TECHNICAL REPORT:

Benchmarking early grade reading skills in Nguni languages

Cally Ardington, Gabrielle Wills, Elizabeth Pretorius, Nicola Deghaye, Nompumelelo Mohohlwane, Alicia Menendez, Nangamso Mtsatse, and Servaas van der Berg

OCTOBER 2020

basic education
Department:
Basic Education
REPUBLIC OF SOUTH AFRICA

RESEP
Research on Socio-Economic Policy

NORC at the University of Chicago

UNIVERSITEIT iYUNIVESITHI STELLENBOSCH UNIVERSITY

100 1918 · 2018

SALDRU
Southern Africa Labour and Development Research Unit

University of Cape Town · iYunivesithi yaseKapa · Universiteit van Stellenbosch

TECHNICAL REPORT:

Benchmarking early grade reading skills in Nguni languages

October 2020

Cally Ardington[1], Gabrielle Wills[2], Elizabeth Pretorius[3], Nicola Deghaye[4], Nompumelelo Mohohlwane[5],

Alicia Menendez[6], Nangamso Mtsatse[7], and Servaas van der Berg[8]

ISBN: 978-1-4315-3413-5

1 SALDRU, University of Cape Town and NORC at the University of Chicago.
2 Research on Socioeconomic Policy (ReSEP), Department of Economics, Stellenbosch University.
3 Research Fellow & Emeritus Professor: Department of Linguistics and Modern Languages, Unisa
4 ReSEP, Department of Economics, Stellenbosch University.
5 Deputy Director of Research Coordination, Monitoring and Evaluation, Department of Basic Education, South Africa.
6 Harris School of Public Policy at the University of Chicago & NORC at the University of Chicago
7 Funda Wande and PhD candidate at Stellenbosch University, ReSEP.
8 ReSEP, Department of Economics, Stellenbosch University and National Research Chair in the Economics of Social Policy.

Abstract

While reading for meaning is the goal of reading, many foundational skills need to be mastered before children can read and understand a text on their own. This report establishes thresholds and benchmarks for foundational early grade reading skills that are necessary (albeit not sufficient) to read for meaning in Nguni languages in the South African context. Our approach to establishing benchmarks integrates theoretical understandings of reading development with an exploratory analysis of early grade reading assessment (EGRA) data in three Nguni languages: isiZulu, isiXhosa and siSwati. The data used is the largest available source of information on early grade reading in these languages, with multiple and comparative assessment data points for nearly 16,400 unique learners in the early grades. By rigorously analysing empirical regularities and reading trajectories across these data, we identify the following context-sensitive benchmarks and threshold: a letter-sounds benchmark at the end of grade 1 of 40 letter-sounds per minute; a lower fluency threshold at the end of grade 2 of 20 correct words per minute and a fluency benchmark at the end of grade 3 of 35 correct words per minute.

About this report

This report presents the key findings and approach used in identifying early grade reading benchmarks and thresholds in three Nguni languages: isiZulu, isiXhosa and siSwati. A summary report version of this detailed technical report is available at www.education.gov.za or www.resep.sun.ac.za.

How to cite this technical report:

Ardington, C., Wills, G., Pretorius, E., Deghaye, N., Menendez, A., Mohohlwane, N., Mtsatse, N. & Van der Berg, S. (2020) *Technical Report: Benchmarking early grade reading skills in Nguni languages.* Stellenbosch: ReSEP, Stellenbosch University. Cape Town: SALDRU, University of Cape Town. Chicago: NORC at the University of Chicago. Pretoria: Department of Basic Education

How to cite the summary report:

Ardington, C., Wills, G., Pretorius, E., Deghaye, N., Mohohlwane, N., Menendez, A., Mtsatse, N. & Van der Berg, S. (2020) *Summary Report: Benchmarking early grade reading skills in Nguni languages.* Stellenbosch: ReSEP, Stellenbosch University. Cape Town: SALDRU, University of Cape Town. Chicago: NORC at the University of Chicago. Pretoria: Department of Basic Education

Message of endorsement from the Department of Basic Education

This report is welcomed as part of the sector's contributions to reading benchmarks in the early grades.

It is seen as a complement to the Reading Benchmarks Design Report also published in 2020, addressing one of the major gaps identified, the absence of reading benchmarks in South African languages. The focus on three of the four Nguni languages - isiZulu, Siswati and isiXhosa - represents a significant population in the country and the basic education sector. The work done by all the researchers, universities and funders involved in the collection of the primary data as well as the analysis for this report are applauded and appreciated. It is my hope that similar work will be undertaken by researchers across the rest of the country for all the remaining languages.

MR HM MWELI

DIRECTOR-GENERAL

DATE: 05/10/2020

Acknowledgments

This report was produced as part of two projects: The Early Learning Programme at ReSEP, Stellenbosch University, funded by the Allan Gray Orbis Foundation Endowment and the project "Reading and socio-emotional skills in challenging school contexts: Evidence from South African primary schools", supported by the Economic and Social Research Council (ESRC) [grant ES/T007583/1].

Additionally, contributions by Cally Ardington and Alicia Menendez to this work have been undertaken as part of the External Impact Evaluation of Story Powered Schools (SPS) that NORC at the University of Chicago is conducting under the USAID Reading and Access Evaluation Contract (PN 7617.010.01 GC-10F-0033M/AID-OAA-M-13-00010). The views expressed in this work do not necessarily reflect the views of the United States Agency for International Development or the United States Government.

In addition to using data collected from the External Impact Evaluation of Story Powered Schools (SPS), this report combines data from multiple early grade reading assessments (EGRA) in South Africa. Without these data, this benchmarking project would not be possible. Accordingly, we acknowledge the funders and implementers of these EGRA related reading projects or programmes. The Early Grade Reading Study II (EGRS II) was conducted by the Department of Basic Education and funded through the United States Agency for International Development (USAID). The Funda Wande evaluation was conducted by SALDRU and funded through the Allan Gray Orbis Foundation Endowment. The Economic and Social Research Council, in a joint initiative with the Department for International Development in the United Kingdom, funded the Leadership for Literacy project [grant ES/N01023X/1] which was conducted by ReSEP. The Zenex Foundation Literacy Project (Zenlit) was made possible by the Zenex Foundation.

Contents

Tables iii
Figures iii
Abbreviations v
Executive summary vi
1. Introduction 1
2. Rationale for developing early reading benchmarks 2
    2.1. Why are reading benchmarks useful? 2
    2.2. The need for language specific benchmarks 3
3. Theoretical framework: Reading as a complex phenomenon 6
    3.1. Implications of reading comprehension theory for establishing reading benchmarks 6
    3.2. Relationships between code-based factors and reading comprehension 7
        3.2.1. The nature of the code set and the development of early decoding skills 8
        3.2.2. The relationship between decoding and reading comprehension 9
    3.3. Pulling the threads together: A developmental view of multiple proficiencies 11
4. Methodology 13
    4.1. Data 13
        4.1.1. Reading measures across datasets 13
        4.1.2. Sample sizes and sample characteristics for identifying reading norms and trends in code-based factors in the early grades 13
    4.2. Data sub-samples used for benchmarking analyses 16
    4.3. ORF text analysis 18
    4.4. Categorisation of comprehension questions 19
    4.5. Our approach to benchmarking 19
5. Preliminary data analysis: Reading norms and trends 22
    5.1. Letter-sound knowledge 22
    5.2. Complex consonant sequences 24
    5.3. Word reading 25
    5.4. Text reading (Oral reading fluency) 26
    5.5. Comprehension 27
    5.6. Relationships between code-based factors 29
6. Benchmarking 31
    6.1. Establishing fluency thresholds 31
        6.1.1. Relationship between speed and accuracy 31
        6.1.2. Relationship between fluency and comprehension 34
        6.1.3. Fluency thresholds and learner profiles 37
        6.1.4. Predictive validity of the fluency threshold and benchmark 39
    6.2. Establishing a letter-sound benchmark 44
        6.2.1. Empirical motivation for a letter-sound benchmark 45
        6.2.2. Relationship between speed and accuracy 46

6.2.3. Predictive validity of the proposed letter-sound benchmark 48
7. Summary and recommendations 52
7.1. Summary of proposed benchmarks and thresholds 52
7.2. Policy recommendations 53
8. References 54
9. Appendix 58

Tables

Table 1: The uses of reading benchmarks at various levels of the education system 3
Table 2: Some differences between Nguni languages, Sotho languages and English 4
Table 3: Grades at which Nguni language EGRAs were conducted, by individual study 14
Table 4: Characteristics of the learner sample across the pooled EGRA dataset by Nguni language 15
Table 5: Sample sizes by language, study and grade used for the Nguni language benchmarking analysis 17
Table 6: Oral reading comprehension score for learners attempting all questions 28
Table 7: Learner characteristics by early grade fluency profiles 39
Table 8: Longitudinal data 40
Table A 1: EGRA study testing components by Nguni language, grade and term 58
Table A 2: Learner and school sample sizes by Nguni language, grade and school term across EGRA studies 59
Table A 3: Sampling design of individual EGRA studies 60
Table A 4: Summary information about ORF texts and associated comprehension question types 65

Figures

Figure E 1: Reading benchmarks and thresholds for early grade reading in Nguni languages ix
Figure 1: Developmental cline in early reading 11
Figure 2: Developmental continuum in early reading in relation to formal grades 12
Figure 3: ORF passage comparisons – total syllables up to 10th and 30th word 18
Figure 4: Total complex consonant sequences up to the 10th and 30th word of each ORF passage 19
Figure 5: Letter-sound score by language, grade and term (pooled EGRA dataset) 23
Figure 6: Comparison of letter-sound scores within studies – 25th, 50th and 75th percentiles and percent scoring zero 23
Figure 7: Proportion of single consonants vs complex consonant sequences correctly sounded 24
Figure 8: Zero scores in single consonant letter-sounds versus complex consonant sequences 25
Figure 9: Isolated (single) word reading by language, grade and term (pooled EGRA dataset) 25
Figure 10: Oral reading fluency* score by language, grade, and term (pooled EGRA dataset) 26
Figure 11: Comparison of oral reading fluency* scores within studies – 25th, 50th and 75th percentiles and percent scoring zero 27
Figure 12: Proportion of learners answering each comprehension question correctly 29
Figure 13: Correlation coefficients across code-based factors 30
Figure 14: Reading speed and accuracy (isiXhosa) 32

Figure 15: Reading speed and accuracy (isiZulu) 32
Figure 16: Reading speed and accuracy (siSwati) 33
Figure 17: Speed distribution for learners reading with at least 95% accuracy 34
Figure 18: Speed distribution for learners reading with less than 95% accuracy 34
Figure 19: Relationship between oral reading fluency and comprehension for learners attempting all questions 35
Figure 20: Relationship between fluency and individual comprehension questions - SPS and FW 36
Figure 21: Relationship between fluency and individual comprehension questions – EGRS and SPS 37
Figure 22: Early grade fluency profiles, isiXhosa samples 38
Figure 23: Early grade fluency profiles, isiZulu and siSwati samples 38
Figure 24: Reading fluency category at assessment II for non-readers at assessment I 41
Figure 25: Reading fluency category at assessment II for learners scoring an ORF of 1 to 19 cwpm at assessment I 41
Figure 26: Reading fluency category at assessment II for learners scoring an ORF of 20 to 34 cwpm at assessment I 42
Figure 27: Relationship between oral reading fluency and written comprehension - SPS 43
Figure 28: Relationship between oral reading fluency and written comprehension -EGRS 43
Figure 29: Correct letter-sounds per minute distribution, isiXhosa samples 44
Figure 30: Correct letter-sounds per minute distribution, isiZulu samples 45
Figure 31: Correct letter-sounds per minute distribution, siSwati samples 45
Figure 32: Relationship between letter-sound knowledge and later oral reading fluency, all languages 46
Figure 33: Relationship between letter-sound speed and accuracy, isiXhosa samples 47
Figure 34: Relationship between letter-sound speed and accuracy, isiZulu samples 47
Figure 35: Relationship between letter-sound speed and accuracy, siSwati samples 48
Figure 36: Development on letter-sound knowledge over time, all languages 49
Figure 37: Reading fluency at assessment II, for learners who achieved the proposed letter-sound benchmark at assessment I 50
Figure 38: Reading fluency at assessment II, for learners who sounded 20 to 39 letters correctly at assessment I 50
Figure 39: Reading fluency at assessment II, for learners who sounded less than 20 letters correctly at assessment I 51
Figure A 1: Relationship between knowledge of complex sequences (2 consonants together) and oral reading fluency 61
Figure A 2: Comparison of isolated word reading score within studies – 25th, 50th and 75th percentiles and percent scoring zero 62

Abbreviations

cpm	Correct per minute
cwpm	Correct words read per minute
DBE	Department of Basic Education
DFID	Department for International Development
EGRA	Early Grade Reading Assessment
EGRS II	Early Grade Reading Study II
ESRC	Economic Sciences Research Council
FW	Funda Wande
LFL	Leadership for Literacy
LOLT	Language of Learning and Teaching
ORF	Oral Reading Fluency
PIRLS	Progress in International Reading Literacy Study
RAN	Rapid automatised naming
SPS	Story Powered Schools
SVR	Simple view of reading
USAID	United States Agency for International Development
Zenlit	Zenex Foundation Literacy Project

Executive summary

In early 2019 South African president Cyril Ramaphosa articulated a new and clear expectation for basic education: every child should be able to read for meaning by age 10 (South African Government, 2019). While reading for meaning is the goal of reading, reading is a complex and hierarchical process. A range of foundational reading subskills need to be mastered before one can comprehend (or understand) what is in a text. For example, knowledge is required of the 'code' of the language in which learners are reading, which we refer to as decoding skills. This report establishes thresholds and benchmarks for some of these foundational early grade decoding skills in 3 Nguni languages: isiZulu, isiXhosa and siSwati.

Why we need reading benchmarks

The most recent Progress in International Reading Literacy Study (PIRLS 2016) showed that 78% of grade 4 learners, mostly tested in their home language, could not reach the low international benchmark – an indicator of the ability to read for meaning (Howie et al., 2017). Reading comprehension assessments such as PIRLS can identify processes of reading comprehension that learners have not mastered. However, they cannot specify which foundational aspects of reading pose problems for learners who struggle to understand what they are reading. To identify foundational decoding problems, early grade assessments of reading are required. In this report, we use early grade assessment data to establish benchmarks and thresholds.

Reading benchmarks and thresholds are numerical measures of proficiency in specific reading skills, that may be used to monitor whether children are on track. They can inform a shared vision of what successful reading looks like at the end of grade 1, 2 and 3. They provide a standard against which teachers can measure learners' reading subskills and identify *early on* learners who are at risk of not learning to read for meaning by age 10. This, in turn, supports remediation at an earlier age. Additionally, as specific learners reach different benchmarks, this can help teachers adapt their instructional focus to meet the learners' needs at their reading level.

While oral reading benchmarks exist in English (Hasbrouck & Tindal, 2006), there is scant research guiding the development of Nguni language reading benchmarks or thresholds. One cannot simply transfer reading benchmarks from English to the Nguni languages due to differences with the phonological, morphological, and orthographical features of African languages. Benchmarking processes need to take account of the linguistic features of the language for which the benchmarks are being developed. At the simplest level, it makes no sense to compare fluency across languages with vastly different word lengths. Beyond that, one needs to allow for language specific accuracy-speed and fluency-comprehension relationships that reflect reading development.

Method

Our approach to establishing benchmarks builds on a theoretical understanding of reading development. The theoretical framework (explained in section 3) then sets the foundation for an exploratory analysis of available early grade reading assessment (EGRA) data in 3 Nguni languages. Drawing on learner assessment data from 5 different studies, we compile the largest available source of information on early grade reading in these languages. When combined, these data contain multiple assessment points which can be compared over time for nearly 16,400 learners in the early grades. This data sample is not nationally representative. However, it provides a very

clear indication of the foundational reading skills of learners in more than 660 predominately no-fee schools, located across 4 provinces and in relatively high poverty contexts.

A key decision in benchmarking is selecting which reading subskills should be benchmarked. We focus on two: letter-sound knowledge which refers to alphabetic knowledge of the written code; and oral reading fluency (ORF). ORF refers to the ability to read words in context with speed, accuracy and prosody.

- Accuracy reflects the percentage of words that are read correctly.
- Speed reflects the number of words that are attempted in a time period.
- Prosody reflects how natural reading sounds (how it conforms to speech rhythms and intonation patterns and reflects punctuation conventions).

Since assessment of prosody is subjective and it is difficult to measure in field studies, measures of ORF typically focus only on speed and accuracy. In this report, the term *fluency* is used to describe reading with speed and accuracy and is typically measured by the number of words correctly read per minute from a passage of text.

Our approach to benchmarking ORF and letter-sounds is based on a conceptualisation of different stages of reading development where different cognitive processes come into play as reading proficiency increases. Within each process, accuracy develops first followed by speed. Our approach aligns with the decoding threshold hypothesis put forward by Wang et al. (2019) where reading comprehension is unlikely to develop until decoding exceeds a lower bound threshold level. There may also be an upper threshold, beyond which there are no additional gains (in comprehension) for increasing decoding skills. This suggests that the relationship between fluency and comprehension will break down at low and high levels of fluency.

Accordingly, our analysis of EGRA data seeks to establish if there are regular patterns and trends in the speed-accuracy and fluency-comprehension relationships across studies, languages, grades, and reading passages. We further establish the validity of these benchmarks and thresholds by examining learners' future reading proficiency levels if they had met specified thresholds or benchmarks at earlier grade points.

Reading norms

As a first step, sample-based reading norms and trends in decoding skills across grades were established using EGRA data from a pooled sample of 16,400 learners. We highlight the following findings:

Letter-sound knowledge: Far too many learners are entering grade 1 with no letter-sound knowledge, despite most having attended grade R. We also identify that 1 in 10 learners in this sample are still unable to sound 1 letter-sound correctly by the end of grade 3. If learners fall behind in acquiring letter-sound knowledge in grade 1 or 2, they are unlikely to make any significant improvement in grade 3 or beyond.

Knowledge of complex consonants: Learners also experience significant difficulty in reading complex consonant sequences (examples include *hl, dl, kh, tsh*, *ndl, gcw, ntsw)* in grade 1 and 2. These sounds feature regularly in Nguni languages and knowledge of these more complex consonant sequences is necessary to read most grade 1 level texts in these languages. Thus, mastery in reading these is required early on. This may require an adjustment to the Nguni language curriculum, which does not include the teaching of complex consonant sequences in grade 1.

Oral reading fluency: The analysis reveals slow but steady increases over time in oral reading fluency. By the end of grade 3, on average, learners can read 19 to 25 correct words per minute (cwpm) from a passage. Sadly, between 15% and 26% of learners in these samples are unable to read 1 word correctly by the end of grade 3.

Our analysis shows that learners' performance in these various reading subskills were closely correlated, particularly in lower grades. But we note the following exception:

Oral reading comprehension: We observe low levels of comprehension even at higher levels of fluency (correct words per minute) from grade 3 upwards. In other words, some learners can read but cannot effectively comprehend what they are reading. This suggests there is not enough focus on teaching learners in the Foundation Phase the skills needed to answer oral and written comprehension questions.

Reading benchmarks and thresholds

Within each reading subskill we found remarkably consistent patterns between speed (number of words/letter-sounds attempted in a minute) and accuracy of reading (the percentage of words / letter-sounds read correctly out of those attempted). Accuracy initially increases rapidly as speed increases, but then flattens and levels off once learners achieve about 95% accuracy (i.e. learners read 95 of every 100 words/letter-sounds attempted correctly). For letter-sounds this levelling off in accuracy occurs at speeds of approximately 40 letter-sounds attempted per minute. For reading words in a passage of connected text, the levelling of accuracy occurs at 22 to 34 words per minute.

All of this analysis, together with expert opinion, allows us to identify the following thresholds and benchmarks:

By the end of grade 1 all learners should be able to read 40 letter-sounds correct per minute.

- This appears to be a good early predictor of oral reading fluency (ORF) later in the Foundation Phase.
- There are few benefits for improving letter-sound knowledge and speed beyond this point.
- Once learners have achieved this level of letter-sound knowledge, decoding instruction should focus on helping learners apply word attack strategies and develop fluency.

By the end of grade 2, all learners should be able to read at least 20 correct words per minute (we have termed this the lower ORF threshold).

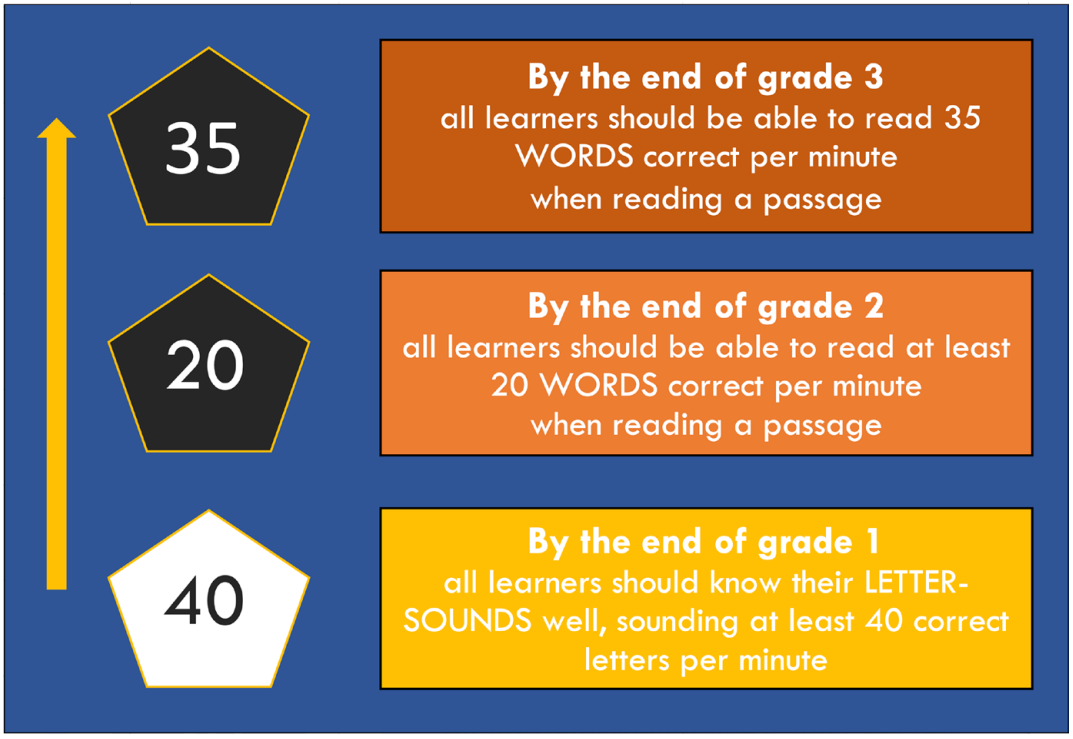
- This is a *minimum* threshold. If learners do not reach this level of fluency, higher order reading skills are very unlikely to develop. Below this threshold we find little evidence that learners can comprehend what they have read, evidenced by very low oral comprehension or written comprehension scores.
- By the end of grade 3, between 53% and 76% of the learners in this sample had reached this grade 2 threshold.

By the end of grade 3, all learners should be able to read 35 correct words per minute.

- This acts as an upper fluency threshold.
- At this level of fluency reading comprehension becomes increasingly possible when learners read on their own.
- Once learners reach this level of fluency, it appears that poor comprehension skills become the limiting factor to further literacy development.

- Therefore, at this milestone teachers should focus on teaching learners the skills and strategies needed to tackle written comprehension while encouraging vocabulary and language development.
- Fluency skills should continue to improve from this milestone.
- Approximately, only a quarter of learners in the EGRA studies analysed had reached this fluency benchmark by the end of grade 3.

Figure E 1: Reading benchmarks and thresholds for early grade reading in Nguni languages



Further research is required

This research is the first of its kind in this context and, as a result, is exploratory. We hope that as more data on early grade reading assessment (EGRA) becomes available, these benchmarks and thresholds will be further tested and corroborated. In particular, we hope that more EGRA data on isiNdebele will be forthcoming to allow the testing of these benchmarks for that language. In the meantime, we suggest that the language structure is similar enough to the three languages tested so that these benchmarks can also be adopted for isiNdebele.

Language benchmarks and thresholds are language- and context-specific. These thresholds apply to the South African context only and only to learners learning to read the Nguni languages.

1. Introduction

In any well-functioning organisation, excellence hinges on clear expectations for acceptable standards of performance. Expectations create clarity and align various moving parts of an organisation around common understandings. In South Africa a general lack of clarity at the level of the school and classroom around minimum levels of acceptable learning standards has contributed towards supressed quality and learning, including very low levels of reading proficiency in the early grades by international and regional comparisons.

In early 2019, acknowledging the reading crisis that faces the nation, South African president Cyril Ramaphosa articulated a new and clear expectation for basic education in South Africa: *every child should be able to read for meaning by aged 10*, aligning with the end of the Foundation Phase (grade 3) (South African Government, 2019). While reading for meaning is the goal of reading, a myriad of foundational skills (building blocks of reading) are necessary for reading and understanding a text on one's own. This report aims to set thresholds and benchmarks for foundational early grade reading skills that are necessary (albeit not sufficient) to read for meaning in Nguni languages.

Reading benchmarks are numerical measures of proficiency in specific reading skills, that may be used to monitor whether children are on track. They are distinct from goals and targets. Goals are often aspirational and communicate a vision of a future outcome, such as: *every child should be able to read for meaning by age 10*. Targets communicate plans to reach these goals within a specific timeframe, such as: *80% of all children should be able to read for meaning by 2030*. In contrast, a benchmark may be: *40 letters correct per minute.* A benchmark such as this would be a way to communicate the speed and accuracy required for mastery of letter-sounds (one of the subskills that contribute to reading for meaning).

Reading thresholds and benchmarks are not decided in an arbitrary manner. They should be informed by scientific literature on reading across language groups. They should be based on strong empirical work and should be sensitive to current realities of learning and curriculum requirements. The use of data to establish benchmarks in this study is guided by language and reading theory, as well as expert advice. We establish benchmarks that are necessary to get learners onto a successful reading trajectory without being so aspirational that no-one can reach them.

This research is an exploratory analysis of early grade reading assessment (EGRA) data in the following Nguni languages: isiZulu, isiXhosa, and siSwati. This is the largest available dataset on early grade reading in these languages and contains multiple reading assessments for nearly 16,400 learners in early grades. These data allow us to provide the first estimates of early reading thresholds and benchmarks in these languages.

The next section sets out a rationale for developing early grade benchmarks. Section 3 lays out the theoretical framework for early reading which underlies the approach adopted in the benchmarking exercise. Section 4 outlines the data and methodology used in the report, which is followed by a descriptive analysis in section 5 to present sample-based reading norms and trends in the 3 Nguni languages. Section 6 shares results of the data analysis from which the benchmarks derive. Section 7 concludes with a summary of the benchmarks and recommendations for their use.

2. Rationale for developing early reading benchmarks

The poor reading performance of learners in South Africa has been well documented for over 15 years. But it is probably the results from the most recent large-scale Progress in International Reading Literacy Study (PIRLS 2016) which attracted the attention of Presidency. It showed that 78% of grade 4 learners, mostly tested in their home language, could not reach the low international benchmark – a signal for the ability to read for meaning (Howie et al., 2017). PIRLS points to extremely poor reading comprehension skills, and the pressing need for schools to improve instruction in this area.

Improving reading comprehension, however, requires an acknowledgement that reading is a complex phenomenon. Numerous knowledge bases, skills and processes that underpin reading comprehension need to develop and coordinate for learners to become proficient at comprehending text when they read on their own. While assessments such as PIRLS can identify that learners are not proficient in reading comprehension, they cannot tell us which reading subskills pose problems for learners who struggle to read. They provide little guidance on the subskills of reading that should be targeted in the classroom. The national school curriculum guides the expected competencies learners are required to master in each grade. But it is silent on how to measure a learner's achievement in specific reading subskills in African languages. To identify foundational decoding problems, assessments of code-based knowledge and skills are needed, which is why early grade assessments of reading are required in combination with reading comprehension data.

2.1. Why are reading benchmarks useful?

To prevent learners from falling behind in the developmental sequence of reading (further described in the next section), we need a shared vision of what reading success looks like at each grade level. Reading benchmarks contribute to a shared vision of reading success in the following ways:

- they provide a set of norms or expectations for reading performance;
- they give specificity to the Foundation Phase curriculum;
- they make teachers aware of developmental milestones that learners should reach to become skilled readers;
- they provide concrete objectives that help teachers, schools and district officials gauge progress; and
- they serve as a form of quality control within an education system so that large numbers of learners do not fall through the cracks.

