BENCHMARKING EARLY GRADE READING SKILLS IN SOUTH AFRICA:

TECHNICAL REPORT

SETSWANA HOME LANGUAGE

15 June 2022

This publication was produced at the request of the United States Agency for International Development. It was prepared independently by Khulisa Management Services, (Pty) Ltd in collaboration with the South African Department of Basic Education

CONTRACT NUMBER: 72067418D00001, ORDER NUMBER: 72067421F00001

Photo: Teaching early grade literacy, Mogokonyane Primary School, North West Province, South Africa

Photo credit: Khulisa Management Services

AUTHORS

Dr Gabrielle Wills (Research On Socio-Economic Policy, Stellenbosch University)

Prof Cally Ardington (University Of Cape Town, SALRDU)

Prof Elizabeth Pretorius (Emerita Professor And Research Fellow In The Department Of Linguistics And Modern Languages, Unisa)

Dr Eileen Pooe (Senior Lecturer, North West University & Setswana National Language Body - PanSALB)

Dr Refilwe Ramagoshi (Research Associate and Senior Lecturer, University of Pretoria)

Ms Lesang Sebaeng (Department of Basic Education)

ACKNOWLEDGEMENTS

Thank you to Ms Nompumelelo Mohohlwane and Ms Lesang Sebaeng from the Department of Basic Education for their valuable contributions to the development of the Learner Assessment passages and tools, training of fieldworkers and quality assurance of the data. In addition, we would like to thank the North West Province officials, school principals, teachers and learners for participating in this study. This study would not have been possible without the commitment of our partners and the Khulisa team, including the fieldworkers and supervisors who spent many hours collecting the data, and the Education Specialist, Ms Benita Williams.

CONTACT DETAILS

Margaret Roper
Khulisa Management Services
26 7th Avenue
Parktown North
Johannesburg, 2196
Telephone: 011-447-6464
Email: mroper@khulisa.com
Web Address: www.khulisa.com

Abbreviations and Acronyms

clspm	correct letter sounds per minute
cwpm	correct words per minute
DBE	Department of Basic Education
DIBELS	Dynamic Indicators of Basic Early Literacy Skills
EC	Eastern Cape
EFAL	English First Additional Language
EGRA	Early Grade Reading Assessment
EGRS	Early Grade Reading Study
ESL	English Second Language
FW	Funda Wande
GoSA	Government of South Africa
HL	home language
KZN	KwaZulu-Natal
LFL	Leadership for Literacy
LOLT	language of learning and teaching
LP	Limpopo Province
LSK	letter sound knowledge
MP	Mpumalanga Province
NW	North West
ODH	orthographic depth hypothesis
ORF	oral reading fluency
PGST	psycholinguistic grain size theory
PIRLS	Progress in International Reading and Literacy Study
RC	reading comprehension
RSP	Reading Support Project
SPS	Story Powered Schools
SVR	simple view of reading
USA	United States of America

Table of Contents

List of Tables

List of Figures

PREAMBLE

This report, produced by Khulisa Management Services (Pty) Ltd. (Khulisa), is submitted under the Data Collection and Analysis for the Early Grade Reading Study (EGRS), the Reading Support Project (RSP) and Language Benchmarking to the United States Agency for International Development (USAID) under PERFORMANCE Indefinite Delivery Indefinite Quantity (IDIQ) Contract Number: 72067418D00001, Order Number: 72067421F00001.

This report derives from the 2021 data collection and analysis for the EGRS I (wave 5 data), the RSP Impact Evaluations and the Language Benchmarking study in two districts in North West Province, South Africa.

A number of reports have been published under this task order and are useful as background.

- Methodology Plan and Study Protocol: Data Collection and Analysis for the EGRS, RSP and Benchmarking. https://pdf.usaid.gov/pdf_docs/42132810ec2c48809efe8ca11e155aff.pdf
- For the full instrument development process refer to the "Report on the Development of Learner Assessment Tools and Contextual Tools"
- The Quality Assurance Surveillance Protocol (QASP). The QASP documents the quality assurance elements of both data collection and analysis. https://pdf.usaid.gov/pdf_docs/PA00Z8SX.pdf
- Task Order 4 Data Collection and Analysis EGRS, RSP, Benchmark and COVID-19: Fieldwork Report https://pdf.usaid.gov/pdf_docs/e4563ed819164a79956698c3a1998964.pdf

As part of this task order, Khulisa conducted additional research on COVID-19 in the schools and two reports were provided. The Preliminary COVID-19 Report submitted in 2021 enabled the DBE to consider the policy implications to prepare for the 2022 school year. https://pdf.usaid.gov/pdf_docs/PA00XGST.pdf. Thereafter, the Consolidated Final COVID-19 Report was submitted in 2022 https://pdf.usaid.gov/pdf_docs/PA00ZBHD.pdf.

The EGRS I Impact Evaluation report and further reports on the EGRS are available on the Department of Basic Education website https://www.education.gov.za/Programmes/EarlyGradeReadingStudy.aspx

Data was analysed to recommend Setswana Home Language (HL) reading benchmarks and English First Additional Language (EFAL) reading benchmarks. The complete technical reports for EFAL and Setswana HL, as well as Summary Reports and Learning Briefs are available on the USAID Development Experience Clearinghouse and the Department of Basic Education Research Repository https://www.education.gov.za/ResearchRepository.aspx.

The methodology for Setting Reading Benchmarks In South Africa is outlined in this report https://pdf.usaid.gov/pdf_docs/PA00X1NZ.pdf..

The data used for Setswana was based on studies funded by the Department of Basic Education, the Department of Planning, Monitoring and Evaluation, the North West Provincial Department of Education, the Initiative for Impact Evaluation, Zenex Foundation, UNICEF, USAID, Anglo American Chairman's Fund.

EXECUTIVE SUMMARY

This report, produced by Khulisa Management Services (Pty) Ltd. (Khulisa), is submitted under the Data Collection and Analysis for the Early Grade Reading Study (EGRS), the Reading Support Project (RSP) and Benchmarking to the United States Agency for International Development (USAID) under PERFORMANCE Indefinite Delivery Indefinite Quantity (IDIQ) Contract Number: 72067418D00001, Order Number: 72067421F00001.

This report derives from the data collection and analysis for the EGRS I[1], the RSP and the Language Benchmarking study in two districts (Dr Kenneth Kaunda and Ngaka Modiri Molema) in North West, South Africa. Multiple assessment points for almost 16,000 Setswana home language learners from 230 quintile 1-3 or no-fee schools[2] in North West Province are available from existing and new data collected through the first Early Grade Reading Study (EGRS I) and Reading Support Project (RSP).

Reading for meaning is a core skill that children must master in the early grades to be able to learn. The South African school curriculum develops from the assumption that children can read for meaning in both their home language and English by the end of the Foundation Phase (Grade 3).

Reiterating the importance of learning to read to be able to learn, South African President Cyril Ramaphosa articulated a clear goal for basic education: every child should be able to read for meaning by age 10, which roughly aligns with the end of the Foundation Phase (South African Government, 2019). This important yet aspirational goal is, however, set against a sobering reality. The 2016 round of the Progress in International Reading Literacy Study (PIRLS) found that 78% of South African Grade 4 readers could not reach the low international PIRLS benchmark – a signal for being able to read for meaning. This compares to 4% among Grade 4 readers internationally. When broken down into language groups, almost 90% of Grade 4 Setswana readers could not read for meaning (Howie et al. 2017).

To read with understanding in African languages or English, various foundational reading subskills need to be mastered before children can, when reading on their own, 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

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

[2] No fee schools cannot charge school fees. These are schools in quintiles 1 to 3, the system DBE used to rate schools according to the income, unemployment and literacy levels in a community. The system is used to determine public funding to schools.

we refer to as decoding skills. Without suitable assessment data to measure these skills, problems acquiring them can go unnoticed, with undeveloped decoding skills showing up only later in very poor written comprehension as reflected in PIRLS results. The poor identification of decoding skills is further perpetuated through a lack of agreed standards as to what constitutes an on-track reading development trajectory across different African or home languages or among learners having to acquire second language reading proficiencies.

What are reading benchmarks and thresholds?

Reading benchmarks and thresholds, provide standards against which teachers can measure learners' reading subskills. These numerical measures of proficiency in specific reading skills can be used to monitor whether children are on track to be able to read with fluency and understanding. They can also be used to identify early on learners who are at risk of not learning to read for meaning by age 10, highlighting where effective remediation should take place (Jukes et al., 2020).

In processes to establish reading benchmarks, it is important to decide which reading subskills should be benchmarked. Developing skills in all subcomponents of reading are important but too many benchmarks can be confusing and hard to track. This report focuses on just two: letter-sound knowledge as a basic skill which refers to alphabetic knowledge of the written code; and oral reading fluency (ORF) referring to the ability to read words in context with speed, accuracy and prosody. These three components to ORF are defined here as follows:

- Accuracy refers to the percentage of words attempted 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).

Assessed measures of ORF typically focus only on speed and accuracy because prosody is subjective and difficult to measure. In this report, the term fluency is used to describe reading with speed and accuracy and is measured as the number of correct words per minute (abbreviated as 'cwpm') read from a passage of text.

The purpose of the proposed letter-sound and fluency thresholds and benchmarks is outlined below:

- **Letter-sound benchmark** - identifies whether learners are developing sufficient alphabetic knowledge that underpins decoding skills necessary for accuracy in reading.

CONTRACT NUMBER: 72067418D00001, ORDER NUMBER: 72067421F00001

- **Oral Reading Fluency (ORF) thresholds** - identify learners who are entering an emergent level of fluency which supports reading accuracy but which is not yet sufficient to support reading with understanding. Reading below a threshold impedes reading development.

- **ORF benchmark** - identifies a minimum fluency level that is necessary for learners to comprehend what they are reading and articulates to teachers a point at which they should concentrate on further developing comprehension skills.

Report aims

This report is dedicated to establishing letter-sound benchmarks and oral reading fluency thresholds and benchmarks for Setswana home language (HL) speakers in the early grades. Although Setswana is one of three official Sesotho-Setswana languages in South Africa, this is the first study to establish reading benchmarks in any Sesotho-Setswana language. This work complements recent research efforts to establish early grade reading benchmarks in Nguni languages (Ardington et al., 2020, 2021a) and English First Additional Language (EFAL) (Wills et al., 2022) in line with aims to establish benchmarks in all official South African languages (DBE, 2020b).

However, it is not sufficient to merely establish standards in home language reading. African home language serves as the language of learning and teaching (LOLT) until the third year (or end of the Foundation Phase) in most South African schools. Then LOLT switches to English from Grade 4 while home language instruction continues. Whereas just a quarter of learners in South Africa are taught in English in the Foundation Phase, 90% are instructed in English from Grade 4 onwards.[3] A separate report[4] focuses on EFAL, establishing fluency thresholds and benchmarks which are contextually relevant for second language English speakers in no-fee schools in South Africa. These thresholds and benchmarks are established to track learners' EFAL decoding skills not only in the Foundation Phase but into the Intermediate Phase, given current evidence revealing slow reading development well into the higher primary grades.

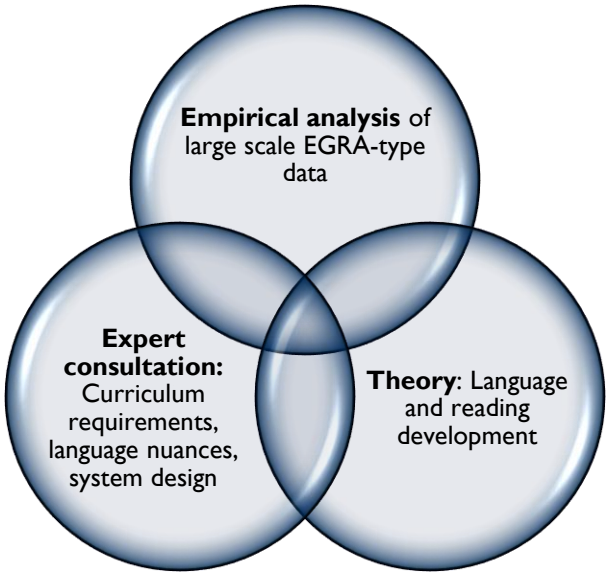
[3] Estimates from the Annual National Assessments of 2013. See https://www.bridge.org.za/wp-content/uploads/2016/06/SPAULL-2016-BRIDGE-reading-presentation.pdf

[4] Wills et al, 2022. The Setting Benchmarks in English First Additional Language Report, and the combined full technical report is available on the USAID DEC https://dec.usaid.gov/dec and the Department of Basic Education Research Repository https://www.education.gov.za/ResearchRepository.aspx.

Data and approach

The context-specific reading thresholds and benchmarks established in this report result from the intersection of three main research components: A review of the theory of language and reading development, empirical analysis of large-scale early grade reading assessment (EGRA) type data and expert consultation.

Figure E 1: Three research components intersect to establish early grade reading thresholds and benchmarks



Any efforts to establish reading thresholds and benchmarks should be grounded in theory about language and how reading develops in different languages. Theory then provides the backbone for an exploratory analysis of decoding skills and their interrelationships, using large-scale early grade reading assessment data.

Key to establishing empirically driven thresholds and benchmarks, is an examination of the relationship between accuracy and speed in reading and then fluency and reading comprehension. These relationships have been well established in alphabetic languages including English (Deno et al., 2001). Accuracy in recognizing letters and words 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). However, the nature of these relationships has been understudied in South African languages. We are sensitive not to impose assumptions about what these relationships look like in Setswana, and rather allow an analysis of empirical regularities and reading trajectories to identify critical thresholds and benchmarks in decoding skills.

CONTRACT NUMBER: 72067418D00001, ORDER NUMBER: 72067421F00001

Traditional approaches to benchmarking reading subskills often focus on identifying a single point or benchmark where decoding skills are sufficiently established to support comprehension (Abadzi, 2012). However, drawing on a 'threshold hypothesis' by Wang et al. (2019), reaching fluency levels as defined by a benchmark may only be attainable once a minimum threshold of proficiency in fluency has developed. Identifying both thresholds (also referred to as grade specific minimum benchmarks) and benchmarks can be a more informative approach to guide reading development and reading instruction. Using longitudinal data, we establish the predictive validity of fluency thresholds for meeting fluency benchmarks (and higher levels of comprehension) in later grades.

From research conception to final report writing, we have ensured that the established reading thresholds and benchmarks in Setswana are contextually relevant through collecting or using suitable existing EGRA-type assessment data for learners in no-fee schools. The data used are described below:

- **Setswana reading data:** Multiple assessment points for almost 16,000 Setswana home language learners from 230 quintile 1-3 or no-fee schools in North West Province are available from existing and new data collected through first Early Grade Reading Study (EGRS I)[5] and Reading Support Project (RSP). These data, particularly new data collected for this study in 2021, are highly suited to establishing Setswana benchmarks.

These data are used not only to establish benchmarks and thresholds, but to carefully examine whether they are attainable by learners at the grades for which they are set, yet aspirational enough for learners to be on a successful reading trajectory. The contextual relevance of the reading thresholds and benchmarks is also established through regular consultation with linguists, home language Setswana specialists and officials from the Department of Basic Education (DBE).

Through this research process, which brings together theoretical knowledge on reading development, empirical analysis, and expert knowledge of language and system realities, we recommend the following reading thresholds and benchmarks in Setswana.

[5] 5 waves of EGRS I data have been collected to track learner outcomes over time. The new data for this study (wave 5 data) includes EGRS I and RSP impact evaluation and benchmarking data for EFAL and Setswana HL

Early grade reading benchmarks: Setswana

Across different grades and assessments, consistent patterns are identified in observing the relationships between speed (number of words/letter-sounds attempted in a minute) and reading accuracy (the percentage of words /letter-sounds read correctly out of those attempted) within each Setswana reading subskill. As observed in Nguni language reading (Ardington et al., 2020, 2021), accuracy and speed initially increase rapidly together, but then this relationship flattens and levels off once learners achieve about 95% accuracy (i.e., learners read 95 of every 100 words/letter-sounds attempted correctly). Consistent with Nguni language analyses, when looking at letter-sounds this levelling off in accuracy occurs at speeds of approximately 40 letter-sounds attempted per minute.[6] However, when reading words in a connected text, the speed at which a flattening occurs in the accuracy-speed relationship is higher in Setswana than Nguni languages due its disjunctive rather than conjunctive orthography.

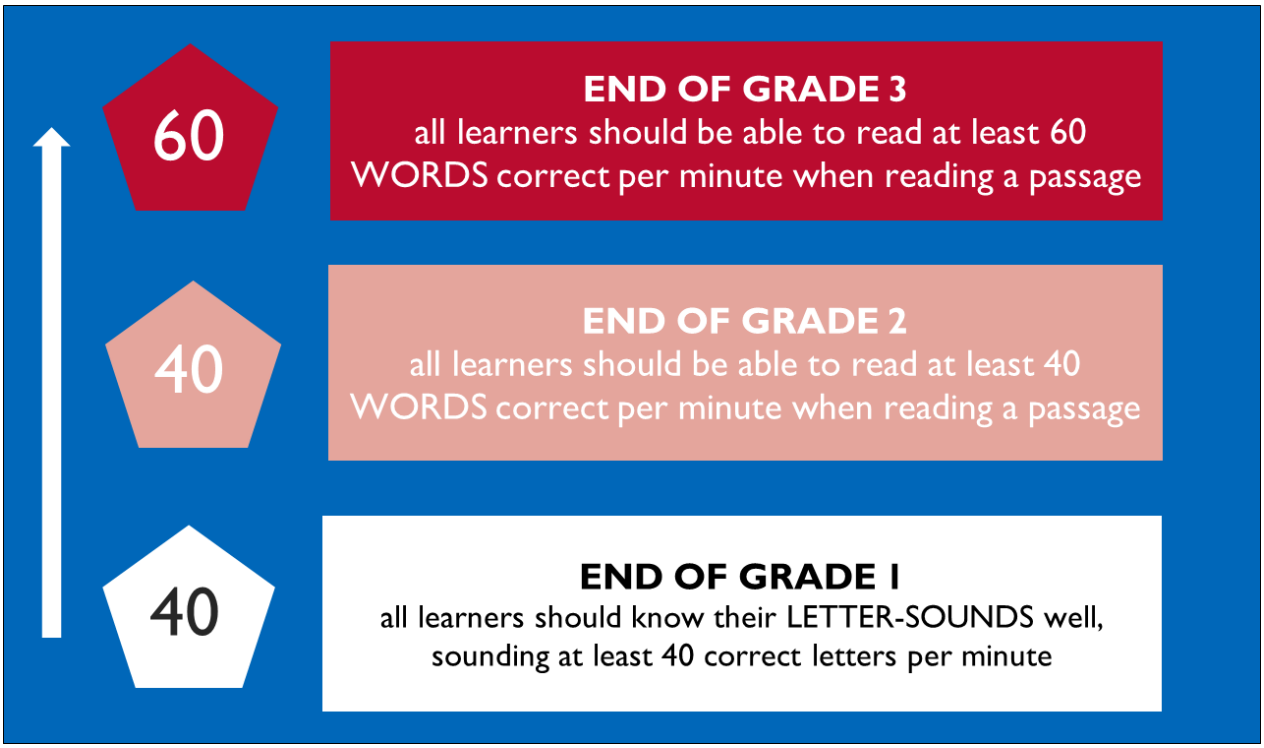
We also examine the relationship between fluency and comprehension. Among grade 3s and 4s, tested across different ORF passages, this relationship is very steep below 40 correct words per minute (cwpm), yet 40 cwpm appears to be a threshold below which learners' comprehension skills have not sufficiently developed to understand what is being read. In a separate analysis that compares learners' fluency to their scores on written comprehension tests, those with ORF scores below 40 cwpm show little evidence that they can comprehend what they have read. The comprehension-fluency gradient then tends to flatten out at around 60 cwpm with diminishing returns to fluency above this point, suggesting that underdeveloped comprehension skills become the key hurdle for learners at or above this fluency level. Typically, only learners that are reading at or above 60 cwpm by the end of the Foundation Phase develop strong comprehension skills as they advance into higher grades.

All this analysis, together with expert opinion, allows us to identify the following Foundation Phase thresholds and benchmarks for Setswana as summarized in Figure E1.

[6] Strong similarities in the relationship between accuracy and speed in letter-sound reading is expected across Nguni and Sotho languages as these are all alphabetic languages.

CONTRACT NUMBER: 72067418D00001, ORDER NUMBER: 72067421F00001

Figure E 2: Reading benchmarks and thresholds for early grade reading in Setswana home language



By the end of Grade 1, all learners taught in Setswana home language should be able to correctly sound 40 letters per minute.

- Beyond reading 40 correct letter-sounds per minute, there are few benefits of improving letter-sound knowledge and speed.
- Although the blending of sounds is integral to phonics instruction, once learners have met this letter-sound knowledge benchmark, decoding instruction should focus on assisting learners in applying word attack strategies and developing fluency.

By the end of Grade 2, all learners taught in Setswana home language should be meeting the Setswana fluency threshold, reading from a grade-appropriate passage at least 40 correct words per minute (cwpm).

- This is a *minimum* threshold. Higher order reading skills are very unlikely to develop if learners do not reach this fluency level. Reaching this fluency threshold by the end of Grade 2 is highly predictive of successful reading by the end of Grade 3 and reaching fluency benchmarks in higher grades.

- Pre-pandemic, approximately 42% of non-repeating learners in the EGRS I/RSP schools had reached the Setswana fluency threshold of 40 cwpm by the end of Grade 2.
- Below 40 words per minute, learners make a lot of mistakes when reading, and even if they don't make mistakes, they have very low oral and written comprehension scores. Quite simply, they are reading too slowly and inaccurately to comprehend what they are reading.

By the end of Grade 3, all learners taught in Setswana home language should be meeting the Setswana fluency benchmark, reading from a grade-appropriate passage at least 60 cwpm.

- This fluency benchmark is indicative of the point at which reading comprehension becomes increasingly possible when learners read on their own.
- Reaching this milestone signals when teachers' focus should hone in on teaching learners the skills and strategies they need to tackle written comprehension while encouraging vocabulary and language development. At this point, data patterns indicate that underdeveloped comprehension skills become the main constraint to further literacy development.
- It is necessary that *all learners* meet this home language benchmark by the end of Grade 3, a critical transition point before the language of instruction shifts from home language to English. Fluency in home language reading strongly supports reading in EFAL. As learners transition into higher grades, fluency skills in Setswana should continue to improve beyond this fluency benchmark.
- However, most learners are not reaching this benchmark by the end of Grade 3. Pre-pandemic, just a quarter (24%) of non-repeating learners in the EGRS I/RSP[7] schools had reached this Setswana fluency benchmark by the end of Grade 3. About a 51-67% of non-repeating learners meet this benchmark by the end of Grade 4, and 52-87% (depending on passage difficulty) of non-repeaters meet this benchmark by the end of Grade 7. This might help explain why Setswana Grade 4 learners performed so poorly in PIRLS 2016.

[7] In 2019 and 2020, the DBE requested USAID's support in proceeding with an expansion of the EGRS. The focus of the expansion was to scale up the coaching intervention, which showed the largest impact in the initial evaluations. The existing Reading Support Project (RSP) was therefore modified to include selected EGRS components. The RSP has since been implemented in 164 of the original 230 EGRS schools, with a further 50 schools serving as controls.

The Setswana grade 1 letter-sound benchmark, Grade 2 fluency threshold and Grade 3 fluency benchmark are not to be viewed as aspirational goals. Rather, **they reflect the minimum level where every learner should be at a particular stage in their schooling to read successfully with meaning**.

Recommendations

This research significantly advances efforts to establish reading benchmarks and thresholds in African languages. To our knowledge, this is the first study to provide benchmarks for reading in a disjunctive orthography in a Sesotho-Setswana language[8]. The body of research will need to be extended until reading skills are benchmarked in all official South African languages. This will require ongoing reading assessment initiatives to expand the available set of data to establish new language benchmarks while validating and testing existing benchmarks with different learner samples and different reading passages.

In contributing to furthering benchmarking initiatives, all new early grade reading data collection initiatives should be preceded by rigorous instrument design and piloting processes to ensure appropriate EGRA-type instruments are administered. In this regard, this Setswana (and EFAL) benchmarking exercise has established best practices to follow through highly collaborative and iterative processes of instrument development. Furthermore, following from the Nguni languages early grade reading benchmark report (Ardington et al., 2020, 2021a), the exploratory statistical methods applied to establish thresholds and benchmarks along a reading development cline are shown here to be applicable to a Sesotho-Setswana language reading.

Moving forward, the value of early grade reading benchmarks in supporting improvements in policy and practice will be further realized once they are linked to a national programme to assess and monitor early grade reading skills (Ardington et al., 2021a:14). Early grade reading assessments (EGRA) should form a critical part of formative assessments in primary schools. The progressive roll-out of EGRA training for teachers by the Department of Basic Education in 2015 should be leveraged to promote EGRA testing in all schools (Maboya, 2020). As teachers conduct EGRA-assessments, guided by threshold and benchmarks, this will help them to identify early-on whether learners are on track, and align their instructional practice with each learner's level of reading development.

Furthermore, if national EGRA-type assessments were introduced, for example through linking this to Systemic Evaluation Programme plans at the Grade 3 and 6 level (DBE, 2020a), thresholds and benchmarks

[8] Although Spaull et al. (2020) provide some tentative benchmarks for Sepedi using a small dataset.

could be used to monitor sector progress in reading in the early grades. Yet merely measuring reading skills will be insufficient for progressive improvements. Addressing the very slow reading or non-development of decoding skills among learners requires significant action and changes in practice at various levels.

