unesco
Institute for Statistics

Benchmarking for Precursor Skills in Reading: SDG 4.1.1a

South Africa Case Study

April 2025

Contents

1. Introduction 4

2. Background: Benchmark Precursor Skills in South Africa 6

2.1 A focus on Foundational Literacy 6

2.2 Assessment Landscape for Measuring Early Grade Reading Proficiency 10

2.3 The Gap: Measuring Precursor Skills for Reading 15

3. Planning Process for Benchmarking 16

3.1 Deciding on Precursor Skills to Benchmark 16

3.2 Design Report for Methodology Selection 18

3.3 Benchmarking for a Multilingual Context 24

3.4 Capacity Requirements for Establishing Benchmarks 26

3.5 Funding for Benchmarking 27

4. Activities Process for Establishing Reading Benchmarks 31

4.1 Development and Review of Assessment Tools 32

4.2 Piloting of Assessment Tool Items 34

4.3 Field Officer Training and Data Collection 36

5. Engaging and Disseminating 40

5.1 Collective Ownership 40

5.2 User Specific Dissemination 41

5.3 Approval and Public Launch of Benchmark Reports 42

6. Usage of Benchmarks 43

6.1 Classroom 43

6.2 Schools and Districts 47

6.3 National and Provincial Administrations 49

6.4 Diffused Uses 51

7. Reflections: Lessons, Gaps, and Next Steps 54

7.1 Theme 1: Importance of benchmarks for literacy strategies 54

7.2 Theme 2: Funding 55

7.3 Theme 3: Diversity and equity 55

7.4 Theme 4: Linguist and Language Specialist Role 56

7.5 Theme 5: Development of Instruments 57
7.6 Theme 6: International and local expressions 58
7.7 Theme 7: Key success factors 59
7.8 Theme 8: International considerations 59
8. Conclusion 60
9. References 61
Annexure A: Interviewee List 63

List of tables

Table 1: South African National Targets, Indicators, and Commitments for Early Literacy 8
Table 2: South African Catalogue of Early Literacy Assessments 13
Table 4: Summary of the South Arican Early Grade Reading Benchmarks (as of April 2025) 26
Table 5: Benchmarking Data Approaches Taken for Reports and Supporting Partners 28
Table 6: Comprehensive List of Tasks Administered in Various Benchmarks Data Collections 32
Table 7: Usage of South African Early Grade Reading Benchmarks 43
Table 8: Publications citing or using South African Established Benchmarks 52

List of figures

Figure 1: Precursory to Derivative Skills in the Reading Process.....16
Figure 2: Accuracy and speed in reading connected text: Sesotho-Setswana languages.....20
Figure 3: Accuracy and speed in reading connected text: Nguni languages.....20
Figure 4: South African Languages of Learning, Teaching and Assessment 24
Figure 5: Different skills needed for reading benchmark development.....27
Figure 6: Activities Process for Establishing Benchmarks 31
Figure 7: Piloting Process for Instrument Finalisation 34
Figure 8: The Six Golden Rules for Task Administration Protocol.....37
Figure 9: Protocol for Administering Letter Sound Recognition Task 38
Figure 10: Theory of Change for Teacher Benchmarks Usage Pilot Study 45

1. Introduction[1]

The United Nations Educational, Scientific and Cultural Organization (UNESCO) Institute for Statistics (UIS) is actively working to enhance the reporting on the Sustainable Development Goal (SDG) 4.1.1a, which measures the percentage of children achieving minimum proficiency in reading and mathematics by grades 2 or 3. Currently, the number of countries reporting on this indicator remains substantially lower when compared to those reporting on primary school enrolment. For example, only 37 countries report on learning outcomes at the grades 2 or 3 level, whereas 203 countries report primary school enrolment. This slow pace in reporting is concerning, as it would take many years to reach comparable levels of enrolment reporting.

In 2024, UIS established a Technical Advisory Group (TAG) to explore precursor skills for reading comprehension and to develop appropriate benchmarks using advanced statistical techniques. Additionally, UIS commissioned case studies to examine how countries have independently established their own benchmarks, supplementing the insights drawn from centralized databases.

South Africa has been identified as one possible case to map alternative approaches to measure reading. This would serve as an example of an effort to measure precursor skills and serve as an option as UIS advises the relevant countries. This document is the report for this case, providing some analysis and progress on the case of South Africa.

The authors of this report were the lead team developing the benchmarks for South Africa for South Africa in their capacities as national Department of Basic Education (DBE) officials in the Research Coordination, Monitoring, and Evaluation directorate. The report draws on the experience of the authors in the process of deciding on the precursor skills to benchmark;

[1] This report was commissioned by the UNESCO Institute for Statistics (UIS) and compiled by Mbali Lesang Sebaeng and Nompumelelo Nyathi-Mohohlwane.

development details; dissemination and possible uses; and finally, reflections and lessons. This report also draws from a desktop review of source documentation and interviews with a range of stakeholders involved in the process at different stages.

2. Background: Benchmark Precursor Skills in South Africa

In the global community's earnest efforts to ensure equal and quality education for all lies the rationale that in achieving this, the ideals of democracy, international corporation and the development of humanity will prevail. One of the critical gates to accessing education- of any quality- is the ability to read. South Africa has spent the last three decades detangling and undoing the damage done across all spheres of society by the apartheid government. One of the casualties has been the education sector. The system inherited from the discriminatory and violent regime was one that showed systemic inequalities that prejudiced the majority of the country's citizens (Department of Basic Education, 2024). The 300 years of colonisation followed by half a century of legal disenfranchisement in resources, training and development of teachers, infrastructure, and research in education in indigenous languages continues to have consequences in the educating of over 80% of the population (Statistics South Africa, 2023). This section covers how South Africa has contended with the importance on early grade reading attainment in light of the historical context in recognition that achieving this will enable the development of the larger society.

2.1 A focus on Foundational Literacy

The foundational years in formal education, typically the first three years, are set out for the intentional instruction of skills that will develop a learner into a proficient reader. This is because, in the subsequent years, learners will rely on this skill to access content and skills in their respective curriculums. The inability to master this skill in the first three grades has serious long-term implications in education, learning, and becoming productive members of society.

In South Africa, Grade 4 marks another pivotal transition for over 70% of learners in publicly funded schools (Sapire and Roberts, 2017). These learners not only transition from learning to read, to reading to learn, but also switch the language they learn in from their home languages

to English. It follows that learners who do not master this transition are at increased risk of academic failure and school dropout. Addressing early literacy deficiencies is essential to ensure learners can read fluently and comprehend texts for use in various contexts. It is for this reason that South Africa has set literacy as a priority in the nation's goals. The table below summarises the key national and international targets and indicators for early literacy in the first six years of formal schooling that South Africa has set in place over the last decade.

Table 1: South African National Targets, Indicators, and Commitments for Early Literacy

National Outcomes	**National Development Plan (NDP) Vision 2030 (adopted 2012)**	
	Goal: Learners home language should be used a s a medium of instruction for longer and English is introduced much earlier in the Foundation phase.	**Target:** 90% of learners in grades 3, 6 and 9 must achieve 50% or more in the Annual National Assessment in literacy, numeracy/ mathematics and science
	Objective: Improve the school system, including increasing the number of students achieving above 50% in literacy and mathematic	**Target:** 2022 Target 600 in SACMEQ
	Action Plan for 2024 Towards the Realisation of Schooling 2030	
	Goal 1 Target: Increase the percentage of learners in Grade 3 who, by the end of the year, have mastered the minimum language and numeracy competencies for Grade 3 in the Annual National Assessments to 95%.	
	Goal 2 Target: Increase the percentage of learners in Grade 6 who, by the end of the year, have mastered the minimum language and mathematics competencies for Grade 6 in the Annual National Assessments to 95%.	
	Goal 7 Target: Improve Grade 6 Southern and Eastern African Consortium for Monitoring Educational Quality (SACMEQ) Assessment average score in languages to 640 by 2030 (2007: 495 2013:538).	
	Medium Term Strategic Framework (MTSF) 2014-2019	
	Impact Target 2019: 75% of learners in Grade 3 achieving at the required level in the Annual National Assessments (ANA) in literacy and numeracy (2013 Baseline: 40% Literacy 50% Numeracy)	**Impact Target 2019**: 75% of Grade 6 learners achieving at the required level in the ANAs in Home Language (2013 Baseline: 68%) First Additional Language (41%), and Mathematics (27%)
	Impact Target 2019: 550 average score obtained by Grade 6 learners in language in the Southern and Eastern African Consortium for Monitoring Educational Quality (SACMEQ) assessment (2007 Baseline: 495)	
	Medium Term Strategic Framework (MTSF) 2019-2024	
	Outcome:10-year-old learners enrolled in publicly funded schools read for meaning	**Target:** All provinces implementing National Reading Plan
	Target: All languages have Grade 1-3 home language literacy lesson plans.	**Target:** 100% of Grade 3 learners who learn through indigenous languages have graded reading books
	Indicator: Grade 3 performance in the new Systemic Evaluation: Reading (targets to be determined after first assessment)	**Indicator:** Grade 6 performance in the new Systemic Evaluation: Maths: Literacy: (targets to be determined after first assessment)

	Target: Improve Grade 6 average score for literacy in SACMEQ to 600 by 2020	**Target:** Improve average score in PIRLS from to 355 by 2021
	Target: 100% of schools have received the Early Grade Reading Assessment tools	**Target:** 100% of foundation phase teachers trained in teaching reading & numeracy
	Medium Term Development Plan (MTDP) 2024-2029	
	Outcome: Improved education outcomes and skills	
	Target: 40% of Grade 3 learners Reading above 'evolving level' as assessed through the South African Systemic Evaluation by 2029/30 (baseline 20%)	
International Commitments	**African Union Goals & Priority Areas of Agenda 2063**	**African Union Strategic Objectives (SO)** *Continental Educational Strategy for Africa* 2016-2025
	Goal 2: Well-Educated Citizens and Skills Revolution underpinned by Science, Technology and Innovation	**SO6:** Launch comprehensive and effective literacy programmes across the continent to eradicate the scourge of illiteracy
	Target: 100% literacy rate by 2025	
	Sustainable Development Goals (SDG) Targets 2030	
	Target 4.1: By 2030, ensure that all girls and boys complete free, equitable and quality primary and secondary education leading to relevant and effective learning outcomes	
	Indicator 4.1.1: Proportion of children and young people by 2030 (a) in grades 2/3; (b) at the end of primary; and (c) at the end of lower secondary achieving at least a minimum proficiency level in (i) reading and (ii) mathematics, by sex	**Indicator 4.1.6:** Administration of a nationally representative learning assessment (a) in Grade 2 or 3; (b) at the end of primary education; and (c) at the end of lower secondary education
	Target 4.5: By 2030, eliminate gender disparities in education and ensure equal access to all levels of education and vocational training for the vulnerable, including persons with disabilities, indigenous peoples and children in vulnerable situations	**Indicator 4.5.2:** Percentage of students in a) early grades, b) at the end of primary, and c) at the end of lower secondary education who have their first or home language as language of instruction

2.2 Assessment Landscape for Measuring Early Grade Reading Proficiency

As can be gleaned from the previous section, South Africa has utilised national, regional, and international indicators to monitor the progress towards achieving literacy in the early years and set medium-term and long-term targets for itself. Below is a brief discussion on the two main sources for targets that continue to be included in sector planning documents for measurement data- that is the National Systemic Evaluations and the Progress in International Reading and Literacy Study (PIRLS). The Annual National Assessment (ANAs) is included in the discussion to provide context for the new Systematics.