Additionally, at various levels of the education system, reading benchmarks can clarify goals and expectations for reading; can contribute to alignment in assessment practices and guide the tracking of progress in reading across the system against national goals as shown in Table 1.

The efficacy of benchmarks lies in their simplicity. Having too many benchmarks puts a burden on teachers, can complicate classroom practice, and they can become an unwieldly or a controlling administrative tool. On the other hand, having only a single benchmark raises questions about where in the developmental sequence it would fit, and ignores the complexity of the reading process and the intricate interactions between lower and higher order skills in reading development. A case is therefore made in this report for the need for more than a single benchmark, while keeping things simple.

Table 1: The uses of reading benchmarks at various levels of the education system

System levels	Clarifying goals and expectations for reading	Clarity & alignment for the effective assessment of reading	Clarifying intervention priorities
 **NATIONAL AND PROVINCIAL ADMINISTRATION**	Benchmarks articulate and communicate an education system's definition of reading proficiency. This promotes alignment of goals across the system with Goal 1 of the DBE Action Plan to 2030.	With a shared understanding of expectations for reading proficiency, the reading assessment process is unified across the system.	With a shared understanding of the size and extent of reading gaps across the system, this paves the way for a constructive intervention response.
 **SCHOOL**	Benchmarks establish standards and targets that school leaders can aim towards and are aligned with national goals for reading proficiency.	Standardises assessment practices across and within schools and ensures that school level assessment is aligned with national goals for reading proficiency.	Clarifies the extent of remedial support required in specific schools and required budget allocations. Focuses intervention responses on the improved teaching of reading and the provision of reading support materials.
 **CLASSROOM**	Benchmarks establish standards and targets that teachers and students can aim towards.	Teachers can determine how many children in their class are on track with their reading.	Target remedial programmes at learners at risk of not being able to read.
 **HOUSEHOLDS**	Benchmarks establish standards for parents against which to assess their children's reading proficiency.	When assessment is linked to standards, and communicated clearly in school reports, this provides meaningful information to parents on how well children read. They can engage in their child's journey to reading proficiency.	Parents and communities are empowered to identify if schools are providing necessary opportunities for their children to learn to read and to partner in remedial programmes.

2.2. The need for language specific benchmarks

Benchmarks for reading exist in various other languages and countries, with particularly well-defined ORF benchmarks for early grade reading in English (Hasbrouck & Tindal, 2006). Until now, detailed work on the development of reading benchmarks in the African languages has not been done in South Africa. One cannot simply transfer reading benchmarks from English due to differences in the phonological, morphological, and orthographical features of African languages. These differences are highlighted in Table 2.

Table 2: Some differences between Nguni languages, Sotho languages and English

NGUNI LANGUAGES	SOTHO LANGUAGES	ENGLISH
Agglutinating Affixes and word stems are "glued" together to form new meanings. E.g. isiZulu: Bakufundisile Stem: fund	**Agglutinating** Affixes and word stems are "glued" together to form new meanings. E.g. Southern Sotho: ba le rutile Stem: rut	**Analytic** Use of helper words, prepositions, and word order to convey meaning. E.g. They taught you.
Transparent/ Shallow There is a one-to-one letter to sound correspondence. E.g. 'g' is pronounced the same in "ngoko" and "jonga" in isiXhosa	**Transparent/ Shallow** There is a one-to-one letter to sound correspondence. E.g. 'g' is pronounced the same in "gape" and "morago" in Sepedi	**Opaque / Deep** The same letter can represent different sounds in different words. E.g. the sound 'g' differs in the words: 'gate' and 'germ'
Conjunctive Morphemes (the smallest meaningful unit) are merged together into single written words. E.g.: isiZulu: Ngiyabathanda	**Disjunctive** Morphemes mostly appear as single words. E.g. Sepedi: ke a ba rata	**Disjunctive** Morphemes mostly appear as single words. E.g. I like them

Early reading development differs from one language to the next. In transparent codes (such as Finnish or Nguni languages), accuracy is reached relatively early. This may result in language comprehension being more strongly related to reading comprehension than decoding at earlier grades than for opaque languages (such as English). The straight-forward nature of letter-sound knowledge in transparent languages may allow learners to achieve mastery of this foundational skill earlier in their reading journey.

However, other features of Nguni languages may delay mastery of basic decoding skills. Nguni languages are agglutinating languages where prefixes or suffixes are "glued" to a word stem to form new meanings. Nguni languages also have a conjunctive orthography where morphemes (the smallest meaningful unit of language) are merged into single written words. This results in both long words and a high degree of visual similarity within and between words, making decoding more challenging. Not all African languages do this. For example, the Sotho languages (Setswana or Sepedi) represent these meaningful units orthographically as separate, smaller words rather than very long words. They have a 'disjunctive' orthography. Like other African languages, Nguni languages also include a large number of complex consonant sequences (examples include *hl, dl, kh, tsh*, *ndl, gcw, ntsw)* and they have a larger code set than English (isiZulu and isiXhosa have about 60 phonemes compared to 44 phonemes in English).

Benchmarking processes clearly need to take account of the linguistic features of the language for which the benchmarks are being developed. At the simplest level, it makes no sense to compare fluency across languages with vastly different word lengths (Spaull, Pretorius & Mohohlwane, 2020). Beyond that, one needs to allow for language specific accuracy-speed and fluency-comprehension relationships that reflect reading development.

Nguni languages also use the Roman alphabet in their writing system. Although research on reading development in African languages is still in the early stages, given the nature of alphabetic writing systems and the way in which the human brain processes written language, it is likely that reading in African languages will exhibit common developmental features associated with both (i) alphabetic writing systems and (ii) with other agglutinating languages. However, there will also be

unique aspects that are specific to reading development in African languages.[9] Benchmarks that derive from early reading development in African languages will thus be unique to those languages, especially with regard to foundational reading skills.

The analysis of EGRA data sets used in this study - containing an extensive number of reading assessments in 3 Nguni languages (isiZulu, isiXhosa and siSwati) - provides a clearer view of how reading development unfolds in these specific African languages, and forms a strong empirical basis from which decisions on benchmarking are derived. Although data analysis may reveal patterns and trends, a theoretical framework helps to interpret and make sense of data, account for the components, processes and their interrelationships, and gives overall coherence to the statistics presented. To this end, the literature review which follows provides a framework that identifies some of the theoretical issues in early reading, and specific issues that may be pertinent to reading in agglutinating languages.

9 Distinguishing what is generic and what is African-language specific in reading development awaits ongoing research.

3. Theoretical framework: Reading as a complex phenomenon

Reading is all about comprehension. Whatever we read, the intent in doing so is to understand the information in the text and construct meaning from it. While the goal of reading is clear, reading itself is a complex phenomenon comprising different components and processes, all of which interact and coordinate together to make it function optimally. Further, there are numerous knowledge bases, skills and processes that underpin reading comprehension and need to develop and coordinate for learners to become good at understanding a text.[10] The following factors are necessary to understand a written text:

- **Linguistic factors:** Knowledge of the language of the text at a sublexical (phonological and morphological), lexical (vocabulary), sentence (morphology and syntax) and discourse-level.
- **Code-based factors:** Knowledge of the written code, its conventions and the skills associated with 'decoding' a text. This is what is meant when we refer to 'decoding' skills. The most foundational component of decoding is alphabetic knowledge (or knowledge of letter-sound relations). Decoding also involves knowledge of syllables and word reading. The next level of decoding is that of reading words in context, also referred to as oral reading fluency (ORF).
- **Text factors:** Knowledge of text conventions; the functions that different genres of text serve; the way information is structured in different genres within and across paragraphs; the role of headings; visuals, etc. Text or topic complexity, topic familiarity and word frequency levels can also affect reading comprehension.
- **Cognitive factors:** A range of lower to higher level cognitive processes such as parsing literal meaning, reasoning, inferencing and perspective taking are required to understand a written text. Additional factors also play a role in reading comprehension: metacognition such as monitoring comprehension and reading strategically (Block & Pressley, 2002; Oakhill, Cain & Elbro, 2015); cognitive resources and constraints such as working memory; processing agility as measured by rapid automatised naming (RAN); attentional control and inhibitory functions (Oakhill, Hartt & Samols 2005; Currie & Cain 2015).

Since there was little information on linguistic and cognitive factors in EGRA studies, our analysis focuses on examining code-based factors in early reading and their interactions. We look at reading comprehension more closely in the next section and then return to a more detailed examination of these code-based factors.

3.1. Implications of reading comprehension theory for establishing reading benchmarks

Reading comprehension entails different dimensions of understanding, ranging from easy literal understandings, to increasingly complex and more abstract levels of understanding (referred to loosely as higher order cognitions). In reading comprehension, *the text* is pivotal to how content is understood and how meaning is constructed.[11] The knowledge that a reader brings to the reading process is very important and affects reading comprehension. Further, the reader's linguistic, cognitive and code-based skills mediate how the reader's knowledge interacts with what is in the text to produce comprehension.

Since the 1980s, cognitive theories of reading comprehension assume that comprehenders construct an abstract representation of the 'state of affairs' described in a text that goes beyond

10 These developmental orchestrations manifest in differences in reading performance across the grades (where grade is a proxy for developmental phase).
11 Different taxonomies of knowledge and understanding (e.g. Bloom, Barrett) are sometimes used to account for reading comprehension. What these knowledge taxonomies have in common is that they are hierarchical (although the relational dynamics between the levels are not always specified).

the literal information (Van Dijk & Kintsch 1983; Kintsch, 1988). This is referred to as the *situation model*. Comprehenders draw on and integrate various linguistic, code, text and general knowledge and cognitive processes to build, monitor, update and modify the situation model during the reading process. But learners who are weak comprehenders struggle with meaning construction and integrative processes (Stafura & Perfetti, 2017; Kim 2017, 2020; Van den Broek & Kendeou, 2017). Three principles underpin cognitive-linguistic theories of reading (Kintsch, 1988; Stanovich, 2000; Stafura & Perfetti, 2017; Kim, 2020; Van den Broek & Kendeou, 2017) with implications for setting reading benchmarks.

i. **Reading is hierarchical**: The processing of information involves multiple levels of increasing abstraction. The development and application of lower-level components and representations are necessary for higher order components and representations.[12] This principle has relevance for the establishment of lower reading thresholds and higher benchmarks to signal milestones at different points in the hierarchy. Having a benchmark at only the higher end of the hierarchy (e.g. fluency to support comprehension) does not show where learners who do not meet the benchmark are struggling lower in the hierarchy. Likewise, having a threshold only at the lower end (e.g. letter-sound knowledge) can help to support reading development to the next level but not necessarily to higher levels in the hierarchy.

iv. **Reading is interactive:** Different components 'speak' to one another and coordinate. Sometimes the relationship is bidirectional. (For example, language proficiency predicts ease of early reading. But once learners can read, and the more they read, the more their language develops). This principle also supports the establishment of more than one reading benchmark to evaluate how well component skills articulate with one another. For example, if a child meets a letter-sound threshold, this provides a good basis for word reading ability. Performing below a letter-sound threshold means that there will be low or no articulation with word reading.

v. **The reading process is dynamic**: Reading comprehension can change as a function of several text factors (grammatical, vocabulary or topic complexity in a text), environmental (parenting practices, exposure to print) or pedagogic factors (inadequate reading instruction, lack of print resources). These factors can influence reading development and performance on a text. With relevance to benchmarking, this principle calls for ensuring consistency across reading assessments used to measure early grade reading (e.g. similar text length and complexity) and context-sensitivity in the setting of benchmarks. For example, it would be inappropriate to derive reading benchmarks with data from high performing, well-resourced schools when most learners come from poor performing and resource-lean schools.

While cognitive factors involved in reading do not differ based on a learner's language or cultural affiliation, there are evidently differences in language types and written codes that may affect the ease of learning to read and the rate of development in the early stages. This necessitates the development of language-specific benchmarks, a notion that gains further impetus as we consider the relationship between code-based factors and reading comprehension in the next discussion.

3.2. Relationships between code-based factors and reading comprehension

In this section, two issues that have a bearing on understanding the nature of early reading development in alphabetic languages are briefly discussed because of their relevance to early reading in the Nguni languages. The first issue relates to the nature of the code set and its influence on the development of early decoding skills. The second relates to the nature of the relationship between decoding skills and reading comprehension.

12 Lower levels of cognitive processing such as working memory, RAN and attentional control are necessary for foundational code-based skills, and both of these are necessary for higher level cognitive processes such as inferencing, perspective taking and comprehension monitoring (Oakhill, Hart & Samols, 2005; Strasser & del Rio, 2014).

Box 1: A brief discussion of code-based factors

The most foundational component of decoding, namely **alphabetic knowledge (or knowledge of letter-sound relations)** has been found to be important in early reading (Adams 1990), in both opaque and transparent languages (Caravolas et al. 2013), and across different language types (Protopopas et al. 2019). This is typically measured through letter-naming and/or letter-sounding tasks, often timed for 1-minute.[13] When children learn letter-sound relations, they develop an explicit awareness of individual speech sounds within words (Ziegler & Goswami 2005). Kim and Piper (2019) in their study of reading across 3 African languages in Kenya (Eastern Bantu agglutinating languages), found letter-sound knowledge in grade 2 had strong direct effects on syllable and word reading, and significant indirect effects on reading comprehension.[14]

The **syllable** is prominent in African languages and builds on alphabetic knowledge since syllables are represented by vowel (V) or consonant vowel (CV) letters. Syllables have an open structure (V or CV), with vowels clearly marking syllable boundaries. Despite the salience of syllables in African languages, syllable reading in these languages has not yet been widely studied. We do not focus explicitly on syllable reading in this study. But we do find remarkably similar relationships between accuracy and speed of reading when breaking-down words into their individual syllables rather than just analysed as whole words.

Word reading is also a foundational decoding skill that is associated with early reading success. Phonemic awareness, letter-sound knowledge and RAN are skills that underlie word reading ability across orthographies (Caravolas et al. 2013; Torppa et al. 2016).

The next level of decoding is that of **oral reading fluency** (ORF) or **text fluency**. We give more attention to this in section 3.2.2.

3.2.1. The nature of the code set and the development of early decoding skills

Writing systems typically represent speech at the sublexical level, using letter symbols (or graphemes) to represent individual sounds (alphabetic writing) or syllables (alphasyllabaries). The nature of this code set has implications for the development of early decoding skills. We highlight 3 important aspects here:

i. **Languages differ in terms of how transparent the grapheme-phoneme relation is. This has implications for reading development.** In many languages there is a one-to-one relationship between phonemes[15] and graphemes[16] (e.g. the Nguni languages). In such cases the orthography is said to be transparent or shallow. However, in some languages this relationship is more complex, where 1 letter can represent different phonemes (e.g. 6 letters (a,e,i,o,u,y) represent about 22 different vowel phonemes in English), or different letters can represent the same phoneme (e.g. the phoneme /f/ is f in *frog,* ph in *phoneme* and gh in *cough).* In such cases the orthography is said to be opaque or deep (e.g. English, French, Portuguese, Japanese Kanji). Katz and Frost (1992) posit that learning to read in languages with opaque orthographies takes longer than transparent ones.[17]

ii. The number of letters used to represent language specific phonemes (i.e. the code set) differs across languages. Typically **the larger the number of symbols in the code set, the longer**

13 The relationship between phonemic awareness and alphabetic knowledge seems to be bidirectional, and some researchers argue that alphabetic knowledge is predictive of phonemic awareness (Vihman 1996).

14 In Finnish, letter-sound knowledge is a strong predictor of early reading, and it remains a predictor of reading ability to the phase where fluency is mastered. Nearly all the letters are taught in the first semester of Grade 1 and children master letter-sounds quickly, reaching a ceiling level by the end of Grade 1 (Aro 2017). In a study, letter-sound knowledge was found to be the strongest predictor of reading skill even in Grade 4 (Leppänen et al. 2008).

15 Phonemes are the most basic sound units that carry meaning in a language. Phonemes can be put together to make words.

16 A grapheme is a way of writing down a phoneme. Graphemes can be made up from 1 letter e.g. p, 2 letters e.g. sh, 3 letters e.g. tch or 4 letters e.g ough.

17 Evidence supporting this comes mainly from European languages, where children learning to read in more transparent languages can reach ceiling levels in decoding by the end of grade 1, whereas English and French children take a year or 2 longer (Seymour, Aro & Erskine, 2003; Caravolas et al., 2013). Research on early reading in Finnish and Turkish have found that children can reach mastery level of these transparent codes within the first year of schooling (Aro, 2017; Miller, Kargin & Guldenoglu, 2014).

it takes to learn them (Nag, 2012, 2017).[18] For example, Finnish has a small code set which is entirely transparent: 25 phonemes and 25 letters of the alphabet to represent them (Aro, 2017). English has a larger code set with 44 phonemes and 26 letters to represent them, but the relationship between the letters and phonemes is not always transparent. The Nguni languages have a larger code set than English; isiZulu and isiXhosa have around 60 phonemes[19] with 26 letters to represent them and many digraphs to represent more complex consonant phonemes, but like Finnish, the orthography is transparent.

iii. **The visual complexity of the orthography can also affect early reading development** (Nag, 2017). Reading words that are visually dissimilar, as in English, is easier than reading words that have complex consonant sequences or that repeat similar visual patterns (Abadzi, 2006). Land (2015) found a high rate of recurring syllables in Zulu text. This increases the visual similarity of syllables within the longer conjunctive word units typical of the Nguni languages. This repetitive visual patterning requires attention to detail and accuracy (and hence more cognitive work) to develop fluency.

Although Nguni languages have a transparent orthography and are thus potentially easier for learning to read, this advantage may be offset by a somewhat larger code set, a complex consonant system and a high degree of visual similarity between syllables.

3.2.2. The relationship between decoding and reading comprehension

In this report we establish a lower threshold and a benchmark for oral reading fluency (ORF) in the Foundation Phase. ORF refers to the ability to read words in context with speed, accuracy and prosody.

- Accuracy reflects the number or percentage of words that are read correctly.
- Speed reflects the number of words that are attempted in a time period.
- Prosody reflects how natural reading sounds (how it conforms to speech rhythms and intonation patterns and reflects punctuation conventions).

Prosody is difficult to measure in field studies because assessing it is subjective. Thus, measures of ORF typically focus only on speed and accuracy. In this report, the term *fluency* is used to describe reading with speed and accuracy and is typically measured by the number of words correctly read per minute from a passage of text.

We benchmark ORF due to the nature of relationship between decoding skills (including ORF) and comprehension. Reading with fluency enables cognitive resources (e.g. working memory or attention) to be allocated to higher level demands needed for reading comprehension. The relationship between fluency and reading comprehension has been well established in English with regard to both accuracy and speed (Deno, Fuchs, Marston & Shin, 2001; Wolf & Katzir-Cohen, 2001). Accuracy has been shown to develop first, and once accuracy is established, reading rates increase as children's mastery of reading increases (Fuchs et al. 2001; Spear-Swerling, 2006). Individual variation occurs mainly in reading rate – weaker readers read accurately, but they read slower than stronger readers. Problems with reading speed also appear to be persistent where weak readers display increasing lags in reading speed.[20]

The same pattern seems to exist in transparent languages, albeit even earlier than in English.[21] In studies of early reading across 3 African languages in Kenya, Kim and Piper (2019) found that fluency was strongly related to reading comprehension.

18 Alphabetic codes tend to have smaller code sets because the code is phonemic; alphasyllabaries are much larger (running into the hundreds) because there are more syllables than phonemes in language; while logographic code sets (e.g. Chinese) are the largest, comprising thousands of characters.

19 Personal communication with Zulu linguistic specialist Lionel Posthumous. However, given the complex phonetic nature of many African consonant sounds, he cautions that there is not consensus yet on the exact number of phonemes.

20 Kim, Park and Wagner (2014) found that for readers in a transparent orthography (Korean), fluency predicted reading comprehension in readers younger than those in English.

21 In some of these studies, poor reading has been shown to be marked by poor speed, not accuracy (Wimmer, 2006). Aro (2017) reports that in Finnish, accuracy is achieved early and remains high – ceiling effects in reading accuracy are found by the end of grade 1.

Pikulski and Chard (2006) and Adams (1990) explain fluency as helping children 'unglue' from text. This implies bidirectional relationships between fluency and reading comprehension. In other words, children who understand what they read tend to read more fluently.[22] As a result, oral reading fluency has been described as the bridge between decoding and reading comprehension (Pikulski & Chard, 2006). However, theories of reading differ as to the nature of the relationship between decoding and comprehension.

3.2.2.1. The simple view of reading

The basic premise of the simple view of reading (SVR) is that reading comprehension can be predicted by the following equation: Reading Comprehension = Decoding x Linguistic Comprehension[23] (Gough & Tunmer, 1986; Hoover and Gough, 1990; Tunmer & Chapman, 2012). The multiplicative nature of this formula means that both decoding and linguistic comprehension are necessary but, on their own, not sufficient for reading comprehension.[24] Hoover and Gough (1990) propose that the balance between the 2 components changes across the grades. In the early stages of reading, decoding (operationalised as word reading, non-word reading or as text fluency) is predicted to have more influence, since children need to master the written code. As decoding skills approach mastery level, their effects on reading comprehension generally decrease, and linguistic comprehension becomes a stronger predictor of reading comprehension (Fuchs et al, 2012; Caravolas et al, 2013; Torppa et al, 2016).

While this SVR relationship has generally been found to hold true across reading in opaque and transparent orthographies (Kendeou et al, 2009; Kendeou, Papadopoulos, & Kotzapoulou, 2013; Kim & Wagner, 2015), a meta-analysis (Florit & Cain, 2011) found differences in the balance between decoding and linguistic comprehension and their influence on reading comprehension.[25] In transparent codes, accuracy is reached relatively early and so decoding (as measured by accuracy) is not as strongly related to reading comprehension as linguistic comprehension. However, when decoding is measured by fluency, then its importance for reading comprehension is found to extend to later grades (Florit & Cain, 2011).

Despite widespread evidence supporting it, the SVR has been criticised for being too simplistic as it does not consider a range of other important linguistic, cognitive and text factors.[26]

3.2.2.2. The decoding threshold hypothesis

The SVR is predicated on a linear relationship between decoding and reading comprehension. The decoding threshold hypothesis of Wang, Sabatini, O'Reilly and Weeks (2019) argues that this relationship is more complex and is more likely to be non-linear. They posit a threshold condition to account for the relationship between decoding and reading comprehension. The relationship between decoding and reading comprehension is only observed when decoding occurs above a lower bound threshold level. For learners who decode below the threshold, reading comprehension is unlikely to develop until decoding can be improved to a level above the decoding threshold (Wang et al. 2019).

In a longitudinal study involving 30,000 learners, the authors tracked learners from grades 5 to 10. Grade 5 learners whose decoding skills were above the threshold showed steady improvement in reading comprehension over the years, while those below the threshold in grade 5 showed

22 Some studies have shown that after controlling for word reading, reading comprehension also makes an independent contribution to ORF in English (Jenkins et al., 2003) and in Korean (Kim, 2015).

23 Linguistic comprehension is defined as the process by which lexical (i.e., word) information, sentences, and discourses are interpreted.

24 In the original studies, decoding was operationalised as word reading ability, while linguistic comprehension referred to language processing at the sentence and discourse level and was operationalised as listening comprehension. Subsequent studies have operationalised decoding as word reading, non-word reading or both, or as text fluency, and linguistic comprehension as vocabulary or as both vocabulary and listening comprehension.

25 For example, in English, decoding accounted for greater variance in reading comprehension for longer, because children learning to read in English take longer to acquire accuracy in the code.

26 Using a direct and indirect effect model, Kim (2016, 2017) has examined pathways of relations between language, decoding, cognitive and literacy skills to reading comprehension in various studies, with Korean and English readers across grades. She argues that the SVR fits the data well, and that word reading and listening comprehension "completely mediate" the relations of language and cognitive skills to reading comprehension (Kim, 2017).

minimal improvement in reading comprehension. No evidence was found of other skills playing a compensatory role to drive reading comprehension when decoding was insufficient. In other words, if threshold decoding skills are not in place, reading comprehension remains stagnant.

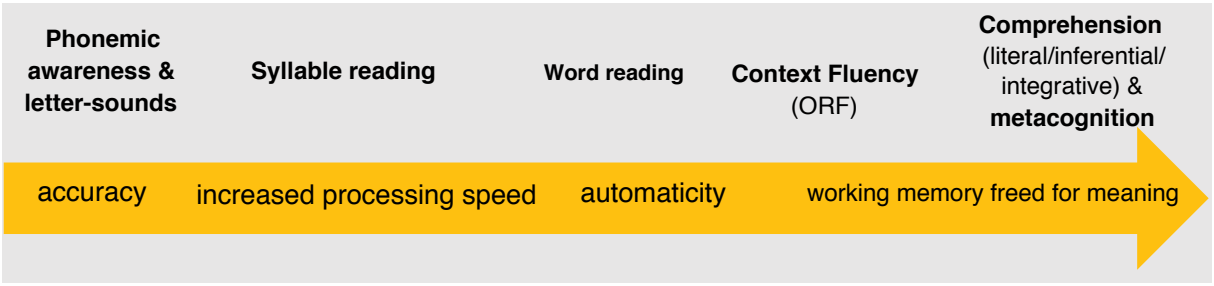
The authors also speculate that there may also be an upper threshold, beyond which there are no additional gains for increased decoding skills. For example, extremely fast decoding does not improve reading comprehension and may cause gaps in understanding. While evidence for the decoding threshold hypothesis in this study came from older learners (grades 5 to 10), it is likely that a decoding threshold applies in the early stages of reading when the relation between decoding and reading comprehension is even stronger. It is also likely that this threshold will be language specific, determined by the features of the language.

3.3. Pulling the threads together: A developmental view of multiple proficiencies

From a cognitive-linguistic perspective, the dynamic, relational and hierarchical nature of skilled reading means that foundational language and code-based knowledge and skills are critical for understanding texts beyond their literal information.

As illustrated in Figure 1, different processes come into play at different stages of development (Stanovich, 2000). A skilled reader in grade 1 is not the same as a skilled reader in grade 3. Furthermore, a learner's developmental progression may be influenced by the nature of the code set of the language in question. Across the components, accuracy develops first, followed by increased processing speed, which then leads to automaticity in processing (processing without effort or conscious attention). This frees up working memory and attention for meaning construction.

Figure 1: Developmental cline in early reading



As the importance of some processes as drivers of reading development diminish as proficiency increases and are replaced by qualitatively different processes, the following types of reader can be distinguished on a cline. There are no clear-cut differences between reader types, thus they are placed on a continuum to show how they sequence as different aspects of proficiency in reading increase:

Non-readers show poor print awareness, poor phonological awareness[27] and have very low letter-sound knowledge (and alphabetic awareness) and immature handwriting. Their ability to read words correctly (accuracy) is minimal. Performing below a minimal letter-sound threshold impedes their ability to decode text.

Emerging readers have developed phonological awareness and acquired some basic knowledge of letter-sounds to enable them to blend letters to form syllables or words. Accuracy is increasing, and with it, increased processing speed to read words in or out of context. However, reading is still halting and effortful, and chunking of words into meaningful phrases is not yet regular. Reading comprehension is limited when they read a text on their own.

27 Phonological awareness is a broad skill that includes identifying and manipulating units of oral language – parts such as words, syllables, and onsets and rhymes. Phonemic awareness refers to the specific ability to focus on and manipulate individual sounds (phonemes) in spoken words.

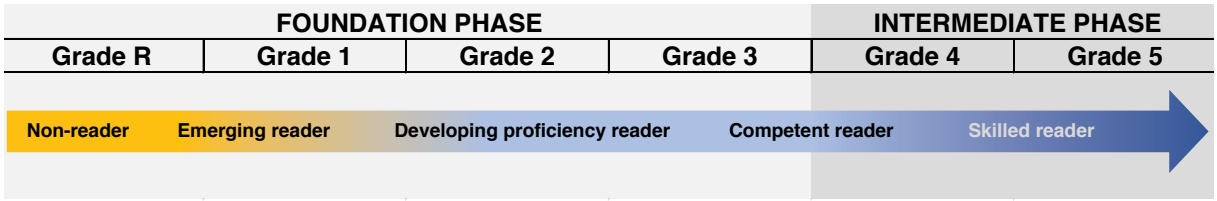
Readers in the next phase are **developing proficiency.** They have more accurate and fluent knowledge of the alphabetic code, which enables them to decode syllables and words in or out of context with greater accuracy (approximating 95% accuracy which refers to correctly reading 95 of 100 words attempted from a passage). Their processing rate increases to a point where some words are read automatically, and they move beyond the level of sounding out words to articulating meaningful phrases. While their decoding skills are not yet fully automatised, they have freed up enough working memory to construct basic meaning from what they read to support comprehension.

