tablets once per week for 30-45 minutes. The teacher prepared by going through the lessons to ensure she could do all activities.				Teachers reported using the program twice per week for 1.5 hours, once in Northern Sotho and once in English, though they averaged only .48 sessions per week over the treatment period.			

6.2 Tablet Integration into the Classroom

Ultimately, three different approaches were observed in three school/grade combinations, each of which created completely different relationships between the teacher, technology and the learner.

In class A2, a group peer learning approach was utilized, with the traditional teacher role transforming into an active facilitator role. Learners were given relative freedom with minimal instruction from the teacher and a very loosely structured classroom environment. It is perhaps as a result of this that 2A learners explored the tablet functions beyond the program to the extent observed, and exhibited strong independent problem-solving skills while using the tablets. However, the lack of planning, structure and reinforcement likely contributed to relatively low gains in literacy in the classroom, though high hours of program use allowed learners to perform well on the method-dependent retention test.

Class B2 engaged in teacher-led classes, with the devices functioning as a learning support material and teachers maintaining the traditional role of authority in the classroom, providing all instruction and corrective feedback, eliciting answers from learners and directing learners on when and what to do with the program at each stage. As the teachers were strong in this school, program activities and literacy were reinforced through the year, yielding good results on both the method-independent EGRA examination and the retention test. However, due to the tight discipline and structure, learners did not explore the tablets to the same extent as 2A and remained unable to perform basic capacity functions on the tablet or problem-solve when difficulties were encountered with the program or the device.

Class A3 used the tablets individually, with learners engaging independently with the devices and devices effectively supplanting the traditional role of the teacher in offering instruction and corrective feedback, while the teacher assumed an observation or monitoring role. Class A3 also showed higher gains during the treatment period than the control period, particularly in word reading and comprehension. However, gains did not approach B2, indicating that teacher-led classes were more effective even than independent work, possibly because teachers were able to supplement the instruction from the program with their own expertise using the teacher-led format.

6.3 Follow-up Activities and Integration into Other Pedagogy

Teachers engaged very differently in terms of preparation and follow-up exercises, as well. Only the teachers of A3 truly prepared for each lesson beforehand by going through all

activities, which is interesting considering they were the least involved in actual program delivery. Both grade 2 classes did no preparation beyond charging the tablets and bringing them to class. However, in terms of linking the program activities to the curriculum, class B2 teachers were the most active, supplementing the programme with writing exercises and linking program activities to exercises in the DBE workbooks. This was followed by class A3, which likewise required writing exercises. Class A2 teachers drew no specific structured connections between the program and traditional pedagogy for learners.

The observed higher retention rates in implicit knowledge in B2 than A2 are likely attributable to two factors. First is the amount and type of reinforcement given in B2. When teachers were asked what follow-up activities to the program learners were asked to perform, teachers in school B immediately identified 'parts of speech' practice exercises in the DBE workbook as well as sentence writing (which would include punctuation practice), while in school A learners were asked to write words they remembered from the program in their notebooks or on the board. The follow up exercises were undertaken independently, meaning that learners in B2 were given individual practice on sentence mechanics, while learners in A2 were not. A second related factor is that learners in school B practiced translating programme knowledge to their own use of language, while for learners in school A these skills were confined to the tablet program and were not used for language production, which likely affected skills uptake and retention.

The opposite effect is notable in grade 3, where initial high scores in B3 drop below the achievement level of A3 on the post-test, likely due to the same phenomenon in reverse: While learners in both schools were given instruction on these elements through DBE workbooks and other exercises, learners in A3 were given additional practice through program use – and, notably, they were required to answer the program exercises individually. This meant that, for example, they were given tutorials on parts of speech every time they answered incorrectly, both giving them incentive to answer correctly and providing individual instruction as many times as needed. These factors likely contributed to the high retention of these items in class A3 as compared to B3.

This may link specifically to lower outcomes on the retention test in A2 than in B2. In fact, the only area in which A2 outperformed class B2 was on story-based factual (explicit) knowledge, which can be explained as due to the pedagogical style of the teachers. Performance on story-based knowledge was lowest of all classes in B2, most likely due to the fact that with the teacher-centred BFI instruction only the teacher's tablet was permitted

sound. This did not enable learners to fully engage the story through the audio-visual medium, which meant that B2 learners engaged story-based facts only through the print medium, while learners in other classes were able to engage the storyline and characters through both audio-visual and print media. Repetition likely played a role in retention. As Bjork and Bjork (1992) propose, higher exposure rates can lead to better storage and retrieval of information, which would explain the lower retention rates of explicit knowledge in B2.

However, teachers leading classes as was done in B2 provided different types of input; for example, additional sentence examples, explanations and code-switching to improve learner understanding of directions or content. This may have contributed to higher uptake scores in B2 than A2 on implicit knowledge tasks in particular.

6.4 Peer Interactions

The results indicate that the most effective use of the technology was made in class B2, which gained the most during the intervention period and also had the largest differential between the control and intervention periods. In this class, the teacher used the device in much the same way as a poster or textbook, leading the classes through each exercise. Using the tablet in this way, the teacher functioned in her traditional role, providing explanation and feedback to learners. The effect was no doubt bolstered by the overall competence of teachers in the school, which is high performing.

The second most effective use of the program was in class A3, which practiced individual use of the tablets with headphones and almost no interference from the teacher. This class gained 32.5 points during the intervention period, which may indicate that learners who engaged the program as if it were the teacher also gained significantly from the interaction.

Class B3 was the only one to gain less during the intervention period than the control period, and used the tablets between 3 and 5 times using a similar group work strategy to class A2, which gained only 13 points during the intervention and had the lowest endline achievement (more than 100 points below all other classes). The evidence thus provided by the study indicates that small and large group work were the least effective uses of the program.

While initially this is counter-intuitive, as peer and collaborative learning are recognized as useful pedagogical tools (Boud et al., 1999), it is theorized that learners engaging in teacher-led tablet classes benefitted from increased difficulty of the exercises as well as lesson discipline, while learners engaging independently benefitted from these aspects as well as relying on their own skills and targeted remedial instruction for individual gaps in knowledge.

However, the learners working in groups can rely on the stronger group members to provide answers, only engaging with material or content they are already familiar with. Observation of other learners providing the correct answers may not be enough engagement for gains in knowledge, skills or competencies. On the other hand, it must be noted that the problem-solving and technical skills of learners in A2 far outstripped those observed in B2, although B2 was stronger in skills attainment.

A second indication of this finding is that independent work is crucial to consolidation of literacy skills. Classes which did not engage in independent work either as part of tablet classes or as follow-up activities to tablet lessons showed lesser gains through the intervention period on EGRA as well as lower achievement on the retention test. Therefore, it can be concluded that learners engaging only in group work are unlikely to gain and/or retain knowledge.

6.5 Conclusory Remarks on the Integrated Analysis

The most significant finding in comparing results is that literacy uptake on a method-independent test did not necessarily translate to literacy uptake on the method-dependent test. Even on skills with significant overlap, phonemic awareness and comprehension, high performance on the method-dependent test did not translate to high ability in the area as scored by the outcomes-based or method-independent test. This may indicate that testing methodology plays a role in affecting learning outcomes. Alternately or in conjunction, it may indicate that curriculum skills mastery is insufficient for literacy outcomes in some situations.

The situation for this study in which this was the case was A2. In all other classes, the intervention was a true supplement to routine classwork, with structure provided by the teacher and the program linked to class activities. In A2, however, program use was a chaotic and disorganized event which occurred in relative isolation from the rest of the curriculum, without the teacher making links between regular classwork and program use.

In other words, while this study was primarily a linguistic study and did not attempt to rate teacher and teaching quality, these could be primary factors influencing the literacy outcomes attained by users of the BFI program.

CHAPTER 7: CONCLUSION

This research study set out to investigate 7 research questions:

1.) How does structured programme technology such as the BFI interact with the components of South Africa's educational system as outlined in the literature review section?

2.) What observed classroom, school and grade factors likely contributed to gains and retention of literacy skills?

3.) What impact does the Northern Sotho BFI program have in grade 2 and 3 on the following literacy skills?
   a. Letter-sound recognition
   b. Decoding of simple words
   c. Reading speed
   d. Comprehension

4.) Is there an optimal age for introducing the BFI program? Is it more effective with older or younger learners?