The Annual National Assessments

The Annual National Assessments (ANAs) were an assessment programme intended to signal achievement levels of learners in literacy and numeracy in grades 1 to 6 and 9. The ANAs were designed, developed, and supplied to schools by the National Department of Basic Education. The assessments were administered in two ways, the first, was sample-based and completed by an external service provider; the second was universal and teachers were tasked with administering, marking, moderating, and reporting (Nuga Deliwe, 2017). The ANAS were administered between 2011 and 2015 in all public schools. The cessation of these assessments was largely due to a critique of the accountability practices that followed their implementation and a perceived lack of support for improvement.

The Systematic Evaluations

After a ten-year absence of an assessment system equivalent to the ANAs for the primary years, the new model of Systemic Evaluations was introduced in 2022 to fill the gap. The new Systemic Evaluations assesses the performance of the education system in learning outcomes for Mathematics and Languages in Grades 3, 6 and 9 against identified targets set out in the Medium-Term Strategic Frameworks (MTSF) and the National Development Plan (NDP). This is a significant effort, assessing over 12000 learners nationally. The baseline statistics for the

Systemics were disseminated in 2024 and the targets against the baseline for 2029/30 have been included in current five-year Medium-Term Development Plan (MTDP) 2024-2029.

Progress in International Reading and Literacy Study

South Africa is one of 57 countries that participate in PIRLS, and one of only two African countries participating in the last cycle. The study is a cross-national survey to monitor reading comprehension levels among Grade 4 learners in their respective languages of learning and teaching. Since the first cycle in 2006, South Africa has seen significant improvements between the 2006 and 2016 cycles (Mohohlwane et al,.2023). In fact, between 2011 and 2016, it achieved the third steepest incline after Morocco and Oman (Gustafsson, 2020) with the proportion of learners reaching the lowest internation reading comprehension benchmark rising from 7% to 22%. Despite these gains, the country's performance remains suboptimal. The COVID-19 pandemic further exacerbated this issue with the percentage of learners who were able exemplify their basic understanding of a grade appropriate text declining to 19% in 2021 (Department of Basic Education, 2023).

Additional Measurements for Early Literacy

The assessments discussed in this report form part of a larger literacy assessment ecosystem. What follows is Table 2: South African Catalogue of Early Literacy Assessments - a comprehensive table of all literacy assessments that are conducted for various purposes and audiences. In addition to the purposes, the table includes each assessment's strengths and limitations; information on the languages, grades, cycle intervals, statistical representation of samples, its usefulness for the classroom and the ease of understanding the data for different stakeholders.

This table was initially produced in a concept note to advocate for the necessity of the last assessment highlighted in green- Funda Uphumelele National Survey (FUNS). The need for a national measurement tool that tracks the system's progress in the attainment of the precursor reading skills in the early grades was identified in the development of the local early grade reading benchmarks. At the time of writing, the data collection for the inaugural Funda Uphumelele National Survey (FUNS) 2025 was completed with data cleaning and reporting

underway. This assessment tool is set to provide the baseline national, provincial, and linguistic statistics of learners' performance against the reading benchmarks. Further details on this can be found in Section 6: Usage of Benchmarks (National and Provincial Administrations) of this report.

Table 1: South African Catalogue of Early Literacy Assessments

Literacy Assessment	Purpose	Strengths	Limitations	Grade & Language	Intervals by years	Comparability over time	Nationally Representative	Provincially representative	Classroom use	Ease of communicating the results
PIRLS: Progress in Reading Literacy Study	Written reading comprehension	○ Measures the ultimate skill of comprehension ○ Nationally representative ○ Lang representative ○ International ○ Secure Assessment	○ Only measures reading comprehension ○ Floor effects – most children do not reach the lower threshold ○ At the end of Grade 4 – late for Foundation Phase remediation	Grade 4 11 Languages	Every 5 years	✓	✓			Point scale is difficult to communicate
EGRA: Early Grade Reading Assessment	Diagnostic test on early skills	○ Focuses on early skills ○ Known by teachers and subject advisors ○ Used for national reading monitoring	○ Use and purpose not fully aligned – classroom vs PED and national ○ Not nationally comparable. ○ Diagnostic teacher info – limited access and reliability Test not secure ○ Not enough tests	Grades 1-3 11 Languages	3 times a year				✓	Scores need updating Termly progress needs interpretation Overall scores need interpretation.
South African Systemic Evaluations	Written comprehension at exit grades	○ National comparison ○ Scaled with anchor items ○ Scale linked to SDG & CAPS	○ Only measures comprehension	Grades 3, 6 and 9 11 Languages	Every 3 years	✓	✓	✓		Point scale with proficiency levels

Literacy Assessment	Purpose	Strengths	Limitations	Grade & Language	Intervals by years	Comparability over time	Nationally Representative	Provincially representative	Classroom use	Ease of communicating the results
Assessment for Learning	Formative feedback	○ Ongoing assessment ○ CAPS aligned ○ Accessible to all teachers	○ Assumes teachers have remedial skills	All grades 11 Languages	Each Topic/ Ongoing				✓	Teachers may understand
SEACMEQ: Southern & East African Consortium for Education Quality	Regional Literacy &Numeracy comparison	○ Regional comparison ○ African curriculum alignment	○ Delay in analysis and reporting ○ Only measures ultimate skill ○ Intermediate phase focus – not Foundation Phase	Grade 6 11 Languages	Every 7-8 years	✓	✓			Well understood at a policy level, regionally and internationally
ELNA: Early Learning National Assessment	Measuring school readiness	○ Pre-schooling measure ○ Pre-literacy skills ○ School readiness alignment	Only for reception years and not all Foundation Phase	Grade R or start of Grade 1 11 Languages	3 years	✓	✓	✓		To be confirmed
Thrive by Five	Measures cognitive dev for 4-5 year old	○ Pre-literacy skills ○ Cognitive dev skills ○ Pre-school focus	○ Only for reception years and not all Foundation Phase	Grades RR & R 11 Languages	3 years	✓	✓	✓		Easy dashboard, policy notes
FUNS: Funda Uphumelele National Survey	10-year-old reading & earlier or later attainment	○ Comparison to new African reading benchmarks by skill and grade ○ Nationally representative ○ Provincial representative ○ Language representative ○ CAPS alignment	○ Limited to selected oral skills ○ Assumes teachers have medial response	Grades 1-4 11 Languages & EFAL	3 years (new)	✓	✓	✓		Percentage reaching benchmarks by skill, grade and lang

Source: Department of Basic Education, Funda Uphumelele National Survey Concept note 2024

2.3 The Gap: Measuring Precursor Skills for Reading

Notwithstanding South Africa's committed to utilising reliable data sources such as PIRLS, SEACMEQ, and the Systematics for monitoring; target-setting; and programme impact evaluations, and assessments, a key knowledge gap persisted. This paucity was because the assessment tools all measure one of the final skills attained in the reading development journey- written comprehension. The earliest this skill can be usefully and appropriately measured is in the final year of the Foundation Phase (Grade 3) before the switch to reading for access to content occurs. Partiality to measuring against this is not an issue for systems that have large proportions of learners effectively reaching the required targets. However, in the case of South Africa, this is far from the reality.

South Africa's PIRLS trend results in all four cycles have never displayed more than a quarter of Grade 4 learners being able to access the lowest order skill in the four-tier comprehension taxonomy: 'focus on and retrieve explicitly stated information' in a text (Department of Basic Education, 2023). This indicates that the reading challenges learners face stem from setbacks in the development of earlier literacy skills taught in the Foundation Phase (Grades R-3). Therefore, measuring exclusively for this skill at a systemic level only provides intel on part of the problem and does not account for earlier skills taught in grades 1, 2, and 3 that contribute to reading proficiency.

To address this, the DBE's Research Coordination, Monitoring, and Evaluation (RCME) directorate, in collaboration with academics, practitioners, funders, and international benchmarking specialists took to exploring how information on performance in the precursory skills relevant to each of the first three grades could be best measured and which of these interdependent skills would be more efficacious to benchmark. What resulted is detailed in the rest of this report and has been adopted as the South African Early Grade Reading Benchmarks for the 11 Languages of Learning, Teaching, and Assessment used in the first three grades. The goal of these benchmarks is to establish what are the critical milestones predictive of future reading comprehension learners should meet at the end of each foundation phase year.

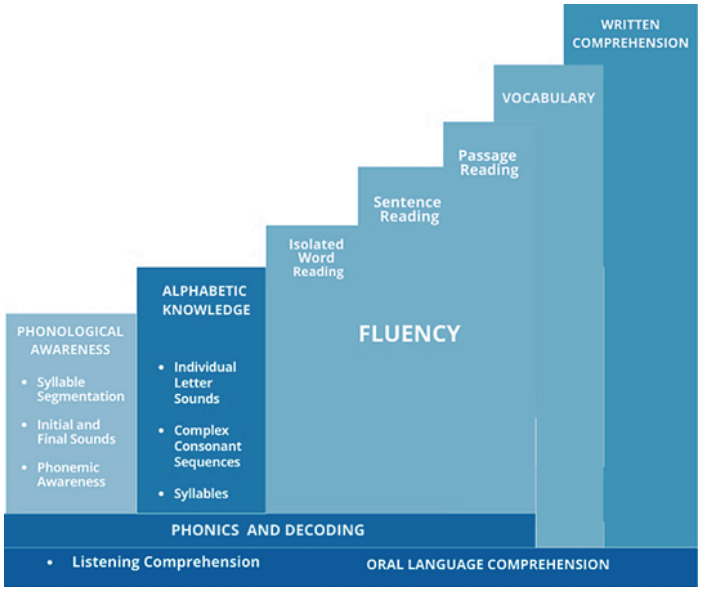
3. Planning Process for Benchmarking

A design process was implemented at the start of the benchmarking and is summarised here. The DBE convened a multi-stakeholder design workshop including RTI International specialists who had led similar processes in other developing countries in Africa. The design phase informed decisions on the precursor skills, grades, methods and experts needed. The process culminated in a design report, Setting Reading Benchmarks report (Khulisa Management Services, 2020).