Competent readers have reached a stage where decoding is accurate (at least 95% accuracy) and largely effortless. Their reading rate is quite advanced for their grade level and they read sentences with natural intonation or prosody. They can read texts containing more complex language and less familiar words; they engage more actively with the text and understand much of what they read. They can respond to questions requiring both integrating information from a specific place in the text (local) with a wider (global) view of the text. Reading becomes a tool for learning – they start learning new things when reading on their own, without mediation from a teacher/adult. They will reread a section of text if comprehension breaks down.

Skilled readers read words in and out of context accurately, effortlessly, and quickly, seldom making decoding mistakes. Their reading is automatised, they chunk words into meaningful phrases and construct and integrate meaning. They are equally good at making local and global inferences across the text. The ability to 'read to learn' comes naturally and they will often voluntarily read for information or pleasure. They readily pick up inconsistencies in a text or discrepancies in perspective.

Figure 2 loosely maps the reading skill cline against grade progression in the early school years.

Figure 2: Developmental continuum in early reading in relation to formal grades



By the time learners exit the Foundation Phase, most of them should be competent readers or at least transitioning from *developing proficiency* to *competent reading*. This developmental sequence shows a general trend; but there will always be exceptions. For example, there may be *emerging readers* in grade R and *competent readers* in grade 2, or a few *non-readers* at the end of grade 1. However, there should not be any *emerging readers* at the end of grade 3. Such a developmental lag would point to challenges in pedagogy and the teaching/learning context.

It is important to note, however, that the developmental cline does not mean that reading instruction should follow this order. CAPS recommends a balanced approach where decoding and comprehension skills are developed in tandem. Activities such as Listening and Speaking, Shared Reading, Group Guided Reading, Paired Reading and Read Alouds help children develop language, comprehension and thinking skills while Phonics is assigned to the development of decoding skills, supported by Handwriting activities. Group Guided Reading can also be used to develop decoding and fluency skills in the early grades. As mastery in decoding increases, the need for targeted phonics instruction diminishes.

4. Methodology

4.1. Data

For this benchmarking exercise, we combined data from different reading and literacy studies that have been implemented recently in South Africa. These include: the Story Powered Schools (SPS) evaluation; the Early Grade Reading Study II (EGRS II) the Funda Wande (FW) evaluation; the Leadership for Literacy (LFL) project and the Zenex Foundation Literacy Project (Zenlit). Each of these studies were conducted between 2017 and 2019 and assessed learners at some point between grades 1 and 6.[28] All 5 of these studies collected Early Grade Reading Assessments (EGRA) in isiXhosa, isiZulu or siSwati (Ardington & Meiring, 2020; Ardington, Hoadley & Menendez, 2019; Department of Basic Education, 2019; Taylor, Wills & Hoadley, 2019; Zenex Foundation, 2018). Furthermore, 4 of these studies are longitudinal in nature, allowing us to track the reading trajectories of children over time. As explained later, this feature is used to establish the predictive validity of our established benchmarks and threshold.

4.1.1. Reading measures across datasets

In its pooled form, we use the combined data in a preliminary analysis to identify sample-based reading norms and trends across grades on specific code-based factors – namely letter-sound knowledge, isolated (or single) word reading and oral reading fluency (ORF). Sub-sets of the pooled data, including reading comprehension performance data, are then used in the benchmarking analysis in section 6.

Across the studies, assessment data were collected on letter-sound knowledge for grades 1 to 4 learners, single word reading for grades 1 to 5 learners, ORF and ORF comprehension for grade 1 to 6 learners. In only 1 study (FW) complex consonant sequences are assessed as a distinct task. Table A1 in Appendix A shows by grade, term and language which home language[29] EGRA subtasks were tested in each study. Except for Zenlit, all studies include item-level data rather than just totals per task and were conducted as one-on-one assessments using Tangerine software.

4.1.2. Sample sizes and sample characteristics for identifying reading norms and trends in code-based factors in the early grades

When pooled, this is the largest and most up-to-date source of early grade reading performance data[30] in Nguni languages that exists, to our knowledge. Almost 16,400 unique learners (with up to 4 assessments at different grade-term points) in more than 660 unique schools in 4 provinces (Eastern Cape, KwaZulu-Natal, Gauteng, and Mpumulanga) are represented in the data. Accounting for multiple waves of assessment data per learner, the dataset contains about 15,100 EGRA learner assessments in isiZulu, 12,000 in isiXhosa and about 8,400 in siSwati. Sample sizes of individual studies vary as shown in Table 3 below. The isiZulu and isiXhosa samples are dominated by the Story Powered Schools (SPS) evaluation sample. Except for the Zenlit study, all studies considered are longitudinal (have more than 1 assessment point per learner) and track the same learners over time.

28 The SPS, FW and EGRS II studies were designed as randomised control trials to assess the effectiveness of different literacy programmes on learner literacy in home language or first additional English. The LFL project was a mixed-methods study to assess the functioning of township and rural schools and to explore how leadership contributes to literacy learning environments. Zenlit, aimed to establish the efficacy of literacy programmes to support the teaching of reading.

29 In most of the studies, English was also tested (or was the primary focus as in EGRS II), but we only focus on home language test results in this report.

30 For code-based reading factors

Table 3: Grades at which Nguni language EGRAs were conducted, by individual study

Study	Grade	Language	No. schools	No. of learners*
SPS	Grades 2-5	isiZulu isiXhosa	188 170	5371 5002
EGRS II	Grade 1-3	isiZulu siSwati	49 131	969 2358
FW	Grade 1 and 2	isiXhosa	59	1187
LFL	Grade 3 and 6	isiZulu	42	510
Zenlit	Grades 1-3 Grades 1-3	isiZulu isiXhosa	12 10	538 450

Notes: *Sample size here reflects the entire sample combined for each study by language and where there are multiple waves of data collection, only wave 1 sample sizes are counted. See Table A2 for a detailed breakdown of sample sizes.

SPS assessed learners at 2 points in time between grades 2 to 5. In EGRS II, each learner was assessed 4 times from grade 1, term 1 to grade 3, term 4. Learners in the FW and LFL studies were tested twice (in terms 1 and 4). The only available siSwati sample is from EGRS II. Appendix Table A2 provides more detail on learner and school sample sizes by grade, language and term of testing for each study.

None of these samples were intended to be representative of early grade learners in any province or language group. Table 4 summarises the sample characteristics and illustrates which groups of schools and learners are most represented.

- Learner reading outcomes are predominately obtained from samples in no-fee charging, Quintile 1 to 3 schools.[31]
- Most learners in the pooled dataset are rural based. The rural and urban location of schools varies by study (as shown in Table 4), but the studies with the largest sample sizes – SPS and EGRS II – are predominately rural based.
- Across the datasets, the learner samples tested in isiZulu are in KwaZulu-Natal, Mpumulanga and Gauteng. All the isiXhosa samples are from the Eastern Cape.

Except for LFL[32], almost all learners in the pooled sample were tested in an Nguni language that matched the Foundation Phase (grades R to 3) language of learning and teaching (LOLT) in their school. Across all samples, 98% of learners were tested in their home language.[33] In all longitudinal data, attrition was below 20%. This suggests sample characteristics do not change significantly at higher grades due to loss to follow-up. Appendix Table A3 provides more detail on how schools and learners were sampled in each study.

31 The South African public-school system categorises schools into 'Quintiles' depending on the wealth of the area in which the school is located. The poorest quintiles 1 to 3 schools do not charge fees, while quintiles 4 and 5 schools are typically fee-charging. Quintile 5 schools tend to serve more affluent students, and are of much higher quality and functioning than schools in the other quintiles.

32 Where 20% of learners in the sample were in English LOLT schools.

33 All learners in the SPS, FW and Zenlit studies were tested in their home language. In the EGRS II study, over 90% of learners were tested in their home language.

Table 4: Characteristics of the learner sample across the pooled EGRA dataset by Nguni language

	isiZulu				isiXhosa			siSwati	
	EGRS II	**LFL**	**ZENLIT**	**SPS**	**ZENLIT**	**SPS**	**FW**	**EGRS II**	**Total**
School province:									
Eastern Cape	-	-	-	-	100%	100%	100%	-	41%
KwaZulu-Natal	-	50%	100%	100%	-	-	-	-	2%
Gauteng	-	50%	-	-	-	-	-	-	38%
Mpumalanga	100%	-	-	-	-	-	-	100%	20%
School location:									
Urban	23%	61%	-	-	90%	-	100%	3%	14%
Rural	77%	39%	100%	100%	10%	100%	-	97%	86%
School wealth:									
Quintile 1	74%	25%	34%	38%	-	71%	-	46%	47%
Quintile 2	24%	25%	49%	40%	-	26%	-	36%	30%
Quintile 3	2%	36%	17%	20%	100%	4%	100%	18%	22%
Quintile 4	-	8%	-	2%	-	-	-	-	1%
Quintile 5	-	6%	-	-	-	-	-	-	0.2%
LOLT:									
Nguni language of testing	97%	78%	100%	100%	100%	100%	100%	100%	99%
English	-	20%	-	-	-	-	-	-	1%
Other	3%	2%	-	-	-	-	-	-	0.3%
Tested in home language:	95%	90%	100%	100%	100%	100%	100%	90%	98%
Female:	46%	54%	48%	49%	56%	50%	50%	46%	49%
Attrition:									
Assessments 1-2	-	9-15%	-	16%		16%	6%	-	
Assessments 2-3	-	-	-	-		-	-	-	
Assessments 1-3	20%	-	-	-		-	-	15%	
N - Sample size (unique learners)	**969**	**510**	**538**	**5371**	**450**	**5002**	**1187**	**2358**	**16385**

Notes: *Sample size here reflects the entire sample combined for each study by language and where there are multiple waves of data collection, only wave 1 sample sizes are counted. LOLT = Language of learning and teaching in the Foundation Phase (grades R to 3).

4.2. Data sub-samples used for benchmarking analyses

For establishing reading thresholds and benchmarks, sub-sets of the data were used[34]. Table 5 summarises the samples used in establishing oral reading fluency (ORF) thresholds, and ORF and letter-sound benchmarks. All the samples are large and drawn from substantial numbers of schools. Large proportions of the learners can read at least 1 word in a connected text ranging from 45% at the end of grade 1 to over 80% by the end of grade 3. There is therefore enough variation to detect patterns in accuracy, speed, fluency and comprehension.

Three samples (shown in Table 5, in the column headed "3-minute oral reading") are used in the analysis of oral reading comprehension for benchmarking. Typically, ORF assessments are limited to 60 seconds. The associated comprehension questions are then asked, but only up to the point in a passage that the learner reached in 60 seconds. This limits the ability to assess the association between fluency and comprehension, as slow readers are not asked more than a few questions (Ardington & Menendez, 2020). In 3 of the study samples learners were timed to 60 seconds and then allowed to read for a further 2-minutes.

Details of the studies which extend the time limit to 3-minutes, including the associated ORF sample sizes and samples attempting all or at least 5 associated ORF questions, are shown in Table 5. In addition to oral reading comprehension, home language written comprehension data is used to explore further the relationship between fluency and comprehension skills at higher grades. This is available in 4 studies.

While most of the studies have control and intervention samples, the benchmarking analysis (and the preliminary analysis of sample norms) is not concerned with disaggregating results by the intervention status. The treatment and control samples are pooled to increase sample sizes and maximise the possible distribution of reading scores.

34 The specific sub-set used for each threshold depended on the EGRA subtasks assessed at specific grades, the availability of item-level data and the specific nature of the tasks administered.

Table 5: Sample sizes by language, study and grade used for the Nguni language benchmarking analysis

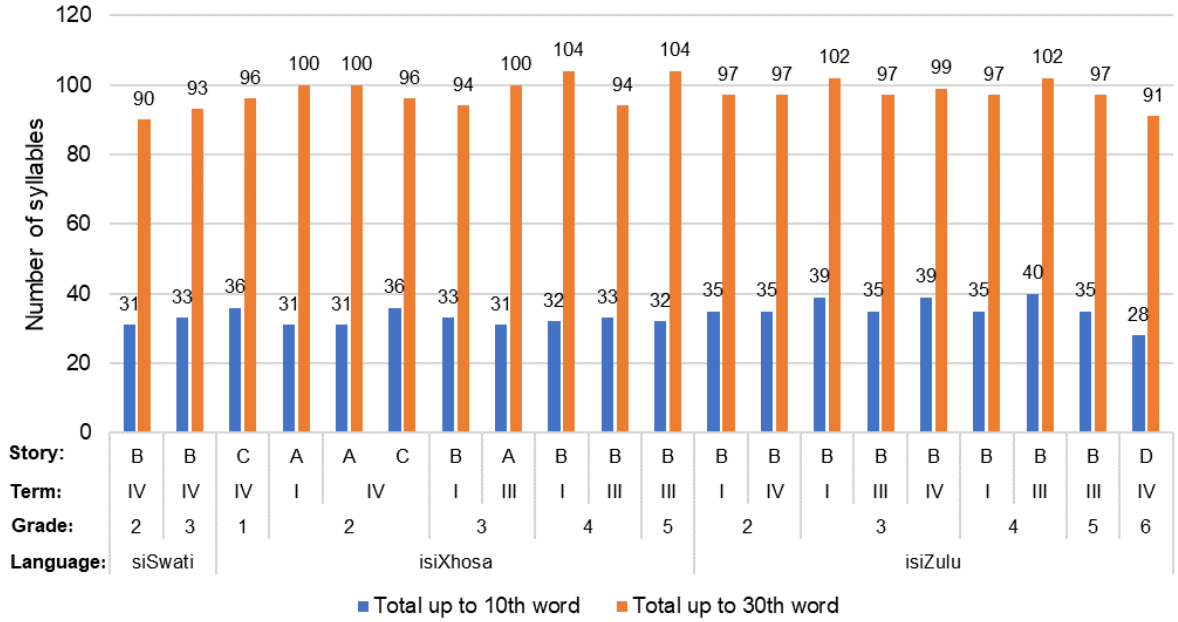
Sample description				Sample sizes by individual task assessed (number of learners)								Number of comprehension questions	
Language	Grade Term	Study	Passage	Letter-sounds (item-level)	ORF (item-level)	ORF >0	ORF > 0 (%)	3-minute oral reading	Attempting all comprehension questions	Attempting at least 5 comprehension questions	Written comprehension	Oral reading	Written
isiXhosa	G1T1	FW	1	595	-	-		-	-	-	-	-	-
	G1T4	FW	1	555	555	251	45%	228	104	193	-	12	-
	G2T1	SPS	1	947	946	535	57%	-	-	-	-	-	-
	G2T1	FW	1	592	592	332	56%	322	154	268	-	12	-
	G2T4	FW	1	561	561	468	83%	444	340	424	-	12	-
	G2T4	FW	2	-	561	431	77%	431	263	414	-	10	-
	G3T1	SPS	1	940	939	632	67%	-	-	-	-	-	-
	G3T3	SPS	1	1410	1410	1204	85%	662	466	625	-	10	-
	G4T1	SPS	1	942	942	780	83%	-	-	-	-	-	-
	G4T3	SPS	1	-	1422	1232	87%	689	562	672	-	9	-
	G5T3	SPS	1	-	1452	1315	91%	720	584	654	1428	9	8
isiZulu	G2T1	SPS	1	925	925	598	65%	-	-	-	-	-	-
	G2T4	EGRS II	1	766	765	433	57%	-	-	-	-	-	-
	G3T1	SPS	1	925	925	717	78%	-	-	-	-	-	-
	G3T1	LFL	1	509	509	405	80%	-	-	-	-	-	-
	G3T3	SPS	1	1477	1475	1293	88%	652	510	601	-	10	-
	G3T4	EGRS II	1	762	762	593	78%	593	514	514	757	5	6
	G3T4	LFL	1	430	430	372	87%	-	-	-	-	-	-
	G4T1	SPS	1	925	925	820	89%	-	-	-	-	-	-
	G4T3	SPS	1	-	1534	1404	92%	730	622	711	-	8	
	G5T3	SPS	1	-	1474	1374	93%	708	560	665	1441	9	8
siSwati	G2T4	EGRS II	1	1995	1994	1252	63%	-	-	-	-	-	-
	G3T4	EGRS II	1	1921	1922	1430	74%	1430	1340	1340	1875	5	6

Notes: G1T1 = Grade 1, Term 1; G2T2 = Grade 2, Term 2, etc. No figures are shown if not applicable.

4.3. ORF text analysis

The texts used to assess ORF and oral reading comprehension were closely analysed to determine their comparability. For assessments from grades 1 to 5, the studies drew on just 3 narrative texts[35]. But the details of the texts (and level of difficulty) varied, even when based on the same story. Table A4 in Appendix B provides a summary of some of the texts used. The texts were remarkably similar in terms of how many syllables were reached by the 10th and 30th words. This is seen in Figure 3. They were also remarkably similar with respect to the number of complex consonant sequences up to the 10th and 30th word as seen in Figure 4. These similarities justify using the various EGRA data together for ORF benchmarking purposes. The technical report provides more detail on the text analysis.

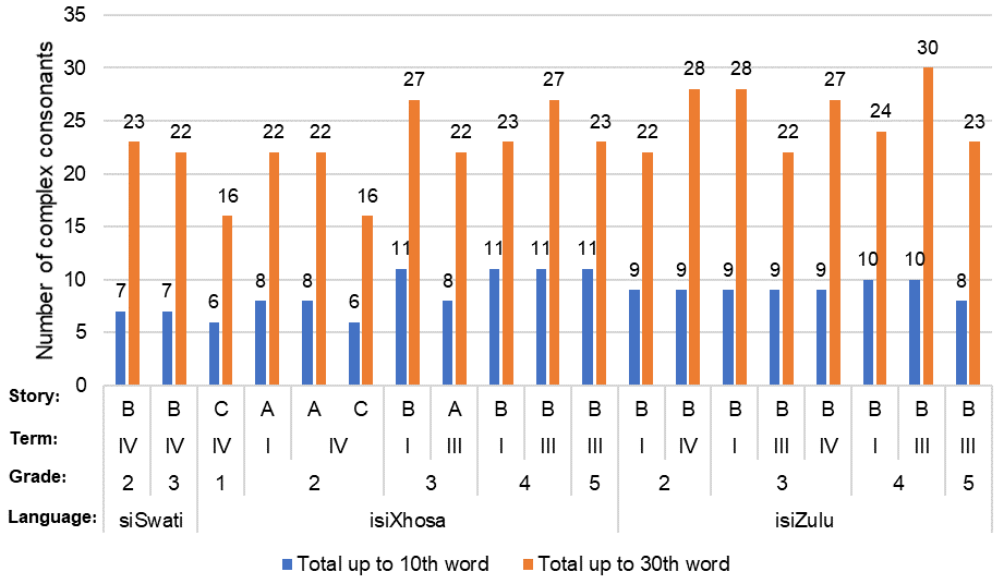
Figure 3: ORF passage comparisons – total syllables up to 10th and 30th word



Notes: A, B, C, D reflect which story was used; Project/Study included: EGRSII, SPS, FW, LFL; Term reflected by I, II, III and IV.

35 The grade 6 LFL text was very different from the grade 1 to 5 texts in other studies. Thus, it is not used for the benchmarking analysis.

Figure 4: Total complex consonant sequences up to the 10th and 30th word of each ORF passage



Notes: A, B, C, D reflect which story was used; Project/Study included: EGRSII, SPS, FW, LFL; Term reflected by I, II, III and IV.

4.4. Categorisation of comprehension questions

To categorise reading comprehension questions across the studies' texts, we adopt the conceptual framework from PIRLS to classify comprehension levels. The PIRLS classification reflects 4 main types of processes involved in meaning construction and integration, as reflected in current cognitive theories of reading comprehension. The 4 comprehension processes focus on: i) retrieving explicitly stated information; ii) making straightforward inferences; iii) interpreting and integrating ideas and information; and iv) evaluating and examining content, language and textual elements. They are operationalised as different types of comprehension questions: literal questions; inferential questions; integrative questions; and evaluative questions. Although there is variation in the type of comprehension questions asked, all questions were typically based on narrative texts and mainly tested lower levels of reading comprehension (for example, straightforward inferences).

Appendix B, including Table A4, provides a discussion of the categorisation of comprehension questions by the 4 comprehension processes. We return to the issue of question difficulty level in section 5.5.

4.5. Our approach to benchmarking

Our approach to benchmarking is based on an understanding that there are various stages of reading development (outlined in our theoretical framework) where different processes come into play, and contribute differentially to performance, as reading proficiency increases.

As explained in the theoretical framework, within each reading process, accuracy develops first, followed by increased processing speed. Then as fluency develops, this frees up working memory for higher order reading and comprehension skills. Therefore, we explicitly analyse both accuracy and speed and their interrelationship. Our approach also aligns with the decoding threshold hypothesis put forward by Wang et al. (2019) where reading comprehension is unlikely to develop

until decoding exceeds a lower bound threshold level. This theory also suggests that there may be an upper threshold, beyond which there are no additional gains in comprehension associated with increasing decoding skills. Therefore, the relationship between fluency and comprehension may break down at low and high levels of fluency.

Our approach seeks to identify a letter-sound benchmark, and an ORF threshold and benchmark:

- The **letter-sound** *benchmark* represents an upper threshold, above which there are diminishing returns to a continued focus on improving letter-sound knowledge. It identifies whether learners are developing the foundational decoding skills (in letter-sound knowledge) necessary for accuracy in reading.
- The **ORF threshold** represents the *minimum* level of fluency that must be achieved, to begin engaging higher order skills in the written mode. It identifies learners who are entering an emergent level of fluency which supports reading accuracy but does not yet allow reading for meaning to develop.
- The **ORF benchmark** represents an upper threshold, above which there are diminishing returns to a continued focus on this decoding skill.

Our method relies predominantly on visualisation and non-parametric statistical techniques to examine the relationships between accuracy, fluency and comprehension with the purpose of identifying critical points in learners' reading trajectories. This exploratory data analysis aims to establish whether, empirically, there are regular patterns and trends between speed and accuracy and between fluency and comprehension and whether these relationships exist across studies, languages, grades and reading passages. If regular, consistent patterns exist this could help identify critical thresholds. These thresholds would be independent of the level of difficulty of the text and comprehension questions asked in relation to the text.

Once potential thresholds are identified, we test them to establish whether these critical points provide meaningful distinctions between learners and whether they align with the stages of reading development. This is done using concurrent data (data from the same grade-point) on related reading skills. We also investigate whether the proposed thresholds are set at levels that are attainable by current learners; sensitive to incremental changes in reading performance in this context and, at the same time, ambitious enough to support meaningful improvements in reading proficiency.

Where data were available to follow learners' reading performance over time, this was used to examine how the proposed thresholds/benchmark points are related to previous and future reading proficiency levels. It is necessary to use reading performance data from later grades to develop benchmarks, given the low levels of reading performance in earlier grades. In other words, data from later grades is used to establish the grade thresholds/benchmarks for an 'on track' successful reading journey.

The establishment of a letter-sound benchmark relied on a combination of expert opinion of letter-sound knowledge required for mastery in decoding, together with insights from the data. Where available, longitudinal data (data for the same learner for 2 timepoints) was used:

- to examine incremental improvements in letter-sound knowledge against baseline scores to identify the point beyond which gains were negligible; and
- to investigate to what extent reaching the letter-sound benchmark is related to future reading proficiency.

The approach to benchmarking adopted here has several advantages over previously used methodologies. Firstly, the non-parametric methods we use make no assumptions about speed-accuracy or fluency-comprehension relationships. Linguistic differences between languages may translate into different fluency-comprehension relationships. Thus, non-parametric methods are particularly useful for languages for which there are no existing benchmarks and little research.

Dominant forms of pedagogy (particularly the attention given to teaching comprehension skills) may also affect the fluency-comprehension relationship. Differences in pedagogy could result in different levels of comprehension across contexts, which are unrelated to reading performance. Secondly, our approach maps meaningfully to the development cline in reading and communicates to teachers the points at which they should focus their instruction. The third advantage is explained in Box 2 below, which compares this approach with typically used benchmarking approaches. The key disadvantage of our method is that it requires quite advanced data analytical skills together with expert judgement. As such, this method is not as systematic, and may be difficult to replicate across languages or studies.

Box 2: An explanation of why we do not benchmark ORF to a fixed comprehension level

Typical benchmarking approaches rely on identifying a comprehension threshold (e.g. at least 80% of questions correct) and then using statistical techniques to identify the fluency levels which are associated with meeting that threshold (see for example RTI International (2017)). These methods are very dependent on the level of difficulty of the comprehension questions and their placement in the text (how far a learner needs to read in the text to attempt these questions).

While comprehension processes are hierarchical, there can be a wide range of difficulty within a single type of comprehension question (eg: literal comprehension questions) making it difficult to assess how comparable passages and sets of questions are. This raises significant methodological challenges as it can make the choice of a cut-off for a desired level of comprehension (eg: 80% of questions correct) somewhat arbitrary. For these reasons, we have not adopted this approach to benchmarking (For more details on this reasoning, see Ardington & Menendez (2020). Rather our approach to setting benchmarks or thresholds is independent of the level of difficulty of the text and comprehension questions.

5. Preliminary data analysis: Reading norms and trends

Before presenting the benchmarking process and analyses, we provide more detail on reading norms and trends in the early grades in the Nguni languages. Drawing on the pooled EGRA data and looking within studies, we provide summary descriptive statistics for EGRA subtasks across grades 1 to 6 for the 3 languages. This is followed by an examination of correlations between the different subtasks. We focus on letter-sound knowledge, reading of single words and then fluency in reading connected text and associated comprehension skills. We reiterate that the sample is not nationally representative. It reflects reading performance for learners in no-fee schools in high poverty contexts.

5.1. Letter-sound knowledge

Letter-sound knowledge generally increases with each grade as seen in Figure 5a, which shows the average number of letters learners can sound correctly at the beginning (Term 1) and end points (Term 3 or 4) of each grade. Figure 5b shows the proportion in each sample that score zero (i.e. they cannot sound 1 letter correctly). We also observe *within* study trends in Figure 6a which plots learner performance at the 25th, 50th (median) and 75th percentile, while Figure 6b shows the percentage of learners scoring zero on letter-sounds.

There is considerable irregularity in letter-sound knowledge patterns across the samples and by grade as revealed in Figure 5 and Figure 6. As a result, we cannot provide a 'norm' across these data. Nevertheless, we observe 5 key observations from the data:

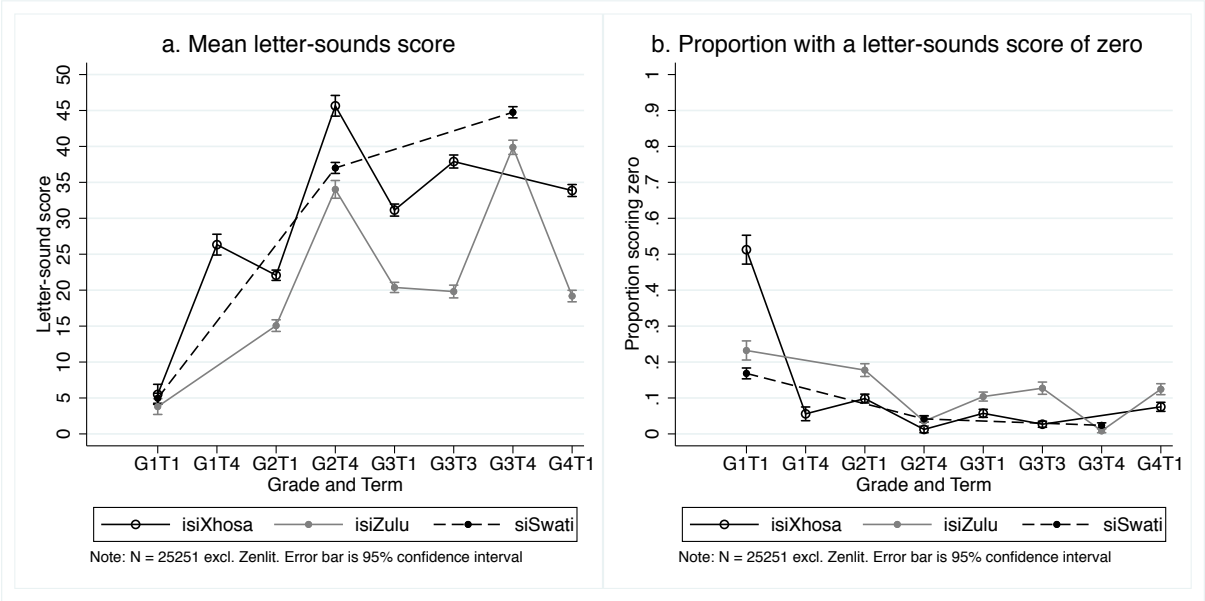
i. **Far too many learners are entering grade 1 with virtually no letter-sound knowledge, despite having attended grade R**. This is seen in Figure 5. At the start of grade 1, on average letter-sound knowledge across the 3 languages is around 5 letters. This problem is particularly pronounced among the Eastern Cape, urban isiXhosa sample (FW): 51% are unable to sound any letter correctly at the start of school and at least 50% do not even *attempt* a letter.
ii. **It is evident that the largest gains in letter-sound knowledge across all samples are acquired in grades 1 and 2.** For example in the first year of school, the isiXhosa FW urban sample make considerable gains in letter-sound knowledge, progressing from a median of 0 at the start of grade 1 to a median of 24 letters cpm by grade 1, term 4. Median letter-sounds cpm more than doubles to 50 letters cpm by the end of grade 2.
iii. **Across all studies gains in letter-sound knowledge are muted in grade 3, even for learners coming off a low base.** For example, the median learner in the isiZulu SPS sample progresses by just 6 letter-sounds cpm from the start of grade 2 (12 letters cpm) to the end of grade 3 (18 letters cpm). Letter-sound knowledge deteriorates overall in grade 4 among the same learners. This suggests it is vital that children master this skill in the first 2 years of school. The small gains in grade 3 are not due to 'ceiling effects' in letter-sound knowledge because the median learner in the isiXhosa Zenlit sample can sound 67 letters cpm by the end of grade 3 (and even this better performing sample underperform internationally[36]). Rather, these findings suggest that there are few opportunities to 'catch-up' this decoding skill after grade 2.
iv. **Despite general gains made in the first 2 years of school, far too many learners are not mastering this basic decoding skill by the end of grade 3.** Substantial proportions of learners across all 3 languages score zero on letter-sound knowledge tasks at the end of the Foundation Phase. At the start of grade 4, around 10% of the isiZulu and siSwati samples cannot sound 1 letter correctly.