Ensuring that all teachers are equipped with the knowledge and resources to effectively teach and assess reading in home languages and English is critical. Both in-service and pre-service training programmes that promote best practice in reading instruction should continue to be tried, tested and evaluated. The growing problem of large class sizes in the early grades also needs to be addressed as individualized reading instruction or even assessment is hindered when class sizes exceed prescribed recommendations (DBE, 2020a:106). More effort also needs to be given to considering how reading can receive higher priority in homes so that children are more exposed to oral language and print at earlier ages. Children are entering school with underdeveloped emergent literacy and language skills (Dawes et al., 2017), making the work of teachers and acquiring decoding skills much harder. There is also little evidence to suggest that decoding skills are being introduced effectively in Grade R.

Finally, resources urgently need to be allocated so that reading remediation programmes are available in all schools, and that schools are equipped with enough personnel and reading resources to support not just a few learners, but many. This is particularly necessary in a context where COVID-19 schooling disruptions have impacted severely on learner's reading development, particularly in the early grades (Ardington et al., 2021b).

1 INTRODUCTION

This Technical Report proposes Setswana Home Language (HL) reading benchmarks for adoption by the South African Department of Basic Education. The report presents pertinent literature that informed the benchmarking study, it describes the benchmarking methodology in detail, and details an analysis of data collected from 225 no-fee schools in North West schools, between September and November 2021.

In the 2016 round of Progress in International Reading and Literacy Study (PIRLS) assessments, 78% of South African Grade 4 readers could not read for meaning, as opposed to 4% of Grade 4 readers internationally. When broken down into language groups, almost 90% of Grade 4 Setswana readers could not read for meaning (Howie et al. 2017). When such large numbers of children cannot understand what they read after four years of schooling, even though most children (70%) were assessed in their home language[9], action needs to be taken to understand why children cannot read and to identify what can be done to support improved reading development.

One contributing reason for poor reading comprehension in both home language and first additional language is that children have not learnt foundational skills required to be able to read and understand what they are reading. Knowledge is required of the 'code' of the language in which learners are reading, which we refer to as decoding skills. Reading comprehension assessments such as PIRLS identify higher order reading skills that many learners have not mastered. However, written language assessments such as PIRLS 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. These assessment data can be used not only to assess levels of decoding skills, but they can also be used to set contextually appropriate milestones to track children's reading development. We refer to these milestones as reading thresholds and benchmarks.

It is not sufficient, however, to merely establish standards to support reading development in home language. African HL language of learning and teaching (LOLT) is maintained until the third year (or end of the Foundation Phase) in most South African schools. Then the language of teaching and learning switches to English from Grade 4 and mother tongue is taught as a subject rather than used as the LOLT. Whereas just a quarter of learners in South Africa are taught in English in the Foundation Phase, 90% are

[9]The mismatch between being assessed in a language different from that spoken at home was largest for the English readers. In the African languages this was much lower (Howie et al. 2017).

instructed in English from Grade 4 onwards.[10] It is therefore vital that on-track reading development occurs in both home language and English First Additional language (EFAL) in the Foundation Phase. If children cannot read for meaning in home language and English by the end of Grade 3, their ability to access the curriculum (or read textbooks) is significantly impeded. Yet little guidance has been provided to South African teachers around reading benchmarks to support reading development in EFAL, despite existing evidence and theory on reading in English.

Box 1: What are reading benchmarks and why are language specific benchmarks needed?

Reading benchmarks and thresholds are numerical measures of proficiency in specific reading skills that can be used to monitor whether children are on track with their reading. Benchmarks and thresholds can inform a **shared vision** of what successful reading looks like at specific grades. They provide a **standard** against which teachers can measure learner's reading subskills and identify early on learners who are at risk of not learning to ready for meaning by age 10. This, in turn, **supports remediation** at an earlier age. They serve as a form of quality control within an education system so that large numbers of learners do not fall through the cracks (Jukes et al. 2020).

Benchmarks for reading exist in many other languages and countries, with well-defined oral fluency norms and benchmarks for early grade reading in English home language (Hasbrouck & Tindal, 2006, 2017). Due to differences in the phonological, morphological, and orthographical features of African languages, one cannot simply apply English reading benchmarks to other languages. Furthermore, across African language groups, one cannot apply the benchmarks from one language group such as Nguni languages to others such as Sotho languages. At the most basic level, it makes no sense to impose the same fluency standards 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. Despite advances in establishing benchmarks in Nguni languages, until now, detailed work on the development of reading benchmarks in Sotho languages has not been done. Furthermore, there are no existing fluency standards in the South African context for EFAL.

Source: Setting Reading Benchmarks In South Africa https://pdf.usaid.gov/pdf_docs/PA00X1NZ.pdf.

1.1 BACKGROUND

Despite the government of South Africa's (GoSA) large investment in basic education, the country continues to face challenges providing a quality education in the majority of schools and its education indicators continue to lag behind those of its peers. In international comparative reading tests, South Africa consistently performs at the bottom with nearly 80 percent of Grade 4 learners unable to read with comprehension in the language of their choice including home language (Howie et al, 2016). The GoSA considers education to be one of its highest domestic priorities and one of the greatest long-term

[10] Estimates from the Annual National Assessments of 2013. See https://www.bridge.org.za/wp-content/uploads/2016/06/SPAULL-2016-BRIDGE-reading-presentation.pdf

challenges facing the country, as is evident in the National Development Plan which states its number one objective as improving the quality of basic education (DBE, 2013).

To support the GoSA, USAID Southern Africa, awarded the PERFORMANCE Indefinite Delivery Indefinite Quantity (IDIQ) to Khulisa Management Services (Khulisa) to provide technical, analytical, advisory, monitoring, evaluation and related support services to assist USAID Southern Africa in **effectively diagnosing needs, and planning, designing, monitoring, evaluating and learning from interventions.** PERFORMANCE helps to fill a critical research gap by providing rigorous analysis in target areas related to improving the quality of language and literacy skills of primary grade learners in South Africa and the region. USAID Southern Africa found there is little data available on the impact of teacher training programs on student learning outcomes, including literacy. Additionally, there is little known about fluency and reading benchmarks for learning in African languages. PERFORMANCE aims to fill this crucial research gap with rigorous evaluations, studies and assessments. Task Order 4 under PERFORMANCE has 12 objectives, three of which relate to establishing learning benchmarks:

- Objective 6: Collect Setswana and EFAL benchmarking data in Grade 6 (and Grades 5 and 4 as may be the case in the existing sample for EGRS), leveraging the EGRS impact data collection effort. Top up the sample of learners if necessary to meet the sample size requirements.

- Objective 8: Collect Setswana and EFAL benchmarking data in Grade 3 (and Grades 2 and 1 as may be the case in the existing sample for RSP) and in Grade 2 (from the new sample), leveraging the RSP impact data collection effort. Top up the sample of learners if necessary to meet the sample size requirements set out below in the section "Evaluation design and methodology"

- Objective 11: Conduct data cleaning and analysis of all relevant data collected in Grades 1-6 to establish reading benchmarks in Setswana and EFAL, in close collaboration with the DBE research team. Provide a high-quality technical report and learning brief in which benchmarks are proposed to the DBE for adoption.

To address the challenge of children not learning to read for understanding, in 2015 the DBE initiated the first Early Grade Reading Study (EGRS I)[11] in two districts in the North West province (districts of Ngaka Modiri Molema and Dr Kenneth Kaunda). The EGRS I evaluated three Setswana HL interventions aimed

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

at improving reading in the early grades: a teacher training intervention, an on-site teacher training and coaching intervention, as well as a parental intervention. The interventions showed significant impacts on learner results when teachers benefitted from training, coaching, and provision of learning materials. In 2018, data was collected from the same sample of learners. This EGRS evaluation showed that the initial impacts of the EGRS on learners' ability to read, continued one year beyond the end of the intervention.

In 2019 and 2020, the DBE requested USAID's support in proceeding with an expansion of the EGRS. The focus of the expansion was to scale up the coaching intervention, which showed the largest impact in the initial evaluations. The existing Reading Support Project (RSP) was therefore modified to include selected EGRS components. The RSP has since been implemented in 164 of the original 230 EGRS schools, with a further 50 schools serving as controls.

The purpose of Task Order 4 is to contribute to the body of research around early grade reading interventions, to make progress towards establishing reading benchmarks in South African Home Languages and English First Additional Language

1.1.1 Aims

In this report, we use existing adapted early grade reading assessment (EGRA) type data and new data collected through the EGRS I and RSP in North West Province to establish letter-sound benchmarks and oral reading fluency thresholds and benchmarks for the early grades in home language Setswana. These data are all drawn from no-fee (Quintiles 1-3)[12] schools so that resulting benchmarks are relevant to the majority of South African learners who are attending less resourced school contexts. This research serves to complement an emerging body of research to establish early grade benchmarks in all official languages in South Africa (see Jukes et al., 2020; Ardington et al., 2020, 2021a).

For the purposes of this benchmarking exercise, we define fluency as the ability to read with speed and accuracy. Although prosody is a component of fluency, it is difficult to measure consistently and reliably in field studies and thus is not considered here. The purpose of the proposed letter-sound and fluency benchmarks is outlined below:

[12] No fee schools cannot charge school fees. These are schools in quintiles 1 to 3, the system DBE used to rate schools according to the income, unemployment and literacy levels in a community. The system is used to determine public funding to schools.

- Letter-sound benchmark - identifies whether learners are developing sufficient alphabetic knowledge that underpins foundational decoding skills necessary for accuracy in reading.

- Fluency threshold - identifies learners who are entering an emergent level of fluency which supports reading accuracy but which is not yet sufficient to read for meaning.

- Oral Reading Fluency (ORF) benchmark - identifies a lower bound fluency level that is necessary for learners to comprehend what they are reading and articulates to teachers a point at which they should concentrate on further developing comprehension skills.

Aligned to three points in the Foundation Phase, we establish a Grade 1 letter-sound benchmark, a Grade 2 fluency threshold and a Grade 3 fluency benchmark for home language Setswana speakers. For English second language speakers we establish a Grade 3 fluency threshold and Grade 5 fluency benchmark. We also work forwards and backwards from these points to establish other grade-specific minimum fluency levels in the Foundation and Intermediate Phases that act as steppingstones or further developmental goals for EFAL reading fluency in primary school. These thresholds/benchmarks should be viewed as establishing the standard or level that *every* learner should attain at grade specific points to move through a successful language and reading trajectory journey in the primary school years.

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

Following Ardington et al. (2020), our approach to setting thresholds and benchmarks integrates theoretical understandings of reading development with a non-parametric analysis of cross-sectional and longitudinal EGRA-type data. The proposed minimum grade specific EFAL fluency levels, are also examined against existing norms for English second language (ESL) speakers in other international contexts.

1.1.2 Report structure

Before detailing our empirical method and results, the next section (section 2) turns to reading development theory that guides our empirical analysis. We provide a theoretical framework for reading development, reviewing three models of early reading development to see how they account for successful reading and the processes that drive it. We also propose an overarching developmental framework of

early reading (section 2.2) against which early reading across all the official languages in South Africa can be mapped with language specific developmental differences identified. Given that reading theories pertain to both Setswana and English.

We describe our empirical approach (Sections 0) and the data used to establish benchmarks/thresholds in Setswana (Section 4 and 5). The phonological and orthographic features of Setswana are outlined (Section 4.1) and an overview is provided of research into early reading in Setswana (Section 4.2). The Setswana benchmarking results are discussed in Section 6 and summarized in Section 7.

The concluding section considers how to increase the efficacy of established reading thresholds and benchmarks through shaping improved policy and practice for improved reading outcomes.

2 THEORETICAL FRAMEWORK FOR EARLY READING DEVELOPMENT

The goal of reading is to understand or comprehend what we read. Yet children in South Africa are struggling to develop this key skill in both home language and English – a skill that is necessary for learners to be able to access the curriculum. To understand why our children are struggling to become readers we need to understand how children become successful readers and what kinds of knowledge and competencies they require to develop on the journey of being able to read with meaning.

Reading itself is a complex process. It requires the development of various skills and knowledge factors including knowledge of linguistic factors[13], text factors[14], code-based factors[15] and the development of a range of lower to higher level cognitive processes. The components needed to become a skilled reader do not necessarily develop simultaneously. Some aspects of the phenomenon may develop first, laying a foundation on which subsequent competencies are built and thrive (Stanovich 2000). It is estimated to take several hundred hours of practice over the years, through regular exposure to written language, to become a skilled reader. Additionally, the extent to which reading development occurs is likely to be supported or impeded by external factors such as socio-economic factors, the home environment, the

[13] Knowledge of the language of the text at a sub lexical (phonological and morphological), lexical (vocabulary), sentence (morphology and syntax) and discourse-level.
[14] 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.
[15] The technical features of a specific writing system and the linguistic features of the language onto which the written symbols are mapped.

schooling context, instructional practice, access to books, nutrition, health, affection, and emotional security.

In the past 70 years, scientific theories of reading have been strongly influenced by reading research centred on English and other Western European languages.[16] Yet in the past thirty years, a broader body of evidence available from different linguistic and writing systems is starting to inform current theories of reading. In particular, findings from studies on reading and particularly early reading in *transparent* orthographies and in other *agglutinating languages* (e.g., Finnish, Turkish, Basque) may be of relevance to early reading in African languages since they share common orthographic and linguistic features.

2.1 READING THEORY: THREE RELEVANT MODELS

2.1.1 Simple view of reading

The simple view of reading (SVR) posits that in order to understand a text, children need to develop in two main skills areas, namely decoding skills (knowledge of the written code) and oral language proficiency (knowledge of the language in which they read) (Gough & Tunmer 1986, Hoover & Gough, 1990). Ability in both skills is necessary for reading comprehension but each are not on their own sufficient to be able to read with understanding. Both skills are important, but in the initial stages of reading development, decoding is a critical skill that is necessary for children to read a text on their own and make sense of it. Once children have acquired relative mastery over the code and can read words accurately and fluently, then its influence diminishes and other skill areas related to language proficiency drive reading comprehension, such as vocabulary.

Basically, the SVR predicts that early skilled readers (e.g., in Grade 1) will show strong decoding skills as well as strong language proficiency. Struggling readers in Grade 1 will not have adequate decoding and/or adequate language proficiency to enable them to read and to understand a text on their own. In Grade 1 there will be variations in decoding skills between readers but by the end of Grade 3, when learners are expected to read fluently on their own with understanding, such differences should have more or less levelled off as mastery in decoding is established. From Grade 3 onwards variations in language proficiency (e.g., vocabulary differences) and other cognitive processes (such as making inferences during reading) should account for variations in reading comprehension.

[16] Including German, Dutch (Germanic language family), French, Spanish, Italian (Romance language family)

Although there is converging evidence across alphabetic languages of the need for strong decoding skills to enable comprehension, in line with the simple view of reading, there are nuanced differences as once decoding is well established, differences in language proficiency (e.g., in vocabulary or morphosyntactic knowledge) become stronger predictors of reading comprehension. For example, in Turkish and Finnish (both transparent, agglutinating languages), accuracy in letter-sounds and word reading is achieved early (end of Grade 1) and remains high (Durgunoğlu & Öney 1999; Leppänen et al. 2008), while in English this develops more slowly. In transparent codes, accuracy is reached relatively early and so decoding as measured by word accuracy is not as strongly related to reading comprehension as linguistic comprehension. However, when decoding is measured by fluency, then its importance in reading comprehension is found to extend to later grades (Florit & Cain 2011), although these patterns can be influenced through pedagogical approaches in the classroom. For example, because accuracy is achieved so early in Finnish, reading rate (fluency) thereafter distinguishes weaker readers from normally developing ones.[17]

2.1.2 The decoding threshold hypothesis

The SVR assumes a linear relationship between decoding and reading comprehension, i.e., the stronger a child's decoding skills are, the more likely they will understand what they read. More recently, Wang, Sabatini, O'Reilly, and Weeks (2019) have proposed a refinement of the SVR by arguing that the relationship is more complex. They posit a decoding threshold hypothesis, which predicts that there are 'threshold conditions' in the relationship between decoding and reading comprehension. In other words, for learners who decode *below* the threshold, reading comprehension is unlikely to develop unless decoding can be improved to a level above the decoding threshold (Wang et al. 2019). They base their claims on the analysis of very large longitudinal data sets in English (over 30,000 learners). The authors 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.

The findings from Ardington et al. (2021a) based on a large dataset of 14,000 readers in three Nguni languages found evidence that learners reading below a decoding threshold were in a non-comprehension

[17] Reading in Finnish, in all schools, is taught via a systematic phonics approach, all the letters are taught in the first term of Grade 1, teachers are well qualified, all early grade teachers are trained in the phonics approach and teach it consistently, schools are well resourced, and struggling readers are identified and remediated early (Aro 2017). See also Kendeou et al. 2013.

zone. They also find that beyond a certain point, increases in fluency do not translate into higher comprehension.

2.1.3 Orthographic depth hypotheses

According to the orthographic depth hypothesis (ODH), the reading process will develop slightly differently depending on the orthography of the language (Katz & Frost, 1992). Alphabetic writing systems can be placed on a continuum of transparency in the way in which sounds are mapped onto letters. Languages with transparent systems have a fairly regular mapping between sounds and letters (i.e., the same letter symbol always represents the same sound) whereas languages with more opaque writing systems have more irregular letter-sound mappings. For example, African languages have transparent (or shallow) orthographies, whereas English has a more opaque (or deep) orthography. Early reading success in alphabetic orthographies is expected irrespective of whether they are transparent or opaque (Alcock et al. 2010; Melby-Lervåg, Lyster, & Hulme, 2012). However, the ODH predicts a quicker process of learning to read in transparent languages than opaque orthographies. Due to regular mapping, learning to read in a transparent orthography will happen more quickly, and novice readers will rely more on phonological processing (i.e., direct phoneme to grapheme mappings).[18]

Another extension of the ODH theory, called the Psycholinguistic Grain Size Theory (PGST) predicts that reading in transparent orthographies will be easier by comparing 'grain size' across languages. 'Grain size' refers to units that readers use to decode words, which can vary from whole word, syllable, morpheme or phoneme-to-grapheme mappings. The PGST proposes that readers of transparent orthographies rely on the mainly small grain size of letter-sounds to read words while readers of opaque orthographies rely also on larger grain sizes such as rimes (e.g., *-at* as in *cat, mat, sat*), syllables or even whole words to read words because of irregularities in the phoneme-grapheme mapping system.

Scientific evidence also seems to support that reading development occurs faster in transparent languages.[19] Based on findings from other transparent languages and agglutinating languages, accuracy in alphabetic knowledge tends to occur early and mastery level (accuracy and fluency) in both letter-sound

[18] For example, children learning to read in transparent orthographies seem to do so more quickly than children learning to read in English (Seymour et al. 2003). The ODH predicts a different trajectory for children learning to read in an opaque orthography: Due to irregular mapping systems, novice readers will take longer to learn to read and will rely not only on phoneme-to-grapheme mapping but will also utilise a lexical strategy (like whole word mapping) to read words.

19 Studies have shown that in transparent languages such as Greek, Finnish and Turkish, children master the code by the end of Grade 1 and rely strongly on phonological processing to do so.

knowledge and word reading should be achieved within the foundational years of schooling. As discussed later, however, this advantage might be offset by the more complex consonant sounds that occur in African languages. Cross-linguistic details still need to be further investigated.

2.1.4 Instructional practice shapes reading development

So far, we have focused on three theories of reading. It is also important to acknowledge that early reading development across language types is likely to be strongly influenced by instructional practice, although studies seldom investigate the role that instruction plays on early reading trajectories.

Early reading instruction in English has been beset by the reading wars, where whole language approaches downplay the role of code-based knowledge and phonological processing, while the phonics approach stresses the importance of explicitly teaching children how the code works and how to blend letter-sounds in sequence so that they can decode words. Such differences in instructional context may subtly affect the early reading trajectories of English readers, for example, novice readers in whole language classrooms may exhibit stronger lexical strategies for decoding because phonological processing is downplayed. It is important, therefore, to examine early reading trajectories in light of possible influences of instructional practices.

2.2 A DEVELOPMENTAL VIEW OF MULTIPLE PROFICIENCIES

Reading development is dynamic and changes over time such that readers' profiles look different at different points in development in transparent and opaque orthographies. Different code- and cognitive-based processes dominate or diminish in importance and are superseded by others as proficiency increases. A skilled Grade 1 reader displays a different profile from a skilled Grade 4 reader (Adams 1990; Stanovich 2000; Castles et al. 2018; Kim 2020; Caravolas et al. 2021).

What this basically means is that a skilled Grade 1 reader is somewhat different from a skilled Grade 4 reader; what manifests as reading competence in the early years of schooling changes from what manifests as reading competence a few years down the line.

Yet, there is converging evidence from different languages which points to decoding accuracy developing first (e.g., in letter-sound knowledge, syllable and word reading), followed by increased processing speed, which then leads to automaticity in processing (i.e., processing without effort or conscious attention). This frees up working memory and attention for meaning construction. The arrow in Figure 1 depicts the direction of this developmental trajectory. However, the points or thresholds at which accuracy or increased alphabetic knowledge lead to automaticity in word reading (in or out of context), thereby

enabling reading comprehension, may differ across languages depending on their linguistic and orthographic features.

Figure 1: Developmental cline in early reading

The influence of some processes as drivers of reading development change as proficiency increases and their role is taken over by qualitatively different processes. Drawing on the work of Stern et al. (2018), different types of readers are distinguished on a cline, specifically because there are no clear-cut differences between them. This is described more fully in Box 1.

Box 1: Developmental continuum of reading adapted from Ardington et al. (2020)

Different types of readers can be distinguished on a continuum.

Non-readers show poor print awareness, poor phonological awareness[1] and have very little 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.

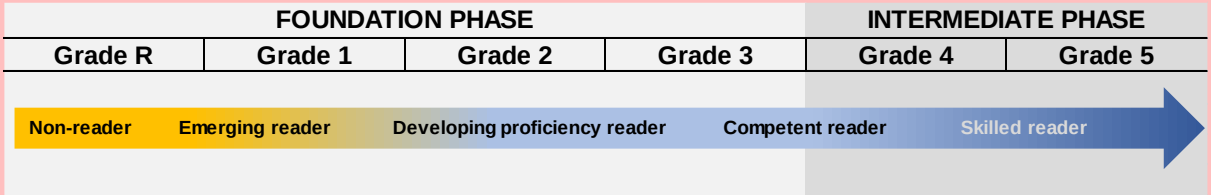
Readers in the next phase are **developing proficiency.** They have more accurate and fluent knowledge of the alphabetic code, which enables them to decode syllables and words in or out of context with greater accuracy (approximating 95% accuracy in home languages which refers to correctly reading 95 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.

The figure below loosely maps the reading skill cline against grade progression in the early school years.

Developmental continuum in early reading in relation to formal grades



By the time learners exit 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 *non-readers* or *emerging readers* at the end of grade 3. Such a developmental lag would point to challenges in pedagogy and the teaching/learning context.

The development framework emphasizes that 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. The three studies on early reading in Setswana discussed in section 4.2 show that this is not yet happening in Grade 2 or 3. This might help to understand why the PIRLS outcomes in Setswana are so low.

In the report that follows, we aim to show in greater detail what code-based thresholds support reading development and reading comprehension in Setswana as a home language.

3 BENCHMARKING METHODOLOGY

3.1 APPROACH

As explained in the theoretical overview, 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. The importance of some earlier processes diminish as proficiency increases and they are replaced by qualitatively different processes. Yet across all the components, accuracy develops first, followed by increased processing speed, which then leads to automaticity (processing without effort or conscious attention) as discussed in section 2.2. This automaticity free-ups cognitive resources (e.g., working memory and attention) to be allocated to constructing meaning from text (LaBerge & Samuels, 1974; Samuels & Flor, 1997, Fuchs *et al.*, 2001; Spear-Swerling, 2006). Following this developmental cline, our empirical work will first examine the relationship between accuracy and speed and then the relationship between fluency and comprehension.

Our approach draws on the idea of non-linearities in the relationship between decoding and reading comprehension where comprehension only occurs above a certain level of decoding proficiency (Wang et al. 2019, Kim 2017, Kim & Wagner 2014). In other words, if decoding skills are below a minimum threshold, reading comprehension remains stagnant. As mentioned above, Wang et al. (2019) speculate that there may also be an upper threshold, beyond which there are no additional gains to comprehension from increased decoding skills. Our empirical work will focus on identifying these critical threshold points in learners' reading development.

A significant body of evidence demonstrates that reading processes differ by language, with variations in the core skills employed by children learning to read (Katz & Frost, 1992; Torppa et al., 2016; Dowd & Bartlett, 2019). This calls for language-specific benchmarks. It also implies that we should be cautious not to impose assumptions about the specific nature of the speed-accuracy or fluency-comprehension relationships in Setswana in our empirical approach.

Our benchmarking approach is also mindful of how benchmarks and thresholds need to be contextually valid. They should not be set so high as to be out of reach for the majority of early grade learners. At the same time, they need to be ambitious enough to establish expectations that are sufficient to support meaningful improvements in early grade reading. They also need to be appropriate for the curriculum context and aligned to the priorities of teaching across school system phases. Our approach to setting benchmarks is anchored to the context through data, with statistical methods that also support contextually relevant thresholds/benchmarks as explained in the next section. Additionally, we have also assessed whether identified thresholds and benchmarks are contextually relevant through consultation with linguistic and curricula experts, including home language Setswana speaking literacy experts.