3.1 Deciding on Precursor Skills to Benchmark

There are a range of early skills that are a precursor to attaining reading proficiency levels necessary for meaningful engagement with written materials. The interaction of these skills is a complex process, and oftentimes language-dependent. The figure below is adapted from the Wills, et al chapter on the foundational skills trajectories observed using empirical longitudinal data (Spaull and Pretorius, 2022). It is a helpful visualisation of how these diverse skills build on and leverage one another to achieve the goal of multimodal comprehension at different levels.

Figure 1: Precursory to Derivative Skills in the Reading Process



When deciding on the skills to benchmark, two aspects were considered, firstly, deciding on the category of skills such as decoding, word reading, passage reading and vocabulary. The next step was determining the specific subset within the skill, such as letter sounds, syllable reading within decoding. Within words, the type of words-sight words and decodable words amongst others; and passage reading including the type of passages such as information, narrative, fiction or non-fiction texts.

The second aspect was measuring the underlying analytical approach associated with proficient reading. These include accuracy, fluency and prosody. Accuracy and fluency lent themselves to more consistent measurement than prosody. A decision was taken to benchmark based on fluency because of the strong relationship between fluency and comprehension. However, accuracy was factored into the measurement. The four categories of reading comprehension linking accuracy and comprehension were considered (DBE 2011, 64). The Independent level with 98% accuracy, the Instructional level at 95%, 90-94% borderline, and below 90% the frustration levels were all adopted as thresholds to consider when determining the benchmarks.

Although evidence on the predictive validity of accuracy was largely based on the English language, this was taken as an initial measure. The relationship for different languages was expected to be affected by linguistic differences such as the relationship between fluency and comprehension, the appropriate speed within each language, and typical fluency development. The data in this process was used to examine these relationships and it was found to be consistent for South African languages.

A decision was made to select constrained skills that are consistent across languages, easy to measure by teachers and fieldworkers and easy to communicate. Caution was taken to limit the number of skills to reduce the potential for confusion. Letter-sounds were chosen for their consistency across languages, and the limited errors in their measurement as well as their finite number. This is in addition to their significance as a contributor to decoding and overall reading development. Passage reading, referred to as Oral Reading Fluency (ORF) was selected as the skill with the most robust rationale for reading. The general existence of ORF

benchmarks for both Home Language and additional language is extensive along with some emerging research for Swahili and Setswana amongst Africa languages. While data collected included information and narrative texts, the differences in benchmarks were stark depending on the type of text used. Narrative, fictional texts were used for the benchmarks after consideration.

The ORF assessment was categorised into ORF fluency followed by ORF comprehension. Following robust discussion on the classification of the ORF comprehension questions, an agreement was reached to follow the PIRLS question taxonomy where questions were classified into four; Literal, Straightforward Inference, Interpret and Integrate, Evaluate and Examine (Mullis and Martin, 2019). Other taxonomies including Bloom's of Barret's would have also been acceptable, the emphasis was on ensuring that interpreting ORF comprehension extended beyond Literal Questions.

3.2 Design Report for Methodology Selection

Different approaches to benchmarks were considered including norm-reference, criteria-referenced, and data-driven approaches. The approach we selected combined these in different ways but was largely data-driven. The skills and competencies benchmarked align with the current national curriculum which specifies criteria for learning in each grade. Secondly, a decision was taken to base the benchmarks on existing learners, rather than aspiration only. We did not only include learners that would reach the benchmarks but rather set aspirations based on deliberately sampled typical learners in the public no-fee paying parts of the education system. This would enable developmental benchmarks that capture a range of performance, making them useful for the bottom-end of learners.

Our analytical approach aligns with the decoding threshold hypothesis put forward by Wang et al. (2019) where there are at least two thresholds. Until decoding reaches a lower bound threshold comprehension is stagnant, however, there may also be an upper threshold above which gains plateau. This suggests a range within which benchmarks may be set based on the data and language. For each of the languages benchmarked, this relationship differed. For instance, for the Nguni languages, we found that speed and accuracy developed at a steep rate

but above 95% accuracy gains in speed plateau (Mohohlwane, et al, 2022). This was at above 20 words per minute for Nguni languages but at 40 words per minute for the Sesotho-Setswana languages. The exact threshold or benchmark differed significantly by language group (Ardington et al. 2020; Wills et al. 2022; Ardington et al. 2022). This showed a non-linear relationship between fluency and accuracy by language group. The statistical method used is a non-parametric analysis of data regularities and thresholds (Ardington et al. 2020; Wills et al, 2022).

To further test this approach, we consider longitudinal data of learners in later grades who read with adequate comprehension in Setswana and examine their performance in earlier grades. This aligns well with the benchmarks. While there is often a critique Dowd & Bartlett, 2019) that fluency benchmarks might encourage speed at the cost of comprehension or be inappropriate for African languages we do not find this. There were very few learners that were both slow and accurate in Nguni and Sesotho-Setswana languages (Ardington et al., 2021). See the figures below.

Figure 2: Accuracy and speed in reading connected text:
Sesotho-Setswana languages

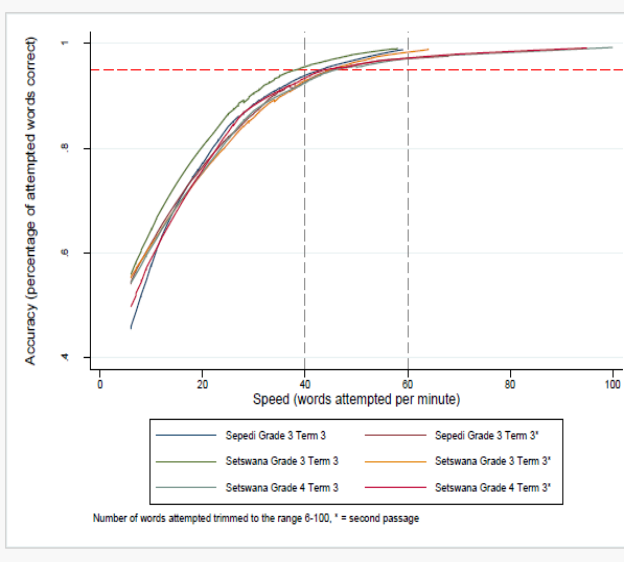
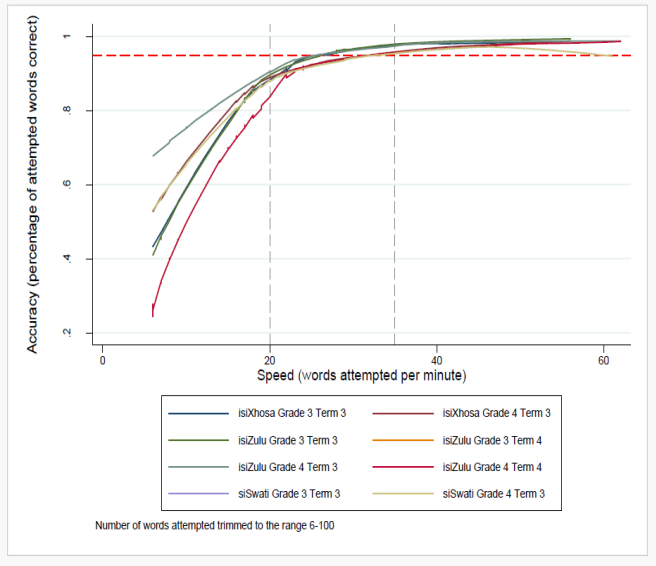


Figure 3: Accuracy and speed in reading connected text:
Nguni languages



Source: Mohohlwane et al. 2023

A further core deviation from the common methods established was in two ways (Khulisa Management Services, 2020; Mohohlwane, 2022), ORF duration and the sample used to benchmark (Jukes et al., 2018a, 2018b). Firstly, benchmarks were typically based on the Early Grade Reading Assessment (EGRA) with learners reading for one-minute then asked questions aligned with how far they read. In low literacy contexts, this severely limits the number of ORF comprehension questions answered and the number of learners used for benchmarking to those at the top end. The resulting benchmarks then limit the developmental nature to track incremental gains; making this approach inappropriate for South Africa. In our approach we marked both how far learners were at 1 minute and extended this to 3 minutes. Resultantly, most learners reached the end of the passage within 3 minutes, allowing them to answer a range of questions and extending the sample we could use to create the benchmarks.

Three data sources were identified for benchmarks, The first, using existing data; the second, topping-up planned data collections to include benchmark-specific skills, protocols and grades; and the third, a unique data collection for benchmarking specifically. All three methods

have been used in South Africa. The languages selected for benchmarking first were chosen based on the availability of existing data. Attempts to consolidate existing data for consideration included communication with the South African Literacy Association members as well as known partners and NGOs working in education. The table below summarises the kinds of data sources used for the development of the different benchmarking reports.