5.) What is the extent of skills uptake from the BFI program, and to what extent are skills gained during the use of the BFI program sustained after use has ended? What types of knowledge, procedural or fact-based, are sustained?

6.) Is there a pattern to retention and non-retention in each task and subtask? In other words, was it possible to predict learners who would and would not retain knowledge?

7.) For tasks included in both tests, does improvement on the Early Grade Reading Assessment (EGRA) correlate in any way to the retention rate?

These questions were investigated using a mixed methods research design including three parts: qualitative research, a method-independent literacy skills test and a method-dependent literacy skills test.

This section summarizes findings for each research question and presents overall conclusions drawn.

Question 1: System Placement and Context

The qualitative component contextualized the intervention in response to question 1: How does structured programme technology such as the BFI interact with the components of South Africa's educational system, outlined in the literature review section?

Data was gathered through observation, interviews and learner drawings and analysed using a general inductive approach, which incorporated the research objectives and a framework derived from data collected. Figure 57 summarizes the relationships elucidated by the qualitative data. Solid lines indicate strong or particularly influential relationships, while dotted lines express infrequent or secondary relationships experienced by some participants.

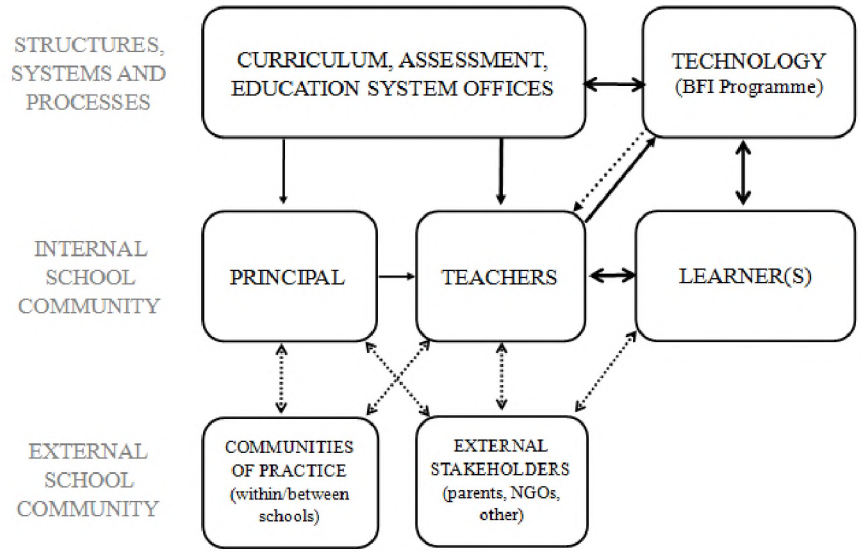


Figure 57: Interactions between Technology and the South African Education System

Five relationships emerged as key to understanding the interactions between technology and the traditional education system: 1) Interactions between policy structures and the technology in schools; 2) Interactions between teachers and the technology; 3) Interactions between teachers and learners; 4) Interactions between learners and the technology; and 5) Interactions between learners while using technology. Secondary or supportive roles were played by external stakeholders, parents and teachers of other schools.

Within the context thus established by the qualitative research, the study narrowed its focus to the primary purpose of investigating linguistics skills gained and retained during program use, to determine which literacy skills, if any, could be enhanced by technology.

Question 2: Classroom, School and Grade Factors Contributing to Observed Outcomes

Using the qualitative data, the question was investigated with a focus on elucidating classroom factors which may have influenced the quantitative results. These included four main factors: 1) Hours of program use; 2) Ways in which the devices were utilized in classroom pedagogy, or pedagogical style of the teacher with the device; 3) Integration of the programme with prescribed and/or regular classroom activities; and 4) Peer interactions while using the device.

These four factors were investigated alongside gains in EGRA and retention test results.

Hours of program use showed no meaningful correlation with either performance or change in performance. Classes with the highest hours of use showed the lowest gains, and classes with the lowest hours of use showing the highest scores and gains. Score ranks by class are outlined in Table 26.

Table 26: Performance Ranking by School and Grade

Class	Tablet Use	Final EGRA	Change in EGRA	Final Retention	Change in Retention
A2	**1**	**4**	**4**	**4**	**4**
A3	**2**	**2**	**2**	**2**	**3**
B2	**3**	**3**	**1**	**1**	**1**
B3	**4**	**1**	**3**	**3**	**2**

This indicated that factors beyond hours of program use influenced outcomes. When observations from the qualitative data were mapped against outcomes, it was found that the least effective instructional method was loosely structured, learner-centred peer learning groups, while the most effective strategy appeared to be teacher-centred traditional teaching observed in school B, grade 2, followed by individual learning.

Boud et al. (1999) advocated peer and collaborative learning as useful pedagogical tools. However, this was not reflected in this study. One possible explanation is that within the context of BFI intervention program, in larger groups weaker learners rely on the capacity of strong learners to complete activities correctly, and are therefore not given negative or instructive feedback which might improve their understanding.

A second explanation emerges in the types of additional activities assigned. The lowest performing class had the least rigorous links to the curriculum, as well as the least diversified activities. The highest-performing class had the most demanding supplementary tasks. This indicates that regular curriculum links and the type of productive language activities given to learners may be instrumental in achieving gains from a program such as the BFI.

Table 27 provides a summary of learner performance by class and school, mapped against qualitative observations of peer interaction, pedagogical style with the devices and additional activities assigned.

Table 27: Summary of Test Performance and Qualitative Data

Class	Final EGRA Score	Change in EGRA score	Final Retention Score	Change in Retention Test	Peer Interaction	Teacher Style	Additional Activities
A2	54.1	13	36.1	-1.1	Large Group Work	Learner-led classes	Workbooks
A3	185.5	32.5	40.6	2.4	Individual	Tablet-led classes	Retelling, writing words, workbooks
B2	180.5	49.8	51.3	11.3	Small Group	Teacher-led classes	Writing sentence, workbooks
B3	220.9	27.2	36.6	5.2	No obs.	No obs.	No obs.

Question 3: Impact of the Technology Intervention

The results of the EGRA examination indicate that the greatest impact of the intervention on literacy skills was improvement in reading speed and comprehension. Significant gains were made in the intervention period over the control period for both of these skills. In reading speed, the average intervention period gain was four times the average gain in the control period; for comprehension the average intervention gain was over three times the average gain in the control period. However, there were higher gains in the control period for both phoneme-grapheme assignment and word reading.

Shiohira and Shezi (2013) performed a needs analysis with early literacy teachers in the North West Province, and found teacher confidence and competency was highest in teaching phonics and word building, and that teachers engaged these activities more frequently and for

longer periods of time than independent reading or writing. Teachers spent less time on higher-order literacy skills such as passage reading and comprehension. It can be conjectured that the BFI program was therefore able to supplement teachers' strengths by engaging learners in extensive reading and comprehension practice.

However, fluency and comprehension gains were not equal across schools or grades, with school A, grade 2 in particular far behind the other classes by the conclusion of the year, and comparable results between grade 3 of school A, the lower performing school, and grade 2 of school B, the higher performing school.

The overall poor performance of class A2 indicates that although the BFI program can provide some improvement, it is not a replacement for traditional teaching for learners at an initial level below decoding ability.

However, it is also important to note that even for learners at this low level of performance, some improvement was noted. Some learners who began the intervention unable to assign phonemes to graphemes – learners who had been in the education system for a minimum of 1.5 years, and in some instances 2.5 years with no success in this area – were able to recognize and read letters. Some learners who had spent a similar amount of time in the classroom but could only read letters advanced to reading words. This provides some evidence of the effectiveness of the program even with remedial learners in some cases; however, it is unclear which additional factors influence ability to learn skills from the program at the early stages of literacy, and overall gains were greater for learners who began the program at least equipped with letter-sound knowledge.

Question 4: Performance by Age

There was no correlation between performance and age in this study, either for gains or changes in gains, in any literacy skill tested. This indicates that there is no advantage or disadvantage to engagement with technology from age 6 to 11.