36 In Indonesia, a nationally representative sample of grade 2 learners read on average 75 letters cpm in Bahasa Indonesian (a language with a transparent orthography and alphabetic system) and less than 1% score zero on letter-sound knowledge (Stern, Dubeck & Dick 2018, p. 67).

v. **'Catch-up' of letter-sound knowledge among weaker learners is not evident.** The weakest learners at the start of grade 1, performing at the 25th percentile have flatter improvement trajectories in the latter grades in letter-sound knowledge compared to learners at the 50th and 75th percentile.

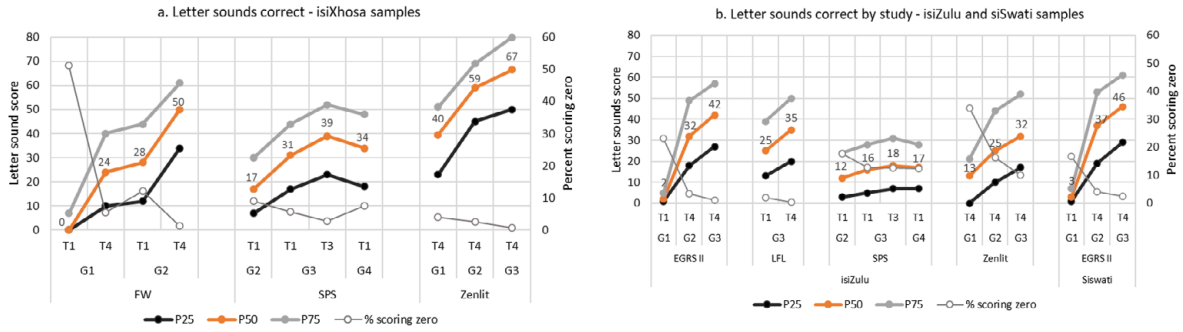
In the Foundation Phase, including the end of grade 3, we also identify that the isiZulu learner samples have less letter-sound knowledge than the isiXhosa learner samples. By the end of grade 2, the median FW learner sounds 50 letters cpm but the median learner in the isiZulu EGRS II sample sounds 32 letters cpm. Within the Zenlit study at the end of grade 3, the median isiXhosa learner sounds 67 letters cpm compared to 32 letters cpm for the median isiZulu learner. Within the SPS study by the end of grade 3, the median isiZulu learner's letter-sound knowledge is lower (18 letters cpm) than the median isiXhosa learner (50 letters cpm). This may relate to different pedagogical approaches to teaching letters in different provinces.

Figure 5: Letter-sound score by language, grade and term (pooled EGRA dataset)



Source: Pooled Nguni EGRA dataset. All letter-sound tasks are timed at 1-minute except for EGRS II, grade 1.

Figure 6: Comparison of letter-sound scores within studies – 25th, 50th and 75th percentiles and percent scoring zero



Notes: We do not disaggregate across control and treatment groups in the case of intervention studies. These are not all longitudinal samples in the strictest sense. The SPS sample includes 2 cohorts. Zenlit is not a panel - rather learners in different grades are observed in the same schools so some gains over grades may be attributed to holding back weaker learners. All letter-sound tasks are timed at 1-minute except for EGRS II, grade 1.

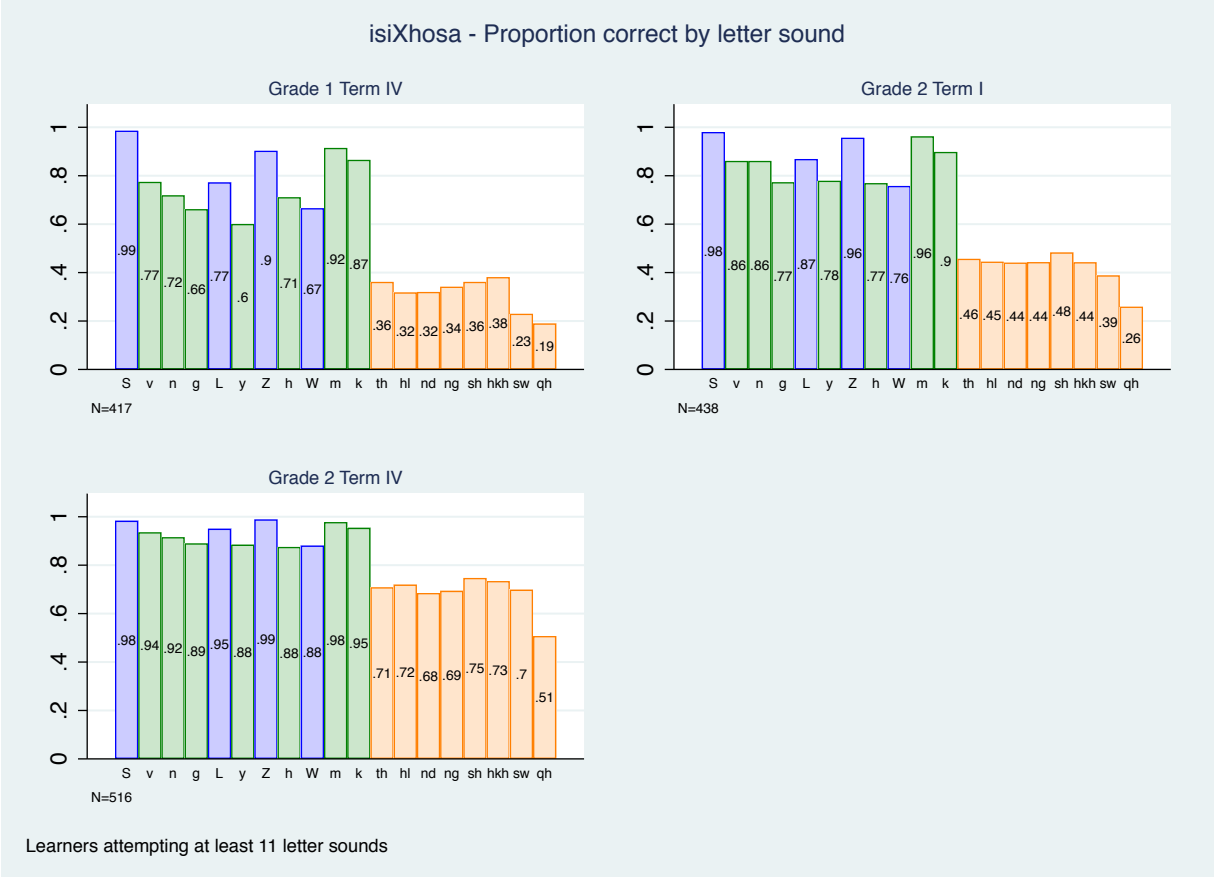
5.2. Complex consonant sequences

Despite weak overall knowledge of letter-sounds, learners find reading single letter-sounds much easier than complex consonant sequences. Figure 7 compares grade 1 and 2 performance on single consonant letter-sounds (upper case blue; lower case green) against knowledge of complex consonant sequences (orange). At the end of grade 1, less than 1 in 3 learners could sound any complex consonant sequences, and by the end of grade 2, still 1 in 4 could not sound any correctly.

The lack of familiarity with complex consonant sequences is also observed in Figure 8 which compares the percentage of learners scoring zero on letter-sounds with those scoring zero on a separate complex consonant sequence assessment. By the end of grade 1, virtually all learners in this isiXhosa sample can sound at least 1 letter correctly, but only half of the learners can sound 1 complex consonant sequence correctly.

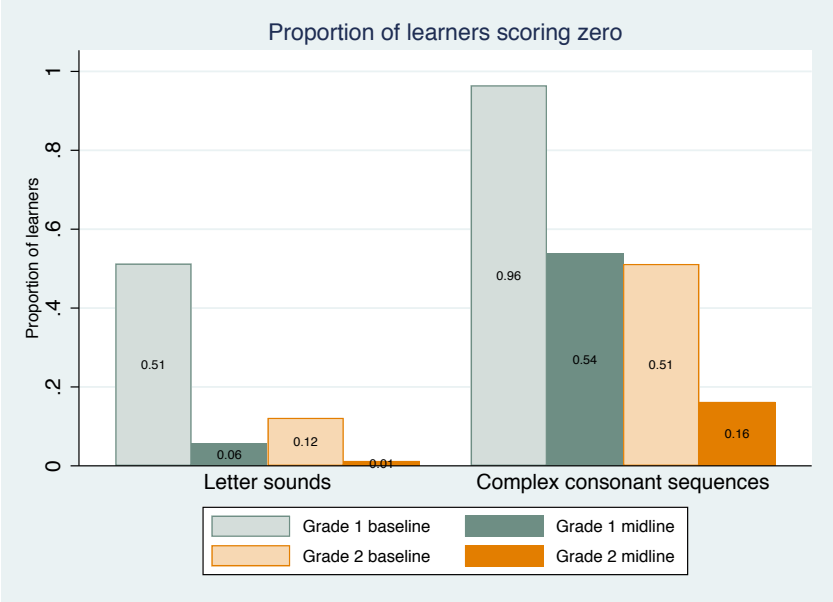
The lack of familiarity with complex consonant sequences is a major inhibitor to reading. Early grade Nguni language texts contain multiple complex consonants even in grade 1 or 2. This is illustrated earlier in Figure 4 which shows that most of the early grade Nguni ORF texts considered in this analysis contain 6 to 11 complex consonants before reaching the 11th word of the passage. Quite simply, it is not possible to even begin reading a passage without knowledge of and automaticity in sounding complex consonant sequences. There is a strong correlation between learners' knowledge of complex consonant sequences and oral reading fluency as illustrated in Figure A 1 in the appendices.

Figure 7: Proportion of single consonants vs complex consonant sequences correctly sounded



Source: Funda Wande evalution, isiXhosa sample. i) grade 1, term 4; ii) grade 2, term 1; iii) grade 2, term 4.

Figure 8: Zero scores in single consonant letter-sounds versus complex consonant sequences

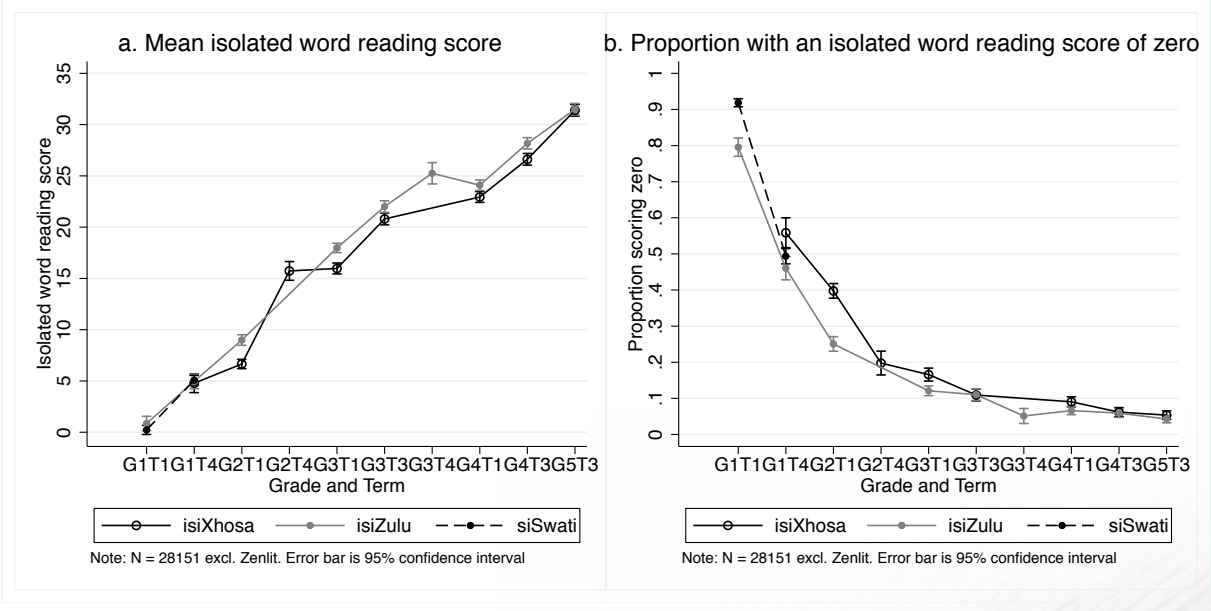


Source: Funda Wande evaluation, isiXhosa sample.

5.3. Word reading

We observe considerable regularity in the acquisition of skills in reading single words across studies and languages. Figure 9a shows learners' estimated mean score (cwpm) in reading words presented out of context (i.e. in a list) at the beginning and end (Term 3 or 4) of each grade. By the end of grade 3 learners can read on average between 21 to 25 cwpm. However, the acquisition of this skill occurs too late in the Foundation Phase. Figure 9b shows that half of the learners (regardless of language) cannot read any isolated words correctly by the end of grade 1. This drops to 20% by the end of grade 2 and 5% to 10% by the end of grade 3. Although the proportion of non-readers declines by grade, as many as 1 of 10 learners in the pooled sample leave the Foundation Phase without being able to read a single word.

Figure 9: Isolated (single) word reading by language, grade and term (pooled EGRA dataset)



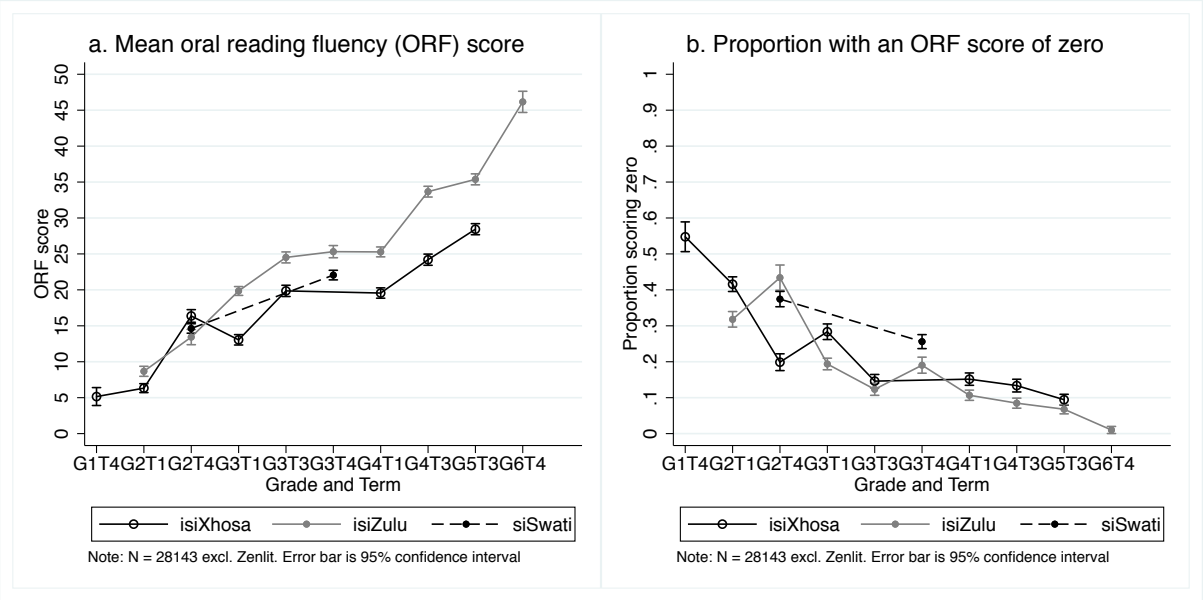
Source: Pooled Nguni EGRA dataset.

5.4. Text reading (Oral reading fluency)

Figure 10a shows learners' oral reading fluency (ORF) mean scores when reading a passage at the beginning and end (Term 3 or 4) of each grade. By the end of grade 3 learners read on average 19 to 25 cwpm depending on the language in question (19 cwpm for isiXhosa, 22 cwpm for siSwati and 25 cwpm for isiZulu).

Figure 10b shows that in this sample around half of the learners in the urban isiXhosa sample cannot read a single word correctly in a grade level text by the end of grade 1. In our samples, around 20% of the isiXhosa learners, 43% of the isiZulu learners and 37% of the siSwati learners still cannot read a word correctly from a passage by the end of grade 2. While ORF begins to emerge in samples in grade 3, far too many learners remain non-readers at the end of grade 3. By the end of the Foundation Phase 15% of isiXhosa learners, 19% of isiZulu learners and 26% of siSwati learners in these samples are unable to read a single word from a passage.

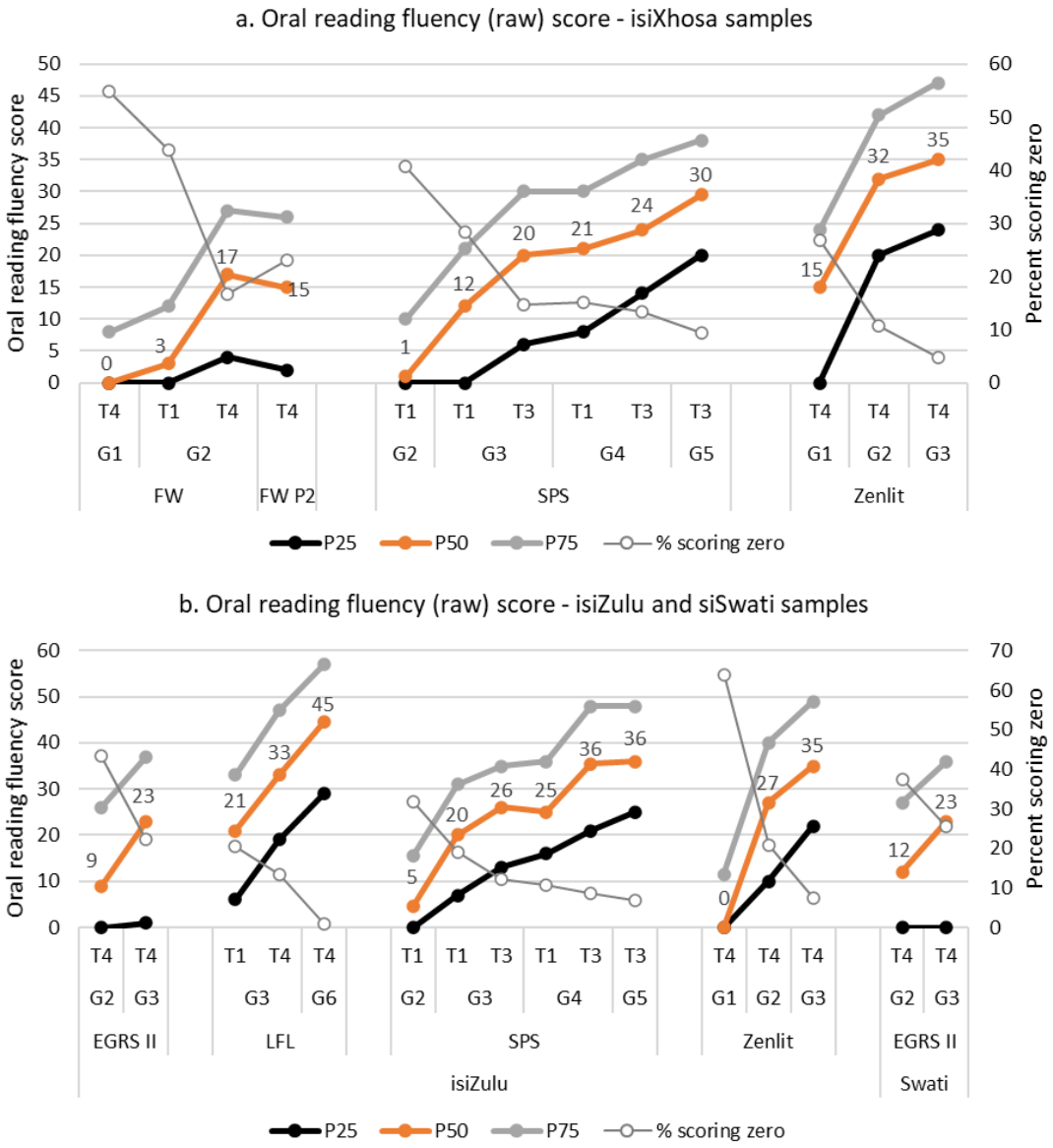
Figure 10: Oral reading fluency* score by language, grade, and term (pooled EGRA dataset)



Notes: The raw score is shown which reflects percentage of words that students get correct within a minute but not adjusted by the time remaining if they finish before 1-minute because 1 study did not record the time 'remaining'.

Despite the very poor reading results overall, some progress is made over the grades. With the availability of ORF scores in isiZulu at end of grade 6, the continued development of fluency in Nguni language reading is observed even beyond grade 4 or 5. Yet what is evident when looking within studies, is that the speed of improvement in reading text is most pronounced in the Foundation Phase, particularly grades 2 to 3 (as seen in Figure 11a and 11b). After the Foundation Phase, gains are more muted (reflected in the SPS sample, for example). Learners need to be instructed effectively in reading in the Foundation Phase. If they fall behind at any point they are at a serious developmental disadvantage. What is concerning is how little progress is made in ORF for learners at the 25th percentile – trends are flatter (across the grades) for these weaker readers than for learners at the 50th or 75th percentile. (In the EGRS II samples, learners at the 25th percentile virtually make no progress at all from grades 2 to 3).

Figure 11: Comparison of oral reading fluency* scores within studies – 25th, 50th and 75th percentiles and percent scoring zero



Notes: The raw score is shown which reflects the percentage of words that students get correct within a minute but not adjusted by the remaining time if they finish before 1-minute because 1 study did not record the time 'remaining'. We do not disaggregate across control and treatment groups in the case of intervention studies. These are not all longitudinal samples in the strictest sense. The SPS sample includes 2 cohorts. Zenlit is not a panel - rather learners in different grades are observed in the same schools at the same point in time. LFL grade 3 learners are a panel, but the grade 6 sample is a distinct group of learners. FW P2 reflects grade 2 FW performance on a second ORF passage.

5.5. Comprehension

As discussed in section 4.2, the analysis of oral reading comprehension is restricted to those samples where learners were given up to 3-minutes to complete reading the passage on which the comprehension questions were based. The first column of Table 6 shows the percentage of learners, in each of these studies, who completed the passage and therefore attempted all the comprehension questions. From grade 3 onwards, at least two thirds of learners in our data samples can complete the passage in the extended 3-minute time limit. Amongst these learners, comprehension scores range from 48% to 77%. There is no clear progression in scores across grades. This is to be expected as the ORF text passages and questions differed across samples.

Table 6: Oral reading comprehension score for learners attempting all questions

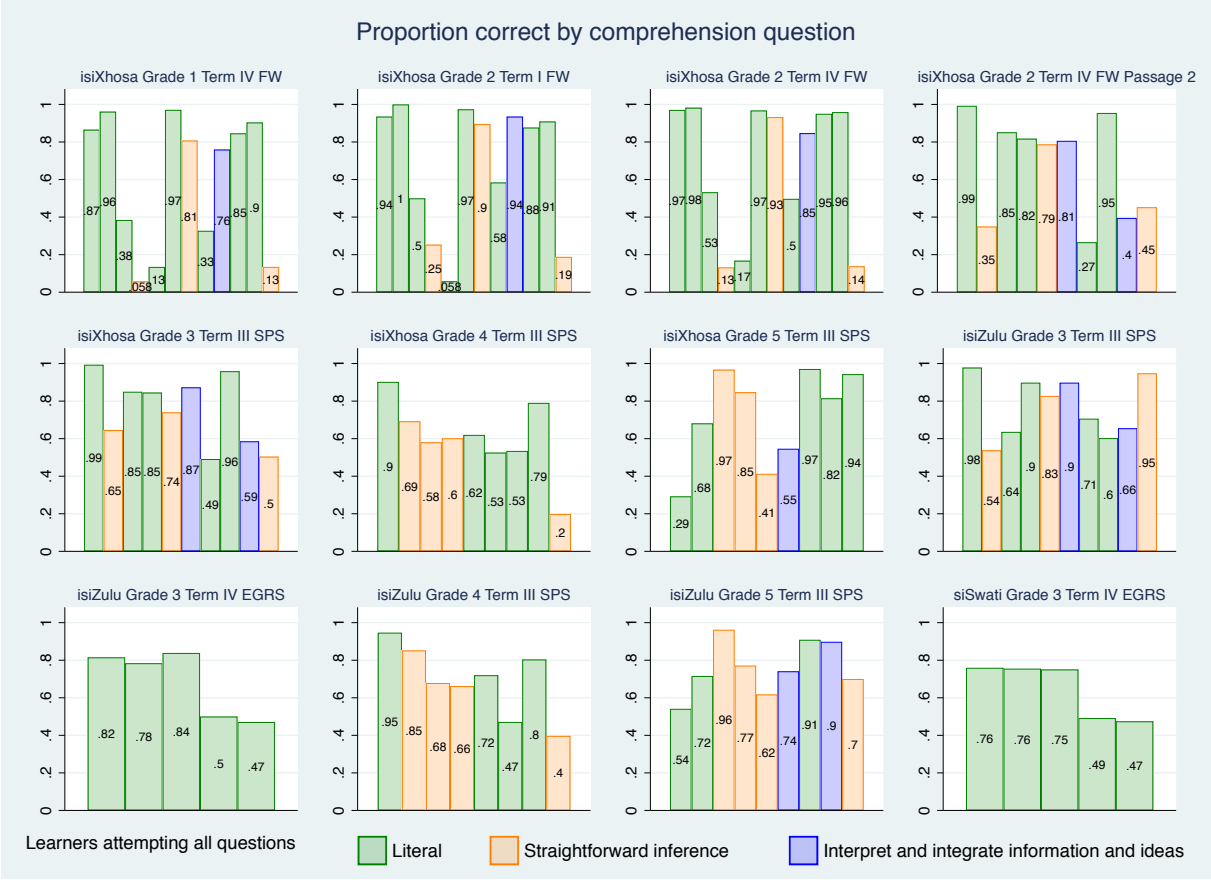
Language	Grade Term	Learners attempting all questions (%)	Comprehension score for learners attempting all questions (%)
isiXhosa	G1T4	17%	59%
	G2T1	26%	68%
	G2T4	57%	67%
	G2T4	47%	48%
	G3T3	57%	75%
	G4T3	68%	61%
	G5T3	72%	65%
isiZulu	G3T3	66%	77%
	G3T4	67%	68%
	G4T3	78%	69%
	G5T3	73%	69%
siSwati	G3T4	70%	65%

The variation in comprehension scores in Table 6 also illustrates the complexity in measuring proficiency in comprehension. This is further demonstrated by examining the scores on individual comprehension questions, which is shown for the various samples in Figure 12[37]. The bar colours indicate the comprehension process engaged in answering the question: literal (green), straightforward inference (orange) or interpret and integrate ideas and information (blue). There is no clear relationship between the hierarchy of the comprehension process and the difficulty of questions. Similar to PIRLS Literacy, there is not an exact mapping between question difficulty and the hierarchy of the comprehension process.[38] There is considerable variation in difficulty (as shown by differences in average scores) within literal comprehension questions, within straightforward inference questions and within the more challenging comprehension questions which require learners to interpret or integrate information. Again, this points to the challenges in anchoring comprehension questions to a required level.