Table 3: Summary of Data Sources used per Benchmarking Reports

Data Source Description	Language Data	Data Collection Years Range	Learner Sample Size (unique learners)	Grades	School Sample Size	School Type	Provinces
Existing Data							
Benchmarks Established: ***Nguni Language Group*** *(2020)*							
Collated early grade reading assessment data from 5 studies: *EGRS II; Funda Wande; Literacy for Leadership; Story Powered Schools;* and *ZenLit*	IsiXhosa (HL) IsiZulu (HL) Siswati (HL)	2017-2019	16400	Grades 1-6	660 schools	Quintile 1-3[2]	Eastern Cape Gauteng KwaZulu-Natal Mpumalanga
Top-Up Data Collection							
Benchmarks Established: ***Sesotho-Setswana Language Group*** *(2022)*							
Room to Read existing data as well as scheduled *Funda Wande* data collection for programme evaluation was extended to increase sample size and grades for the purpose of benchmarking.	Sepedi (HL)	2015-2021	24686	Grades 1-4, and 6-7	429 schools	Quintiles 1-3	Free State Limpopo North West
Existing data for Sesotho from *Save the Children*	Sesotho (HL)						
Top-up to existing longitudinal on EGRS I and RSP data during the Sustainability Impact Evaluation data collection in 2021	Setswana (HL)						
Benchmarks Established: ***Afrikaans Home Language*** *(2022)*							
Scheduled *Funda Wande* data collection for programme evaluation was extended to increase sample size and grades for the purpose of benchmarking.	Afrikaans (HL)	2022	735	Grades 1, 4, and 7	100 schools	Quintiles 1-5	Western Cape

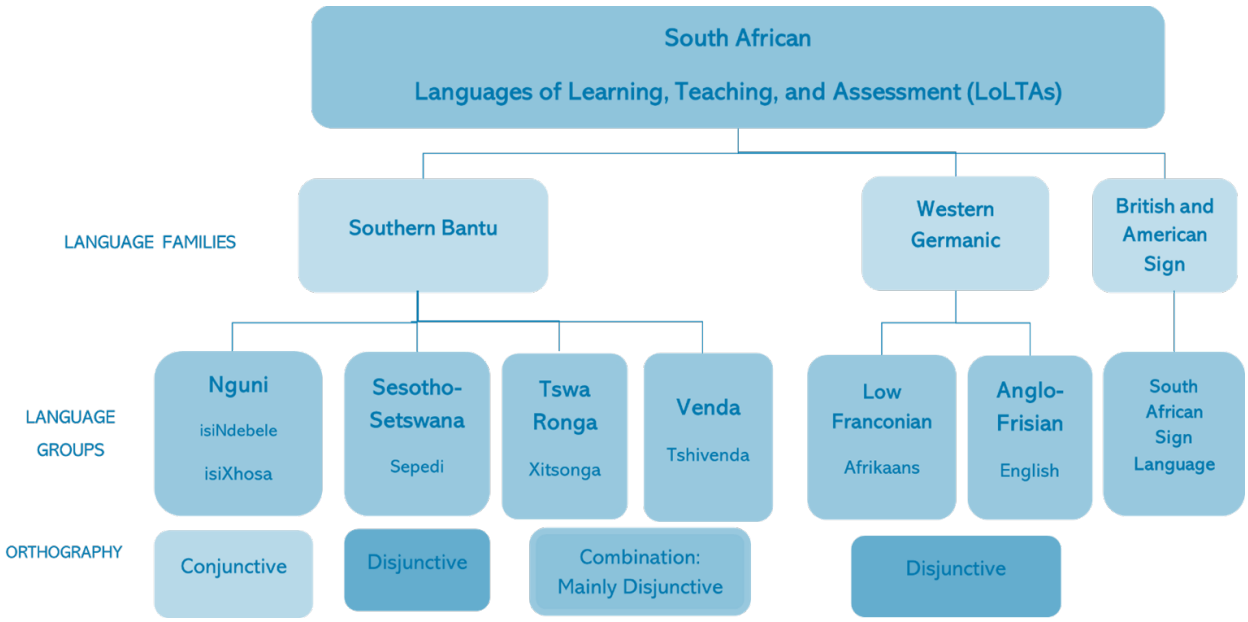
[2] The quintile system in South Africa is a system to categorise public schools for public funds disbursement based on socio-economic status of the communities they serve. Schools in quintiles 1-3 receive the more support per learner and are no-fee paying schools.

Benchmarks Established: ***English First Additional Language*** *(2022)*							
Collation of five datasets that included existing data and new data collected with the extended use for benchmarks including: Top-up to existing longitudinal on EGRS I and RSP data during the Sustainability Impact Evaluation data collection in 2021.	English (FAL)	2015-2021	23 000	Grades 2-7	900 schools	Quintiles 1-3	Eastern Cape Gauteng KwaZulu-Natal Limpopo Mpumalanga North West
Benchmarks Established: ***English Home Language*** *(Upcoming 2025)*							
Available EGRA data from the Western Cape Provincial Education Department	English (HL)	2022-2023	500917	Grades 1-3	914 schools	Quintiles 1-5	Western Cape
Top up using the data collected for the Funda Uphumelele National Survey		2025	3040	Grades 1-4	76 schools		All 9 provinces
Primary Collection							
Benchmarks Established: ***Xitsonga Home Language*** *(2023)*							
Dedicated data collection primarily for benchmarking of Xitsonga.	Xitsonga (HL)	2022	7691	Grades 2-4, and 6	120 schools	Quintiles 1-3	Limpopo
Benchmarks Established: ***Tshivenḓa Home Language*** *(2023)*							
Dedicated data collection primarily for benchmarking of Tshivenḓa.	Tshivenḓa (HL)	2023	3157	Grades 3, 4, and 7	60 schools	Quintiles 1-3	Limpopo

3.3 Benchmarking for a Multilingual Context

South Africa is a multilingual context with 12 official languages, 11 of which are spoken. The languages are distinct in orthography – writing conventions including spelling and grammar; and morphology – how words relate to each other to make meaning. These differences are significant in determining the relationship between language components, linguistic strategies, and appropriate pacing when reading. The figure below provides groupings of the 12 official South African languages used for instruction.

Figure 4: South African Languages of Learning, Teaching and Assessment



The languages may be described as follows: Nguni languages have a conjunctive orthography - one word may be a whole sentence; Sesotho-Setswana languages have a disjunctive orthography - short word segments with prefixes and suffixes written separately and Tshivenda and Xitsonga - mostly disjunctive orthography but with compounding features. While English and Afrikaans are mostly disjunctive. Each of these languages have other distinct features that affect reading such as double vowels for Afrikaans, complex consonants for Nguni languages and diacritics for Tshivenda.

Based on this, a decision was taken to benchmark each language separately and then compare these benchmarks within each African language family. If consistent patterns emerged, then a language family benchmark could be adopted. This was done and based on the similarity, a total of 6 categories were benchmarked out of 11 spoken official languages.

The benchmarking design work started in 2019 with the first set of benchmarks finalised in 2020. To date, early-grade reading benchmarks have been developed for the Nguni language group (isiZulu, isiNdebele, Siswati and isiXhosa); the Sesotho-Setswana language group (Sepedi, Setswana and Sesotho); Afrikaans, Xitsonga, and Tshivenda. All of these were developed at the Home Language level with English as the outstanding language. Plans to establish the English Home Language benchmarks are underway for finalisation in 2025. However, English benchmarks were developed at the First Additional Language level as the only language at the additional language level.

The benchmarks established are expressed as the minimum number of words read correctly per minute (CWPM) in a grade-appropriate narrative passage for grades 2, 3, and in some cases where data was available, grades 4 and 6. The minimum benchmark set for the end of Grade 1 is the number of single letters that should be corrected sounded in one minute (CLSPM). The benchmarks were set with the correlation to comprehension attainment factored in. They are therefore the minimum thresholds learners need to reach in grades 1, 2 and 3 to be on track to read with adequate comprehension by the age of ten. The table below provides a summary of the benchmarks.

Table 2: Summary of the South Arican Early Grade Reading Benchmarks (as of April 2025)

Language/Language Group	By the end of this grade, all learners should be able to read* at least. * Sound-out letter sounds for Grade 1				
	Grade 1	Grade 2	Grade 3	Grade 4	Grade 6
Nguni Language Group (Home Languages) (IsiNdebele, IsiZulu, isiXhosa, and Siswati)	**40** CLSPM	**20** CWPM	**35** CWPM		
Sesotho-Setswana Group (Home Languages) (Sepedi, Sesotho, and Setswana)	**40** CLSPM	**40** CWPM	**60** CWPM		
Afrikaans Home Language	**40** CLSPM	**50** CWPM	**80** CWPM		
Tshivenḓa Home Language	**40** CLSPM	**35** CWPM	**55** CWPM		**90 CWPM**
Xitsonga Home Language	**40** CLSPM	**30** CWPM	**40** CWPM	**50** CWPM	**85 CWPM**
English First Additional Language	**Not Applicable**	**30** CWPM	**50** CWPM	**70** CWPM	**90 CWPM**

Source: Department of Basic Education, Summary of benchmarks, 2024

3.4 Capacity Requirements for Establishing Benchmarks

A range of expertise was identified to develop the benchmarks successfully. The figure below illustrates these, and their skills are briefly discussed next.

Early grade reading experts are subject matter experts. They can evaluate the rigor of the research and assure the quality of the assessment tools. For the benchmarking approaches set out in this report, these experts must agree with scientific evidence about the importance of automaticity.

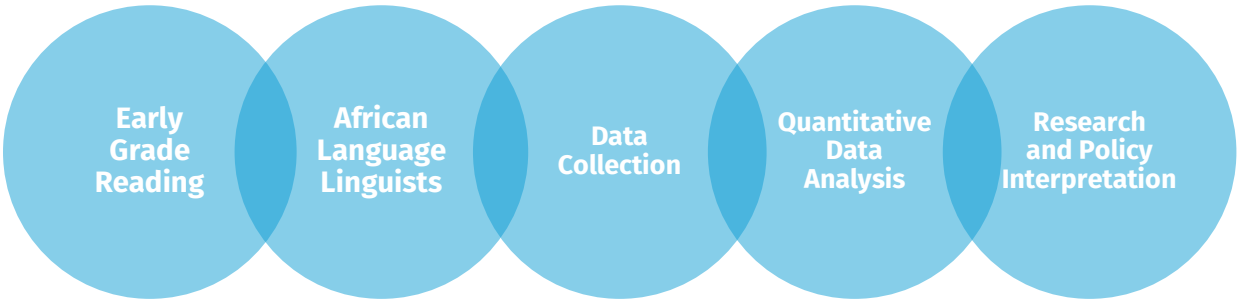
African language linguists understand the orthography and morphology of African languages, the standardised version and variations as well as how to ensure constructs are well communicated. They often acted as writers, originating stories and versioning across languages. We found that experience as both linguists and teacher trainers for languages was the ideal experience.

Data collection management and the ability to recruit the appropriate number and language-specific fieldworkers were core, fieldworkers with home language level proficiency across the languages were required. In addition, the fieldworkers needed to be computer literate with the ability to be trained in new software and the stamina to collect data from multiple schools across selected provinces.