3.2 STATISTICAL METHOD AND SKILLS

3.2.1 Exploratory non-parametric techniques

Following Ardington et al. (2020, 2021a) and Jukes et al. (2020), we engage in exploratory data analysis, guided by reading development theory and expert linguistic knowledge. To avoid imposing *a priori* assumptions about the nature of understudied reading development in Setswana, we use non-parametric techniques to explore the accuracy-speed and fluency-comprehension relationships with the purpose of identifying critical points in learners' reading trajectories.

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. We do this 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 can be achieved 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. In other words, both backwards and forwards analysis of the data are used in a system of checks and balances to verify the results.

3.2.2 Advantages of non-parametric methods over other benchmarking approaches

Widely-used approaches to benchmarking rely on identifying a fixed comprehension threshold (e.g., at least 80% of questions correct) and then apply statistical techniques to identify the fluency levels associated with meeting that comprehension threshold (Room to Read, 2018; Abadzi, 2012; RTI, 2010). Our approach to benchmarking has some advantages over these previously used methodologies (see Ardington et al., 2021a; Jukes et al. 2020).

First, non-parametric methods make no assumptions about the speed-accuracy or fluency-comprehension relationships which can be affected by the linguistic differences between languages and pedagogy. A

pedagogy that focuses little on teaching comprehension skills can also result in lower comprehension scores in that context which are independent of students' reading skills.

Second, our thresholds are invariant to the serious challenges of establishing the appropriate level of comprehension questions. Traditional benchmarking methods assume that a fixed level of comprehension is a comparable construct across passages and languages. Our established thresholds are invariant to the challenges of cross-text comparability of comprehension questions or the impact of the placement of the comprehension questions in relation to the text. We illustrate the challenge of establishing comprehension difficulty using the Setswana data in the analysis that follows.

Third, our approach does not depend on having large samples of learners with advanced comprehension skills (Abadzi, 2012). Research on Nguni languages shows the relationship between fluency and comprehension flattens out at fairly low levels of comprehension. This suggests that beyond a certain fluency point, poor comprehension skills become the limiting factor and we may find insufficient numbers of learners achieving proficient (e.g., 80%) comprehension levels to support the identification of benchmarks. Our approach is not dependent on a fixed level of comprehension, so that low comprehension skills are not a limitation. Our identification of critical thresholds in the accuracy-speed and fluency-comprehension gradients relies on examining the full distribution of these relationships whereas traditional methods only focus on these relationships around the specific comprehension cut-off.

3.2.3 Establishing ORF benchmarks

3.2.3.1 Examining the relationship between speed and accuracy

Existing research on Nguni languages in the early grades is indicative of a non-linear relationship between reading accuracy and reading speed. Accuracy and speed initially increase together steeply but eventually accuracy does not improve as much with additional increments in speed – the relationship flattens out. The point at which this relationship tappers off can inform a reading threshold. Then examining the reading speed at which sufficient accuracy is achieved can inform what benchmark should be established.

Preliminary analysis of the wave 4 EGRS I ORF data, and pilot data, suggested a similar non-linear relationship in Setswana, however given the shorter nature of word length in the disjunctive orthography of Setswana (compared to the Nguni languages), tentative analysis suggests that the point at which 95% accuracy will be reached will be much higher than in the Nguni languages (Spaull et al., 2020). We examine this further using a larger dataset.

3.2.3.2 Examining the relationship between fluency and comprehension

In the same way that we examine the relationship between speed and accuracy, we then explore non-parametrically the relationship between fluency (a measure of both speed and accuracy) and comprehension. We seek to establish if there are regular patterns that exist in this relationship, and whether critical thresholds can be identified. For this analysis, we use samples that are allowed 3-minutes to read a passage and complete a subset of the ORF comprehension questions. Although we examine a fluency-comprehension relationship, our approach avoids benchmarking fluency to a fixed comprehension level.

3.2.3.3 Validity checks – concurrent and predictive validity

We then establish the concurrent validity of the proposed fluency threshold and benchmark by examining how they align against the performance of the same learners on written comprehension assessments. We then use longitudinal data to examine whether meeting specified thresholds or benchmarks at earlier grade points is predictive of learners' future fluency and comprehension proficiency levels. Given the low levels of reading proficiency, we use data from later grades to understand what a successful trajectory could look like, working forwards and backwards to establish the grade thresholds/benchmarks for an 'on track' successful reading journey.

3.2.4 Establishing letter sound benchmarks

To set a letter-sound benchmark in Nguni languages, Ardington et al. (2020) relied on a combination of empirical insights from a large reading dataset in Nguni languages and expert opinion of letter-sound knowledge required for mastery in decoding, grounded in the theory of reading. The authors motivate that by the end of Grade 1, learners should be reading 40 correct letter sounds per minute (clspm) or more. Like Nguni languages, Sesotho-Setswana languages are alphabetic languages. Despite differences in pronunciation, one wouldn't expect significant differences in the process of letter sound acquisition across these language groups. To find empirical support for this hypothesis, we repeat the analysis in the Nguni-language benchmarking report using Setswana data. We initially explore the relationship between accuracy and speed in sounding out letters. Then longitudinal data (data for the same learner for two timepoints) is used to examine incremental improvements in letter-sound knowledge against baseline scores to identify whether there is a point beyond which gains in letter-sound knowledge are negligible.

3.3 INSTRUMENT DEVELOPMENT PROCESS: INSTRUMENTS THAT SUPPORT BENCHMARKING

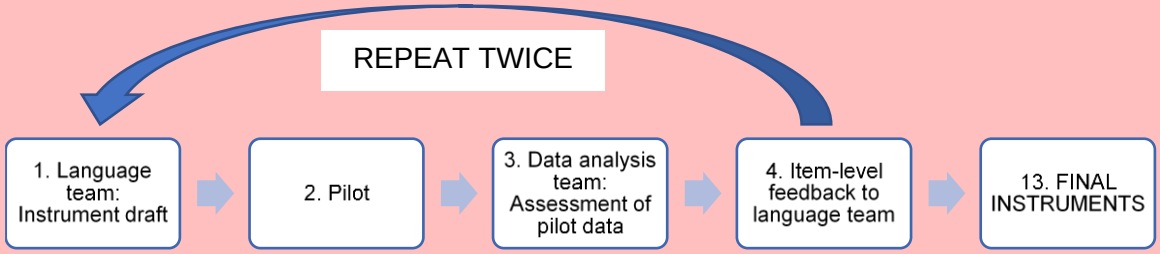
As discussed above, the choice of statistical methods and approaches used in benchmarking can shape the outcomes of that process. However, equally important is the availability of suitable assessment data with which to apply these methods.

Prior to the collection of EGRS I wave 5 data, the waves 1-4 ORF assessments were not suitable for examining the relationship between reading fluency and comprehension. To effectively model this relationship, there must be a large enough sample that read far enough in the text to be able to answer all reading comprehension questions. This requires that enough time is provided to learners to read the text, and that the text is not too long. In earlier EGRS waves (and many other reading studies) ORF assessments only allowed reading for a minute.[20] As a result too few learners had the chance to answer all comprehension questions.

Box 2: Piloting process – key to establishing appropriate language assessments to support benchmarking

To improve the validity and relevance of literacy assessments used for benchmarking processes, in 2021, Khulisa Management Services conducted three pilots to determine the appropriateness of instruments developed to assess literacy skills in no-fee schools in South Africa (see the Appendix Table A 1). After each round of piloting, the data from the assessments was analysed at the item-level and presented to a language team including linguists, DBE officials and home language Setswana experts in order to guide the iterative improvement of language instruments (or assessments). The process of development involved 13 steps as shown below.

Instrument development process



The key aims of the piloting were to make sure that 1) the instruments are set at the right level with no floor or ceiling effects, 2) the language is appropriate for the context with comprehension questions asked that are unambiguous and 3) that the length of the assessment (and specific reading passages) are appropriate for the learners and the evaluation context. In particular, key purposes of the pilots were to assess the suitability of the length of the passages of connected text used to measure oral reading fluency (ORF) and to assess the difficulty level of comprehension questions. Importantly, assessments were also designed and adjusted to account for poorer reading performance expected during a pandemic period.

The instruments[21] were piloted for data collection in 2021, which allowed for both 1-minute and 3-minute timings for the oral reading assessment. As a desired outcome of the instrument development process (and piloting iterations), most learners should be able to finish the passage within 3-minutes so they can

20 Learners are then only asked comprehension questions related to the parts of the passage that they have read within the time limit. This creates an artificial relationship between fluency and comprehension.

21 Refer to the "Report on the Development of Learner Assessment Tools and Contextual Tools on the USAID DEC (Bisgard et al., 2021)

attempt all the comprehension questions. At each stage of the piloting process (described in Box 2), the analysis team assessed passage length against the percentage of learners reading the entire passage in 3-minutes. Typically, the length of passages was found to be too long. Consequently, they were reduced with each pilot iteration and more appropriate passage lengths were finally achieved to support the availability of larger sample sizes reading the whole passage at final data collection. The piloting process was also used to assess and adjust the difficulty level of the comprehension questions asked about the passage. Pilot data was checked for irregularities in results, to highlight any instrument wording that may be ambiguous. The pilot data was also checked to ensure that the scores across all the questions aligned with expectations in terms of the comprehension processes that each question was tapping into. Various alterations were made at each piloting stage to improve the appropriateness of the comprehension questions.

In summary, three rounds of piloting integrated with detailed analytical processing of data was vital to the finalization of the assessment instruments and the quality of the data used for benchmarking purposes. The piloting process also demonstrated the efficacy of data analysts working together with linguistics and language experts for the development of appropriate literacy assessments.

BENCHMARKING READING SUBSKILLS IN SETSWANA

4 READING IN SETSWANA

4.1 LINGUISTIC AND ORTHOGRAPHIC FEATURES OF SETSWANA

Setswana is one of the 11 official languages in South Africa. It is most commonly spoken in the north-western parts of South Africa, where the country borders with Botswana. Setswana is also Botswana's national language where as many as 70% of Botswana's population speak it although there are variations across countries in how it is spoken.[22] In addition, there are small groups speaking variations of Setswana in Namibia and Zimbabwe.

Setswana belongs to the family of Sesotho-Setswana languages and is closely related to Sepedi or otherwise known as Northern Sotho (spoken widely in Limpopo Province), and Sesotho or otherwise known as Southern Sotho (spoken in Free State Province and Lesotho). The Census 2011 indicates that 2,826,464

22 https://en.wikipedia.org/wiki/Tswana_language

people speak Setswana as their first language in the North West Province,[23] and 9.4% of the South African population speak Setswana outside of households (2018).[24] As an indication of the extent to which Setswana is spoken as a home language in South African schools, at the Grade 4 level about 7% of learners were tested in Setswana in the 2016 PIRLS assessment (Howie et al., 2017).

Setswana's vowel system is relatively small and straightforward (with just 7 vowels as described in Box 2) compared to the more complex system of about 20 different vowels in English. Yet overall Setswana and other Sotho languages have a larger code set than English, comprising a large number of simple and more complex consonants, as reflected in the digraphs, trigraphs and complex consonant clusters to be mastered in these orthographies.

[23] Census in Brief updates 28 Oct 2012, https://www.statssa.gov.za/census/census_2011/census_products/Census_2011_Census_in_brief.pdf

[24] https://www.statista.com/statistics/1114302/distribution-of-languages-spoken-inside-and-outside-of-households-in-south-africa/

Box 2: Vowel set and combination vowels in Setswana

Setswana has just 7 basic vowels. By comparison Nguni languages typically have 5 vowels (Katz, 2020)

Vowels in Setswana		Example	
		Setswana	English
a	as in	rata	like, want, love
e	as in	lema	plough
ê	as in	rêma	chop
i	as in	dira	do
o	as in	motho	person/human-being
ô	as in	tôrô	dream
u	as in	khudu	tortoise

In Setswana, the circumflex sign/diacritic mark should be used to differentiate the ê from e and ô from o. The following are examples of words to be distinguished by using diacritics (Department of Education and Training, 1988:6):

pholo (ox), phôlô (health) and pholô (harvest)
lema (plough) lêma (spoil a child or shape horns)

Setswana unlike English, does not have diphthongs, but has a combination of some basic vowels and consonants as follows:

Combination Vowel	Example	
	Setswana	English translation
ae	mae	eggs
ao	maoto	feet
êi	êiye	onion
ia	diatla	hands
oa	boatla	careless
oê	Mokoêna	A common surname

Setswana also has the following semi-vowels:

w as in wena (you), bolawa (be killed)

y as in yo (this one referring to personal class only) and

y as in ya (to go)

Because African languages have largely transparent orthographies, this should in principle confer an advantage on learning to read in African languages as proposed by the ODH theory. However, this advantage might be offset by the more complex consonant sounds that occur in African languages. For example, Table 1 shows the simple consonants in Setswana while Table 2 provides a list of complex consonants in Setswana. The complex consonants consist of two, three (or even four consonants) and need to be pronounced as a blended sound (as in ngw and tshw). These occur in both Nguni and Sesotho-Setswana languages (Katz, 2020). As with Nguni languages, knowledge of these consonant clusters is a key foundational reading skill that should be mastered in the early grades. For this reason, in the analysis section, we report on knowledge of complex consonants among EGRS I and RSP learners.

Table 1: Simple Consonants in Setswana

Simple Consonant	Sounds like			Example	
				Setswana	English translation
b	b	in	brother	baba	enemies
d	d	in	drain	dira	work
f	f	in	after	fêpa	feed
g	g	in	gate	gagaba	crawl
k	k	in	Kampala	kala	branch
l	l	in	lap	lema	plough
m	m	in	madam	madi	blood/money
n	n	in	snare	nare	buffalo
p	p	in	pink	pina	song
j	j	in	Johannes	ja	eat
r	r	in	drill	rata	love
s	s	in	sing	sila	grind
t	t	in	Afr. taal	tiro	job
h	h	in	hoot	hutsa	curse
w	w	in	wet	wela	fall into
y	y	in	yell	bolaya	kill

Notes: Previously, Setswana did not have consonants such as c, v, x and z. More recently the Setswana National Language Body has agreed to incorporate these consonants when writing and reading, for example: v as in vene and not bene (van in English), vote as in voutu and not boutu; x as in nxae instead of ntlae (pardon), nce-nce-nce (sound of clock or watch ticking) and z as in zama-zama (from Nguni meaning to try) and not 'sama-sama' in Setswana

Table 2: Complex Consonants in Setswana

Complex consonant	Sounds like			Example	
				Setswana	English translation
ng	ng	in	ring	ngaka	doctor
kg	g	in	-	kgomo	cow
kh	k	in	king	khiba	apron
ph	p	in	path	phala	impala
šw	sh	in	shilling	mašwi	milke
th	t	in	term	thata	strong
tl	tl	in	butler	batla	seek
ts	ts	in	rats (Afrikaans)	tsela	path
ny	ny	in	canyon	nyatsa	despise/undermine
tš	tj	in	tjank (Afrikaans)	ntšwa	dog
tŝ				botŝarara	sourness
nn				monna	man
rr				rra/rre	my father
mp				mpa	stomach
tlh				tlhako	hoof
tsh	ts	in	rats (English)	tshaba	run away
tšh	ch	in	child	setšhaba	nation
tshw				tshwana	same as
mph				mpha	give me
ngw				ngwana	child

In contrast to Nguni languages, all the Sesotho-Setswana languages are written disjunctively. By example, the phrase *O ka tsamaya* (You may go) in Setswana would be written conjunctively in isiZulu as *Ungahamba*. For this reason, we expect that benchmarks and thresholds for early grade reading in Setswana will exceed those established for Nguni languages (Ardington et al., 2020, 2021a). This is also implied in the work of Spaull, Pretorius and Mohohlwane (2020) that suggest tentative benchmarks in Sepedi (a Sotho language) that are higher by more than double the number of correct words per minute compared with isiZulu benchmarks.[25]

25 In Spaull, Pretorius and Mohohlwane (2020) the isiZulu learners reading at 21 cwpm or faster read with 95% accuracy or higher. In contrast, 95% accuracy is associated with reading at 51 cwpm or faster in Sepedi (Northern Sotho).

Although the disjunctive orthography of the Sesotho-Setswana languages yields many single syllable 'V' or 'CV' morphemic word units (e.g., in Sepedi - *a, o, ka, ke, sa, se, ga, go)* there are also longer multisyllabic words (e.g., in Sepedi - *botlhabatsatsi, ditlhatlhagangwa)* and many syllables within words that display strong visual similarity and are therefore more difficult for novice readers to tell apart (Land 2015; Pretorius 2018). These linguistic and orthographic features might delay mastery in decoding skills. This can be exacerbated if early reading instruction is not well taught; if children are left to their own devices to figure out the complex code; and if they are not given opportunities to read on a regular basis to practice their fledgling decoding skills.

Another key feature of Setswana is that it is a tone (or otherwise known as tonal) language, which makes use of high (H) and low (L) tones. Tones can alter the meaning of a word or expression completely. For example: *nòkà* (waistline), *nókâ* (putting salt in) and *nókâ* (river) appear similar but differ in the use of tones which change the meaning of the word. Tonal differences may also confuse readers because although Setswana is an alphabetic language, the pronunciation of the letters, especially consonants, are slightly different from English. For example, 'b' is pronounced 'bee'; 'm' is pronounced 'mmm'; and 'k' is pronounced as 'ka'. Due to the challenge of measuring learner's correct use of tones in EGRA-studies this is not considered in the analysis sections.

4.2 Review of studies of early reading development in Setswana

The body of reading evidence on African languages, and particularly Setswana remains limited. When considering how reading develops in transparent, agglutinating African languages, it is prudent to be mindful of the sources from whence our knowledge of reading derives. When applied to different languages and contexts, claims about how reading develops as derived from other language contexts must be tested. Through evidence-based testing of claims, we can gain a deeper understanding of what is generic about reading across different linguistic families and orthographic systems and what is more nuanced and language specific, and to draw instructional implications from such findings.

Three existing studies of early reading development in Setswana are discussed here. We examine whether the findings support or dispute any of the claims made about the role of phonological processing in decoding, the role of decoding and language proficiency in early reading, and the role of orthographic depth and grain size.

A study by Lekgoko and Winskel (2008) compared the relationship of phonemic awareness and letter knowledge to word reading of 35 Grade 2 children[26] who had learned to read in Setswana (Grade 1) and then in English (Grade 2) in two rural state schools in Botswana. The children were tested on four measures of early reading in both languages as shown in Table 3. The study found that performance on all four measures of foundational reading in Setswana was low even though the children had been in school for 1½ years. Children performed comparatively better in phonemic awareness and letter-sound knowledge in English.

Table 3: Findings from Lekgoko & Winskel (2008). Small study of Grade 2 children in Botswana

	Setswana	*English*
phonemic awareness (phoneme deletion)	4.6	6.5
letter-sound knowledge	11.3	22.4
word reading (using number words)	28.6	
nonword reading[27]	27.2	
Sample size	35	

Notes: Of the original 50 children included in the sample, 28% (14) were excluded because of floor effects - they did not have letter knowledge and could not read at all. The tests were adapted according to the linguistic and orthographic features of each language.

The findings also pointed to a complex pattern of within-language relationships. For example, phonemic awareness correlated moderately with word and nonword reading in Setswana (.45** and .46** respectively) but letter-sound knowledge did not. Word and nonword reading tasks were extremely highly correlated (.99**) in Setswana but not in English, which suggests a similar strategy in word reading in Setswana, a finding which concurs with evidence from languages with more transparent mapping between phonology and orthography. Phoneme awareness accounted for a significant proportion of unique variance, for both word and nonword reading (22% nonword reading, 24% word reading), but letter-sound knowledge did not. The results of the study confirm other findings that show that children with good phonemic skills learn to read more easily than children with poor phonemic skills. Phonemic awareness contributes to decoding skills.

[26] The participants were assessed midway in Grade 2 and had a mean of 8.4 years of age.
[27] The use of number words to assess word reading may possibly not yield as robust results as a word list comprising random high frequency words or increasing word length.

The learners knew double the number of letters in English than in Setswana (22.4 vs 11.3), and in English letter knowledge accounted for a significant proportion of unique variance for both word reading and nonword reading (55% and 26% respectively). This suggests support for the decoding threshold hypothesis, which predicts that a relationship does not obtain when performance is below a threshold level. Knowing only 11 letters in Grade 2 in Setswana seems too little to productively support the development of decoding skills.

While this is a small-scale study and caution is advised, the results generally confirm what previous research has shown about the development of different decoding skills in early reading trajectories across transparent/opaque orthographies. The study does not explain how reading is taught across languages in Botswana's schools, but the low reading performance suggests that instructional practices are not developing strong foundational decoding skills in either language.

In a study by Malda et al. (2014), the researchers were interested in comparing similarities/differences across opaque (English) and transparent orthographies (Afrikaans and Setswana). The reading abilities of 358 Grade 3 learners were profiled in English, Afrikaans and Setswana, using a battery of tests that tap into different linguistic, cognitive and code-based components of reading to identify strengths and weakness in reading development. Only the Setswana results are discussed here (n=109 learners).

The overall findings indicated that the pathways between cognitive and reading skills were similar across the three orthographies, which suggest that the "main road to reading" is the same for children learning to read in alphabetic languages, irrespective of orthography. However, in line with theories about orthographic depth, differential effects were found in terms of the strength of the association between the sets of skills. Overall, the scores on the various reading measures were lowest in the Setswana group. For example, the mean for reading comprehension was low (7.7, compared to 12.7 and 17.6 in English and Afrikaans respectively), while the mean ORF score was 37.4 cwpm, indicating slow reading. As in Lekgoko and Winskel study, although performance was better on word than nonword reading, robust correlations were found between these two decoding measures (.76**), suggesting reliance on phonological processing during decoding.

As in the previous study and in line with theories of orthographic depth, phonological awareness played a larger role in the more transparent orthographies of Afrikaans and Setswana, while vocabulary and working memory seemed to play a stronger role in reading in English.

Although not tested directly in the study, the findings suggest support for the simple view of reading (SVR), which predicts that reading comprehension is compromised if decoding skills are not well developed. The

low foundational decoding skills in Setswana are noteworthy, where both word and ORF scores showed low correlations with reading comprehension, but not in English and Afrikaans, where the correlations were higher.

The overall low reading scores in Setswana and low correlations between the component measures also tentatively suggest support for the decoding threshold hypothesis – relationships between decoding and reading comprehension do not obtain when children read below a threshold level. Although there were no norms or benchmarks for word reading in the Sesotho-Setswana languages at the time of the Malda et al. study, reading at 37 cwpm in Setswana in Grade 3 seems excessively slow for a disjunctive orthography and suggests that foundational decoding skills were not well developed in this group of learners. The authors suggest that one of the reasons for poor foundational reading skills "could be that reading instruction is not following an appropriate phonics approach, which may delay the point where decoding skills are no longer sources of individual differences" (p42). The results from the current Setswana benchmarking project show that reading at 37 cwpm is below the minimum threshold of 40 cwpm in Grade 2, below which learners struggle to read with meaning.

In the analysis section we report on longitudinal Setswana data from all waves of the EGRS I study in North West Province, at five different time points. But given that data from the earlier waves is documented (DBE, 2017), we note the following main findings from the earlier EGRS I waves. While these results do show a slow improvement of foundational reading scores over two years of Foundation Phase schooling, the reading trajectory in Setswana seems inordinately slow, for example, by the end of Grade 1 learners only know on average 23 clspm (albeit better than the 11 letters by mid-Grade 2 in the Botswana study), can only read on average 6 cwpm, and have an average ORF of 8 cwpm; by the end of Grade 2, this increases to an average of almost 40 clspm, word reading of 19 cwpm, and an ORF of 25 cwpm. With such poor foundational decoding skills, it is perhaps not surprising (as predicted by the SVR) that by the end of Grade 2 Setswana readers could still not read for meaning.

An interesting finding from the EGRS I report (DBE, 2017) on correlations between subtests possibly provide indirect evidence in support of the decoding threshold hypothesis. Correlations between the subtests were found to be much lower at baseline than at Wave 2[28], and the composite score at baseline shows a much lower correlation to Wave 2 (0.25) and Wave 3 (0.22) than the high correlation of 0.72 between Waves 2 and 3. Although this is attributed to the improved quality of fieldwork during Waves 2

[28] No correlation between subtests provided for Wave 3

and 3, these low correlations may also provide indirect support for the threshold hypothesis – the early reading scores are still too low to reflect a relationship. Although we qualify that this finding could also be driven by floor effects with many learners scoring zero at baseline. By implication, it means there is no way of discriminating between learners at baseline, resulting in low correlation with later measures.

Even though not much research has been done on early reading in Setswana, the evidence from these three studies enables us to identify the following preliminary trends:

- Children who learn to read in Setswana as a home language seem to do so in ways that are not dissimilar across other alphabetic languages, especially transparent orthographies.

- As in other transparent alphabetic orthographies, phonological processing is important in early Setswana reading and shows a relationship to word reading.

- As in other transparent alphabetic orthographies, phonological processing in word and nonword reading is similar, suggesting a phonological processing route during decoding.