37 Again, this analysis is restricted to learners who attempted all the questions.

38 PIRLS classifies question difficulty according to the lowest benchmark (low international, intermediate, high, advanced) at which a specified minimum percentage of learners answer questions correctly. For example, a multiple-choice question is classified as low international benchmark if at least 65% of learners with scores between 390 and 410 answer the question correctly. While most literal questions are anchored at the low international benchmark, there are literal questions at every level of difficulty. Within the other comprehension processes, questions also range in difficulty level.

Figure 12: Proportion of learners answering each comprehension question correctly



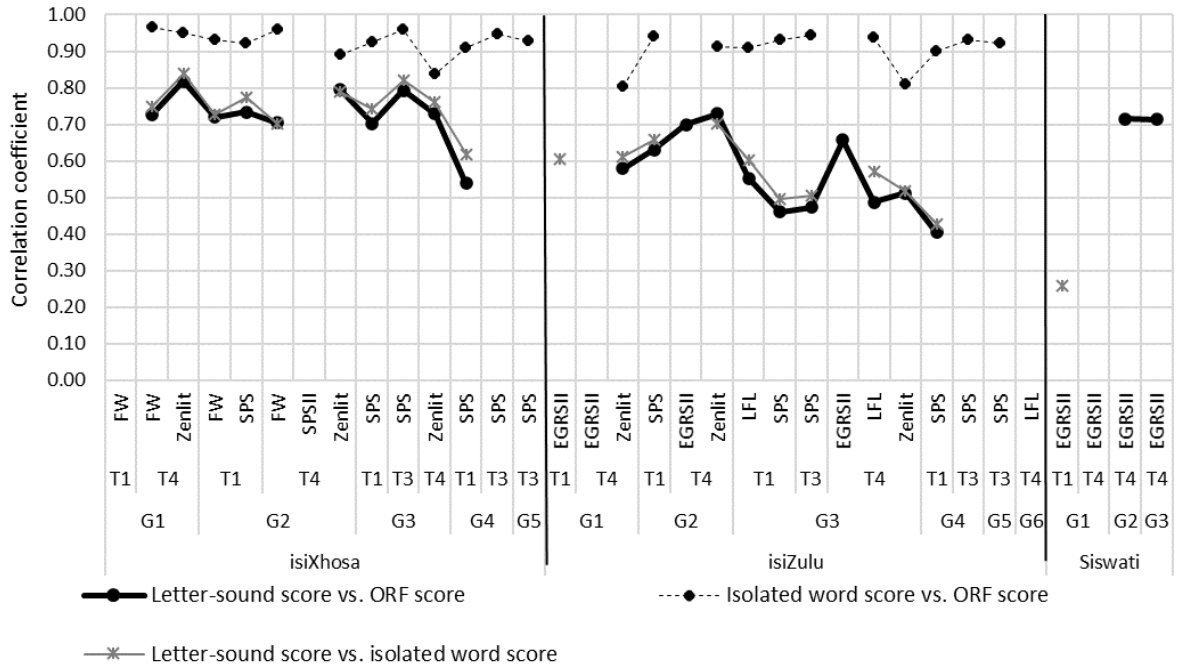
5.6. Relationships between code-based factors

Skills across various code-based factors (letter-sound knowledge, reading isolated words or words in a passage) are highly interrelated. Figure 13 graphically plots correlation coefficients across learner performance in letter-sound knowledge, wording reading and ORF by grade and term.

Isolated or single word reading and ORF follow a very close relationship (almost 1 to 1) across all grades and languages. This implies that setting both isolated word reading and ORF benchmarks is unnecessary. For this reason, in this analysis, thresholds and benchmarks are not set for isolated word reading.

The correlation between letter-sound knowledge and reading words in a passage (or as single, isolated words), starts off as very strong but declines at higher grades. The decline does not mean letter-sound knowledge becomes less important. Rather it suggests that, as decoding skills improve (at higher grades), other skills become more important for fluency when reading from a connected text. Further, there is reduced variation in letter-sound knowledge at higher grades as mastery increases. This may limit detectable correlations.

Figure 13: Correlation coefficients across code-based factors



6. Benchmarking

In this section, we present the analysis to establish ORF thresholds and benchmarks. This benchmarking analysis also provides more in-depth insight into letter-sound knowledge, ORF and comprehension skills as well as the interrelationships between them.

6.1. Establishing fluency thresholds

Our analysis process follows the developmental cline of reading (see Figure 2 in section 3). The relationship between speed and comprehension is moderated by accuracy as reading errors reduce speed and clutter working memory. Accuracy needs to develop before reading speed can be increased, and automaticity can be developed. Thus, we begin with a focus on accuracy.

6.1.1. Relationship between speed and accuracy

The relationship between speed and comprehension is moderated by accuracy, with errors both reducing speed and cluttering working memory. The accuracy-speed relationship is likely to differ depending on the characteristics of the language. Therefore, we investigate the relationship between oral reading speed and accuracy for each of the Nguni languages in Figures 14 to 16. The relationship is displayed using locally weighted polynomial regressions. Reading speed is measured by the number of words attempted (in the time limit) from a passage. Accuracy is measured by the percentage of those words correctly attempted. The figures show the average accuracy associated with each level of speed. The figures include grey dashed vertical reference lines at 20 and 35 words per minute and a red horizontal line representing accuracy of 95% (i.e. for every 100 words attempted, the learner gets 95 correct).

Across all languages, grades and reading passages, there is a remarkably consistent pattern: initially accuracy and speed increase steeply together, and then the relationship tends to flatten off. This flattening occurs when accuracy levels reach around 95%. For example, as shown in Figure 14, at the beginning of grade 3 isiXhosa learners who are attempting around 10 words per minute are getting every third word incorrect, on average. Accuracy and speed improve steeply and grade 3 isiXhosa learners who reach 95% accuracy are typically reading at a speed of around 26 words per minute. Thereafter, we see little change in accuracy as speed increases further, suggesting that an accuracy threshold has been reached. Across all the samples, the speed at which average accuracy reaches 95% ranges between 22 and 34 words per minute. The same patterns are observed for all 3 languages: isiXhosa, isiZulu and siSwati.

Figure 14: Reading speed and accuracy (isiXhosa)

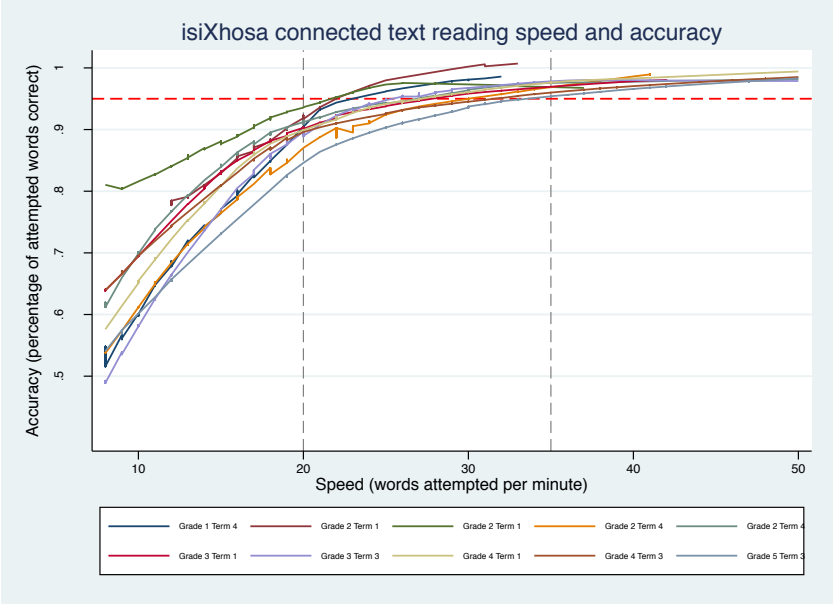


Figure 15: Reading speed and accuracy (isiZulu)

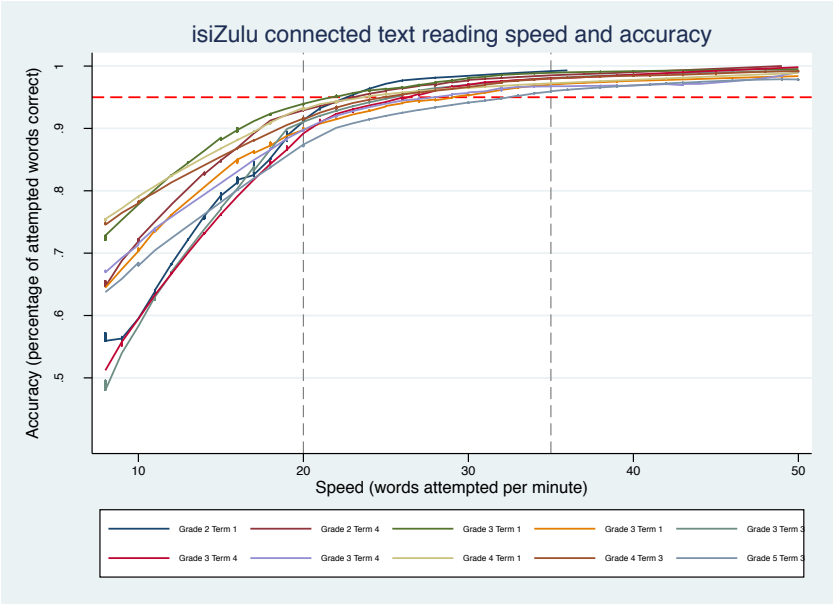
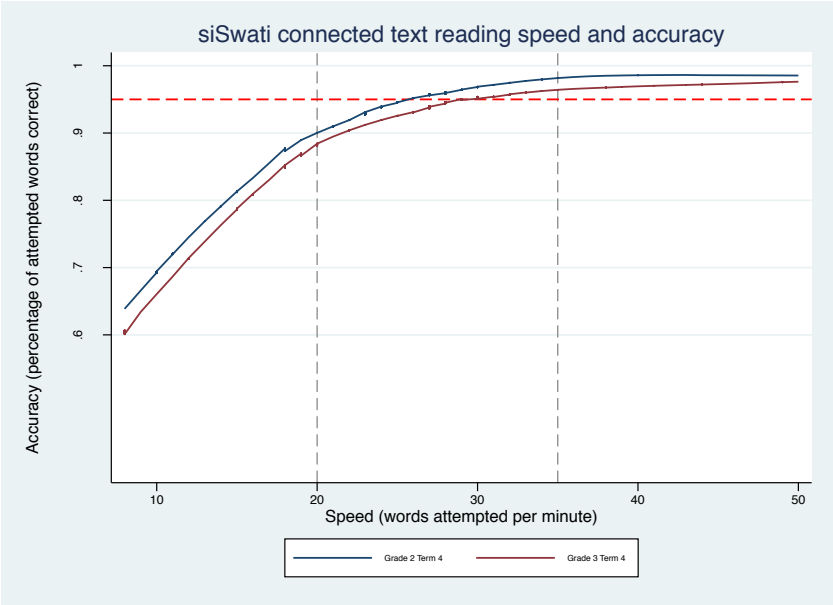


Figure 16: Reading speed and accuracy (siSwati)



Next, we investigate the distribution of reading speed among accurate readers (those learners who could read with at least 95% accuracy). This is seen in Figure 17 showing box and whisker plots of the distribution of words attempted for the sub-sample of learners who read with at least 95% accuracy. For each sample, the lower and upper edge of the box represent the 25th and 75th percentile of the distribution respectively, while the median is indicated by the horizontal line within the box. The figure includes grey dashed reference lines at 20 and 35 words per minute. Amongst accurate readers, learners in grade 1 and the beginning of grade 2 tend to be considerably slower than those in higher grades. From grade 3 upwards, the 25th percentile of speed tends to be above 20 words attempted per minute. This means that there are very few accurate readers who read slower than 20 words per minute who are reaching accuracy levels of 95% or higher.

Figure 18 shows the analogous figure for learners that do not achieve 95% accuracy. Throughout the Foundation Phase, there are almost no readers with poor accuracy who are managing to read at speeds over 20 words per minute. In the Intermediate Phase, inaccurate readers are almost all reading below 35 words per minute.

Learners reading slower than 20 words per minute, have not yet reached accuracy levels to support automaticity and would benefit from instruction focussed on improving their decoding skills and fluency. Until they reach this threshold, it is likely that the development of higher order skills, including comprehension, will stagnate.

Figure 17: Speed distribution for learners reading with at least 95% accuracy

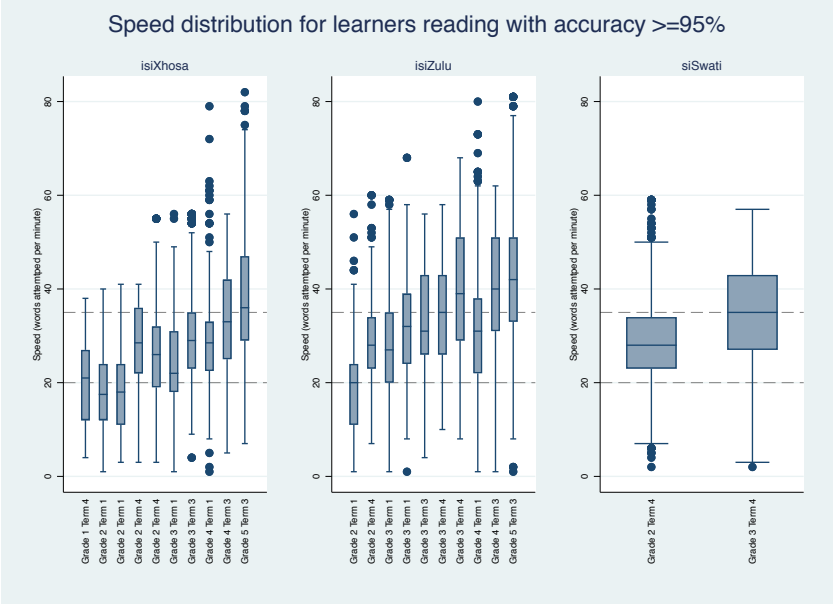
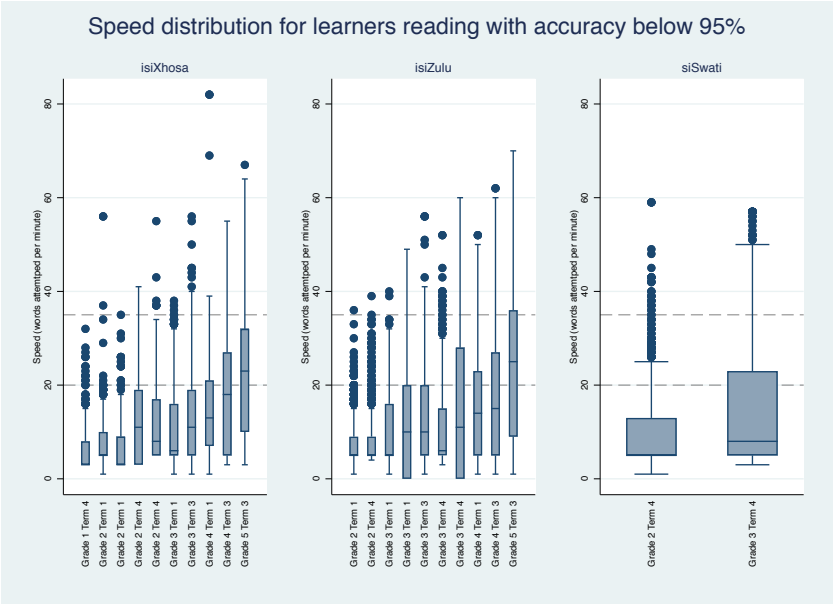


Figure 18: Speed distribution for learners reading with less than 95% accuracy



6.1.2. Relationship between fluency and comprehension

We now examine the relationship between oral reading fluency (ORF) and oral reading comprehension. As discussed previously, ORF is measured by the number of correct words per minute (cwpm) read from a passage of connected text.

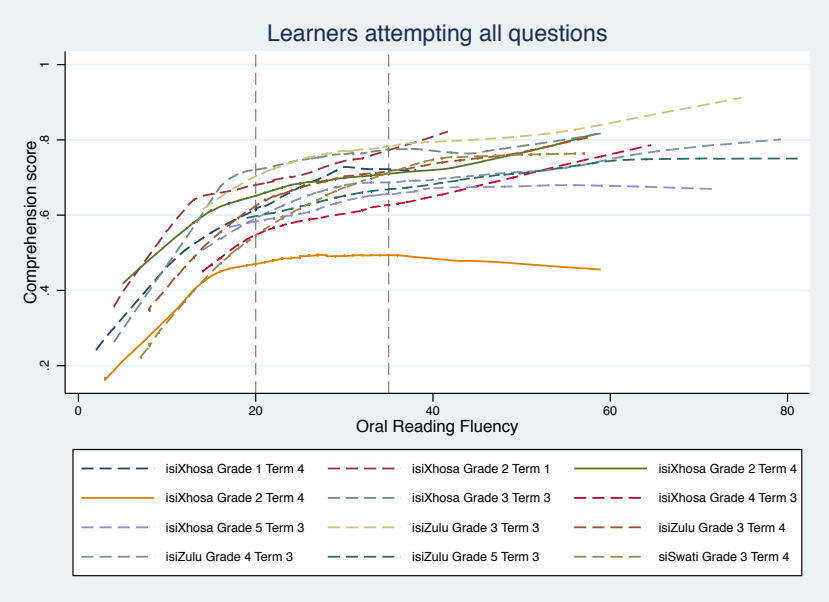
In EGRA, learners are only asked comprehension questions relating to the parts of the passage that they have read. This is problematic in samples where a large portion of learners do not finish the passage within the 60 second time limit. This introduces a mechanistic relationship between fluency and comprehension in such samples. To overcome this challenge, we focus only on those studies that allowed learners an additional 2-minutes to complete the passage. We restrict the sample to learners who read far enough in the extended time to attempt all comprehension questions.

Figure 19 plots the average comprehension score at each level of fluency separately for 1 siSwati, 7 isiXhosa and 4 isiZulu samples. Although there are differences in the average comprehension level between samples, the fluency-comprehension gradient is similar across studies. It is initially very steep and then flattens out. The figure includes reference lines at 20 and 35 cwpm. Below 20 cwpm, the gradient is very steep[39]. Learners reading below 20 cwpm tend to have poor comprehension. Across samples, in this low accuracy and low fluency zone, learners would benefit from instruction that improves their decoding skills. Fluency of below 20 cwpm appears to be a threshold below which comprehension skills are unlikely to develop.

Learners reading between 20 and 35 cwpm have reached an accuracy threshold. Here increasing speed and automaticity is associated with improvements in comprehension. Above 35 words per minute, the comprehension-fluency gradient tends to flatten with diminishing returns to fluency. But this flattening occurs at fairly low comprehension levels (between 60% and 80% of comprehension questions correct in most samples) suggesting that underdeveloped comprehension skills become the key hurdle for learners at these fluency levels. In this fluency zone, the instructional focus and support should emphasise strengthening reading comprehension skills through vocabulary development and deeper engagement with text.

The variability in the level of comprehension scores between samples highlights the challenge with ensuring equivalence of difficulty across sets of comprehension questions. In one of the studies, the same learners were assessed on 2 different passages with accompanying comprehension questions. These learners are represented by the solid orange and dashed red line in Figure 19. While there is a very high correlation between the oral reading fluency scores on both passages (0.94), the comprehension scores differ dramatically.

Figure 19: Relationship between oral reading fluency and comprehension for learners attempting all questions



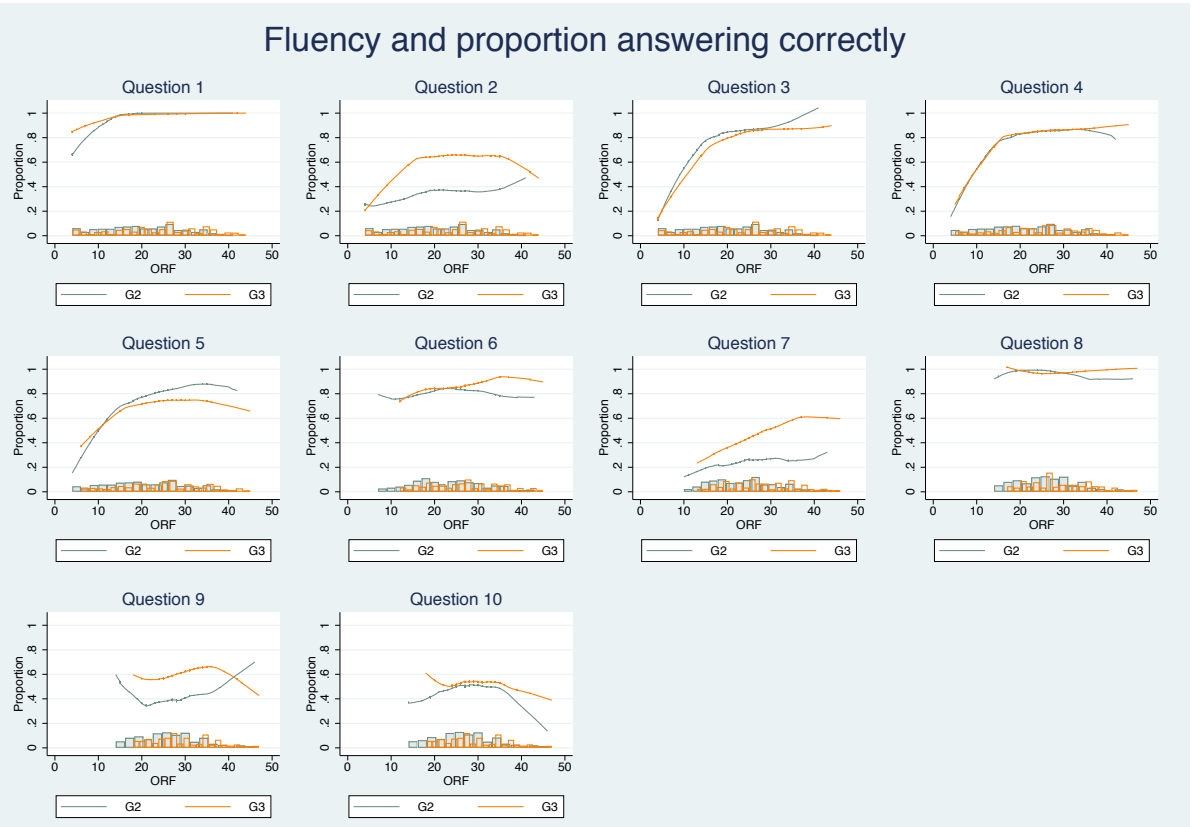
We explore this further by examining the relationship between fluency and individual comprehension questions in cases where the same story was used as an ORF passage in more than one study. Figure 20 and Figure 21 compare ORF and reading comprehension scores (at the question level) between studies which use the same story. The histogram bars reflect each grade sample's distribution of ORF scores. The lines are locally weighted polynomial regressions of ORF against the proportion getting the comprehension question correct.

39 Reference lines are shown in the figures at 20 and 35 cwpm.

This analysis reveals considerable variation in the proportion of learners answering comprehension questions correctly (and by implication in the difficulty level of questions) and a concave fluency-comprehension gradient. Notwithstanding the substantial differences in comprehension difficulty, the fluency-comprehension gradient is remarkably consistent and aligns with the notion of thresholds in the developmental cline of reading. There are clear regular patterns in the data across these Nguni languages, grades and reading passages to support the identification of:

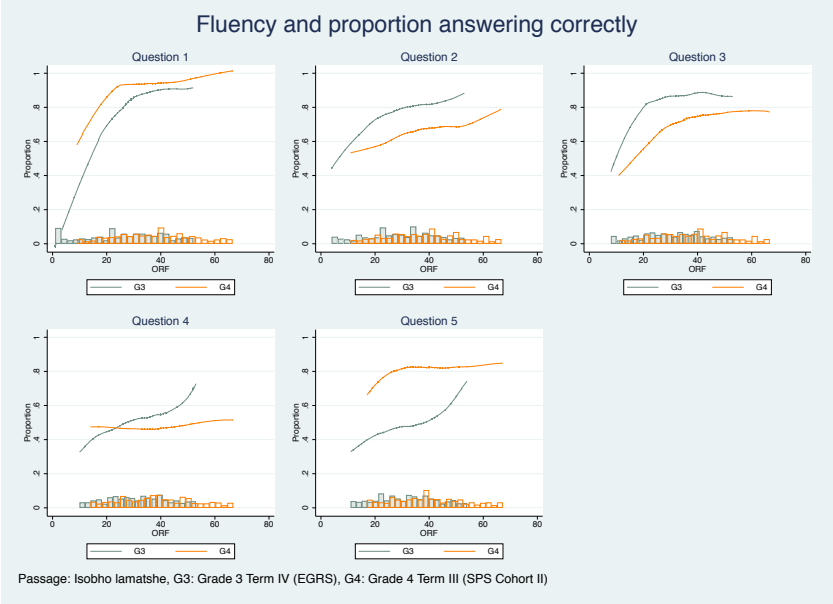
- a lower threshold at around 20 cwpm below which teaching should focus on improving decoding skills; and
- a higher benchmark at around 35 cwpm above which teachers' attention should focus on the strengthening of reading comprehension skills.

Figure 20: Relationship between fluency and individual comprehension questions - SPS and FW



Notes: The histogram bars reflect each grade sample's distribution of ORF scores. The lines are locally weighted polynomial regressions of ORF against the proportion getting the comprehension question correct.

Figure 21: Relationship between fluency and individual comprehension questions – EGRS and SPS



Notes: The histogram bars reflect each grade sample's distribution of ORF scores. The lines are locally weighted polynomial regressions of ORF against the proportion getting the comprehension question correct.

6.1.3. Fluency thresholds and learner profiles

Figure 22 and Figure 23 show the percentage of learners falling into the following fluency categories:

i. unable to read 1 word (shown in blue),
ii. reading less than 20 cwpm (shown in red),
iii. reading 20 to 34 cwpm (shown in green) and
iv. reading at least 35 cwpm (in yellow).

Within each language, the rows are organised by grade and term. Although there are differences between studies, the general progression is clear. In grade 1 and 2:

- between 23% and 55% of learners are unable to read a single word correctly;
- most learners are reading at rates below the lower threshold of 20 cwpm; and
- very few learners reach the upper threshold of 35 cwpm.

By the end of grade 3, most learners (53% to 76%, depending on the sample) have reached the lower threshold (20 cwpm) but only a quarter have reached the benchmark (35 cwpm)[40]. The results suggest the thresholds are set at a level that are practically useful: they are not set so high as to be completely out of reach for current Foundation Phase learners, while at the same time they are high enough to steer the improvement of reading skills over time to a level more appropriate for the demands of the curriculum. These thresholds should be viewed not as aspirational goals but rather indicate the minimum level where every learner at a particular stage in their schooling should be.