Data analysis, writing and interpretation skills were needed to make sense of the data and to set the benchmarks. This required collaboration with all other experts in the process, to ensure that the data interpretation aligns with reading theory, linguistic rules and test development decisions.

A decision was taken early on to work with different teams from across the country for the respective languages. By the conclusion of the work, experts from more than 8 universities were collaborators, particularly as linguists. Core to the process was recognising the need to guide the various experts in terms of expectations, development and outputs based on their prior experience with this work. This awareness required planning and additional time protected for the iterative learning processes.

Figure 5: Different skills needed for reading benchmark development



3.5 Funding for Benchmarking

A defining feature of the benchmarking process in South Africa is the collaboration between the national Department of Basic Education, Provincial Education Departments, donors, linguists, data collection service providers and analysts. A deliberate effort was taken to pool

funds or to allocate areas of funding efficiently, avoiding duplication and the development of "donor A" versus "donor B" benchmarks, or "government" versus "NGO" benchmarks.

The costs and effort differed by the approach used. For Nguni and English First Additional languages, benchmark-specific costs were limited to the analysis and report writing costs as the data collection had already taken place; while the costs for the top-up approach and a dedicated data collection differed. The second most expensive was the top-up data collection costs for the Sesotho-Setswana and Afrikaans benchmarks with cost drivers being the additional development and grade to be assessed. The most expensive was the dedicated data collection where instruments, data collection and analysis were explicitly for the purpose of benchmarking, this was for Xitsonga and Tshivenḓa.

Donors contributed to the expert hires, fieldwork data collection and external analysis. The DBE contributed conceptual guidance, instrument development leadership and physical development, fieldworker training, data analysis and report writing. This was within the standard DBE funding mechanisms and entailed prioritisation of this work for the officials involved. The table below summarises partners involved.

Table 3: Benchmarking Data Approaches Taken for Reports and Supporting Partners

Benchmarking Data Sourcing Approach	Benchmark Report Completion	Data Sources, Project Partners, and Funders
1. Nguni Language Group report		
Re-analysis of Existing Data	**October 2020**	**Benchmark Reports:** Technical Report- Benchmarking Early Grade Reading Skills in Nguni Languages; and Summary Report- Benchmarking in the Nguni Languages Department of Basic Education; NORC and Harris School of Public Policy (University of Chicago); ReSEP (Stellenbosch University), SALDRU (University of Cape Town), University of South Africa (UNISA).
		Data Sources and Funders: *Early Grade Reading Study II:* Department of Basic Education; Mpumalanga Provincial Education Department; Class Act; Khulisa Management Services; University of the Witwatersrand funded by USAID. *Funda Wande Evaluation:* Eastern Cape Provincial Education Department; Funda Wande (Eastern Cape); and SALDRU (University of Cape Town) funded

		by the Allan Gray Orbis Foundation Endowment. *Leadership for Literacy:* Department of Basic Education; ReSEP (Stellenbosch University), funded by the Economic and Social Research Council (ESRC). *Story Powered Schools Impact Evaluation*: NORC (University of Chicago); Dr. Alicia Menendez and Dr. Cally Ardington funded by USAID. *Zenex Foundation Literacy Project (ZenLit):* KwaZulu-Natal, Eastern Cape and Western Cape Provincial Education Departments; funded by the Zenex Foundation.
2. English (First Additional Language) report		
Collection of Top-Up Data	**April 2022**	**Benchmark Reports:** Technical Report- English as a First Additional Language Benchmarking; and Summary Report- English as a First Additional Language Benchmarking Department of Basic Education (DBE); North West Provincial Education Departments; ReSEP (Stellenbosch University), SALDRU (University of Cape Town), University of Pretoria (UP); University of the North West (NWU); University of South Africa (UNISA); iKapa Data; Khulisa Management Services funded by USAID
		Data Sources and Funders: *Early Grade Reading Study I* Department of Basic Education (DBE), Department of Planning, Monitoring and Evaluation, (DPME); Gauteng and North West Provincial Education Departments; Class Act; Human Sciences Research Council (HSRC); International Initiative for Impact Evaluation (3ie); Khulisa Management Services; University of the Witwatersrand funded by Anglo American Chairman's Fund, USAID; UNICEF, and Zenex Foundation. *Early Grade Reading Study II:* Department of Basic Education; Mpumalanga Provincial Education Department; Class Act; Khulisa Management Services; University of the Witwatersrand funded by USAID. *Early Grade Reading Programme Evaluation:* Department of Basic Education (DBE); North West Provincial Education Department; Molteno Institute for Language and Literacy, Social Surveys Africa funded by Hempel Foundation, UNICEF and Zenex Foundation. *Funda Wande Evaluation:* Funda Wande (Limpopo); Limpopo Provincial Education Department; iKapa Data; and SALDRU (University of Cape Town) funded by the Allan Gray Orbis Foundation Endowment. *Leadership for Literacy:* Department of Basic Education; ReSEP (Stellenbosch University), funded by the Economic and Social Research Council (ESRC). *Story Powered Schools Impact Evaluation*: NORC (University of Chicago); Dr. Alicia Menendez and Dr. Cally Ardington funded by USAID.
3. Sesotho-Setswana Language Group report		
Collection of Top-Up Data	**June 2022**	**Benchmark Reports:** Sesotho-Setswana Language Group Benchmarks Report; Technical Report-Sepedi Early Grade Reading Benchmarks; Technical Report-Setswana Early Grade Reading Benchmarks; and Summary Report-Setswana Early Grade

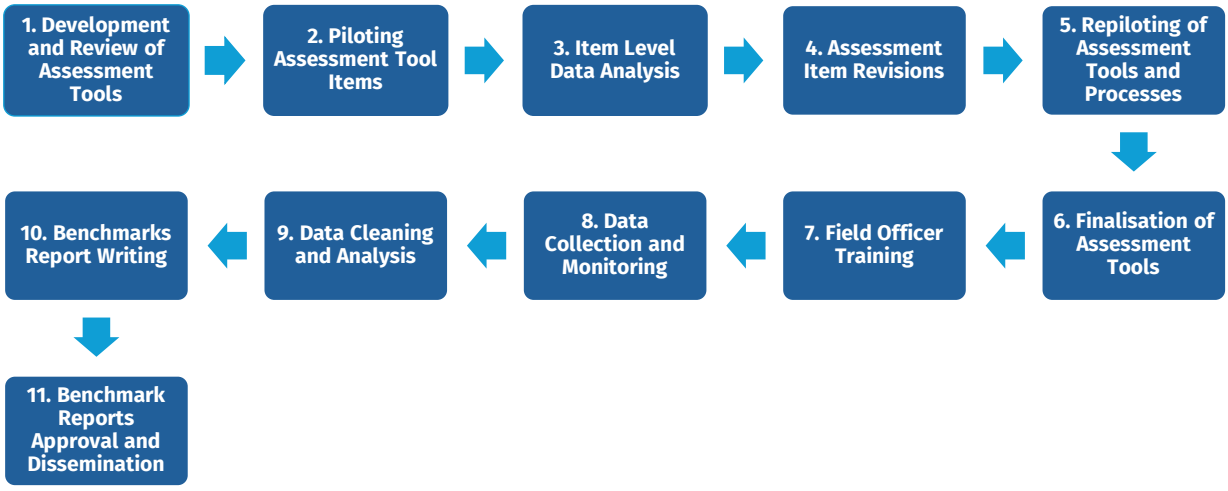
		Reading Benchmarks; Department of Basic Education (DBE); North West and Limpopo Provincial Education Departments; Funda Wande; Room to Read; ReSEP (Stellenbosch University), SALDRU (University of Cape Town), University of the Free State (UFS); University of Limpopo; University of Pretoria (UP); University of the North West (NWU); University of South Africa (UNISA); iKapa Data; Khulisa Management Services funded by USAID and Zenex Foundation.
		Data Sources and Funders: *Early Grade Reading Study I* Department of Basic Education (DBE), Department of Planning, Monitoring and Evaluation, (DPME); Gauteng and North West Provincial Education Departments; Class Act; Human Sciences Research Council (HSRC); International Initiative for Impact Evaluation (3ie); Khulisa Management Services; University of the Witwatersrand funded by Anglo American Chairman's Fund, USAID; UNICEF, and Zenex Foundation. *Funda Wande Evaluation:* Funda Wande (Limpopo); Limpopo Provincial Education Department; iKapa Data; and SALDRU (University of Cape Town) funded by the Allan Gray Orbis Foundation Endowment. *Room to Read EGRA Data:* Room to Read (Limpopo); Limpopo Provincial Education Department. *Save the Children Literacy Boost Study*: Save the Children; and Free State Provincial Education Department
4. Afrikaans Language report		
Collection of Top-Up Data	**October 2022**	**Benchmark Report:** Afrikaans Early Grade Reading Benchmarks Report Department of Basic Education (DBE); Western Cape Provincial Education Department, SALDRU (University of Cape Town), Stellenbosch University (SU); and iKapa Data.funded by Funda Wande.
5. Xitsonga Language report		
Dedicated data collection	**October 2023**	**Benchmark Reports:** Xitsonga EGR Benchmarks Technical Report Department of Basic Education;(DBE); Limpopo Provincial Education Department; Data Innovators (DI), (University of the Free State (UFS); University of KwaZulu-Natal; University of Limpopo; Decipher Data; and Word Bank funded by Old Mutual Foundation.
6. Tshivenda Language report		
Dedicated data collection	**November 2023**	**Benchmark Reports:** Tshivenda Benchmark Technical Report Department of Basic Education (DBE); Limpopo Provincial Education Department, University of South Africa (UNISA), Benita Williams Evaluation; Firdale Consulting; and Khulisa Management Services funded by USAID; and Zenex Foundation.

4. Activities Process for Establishing Reading Benchmarks

The next step in establishing the benchmarks for the different languages was conducting a scoping exercise[3] to see what reading data was available in the country for each the languages and grades. These were data from studies, programme evaluations, and other purposes.

There are different ways that this skill can be assessed, but the most replicable and comparable method employed by the various data proprietors with the most extensive data was measurement using EGRA-type assessments. The Nguni Language Group benchmarks were established exclusively from existing available data. However, instances where there was either no data or insufficient data of this kind available, necessitated having to collect similar data to be analysed using the agreed upon methodology. The process applied for the iterations where top-up or primary data collection was needed is charted below.