However, it was noted in the qualitative observations that grade 2 learners had more difficulty with the fine motor skills necessary for operations such as drag-and-drop, and would have benefitted from program formatting which counted 'close drops' as correct.

Question 5 and 6: Skills Uptake and Retention

The retention test focused on three skill areas required by the CAPS curriculum (DBE, 2011) and included in the BFI program: phoneme manipulation, grammar and sentence structure

and reading comprehension. In the interest of measuring explicit knowledge gained and retained from the program, facts from the contents were also tested.

Overall, a 50 percent average was not achieved in any skill category, indicating overall low performance on the paper-based test. The lowest performance was on explicit knowledge, or factual knowledge presented in the BFI program, at 28 percent. This indicates that while the BFI program may have been effective in improving literacy skills such as reading speed and comprehension, it was not an effective vehicle for attaining explicit or factual knowledge.

Based on prior research, expectation was that scores would be higher on the retention pre-test, with a decline in knowledge on the post-test. However, for all implicit skill categories, achievement was higher on the post-test. Explicit factual knowledge was the only area with a decline between tests, and that of only 1 percent. Further, rates of regression (learners who "forgot" knowledge) and rates of progression (learners who "learned" rather than "remembered") were inconsistent, with higher rates of progression on many tasks, particularly for learners in the higher performing school.

There is a noted effect in the literature that testing itself is conducive to learning and improves the retrieval of information (Toppino & Cohen, 2009; Wheeler et al., 2003; Bjork & Bjork, 1992); the first conclusion of this research supports such prior literature with regard to overall performance. However, the rates of regression and – in particular – progression indicate that the process of information retrieval is complex and not well understood (Toppino & Cohen, 2009), and not only was it impossible to predict learners who would retain knowledge between the pre and post-tests, it was impossible to predict learners who would *gain* knowledge between the testing periods. Ultimately, the conclusion of this research is that single session testing may not be an adequate measure of knowledge uptake or retention, and that overall retention of knowledge was higher than expected based on prior research by Carpenter et al. (2007), which showed up to a 63% decay on both implicit and explicit knowledge over 42 days, or 35 days fewer than the retention delay in this study.

Question 7: Correlations between the EGRA and Retention Tests

This question was meant to determine whether or not a method-dependent test would link to literacy skills. Ultimately, it was found that the two could not be meaningfully compared, for two reasons.

First, the assessment practice of teachers in the early grades is to read tests in their entirety to learners, who select or write their answer after being read the questions. This contributed to higher performance on the retention test than should have seemed feasible based on EGRA reading ability in class A2.

Second, the EGRA was meant to measure early literacy ability, and so the reading and comprehension was presented at late grade 1 or early grade 2 level. Therefore, it was possible to compare performance across grades. However, high performing learners, mostly in grade 3, encountered a ceiling effect on the test, in which their true capacity could not be measured as the test was too easy for them. The retention test, on the other hand, was presented at grade level for grade 2 and 3. This led to lower performance on the retention tests for high performing EGRA classes.

A different testing structure and instruments would have to be designed and developed to draw comparisons between method-dependent skills gained and overall literacy achievement.

7.1 Other Observations and Directions for Further Research

7.1.1 Reading Speed

Reading speed showed a strong positive correlation with comprehension, which indicates that reading speed is a good predictor of comprehension performance in Northern Sotho. This echoes the results of research in English which indicate that reading speed is a good predictor of comprehension (Schilling et al., 2007; Buck & Torgesen, 2003).

The data also begins to point to a reading threshold in Northern Sotho, beyond which learners are most likely to be reading with at least basic comprehension. The data indicates that such a threshold occurs between 20 and 24 correct words per minute. There are two instances of data points with higher reading speeds and 0 comprehension. It is unclear what caused these inconsistencies, but they make up only 2 percent of data points with reading speeds over 24 cwpm, and therefore do not follow the general trend. Additionally, in other testing periods both learners scored at least a 60 percent on comprehension, which indicates that context factors played a role in poor performance.

If a larger sample and variety of reading passages confirms that the Northern Sotho reading threshold is within this range, it means that learners reading at the relatively slow pace of one word every three seconds are in general reading with comprehension. This supplements the body of evidence presented by Land (2015, 2016), Van Rooy and Pretorius (2013) and Hefer (2013) showing slower reading speeds in Bantu languages than English, even for first-language Bantu speakers, and suggests that slower reading speeds do not necessarily reflect low comprehension. This challenges the benchmarks established by NEEDU (2012) of 70 cwpm for grade 2, based on English outcomes, and may mitigate the initially alarming result that over 70 percent of the higher performing learners in grade 2 tested could not achieve that benchmark. Further, this finding could have implications for the applicability to Northern Sotho of 'chunking' theories, which are based on a shared principle that the human brain has a limited short-term capacity of around seven items of information held for twelve seconds (Baddeley, 2003).

Further and more extensive research on more types and difficulties of texts is necessary to determine if the emerging fluency threshold of 20-24 cwpm is in fact adequate for comprehension. If this threshold is confirmed, it may indicate that Northern Sotho early readers are able to leverage context more effectively than their English-speaking counterparts. This aspect could be examined in conjunction with the syntactic and morphological features

of the language in order to determine if there are and/or which features are relevant to assisting learners in leveraging context.

7.1.2 Error

Additionally, the results of the study support prior research in English, which has shown a negative correlation between oral reading errors and achievement (Blaxall & Willows, 1984; Anderson et al., 1979), particularly in lower grades (Anderson et al., 1988). The results of this study also show a significant negative correlation between error rates and comprehension.

A high average number of errors was observed for learners reading between 20 and 40 correct words per minute, contributing to high degrees of variance which affected the strength of the correlation. This is not unexpected, as learners in this range have reached a degree of fluency and are beginning to challenge the text, but have not yet fully attained whole-word recognition or the 'fast-track' neurological reading pathway which enables them to bypass phoneme-based (or, possibly, syllable-based) strategies (Coltheart et al., 2001). These learners also contributed to a weak but significant positive correlation between error rates and reading speed.

However, while the data did show that learners who attain at least 60 percent on the comprehension were less likely to have error than those with lower scores, error rates were relatively high even among learners who scored well in comprehension, with one quarter of participants who scored above 50 percent making errors in passage reading, and some participants at error rates of over 6 percent able to answer more than half of comprehension questions correctly.

Further research could determine more concretely how error impacts comprehension in Northern Sotho.

7.1.3 Phonemic Strategies of Early Northern Sotho Readers

Overall, Northern Sotho early readers could be mapped along similar trajectories to the literacy attainment levels advanced by Wolf (2008) and discussed at length in Chapter 4. And yet, while Northern Sotho early readers demonstrated mastery of letter-sound relationships, they were mostly unable to break words into phonemes in Northern Sotho, exhibiting a strong preference for syllabic segmentation, even after intense phonemic instruction in the BFI program. In fact, scores for syllabification were an average of four times higher than scores for sound segmenting.

As discussed by Modipa et al., (2010) Northern Sotho for the most part has a strict consonant-vowel sound structure, with a limited set of possible syllables, all composed of a consonant sound and an accompanying vowel (with the exception of 'ng', which can be used to close a syllable at the end of a word only), and clear syllable breaks after each vowel. As a result of this structure, speakers of the language seem to adapt rather quickly to the idea of words as made up of syllables.

A second related point is traditional pedagogy. As discussed by Zeigler and Goswami (2005), there is a strong preference for syllabic segmentation in pre-literate learners of many languages, and evidence of a reflexive relationship between pedagogy and perceptions of language which can affect how speakers define breaks in words. Thus, learners shift towards phonemic representation in a reflexive relationship with literacy training, a conclusion also supported by Tamaoka and Terao (2004) and Inagaki et al., (2000), who found that undergoing pedagogy shifted Japanese readers from mixed syllabic and moraic representation to strict morae. However, although Northern Sotho is written in an alphabetic script and current recommended practice is based on phonemes, historically teachers taught learners to read using a syllabic approach, with the components of words presented not a phonemes but as syllables. Thus, pedagogy may be reinforcing syllabic representations of words for Northern Sotho learners.