- All three studies indicate that foundational decoding skills, as reflected in phonological awareness, letter sound knowledge and word/nonword reading are low, suggesting that children who learn to read in Setswana as home language are getting off to a slow start. Alphabetic knowledge is low and word reading in and out of context is low and slow, suggesting that decoding skills have not yet been developed to the point where they productively support reading comprehension, i.e., the automaticity needed for fluent reading has not yet developed. This has implications for the way in which early reading instruction is taught in classrooms.

All these findings provide direct or indirect evidence that largely support the claims about early reading development in alphabetic writing systems as expressed in the SVR, decoding threshold and orthographic depth models of reading.

None of the studies assessed differences in alphabetic knowledge according to simple/complex consonants, nor did they assess syllable reading. The studies also did not explore the relationship between accuracy and speed in foundational decoding skills over time, threshold levels, and the role of language proficiency in early reading has not been extensively examined. These are avenues that await further research.

5 SETSWANA DATA

5.1 BACKGROUND

In 2015, the EGRS I in North West province started tracking the reading outcomes of a Grade 1 learner cohort in no-fee schools in two North West districts to support an impact evaluation of teacher and coaching interventions aimed at improving the teaching of Setswana. As the EGRS intervention ended, the RSP commenced in a subset of the original EGRS I schools in 2018 leading to further data collection.

Although RSP and EGRS I data collection was originally intended to identify the impacts of these reading interventions, the most recent 2021 round of data collection was also designed with benchmarking in mind. Both sampling and instrument development processes were adjusted to support benchmarking. Although longitudinal samples are the samples of interest for the RSP and EGRS I impact evaluations, samples were 'topped-up' and additional grades that were not exposed to reading interventions were also assessed.

Table 4 distinguishes learner and sample sizes across data collection waves for the EGRS I and RSP study used in this report. It highlights the grades and terms at which learners were tested. With five rounds of EGRS I data collection and two rounds of data collection for RSP, data on reading assessments in Setswana are available for Grades 1 to Grade 7 learners over the years 2015 to 2021. When considering all sample sizes across waves, this is the largest existing dataset to our knowledge on reading outcomes in Setswana, with assessments for almost 16,000 unique learners from 230 schools. Of this total sample, different sub-samples are used for different aspects of the benchmarking analysis.

Table 4 also distinguishes longitudinal samples from additional samples that were added to support benchmarking in 2021. The benchmarking samples boost our ability to detect fluency-comprehension relationships, while the longitudinal components of the EGRS and RSP data collection allow us to assess the predictive validity of proposed threshold and benchmarks.

Although RSP and EGRS I include control and intervention samples, the benchmarking analysis is not concerned with disaggregating results by intervention status. The treatment and control samples are pooled to increase sample sizes and maximize the possible distribution of reading scores.

Table 4: EGRS I and RSP – Full sample sizes by year, term, and grade across different study components

		2015 Term 1	2015 Term 4	2016 Term 4	2018 Term 3	2021 Term 3
EGRS I	Original evaluation sample tracked	Grade 1 *nl = 4538* *ns = 230*	Grade 1 *nl = 4162* *ns = 230*	Grade 2 & repeaters *nl = 3777* *ns = 230*	Grade 4 & repeaters *nl = 3422* *ns = 225*	Grade 7 & repeaters *nl = 2363* *ns = 214*
	Benchmarking 'top-up' sample					Grade 7 *nl = 891* *ns = 181*
	Sustainability sample*				Grade 3 *nl = 2133* *ns = 212*	
RSP (subset of EGRS I schools)	Original evaluation sample tracked				Grade 1 *nl = 4201* *ns = 212*	Grade 4 & repeaters *nl = 2534* *ns = 206*
	Benchmarking 'top-up' sample					Grade 4 *nl = 752* *ns = 157*
	New benchmarking sample					Grade 3 *nl = 3336* *ns = 209*

Notes: *This sustainability sample was added in EGRS I in 2018 to assess whether teacher coaching and training impacts were sustained a year after the intervention ended. ^Repeaters refer to original Grade 1 cohort learners tracked into subsequent waves but have repeated years and are thus in earlier grades than most of the original cohort. nl = number of learners; ns = number of schools.

5.2 SETSWANA SAMPLE CHARACTERISTICS

The EGRS I learner samples are not designed to be representative of all early grade readers in Setswana. Nevertheless, the sample is large enough to provide a good indication of Setswana reading skills of learners in no-fee (quintile 1-3) schools in South Africa. The learners are all located in no-fee schools, situated in two districts in North West Province.

Roughly 77% of the original EGRS I schools are in rural areas, and while half are categorized as the poorest quintile 1 schools, a further 28% are quintile 2 and 23% are quintile 3 schools (see Table 5). For the available learner samples assessed from the original EGRS I cohort, and the RSP cohorts in 2021, Table 6 identifies what percentage of these learners are female and for what percentage Setswana is their home language. Regardless of which sample is considered, for at least 91% of the learner sample, Setswana is their home language. The gender balance is slightly skewed towards boys than girls. However, this is unlikely to affect the benchmarks selected as they are not drawn from averages but from examining a full distribution of reading outcomes.

Table 5: School level characteristics of the EGRS I original school sample (2015)

Characteristics	% of schools
Rural	77%
Quintile 1	49%
Quintile 2	28%
Quintile 3	23%
Observations (N)	230

Table 6: Learner level characteristics of EGRS I and RSP samples at first wave or 'baseline'

		Female	Setswana home language
EGRS I	Original evaluation sample tracked	43%	93%
	Benchmarking 'top-up' sample	45%	92%
	Sustainability sample*	49%	N.A.
RSP	Original evaluation sample tracked	47%	92%
	Benchmarking 'top-up' sample	46%	90%
	New benchmarking sample	50%	93%

Notes: *This sustainability sample was added in EGRS I in 2018 to assess whether teacher coaching and training impacts were sustained a year after the intervention ended.

5.3 SETSWANA ASSESSMENTS

Across RSP and EGRS I data collection waves, several decoding processes were assessed using a battery of tests. This report is primarily focused on setting fluency and letter-sound knowledge benchmarks. Thus, we focus predominately on the outcomes of oral reading fluency and letter-sound knowledge assessment tasks.

Before we summarize performance levels in these two reading components, we briefly explain these assessment tasks.

5.3.1 Setswana letter sound knowledge assessments

Timed letter-sound knowledge assessments are available for Grades 1 to 4 learners across different waves of RSP/EGRS I as seen in appendix Table A 2. Across years, the letter-sounds tested were very similar up to the 20th item. In some years, knowledge of complex consonant clusters and diacritics were tested

within the letter-sound knowledge assessment. In 2021, knowledge of complex consonants (with two up to four complex consonant sounds) and diacritics were separated out as a distinct assessment from testing simple vowel and consonant letter sounds.

5.3.2 Setswana oral reading fluency, oral reading comprehension and written assessments

Setswana assessments in oral reading fluency, with related ORF comprehension questions, are available at the Grades 1-4 and Grade 7 level. In Grades 1-4, the ORF passages assessed were narrative texts. Narrative and informational text passages were developed for assessing Grade 7 learners in 2021.

Although similar stories are often used across grades and years, word length and difficulty tend to differ even where the same story is used as seen in Table 7. In the most recent 2021 data collection, more than one Setswana ORF passage was developed and assessed per grade. With two passages, it is possible to determine how the level of the accuracy-speed relationship or fluency-comprehension relationship may be affected by the nature and difficulty of the passage (and related comprehension questions).

For almost all the ORF assessments shown in Table 7, comprehension questions are asked in relation to the text read. Written comprehension assessments were also administered to Grade 3 and 4 learners in 2018 and to Grade 4 and 7 learners in 2021. Compared to 2018, learners were given more time to complete the written comprehension in 2021, increasing the available variation in written comprehension scores that can be used for analysis.

Table 7: Setswana oral reading fluency and written comprehension assessments

Study details			Oral reading fluency				Written comprehension		
Grade	year	Term	Time allowed	Passage description	Max possible words	N compre-hension questions	Time allowed	Passage description	N compre-hension questions
1	2018	III	60s	Narrative passage A	62	6			
1	2015	IV	60s	Narrative passage A	62	0			
2	2016	IV	60s	Narrative passage A	66	4			
3	2018	III	60s	Narrative passage B	159	8	12 mins	Narrative passage C	7
3	2021	III	60s+ 120s	Narrative passage A	58	7			
			60s+ 120s	Narrative passage D	64	6			
4	2018	III	60s	Narrative passage B	159	8	15 mins	PIRLS literacy passage	7
			60s	Narrative passage C	217	N/A			
4	2021	III	60s+ 120s	Narrative passage B	130	8	15 mins	Narrative passage C	7
			60s+ 120s	Narrative passage D	118	7			
7	2021	III	60s+ 120s	Informational text A	190	7	25 mins	PIRLS literacy passage	7
			60s+ 120s	Narrative passage E	261	8			

In the development of ORF and written comprehension questions, and in the analysis of these data, close attention was paid to which types of comprehension processes involved in meaning construction and integration were being assessed. In the instrument development stage, questions were classified using a conceptual framework from the PIRLS. 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 operationalized through 4 types of questions: literal questions; inferential questions; integrative questions; and evaluative questions. Although there is variation in the type of comprehension questions asked, questions most commonly test lower levels of reading comprehension (for example, literal or straightforward inferences).

5.4 SETSWANA READING NORMS: LETTER-SOUND KNOWLEDGE, COMPLEX CONSONANTS AND FLUENCY

For learners attempting the letter-sound and oral reading fluency tasks, average performance levels on these decoding skills are summarized in Table 8. Table 9 then presents results from the complex consonant/diacritic assessment added in 2021. In addition, Figure 2 illustrates the percentage of learners not reading one word correctly from a connected text (i.e., ORF is zero). The tables and figure reveal some clear patterns and trends.

Table 8: Mean ORF scores and letter-sounds per minute in Setswana, EGRS I and RSP data (2015-2021)

Grade	Term	Year	Passage	Oral reading fluency (ORF) (cwpm)		Letter sounds per minute (clspm)	
				N	Mean	N	Mean
1	1	2015				4 452	4.9
1	3	2015	1	4 131	4.9	4 159	15.1
1	4	2018	1	3 551	9.1	4 088	22.8^
2	4	2016	1	3 712	26.0	3 712	40.2
3	3	2018	1	2 107	38.3	2 112	42.3^
3	3	2021	1	3 209	28.3	3 305	45.6
3	3	2021	2	3 190	32.6		
4	3	2018	1	3 287	48.4	3 295	41^
4	3	2021	1	3 230	35.7	3 256	45.9
4	3	2021	2	3 214	32.7		
7	3	2021	1	3 219	54.5		
7	3	2021	2	3 211	82.0		

Notes: Sample includes learners who cannot read one word. ^Letter-sound assessment in 2018 included complex consonants.

Table 9: Letter-sound knowledge vs. knowledge of complex consonants and diacritics, EGRS I and RSP 2021

	Single Letter sounds		Complex consonants and diacritics	
	% scoring zero	correct letter-sounds per minute (clspm)	% scoring zero	correct items per minute
Grade 3, term 3 of 2021	4	45.6	19	19.0
Grade 4, term 3 of 2021	2	45.9	15	22

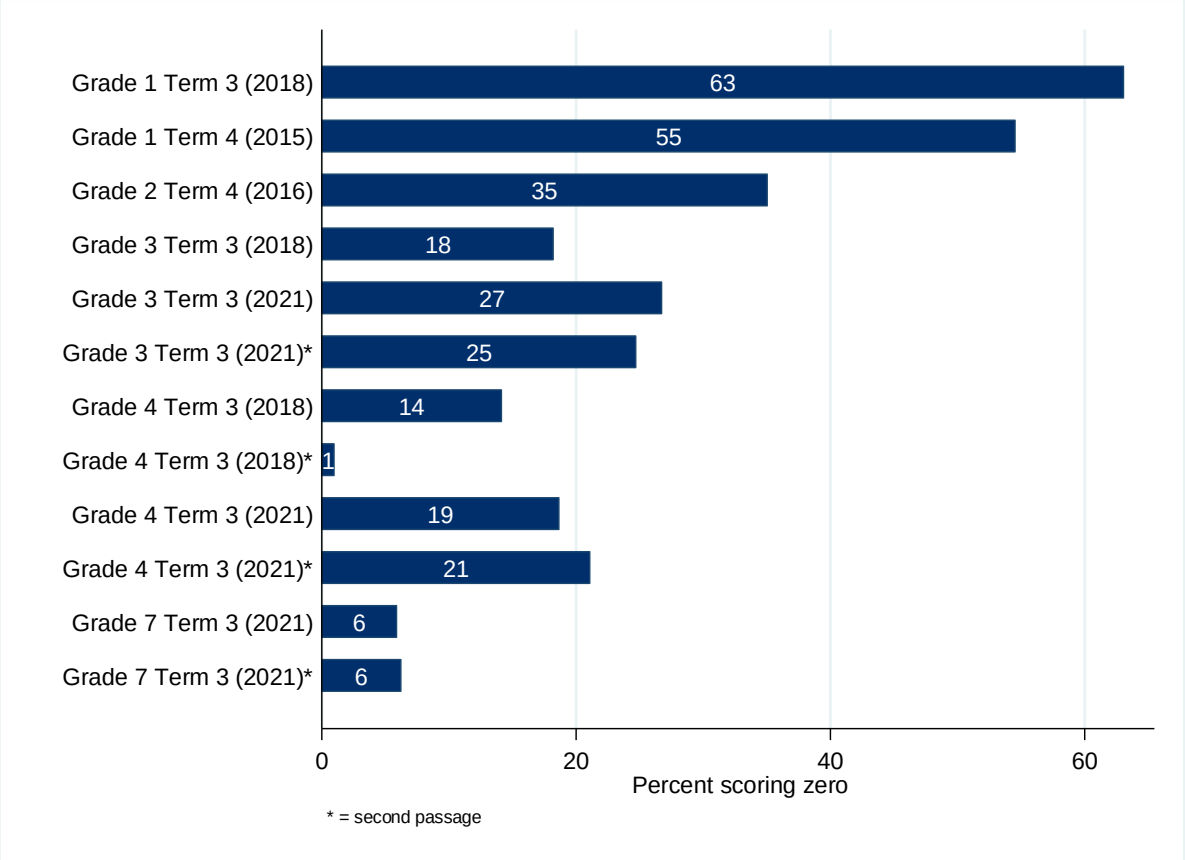
We observe the following patterns in letter-sound knowledge and knowledge of complex consonants and diacritics:

- Far too many learners are entering Grade 1 with very limited letter-sound knowledge, despite having been in Grade R, with letter-sound knowledge averaging 5 clspm at the start of Grade 1.
- However, by the end of Grade 2, learners know on average 40 clspm (in line with benchmark as explained later).
- Very little improvement in letter-sound knowledge is observed beyond Grade 2, which is expected as letter-sound knowledge is a constrained skill (because there are a limited number of alphabetic symbols to learn).
- Learners find reading single letter-sounds much easier than reading complex consonant clusters and diacritics (see Table 9). By term 3, learners in Grade 3 and 4 correctly sound less than half as many complex consonant clusters / diacritics as they do single letter sounds. Of Grade 4s in 2021

(who are probably performing at a Grade 3 pre-pandemic level), 15% cannot correctly sound one complex consonant cluster or diacritic, despite this being a fundamental decoding skill in Setswana.

- These trends are roughly in line with what is observed among Nguni language samples from no-fee schools (see Ardington et al., 2020).

Figure 2: Percentage of learners scoring zero on ORF assessments in Setswana, EGRS I and RSP samples (2015-2021)



The oral reading fluency outcomes in Table 8 and Figure 2 reveal interesting patterns in reading development:

- Among the Setswana samples, the ability to read connected text from a passage is only emerging at the end of Grade 2 with average fluency at 26 cwpm and the percentage of non-readers (i.e., ORF is zero) at 35%. This compares to reading less than 10 cwpm on average at the end of Grade 1, where 55% are non-readers.
- As expected, the percentage of non-readers declines with each grade suggesting fluency continues to develop into the Intermediate and Senior Phase. At the end of Grade 7 learners are reading on

average 55 – 82 cwpm depending on the difficulty of the passage (and non-reader percentages decline to around 6%).

- Although letter-sound knowledge and fluency in home language improves across grades, these skills develop too slowly. These skills are not being mastered by the end of the Foundation Phase when the curriculum requires that learners can read for meaning not only in their home language but in English. This is consistent with the evidence discussed earlier in section 4.2. By the end of Grade 3, 18-27% of learners still cannot read one word correctly from a Setswana text, and this does not improve much by the end of Grade 4 (14-19% non-readers in 2018). Later, we suggest 60 cwpm as a Setswana fluency benchmark that should be met at the end of the Foundation Phase. By comparison, at the end of Grade 3, Setswana learners in a pre-pandemic period were on average reading at half that pace between 33 and 38 cwpm.
- The results are suggestive of COVID-19 disruptions to reading development. Grade 3 ORF averages in 2021 are lower compared to 2018. Furthermore, Grade 4 ORF averages in 2021 are slightly *lower* than Grade 3 ORF averages in 2018 despite administering virtually the *same* passage (36 vs. 38 cwpm). A more detailed analysis of learning losses, however, is beyond the scope of this report.

5.5 DATA SUB-SAMPLES USED TO ASSESS THE FLUENCY-COMPREHENSION RELATIONSHIP

For establishing reading thresholds and benchmarks, sub-sets of the data in Table 4 are used. To examine the fluency-comprehension relationship, we only use the 2021 EGRS I data where learners are given 3-minutes to complete a passage. This analysis sample is further restricted to learners who read far enough to attempt all (or a subset) of the ORF comprehension questions. This is necessary to avoid detecting a mechanistic or data-imposed relationship between fluency and comprehension, where learners are only asked comprehension questions relating to the parts of the passage that they have read.

Longer passage lengths at the Grade 4 and 7 level compared to the Grade 3 level, made it harder for learners to complete the passages in 3-minutes. For this reason, in Grades 4 and 7 a sample that complete a subset (at least 5) of the comprehension questions is considered. Of learners attempting all or a subset of questions, Table 10 shows what percentage read at least one word correctly, their average comprehension scores are shown as well as the percentages scoring over 80% or 60% correct for comprehension. Large proportions of learners (ranging from 72-80%) in Grades 3 to 7 attempt a subset of the ORF comprehension questions and correctly read at least 1 word in Setswana from a connected text (see second panel of Table 10). There is therefore enough variation to detect patterns in fluency and comprehension.

Although one may expect comprehension scores to rise by grade, this is not the case among those who are able to read far enough in the passage to complete the comprehension questions. This happens because comprehension scores are dependent on the difficulty level of the comprehension questions asked and how they relate to the passage. This is discussed further in Box 3, of section 6.2.

Table 10: EGRS I/RSP 2021 samples used to assess the fluency-comprehension relationship in Setswana

	ORF text and word count	ORF > 0	Learners attempting all questions				Learners attempting subset of questions			
			% of learners with ORF > 0	Mean compre-hension score (%)	% scoring 80%+ for compre-hension	% scoring 60%+ for compre-hension	% of learners with ORF >0	Mean compre-hension score (%)	% scoring 80%+ for compre-hension	% scoring 60%+ for compre-hension
Grade 3 Passage 1	58	2,349	77%	55%	26%	46%	77%	55%	26%	46%
Grade 3 Passage 2	64	2,401	80%	47%	28%	42%	80%	47%	28%	42%
Grade 4 Passage 1	130	2,626	48%	50%	19%	45%	76%	56%	33%	50%
Grade 4 Passage 2	118	2,534	58%	32%	5%	14%	80%	36%	11%	24%
Grade 7 Passage 1	190	3,028	52%	32%	6%	14%	72%	36%	14%	32%
Grade 7 Passage 2	261	3,009	46%	45%	13%	36%	72%	48%	18%	34%

6 BENCHMARKING RESULTS: SETSWANA

The benchmarking results presented in this section provide more in-depth analysis of ORF, comprehension skills and letter-sound knowledge and the interrelationships between them. In the analysis that follows, letter-sound knowledge is measured as the number of correct letters sounded per minute while oral reading fluency is measured as the number of correct words sounded per minute from a connected text.

6.1 ESTABLISHING A FLUENCY THRESHOLD AND BENCHMARK: AN ANALYSIS OF READING SPEED AND ACCURACY

Following the developmental cline of reading (as seen in Figure 2 in Section 2.2), accuracy moderates the relationship between speed and comprehension. If errors are made when reading, this reduces reading speed and clutters working memory. For this reason, our analysis commences with a focus on reading accuracy, because this needs to develop before speed and automaticity improves. In the analysis that follows:

- 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.

Since accuracy in reading in transparent languages should occur more easily than in opaque languages such as English, we set the point defining a sufficient accuracy level quite high at 95% (i.e., getting 95 of every 100 words attempted correct) which follows Betts (1946).

The relationship between speed and accuracy is displayed using locally weighted polynomial regressions in Figure 3. Across all grades and reading passages, a consistent pattern emerges where initially accuracy and speed increase quite steeply together, but then the relationship flattens off. Typically, this flattening occurs when accuracy levels reach around 95% (shown by a red horizontal line in the figure). For example, at the end of Grade 3 in 2018, Setswana learners who are attempting around 8 words per minute are getting every 2nd word wrong. Yet when attempting 20 words per minute, they get every 4th word wrong. Accuracy and speed improve rapidly together and Grade 3 Setswana learners who reach 95% accuracy are typically reading at a speed of around 44 words per minute. The speed at which average accuracy reaches 95% ranges between 36 and 61 words per minute across all grade samples.

We next investigate the distribution of reading speed among accurate readers and by contrast among non-accurate readers. This is depicted in box and whisker plots of the distribution of words attempted for the sub-sample of accurate readers (who read with at least 95% accuracy) shown in Figure 4 and inaccurate readers (who read at less than 95% accuracy) as shown in Figure 5. The lower and upper edge of the boxes represent the 25th and 75th percentile of the distribution respectively – i.e., 50% of each sample learners have reading speeds in this band. The median is indicated by the horizontal line within the box. The figures include grey dashed reference lines at 40 and 60 words per minute.

Amongst accurate readers, learners in Grade 1 tend to read considerably slower than learners in higher grades which is consistent with a developmental cline. From the end of Grade 2, at the 25th percentile, reading speed tends to lie around or above 40 words attempted per minute. This means that there are very few accurate readers who read slower than 40 words per minute who are reaching accuracy levels of 95% or higher. By contrast, in the analogous Figure 5, showing reading speeds for learners that do not achieve 95% accuracy, there are almost no readers in the Foundation Phase with poor accuracy who are managing to read at speeds over 40 words per minute. In fact, in the Intermediate Phase, by the end of Grade 4, inaccurate readers are almost all reading below 40 words per minute. By the end of primary schooling, there are very few inaccurate readers who are reading more than 60 words per minute.

A common criticism of setting fluency benchmarks is that by encouraging speed, this ignores the possibility that there are students that read slowly but with accuracy (Dowd & Bartlett, 2019). In the evidence shown

from this large Setswana data set, there are very few learners who read slowly but accurately and even fewer with poor accuracy but reasonable speed.