Figure 6: Activities Process for Establishing Benchmarks



[3] Mapping of South African Language Reading Data Available (2020)

4.1 Development and Review of Assessment Tools

Although the fluency benchmarks agreed on were limited to letter sound recognition and oral reading of a passage, it was important to survey other related skills. The rationale for inclusion of orally administered comprehension questions based on the passage used to determine fluency has already been covered in the previous section. The other skills were assessed to analyse the relationship between these skills and fluency in the specific language. This information was used to contextualise the fluency thresholds set. Depending on the opportunities and constraints present in the different rounds of data collection, a selection of the tasks in the table below were included. All cycles, however, included the additional tasks of rapid object naming, isolated word reading and written comprehension.

Table 4: Comprehensive List of Tasks Administered in Various Benchmarks Data Collections

Oral Assessments				
Assessment Task	**Administration Tool**	**No. Task Components**	**Task Duration**	**Grades Assessed**
Expressive Vocabulary	Oral Prompt	10 Items	UNTIMED	1
Rapid Object Naming	Learner Chart	36 Objects	20 SEC	1-3
Listening Comprehension	Oral Prompt	7-12 Questions	UNTIMED	1-2
Phonemic Awareness	Oral Prompt	5-7 Prompts	UNTIMED	1-2
Syllabic Awareness	Oral Prompt	5-7 Prompts	UNTIMED	1-2
Letter Sound Knowledge Task	Learner Chart	60 Single Letter Chart	60 SEC	1-3
Complex Consonants Task	Learner Chart	20-30 Item Chart	60 SEC	1-4
Syllable Reading	Learner Chart	20-30 Item Chart	60 SEC	1-4
Isolated Word Reading Task	Learner Chart	60 Word Chart	60 SEC	1-4
Oral Reading Fluency (ORF)	Learner Chart	No. of words passage determined by language, grade and pilot performance	60 + 120 SECONDS	1-7
ORF Comprehension	Oral Prompts with Reference to Chart	7-12 Questions	UNTIMED	1-7
Written Assessments				
Assessment Task	**Administration Tool**	**No. Task Items**	**Task Duration**	**Grades Assessed**
Written Response Comprehension	Written Passage and Questions	7-18 Questions	20 -40 MINUTES	3-7
Vocabulary	Written Prompts	10 Items	10 MIN	4-7

Content Sources for Assessment Task

Assessment tasks generally fell into three categories. The first group is tasks sourced from previous existing datasets having gone through the testing and verification protocols determined by the relevant researchers. If the original task was in the language relevant to the language being surveyed, it was included in the assessment pack for piloting as is or with minor editing. If the original task, however, was sourced from a language different from the required language, the first step was to version or translate the task to the needed language. The second group were tasks that existed in draft or untested versions. For these, the two linguists and foundation phase specialists appointed per language were responsible for reviewing, revising and/or the re-developing the task to fit the purpose of benchmarking to the level that they were 'pilot ready'. The last category were tasks that needed to be newly originated.

In all three instances, the two linguists contracted were given guidance and scaffolding by the RCME officials responsible for the development of benchmarks. This began with an instrument development workshop where example resources from previous studies, task specifications, and expectations for development were shared and implemented. These tasks were submitted for review and finalised for piloting. The linguists and Foundation Phase specialists appointed were individuals who are accomplished in their respective fields, however, as alluded to, the issue of the dearth in research in African languages in general and their instruction and assessment of reading skills in particular warranted this level of involvement of the benchmarks research lead and guidance in the development of the instruments.

One of the benefits of staggering the benchmarking of each of the languages was that the subsequent languages benchmarked benefited from the lessons learned and resources developed in the preceding rounds. This was especially relevant in the development of Oral Reading Fluency passages. For these, there was effort to retire passages that had been in circulation for years and to version newly developed passages into new languages. Collaboration with Zenex Foundation's *Ulwazi Lwethu* project that had commissioned a coalition of organisations working in reading and early literacy to develop open-source readers. As these were yet to be published, the benchmarks development team was given

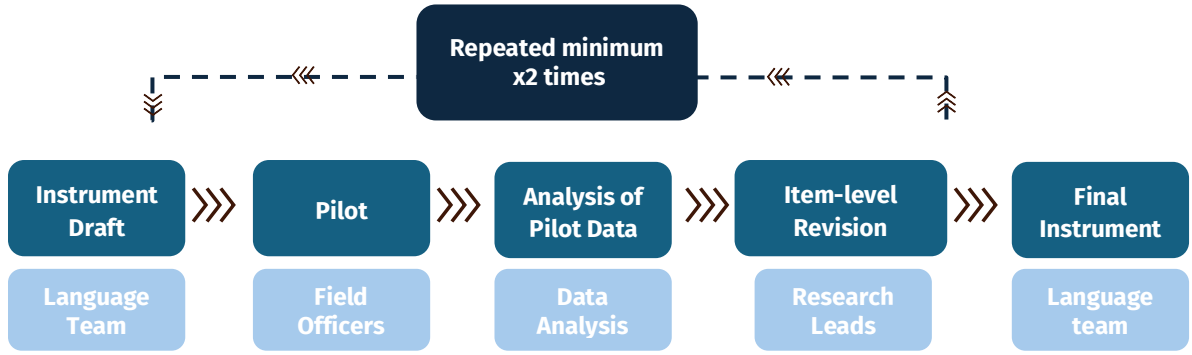
access to the passages originated in the different languages to edit and modify for the purpose of creating ORF tasks.

Another task worth mentioning is the written comprehension task. For this, the released PIRLS passages, accessed through application to the Centre for Evaluation and Assessment (CEA) that runs PIRLS in South Africa. While PIRLS has a rigorous validation and quality assurance process in place, the benchmarking language specialists were required to do further language and context review for the passage and questions and document any amendments they made to the original.

4.2 Piloting of Assessment Tool Items

Ahead of any data collection, the instruments were piloted in schools with learners in order to ensure that the final resulting final versions were appropriate for purpose. In each instance, piloting of the assessment items happened two times. This process involved various collaborators who led the different aspects of the process. The figure below depicts the leads for the different phase of piloting; however, it is worth mentioning that each component had the support of other collaborators and did not happen in silos. For instance, after concluding the pilot data analysis, the data team would present the findings in the workshop with the language team that comprised of the linguists and the benchmarking research team from RCME. These workshops would conclude with the decisions on which aspects of the tests needed to be revised and what in particular needed revision. The linguists would then enact those decisions.

Figure 7: Piloting Process for Instrument Finalisation



The key aims of the piloting / revisions

1. Developing new items that are language-specific
2. Ensuring assessments are set at the right level. No floor or ceiling effects
3. Appropriate language for the context with questions and instructions that are unambiguous
4. Appropriate length of the assessment (and specific reading passages)
    - assess the suitability of the length of the passages of connected text used to measure oral reading fluency (ORF) assess the difficulty level of comprehension questions
5. Appropriate duration of the complete assessment per grade

Capacity Gaps

Apart from the good practice of collaboration when doing work that requires and affects multiple stakeholders, the collaborative efforts were also a necessity. One of the challenges in doing the benchmarking work continues to be the limited skills, capacity, knowledge, and experience available in South Africa. For example, data collection service providers would typically be responsible for the training of the field officers they contract, however, the technical knowledge of how to conduct the assessments has had to be led by the benchmarking research team. Similarly, direction of what revisions needed to be made on the different items tested was led by the research team. One of the linguists interviewed reflected that the one skill she would have benefited from having ahead of joining her benchmarking language team was quantitative analysis- or at the very least, an understanding of the results. Her expertise, like many language practitioners, lie in the qualitative field.

Being aware of the various skills gaps for research in African languages in the country, one of the strategies we invested was contracting a junior language practitioner to be part of each of the language team. This was someone who was currently pursuing a post-graduate degree in African languages or linguistics. When proposing this role to the different funders, it was made explicit that the role was a capacity development role and therefore the goal would be for the incumbent to learn from the senior linguists and expose them to the kind of work that is being conducted in the African language education space. One of the long-term goals beyond their involvement in the establishment of the benchmarks was to pique their interest enough for them to direct their research or future plans in the field.

4.3 Field Officer Training and Data Collection

The training of the field officers for each round was a collaboration between the service provider appointed for the data collection, the benchmarking research team, and the language specialists who developed the instruments. The main data collection training was between four and six days- largely dependent on the number of tasks for data collection. There were three components to the training. First, there was an orientation to the project- the purpose, the survey tools, and the logistics. This included the sampling protocol to be administered at each school. While the number of learners differed across the rounds of data collection per language, the sampling protocol remained consistent to the extent that learners were selected from a class list of learners present on the day across the grade using the same formula.

The second component was the training on the administration of the actual tools. This consumed most of the training time as it involved the instruction on how the data collection software works; the tasks and how to administer them. Lastly, practice time for administration in smaller groups and a day out in schools for simulation of an actual field day were done. More on the administration process is covered in the next section on quality assurance procedures taken.