This tendency is likely compounded by the difficulties of the Roman alphabet in addressing the digraphs and trigraphs utilized to represent the additional consonant sounds in Northern Sotho: kg, ph, hlw and so on. Many learners do not seem to be able to draw the links between the phonemes of the language and their multi-consonant grapheme representations; the prevalence of errors on digraphs and trigraphs on the retention tests indicates that pedagogy emphasizes spelling of words in Northern Sotho as amalgamations of single letters rather than sounds. Even the BFI program, which engages a number of digraphs and trigraphs as phonemes, has conflicts in pedagogy. For example, after words are introduced and broken into their composite phonemes, the typing and word-building activities require learners to work with letters rather than phonemes, as there is no single-key 'kg' or 'ng' on the computer keyboard. This results in a mixed pedagogy in which phonemes are taught but word-building is based on letters rather than sounds, which may lead to confusion on the part of learners regarding phonemes and letters, which may compound an intuitive reliance on syllables rather than sounds.

However, interestingly, a reliance on syllable segmenting does not mean learners cannot read, and even read words which are unfamiliar at times. It is possible that learners are mainly reading using a whole-word reading strategy, but only one learner in the dataset was able to read individual words without being able to assign phonemes to graphemes. It is possible learners were using a *modified phonemic approach*, in which they engaged words based on the sounds contained, but internally defined and broke apart words into 'sounds' based on syllables rather than individual phonemes. Further research would be necessary to determine if this is the case, as well as which type of segmentation or pedagogy results in more efficient or effective reading in Northern Sotho.

7.1.4 Retention of Knowledge

The results of the retention test also entice further investigation. Even if testing could account for *gains* in knowledge, the fact that these gains persisted over eleven weeks of no instruction means that either the testing effect (Toppino & Cohen, 2009; Wheeler et al., 2003) or the 'storage capacity' of knowledge (Bjork & Bjork, 1992) is much greater than previously considered. Further research would be necessary to determine what influenced improvement in scores between pre and post-tests with no additional instruction, especially considering that gains were made in some classes in explicit knowledge as well as implicit or skills-based knowledge.

Pace of Learning as a Factor in Retention

A potential factor influencing retention has to do with the pace of learning throughout the year. When comprehension was considered, A3 made more than twice the gains of B3 on the EGRA test, starting from a lower benchmark and ending on par with B3. This necessitates a higher pace of learning, with less time for revision throughout the year, which could ultimately contribute to a higher rate of regression as learners may essentially be working more from short-term memory (Baddeley, 2003).

However, the scope of the research does not allow for conclusions to be drawn in this regard, and further research would be necessary to determine how pace of learning affects retention of literacy skills.

7.2 Final Remarks

The findings of the two quantitative evaluations, integrated with classroom observations, indicate that a best-practice based technological intervention can be successful in improving early grade literacy skills, to a greater or lesser extent dependent upon variables such as prior ability and method of teacher engagement and interaction with the technology. The highest test results were found when teachers engaged the technology as a support material with core instruction still in the hands of the teacher, closely followed by an independent learning model. Learner-led group work had the lowest performance returns.

Phoneme-grapheme assignment and simple word decoding showed higher gains through traditional pedagogy than with the integration of technology. Instruction provided through technology was significantly correlated to improvement in reading speed and comprehension ability, higher order skills which are often neglected in favour of phonics and word-building classroom activities.

Performance on the delayed retention test was significantly higher than performance on a baseline retention or skills uptake test for implicit or skills-based knowledge tested, while there was a slight decay between the two tests on explicit or factual knowledge. Scores on explicit knowledge on both tests were lower than scores on implicit knowledge tasks. This indicates that there was overall a high rate of retention of skills gained or practiced through the ICT program, while the program was less effective at teaching factual information.

Finally, the observations of the research conducted suggest a number of interesting avenues for further research, including more in-depth analyses of early reading processes and strategies in Northern Sotho literacy learners and further research into retention of linguistics skills over time.

APPENDIX A: Memorandum of Agreement between the Molteno Institute for Language and Literacy and the Limpopo Provincial Department of Education

This research was conducted under the funding and structures provided by the United States Agency for International Development (USAID) in partnership with the Limpopo Provincial Board of Education and the National Department of Basic Education.

The following Memorandum of Agreement (MOA) was signed by Limpopo Provincial Board of Education official representatives and the Molteno Institute of Language and Literacy. The MOA grants permission for implementation and research related to the Bridges to the Future Intervention in Limpopo Province.

MEMORANDUM OF AGREEMENT

Entered into between

The **Limpopo Department of Education**
Herein represented by Mr. Morebudi Jackson Thamaga in his capacity as Head of Department and duly authorized thereto

and

Molteno Institute for Language and Literacy
Herein represented by Mr Masennya Phineas Dikotla in his capacity as Chief Executive Officer and duly authorized thereto

TABLE OF CONTENTS

1.	PREAMBLE	3
2.	DEFINITIONS	3
3	INTERPRETATIONS	3
4.	OBJECTIVES OF THE SUPPORT	4
5.	BACKGROUND	5
6.	COMMENCEMENT AND DURATION OF THE SUPPORT	5
7.	FUNDING	5
8.	ROLES AND RESPONSIBILITIES OF THE PARTIES	6
9.	ADDRESSES FOR SERVICE OF NOTICES AND PROCESSES	6 - 7
10.	AMENDMENTS	7
11.	SIGNATURES	8

1. PREAMBLE

The Limpopo Department of Education and Molteno Institute for Language and Literacy hereby agree to form a partnership for the purpose of improving the quality of teaching and learning in rural schools within Limpopo.

This Memorandum of Agreement provides an understanding of the two parties in respect of the roles and responsibilities of the two parties in promoting this collaboration agreement.

WHEREAS the Limpopo Department of Education is responsible for the provision of resources for effective teaching and learning and

WHEREAS Molteno Institute for Language and Literacy contributes towards improved performance in the language and literacy levels.

NOW THEREFORE, the parties wish to confirm their relationship with this Memorandum of Agreement as follows:

2. DEFINITIONS

In this agreement, unless clearly inconsistent with or otherwise indicated by the context:

"MoA" means this Memorandum of Agreement;

"Molteno" means Molteno Institute for Language and Literacy;

"The LDoE" means the Limpopo Provincial Department of Education; and

"The Parties" means the LDoE and Molteno Institute.

3. INTERPRETATIONS

3.1. The head notes to the various clauses of this MoA and the index are inserted for reference purposes only, and shall in no way govern or affect the construction of the MoA.

3.2. This document shall be deemed to constitute the sole memorandum of agreement between the Parties, with reference to its subject matter, and shall cancel and negate any prior verbal or written communications relating to such subject matter, whether expressed or implied, including any letters, memoranda or minutes.

3.3. This Agreement shall be deemed a contract made in South Africa, subject to and interpreted in accordance with the laws of South Africa and subject to jurisdiction of the South African courts.

3.4. Words importing persons shall include bodies corporate, and vice versa.

3.5 The singular shall include the plural, and vice versa, and reference to any gender shall include the other gender.

3.6 Any reference to a statutory provision shall include a reference to that provision as modified, amended, replaced, or re-enacted from time to time.

3.7 If any provision in a definition is a substantive provision, conferring rights or imposing obligations on any Party, effect shall be given to it as if it were a substantive clause in the body of the MoA, notwithstanding that it is only contained in the definitions clause.

3.8 Notwithstanding any other provision of this agreement to the contrary, should any of the parties be in breach of any material obligation imposed on it in terms of this agreement and fail to remedy such breach within 14 days of receipt of a written notice to from the innocent party, the innocent party shall be entitled to forthwith cancel the agreement without prejudice to any other rights it may have in terms of this agreement or in law.

3.9 Any dispute that arises from this agreement shall be submitted to and decided in accordance with the rules of the arbitration foundation of South Africa. The arbitration shall be held in Johannesburg, unless otherwise agreed by the parties. Notwithstanding anything to the contrary contained herein, this clause shall not preclude either party from obtaining interim relief on an urgent basis from a court of competent jurisdiction pending the decision of the arbitrator or from instituting in any court of competent jurisdiction any proceedings for an interdict or any other injunctive relief.

4. OBJECTIVES OF THE SUPPORT

4.1 The support provided by Molteno is intended to cultivate a sense of ownership within the targeted schools and Limpopo Department of Education.