40 This varies from 15% to 47% depending on the individual samples.

Figure 22: Early grade fluency profiles, isiXhosa samples

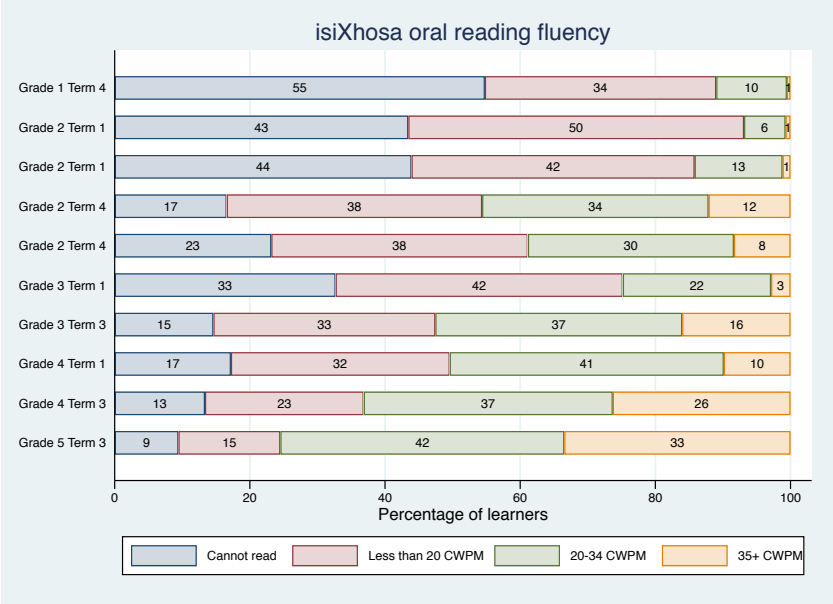
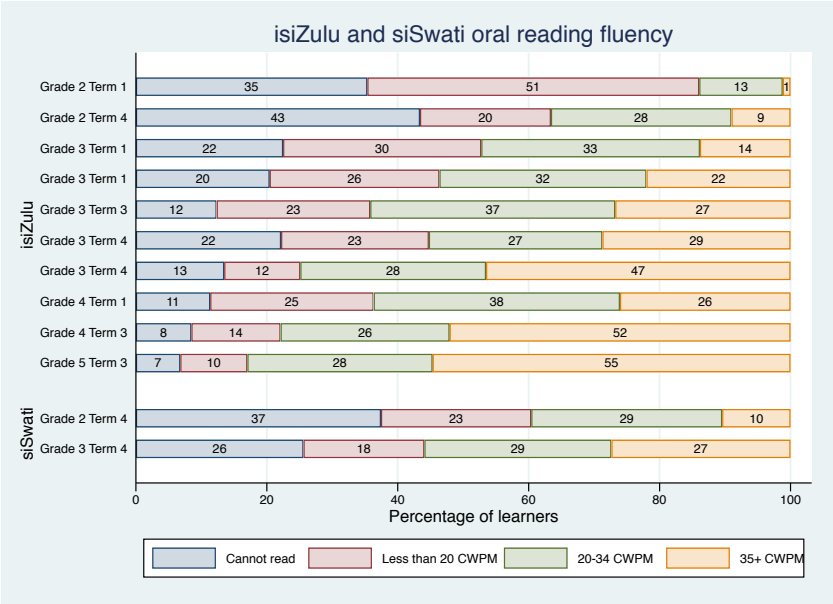


Figure 23: Early grade fluency profiles, isiZulu and siSwati samples



Next, we investigate whether our proposed thresholds correspond to meaningful and distinguishable zones along the reading development cline (see Figure 2). In Table 7, we combine all the samples within a language and summarise accuracy, comprehension and letter-sound knowledge by the same 4 fluency categories.

The table shows that learners below the lower threshold (less than 20 cwpm):

- have low levels of accuracy and very poor comprehension scores;
- tend to perform poorly on letter-sound knowledge; and
- would benefit from instruction on letter-sound knowledge and on improving their word decoding.

Among the group of learners who meet the lower threshold, but not the upper threshold (reading between 20-34 cwpm)

- levels of accuracy have improved with over 70% of these learners achieving at least 95% accuracy;
- comprehension skills are developing but remain poor; and
- accuracy has improved (over 70% of these learners achieving at least 95% accuracy).

Finally, the group of learners who meet the benchmark of 35 cwpm (upper threshold):

- are accurate readers;
- with comprehension scores in the range of 59% to 74%; and
- would benefit from instruction that focuses on improving their comprehension skills through increasing vocabulary and critical engagement with text.

Table 7: Learner characteristics by early grade fluency profiles

	isiXhosa	isiZulu	siSwati
Cannot read: ORF = 0			
Mean correct letter-sounds per minute	12.8	10.8	18.3
% unable to sound 1 letter	12%	20%	10%
ORF = 1 to 19 cwpm			
Mean correct letter-sounds per minute	31.1	20	39.5
% with at least 95% accuracy	19%	25%	19%
Comprehension (% of total correct)	21%	18%	21%
Comprehension (% of attempted correct)	47%	51%	32%
ORF = 20 to 34 cwpm			
Mean correct letter-sounds per minute	47.8	29	53.2
% with at least 95% accuracy	71%	78%	76%
Comprehension (% of total correct)	46%	46%	53%
Comprehension (% of attempted correct)	65%	73%	62%
ORF = 35+ cwpm			
Mean correct letter-sounds per minute	57.2	36.3	61.8
% with at least 95% accuracy	87%	90%	84%
Comprehension (% of total correct)	59%	62%	74%
Comprehension (% of attempted correct)	73%	78%	74%

6.1.4. Predictive validity of the fluency threshold and benchmark

We now turn to the longitudinal data to investigate the predictive validity of the proposed ORF thresholds (the extent to which meeting the proposed ORF thresholds in the first assessment predicts ORF in the second assessment). Table 8 summarises the data that allow us to follow learners' reading performance over time, indicating the term and grade at the first and second assessments. The time between the first and second assessment varies from 3 to 6 school terms across the various studies. For example, using EGRS II data, we can follow siSwati learners for 12 months and compare the grade 3 outcomes between learners who met and did not meet the proposed ORF thresholds at the end of grade 2.

Table 8: Longitudinal data

Study	Languages	Assessment I	Time between assessments	Assessment II
EGRS II	isiZulu, siSwati	Grade 2 Term 4	4 school terms	Grade 3 Term 4
LFL	isiZulu	Grade 3 Term 1	3 school terms	Grade 3 Term 4
FW	isiXhosa	Grade 1 Term 1	3 school terms	Grade 1 Term 4
FW	isiXhosa	Grade 2 Term 1	3 school terms	Grade 2 Term 4
SPS	isiZulu, isiXhosa	Grade 2 Term 1	6 school terms	Grade 3 Term 3
SPS	isiZulu, isiXhosa	Grade 3 Term 1	6 school terms	Grade 4 Term 3
SPS	isiZulu, isiXhosa	Grade 4 Term 1	6 school terms	Grade 5 Term 3

Figure 24, Figure 25 and Figure 26 below categorise learner fluency at the second assessment for 3 groups of learners:

i. non-readers (0 cwpm) at assessment I (see Figure 24);
ii. those not meeting the lower threshold (1-19 cwpm at assessment I) (see Figure 25); and
iii. those meeting the lower threshold but not the upper benchmark (20 to 34 cwpm at assessment I) (see Figure 26).

Figure 24 shows that many learners (58% to 74%) who were non-readers at the first assessment (i.e. unable to read at all at the first assessment) were still unable to read 1 word by the second assessment, as shown by the blue bars. The percentage of these learners still not reading increases by grade implying that learners who were unable to read a word in higher grades are increasingly selected from a group of learners with chronic reading difficulties. However, a sizeable portion of these non-readers have begun to read slowly by the second assessment (3 to 6 terms later). But most are not yet reaching the lower threshold (of 20 cwpm) at assessment II (green bars). Only a very small percentage are reaching the benchmark of 35 cwpm at the second assessment.

Among learners who were reading below the lower threshold (1-19 cwpm) at the first assessment (shown in Figure 25), most had reached that threshold 3 to 6 school terms later.

Finally, Figure 26 shows that reaching the lower threshold (20 to 24 cwpm) by the first assessment is predictive of reaching the upper benchmark (35 cwpm) by the second assessment. Of those learners who reached the lower threshold by the first assessment, 51% to 78% were reading at least or above the fluency benchmark of 35 cwpm by assessment II. These results suggest that the proposed ORF thresholds are useful as they predict reading performance in later grades.

Figure 24: Reading fluency category at assessment II for non-readers at assessment I

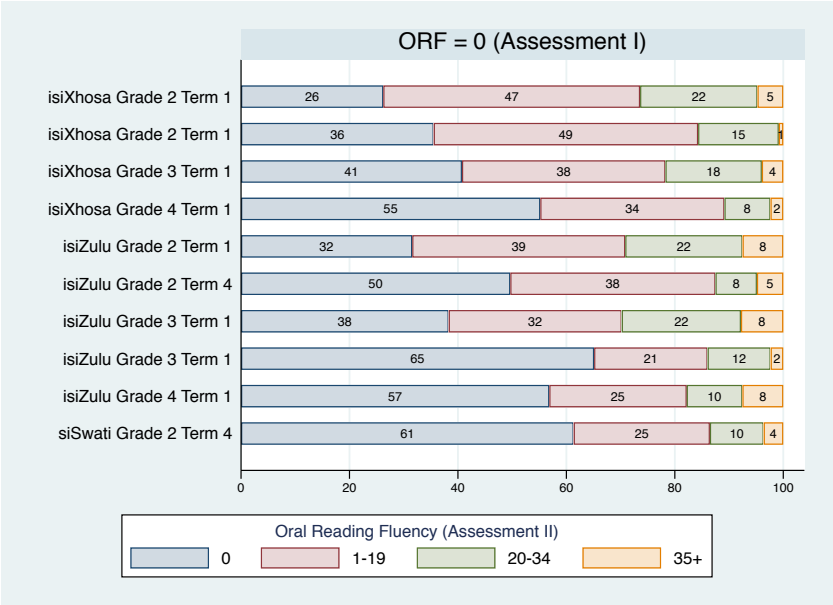


Figure 25: Reading fluency category at assessment II for learners scoring an ORF of 1 to 19 cwpm at assessment I

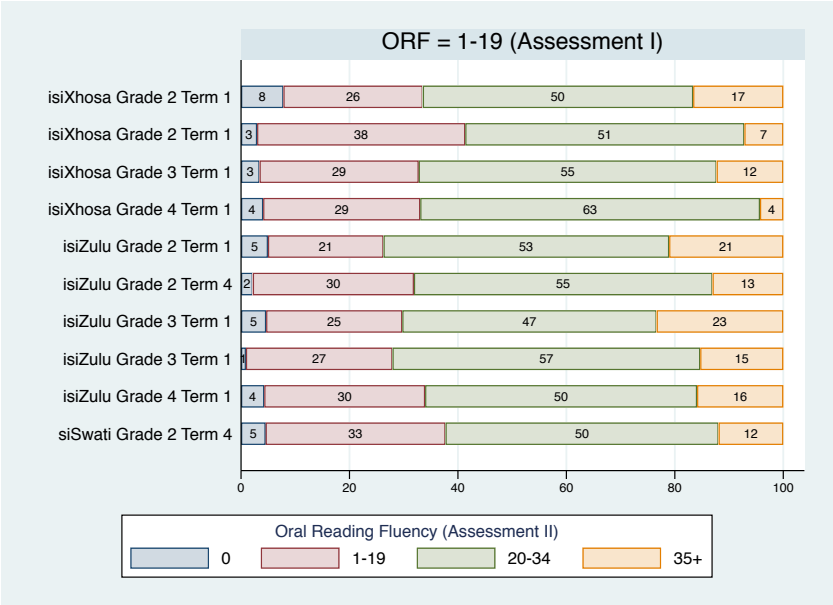
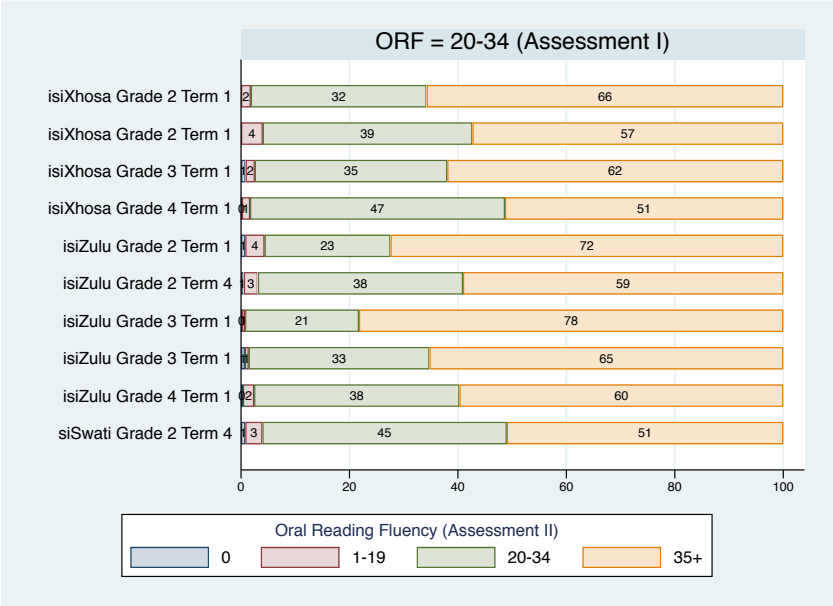


Figure 26: Reading fluency category at assessment II for learners scoring an ORF of 20 to 34 cwpm at assessment I



In addition to oral reading comprehension, 2 of the studies included a written comprehension exercise. We can therefore examine the relationship between oral reading fluency and scores on the written comprehension exercise as an indicator of the concurrent validity of the fluency thresholds. The longitudinal nature of these studies also allows us to investigate the predictive validity of the thresholds, i.e. the relationship between oral reading fluency at the first assessment and performance on the written comprehension 4 to 6 school terms later.

We begin with the predictive validity process using SPS data in Figure 27. The box plot in the left panel shows the ORF distribution at the first assessment by the written comprehension score at the second assessment. ORF was measured at the beginning of grade 4 and written comprehension in the third term of grade 5. The maximum a learner could score on the written comprehension was 3. The dashed grey lines indicate the lower 20 cwpm threshold and the 35 cwpm benchmark. For both isiXhosa and isiZulu, there are very few learners who had not met the lower fluency threshold by grade 5, scoring at least 5 for the written comprehension. Moving to the right panel, we examine the relationship between concurrent fluency and written comprehension. Here, isiXhosa learners scoring at least 5 for the written comprehension tend to be reading around the upper fluency benchmark while most isiZulu learners have met the upper threshold.

Figure 28 shows analogous figures from the EGRS II study where we examine the relationship between previous (end of grade 2) and current (end of grade 3) fluency and written comprehension at the end of grade 3. Results are very similar to those from the SPS study. Learners scoring at least 5 out of 6 for the written comprehension, were meeting the lower fluency threshold in the previous year and have current ORF levels around the benchmark of 35 cwpm.

In summary, the longitudinal data provide support for the predictive validity of the proposed fluency threshold and benchmark.

Figure 27: Relationship between oral reading fluency and written comprehension - SPS

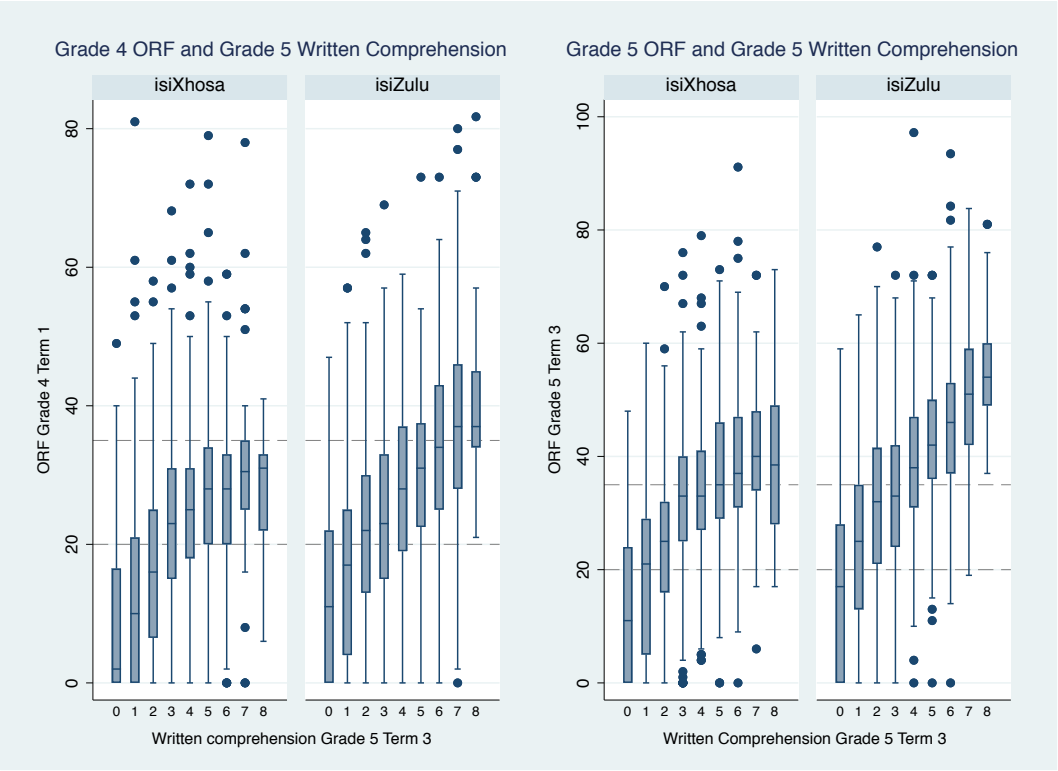
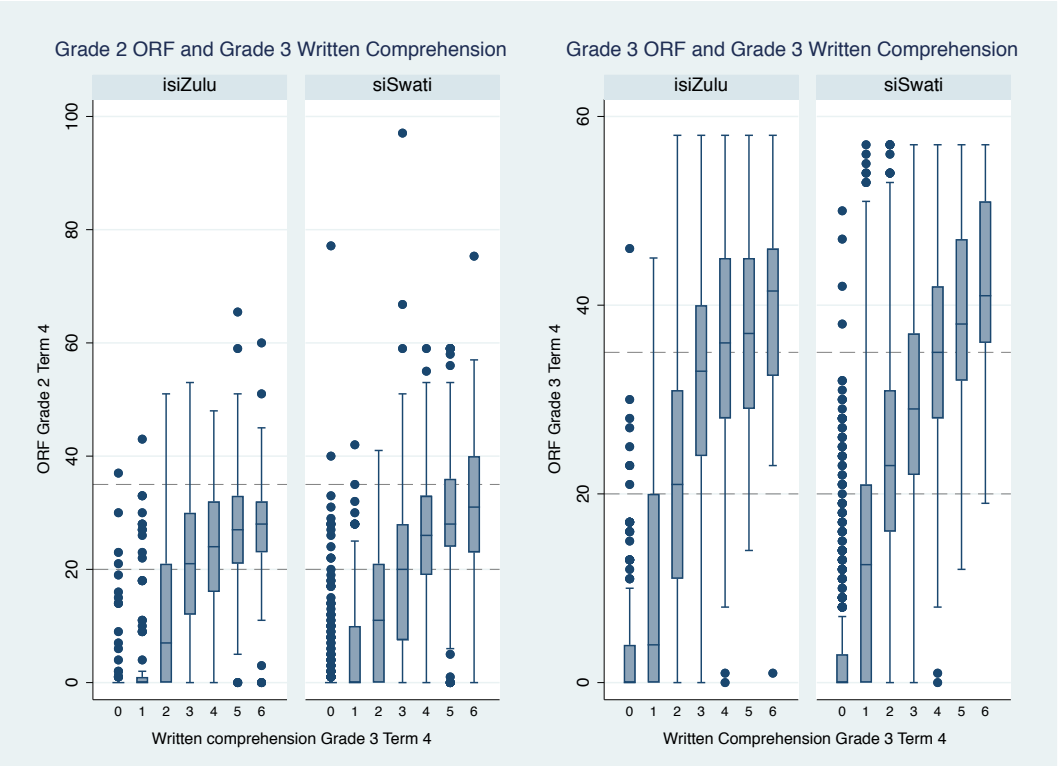


Figure 28: Relationship between oral reading fluency and written comprehension -EGRS



6.2. Establishing a letter-sound benchmark

Having identified a fluency threshold and benchmark, we now work backwards along the development cline to consider what letter-sound benchmark could support the acquisition of fluency skills. In the selection of an appropriate letter-sound benchmark, we are guided by a combination of insights from the data, expert opinion grounded in the theory of reading and taking cognisance of the curriculum requirements.

We begin by documenting the existing levels of letter-sound knowledge across grades and then examine the relationship between speed and accuracy in sounding letters. Thereafter, we turn to the longitudinal data to investigate how this foundational skill develops over time. Having identified a potential letter-sound threshold, we show how this threshold maps to future oral reading fluency.

The distribution of correct letter-sounds per minute is shown separately for each language sample in Figure 29, Figure 30 and Figure 31 grouped by study. We see much greater variation between studies in letter-sound knowledge than for oral reading fluency. This seems to be driven by higher homogeneity between learners within schools on the letter-sound task versus the oral reading task. A possible explanation for this is that differences in pedagogical practice are likely to manifest more clearly in constrained skills that are more responsive to repetition and drill. Looking across the studies, letter-sound knowledge appears to be particularly weak in the rural isiZulu schools of the SPS impact evaluation.

Benchmark setting is a balancing act. Benchmarks for lower order skills need to be high enough to map to the threshold necessary to support the development of higher order skills. At the same time, they need to low enough to be responsive to incremental change. Letter-sound knowledge is a constrained skill and mastery involves sufficient accuracy and speed to facilitate automaticity. We therefore decided to exclude the isiZulu samples from the SPS study in our data analysis.

Figure 29: Correct letter-sounds per minute distribution, isiXhosa samples

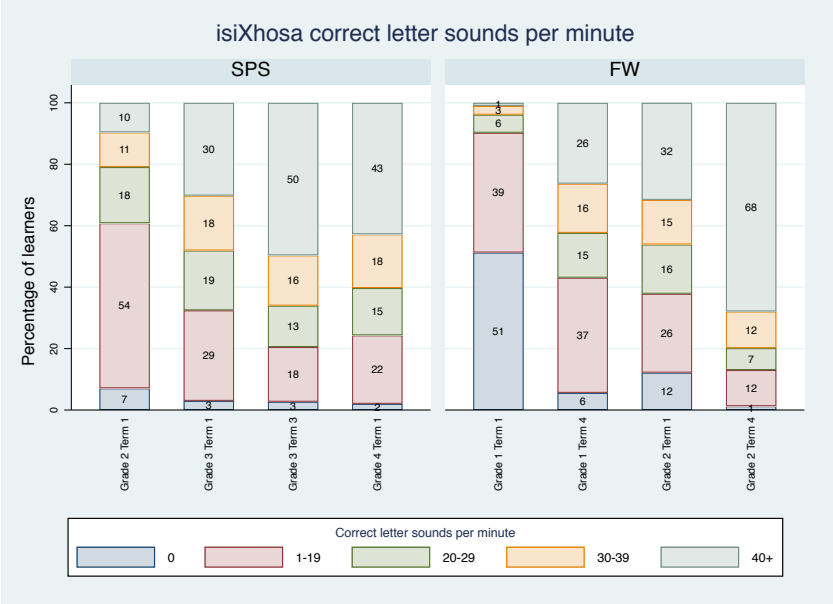


Figure 30: Correct letter-sounds per minute distribution, isiZulu samples

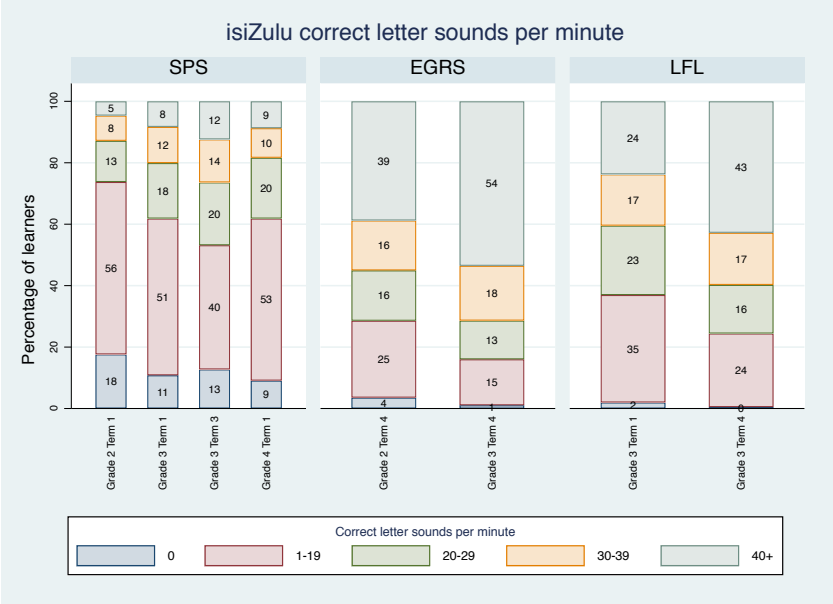
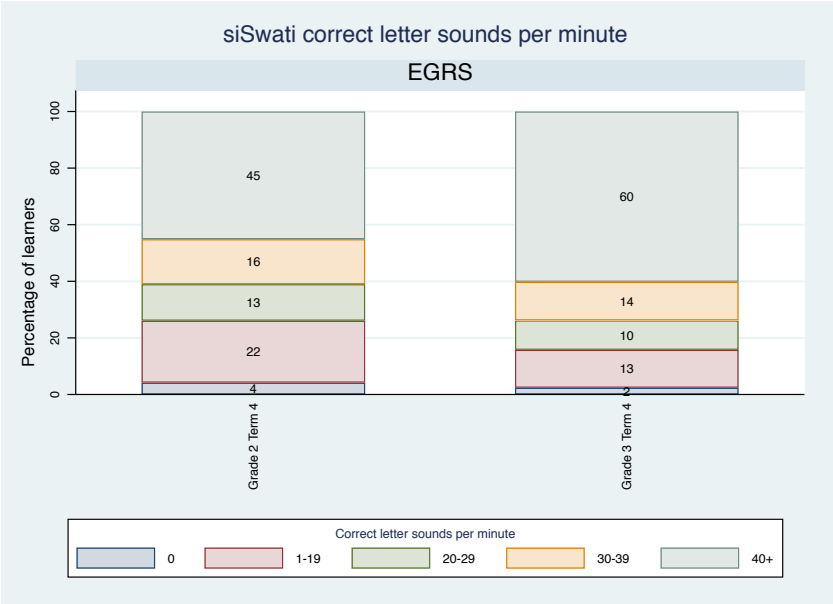


Figure 31: Correct letter-sounds per minute distribution, siSwati samples



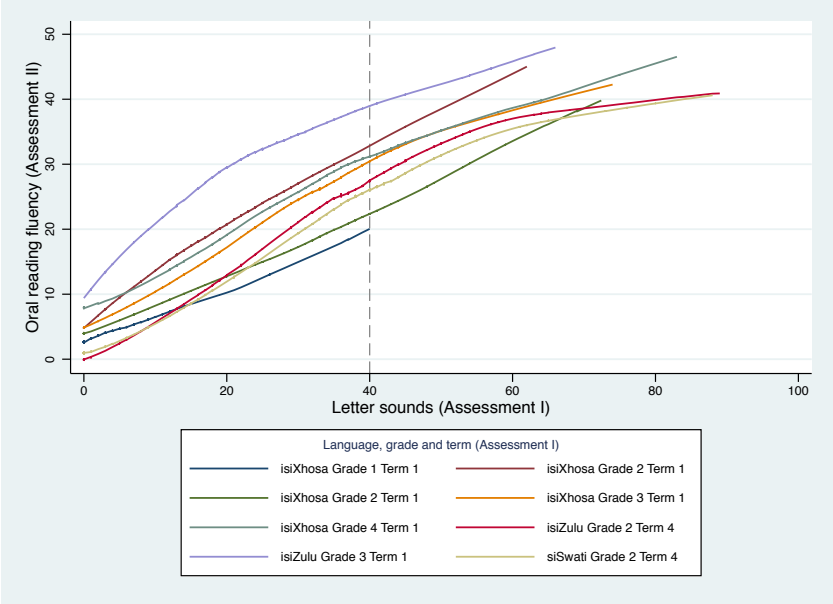
6.2.1. Empirical motivation for a letter-sound benchmark

We first consider whether a letter-sound benchmark is a useful predictor of later fluency skills. In section 5.6, we documented a reasonably high correlation between concurrent oral reading fluency and letters sounded correctly in a minute. We can exploit the longitudinal data to examine the relationship between earlier letter-sound knowledge and later oral reading fluency.

For learners who were assessed twice, Figure 32 shows the relationship between correct letter-sounds per minute at the first assessment and oral reading fluency at the second assessment. Despite differences in the level and slope of the lines, there is a clear pattern: learners with better letter-sound knowledge at the first assessment have higher ORF scores when assessed again 3 to

6 terms later. This suggests that reaching a letter-sound knowledge benchmark in the early grades would be a good indicator of whether a learner is on track to develop ORF in later grades. This provides a solid motivation for developing a letter-sound knowledge benchmark.

Figure 32: Relationship between letter-sound knowledge and later oral reading fluency, all languages



6.2.2. Relationship between speed and accuracy

But what would be an appropriate letter-sound benchmark? In answering this question, we again examine the relationship between speed and accuracy but now in relation to sounding letters. Figures 33 to 35 plot the relationship between the number of letter-sounds attempted (speed) and the percentage of these attempted letter-sounds that the learner reads correctly (accuracy) for each of the Nguni languages. The figures confirm the following patterns with respect to reading letter-sounds:

- Learners with low speed tend to have low accuracy.
- Accuracy improves steadily with speed to a point, but beyond this point there are no further improvements in accuracy.
- The letter-sound speed-accuracy gradient tends to flatten around 40 letter-sounds per minute (indicated by the dashed grey line).