Quality Assurance Procedures

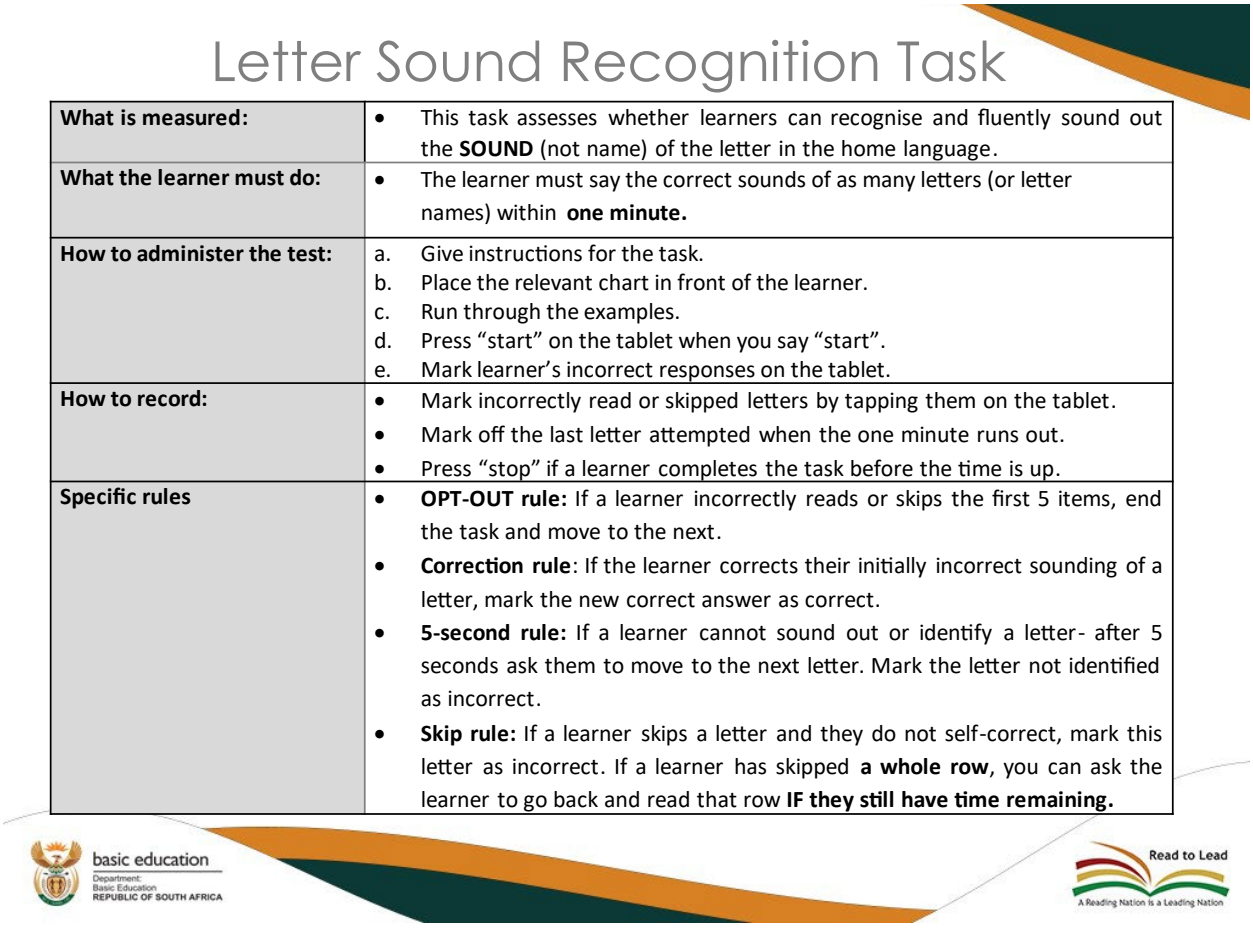
One of the highest priorities when conducting large scale data collections is ensuring that the data collected is of good quality. There were various measures taken at the different stages of both in preparation and during data collection to ensure this.

Ahead of fieldwork, the field officers were provided with a script to follow for each of the tasks. They were trained to give the instructions exactly as they appear and not to correct learners outside of the examples portion of administration. Over and above that, there were a set of protocols called the Six Golden Rules applicable in different combinations in the administration of the tasks. During the training, the field officers applied the rules during the practice sessions. This was to ensure that learners are assessed fairly, consistently, and that their actual competencies are captured regardless of who administers the task with them.

Figure 8 below is a training presentation slide showing the golden rules followed by another example slide summarising the protocol of a task (Figure 9).

Figure 8: The Six Golden Rules for Task Administration Protocol

Figure 9: Protocol for Administering Letter Sound Recognition Task



Letter Sound Recognition Task

What is measured:	• This task assesses whether learners can recognise and fluently sound out the **SOUND** (not name) of the letter in the home language.
What the learner must do:	• The learner must say the correct sounds of as many letters (or letter names) within **one minute.**
How to administer the test:	a. Give instructions for the task. b. Place the relevant chart in front of the learner. c. Run through the examples. d. Press "start" on the tablet when you say "start". e. Mark learner's incorrect responses on the tablet.
How to record:	• Mark incorrectly read or skipped letters by tapping them on the tablet. • Mark off the last letter attempted when the one minute runs out. • Press "stop" if a learner completes the task before the time is up.
Specific rules	• **OPT-OUT rule:** If a learner incorrectly reads or skips the first 5 items, end the task and move to the next. • **Correction rule**: If the learner corrects their initially incorrect sounding of a letter, mark the new correct answer as correct. • **5-second rule:** If a learner cannot sound out or identify a letter- after 5 seconds ask them to move to the next letter. Mark the letter not identified as incorrect. • **Skip rule:** If a learner skips a letter and they do not self-correct, mark this letter as incorrect. If a learner has skipped **a whole row**, you can ask the learner to go back and read that row **IF they still have time remaining.**

To ensure that the enumerators in training were retaining the different components, the practice sessions were always followed by Iter-Rater Reliability tests. These were tests where all the enumerators listened to a recording of a learner or trainer doing the task and they had to capture the responses on the tablet and submit their individual responses. Their responses would then be analysed against a predetermined memo, checking for their accuracy in capturing of the correct and incorrect items and start and stop timers. These results would then be presented at a subsequent training session. The information would be used for the trainees to know what aspects they needed to practice and for components where the results showed a high proportion of trainees making the same mistakes, the trainers would use this

as a signal of what needed retraining. A panel of results was kept for each trainee and used at the end of the training to determine who would be appointed for fieldwork. The training team also made site visits in the first week of data collection to make sure that due process was being followed and correct any deviations observed in the field. A monitoring tool was used and shared at the daily check-out meetings held in the first week of data collection.

Additional capacity in the form of Field Coordinators and Team Supervisors were added by the data collection service providers during fieldwork. Their roles included daily check-ins to execute various quality assurance activities. This was also important as mop-ups were arranged with schools in cases where the sample quality was not reached. For the written assessment marking and capturing, the supervisors and coordinators verified and moderated the marks captured by the fieldworkers to ensure accuracy after field teams had group marking sessions at the end of each school day. Supervisors would then address any issues and variation that arose during the marking session. The linguists would be the last step of quality assurance of the written assessment, they moderated a sample of marked written assessments to check for accuracy and consistency.

A live data tracking dashboard evolution took place with each new iteration of benchmark data collection. In the initial rounds, data from day 1 and 2 were analysed by the service providers to identify themes of errors observed during the first 2 days of data collection. This approach was used to enable the benchmarks research team to identify any issues, inconsistency in understanding as well as inaccurate data capturing. The dashboard in its current form provides the live submissions of individual field officers and teams; tracking the location of where they are collecting data, their duration per task for productivity's sake, and the quality of the data they are collecting. This live data dashboard allowed for on the spot intervention and correction. It has become a helpful tool even beyond data collection but also giving insight during data cleaning and analysis.

5. Engaging and Disseminating

The research team made upfront decisions on engaging different stakeholders at various stages of the benchmark development and use. The decisions on which stakeholder to engage was iterative and reimagined over time. This section summarises some of the core decisions and practical steps taken to achieve this.

5.1 Collective Ownership

The mandate of the Department of Basic Education is to develop national policy and guidance for the sector. Collective ownership would best attain this, enabling the benchmarks to serve as a literacy bridge rather than a way to further fragment the sector. To facilitate this, internal consultation processes were designed to ensure consensus *within* the DBE before engaging externally. Practically, this meant co-hosting and jointly reviewing the design process and report with officials responsible for curriculum. Agreements were also made and documented on four aspects 1) developing language-specific benchmarks, 2) skills to be benchmarked, 3) the use of an ORF question taxonomy to align with curriculum expectations and areas of curriculum strengthening, and 4) language sequencing – which language to start with.

As mentioned earlier, the recruitment of linguists and analysis capacity for the benchmarks deliberately targeted different universities, particularly those teaching the range of languages. This created a national university-wide collective ownership of the work.

Once the benchmarking process started a Reference Group consisting of provincial and district representatives responsible for language support in the selected province was convened. Practically, this meant that for a language such as Afrikaans, district representatives were invited from Northern Cape and Western Cape as the two languages with the highest number of speakers. This forum served as a way to engage the province and district, validate school languages as reported in administrative data, and support access to schools where challenges were experienced. By the end of the process these officials served as a way of disseminating the benchmarks to schools.

The message of collective ownership was also shared with donors, all donors interested in this work and literacy more broadly were invited to partner and co-fund; they were explicitly informed that the benchmarks would be DBE benchmarks and not referred to by donor. Practically this meant allocating the donors to different languages depending on the resources they have available, planning the timing for funding and jointly developing co-funding agreements and management systems.

Finally, all the benchmarks were formally approved by the Director-General with a foreword included in each report and formal letters written to each provincial head informing them of the reports. This process and endorsement are consistent with the processes followed for other national reports such as the Annual Performance Plan amongst others.

5.2 User Specific Dissemination

As mentioned earlier, data collected throughout the benchmarking process extended beyond the skills benchmarked. This enabled correlations across different skills such as phonemic awareness, syllable reading and letters although only letters were benchmarks. The use of the additional data helped validate the relationship with letters and the appropriateness of letter benchmarks. In addition, different users, such as academics, and curriculum officials requires analysis and discussion of these relationships for the development and support of the curriculum.

For brevity and simplicity, a decision to benchmark *one* skill per grade publicly and across language groups (where possible) was made to avoid confusion and manage communication. Further analysis moving beyond learner meeting and not meeting the benchmarks in a binary way was made, more helpful categories have been developed over time to measure those far from the benchmarks, those close to meeting the benchmarks, and those meeting them. This proved more helpful for provincial and teacher communication and developmental responses.

5.3 Approval and Public Launch of Benchmark Reports

Formal Director-General approval of the reports was followed by public release launches for the reports. The reports approved after the Nguni Language Group report were launched in hybrid livestreamed events. The recordings can be accessed on the department's YouTube channel hyperlinked below.

1. Early Grade Reading Benchmarks Launch for Sesotho-Setswana Language Group, English First Additional Language; and Afrikaans: 09 November 2022
2. Early Grade Reading Benchmarks Launch for Tshivenḓa and Xitsonga: 16 November 2023

6. Usage of Benchmarks

Deciding *how* the early reading skills benchmarks should be measured was crucial as the result would affect the communication to and the usability by the sector. By design, the benchmarks were to serve in the three categories of direct actors in the education sector: national and provincial administrations, districts and schools, and teachers in the classroom as summarised in Table 7. However, other education stakeholders such as donors; Non-Governmental Organisations (NGOs) working in early literacy, researchers; Initial Teacher Education training institutions; and the international community can also find use for this tool. The various effected and proposed implementation of use since the development of benchmarks as discussed below.