Figure 3: Reading speed and accuracy (Setswana)

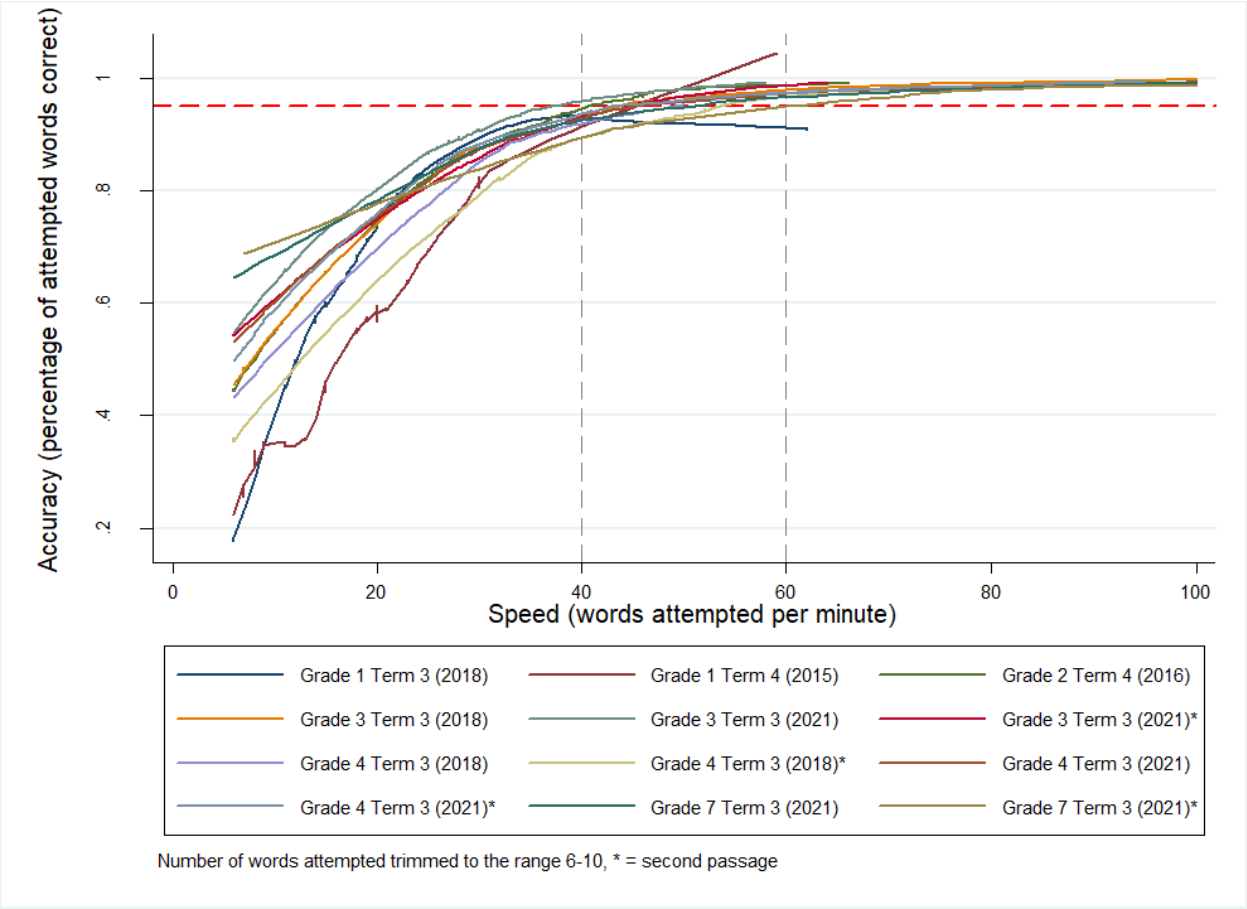


Figure 4: Speed distribution for learners reading with at least 95% accuracy in Setswana

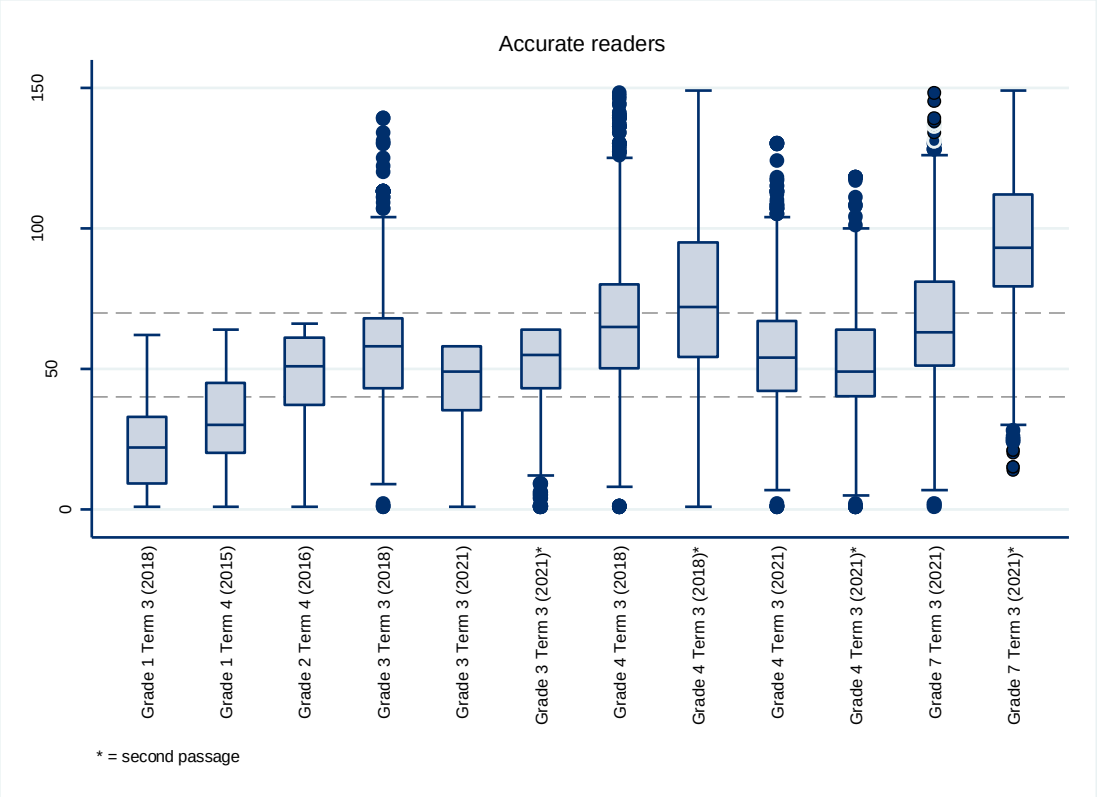
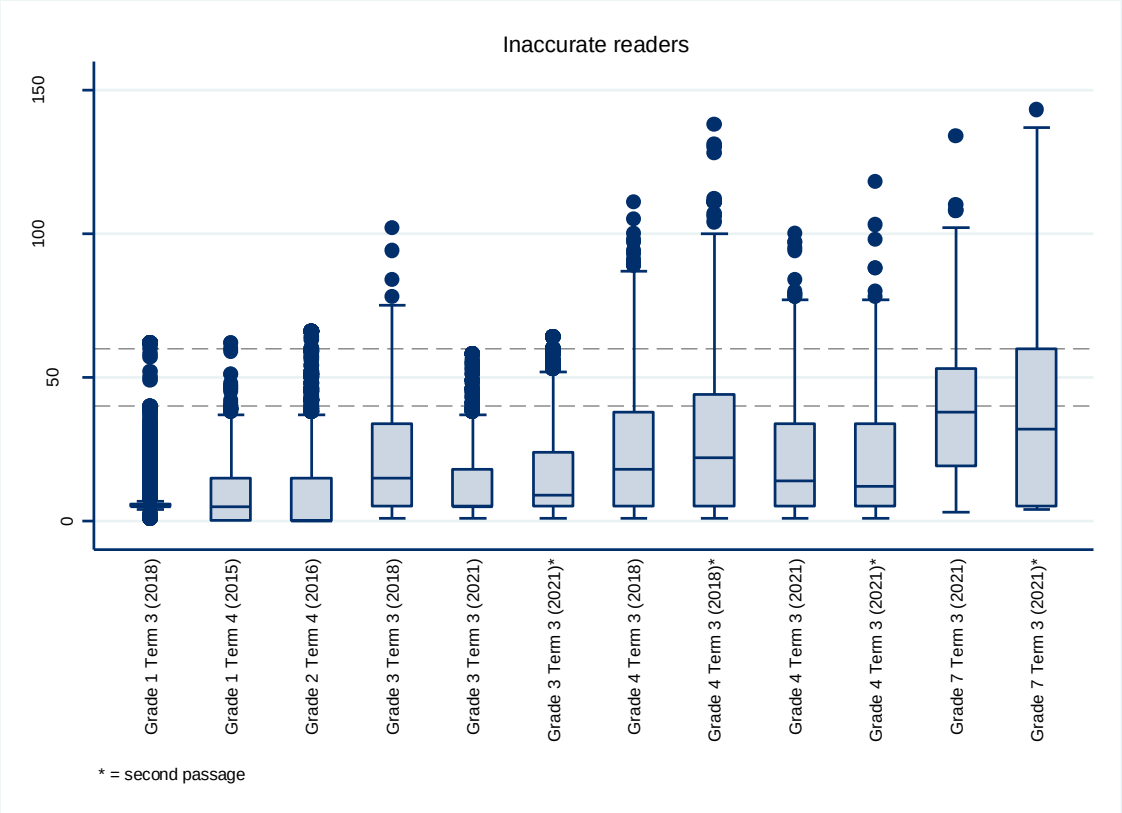


Figure 5: Speed distribution for learners reading with less than 95% accuracy in Setswana



The previous two figures, in conjunction with Figure 3 showing the non-linear relationship between fluency and accuracy, provide the first piece of evidence in support of a Setswana minimum fluency threshold of 40 cwpm. If learners are reading slower than 40 words per minute in Setswana, they have not yet reached accuracy levels supportive of automaticity in reading. These learners would benefit from instruction focused on improving their decoding skills and fluency. It is likely that the development of higher order skills will stagnate, including comprehension, until they reach this threshold.

6.2 RELATIONSHIP BETWEEN FLUENCY AND COMPREHENSION

The next step in our analysis is to examine whether the threshold and benchmark are logical reading points in lieu of the goal of reading: to read with comprehension. For this reason, we now examine the relationship between ORF and oral reading comprehension in reference to the threshold and benchmark. In addition to oral reading comprehension, home language written comprehension data is used to further explore the relationship between fluency and comprehension skills at higher grades.

6.2.1 Relationship between fluency and comprehension

Figure 6 plots the average comprehension score at each level of fluency separately for the Grade 3-7 Setswana samples attempting a subset (at least 5) of the ORF comprehension questions. (The analogous figure for the sample attempting all questions is shown in Figure A 4.) The figures include reference lines at 40 and 60 cwpm.

Although there are differences in the average comprehension levels between samples, the fluency-comprehension gradient is similar in Grades 3 and 4 across passages as seen in Figure 6. In Grades 3 and 4, the gradient is non-linear. Fluency below 40 cwpm appears to be a threshold below which comprehension skills are unlikely to develop. At all grade levels, learners reading below 40 cwpm tend to have very poor comprehension. Across samples, in this low accuracy and low fluency zone, learners would benefit from instruction that improves their decoding skills.

Looking at the Grade 3 and 4 relationship, learners reading between 40 and 60 cwpm appear to have reached an accuracy threshold. In this zone, increasing speed and automaticity is associated with improvements in comprehension. The comprehension-fluency gradient then tends to flatten out at around 60 cwpm, with diminishing returns to fluency. But this flattening occurs at fairly low comprehension levels (between 50% and 75% of comprehension questions correct in the Grades 3-4 samples) suggesting that underdeveloped comprehension skills become the key hurdle for learners at these fluency levels.

Box 3: Benchmarks should not be fixed to a specific level of comprehension. Illustrating the challenge of establishing equivalence of difficulty across sets of comprehension questions using Setswana data

Traditional approaches to benchmarking depend on identifying a fixed comprehension level (e.g., at least 80% of questions correct) and then applying statistical techniques to identify the fluency levels associated with meeting that comprehension threshold (Room to Read, 2018; Abadzi, 2012, RTI, 2017). Those approaches assume that 80% comprehension is a defined construct with equivalent meaning across passages. In the following discussion, we use the Setswana reading data to show how problematic this assumption is. This reinforces the need to detach empirical benchmarking methods from fixed comprehension levels (Ardington et al. 2021a:11).

In the 2021 EGRS I / RSP data collection, grades 3, 4 and 7 learners were assessed on two different Setswana ORF passages with accompanying comprehension questions. Comparing the graphed lines across the two grade 4 passages in Figure 6 (yellow and red solid lines) is illustrative of the challenge of establishing equivalence of difficulty across sets of comprehension questions. Despite a very high correlation between the oral reading fluency scores on both passages (0.91), the comprehension scores considerably differ. Linking a statistical process to a fixed comprehension process is therefore too sensitive to the cross-text comparability of comprehension questions.

While comprehension processes are hierarchical, question difficulty ranges within a single comprehension question typology (e.g., literal comprehension questions). The figure below shows how learners perform on individual comprehension questions classified as per PIRLS by the underlying comprehension process they aim to assess. The bar colours indicate the types of comprehension processes engaged in the questions: literal (blue), straightforward inference (red), interpret and integrate ideas and information (grey) and examine or evaluate (dark blue). There is no clear relationship between the hierarchy of the comprehension process and the difficulty of questions. 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, integrate or examine information. Even where different passages assign an equal mix of comprehension question types, question difficulty ranges even within a single comprehension typology. There are challenges in anchoring comprehension questions to a required level, presenting significant limits to creating comparable reading comprehension assessments. This also implies that deciding on a chosen cut-off point for a desired comprehension level (e.g., 80 % of questions correct) is somewhat arbitrary. Benchmarks established using a fixed comprehension level approach would be biased by the difficulty level of the comprehension questions.

Proportion of learners answering each comprehension question correctly

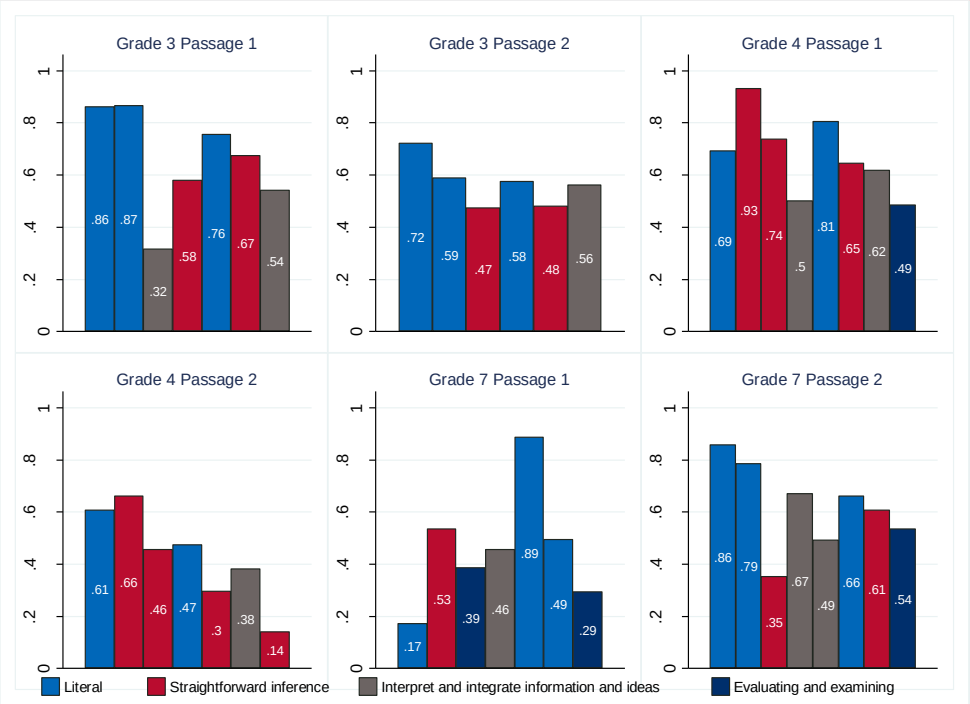
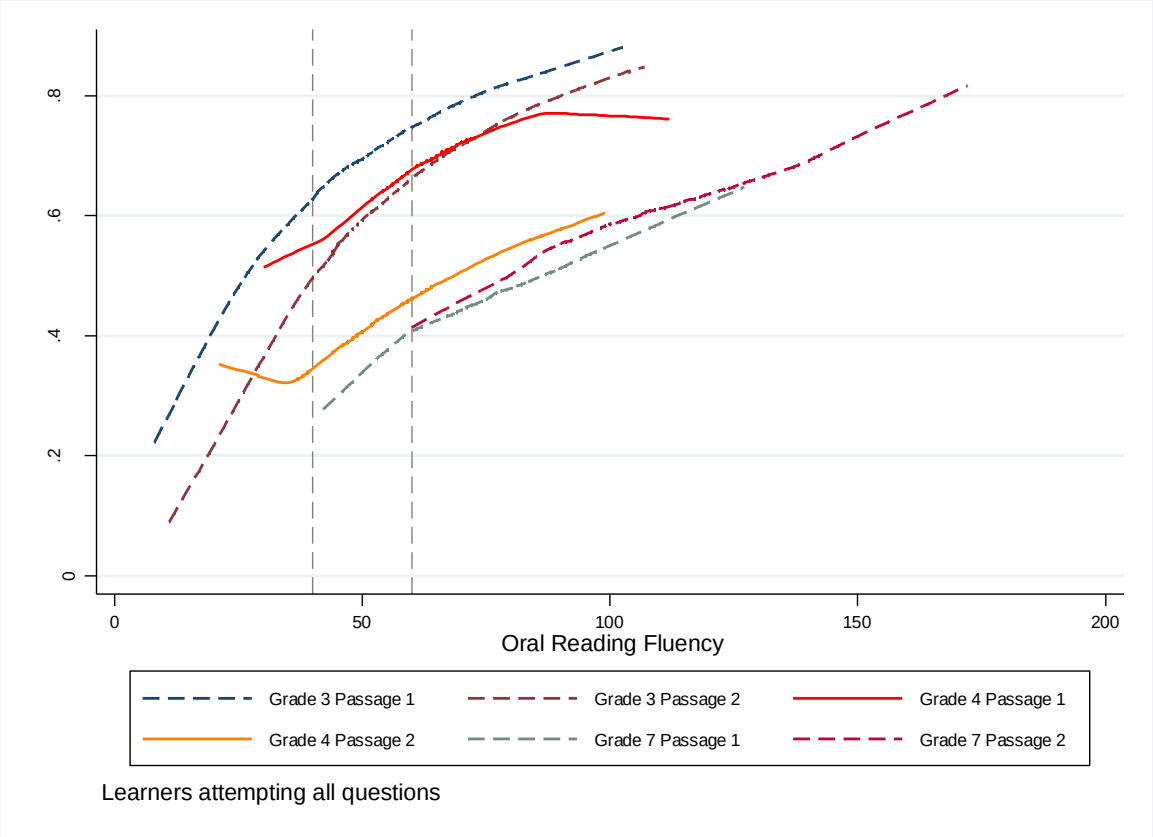


Figure 6: Relationship between fluency and comprehension for students attempting at least a subset of comprehension questions



Source: EGRS I /RSP 2021 data.

The concave relationship between fluency and comprehension is a key reason for establishing both a reading threshold and benchmark. Even when analysing more closely individual comprehension questions, the concave relationship between fluency and comprehension typically holds, particularly at the Grade 3-4 level. This is explored further by examining the relationship between fluency and individual comprehension questions for all grades, and passages tested (passage 1 results are shown in Figure 7 to Figure 9 while passage 2 results are in the appendix as Figure A 1 to Figure A 3). The lines reflect locally weighted polynomial regressions of the relationship between ORF and the proportion correctly answering each comprehension question. The histogram bars reflect each grade sample's distribution of ORF scores for learners attempting each question. Notwithstanding the substantial differences in comprehension difficulty across individual questions, the fluency-comprehension gradient is remarkably consistent and aligns with the notion of thresholds in the developmental cline of reading.

The fluency-comprehension relationship is expected to be slightly different at the Grade 7 level, as learners are at a more advanced stage in their reading development (seen in the majority of Grade 7 sample

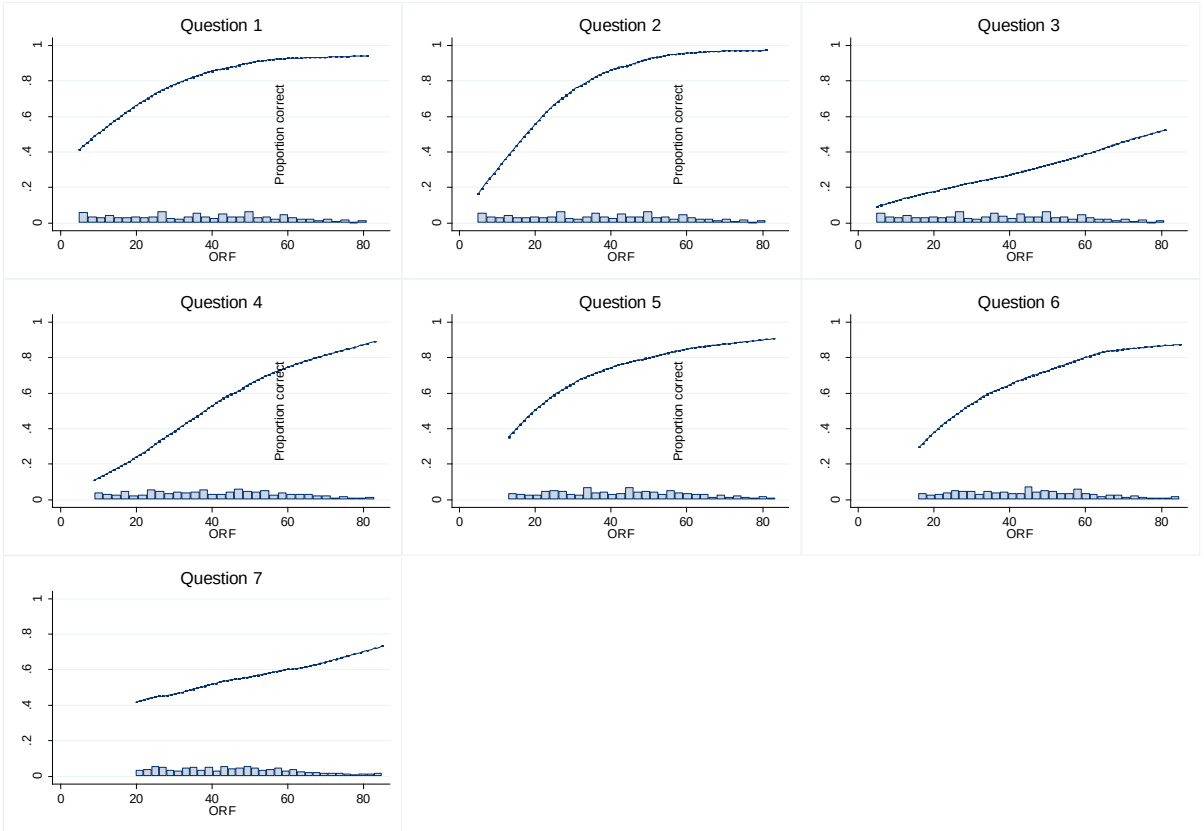
learners reading above the 40 cwpm fluency threshold). Although a slightly more linear relationship is observed for Grade 7 learners overall in Figure 6, the analyses of individual questions reveal a concave relationship on many of the questions. On the Grade 7, passage 1, a steeper gradient is observed as learners progress from reading between 40 to 60 cwpm, implying that once the threshold is met, comprehension advances more rapidly even at higher grade levels. The gradient between fluency and comprehension becomes flatter (and the levels are low) even where learners read at or above 60 cwpm.

At all grade levels, patterns suggest that in a fluency zone, where learners have reached the 60 cwpm benchmark, the instructional focus and support should emphasize strengthening reading comprehension skills through vocabulary development, the application of reading comprehension strategies and deeper engagement with text.

In summary, there are regular patterns in the Setswana reading data across grades and reading passages to support the identification of:

- a lower fluency threshold at around 40 cwpm below which teaching should focus on improving decoding skills; and
- a higher fluency benchmark at around 60 cwpm above which teachers' attention should hone in on strengthening reading comprehension skills and vocabulary, while continuing to support enhanced fluency.

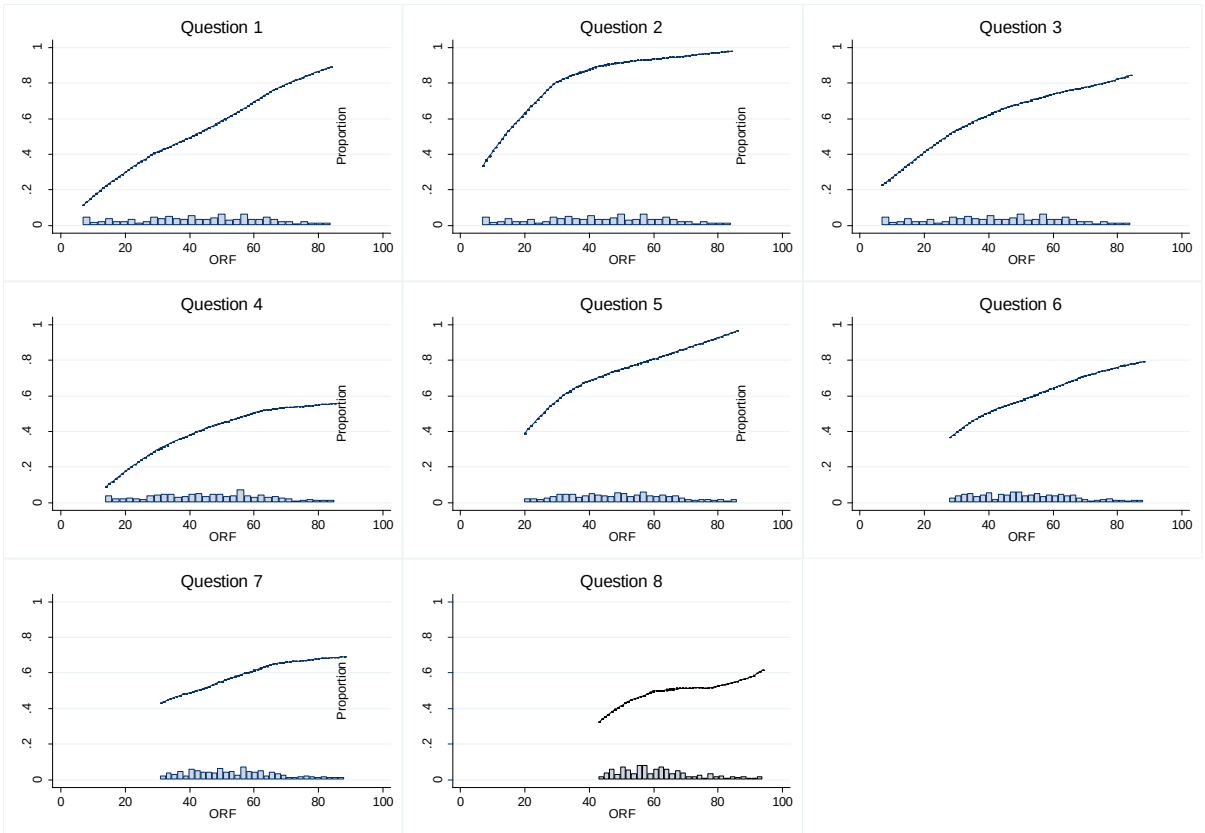
Figure 7: Relationship between fluency and individual comprehension questions: Grade 3, passage 1



Grade 3 Passage 1

Source: RSP 2021, own calculations. 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.