4.2 The support is aimed at contributing towards provision of knowledge, skills and resources for promoting effective teaching and learning.

4.3 Document and communicate best practices that can be disseminated to all the schools in the Limpopo Province.

5. BACKGROUND

The Department of Basic Education has declared raising literacy levels as one of its top priorities. In 2010, the Minister of Basic Education stated that "the levels and quality of educational outcomes achieved by our learners are evidently still far below our national target". The findings based on the SAMEQ and PIRLS report showed that Limpopo province continued to be at the lowest end of literacy and other basic skills. Molteno, through its Bridges to the Future (BFI SA) program will develop lessons to cater for foundation phase difficulty (grades 1–3). The program was previously piloted in the ABET centres with adult learners, and was found to be a success. It will now be implemented with primary schools. Furthermore, the lesson development will follow the

tried-and-tested BFI-SA methodology, and will offer a variety of Limpopo languages, namely Sepedi, Tshivenda, Xitsonga, and English. The new expanded BFI-SA program will be implemented at 50 primary schools from 2013 to 2014.

The Bridges to the Future Initiative in South Africa (BFI-SA) basically aims at making a real impact on reading in young children in South Africa. BFI-SA differentiates itself through the use of interactive computer-based learning in a complementary fashion to existing literacy learning programmes, combined with offering the learner a choice of language of instruction.

6. COMMENCEMENT AND DURATION OF THIS SUPPORT

This MoA comes into operation on the date of the last party signing and subsists for the duration of the partnership unless terminated by either Party by one month notice in writing to the other withdrawing from the partnership and terminate its participation under this MoA.

Mottero will give support to the LDoE for the duration of the period for which funds will be available and mandate by Mottero Board to support literacy development is still in effect.

7. FUNDING

Mottero shall utilize the funds as provided by funders to support curriculum through language and literacy programmes.

8. ROLES AND RESPONSIBILITIES OF THE PARTIES

8.1 Nature of collaboration

- Mottero Institute is leading and coordinating a consortium of two organisations in the implementation of a technology-based early literacy intervention. the organisations are Trydan Interactive and Learn International.
- Train and mentor teachers in the use of technology for teaching early literacy.
- Assist with administration, promotion and marketing of the support.
- Report on all activities that form part of this MoA.
- Mottero in conjunction with Limpopo Department of Education will identify the participating schools.
- The LDoE will ensure schools participate and implement the project

8.2 Mottero Institute for Language and Literacy shall:

- Develop the learning contents and provide learning materials including installation of the software.
- Provide teacher training and mentoring.
- Conduct monitoring and evaluation

8.3 Limpopo Department of Education shall:

- Ensure that environments are conducive for effective teaching and learning.
- Ensure that there is proper maintenance of infrastructure and resources.
- Ensure that there is security of resources.
- Ensure that sufficient educators; administrative, school, circuit and district management resources are in place to for optimal performance and sustainability.
- Ensure that relevant data is available and relevant stakeholders are accommodated.
- Provide support in the establishment of Steering Committees for implementation of identified specific projects.
- Verify output indicators with regards to linguistic accuracy and alignment of curriculum

9. **ADDRESSES FOR SERVICE OF NOTICES AND PROCESSES**

For the Department:

Office of the Head of Department

Corner 113 Biccard & 24 Excelsior Streets

POLOKWANE

0699

Marked for Attention of: Donor Funding

Fax: 086 560 6095 or 015 297 5521

E-mail: lladib@edu.limpopo.gov.za

And

For Molteno Institute for Language and Literacy

Office of the Chief Executive Officer

49 Jorissen Street

10th Floor Orion House

Braamfontein

2017

Fax: 011 339 3555

Email: mdikolla@molteno.co.za

9.1 Either party may from time to time decide to vary its address for the service of notices and processes by giving written notice to the other party and such notice must be posted by prepaid registered post or hand delivered to the other party's addresses for service of notices and processes and shall be presumed, until the contrary is proven, to have been received by that party on the 10th day after the day of posting or on the date of delivery, as the case may be.

9.2 Matters or other communication to be given to either of the parties in terms of this agreement is valid and effective only if it is given in writing.

9.3 Any notice in terms of this agreement to either party which is—

9.3.1 sent by prepaid registered post in correctly addressed envelope to the address specified for it is deemed to have been received, unless the contrary is proved, within 14 days from the date on which it was posted;

9.3.2 delivered to the party by hand at the address specified for it in clause 7 is deemed to have been received on the day of delivery, provided that it has been delivered to a responsible person during ordinary business hours; or

9.3.3 sent by facsimile or e-mail to a party at the telefax numbers and e-mail address specified is deemed to have been received, unless the contrary is proved, within four hours of transmission if it is transmitted during normal business hours of the receiving party or within 12 hours of the first business day of the receiving party after it is transmitted, if it is transmitted outside business hours.

10. AMENDMENTS

Amendments to this MoA must be in writing and shall only take effect when signed by both parties. Project specific requirements shall be annexed to this MoA.

THE MEMORANDUM OF AGREEMENT BETWEEN THE LIMPOPO DEPARTMENT OF EDUCATION AND MOLTENO INSTITUTE FOR LANGUAGE AND LITERACY.

SIGNED AT POLOKWANE ON THIS 19TH DAY

OF APRIL 2013

LIMPOPO DEPARTMENT OF EDUCATION

AS WITNESSES

1. _____

2. _____

THE MEMORANDUM OF AGREEMENT BETWEEN THE LIMPOPO DEPARTMENT OF EDUCATION AND MOLTENO INSTITUTE FOR LANGUAGE AND LITERACY.

SIGNED AT Johannesburg ON THIS 25TH DAY

OF APRIL 2013

MOLTENO INSTITUTE FOR LANGUAGE AND LITERACY.

AS WITNESSES

1. _____

2. _____

JS

RBR

REFERENCES

Aadzi, H. (2008). Efficient learning for the poor. *International Review of Education, 54,* 581-604.

Adams, M. (1990). *Beginning to Read: Thinking and Learning about Print.* Cambridge: MIT Press.

Ahlsén, E. (2006). *Introduction to Neurolinguistics.* Philadelphia: John Benjamins Publishing Company.

Alerby, E. (2000). A way of visualising children's and young people's thoughts about the environment: A study of drawings. *Environmental Education Research, 6(3),* 205-222.

Anderson, D., Lorch, E., Field, D., & Sanders, J. (1981). The effects of TV program comprehensibility on preschool children's visual attention to television. *Child Development,* 151-175.

Anderson, L., Evertson, C., & Brophy, J. (1979). An experimental study of effective teaching in first-grade reading groups. *Elementary School Journal,* 193-223.

Anderson, R., & Freebody, P. (1981). Vocabulary knowledge. In *Comprehension and Teaching: Research Reviews* (pp. 77-117). Newark, DE: International Reading Association.

Anderson, R., & Nagy, W. (1995). *Metalinguistic Awareness and Literacy Acquisition in Different Languages.* Chicago: Center for the Study of Reading. Retrieved 3 15, 2015, from https://www.ideals.illinois.edu

Anderson, R., Wilkinson, I., Mason, J., Shirey, L., & Wilson, P. (1988). Do errors on classroom reading tasks slow growth in reading? *The Elementary School Journal, 88(3),* 266-280.

Armstrong, P. (2009). *Teacher Pay in South Africa: How Attractive is the Teaching Profession?* Stellenbosch: University of Stellenbosch.

Averell, L., & Heathcote, A. (2009). *Long term implicit and explicit memory for briefly studied words.* Proceedings of the 31st Annual Conference of the Cognitive Science

Society. Austin TX: Cognitive Science Society. Retrieved September 2016, 15, from http://www.newcl.org/publications/averell-cpc09.pdf

Baddeley, A. (2003). Working memory and language: An overview. *Journal of Communication Disorders, 36(3),* 189-208.

Bar-Kochva, I., & Breznitz, Z. (2012). Does the reading of different orthographies produce distinct brain activity patterns? An ERP study. *PLoS ONE, 7(5).* doi:10.1371/journal. pone.0036030

Baumann, J. (2005). Vocabulary-comprehension relationships. In B. Maloch, J. Hoffman, D. Schallert, C. Fairbanks, & J. Worthy (Eds.), *Fifty-fourth Yearbook of the National Reading Conference* (pp. 117-131). Oak Creed, WI: National Reading Conference.