Figure 33: Relationship between letter-sound speed and accuracy, isiXhosa samples

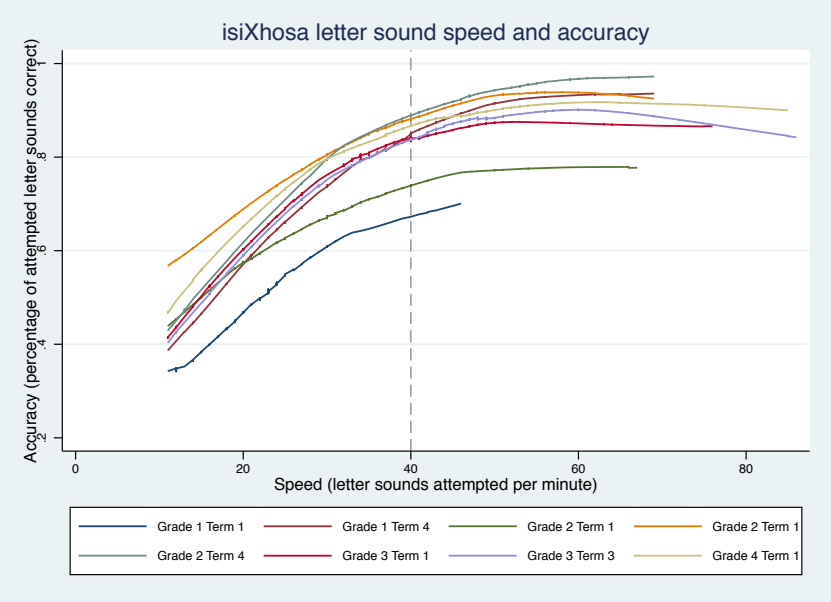


Figure 34: Relationship between letter-sound speed and accuracy, isiZulu samples

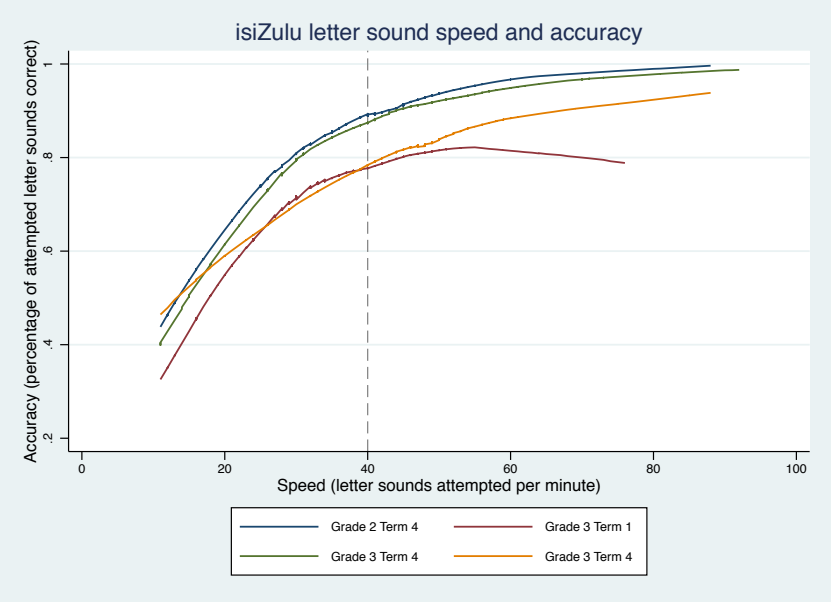
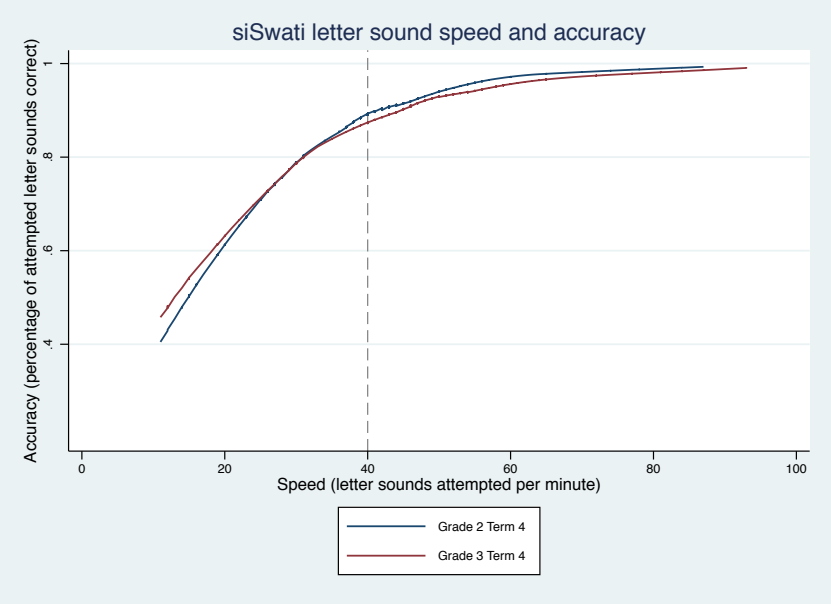


Figure 35: Relationship between letter-sound speed and accuracy, siSwati samples

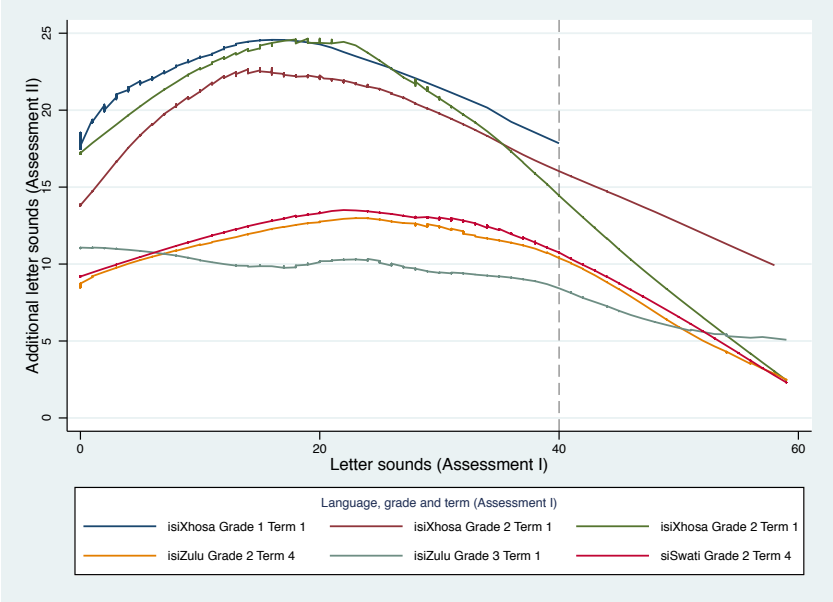


6.2.3. Predictive validity of the proposed letter-sound benchmark

The importance of foundational skills such as letter-sound knowledge tends to diminish as reading proficiency develops. In line with Wang's decoding threshold hypothesis, there is likely a point of letter-sound recognition speed beyond which there are no additional dividends for word decoding. Using the longitudinal data, we examine whether there is a point at which learners' letter-sound speed tends to stagnate. Figure 36 shows the relationship between correct letter-sounds per minute at the first assessment and the improvement (measured in additional letter-sounds per minute) by the next assessment 3 to 6 school terms letter. Across the distribution, improvements in the isiXhosa samples tend to be around 10 letters higher than the isiZulu and siSwati samples. Gains in letter-sounds per minute tend to increase with baseline performance and then begin to decline. For all samples, improvements in letter-sounds per minute are lower for learners who could correctly sound 40 letters per minute at the first assessment than for those who could only correctly sound 20 letters.

The flattening accuracy-speed gradient and the diminishing improvements in letter-sound knowledge over time, validates setting the letter-sound benchmark at around 40 correct letter-sounds per minute.

Figure 36: Development on letter-sound knowledge over time, all languages



We examine how this letter-sound benchmark of 40 cpm maps to future oral reading fluency in Figures 37 to 39. Figure 37 shows fluency at assessment II for learners who met the letter-sound benchmark at the first assessment. Figure 38 shows fluency at assessment II among learners who were correctly sounding out 20 to 39 letters cpm at the first assessment. Figure 39 shows learners who had low or no letter-sound proficiency when first assessed.

- Among learners who met the proposed letter-sound benchmark of 40 at assessment I, most had reached the fluency benchmark of 35 cwpm when they were re-assessed 3 to 6 school terms later. If they missed the fluency benchmark of 35 cwpm, most had at least reached the lower fluency threshold of 20 cwpm.
- In most samples, among learners who were achieving 20 to 39 letter-sounds cpm at the first assessment, most met the lower fluency threshold of 20 cwpm by the second assessment.
- Among learners who had low letter-sound proficiency (0 to 19 letter-sounds cpm) when first assessed, a sizeable portion were unable to read 1 word at the next assessment. Very few reached the fluency benchmark of 35 cwpm when re-assessed.

In summary, the letter-sound benchmark of 40 letters sounded correctly per minute appears to be a good predictor of oral reading fluency in later grades.

Figure 37: Reading fluency at assessment II, for learners who achieved the proposed letter-sound benchmark at assessment I

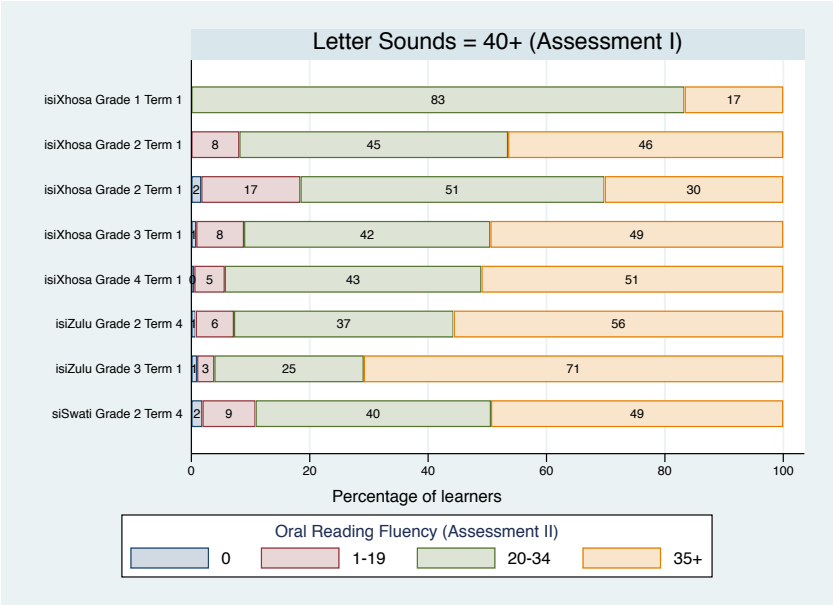


Figure 38: Reading fluency at assessment II, for learners who sounded 20 to 39 letters correctly at assessment I

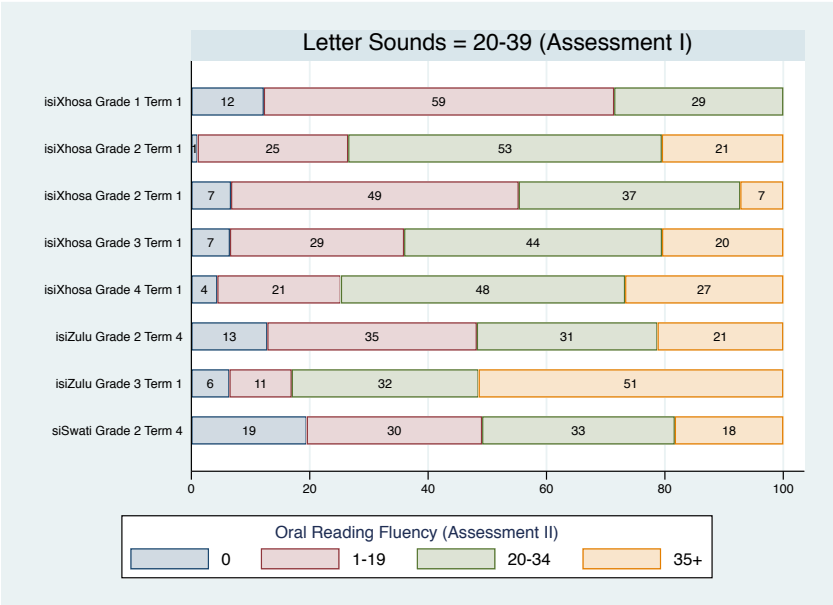
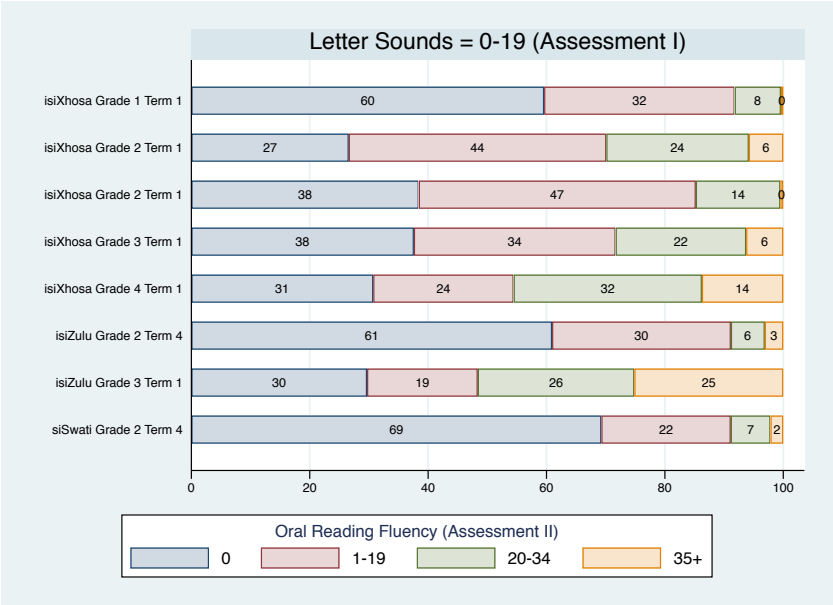


Figure 39: Reading fluency at assessment II, for learners who sounded less than 20 letters correctly at assessment I



7. Summary and recommendations

This research aimed to establish early grade reading benchmarks in Nguni languages. Through an exploratory analysis of available data, positioned within a theoretical framework that outlines the reading development process, the report identifies a letter-sound benchmark, an oral reading fluency (ORF) threshold and an ORF benchmark to guide the teaching of reading in the Foundation Phase. Given the close relationship between learner proficiency in single word reading and ORF in the available data, it was unnecessary to set a word reading benchmark.

As a preliminary step in the analysis, current reading norms in Nguni languages for predominately rural-based learners in no-fee schools were identified from available EGRA data sources. By the end of grade 3, across the three languages, the learner sample read an average of 19 to 25 cwpm from a passage. The proportion of learners in these data samples that are non-readers by the end grade 3 remains unacceptably high (15% to 26% were unable to read 1 word from a passage).

7.1. Summary of proposed benchmarks and thresholds

Based on the exploratory data analysis presented in this report, we propose the following threshold and benchmarks for early reading success in Nguni languages.

- A letter-sounds benchmark of 40 letter-sounds correct per minute (cpm), which should be reached by all learners at the end of grade 1.
- A lower ORF threshold of 20 words read correctly per minute (cwpm) from a connected text. All learners should reach this minimum threshold by the end of grade 2.
- An ORF benchmark of 35 cwpm, which should be attained by all learners by the end of grade 3.

Learners who reach the letter-sounds benchmark of 40 letters cpm are in a much stronger position to read words in and out of context, and decoding instruction can focus on developing fluency. There are few remaining benefits to improving letter-sound knowledge and speed beyond this point. Meeting this letter-sound benchmark is shown to be an early predictor of whether a learner is on track to reach our recommended ORF threshold and benchmark in the later grades.

The ORF threshold of 20 cwpm is a *minimum* threshold which learners must reach so that instructional focus can facilitate the engagement of higher order reading skills. Below this threshold we find little evidence that learners can comprehend what they have read, evidenced by very low oral reading comprehension or written comprehension scores. Attention to the development of fluency and prosody should continue so that reading aloud sounds increasingly natural.

The ORF benchmark of 35 cwpm acts as an upper threshold. Once learners reach this benchmark when reading connected text, attention to prosody continues but the main instructional focus should be on developing reading comprehension strategies and improving vocabulary, to empower learners to engage critically with texts. Learners who reach this benchmark will be able to transition to silent reading more easily.

Currently, by the end of grade 3, most learners (53% to 76%, depending on the sample) have only reached the lower threshold of 20 cwpm. Approximately a quarter of learners have reached the benchmark of 35 cwpm and would benefit from a strong instructional focus on comprehension skills.

In addition to being empirically and theoretically grounded, we believe the suggested threshold and benchmarks are valid for the South African context for 3 main reasons:

i. They are neither too aspirational nor too low. They are not so high as to be completely out of reach for current Foundation Phase learners in poor South African school contexts - a fact that we show in the data. Yet, they establish expectations that are also sufficiently high enough to move teachers, classrooms and schools on a journey of reading excellence, aligning with the presidential goal that all learners should be able to read for meaning by the end of grade 3.
ii. Strong theoretical foundations and predictive validity: These thresholds meet other criteria for a "good" threshold as they are theoretically sound, and they predict reading performance in later grades.
iii. Importantly, as explained in the methodology section, we made no assumptions about the fluency-comprehension relationship in the context of South African classrooms and in the Nguni languages. Rather, we allowed the data to guide us in the selection of appropriate thresholds and benchmarks. This process is independent of the difficulty level of comprehension questions. A weakness of other benchmarking approaches is that, while the analysis is easier to perform, associated results are highly sensitive to the difficulty and placement of comprehension questions in the reading assessment.

7.2. Policy recommendations

The Nguni language benchmarks and threshold proposed in this report provide standards to establish whether learners are on a healthy reading trajectory; guiding teachers to identify learners who are at risk of falling behind in the developmental sequence of reading. In particular, the letter-sound threshold provides a tool for early identification of at-risk learners which can be easily and quickly administered. The proposed benchmarks and threshold could be used to guide classroom assessment of letter-sound knowledge and oral reading fluency.

The very high number of learners entering grade 1 with no letter-sound knowledge suggests that the quality of grade R instruction must be evaluated. From the limited data available, this analysis also found that learners find complex consonant sequences much more difficult than reading and sounding single letter-sounds. Because complex sounds occur frequently in the Nguni languages, knowledge of complex consonant sequences is required by the end of grade 1 to enable learners to keep on track with the proposed ORF threshold and benchmarks by grade 2 and 3. Resolving this would require adaptation of the curriculum which guides Nguni language instruction. Currently the teaching of complex consonant sequences is not included in the grade 1 curriculum.

In closing, we return to the goal of reading: *reading for meaning*. Our EGRA sample data confirms patterns seen in national PIRLS testing of significant deficits in learners' acquisition of comprehension skills. While the teaching of foundational reading skills is fundamental to reading proficiently, the benchmarks and threshold proposed are a necessary but not sufficient condition for reading for meaning. In other words, even the ORF benchmark of 35 cwpm does not represent the point where learners are able to read for meaning on their own. However, at this level of fluency, comprehension is a possibility, if the skills to answer oral and written comprehension are taught. It is evident from our analysis that more attention needs to be given in the Foundation Phase to the teaching of comprehension skills. But the teaching of comprehension skills, and higher order skills may have to be differentiated, depending on where learners are on the developmental sequence of reading. While it may be a challenge to provide learner-specific guidance in Foundation Phase classroom contexts with large class sizes, group-guided reading provides the opportunity for differentiated learner strategies to meet each learner along their reading development path. We supply another instructive signal to support this: for those learners who start approaching the ORF benchmark of 35 cwpm, this is the signal to boost the teaching of reading comprehension skills.

8. References

Abadzi, H. (2006) *Efficient learning for the poor.* Washington: The World Bank.

Adams M. (1990) *Beginning to read: Thinking and learning about print.* Cambridge, MA: MIT Press.

Ardington, C. & Meiring, T. (2020) *Impact evaluation of Funda Wande coaching intervention midline findings*. Cape Town: SALDRU, University of Cape Town.

Ardington, C. & Menendez, A. (2020) *Comments on Design Report B: Reading Benchmarking in South Africa.* Unpublished report submitted to USAID. Cape Town.

Ardington, C., Hoadley, U. & Menendez, A. (2019) *Impact Evaluation of USAID/South Africa Story Powered School Program Endline Report.* Washington, DC: USAID Reading and Access Activity.

Aro, M. (2017) Learning to read Finnish. In Verhoeven, L & Perfetti, C. (eds). *Learning to read across languages and writing systems* (pp 416-435). Cambridge: Cambridge University Press.

Block, C.C. & Pressley, M. (eds.) (2002) *Comprehension instruction: Research-based best practices.* New York, NY: Guilford.

Caravolas, M., Lervåg, A., Defior, S., Seidlova-Makora, G. & Hulme, C. (2013) Different Patterns, but Equivalent Predictors, of Growth in Reading in Consistent and Inconsistent Orthographies. *Psychological Science*, 24(8), 1398-1407.

Department of Basic Education (2019) The Secondary Early Grade Reading Study. Year 1 Report. Learner performance after the first year of implementation. Pretoria: Department of Basic Education.

Currie, N.K., & Cain, K. (2015) Children's inference generation: The role of vocabulary and working memory. *Journal of Experimental Child Psychology*, *137*, 57–75. https://doi.org/10.1016/j.jecp.2015.03.005

Deno, S.L., Fuchs, L.S., Marston, D. & Shin, J. (2001) Using curriculum-based measurement to establish growth standards for students with learning disabilities. *School Psychology Review*, 30(4), 507–524.

Florit, E. & Cain, K. (2011) The simple view of reading: Is it valid for different types of alphabetic orthographies? *Educational Psychology Review,* 23, 553–576. https://doi.org/10.1007/s10648-011-9175-6

Fuchs, L.S., Fuchs, D., Hosp, M.K. & Jenkins, J.R. (2001) Oral reading fluency as an indicator of reading competence: A theoretical, empirical, and historical analysis. *Scientific Studies of Reading,* 5(3), 239–256.

Fuchs, D., Compton, D., Fuchs, L., Bryant, J., Hamlett, C. & Lambert, W. (2012) First-grade cognitive abilities as long-term predictors of reading comprehension and disability status. *Journal of Learning Disabilities,* 45, 217–231. https://doi.org/10.1177/0022219412442154

Gough, P. & Tunmer, W. (1986) Decoding, reading, and reading disability. *Remedial and Special Education*, 7, 6-10.

Hasbrouck, J. & Tindal, G.A. (2006) Oral reading fluency norms: a valuable assessment tool for reading teachers. *The Reading Teacher*, 59(7), 636-644.

Hoover, W.A. & Gough, P.B. (1990) The simple view of reading. *Reading and Writing: An Interdisciplinary Journal*, 2, 127–160. https://doi.org/10.1007/BF00401799

Howie, S., Combrinck, C., Roux, K., Tshele, M., Mokoena, G., & McLeod Palane, N. (2017) *Progress in International Reading Literacy Study 2016 - South African Childrens' Reading Literacy Achievement*. Pretoria: Centre for Evaluation and Assessment

Jenkins, J.R., Fuchs, L.S., van den Broek, P., Espin, C. & Deno, S.L. (2003) Sources of individual differences in reading comprehension and reading fluency. *Journal of Educational Psychology*, *95*(4), 719–729. https://doi.org/10.1037/0022-0663.95.4.719

Katz, L., & Frost, R. (1992) The reading process is different for different orthographies: The orthographic depth hypothesis. In R. Frost & L. Katz (eds.), *Orthography, phonology, morphology, and meaning* (pp. 67–84). Amsterdam: Elsevier Science Publishers.

Kendeou, P., Papadopoulos, T.C. & Kotzapoulou, M. (2013) Evidence for the early emergence of the simple view of reading in a transparent orthography. *Reading and Writing: An Interdisciplinary Journal*, 26, 189–204. https://doi.org/10.1007/s11145-012-9361-z

Kendeou, P., van den Broek, P., White, M. J. & Lynch, J.S. (2009) Predicting reading comprehension in early elementary school: The independent contributions of oral language and decoding skills. *Journal of Educational Psychology*, 101, 765–778. https://doi.org/10.1037/a0015956

Kim, Y-S.G., Park, C., & Wagner, R.K. (2014) Is oral/text reading fluency a "bridge" to reading comprehension? *Reading and Writing*, *27*(1), 79–99. https://doi.org/10.1007/s11145-013-9434-7

Kim, Y-S.G. (2015) Language and cognitive predictors of text comprehension: Evidence from multivariate analysis. *Child Development*, 86, 128–144. https://doi.org/10.1111/ cdev.12293

Kim, Y-S.G., & Wagner, R.K. (2015) Text (oral) reading fluency as a construct in reading development: An investigation of its mediating role for children from grades 1 to 4. *Scientific Studies of Reading*. Advance online publication. doi:10.1080/10888438.2015.1007375

Kim, Y-S.G. (2017) Why the simple view of reading is not simplistic: Unpacking the simple view of reading using a direct and indirect effect model of reading (DIER). *Scientific Studies of Reading*, 21, 310–333. http://dx.doi.org/10.1080/10888438.2017.1291643

Kim, Y-S.G. (2020) Hierarchical and Dynamic Relations of Language and Cognitive Skills to Reading Comprehension: Testing the Direct and Indirect Effects Model of Reading (DIER) *Journal of Educational Psychology*, 112(4), 667– 684. http://dx.doi.org/10.1037/edu0000407

Kim, Y-S.G. & Piper, B. (2019) Component skills of reading and their structural relations: evidence from three sub-Saharan African languages with transparent orthographies. *Journal of Research in Reading,* 1–23.

Kintsch, W. (1988) The use of knowledge in discourse processing: A construction-integration model. *Psychological Review*, 95, 163–182. https://doi.org/10.1037/0033-295X.95.2.163

Land, S. (2015) *Reading isiZulu: Reading processes in an agglutinative language with a transparent orthography*. Unpublished doctoral thesis, University of KwaZulu-Natal, Pietermaritzburg, South Africa.

Miller, P., Kargin, T. & Guldenoglu, B. (2014) Differences in the reading of shallow and deep orthography: Developmental evidence from Hebrew and Turkish readers. *Journal of Research in Reading* 37(4), 409–432. https://doi.org/10.1111/j.1467-9817.2012.01540.x

Mullis, I.V.S. & Martin, M.O. (2015) PIRLS 2016 Assessment Framework (2nd ed.). TIMMS & PIRLS International Study Center and IEA, United States.

Nag, S. (2007) Early reading in Kannada: The pace of acquisition of orthographic knowledge and phonemic awareness. *Journal of Research in Reading*, 30, 7–22. https://doi.org/10.1111/j.1467-9817.2006.00329.x

Nag, S. (2017) Learning to read alphasyllabaries. In K Cain, DL Compton & RK Parilla (eds.), *Theories of reading development* (pp. 75-97). Amsterdam: John Benjamins Publishing Company.

Oakhill, J., Hartt, J. & Samols, D. (2005) Levels of comprehension monitoring and working memory in good and poor comprehenders. *Reading and Writing*, *18*, 657–686. https://doi.org/10.1007/s11145-005-3355-z

Oakhill, J., Cain, K. & Elbro, C. (2015) *Understanding and teaching reading comprehension.* London: Routledge.

Pearson, D.P., Roehler, L.R., Dole, J.A., & Duffy, G.G. (1990) *Technical Report No. 512. Developing expertise in reading comprehension: What should be taught? How should it be taught?* Center for the Study of Reading. College of Education, University of Illinois at Urbana-Champaign.

Pikulski, J.J. & Chard, D.J. (2006) Bridge between decoding and reading comprehension. *The Reading Teacher*, *58*(6), 510–519. https://doi.org/10.1598/RT.58.6.2

Pretorius, E.J. (2005) ESL learner differences in anaphoric resolution: Reading to learn in the academic context. *Applied Psycholinguistics,* 26, 521-539.

Pretorius, E.J. (2006) The comprehension of logical relations in expository texts by students who study through the medium of ESL. *System* 34(3), 432-450.

Raphael, T., Highfield, K. & Au, K.H. (2006) *Question answer relationships.* New York: Scholastic.

RTI International (2017) *All Children Reading-Asia: EGRA Benchmarks and Standards Research Report*.

Seymour, P.H.K., Aro, M. & Erskine, J.M. (2003) Foundation literacy acquisition in European orthographies. *British Journal of Psychology*, 94, 143–174.

South African Government (2019) *State of the Nation Address by Cyril Ramaphosa*. 20 June 2019. https://www.gov.za/speeches/2SONA2019.

Spaull, N., Pretorius, E.J. & Mohohlwane, M. (2020) Investigating the Comprehension Iceberg: Developing empirical benchmarks for early grade reading in agglutinating African languages. *South African Journal for Early Childhood Education* 10(1), a773. https://doi.org/ 10.4102/sajce.v10i1.773

Spear-Swerling, L. (2006) Children's reading comprehension and oral reading fluency in easy text. *Reading and Writing: An Interdisciplinary Journal, 19,* 199-220.