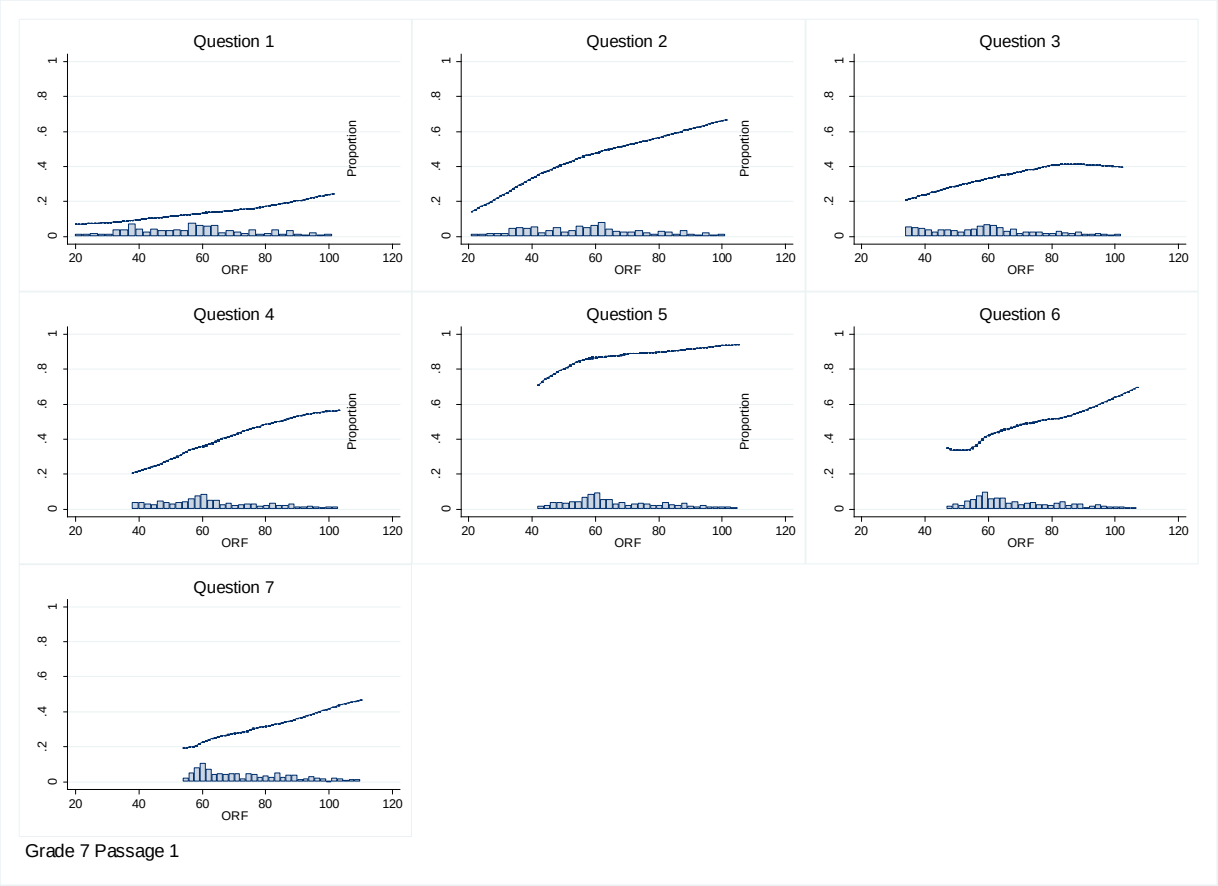
Figure 8: Relationship between fluency and individual comprehension questions: Grade 4, passage 1



Grade 4 Passage 1

Source: RSP, 2021, own calculations. Notes: See figure above.

Figure 9: Relationship between fluency and individual comprehension questions: Grade 7, passage 1



Source: EGRS, 2021, own calculations. Notes: See figure notes above.

6.3 SETSWANA FLUENCY THRESHOLDS: ATTAINABILITY AND LEARNER PROFILES

We now consider how the threshold and benchmarks can be used to distinguish learners into reading profiles and explore whether the threshold and benchmark are attainable for learners in no-fee schools.

Figure 10 shows the percentage of learners by grade and term falling into the following fluency categories:

- Non-reader: unable to read 1 word (shown in dark blue),
- Reading below the threshold: reading less than 40 cwpm (shown in light blue),
- Reading at or above the threshold but below the benchmark: 40 to 59 cwpm (shown in red) and
- Reading at or above the benchmark: at least 60 cwpm (shown in dark red).

We have excluded any repeaters, so that the assessment level matches the grade in which learners currently are. The general progression pre-COVID is clear:

- By the end of Grade 1, 55% of learners are unable to read a single word correctly (non-readers), a further 39% are reading slower than the minimum threshold of 40 cwpm.

- By the end of Grade 2, there are half as many non-readers at 29% compared to Grade 1. About 58% (29% + 29%) are reading below the minimum threshold of 40 cwpm, while 42% reach the threshold. A very small percentage at 15% are meeting the benchmark.

- By the end of Grade 3 (measured pre-pandemic in 2018) the percentage of non-readers remains high at 18%, although almost a half of learners (51%) meet the minimum threshold and almost a quarter (24%) of the sample meet the benchmark.

- By the end of Grade 4, there are very few non-readers at just 1-6%. The majority are meeting the minimum threshold (77%) and a 51-67% are reading at or above the benchmark of 60 cwpm.

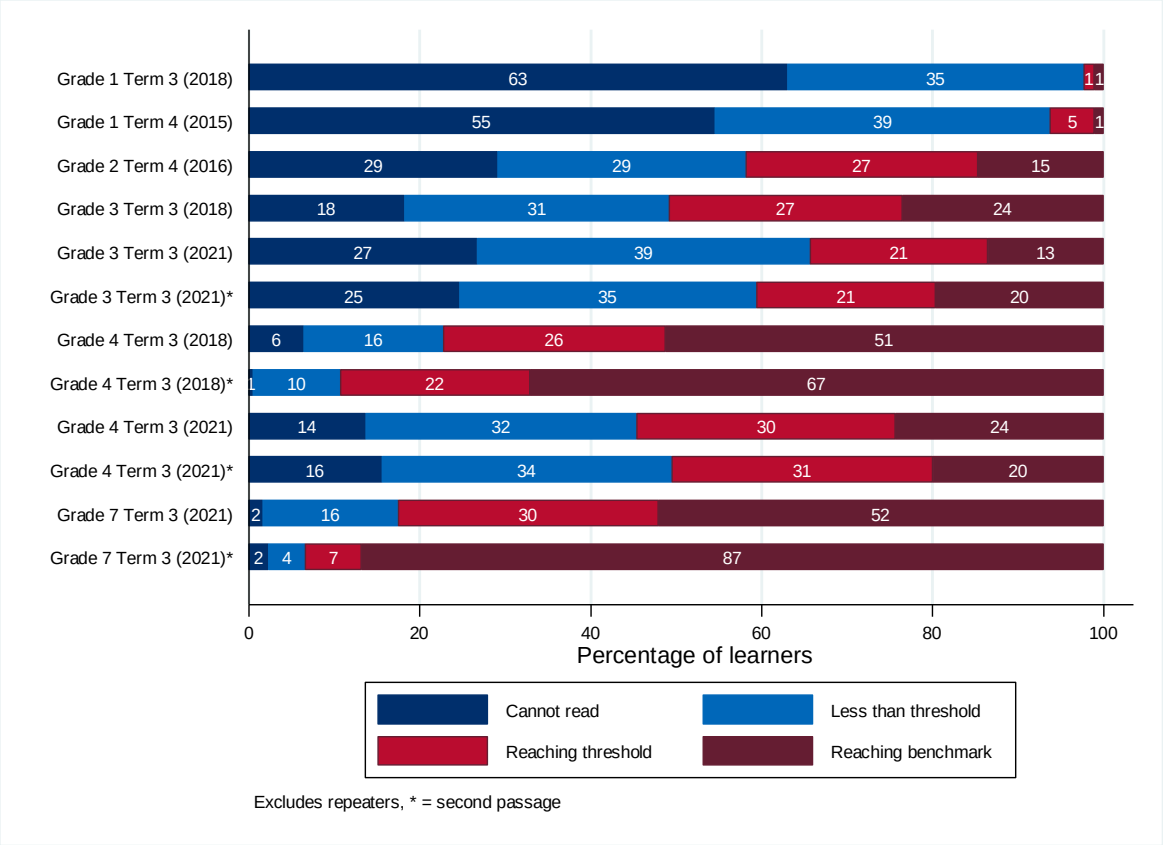
Unfortunately, the Grade 3 and 4 reading profiles have been significantly worsened in a pandemic period as seen in the 2021 comparative profiles.

- At the Grade 4 level, just 20-24% were meeting the benchmark in 2021 compared to around a half meeting the benchmark pre-pandemic.

By the end of primary school, the majority (82% or more) of Grade 7s are reading at or above the minimum Grade 2 threshold. Depending on passage difficulty, 52-87% of Grade 7s are meeting the Grade 3 benchmark of 60 cwpm even though the Grade 7s were assessed in a pandemic period. A possible reason for this is that the acquisition of their foundational early grade reading skills occurred before the pandemic.

The results confirm that even at the lower pandemic reading levels, the thresholds are set at levels that are practically useful: they are attainable for a large enough sample of current Foundation Phase and Intermediate Phase learners, while at the same time they are high enough to encourage reading development to a level more appropriate for the demands of the curriculum.

Figure 10: Early grade fluency profiles, Setswana samples



Source: EGRS I, 2015-2021; RSP 2018-2021. Own calculations.

To what extent do the proposed thresholds and benchmarks correspond to meaningful and distinguishable zones along the reading development cline? We examine this in Table 11 by exploring reading profiles following Stern et al (2018). Reading profiles are explored by combining all the samples and summarizing accuracy, comprehension and letter-sound knowledge across 4 fluency categories.

Learners who cannot read one word (0 cwpm):

- Their letter sound-knowledge is highly underdeveloped, averaging just 14 clspm;
- 1 of every 5 learners cannot sound one letter correctly in a minute.

Learners reading below the fluency threshold (less than 40 cwpm):

- have low very levels of accuracy with just a third reaching 95% accuracy;
- have very poor comprehension scores;
- would benefit from instruction on letter-sound knowledge and on improving their word decoding.

Table 11: Learner characteristics by early grade Setswana fluency profiles

CANNOT READ: 0 CWPM	
Mean correct letter-sounds per minute	13.9
% unable to sound 1 letter	20%
READING BELOW THRESHOLD: 1-39 CWPM	
Mean correct letter-sounds per minute	36.9
% with at least 95% accuracy in word reading	33%
Comprehension (% of total correct)	23%
Comprehension (% of attempted correct)	35%
MEETS THRESHOLD BUT NOT BENCHMARK: 40 - 59 CWPM	
Mean correct letter-sounds per minute	54.5
% with at least 95% accuracy in word reading	82%
Comprehension (% of total correct)	48%
Comprehension (% of attempted correct)	59%
MEETS BENCHMARK: 60+ CWPM	
Mean correct letter-sounds per minute	58.8
% with at least 95% accuracy in word reading	94%
Comprehension (% of total correct)	56%
Comprehension (% of attempted correct)	66%

Among the group of learners who meet the lower threshold (reading between 40-59 cwpm), but not the benchmark

- letter-sound knowledge is well established (averaging 55 clspm);
- levels of word accuracy have improved with 82% of these learners achieving at least 95% accuracy;
- comprehension skills are developing but remain weak.

Finally, among the group of learners who meet the benchmark of 60 cwpm:

- letter-sound knowledge is very well established (averaging 59 clspm);
- they are almost all accurate readers (94% of them reach 95% accuracy levels);
- with comprehension scores in the range of 56% to 66%; and

- would benefit from instruction that focuses on improving their comprehension skills through increasing vocabulary and critical engagement with text.

6.4 Predictive and concurrent validity of the Setswana fluency threshold and benchmark

6.4.1 Predicting future fluency

The longitudinal nature of the EGRS I sample allows us to investigate the predictive validity of the proposed ORF threshold and benchmark. We illustrate this in Figure 11 and Figure 12 by tracking the fluency profiles of EGRS I learners from Grade 2 (term 4) to Grade 4 (term 3), and then Grade 4 (term 3) to Grade 7 (term 3). In each figure, we distinguish learners into initial fluency categories: non-readers (0 cwpm), reading below the threshold (1-39 cwpm), and meeting the threshold (40-59 cwpm). Learners who were already reaching the benchmark at the initial point are excluded from the figures. By initial fluency category, we then identify their fluency category in a later grade assessment. Three clear patterns emerge when we consider the Grade 2-4 and Grade 4-7 transition patterns in Figure 11 and Figure 12.

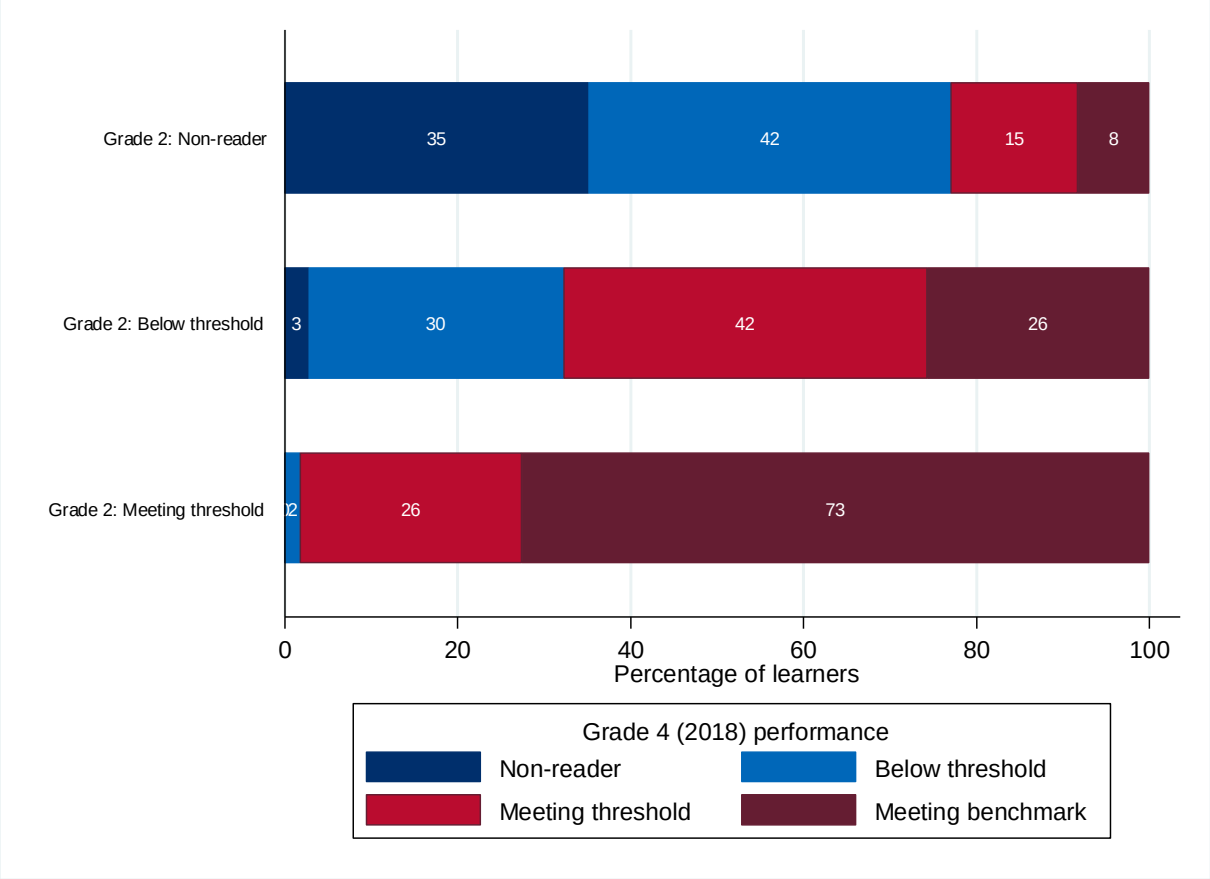
1. **Non-readers stagnate.** About 35% of Grade 2 learners who were non-readers were still unable to read one word by Grade 4. A sizeable portion of these Grade 2 non-readers begin to read slowly by Grade 4, but most are not yet reaching the lower threshold (of 40 cwpm). By Grade 4, only 23% percent of Grade 2s reach the threshold, and just 8% meet the Grade 3 benchmark. A similar picture holds in the Grade 4-7 transition. About 46% of Grade 4 learners who were non-readers were still unable to read one word by the end of primary school. However, a sizeable portion of these non-readers have begun to read slowly by Grade 7, but most are not yet reaching the lower threshold (of 40 cwpm). Only a small percentage (12%) meet the benchmark by the time they leave primary school.

2. **Slow readers can attain the lower threshold.** Among Grade 2 learners who were reading below the lower threshold (1-39 cwpm) in Grade 2, the majority (68%) had reached that threshold by Grade 4, with just over a quarter 26%) meeting the benchmark. Among learners who were reading below the lower threshold (1-39 cwpm) in Grade 4, the majority (73%) had reached that threshold by Grade 7, and 45% meet the benchmark. However, such a slow pace of reading development is unlikely to support learning in primary school.

3. **Meeting the threshold is highly predictive of meeting the benchmark.** An encouraging picture emerges for those meeting the threshold by the end of Grade 2. By the time they reach Grade 4, 73% of this group are meeting the benchmark. At the Grade 4 level, of learners meeting the threshold, almost all (91%) meet the benchmark by the end of Grade 7. The threshold of 40 cwpm

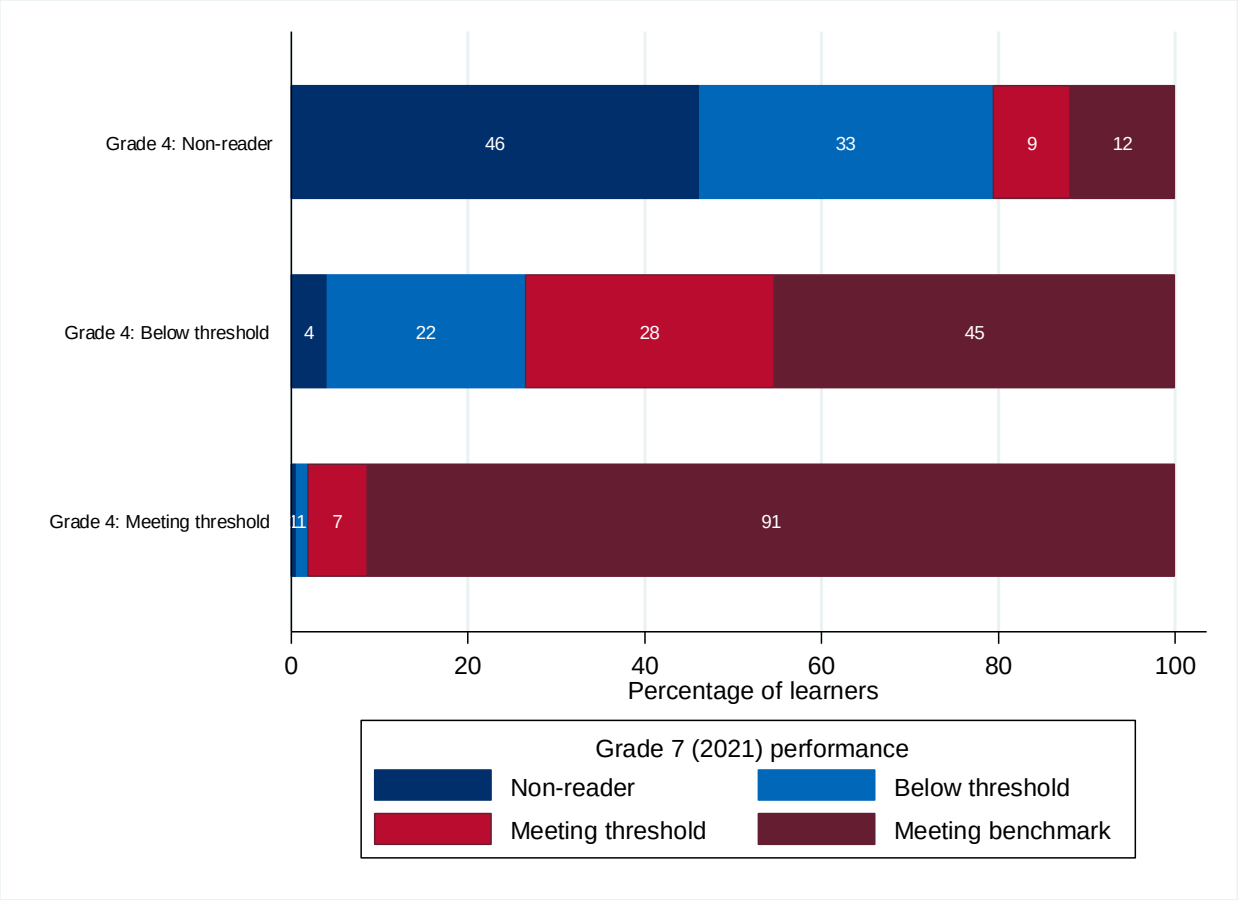
clearly signals a point at which reading development can take off, and a key milestone in being able to meet the benchmark of 60 cwpm.

Figure 11: Fluency in Grade 4 by learners' fluency profile in Grade 2, Setswana



Source: EGRS I (longitudinal sample 2016 to 2018), own calculations.

Figure 12: Fluency in Grade 7 by learner's fluency profile in Grade 4, Setswana



Source: EGRS I (longitudinal sample 2018 to 2021), own calculations.

6.4.2 Fluency and written comprehension: concurrent and predictive validity

In addition to investigating the predictive validity of the threshold in relation to future fluency, we investigate the validity of the fluency threshold in predicting learners' current and future comprehension skills. The additional written comprehension assessments designed in 2021 are particularly useful for this purpose.

Figure 13 shows the concurrent relationship between Setswana fluency (y-axis) and Setswana written comprehension scores (x-axis) in Grades 3 and 4 respectively. The dashed grey lines represent the lower 40 cwpm threshold and the 60 cwpm benchmark. Both figures show a similar pattern. When learners are reading at or above the fluency threshold of 40 cwpm, the majority tend to get at least half of the written comprehension questions correct. As they reach the fluency benchmark of 60 cwpm, they are virtually getting full marks for written comprehension. The concave fluency-comprehension relationship in both figures is also noted, which agrees with the patterns observed earlier in examining the relationship between oral reading fluency and comprehension questions asked on the related passage. In other words,

there tends to be diminishing returns to fluency as learners reach the 60 cwpm benchmark, reinforcing the need for the teaching of comprehension skills.

Figure 13: Relationship between Grade 3 oral reading fluency and Grade 3 written comprehension in Setswana, EGRS I (2018)

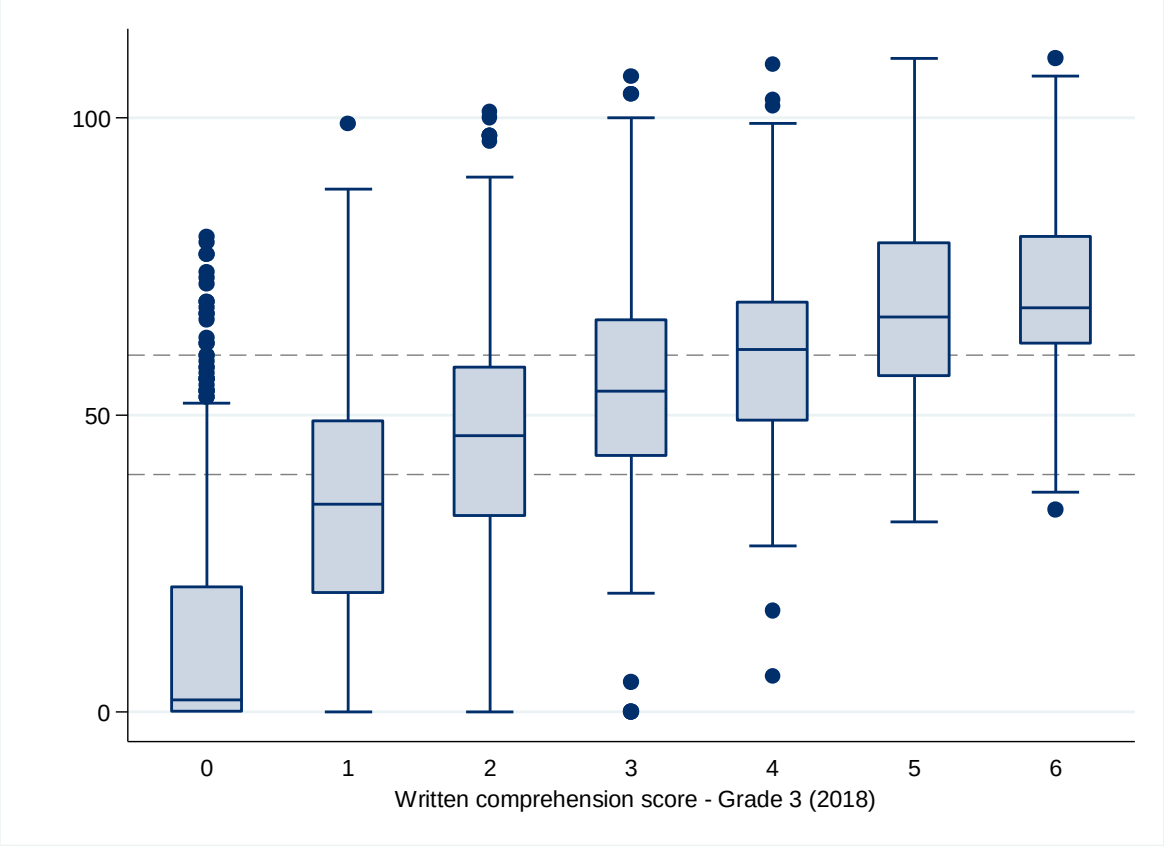


Figure 14 and Figure 15 subsequently consider the relationship between current fluency and future written comprehension. Learners' Setswana fluency in the first year of their Intermediate Phase (Grade 4) is highly related to how they perform in Setswana written comprehension by the end of primary school as seen in Figure 14. The majority of Grade 4 learners reading at or above the 40 cwpm mark are getting more than half (6 of 11) of the written comprehension questions correct by the end of Grade 7, while those reading at or above the benchmark in Grade 4 are close to getting full marks for comprehension.