Bembry, K., Jordan, H., Gomez, E., Anderson, M., & Mendro, R. (1998). *Policy implications of long-term teacher effects on student achievement.* Dallas: Dallas Public Schools. Retrieved 5 23, 2016, from http://www.dallasisd.org/cms/lib/TX01001475/ Centricity/Shared/evalacct/research/articles/Bembry-Policy-Implications-of-Long-Term-Teacher-Effects-on-Student-Achievement-1998.pdf

Bikitsha, N., & Katz, J. (2013). *An analysis of the Xhosa Home Language CAPS for the Foundation Phase, and the implications thereof for the teaching, learning and assessment of phonics and meaningful reading in Xhosa.* Johannesburg: Presentation at the National Conference of The Reading Association of South Africa.

Bjork, R., & Bjork, E. (1992). A new theory of disuse and an old theory of stimulus fluctuation. In A. Healy, S. Kosslyn, & R. Shiffrin (Eds.), *From Learning Processes to Cognitive Processes: Essays in Honor of William K. Estes* (pp. 35-67). Hillsdale: Erlbaum.

Black, P., & William, D. (2001). *Inside the Black Box: Raising Standards through Classroom Assessment.* London: BERA. Retrieved August 10, 2016, from https://weaeducation. typepad.co.uk/files/blackbox-1.pdf

Blaxall, J., & Willows, D. (1984). Reading ability and text difficulty as influences on second graders' oral reading errors. *Journal of Educational Psychology, 76,* 330-341.

Bomengen, M. (2010, September 23). *Reading Horizons.* Retrieved September 21, 2015, from http://www.readinghorizons.com/blog/post/2010/09/23/what-is-the-whole-language-approach-to-teaching-reading.aspx

Boud, D., Cohen, R., & Sampson, J. (1999). Peer learning and assessment. *Assessment and Evaluation in Higher Education, 24(4),* 413-426.

Brice-Heath, S. (1982). What no bedtime story means: Narrative skills at home. *Language in Society, 11(1),* 49-76.

Buck, J., & Torgesen, J. (2003). *The Relationship between Performance on a Measure of Oral Reading Fluency and Performance on the Florida Comprehensive Assessment Test.* Tallahassee: Florida Center for Reading Research.

Carpenter, S., & DeLosh, E. (2005). Application of testing and spacing effects to name learning. *Applied Cognitive Psychology, 19,* 619-636.

Carpenter, S., Pashler, H., Wixted, J., & Vul, E. (2007). *The Effects of Testing on Learning and Forgetting.* San Diego: UCSD. Retrieved September 15, 2016, from http://laplab.ucsd.edu/articles/carpenter_etal_inpress_memcog.pdf

Cartwright, N. (2007). *Are RCTs the Gold Standard?* London: Contingency And Dissent in Science Project, Centre for Philosophy of Natural and Social Science.

Chafe, W. (1988). Linking intonation units in spoken English. In J. Haiman, & S. Thompson (Eds.), *Clause Combining in Grammar and Discourse* (pp. 1-27). Amsterdam: Benjamins.

Chall, J. (1967). *Learning to Read: The Great Debate.* New York: McGraw-Hill.

Chisholm, L., Hoadley, U., Kivilu, M., Brookes, H., Prinsloo, C., Kgobe, A., Rule, S. (2005). *Educator Workload in South Africa.* Cape Town: HSRC Press.

Clay, M. (1991). *Becoming literate: The construction of inner control.* Auckland, NZ: Heinemann.

Clay, M. (2004). Talking, reading and writing. *Journal of Reading Recovery, 3(2),* 1-15.

Clay, M. (2005). *An Observation Survey of Early Literacy Achievement.* Auckland, NZ: Heinemann.

Clymer, T. (1968). What is "reading"? Some current concepts. In H. Richie & H. Robinson (eds.), *Innovation and Change in Reading Instruction* (pp 7-29). Chicago: University of Chicago Press.

Coltheart, M., Rastle, K., Perry, C., & Langdon, R. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review, 108(1)*, 204-256.

Combrinck, C., van Staden, S., & Roux, K. (2014). Developing early readers: Patterns in introducing critical reading skills and strategies to South African children. *Reading and Writing, 5(1), 1-9.* Retrieved August 2015, from http://dx.doi.org/10.4102/rw.v5i1.45

Creswell, J. (1994). *Research Design: Qualitative and Quantitative Approaches.* Thousand Oaks, CA: Sage.

Creswell, J. (2003). *Research Design: Qualitative, Quantitative, and Mixed Approaches.* Thousand Oaks, CA: Sage.

Creswell, J. (2013). *Steps in Conducting a Scholarly Mixed Methods Study.* Lincoln: Discipline-Based Education Research Group.

Creswell, J., & Plano Clark, V. (2011). *Designing and Conducting Mixed Methods Research* (2nd ed.). Thousand Oaks, CA: Sage.

Cristia, J., Cueto, S., Ibarran, P., Santiago, A., & Severin, E. (2012, February). *Technology and child development: Evidence from the one laptop per child program.* Inter-American Development Bank. Retrieved September 13, 2015, from https://publications.iadb.org

Cronbach, L. (1957). The two disciplines of scientific psychology. *American Psychologist, 12,* 671-684.

Cutler, A., & Swinney, D. (1987). Prosody and the development of comprehension. *Journal of Child Language, 14(1),* 145-167.

de Klerk, C. (2000). Language shift in Grahamstown: A case study of selected Xhosa speakers. *International Journal of Social Languages, 146,* 87-110.

Diemer, M., van der Merwe, K., & de Vos, M. (2016). The development of phonological awareness literacy measures for isiXhosa. *Southern African Linguistics and Applied Language Studies, 33(3).* 325-341

Department of Basic Education. (2007). *Education White Paper 7.* Pretoria: DBE.

Department of Basic Education. (2011a). *Curriculum Assessment Policy Statements.* Pretoria:DBE. Retrieved from www.education.gov.za

Department of Basic Education. (2011b). *Integrated Strategic Planning Framework 2011 - 2025.* Pretoria: DBE.

Department of Basic Education. (2011c). *Report on the 2008 and 2009 Annual Surveys for Ordinary Schools.* Pretoria: DBE.

Department of Basic Education. (2016). Retrieved September 10, 2016, from www.education.gov.za/

Doczi, M. (2000). *ICTs and Social and Economic Inclusion.* Retrieved from http://www.med.govt.nz/pbt/infotech/ictinclusion/ictinclusion.pdf

Duncan, P. (2013). *Drawing as a Method for Accessing Young Children's Perspectives in Research.* Sterling: University of Sterling.

Educational Access Television. (2009). Retrieved September 13, 2015, from http://educationalaccesstv.com/

Ehri, L. (2005). Sight word reading. *Scientific Studies of Reading, 9(2),* 167-188.

Ehri, L., Nunes, S., Stahl, S., & Willows, D. (2001). *Systematic phonics instruction helps students learn to read: Evidence from the National Reading Panel's meta-analysis.* Toronto: Ontario Institute for Studies in Education.

Eldén, S. (2012). Inviting the messy: Drawing methods and 'children's voices'. *Childhood, 20(1),* 68-81.

eNCA. (2015, October 11). *Bridging the digital connectivity divide in SA schools.* Retrieved December 19, 2015, from eNCA: https://www.enca.com/technology/bridging-digital-connectivity-divide-sa-schools

Fountas, I., & Pinnell, G. (1996). *Guided Reading: Good First Reading for All Children.* Portsmouth: Heinemann.

Guba, E. (1990). The alternative paradigm dialog. In E. Guba (Ed.), *The Paradigm Dialog* (pp. 17-27). Newbury Park, CA: Sage.

Guthrie, M. (1970). *Collected Papers on Bantu Linguistics.* Westmead: Gregg International Publishers Ltd.

Hanushek, E., Kain, J., & Rivkin, S. (2004). Why public schools lose teachers. *Journal of Human Resources, 39,* 326-354.

Harrison, L., Clarke, L., & Ungerer, J. (2007). Children's drawings provide a new perspective on teacher-child relationship quality and school adjustment. *Early Childhood Research Quarterly, 22(1),* 55-71.