Stafura, J.Z. & Perfetti, C.A. (2017) Integrating word processing with text comprehension: Theoretical frameworks and empirical examples. In K Cain, DL Compton & RK Parrila (eds.). *Theories of reading development* (pp. 9-32). Amsterdam: John Benjamins Publishing Company.

Strasser, K. & del Rio, F. (2014) The role of comprehension monitoring, theory of mind, and vocabulary depth in predicting story comprehension and recall of kindergarten children. *Reading Research Quarterly*, 49, 169–187.

Stern, M.B., Dubeck, M.M. & Dick, A (2018) Early Grade Reading Assessment (EGRA) data for targeted instructional support: Learning profiles and instructional needs in Indonesia. *International Journal of Educational Development*, 61, 64-71.

Stanovich, K.E. (2000) *Progress in understanding reading: Scientific foundations and new frontiers*. New York: The Guildford Press.

Taylor, N., Wills, G. & Hoadley, H. (2019) Addressing the 'leadership conundrum' through a mixed methods study of school leadership for literacy. *Research in Comparative and International Education* 14 (1). 30-53 https://doi.org/10.1177/1745499919828928.

Torppa, M., Georgiou, G.K., Lerkkanen, M-K, Niemi, P., Poikkeus, A-M., & Nurmi, J.E. (2016) Examining the simple view of reading in a transparent orthography: A longitudinal study from kindergarten to grade 3. *Merrill Palmer Quarterly*, 62, 179–206. http://digitalcommons.wayne.edu/mpq/vol62/iss2/4/

Tunmer, W.E. & Chapman, J.W. (2012) The simple view of reading redux: Vocabulary knowledge and the independent components hypothesis. *Journal of Learning Disabilities*, 45, 453–466. http://dx.doi.org/10.1177/0022219411432685

Van den Broek, P. (1997) Discovering the cement of the universe: the development of event comprehension from childhood to adulthood. In: Van den Broek, P., Bauer, P.J., Bourg, T. (eds.), *Developmental Spans in Event Comprehension and Representation*. Lawrens Erlbaum Associates, Mahwah, NJ.

Van den Broek, P. & Kendeou, P. (2017) Development of reading comprehension: Change and continuity in the ability to construct coherent representations. In K Cain, DL Compton & RK Parrila (Eds.), *Theories of reading development* (pp. 283-308). Amsterdam: John Benjamins Publishing Company.

Van Dijk, T.A. & Kintsch, W. (1983) Strategies of Discourse Comprehension. New York: Academic Press

Vauras, M., Kinnunen, R., Kuusela, L. (1994) Development of text-processing skills in high-, average-, and low-achieving primary school children. *Journal of Reading Behavior,* 26, 361–389.

Vihman, M.M. (1996). *Applied language studies. Phonological development: The origins of language in the child.* Blackwell Publishing.

Wang, Z., Sabatini, J., O'Reilly, T. & Weeks, J. (2019) Decoding and Reading Comprehension: A Test of the Decoding Threshold Hypothesis. *Journal of Educational Psychology*, 111(3), 387-401.

Wimmer, H. (2006) Don't neglect reading fluency! *Developmental Science*, 9, 447-448.

Wolf, M. & Katzir-Cohen, T. (2001). Reading fluency and its intervention. *Scientific Studies of Reading*, 5(3), 211–239.

Zenex Foundation (2018) Leveraging our collective wisdom to address the literacy challenge. Conference Proceedings of the Literacy Symposium, 24 October 2018. https://www.zenexfoundation.org.za/informationhub/publications/early-grades-numeracy-literacy

Ziegler, J.C & Goswami, U. (2005) Reading acquisition, developmental dyslexia, and skilled reading across languages: a psycholinguistic grain size theory. *Psychological Bulettin*, 131, 3–29. http://dx.doi.org/10.1037/0033-2909.131.1.3

9. Appendix

Appendix A: Additional tables and figures

Table A 1: EGRA study testing components by Nguni language, grade and term

General study features						EGRA tasks assessed				ORF features		
Language#	Study	Grade	Term tested	Assessment (wave) and year	Assessment in Tangerine	Letter-sounds	Separate complex consonants	Single word reading	ORF passage	ORF time limit (minutes)	ORF Reading for additional 2-mins	ORF comprehension
isiXhosa	FW	1	I	Wave 1 2018	X	X	X	-	-	-	-	-
	Zenlit	1	IV	2017	-	X	-	X	X	1	-	X
	FW	1	IV	Wave 2 2018	X	X	X	X	X	1	X	X
	SPS	2	I	Wave 1, 2017/2018	X	X	-	X	X	1	X	X
	FW	2	I	Wave 1 2018	X	X	X	X	X	1	X	X
	Zenlit	2	IV	2017	-	X	-	X	X	1	-	X
	FW	2	IV	Wave 2 2018	X	X	X	X	X	1	X	X
	SPS	3	I	Wave 1, 2017/2018	X	X	-	X	X	1	X	X
	SPS	3	III	Wave 2, 2018/2019	X	X	-	X	X	1	X	X
	Zenlit	3	IV	2017	-	X	-	X	X	1	-	X
	SPS	4	I	Wave 1, 2017/2018	X	X	-	X	X	1	X	X
	SPS	4	III	Wave 2, 2018/2019	X	-	-	X	X	1	X	X
	SPS	5	III	Wave 2, 2018/2019	X	-	-	X	X	1	X	X
isiZulu	EGRS II	1	I	Wave 1, 2017	X	X*	-	X	-	-	-	-
	EGRS II	1	IV	Wave 2, 2017	X	-	-	X	-	-	-	-
	Zenlit	1	IV	2017	-	X	-	X	X	1	-	X
	SPS	2	I	Wave 1, 2017/2018	X	X	-	X	X	1	X	X
	EGRS II	2	IV	Wave 3, 2018	X	X	-	-	X	1	-	X
	Zenlit	2	IV	2017	-	X	-	X	X	1	-	X
	LFL	3	I	Wave 1, 2017	X	X	-	X	X	1	-	X
	SPS	3	I	Wave 1, 2017/2018	X	X	-	X	X	1	X	X
	SPS	3	III	Wave 2, 2018/2019	X	X	-	X	X	1	X	X
	EGRS II	3	IV	Wave 4, 2019	X	X	-	-	X	1	X	X
	LFL	3	IV	Wave 2, 2017	X	X	-	X	X	1	-	X
	Zenlit	3	IV	2017	-	X	-	X	X	1	-	X
	SPS	4	I	Wave 1, 2017/2018	X	X	-	X	X	1	X	X
	SPS	4	III	Wave 2, 2018/2019	X	-	-	X	X	1	X	X
	SPS	5	III	Wave 2, 2018/2019	X	-	-	X	X	1	X	X
	LFL	6	IV	Wave 2, 2017	X	-	-	-	X	1	-	X
siSwati	EGRS II	1	I	Wave 1, 2017	X	X*	-	X	-	-	-	-
	EGRS II	1	IV	Wave 2, 2017	X	-	-	X	-	-	-	-
	EGRS II	2	IV	Wave 3, 2018	X	X	-	-	X	1	-	X
	EGRS II	3	IV	Wave 4, 2019	X	X	-	-	X	1	X	X

Notes: *All letter-sound tasks are timed to 1-minute with the exception of EGRS II.

Table A 2: Learner and school sample sizes by Nguni language, grade and school term across EGRA studies

				Learner sample size			School sample size		
	Language*	Grade	Term tested	Total learner sample	Learner sample - treatment	Learner sample - control	Total schools	Schools - treatment	Schools - control
EARLY GRADE READING STUDY II (EGRS II)									
Wave 1, 2017	S	1	I	2358	1347	1011	130	73	57
Wave 1, 2017	Z	1	I	969	521	448	50	27	23
Wave 2, 2017	S	1	IV	2146	1231	915	129	73	56
Wave 2, 2017	Z	1	IV	922	490	432	51	27	24
Wave 3, 2018	S	2	IV	1999	1167	832	130	74	56
Wave 3, 2018	Z	2	IV	766	405	361	49	26	23
Wave 4, 2019	S	3	IV	1922	1091	831	130	73	57
Wave 4, 2019	Z	3	IV	762	400	362	50	27	23
LEADERSHIP FOR LITERACY (LFL)									
Wave 1, 2017	Z	3	I	509	-	-	42	-	-
Wave 2, 2017	Z	3	IV	430	-	-	42	-	-
Wave 2, 2017	Z	6	IV	369	-	-	42	-	-
ZENLIT									
2017	Z	1	IV	180	121	59	12	8	4
2017	Z	2	IV	179	119	60	12	8	4
2017	Z	3	IV	179	119	60	12	8	4
2017	X	1	IV	150	90	60	10	6	4
2017	X	2	IV	150	90	60	10	6	4
2017	X	3	IV	150	90	60	10	6	4
STORY POWERED SCHOOLS (SPS)									
Wave 1, 2017/2018	X	2	I	1676	853	823	171	86	85
Wave 1, 2017/2018	X	3	I	1671	848	823	171	86	85
Wave 1, 2017/2018	X	4	I	1685	860	825	171	86	85
Wave 1, 2017/2018	Z	2	I	1780	924	856	188	97	91
Wave 1, 2017/2018	Z	3	I	1784	924	860	188	97	91
Wave 1, 2017/2018	Z	4	I	1808	929	879	188	97	91
Wave 2, 2018/2019	X	3	III	1420	706	714	170	85	85
Wave 2, 2018/2019	X	4	III	1431	720	711	170	85	85
Wave 2, 2018/2019	X	5	III	1460	731	729	170	85	85
Wave 2, 2018/2019	Z	3	III	1477	778	699	188	97	91
Wave 2, 2018/2019	Z	4	III	1535	789	746	188	97	91
Wave 2, 2018/2019	Z	5	III	1474	738	736	188	97	91
FUNDA WANDE (FW)									
Wave 1 2018	X	1	I	595	301	294	59	30	29
Wave 1 2018	X	2	I	592	302	290	59	30	29
Wave 2 2018	X	1	IV	555	276	279	58	30	29
Wave 2 2018	X	2	IV	561	283	278	58	30	29

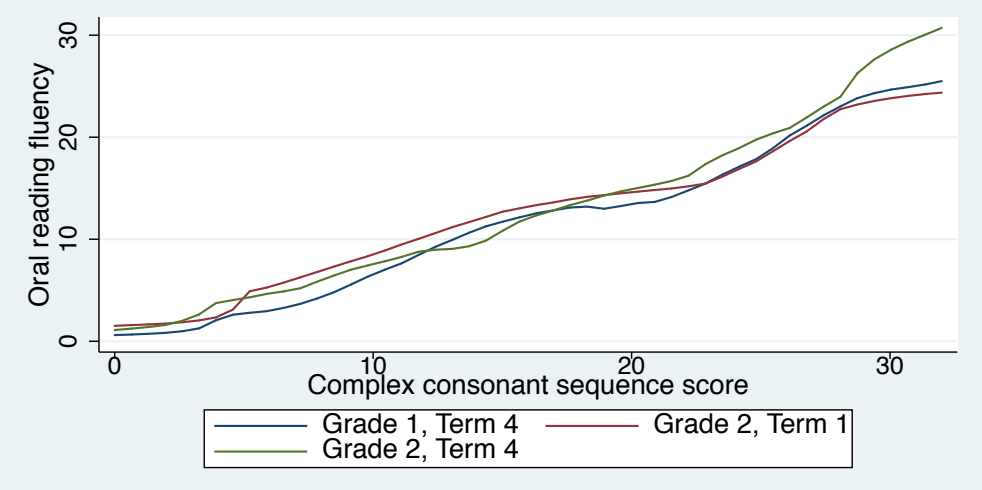
Notes: *X = isiXhosa, Z = isiZulu, S = siSwati

Table A 3: Sampling design of individual EGRA studies

	Early Grade Reading Study (EGRS II)
Sampling of schools:	Representative of districts of Ehlanzeni and Gert Sibande in Mpumulanga province. Random allocation of qualifying schools to treatment and control groups.
Sampling within schools:	Random selection of learners from each school. +/ 20 learners per school at baseline (wave1).
Stratification:	Ten strata based on school size, performance in the Annual National Assessments (ANA) and by school wealth (DBE Quintile).
Implementer & Funder:	Implementer: Department of Basic Education. Funder: USAID
	Story Powered Schools (SPS)
Sampling of schools:	Within each district, the sampling team worked with DBE officials to create clusters of 10 schools for cohort I, and 4 schools for cohort I. Clusters were randomly assigned to treatment and control. 4 schools (cohort I)/2 schools (cohort II) were randomly selected from each cluster for inclusion in the evaluation.
Sampling within schools:	Random selection of one grade 2, one grade 3 and one grade 4 class per school. Random selection of 5 girls and 5 boys within each selected class
Stratification:	7 strata based on ECDoE and KZNDoE districts
Implementer & Funder:	Implementer: NORC at the University of Chicago. Funder: USAID
	Funda Wande (FW)
Sampling of schools:	Districts: Buffalo City, NMMB, Saartjie Baartman.
	District managers identified schools that should be invited to apply for the Funda Wande programme. Criteria were no-fee schools, isiXhosa LOLT, no major literacy intervention. District managers identified 93 schools, which were invited to apply. Of these, 77 returned the application form and were self-described as motivated to take part in the study. Funda Wande then screened the applications to exclude schools with chronic management problems, severe overcrowding or fewer than 20 learners per grade. Of the applications, 64 schools were considered eligible for the programme. The list included one quintile 1 school but all the rest of the schools were classified as quintile 3. In the interest of having treatment and control schools as similar as possible, the quintile 1 school was removed as it was also fairly remote.
Sampling within schools:	One grade 1 and one grade 2 class were randomly selected. Within each class, 10 learners were randomly selected
Stratification:	8 strata based on Eastern Cape DoE circuits
Implementer & Funder:	Implementer: SALDRU at the University of Cape Town.
	Leadership for Literacy project (LFL)
Sampling of schools:	Matched pairs design. For the isiZulu sample: 42 potentially higher performing no-fee or low fee schools (with largely isiZulu home language students) matched to under-performing pairs of schools using the Annual National Assessments. 22 of the schools are in KwaZulu-Natal, and 20 in Gauteng.
Sampling within schools:	Students from one grade 3 and one grade 6 class per school were sampled. Originally 10 students from each class were to be randomly sampled for the one-on-one assessments. But by wave 2 fewer or higher numbers of randomly sampled learners are achieved: 4-15 students per class in grade 6 and 6-15 per class in grade 3.
Stratification:	None
Notes on exclusions:	There is one school (schoolid 2400) with extremely high grade 6 ORF scores for all 10 learners tested. It's not clear if this school should be deleted from the grade 6 measures due to issues with testing or whether these are real reading outcomes. We keep this school in the sample.

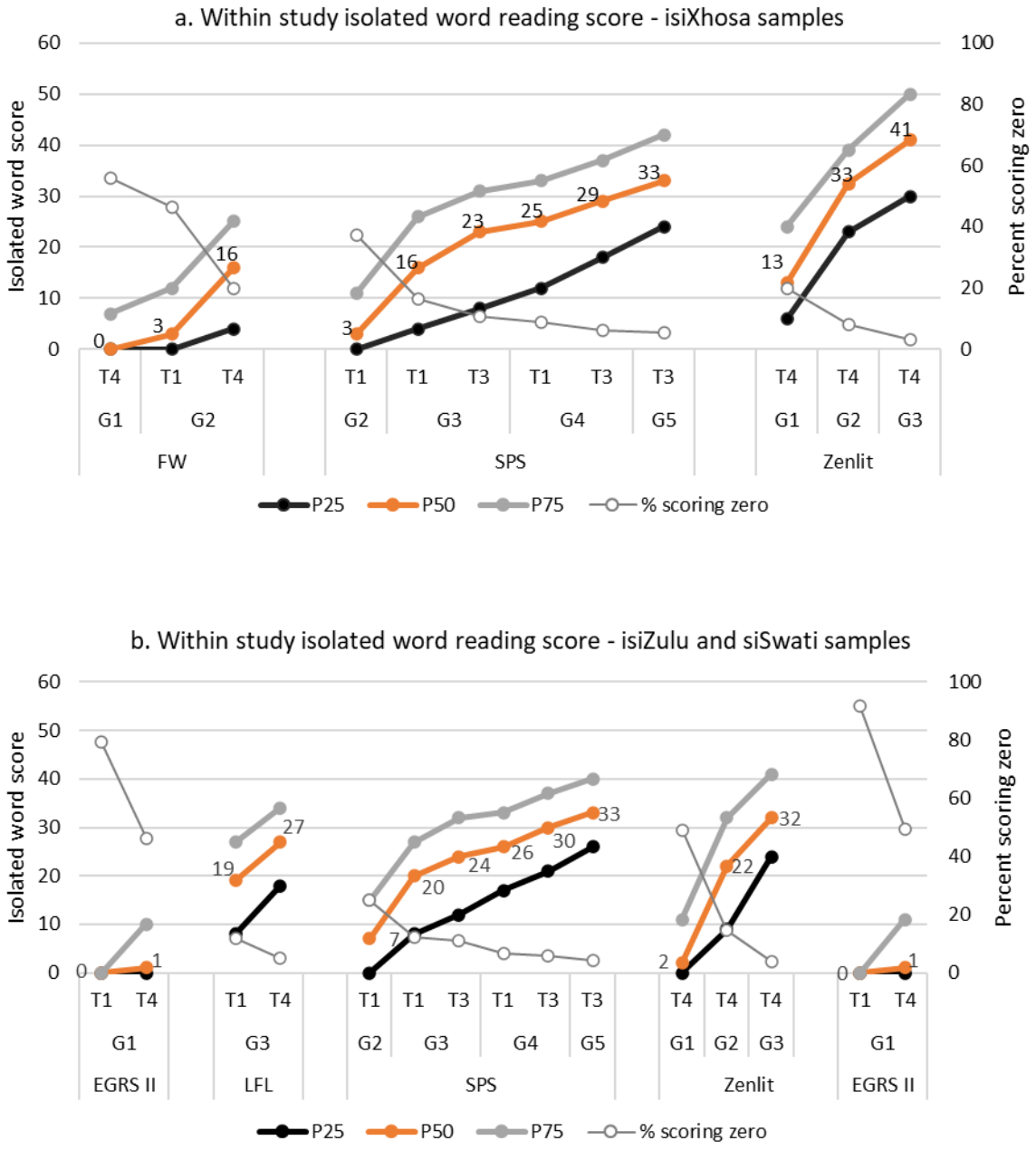
Implementer & Funder:	Implementer: Research on Socio-economic Policy (ReSEP). Funder: Economic Sciences Research Council (ESRC/DFID)
	Zenex Foundation Literacy Project (Zenlit)
Sampling of schools:	Purposive sampling of test schools. Control schools were matched on LOLT, region, ANA scores, socio-economic status and quintile. isiXhosa schools are all in the Eastern Cape (mainly from Nelson Mandela Bay). isiZulu schools are in KwaZulu-Natal (Ilembe District).
Sampling within schools:	Achieved sample per school: 15 grade 1 learners; 14-20 grade 2 learners and 14-16 grade 3 learners.
Stratification:	None
Notes on exclusions:	None
Implementer & Funder:	Funder: Zenex Foundation.

Figure A 1: Relationship between knowledge of complex sequences (2 consonants together) and oral reading fluency



Source: Funda Wande evaluation, isiXhosa sample.

Figure A 2: Comparison of isolated word reading score within studies – 25th, 50th and 75th percentiles and percent scoring zero



Source: Pooled EGRA dataset. Notes: We do not disaggregate across control and treatment groups in the case of intervention studies. These are not all longitudinal samples in the strictest sense. The SPS sample includes 2 cohorts. Zenlit is not a panel - rather learners in different grades are observed in the same schools at the same point in time.

Appendix B: Categorisation of comprehension questions

Since this benchmarking study combined EGRA data from multiple reading studies, it was important to ensure uniformity in the categorisation of the reading comprehension questions across the various assessments. The conceptual framework for classifying different kinds of questions in reading comprehension assessment is based on the PIRLS framework, where 4 categories of question types are used, ranging in cognitive demand from easy, literal questions to more challenging inferential and evaluative questions. These 4 levels are set out in the box below (taken from Howie et al. 2017).

1. Focus on and Retrieve Explicitly Stated Information (Literal)

In focusing on and retrieving explicitly stated information, readers locate and understand content that is relevant to the question. Items testing this process require the reader to focus on the text at the word, phrase and sentence level for the purpose of constructing meaning. The process may also require the reader to focus on and retrieve pieces of information from across the text (Mullis & Martin, 2015).

2. Make Straightforward Inferences (SI)

The ability to 'make straightforward inferences' that are not explicitly stated allows readers to move beyond the surface of texts and to resolve gaps in meaning. Some of these inferences are straightforward in that they are based primarily on information that is contained in the text and readers must connect 2 or more ideas. The ideas themselves may be explicitly stated, but the connection between them is not, and must be inferred. Skilled readers will connect 2 or more pieces of information and recognise the relationship even though it is not stated in the text (Mullis & Martin, 2015).

3. Interpret and Integrate Ideas and Information (I&I)

As with the more straightforward inferences, readers who are engaged in interpreting and integrating ideas and information in text may focus on local or global meanings. As readers interpret and integrate, they construct meaning by integrating personal knowledge and experience with meaning that resides within the text. In this way, readers draw on their understanding of the world, as well as their background knowledge and experiences, more than they do for straightforward inferences and make connections that are not only implicit, but that may be open to some interpretation based on their own perspective (Mullis & Martin 2015).

4. Evaluate and Examine Content, Language and Textual Elements (E&E)

As readers evaluate the content and elements of a text, the focus shifts from constructing meaning to critically considering the text itself. In evaluating and critiquing elements of text structure and language, readers draw upon their knowledge of language usage to reflect on and judge the author's language choices and devices for conveying meaning. Using past reading experience and familiarity with the language and text structure, readers evaluate the visual and textual features used to organise the text (Mullis & Martin 2015).

Sometimes questions may seem to straddle 2 categories, posing challenges as to how they should be categorised. To this end, 4 criteria were used to serve as guidelines to clarify distinctions between categories and to ensure consistency in assigning questions to their relevant categories. Based on the seminal work on reading comprehension by David Pearson and colleagues at the Centre for Reading, University of Illinois during the seventies and eighties (Pearson et al, 1990), the central premise underlying these criteria is that the question type and its difficulty level is determined in relation to the task demands of the question and the source of the information needed to answer it.

- **Source of information:** In reading comprehension, there are basically 2 sources of information from which answers to questions can derive – the *text* being read (text based or textual information) and the general knowledge that the *reader* brings to the reading task (reader based or personal information – also referred to as background information).
- **Explicitness of textual information:** The information needed to answer a question may be *explicitly* stated in a text (i.e. it can be located precisely in a specific part of the text) or it may be *implicit* in the text (i.e. the answer is somewhere in the text but it needs to be inferred). Inferring answers from textual information places higher cognitive demands on the reader than locating information in a specific part of the text. Raphael, Highfield and Au (2006) refer in practical terms to explicit text information as **'right there'** – readers can put their finger on the information in the text to answer the question, while implicit text information is referred to as **'think and search'** – the information is implied somewhere in the text, but the reader has to put together information from different parts of the text to answer the question (i.e. make connections or inferences between text parts).
- **Distance between connected textual information:** Research has shown that it is easier to make inferences or connections between adjacent or local parts of text than between more remote or global parts of text (Vauras, Kinnunen & Kuusela 1994; Van den Broek 1997. Pretorius 2005, 2006). The criterion of **local** versus **distant** position was used to distinguish between connecting information from (i) within a single sentence or from adjacent sentences (local), as opposed to (ii) connecting information across sentences within a paragraph, across paragraphs or across the whole text (global).
- **Relationship between reader-based and text-based information:** Although all reading comprehension relies on the knowledge that a reader brings to a text (the reader's linguistic knowledge, knowledge of texts and literate conventions, general background knowledge, extent of topic-specific knowledge, etc), reading comprehension in particular engages reader-based knowledge with information in the text being read in particular ways. Raphael et al. (2006) refer in practical terms to reader-based knowledge as **'in my head'** information, and they then further distinguish between **'text and me'** and **'on my own'** information. More challenging questions that require readers to infer, integrate or evaluate information across a text (i.e. globally) rely more on implicit textual information combined with reader-based information, i.e. **'text and me'** information. When information to an answer is not found in the text at all it can be categorised as an **'on my own'** response. Here the question relates specifically to readers' personal experiences, ideas and feelings, and all responses will be unique to particular readers. These are typically the kinds of questions that teachers will pose to learners before reading a text, to activate their background knowledge. A reader does not have to read the text to answer this kind of question. Because these kinds of questions are very open-ended, auto experiential and involve no true 'right or wrong' answers, they are seldom asked in formal reading comprehension assessments, and so they will not concern us here.

The PIRLS question categories are arranged hierarchically from 1 to 4 in terms of increasing cognitive demands. If uncertainty arises as to which category to assign a question, the doubt usually revolves around adjacent categories rather than those further apart. For example, it is easier to distinguish between literal (1) and integrative (3) or evaluative (4) questions than between literal (1) and straightforward inferences (2), or between straightforward inferences (2) and integrative (3) questions.

The 4 criteria outlined above help to clarify distinctions between the 4 PIRLS question types:

- To clarify the distinction between *Literal* and *Straightforward Inferences (SI)*, the criterion of **explicit information** in the text was used. Readers can put their finger on the answer in the text. The answer is not negotiable - it is unanimous and **'right there'.**
- To clarify the distinction between *Straightforward Inferences (SI)* and *Integrate and Interpret (I&I)* the criterion of **local versus distant** position was used. A question was classified as SI if it involved some thinking and searching to make a connection or inference between adjacent

sentences. If the inference was made across several sentences (i.e. non-adjacent sentences), then the question was classified as an I&I, since connecting information globally requires greater effort of integration. In addition, although all inferences rely on some kind of knowledge, straightforward, local inferences rely more readily on **textual information** in adjacent chunks of information (**in the text**). Whereas with I&I questions, text information combined with own personal knowledge form the basis for integration and interpretation (**'text and me'**).

- To clarify the distinction between *Integrate and Interpret (I&I)* and *Evaluate and Examine (E&E)*, the criterion of **'text and me'** was used to distinguish whether the connection made by the reader was more of interpreting global connections (I&I) or making global connections of an evaluative nature (E&E).

Table A4 below provides a summary of some of the texts used in the study, together with information concerning the grades for which they were used, their text length and the number and types of questions they comprised. Details about the criteria as they applied to each text at each grade level are available on request from the authors, together with evidence as to where the information for answering the question was positioned. As can be seen, nearly all the questions are similarly categorised across the different language versions; there were only 2 differences in the Grade 4 and 5 isiXhosa and isiZulu versions of *Isuphu yelitye/Isobho lamatshe* from the Story Power Schools (SPS) Evaluation. These were accordingly captured in the data and do not affect the statistical analysis of the data.

Table A 4: Summary information about ORF texts and associated comprehension question types

Study	Name of ORF text	Grade	Language	Text length (number of words)	Number of questions	Literal	SI	I&I	E&E
FW Midline 2019	*USipho*	1 & 2	isiXhosa	41	12	1,2,3,5, 6, 8,10, 11	4,7,12	9	
FW Midline 2019	*UJabu nenja*	2	isiXhosa	56	10	1,3,4,7,8	2, 5,10	6, 9	
SPS Endline	*UJabu nenja*	3	isiXhosa	56	10	1,3,4,7,8	2, 5,10	6, 9	
			isiZulu	56	10	1,3,4,7,8	2, 5,10	6, 9	
SPS Endline	*Isuphu yelitye*	4	isiXhosa	56	9	1,5,6,7,8,	2,3,4,9		
	Isobho lamatshe		isiZulu	62	8	1,5,6,7	2,3,4,8		
SPS Endline	*Isuphu yelitye*	5	isiXhosa	82	10	1,2,7,8,9	3,4,5	6	10
	Isobho lamatshe		isiZulu	81	10	1,2,7	3,4,5, 9	6, 8	10

Notes: SI = Straightforward inference. I & I = Integrate and Interpret; E & E = Evaluate and Examine

Published by the Department of Basic Education

222 Struben Street

Private Bag X895, Pretoria, 0001

Telephone: 012 357 3000 Fax: 012 323 0601

ISBN Number: 978-1-4315-3412-5

website: www.education.gov.za

facebook: www.facebook.com/BasicEd

twitter: www.twitter.com/dbe_sa

Funders of this report

Research jointly supported by the ESRC and DFID