The importance of meeting the fluency threshold by the end of Grade 2 in equipping learners with the reading comprehension skills they need for secondary school success is starkly displayed in Figure 15. It shows the relationship between Grade 2 fluency and end of primary written comprehension performance. Almost all learners reading below the threshold at the end of Grade 2 are failing home language written comprehension by the end of primary school. By contrast the majority of those reading at or above the threshold of 40 cwpm at the end of Grade 2, are scoring above 80% (9 of 11) for written comprehension

by the end of primary school. They are clearly equipped with the reading skills they need to be able to understand fully what they are reading later on. Reading at or above the threshold by the end of Grade 2, is an important milestone to be able to read for meaning and to learn in later grades.

Figure 14: Relationship between Grade 4 oral reading fluency and written comprehension performance in Grade 7 in Setswana, EGRS I (2018 and 2021)

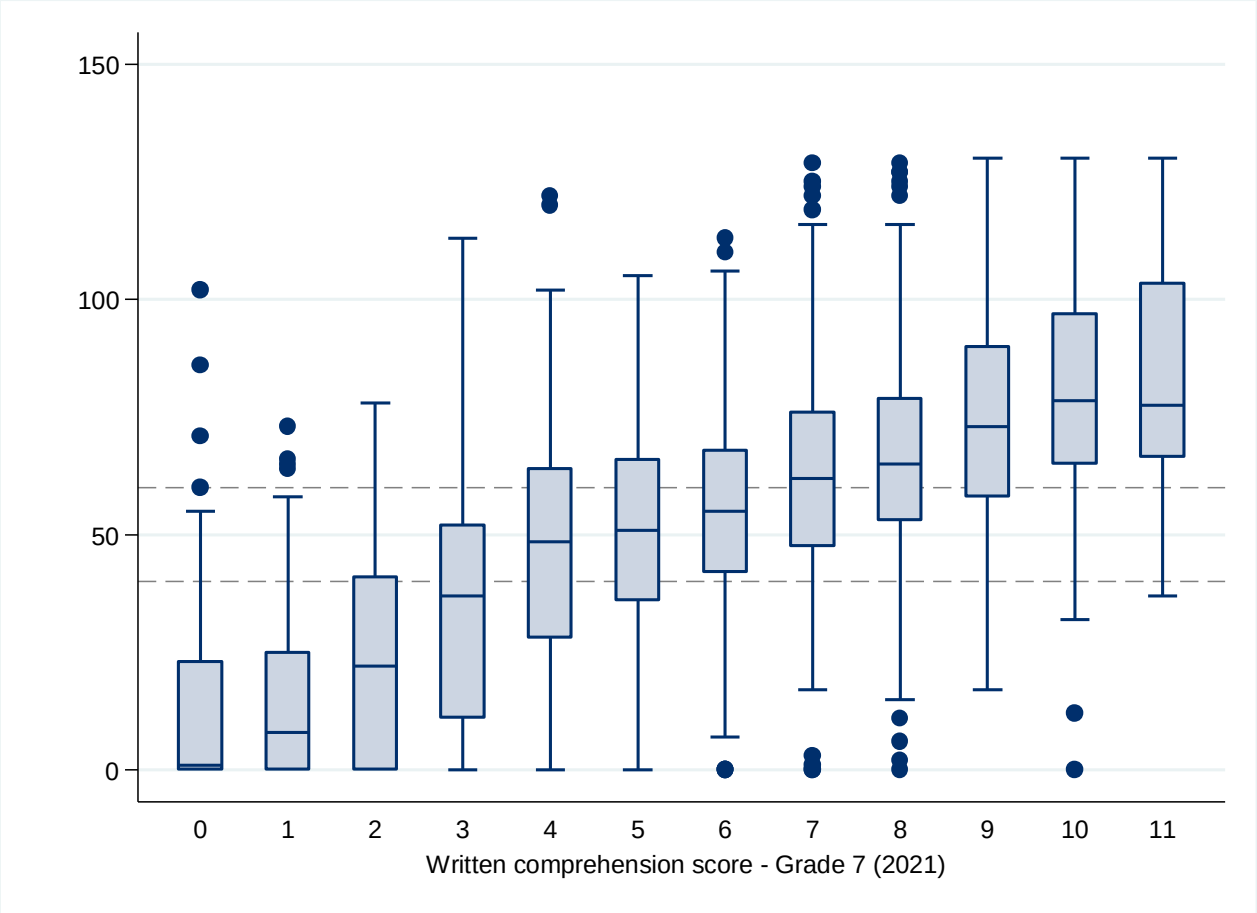
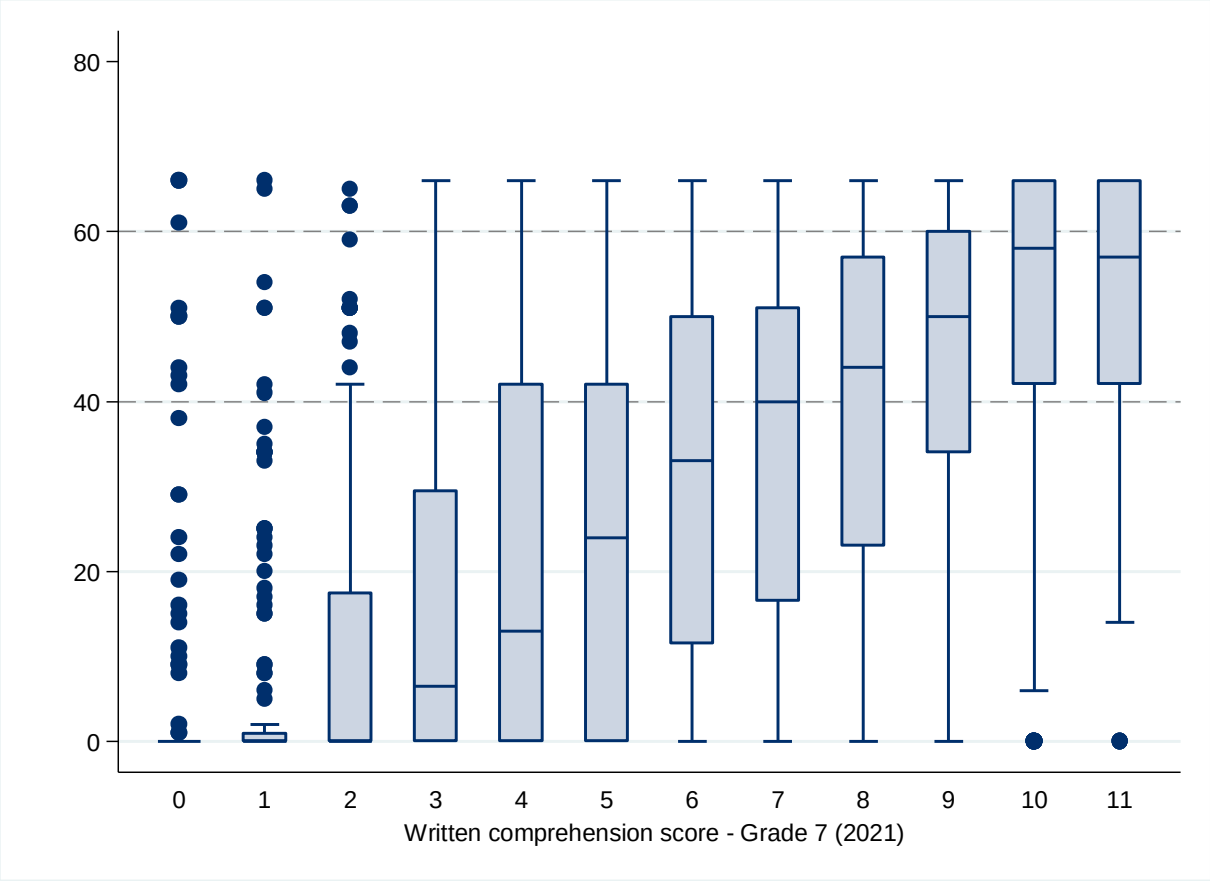


Figure 15: Relationship between Grade 2 oral reading fluency and written comprehension performance in Grade 7 in Setswana (EGRS I 2016 to 2021)



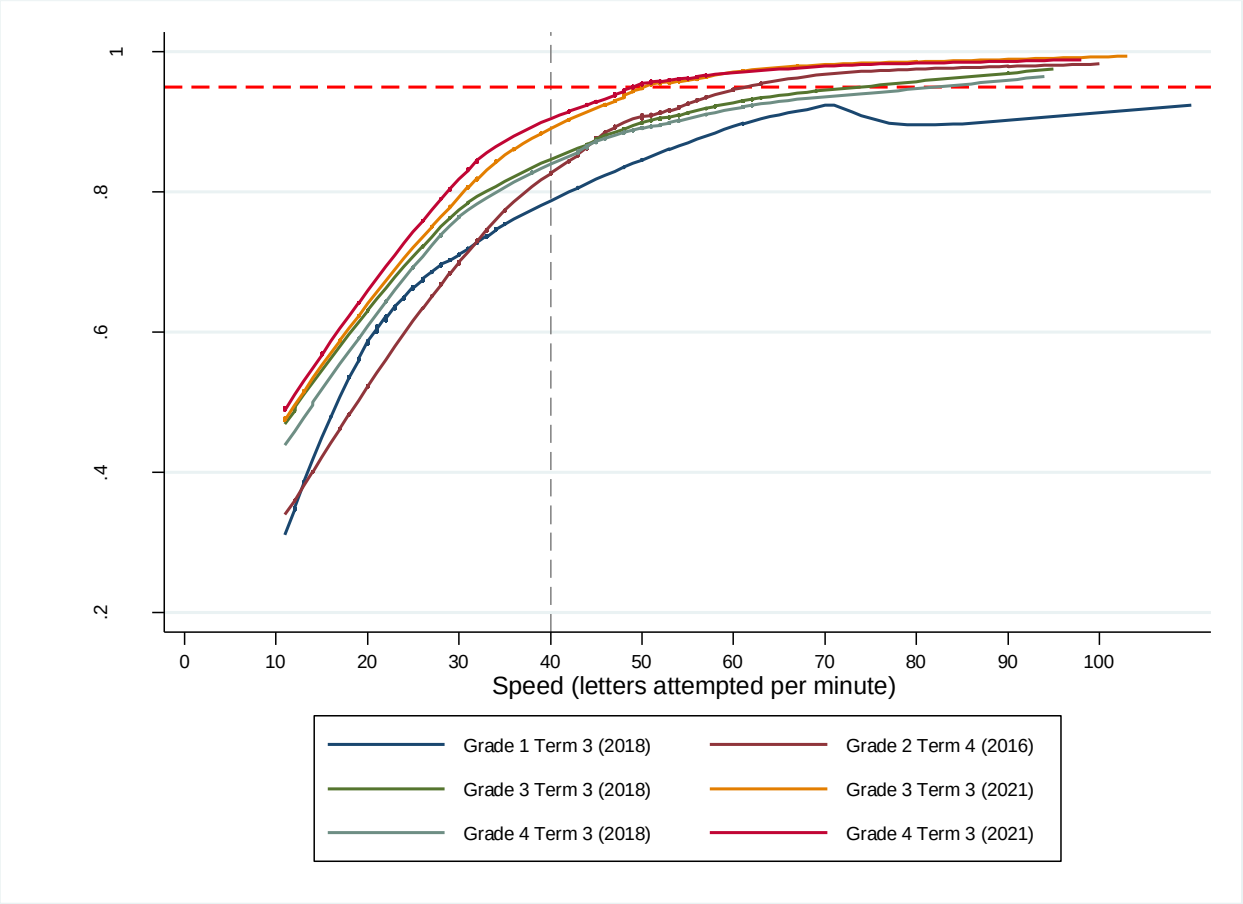
6.5 ESTABLISHING A LETTER-SOUND BENCHMARK IN SETSWANA

Discussed earlier in section 2, reading development is a hierarchical process. Working backwards along the development cline, we consider what letter-sound benchmark could support the acquisition of fluency skills. Specifically, we consider whether a letter-sound benchmark of 40 correct letter sounds per minute (clspm) in Grade 1 suggested for Nguni languages is also appropriate for Sesotho-Setswana languages (where both language groups are alphabetic in nature). To do this, we repeat the analysis provided in the Nguni language benchmarking report (Ardington et al. 2020) using Setswana reading data.

6.5.1 Relationship between speed and accuracy in alphabetic knowledge

Is 40 clspm also an appropriate letter-sound benchmark for Setswana? To answer this question, we look at the relationship between the speed at which learners can sound letters and the accuracy with which they do this in Figure 16.

Figure 16: Letter-sounds – the relationship between speed and accuracy, Setswana



Source: EGRS I (2016-2021), RSP (2018-2021), own calculations.

Consistent with what is found for letter-sound reading in Nguni languages, the following patterns with respect to reading letter-sounds are found in Setswana:

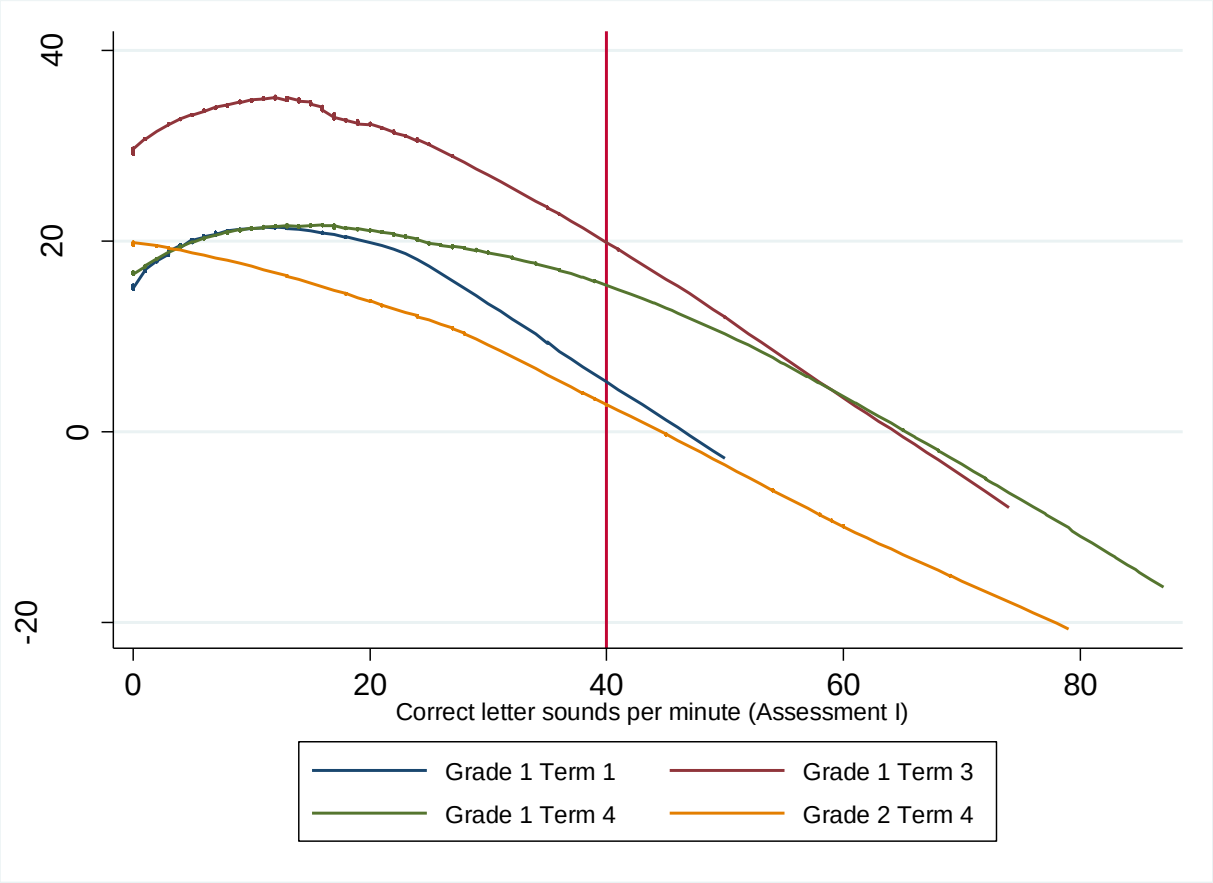
- Learners with low speed tend to have low accuracy.
- Accuracy improves steadily with speed to a point, but beyond this point there are no further improvements in accuracy.
- The letter-sound speed-accuracy gradient tends to flatten around 40 letter-sounds per minute (indicated by the dashed grey line).

Analysis of longitudinal data provides support for the idea that there is letter-sound recognition speed beyond which learners show little to no improvement in support of the decoding threshold hypothesis.

We examine whether there is a point at which learners' letter-sound speed tends to stagnate. This is seen in Figure 17 plotting the relationship between the gain in letter-sound reading at a second assessment point (assessment II) against letter-sound reading at an initial assessment (assessment I). Gains in letter-sounds per minute tend to increase with baseline performance and then begin to decline (consistent with what is found for Nguni languages). There is less improvement in letter-sounds per minute for learners who could correctly sound 40 letters per minute at the first assessment than for those who were not meeting this letter-sound benchmark at the first assessment. An examination of the improvement from the end of Grade 2 to the end of Grade 4 (yellow line) also suggests that improvements in letter sounds diminish as learners exit the Foundation Phase. At every baseline letter-sound level, there is substantially less improvement from the end of Grade 2 than from the end of Grade 1.

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

Figure 17: Development on letter-sound knowledge over time, Setswana



Benchmarks for lower order skills need to be high enough to support the development of higher order skills. At the same time, they need to be low enough to be responsive to incremental change. They should also be attainable. In Figure 18 we distinguish learner by grade samples into four categories: cannot read any letter correctly, sounding fewer letters correctly than the benchmark (1-39 clspm) and meeting the benchmark (40 clspm).

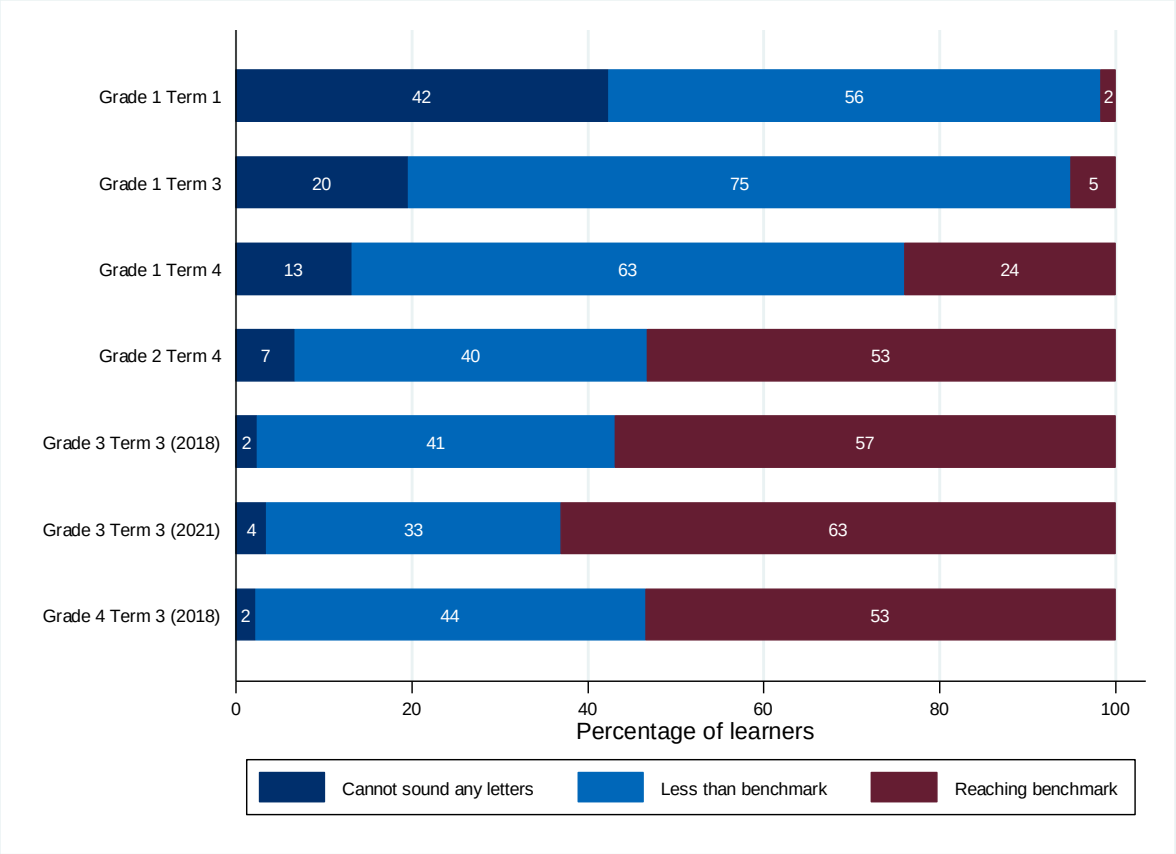
- By the end of Grade 1 (in a pre-COVID year), about a quarter (24%) of Grade 1s were meeting the letter-sound benchmark. The benchmark is attainable. However, the majority of learners are acquiring letter-sound knowledge too slowly, with 13% unable to sound one letter correctly.

- By the end of Grade 2, over half (53%) of learners pre-COVID were meeting the 40 clspm benchmark.

- By the end of Grade 3 (and into Grade 4), the letter-sound distribution does not improve very much relative to the Grade 2 distribution, with 53-63% meeting the 40 clspm benchmark. Teachers are required by the curriculum to move on towards teaching higher order skills with each grade, yet this basic skill is not being mastered by learners with around 38-46% unable to meet the benchmark by the end of Grade 3 (and Grade 4).

It is also worth highlighting that while the 2021 letter-sound distributions appear to be better compared to pre-COVID assessments, the letter-sound assessment was much easier in 2021 relative to 2018 as it excluded complex consonants and diacritics (consistent with the pre-2018 EGRS I assessments).

Figure 18: Correct letter-sounds per minute distribution, Setswana



Source: EGRS I (2015-2021), RSP (2018-2021), own calculations. Notes: The letter-sound assessment was much easier in 2021 relative to 2018 as it excluded complex consonant clusters and diacritics

7 SUMMARY OF SETSWANA READING THRESHOLDS AND BENCHMARKS

In concluding this section, we establish the following thresholds and benchmarks for early grade reading sub-skills in Setswana:

- By the end of Grade 1, all learners should be able to correctly sound 40 letters per minute.
- By the end of Grade 2, all learners should be meeting the fluency threshold, correctly reading from a passage at least 40 words per minute.
- By the end of Grade 3, all learners should be meeting the fluency benchmark, correctly reading from a passage at least 60 words per minute.

These benchmarks complement recent efforts to provide greater specificity in the teaching of African languages in South Africa (DBE, 2020b). We now turn to English, and the subskills required to develop

adequate reading skills in the language of teaching and learning as learners transition into the Intermediate Phase.

8 RECOMMENDATIONS

Although this research advances the establishment of reading benchmarks and thresholds in South African languages, the true value of this body of research in supporting policy and practice will only be realized when linked to a national programme to assess and monitor early grade reading skills (Ardington et al., 2021a:14). If national EGRA-type assessments were introduced, the threshold and benchmarks could be used to monitor sector progress in reading in the early grades. Currently, early grade reading is not being systematically measured in South Africa at district, provincial or national levels. Without credible measurement of foundational early reading skills, it will not be possible to track reading progress. This leads us to key recommendations.

1. **Implement a national system to test and monitor early grade reading skills:** Measurement of early reading skills would bring into focus the importance of early reading skills, reshaping policy priorities towards developing reading competency and directing resources to promote reading development. Written assessments of reading comprehension, typically used in primary school testing, are unable to detect foundational reading skills. The proposed Systematic Evaluations – a sample-based testing system (DBE, 2020a) – could be bolstered to include oral reading fluency assessments in addition to written assessments. At the primary grades, the Systemic Evaluation tests will be aimed annually at the Grade 3 and 6 level. Established thresholds and benchmarks can be applied at these grade levels to identify and track over time what proportion of learners can meet them and to remediate learners who do not meet them.

2. **Continue to collect early grade reading data to establish threshold and benchmarks in all South African languages:** Thresholds and benchmarks for reading are only likely to be linked to a national testing system if they are available in all official South African languages. Early grade reading data collection exercises will need to be repeated in other languages as identified in Table 22 below. It would also be prudent to re-evaluate current benchmarks and thresholds as more data is made available, with different samples from different provinces and in lieu of the possibility that reading norms can shift over time.