Hart, B., & Risley, T. (2003). The early catastrophe: The 30 million word gap by age 3. *American Educator.* Retrieved 11 1, 2011, from www.aft.org//sites/default/files/periodicals/TheEarlyCatastrophe.pdf

Hasbrouck, J., & Tindal, G. (2006). *Oral reading fluency: 90 years of measurement.* Eugene, OR: Behavioral Research and Teaching (University of Oregon).

Haynie, W. (2007). Effects of test taking on retention learning in technology education: A meta-analysis. *Journal of Technology Education, 18(2)*, 24-36.

Hefer, E. (2013). Television subtitles and literacy: Where do we go from here? *Journal of Multilingual and Multicultural Development, 34(7).*

Heilman, A. (2006). *Phonics in Proper Perspective* (10th ed.). Boston: Pearson Education.

Higgins, S., Xiao, Z., & Katsipataki, M. (2012). *The Impact of Digital Technology on Learning: A Summary for the Education Endowment Foundation*. Education Endowment Foundation.

Hogan, T. (2010). A short report: Word-level phonological and lexical characteristics interact to influence phoneme awareness. *Journal of Learning Disabilities, 43(4),* 346-356. doi:10.1177/0022219410369083

Howie, S., Van Staden, S., Tshele, M., Dowse, C., & Zimmerman, L. (2012). *PIRLS 2011: South African children's Reading Literacy Achievement Summary Report.* Pretoria: Centre for Evaluation and Assessment.

Howie, S., Venter, E., can Staden, S., Zimmerman, L., Long, C., du Toit, C., . . . Archer, E. (2008). *PIRLS 2006 Summary Report: South African Children's Reading Achievement.* Pretoria, SA: Centre for Evaluation and Assessment.

Isaacs, S. (2007). *Survery of ICT and Education: South Africa Country Report.* World Bank.

Jackson, N., & Doellinger, H. (2002). Resilient readers? University students who are poor recoders but sometimes good text comprehenders. *Journal of Educational Psychology, 94,* 64-78.

JET. (forthcoming). *Initial teacher education: A review of Bachelor of Education programme outcomes in higher education institution teacher training programmes in literacy and mathematics.* Johannesburg: JET.

Johnson, R., & Onwuegbuzie, A. (2004). Mixed methods research: A research paradigm whose time has come. *Education Researcher, 33(7),* 14-26.

Jonassen, D., Campbell, J., & Davidson, M. (1994). Learning with media: Restructuring the debate. *Educational Technology: Research and Development, 42(2),* 31-39.

Karlin, A. (1985). Intonation in oral reading and reading comprehension. *Reading Horizons, 25(3),* 169-175.

Kessler, S. (2012). *2.5 Million Laptops Later, One Laptop Per Child Doesn't Improve Test Scores.* Mashable.

Kim, J. (2008). Research and the reading wars. In F. Hess (Ed.), *When Research Matters: How Scholarship Influences Education Policy* (pp. 89-111). Cambridge, MA: Harvard Education Press.

Komulainen, S. (2007). The ambiguity of the child's 'voice' in social research. *Childhood, 14(1),* 11-28.

Konza, D. (2011). *Research into Practice: Literacy is Everyone's Business.* Government of South Australia: Department of Education and Children's Services. Retrieved from http://www.decd.sa.gov.au/literacy/files/links/UtRP_1_2_v2.pdf

Krcmar, M., Grela, B., & Lin, K. (2007). Can toddlers learn vocabulary from television? An experimental approach. *Media Psychology, 10,* 41-63.

Kuhn, M., & Stahl, S. (2003). Fluency: A review of developmental and remedial practices. *Journal of Educational Psychology, 95(1),* 3-21.

Kulik, J. (1994). Meta-analytic studies of findings on computer-based instruction. In E. Baker & H. J. O'Neil, *Technology Assessment in Education and Training.* Hillsdale, NJ: Lawrence Erlbaum.

LaBerge, D., & Samuels, S. (1974). Towards a theory of automatic information processing in reading. *Cognitive Psychology, 6,* 293-323.

Ladbrook, M. (2009). *Challenges Experienced by Educators in the Implementation of Inclusive Education in Primary Schools in South Africa.* Pretoria: UNISA.

Land, S. (2015, January). *Reading isiZulu: Reading processes in an agglutinative language with a transparent orthography.* Retrieved August 2, 2016, from https://www.researchgate.net/publication/280597179

Land, S. (2016). Automaticity in reading isiZulu. *Reading & Writing, 7(1).*

Lemann, N. (1997, November). The Reading Wars. *Atlantic Monthly,* pp. 128-134.

Literat, I. (2013). A pencil for your thoughts: Participatory drawing as a visual research method with young children. *International Journal of Qualitative Methods,* 84-98.

Lopes, J., Silva, M., Moniz, A., & Zibulsky, J. (2015). Prosody growth and reading comprehension: A longitudinal study from 2nd through the end of 3rd grade. *Revista de Psicodidáctica, 20(1),* 5-23.

Louwrens, L., & Poulos, G. (2006). The status of the word in selected conventional writing systems - the case of disjunctive writing. *Southern African Linguistics and Applied Language Studies, 24(3),* 389-401.

McCaffrey, D., Lockwood, J., Koretz, D., & Hamilton, L. (2003). *Evaluating value added models for teacher accountability*. Santa Monica: RAND Corporation (Carnagie). Retrieved May 23, 2016, from www.rand.org/pubs/monographs/2004/RAND_MG158.pdf

Modipa, T., Davel, M., & De Wet, F. (2010). Acoustic modelling of Sepedi affricates for ASR. *Natural Language Processing - Speech Recognition and Synthesis, J.5,* 394-398. Retrieved from http://www.meraka.org.za/pubs/Modipa10Acoustic.pdf

Mokhtari, K., & Reichard, C. (2002). Assessing learners' metacognitive awareness of reading strategies. *Journal of Educational Psychology, 94(2),* 249-259.

Mullis, I., Martin, M., Foy, P., & Drucker, K. (2012). *PIRLS 2011 International Results in Reading.* Chestnut Hill: Boston College.

Nagel, T. (1986). *The View from Nowhere.* New York: Oxford University Press.

Nagy, W. (1988). *Teaching Vocabulary to Improve Reading Comprehension.* Chicago: ERIC/EECE.

National Telecommunications and Information Administration (NTAI). (1999). *Falling through the net: Defining the digital divide.* Department of Commerce. NTAI. Retrieved from http://www.ntia.doc.gov/ntiahome/fttn99

NEEDU. (2012). *NEEDU National Report 2012: The State of Literacy Teaching and Learning in the Foundation Phase.* Pretoria, SA: Department of Basic Education National Education Evaluation and Development Unit.

Nwaerondu, N. G., & Thompson, G. (1987). The use of educational radio in developing countries: Lessons from the past. *Journal of Distance Education,* 43-54.

OECD. (2001). *Understanding the Digital Divide*. Paris: OECD Publications.

OECD. (2008). *Reviews of National Policies for Education: South Africa.* Paris: OECD. Retrieved from http://www.oecd.org/southafrica/reviewsofnationalpoliciesforeducation-southafrica.htm

OECD. (2015). *Students, computers and learning: Making the connection*. Paris: PISA, OECD Publishing.

Olsen, W. (2004). Triangulation in social science research: Qualitative and quantitative methods really can be mixed. In M. Holborn, & M. Holborn (Ed.), *Developments in Sociology*. Ormskirk: Causeway Press. Retrieved August 24, 2011, from http://www.federica.eu/users/9/docs/amaturo-39571-01-Triangulation.pdf

Patton, M. (1990). *Qualitative Evaluation and Research Methods* (2nd ed.). Newbury Park: Sage.

Pence, K., & Justice, L. (2008). *Language Development from Theory to Practice.* Upper Saddle River, NJ: Pearson Education.

Perfetti, C. (2003). The universal grammar of reading. *Scientific Studies of Reading, 7,* 3-24.

Perfetti, C., Beck, I., Bell, L., & Hughes, C. (1987). Phonemic knowledge and learning to read are reciprocal: A longitudinal study of first grade children. *Merrill-Palmer Quarterly, 33,* 283-319.