Table 12: Current status of DBE donor/partner planned early grade reading benchmarking in all official languages

Language	Language group	Establishing reading benchmarks: current status
Sepedi	Sotho-Tswana languages	in progress - establish whether Setswana benchmarks are applicable
Sesotho		available - establish whether Setswana benchmarks are applicable
Setswana		complete (current report)
Siswati	Nguni languages	complete
isiNdebele		data not available – but Nguni language benchmarks assumed
isiZulu		complete
isiXhosa		complete
Xitsonga		data collection planned for August 2022
Tshivenda		data collection planned for 2023
Afrikaans		data collection in progress
English (Additional language)		complete (current report)
English (home language)		norms available from international contexts

4. **All forthcoming early grade reading data collection programs should ensure evidence-based best practices are followed so that new assessments can support benchmarking exercises.** Organizations considering collecting early grade reading assessment data for independent research projects or evaluations could collaborate to support national reading benchmarking exercises. Their data could be used for benchmarking purposes if appropriate protocols and processes are followed**.** This includes applying a 3-minute time allocation rule to ORF assessments. The EGRS I wave 5 process of instrument development, with multiple rounds of testing and suitable protocols to support the collection of assessments of decoding skills provides a best practice scenario for how this should be done. The collaborative work of linguists, data analysts, home language specialists and DBE officials was critical throughout this process, and in establishing the final benchmarks/thresholds.

5. **Ensure that EGRA-type assessments are included as a critical aspect of formative assessment practice in primary schools:** EGRA-type assessments should be ongoing in all primary schools. These EGRA-assessments are not only relevant for the Foundation Phase grades but should be applied at the Intermediate Phase level, especially given the devastating learning losses associated with COVID-19 related school disruptions (Ardington et al., 2021b). Our analysis at the Grade 7 level was indicative of significant percentages of learners leaving primary school not having met the grade fluency benchmark of 60 cwpm in Setswana (13-48%) and Grade 5 fluency benchmark of 90 cwpm in EFAL (37%). The progressive roll-out of EGRA training for teachers by the Department of Basic Education in 2015 should be leveraged to promote EGRA testing, equipping teachers with the skills to administer these assessments (Maboya, 2020). The newly established benchmarks should be connected this, as they provide a framework for teachers to easily interpret the results from the EGRA assessment. As teachers conduct

EGRA-assessments, guided by thresholds and benchmarks, this will help them to identify early-on whether learners are on track, and align their instructional practice with each learner's level of reading development.

6. **Preservice Initial Teacher Education programs should also reflect familiarity with EGRA assessment procedures**: To support the inclusion of EGRA-type assessments in formative assessment and its effective implementation in schools, new teachers entering the system need to be equipped to assess early reading development and identify and remediate struggling readers through appropriate preservice Initial Teacher Education programs.

7. Resources to promote reading development through a multi-sectoral approach need to be consistently prioritized in policy and in budgets: Key resources required to promote reading development and instil an enjoyment of reading include training teachers in how to teach reading effectively through improved pre-service and in-service training; and ensuring that learners have sufficient and suitable books to read. Relatedly, the growing problem of large class sizes in the early grades also needs to be addressed as individualized reading instruction or even assessment is hindered when class sizes exceed prescribed recommendations (DBE, 2020a:106). More effort also needs to be given to considering how reading can receive higher priority in homes so that children are more exposed to oral language and print at earlier ages. Children are entering school with underdeveloped emergent literacy and language skills (Dawes et al., 2017), making the work of teachers and acquiring decoding skills much harder. There is also little evidence to suggest that decoding skills are being introduced effectively in Grade R.

8. **Urgent allocation of resources for large-scale reading remediation programs in all schools:** Finally, resources urgently need to be allocated so that reading remediation programs are available in all schools, and that schools are equipped with enough personnel and resources to support not just a few learners, but many. Teacher assistant support models have proven to be effective in initial pilot projects (Ardington & Henry, 2021), and could be further experimented with as a working model to support reading remediation at scale. Effective remediation is particularly necessary in a context where COVID-19 schooling disruptions have impacted severely on learner's reading development, particularly in the early grades (Ardington et al., 2021b).

REFERENCES

Abadzi, H., 2012. Developing Cross-Language Metrics for Reading Fluency Measurement, Developing Cross-Language Metrics for Reading Fluency Measurement. https://doi.org/10.1596/26819

Alcock, KJ, Ngorosho, D, Deus, C & Jukes, MCH. 2010. We don't have language at our house: Disentangling the relationship between phonological awareness, schooling and literacy. British Journal of Educational Psychology, 80(1), 55-76. Doi:IO.I348/000709909X424411

Anderson, N.J. 1999. Improving reading speed: Activities for the classroom. English Teaching Forum 37: 2–5.

Ardington, C. & Henry, J. 2021. Funda Wande Limpopo Impact Evaluation. Midline Report. SALDRU, University of Cape Town. Cape Town.

Ardington, C., Wills, G., Pretorius, E., Deghaye, A., Mohohlwane, & N., Menendez. 2021a. Benchmarking oral reading fluency in the early grades in Nguni languages. International Journal of Educational Development, 81, 102433. DOI: 10.1016/j.ijedudev.2021.102433

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.

Ardington, C.; Wills, G. & Kotze, J. 2021b. COVID-19 learning losses: Early grade reading in South Africa. International Journal of Educational Development, vol 86, 102480.

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.

Betts, E. A. 1946. Foundations of reading instruction, with emphasis on differentiated guidance. American Book Co.

Castles, A., Rastle, K. & Nation, K., 2018, 'Ending the reading wars: Reading acquisition from novice to expert', Psychological Science in the Public Interest 19(1), 5–51. https://doi.org/10.1177/1529100618772271

Dale, E.& Chall, JS. 1948. A formula for predicting readability and A formula for predicting readability: Instructions. Educational Research Bulletin, XXVII (January 21 and February 18, 1948), 11-54.

Dawes, A., Biersteker, L., Girdwood, E., Snelling, M. & Tredoux, C. 2017. Innovation Edge Briefing Document. The Development and Age Validation of the Early Learning Outcomes Measure (ELOM). Available: http://elom.org.za/wp-content/uploads/2017/07/ELOM-Briefing-Document.pdf

Deno, SL, Fuchs, LS, 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.

Department of Education and Training. 1988. Setswana. Terminology and orthography, no. 4. Government Printer Pretoria, South Africa.

DBE, 2011. Curriculum ad Assessment Policy Statement. Foundation Phase Grades 1-3. English First Additional Language. Department of Basic Education, South Africa. Pretoria. https://www.education.gov.za/Portals/0/CD/National%20Curriculum%20Statements%20and%20Vocational/CAPS%20ENGLISH%20FAL%20GR%201-3%20FS.pdf?ver=2015-01-27-155321-957

DBE, 2017. Summary Report: Results of Year 2 Impact Evaluation. The Early Grade Reading Study (EGRS). Department of Basic Education, South Africa. Pretoria. https://www.education.gov.za/Portals/0/Documents/Reports/EGRS%20Summary%20Report.pdf?ver=2017-08-17-090215-583

Department of Basic Education and the University of the Witwatersrand. 2020. Early Grade Reading Study 2017-2019, Waves 1-4 Merged [dataset]. Version 1. Pretoria: DBE and Wits [producers], 2020. Cape Town: DataFirst [distributor], 2020. DOI: https://doi.org/10.25828/qwx3-4m77

DBE, 2020a. Action Plan to 2024. Towards the realisation of Schooling 2030. August 2020. Department of Basic Education. Pretoria, South Africa.

DBE, 2020b. National Framework for the Teaching of Reading in African Languages in the Foundation Phase. Department of Basic Education. Pretoria, South Africa.

Dowd, A.J., Bartlett, L., 2019. The Need for Speed: Interrogating the Dominance of Oral Reading Fluency in International Reading Efforts. Comp. Educ. Rev. 63, 189–212. https://doi.org/10.1086/702612

Durgunoğlu, A & Öney, B. 1999. A cross-linguistic comparison of phonological awareness and word recognition. Reading and Writing: An Interdisciplinary Journal, 11, 281–299.

EGRS II. 2018. Year 2 report: Learner performance after the second year of implementation. Pretoria: Department of Basic Education.

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. doi:10.1007/s10648-011-9175-6

Fuchs, LS, Fuchs, D, Hosp, MK & Jenkins, JR. 2001. Oral reading fluency as an indicator of reading competence: A theoretical, empirical, and historical analysis. Scientific Studies of Reading, 5(3), 239–256.

Geva, E. & Zadeh, Z.Y. 2006. Reading efficiency in native English-speaking and English-as-a-second-language children: The role of oral proficiency and underlying cognitive-linguistic processes. Scientific Studies of Reading 10(1): 31–57.

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. Read. Teach. 59, 636–644. https://doi.org/10.1598/RT.59.7.3

Hasbrouck, J. & Tindal, G., 2017. An update to compiled ORF norms (Technical Report No. 1702). Eugene, OR, Behavioral Research and Teaching, University of Oregon.

Hoover, WA & Gough, PB. 1990. The simple view of reading. Reading and Writing: An Interdisciplinary Journal, 2, 127–160. doi: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.

Jimerson, S. R., Hong, S., Stage, S., & Gerber, M. 2013. Examining oral reading fluency trajectories among English language learners and English speaking students. New Approaches in Educational Research, 2(1), 3–11.

Jukes, M., Pretorius, E. Schaefer, M. Tjasink, K. Roper, M. Bisgard, J. Mabhena, N. 2020. Setting Reading Benchmarks in South Africa. Khulisa Management Services: Johannesburg. Department of Basic Education: Pretoria. Available from: https://pdf.usaid.gov/pdf_docs/PA00X1NZ.pdf.

Katz, J., 2020. Back to basics - comparing the orthographic, phonic and grammatical features of English and African languages to improve literacy teaching. Presentation at PrimTEd Literacy Working Group Seminar. Materials for literacy teacher programmes.

Katz, L., Frost, R., 1992. The Reading Process is Different for Different Orthographies: The Orthographic Depth Hypothesis, in: Frost, R., Katz, L. (Eds.), Orthography, Phonology, Morphology, and Meaning. Elsevier Science Publishers, Amsterdam, pp. 67–84. https://doi.org/10.1016/S0166-4115(08)62789-2

Kendeou, P, Papadopoulos, TC & 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. doi:10.1007/s11145-012-9361-z

Kaminski RA & Good RH. III. 1996. Toward a technology for assessing basic early literacy skills. School Psychology Review, 25(2), 215-227.

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

Kim, YG. 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, YG. (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

Kilpatrick, DA 2012. Essentials of assessing, preventing and overcoming reading difficulties. Hoboken, NJ: John Wiley & Sons.

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.

LaBerge, D., Samuels, S.J., 1974. Toward a theory of automatic information processing in reading. Cogn. Psychol. 6, 293–323. https://doi.org/10.1016/0010-0285(74)90015-2

Lekgoko, O. & Winskel, H. 2008. Learning to read in Setswana and English: cross-language transference of letter-knowledge, phonological awareness and word reading skills. Perspectives in Education, 26(4): 57 - 73.

Leppänen, I, Aunola, K, Niemi, P & Nurmi, J-E. (2008). Letter knowledge predicts Grade 4 reading fluency and reading comprehension. Learning and Instruction, I8, 548-564.

Lesaux, N. K. and L. S. Siegel (2003). The development of reading in children who speak English as a second language. Developmental Psychology 39 (6): 1005-1019.

Lipka, O. & Siegel, L.S. 2007. The development of reading skills in children with English as a second language. Scientific Studies of Reading 11(2): 105–131.

Liswaniso, B.L. 2021. The design and effects of a catch-up reading intervention for Grade 5 teachers and learners in Namibia. Unpublished PhD thesis, University of South Africa.

Maboya, M., 2020. Report on Foundational Skills of Literacy and Numeracy. Curriculum Policy, Support and Monitoring, Department of Basic Education, Pretoria, South Africa.

Malda, M, Nel, C & van de Vijver, FJR. 2014. The road to reading for South African learners: The role of orthographic depth. Learning and Individual Differences, 30: 34-45. http://dx.doi.org/10.1016/j.lindif.2013.11.008

Makaure, ZP. 2021. The contribution of phonological processing skills to early literacy development in Norther Sotho-English bilingual children: A longitudinal study. Unpublished doctoral thesis, University of South Africa, Pretoria.

Melawana ya Mokwalo le Mopeleto SETSWANA 2008. PanSALB. Arcadia. Ramagoshi, R.M. 2020. A re Bueng Setswana.

Melby-Lervåg, M, Lyster, S & Hulme, C. 2012. Phonological skills and their role in learning to read: A meta-analytic review. Psychological Bulletin, 138(2):322-52. doi: 10.1037/a0026744.

Menendez, A & Ardington, C. 2018. Impact Evaluation of USAID/South Africa Story Powered School Programme – Baseline.

Mutema, F. 2021. The development of reading literacy skills in the early years of primary schooling: a case of four Zimbabwean schools. Unpublished doctoral thesis, University of South Africa, Pretoria.

Piper, B. 2009. Integrated Education Program: Impact Study of SMRS Using Early Grade Reading Assessment in Three Provinces in South Africa. RTI International. Research Triangle Park, NC.

Piper, B., & Zuilkowski, SS. 2015. Assessing reading fluency in Kenya: Oral or silent assessment? International Review of Education, 61(2): 153-171. DOI 10.1007/s11159-015-9470-4

Pretorius, EJ. 2014. Supporting transition or playing catch up in Grade 4? Implications for standards in education and training. Perspectives in Education 32(1): 51-76.

Pretorius, E.J., Spaull, N., 2016. Exploring relationships between oral reading fluency and reading comprehension amongst English second language readers in South Africa. Read. Writ. 29, 1449–1471. https://doi.org/10.1007/s11145-016-9645-9

Pretorius, EJ. 2018. Getting it right from the start: some cautionary notes for early reading instruction in African languages. In N Spaull & JP Comings (Eds). Improving Early Literacy Outcomes: Curriculum, Teaching and Assessment (pp63-80). Leiden: IBE/BRIL

Psychological Services, Broward County Public Schools. 2012. Oral Reading Fluency Data for English Language Learners (ELL). August 2012.

Room to Read, 2018. Setting Data-Driven Oral Reading Fluency Benchmarks Guidance Note. Research Triangle Park, NC.

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

Samuels, S.J., Flor, R.F., 1997. The importance of automaticity for developing expertise in reading. Read. Writ. Q. https://doi.org/10.1080/1057356970130202

Seymour, PHK, Aro, M & Erskine, JM. 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, 2019." June 20. https://www.gov.za/speeches/2SONA2019.

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

Spaull, N., Pretorius, E. & Mohohlwane, N., 2020, 'Investigating the comprehension iceberg: Developing empirical benchmarks for early-grade reading in agglutinating African languages', South African Journal of Childhood Education 10(1), a773. https://doi.org/ 10.4102/sajce.v10i1.773

Stanovich, KE. 2000. Progress in Understanding reading: Scientific foundations and new frontiers. New York: The Guildford Press.

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

Torppa, Georgiou, Lerkkanen, Niemi, Poikkeus, Nurmi, 2016. Examining the Simple View of Reading in a Transparent Orthography: A Longitudinal Study From Kindergarten to Grade 3. Merrill. Palmer. Q. 62, 179. https://doi.org/10.13110/merrpalmquar1982.62.2.0179

University of Oregon. 2021. 8th Edition of Dynamic Indicators of Basic Early Literacy Skills (DIBELS) Administration and Scoring Guide, 2021 Edition. Eugene, OR: University of Oregon. https://dibels.uoregon.edu/sites/dibels1.uoregon.edu/files/2021-10/UO_Dibels8_Scoring_Guide_100121.pdf

Wang, Z., Sabatini, J., O'Reilly, T., Weeks, J., 2019. Decoding and Reading Comprehension: A Test of the Decoding Threshold Hypothesis. J. Educ. Psychol. 111, 387–401.

Wills, G. & van der Berg, S. 2020. Measuring School Leadership and Management and Linkages with Literacy: Evidence from rural and township primary schools in South Africa. Educational Management Administration and Leadership. 45 (5) 708-731. DOI: 10.1177/1741143220915923

Zenex Literacy Project: Learner reading assessment results. 2017. ERA: Stellenbosch University.

APPENDIX

Instrument development: Pilot samples

Piloting of the learner instruments was conducted in three rounds. Table A1 summarizes the sample size by grade for each round of piloting. The first pilot was very small with a target of around 30 learners in each grade and was conducted in two schools in Gauteng. Unfortunately, this meant that around half the assessed learners were not Setswana home language speakers. We decided to include all learners in the analysis for the first pilot as the non-Setswana home language learners actually performed slightly better than the Setswana home language learners and we were concerned about further reductions to the already small sample. The second and third pilots were larger and were conducted in schools in the North-West province. The percentage of Setswana home language speakers increased to between 88% and 95% of the samples for each grade. Given the purpose of the pilot, the assessments were suspended if learners were unable to read at least one word of the first passage. This allowed more time to assess learners who were able to read. Although 44% of the Grade 3 learners in pilot 2 were unable to read a word, the fieldwork team were still able to assess 62 learners who could read

Table A 1. Pilot samples

	Pilot 1			Pilot 2			Pilot 3		
	Learners	% Setswana home language	% non-readers	Learners	% Setswana home language	% non-readers	Learners	% Setswana home language	% non-readers
Grade 3	30	50%	3%	111	88%	44%	125	93%	16%
Grade 4	27	52%	10%	74	91%	20%	105	89%	11%
Grade 7	29	45%	0%	85	95%	8%	95	95%	1%

Setswana data: additional tables

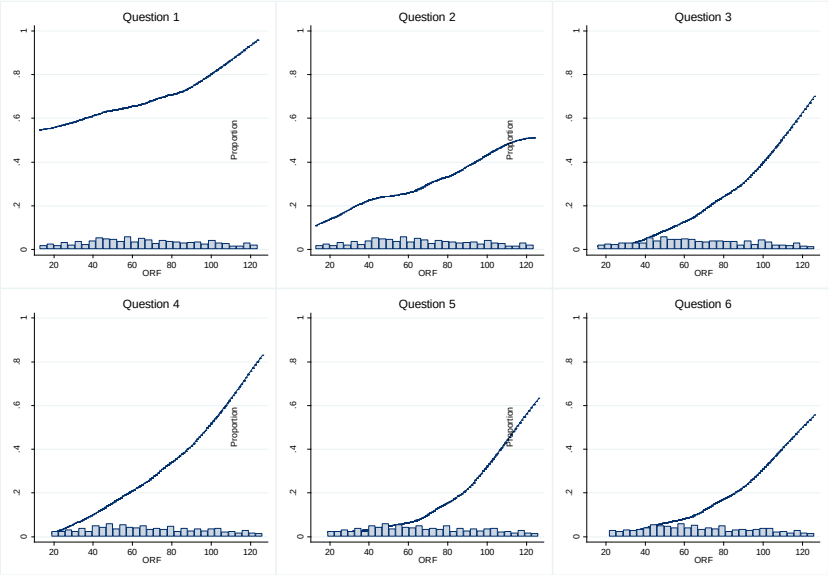
Table A 2: Setswana letter-sound knowledge and complex consonant and diacritic assessments in EGRS I and RSP

Details			Letter sound knowledge			Separate complex consonants & diacritic test
Grade	year	Term	Time allowed	Includes complex consonants & diacritics	Max possible items	
I	2015	I	60s	No	110	
I	2018	III	60s	Yes	110	
I	2015	IV	60s	No	110	
2	2016	IV	60s	No	110	
3	2018	III	60s	Yes	110	
3	2021	III	60s	No	110	60s allowed
4	2018	III	60s	Yes	110	
4	2021	III	60s	No	110	60s allowed

*s = seconds

Setswana benchmarking results: Additional tables and figures

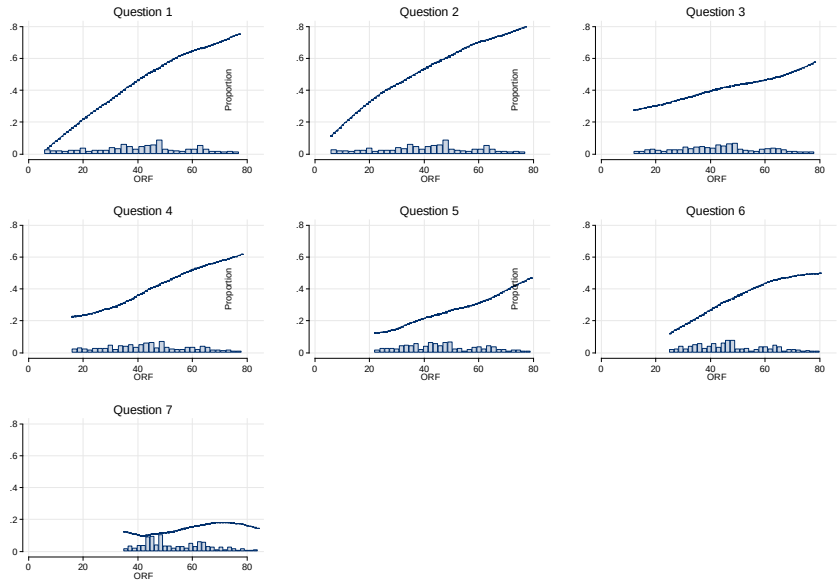
Figure A 1: Relationship between fluency and individual comprehension questions: Grade 3, passage 2, Setswana



Grade 3 Passage 2

Source: RSP 2021, own calculations. 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.

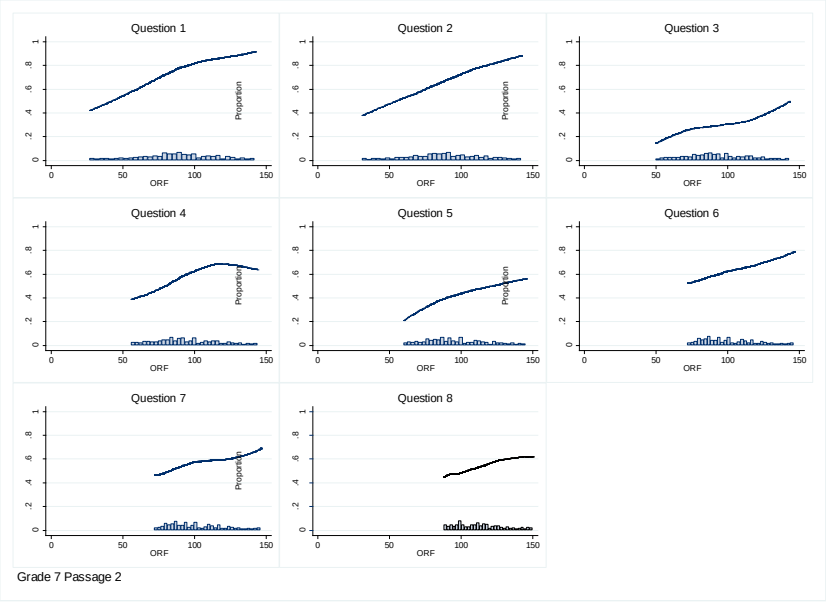
Figure A 2: Relationship between fluency and individual comprehension questions: Grade 4, passage 2, Setswana



Grade 4 Passage 2

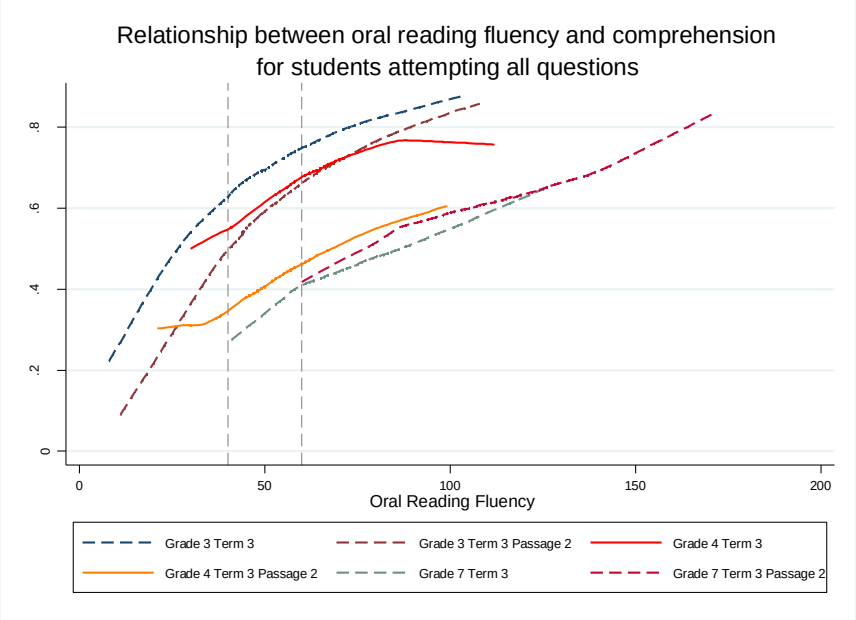
Source: EGRS I 2021, own calculations. Notes: The histogram bars reflect each grade sample's distribution of ORF scores. The lines are locally weighted polynomial regressions of ORF against the proportion getting the comprehension question correct.

Figure A 3: Relationship between fluency and individual comprehension questions: Grade 7, passage 2, Setswana



Source: EGRS I 2021, own calculations. 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 A 4: Relationship between fluency and comprehension for students attempting all comprehension questions



Source: EGRS I /RSP 2021 data.