Pitsoe, V. (2013). Teacher attrition in South Africa: Trends, challenges and prospects. *Journal of Social Sciences, 36(3),* 309-318.

Pretorius, E. (2002). Reading and applied linguistics - a deafening silence? *Southern African Linguistics and Applied Language Studies, 20(1-2),* 91-103.

Pretorius, E., & Currin, S. (2010). Do the rich get richer and the poor poorer? The effects of an intervention programme on reading in the home and school language in a high poverty multilingual context. *International Journal of Educational Development,* 67-76.

Pretorius, E., & Mokhwesana, M. (2009). Putting reading in Northern Sotho on track in the early years: Changing resources, expectations and practices in a high poverty school. *Southern African Journal of African Languages, 1,* 54-73.

Prinsloo, D. (2009). Current lexocography practice in Bantu with specific reference to the Oxford Northern Sotho School Dictionary. *International Journal of Lexicography, 22(2),* 151-178.

Probert, T., & de Vos, M. (2016). Word recognition strategies amongst isiXhosa/English bilingual learners: The interaction of orthography and language of learning and teaching. *Reading and Writing, 7(1).* Retrieved from http://dx.doi.org/10.4102/rw.v7i1.84

Rayner, K. (1975). Eye movements in reading and information processing. *Psychological Bulliten, 85,* 618-665.

Rayner, K., & Sereno, S. (1994). Eye movements in reading: Psycholinguistic studies. In M. Gernsbacher (Ed.), *Handbook of Psycholinguistics* (pp. 57-81). San Diego: Academic Press.

Roediger, H., & Karpicke, J. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science, 17(3),* 249-255. doi: 10.1111/j.1467-9280. 2006.01693.x

Rose, G. (2007). *Visual Methodologies: An Introduction to the Interpretation of Visual Materials.* London: Sage.

Ross, S., & Morrison, G. (2004). Experimental research methods. In D. Jonassen (Ed.), *Handbook of Research for Educational Communications and Technology* (pp. 1021-1043). Mahwah, NJ: Lawrence Erlbaum Associated Press.

RTI International. (2009). *Early Grade Reading Assessment Toolkit.* The World Bank.

Rudwick, S. (2008). 'Coconuts' and 'oreos': English-speaking Zulu people in a South African township. *World Englishes, 27(1),* 101-116.

Runquist, W. (1983). Some effects of remembering on forgetting. *Memory and Cognition, 11,* 641-650.

Schacter, J. (1999). *The Impact of Education Technology on Student Achievement: What the Most Current Research Has to Say.* Milken Exchange on Education Technology.

Schantz, P., & Zimmer, J. (2005, March). Why Johnny Can't Read: 50 years of Controversy. *History of Reading News,* pp. 1-4.

Schilling, S., Carlisle, J., Scott, S., & Zeng, J. (2007). Are fluency measures accurate predictors of reading achievement? *Elementary School Journal,* 429-448.

Schramm, W., Lyle, J., & Parker, E. (1961). *Television in the Lives of Our Children.* Standford: Standford University Press.

Schreiber, P. (1991). Understanding prosody's role in reading acquisition. *Theory into Practice, 30(3),* 158-164.

Shiohira, K. (2014a). *A Situational Analysis of non-Fresh Start Schools in Libode and Mount Frere Districts.* Centurion: National Education Collaboration Trust.

Shiohira, K. (2014b). *The broken triangle: Academic research, policy and development practitioners.* Toronto: Comparative and International Education Society Conference Proceedings.

Shiohira, K. (2015). Population and Education. In M. Makiwane, & G. Chimere-Dan (Eds.), *Eastern Cape: The State of the Population 2014* (pp. 132-147). Bisho: Department of Social Development.

Shiohira, K., & Shezi, D. (2013). *Northwest Needs Assessment: An Investigation into the Literacy Practices and Challenges of Foundation Phase Teachers.* Johannesburg: Molteno Institute for Language and Literacy.

Smith, F. (1986). *Insult to Intelligence: The Beaurocratic Invasion of Our Classrooms.* Toronto: Pearson Education.

Smith, J. (1983). Quantitative versus qualitative research: An attempt to clarify the issue. *Education Researcher, 12,* 6-13.

Snow, C. (2008). What counts as literacy in early childhood? In *Blackwell Handbook of Early Childhood Development.* Oxford, UK: Blackwell Publishing.

South African Board of Education. (2006). *2007-2011 Strategic Plan.* Retrieved September 2016, from Department of Basic Education: http://www.gov.za/documents/department-education-strategic-plan-2007-2011

Statistics South Africa. (2015). Retrieved September 21, 2015, from http://www.statssa.gov.za

Sullivan, G. (2011). Getting off the "gold standard": Randomized control trials and education research. *Journal of Graduate Medical Education, 3(3)*, 285-289.

Tashakkori, A., & Teddlie, C. (1998). *Mixed Methodology: Combining Qualitative and Quantitative Approaches.* Thousand Oaks, CA: Sage.

Taylor, N., Fleisch, B., & Shindler, J. (2008). *Changes in Education since 1994*. Pretoria: National Office of the Presidency. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.606.4651&rep=rep1&type=pdf

Taylor, N., Muller, J., & Vinjevold, P. (2003). *Getting Schools Working: Research and Systemic School Reform in South Africa.* Cape Town: Pearson Education.

Thomas, D. (2003). A general inductive approach for qualitative data analysis. Retrieved: http://www.frankumstein.com/PDF/Psychology/Inductive%20Content%20Analysis.pdf

Tolani-Brown, N., McCormac, M., & Zimmermann, R. (2009). An analysis of the research and impact of ICT in education in developing country contexts. *Journal of Education for International Development, 4(2),* 1-12.

Toppino, T., & Cohen, M. (2009). The Testing Effect and the Retention Interval: Questions and Answers. *Experimental Psychology, 56(4),* 252-257.

UNESCO. (2008). *Mother Tongue Matters: Local language as a key to effective learning*. Paris: UNESCO.

UNESCO. (2014). *Mobiles for Reading: A landscape research review*. JBS International, Inc.

Van der Berg, S. (2003). *Determinants of Education Quality: A summary of findings from the Western Cape Primary School Pupil Progress survey of 2003*. Johannesburg: JET.

van Rooy, B., & Pretorius, E. (2013). Is reading in an agglutinating language different from an analytic language? An analysis of isiZulu and English reading based on eye movements. *Southern African Linguistics and Applied Language Studies , 31(3),* 281-297.

Vogel, J., Vogel, D., Cannon-Bowers, J., Bowers, C., Muse, K., & Wright, M. (2006). Computer gaming and interactive simulations for learning: A meta-analysis. *Journal of Education Computing Research, 34(3),* 229-243.

Wagner, R., & Torgesen, J. (1987). The nature of phonological processing and its role in the acquisition of reading skills. *Psychological Bulletin, 101(2),* 192-212.

Wainer, J., Dwyer, T., Dutra, R., Covic, A., Magalhaes, V., Ferreira, L., . . . Claudio, K. (2008). Too much computer and internet use is bad for your grades, especially if you are

young and poor: Results from the 2001 Brazilian SAEB. *Computers and Education, 51,* 1417-1429.

WestEd. (2002). *The Learning Return on Our Education Investment: A review of findings from research.* San Francisco: WestEd.

Wheeler, M., Ewers, M., & Buonanno, J. (2003). Different rates of forgetting following study versus test trials. *Memory, 11,* 571-580.

Williams, E. (1996). Reading in two languages in year five in African primary schools. *Applied Linguistics, 17(2),* 182-209.

Wilsenach, C. (2015). Receptive vocabulary and early literacy skills in emergent bilingual Northern Sotho-English children. *Reading and Writing, 6(1).* Retrieved from http://dx.doi.org/10.4102/rw.v6i1.77

Wolf, L., & MacKinnon, S. (2002). *TechKnowLogia.* Retrieved 9 21, 2015, from TechKnowLogia: www.techknowlogia.org

Wolf, M. (2008). *Proust and the Squid: The Story and Science of the Reading Brain.* New York: Harper Perennial.

Zeigler, J., & Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: A psycholinguistic grain size theory. *Psychological bulletin, 131(1),* 3-29.