Table 5: Usage of South African Early Grade Reading Benchmarks

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

Source: Department of Basic Education, Benchmark design report 2020

6.1 Classroom

Early Grade Reading Assessment and Tool Update

In 2015, the national education department adopted and adapted USAID's Early Grade Reading Assessments (EGRA) diagnostic tool developed by Research Triangle Institute (RTI). To date the programme has been distributed to Grade 1-3 teachers in over 12000 schools across the country. Learners' attainment in the early skills of oral fluency in letter sound recognition,

word reading and passage reading with comprehension are measured akin to the benchmarked skills.

However, prior to the establishment and adoption of benchmarks, learners' skills were diagnosed, and proficiencies categorise using criteria established by a selection of curriculum specialists and subject advisors for each of the 11 Home Languages used for learning, teaching, and assessment including English as a First Additional Language (EFAL). The EGRA criteria has since been updated to align to the benchmarks ahead of the next phase rollout.

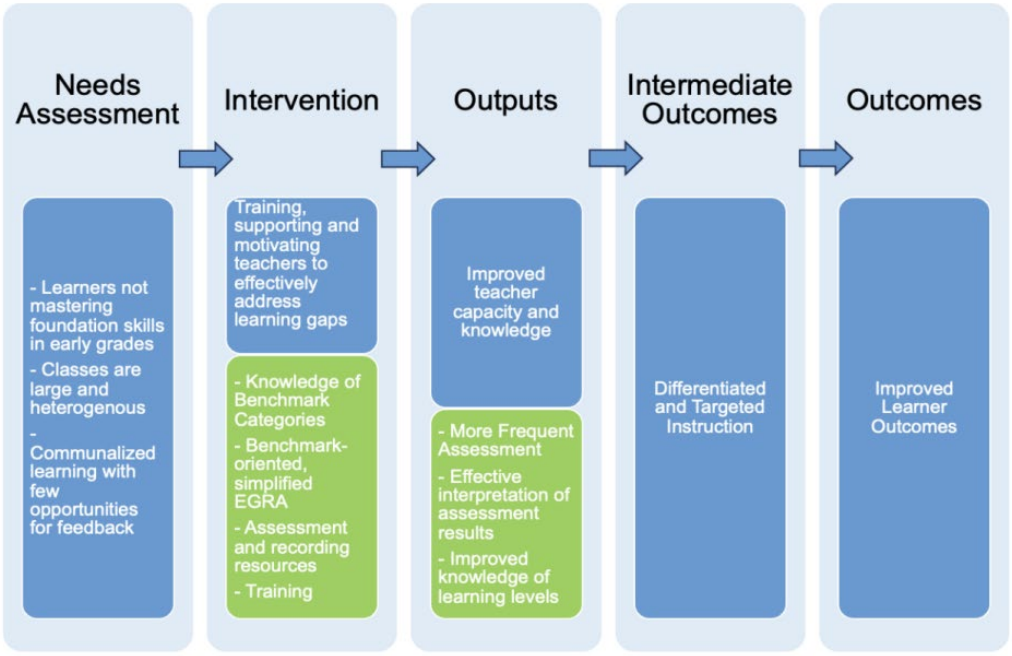
Teachers' Usage of Benchmarks in EGRAs: Pilot Study

In the monitoring of the implementation of EGRAs in schools, the curriculum directorate found that teachers experienced a few challenges in utilising the toolkit as intended:

a. Teachers reported struggling to interpret their learners' results using the criteria for remediation purposes;
b. The training model utilised was a scaffolded 'train the trainer' model. There was thus an inconsistency in the training provided to teachers
c. Teachers expressed that in large class sizes- as many as 50 learners, it was a time-consuming exercise to conduct the assessments with all learners to ascertain their individual reading proficiencies.
d. The expected frequency of assessments (thrice a year) further compounded on the time constraints;
e. Teachers were additionally found to be misappropriating the toolkit resources as learning and teaching materials and not a diagnostic tool. In a considerable number of instances, learners were found to have memorised the assessment tasks; and
f. There was a shortage of resources in that the same toolkit that teachers (mis)used in their classrooms were the same as those used for monitoring by subject advisors.

In 2023 the RCME directorate along with collaborators at the University of Cape Town conducted a pilot study that aimed to investigate how then newly established benchmarks could be used in average South African classrooms. The study's theory of change can be seen in the figure below and the report for a more comprehensive description of the study can be accessed here.

Figure 10: Theory of Change for Teacher Benchmarks Usage Pilot Study



Source: Teacher Benchmarks Usage Pilot Study Research Findings (Lobelo et al., 2023)

For the purposes of this report, below is the summary of the intervention design that pertain to the issues listed above as well as the findings and learnings.

Intervention 1: *Simplify administration of EGRA and align this to the benchmark development categories, Simplify the recording of learner proficiencies to reduce the cognitive burden and the time taken.*

Key Finding/ Lessons: Ninety-five (95%) percent of the teachers who were in the high-use treatment arm of the study found the simplified EGRA to be easier to administer than the original EGRA.

Intervention 2: *Conduct on-site training directly to teachers on newly established benchmarks and correct administration of EGRA*

Key Finding/ Lessons: Despite the top-up EGRA use training, the qualitative data showed that a lack of understanding of assessment practices for diagnostic purposes persisted. Instances of teachers using the resources for instructional purposes were observed.

In especially large classrooms, where these assessments are orally administered, learners waiting to be assessed quickly learnt to recite the contents while their classmates were being assessed, thus making the validity of the results unreliable. The request for additional test items was made.

Intervention 3: *Capacitate teachers on interpretation of benchmarks aligned EGRA results categories for proficiency diagnostic purposes.*

Key Finding/ Lessons: At the baseline teachers on average over-estimated the reading proficiency of 75% of their learners, at endline, teachers who effectively applied the intervention ranking of their learners' abilities had an average of 0,7 correlation to the respective learners' actual results. Teachers in the control group declined in their knowledge of learners' reading skills ability.

Intervention 4: *Direct teachers to resources on for learner remedial interventions based on skills gaps identified*

Key Finding/ Lessons: While teachers were referred to resources to for the remediation of skills gaps identified in learners, the intervention design assumed teachers could rely on their professional expertise and experience for intervention. Teachers at endline reported needing support in not only identification of skills gaps after they correctly categorise their learners, but additional capacity and resources for remediation strategies.

Table 8: Example of proficiency categories for Nguni Language ORF CWPM scores

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

In the interviews conducted for this report with one of the linguists also involved in initial teacher training this gap was confirmed. Her university conducted a survey with other teacher training institutions to see how they teach pre-service teachers reading pedagogy for African languages. Their findings were that different it is taken for granted that language teachers know how to teach early reading skills. Of the institutions surveyed, there was none that had a pre-service teacher reading programme for African languages.

6.2 Schools and Districts

Western Cape Province EGRA Data Dashboard

The Early Grade Reading Assessments continues to be a key strategy for identifying gaps in learners' reading skills in the early grades despite its limitations. Since its adoption and national roll-out a decade ago, the Western Cape Provincial Education WCED PED has been a forerunner in implementation, use and innovation.

According to the two PED officials interviewed for this case study, the driving reason for the institutionalisation was that in their commitment to being a data driven province, there was a gap in sources for reporting on early grade reading. Between 2017 and 2019, all 10000 foundation phase teachers were trained by the province on administering EGRAs and reporting. Districts are responsible for continuous top-up training for novice teachers entering the system

Modifications were made to the EGRAs for ease of use and to encourage uptake. These changes included reducing the number of tasks administered as well as only conducting the

assessment twice a year instead of three. Teachers were also allocated time in the timetable for the administration. The first assessment in January is conducted in the first two weeks scheduled for baseline assessments and the second assessment in October is recommended to happen in the Group Guided Reading (GGR).

The establishment and dissemination of the benchmarks was readily received by the provincial department. A circular to all districts announcing the adoption of these benchmarks, replacing the criteria previously used and their incorporation into EGRA reporting processes was communicated. Previously schools implemented EGRA for compliance reasons. However, with the language specific benchmarks, confidence in the relevance of this assessment to learner needs is seen.

As the province continually prioritise the upskilling of teachers, especially new teachers who are coming into the system, misconceptions about the benchmarks and EGRA are addressed in these top-up trainings to ensure effective implementation of the EGRA. In addition, the province developed an interactive poster for EGRA in every language. It includes links to tools, recording sheets, an EGRA administration video, and a link to the dashboard to facilitate easier access for teachers even after training. Another component in these training sessions is capacitating teachers with best practice and strategies for remediation.

Perhaps one of the innovative developments by the WCED has been the EGRA dashboard. It is a live platform that summarises the EGRA data that teachers report on twice a year. Data can be filtered and viewed by school, circuit, district, and province depending on access control. Subject advisors are expected to draw data from the dashboard before visiting schools to engage with teachers about the performance data, highlighting areas where learners have not met benchmarks. The teachers are encouraged to utilise the data to implement interventions for learners who do not meet benchmarks with support from subject advisors. The requirement to report on this data and the use of the dashboard data has increased accountability. Currently, 98% of schools in the province report on reading using EGRA data.

Concerns about the validity and accuracy of the learner data teachers report on were allayed bt the correlation analysis the province conducted of these EGRA results against the

provincially conducted Systemic Assessments as well as the PIRLS results. The self-reported classroom level EGRA results were found to have correlation with the provincial Systemic Assessments and the province's results on PIRLS.

The interviewees expressed that the benchmark areas of improvement the national department needed to address were firstly, the long-standing request to extend EGRAs to grade 4, as this stage is still crucial for literacy development. Secondly, support to the province in addressing the challenge in aligning the current EGRA tools with the synthetic phonics approach that the province has adopted. Thirdly, there is ongoing need to capacitate subject advisors on the effective use of the dashboard to interpret the data for use in assisting teachers implement effective strategies. Finally, there remains a need for advocacy and pace-setting tools to support teachers in exceeding the minimum thresholds that are the benchmarks.

6.3 National and Provincial Administrations

Setting and Strengthening Reading Norms and Standards

The representative from the Curriculum Implementation and Quality Improvement directorate interviewed for this report expressed that the next iteration of the benchmarking process should involve the extension of the fluency benchmarks for other constitutionally recognised languages such as Khoi, San, and Nama languages; and extension up to Grade 6 for all languages. Additionally, there is a request to utilise similar rigorous methodologies or extrapolate from the large-scale data now available to establish other reading proficiency norms including passage lengths for written comprehension texts.

Current ongoing benchmarks usage support to the curriculum section within the national department includes incorporating the developed benchmarks and phonics into the upcoming revised curriculum, ensuring that they are referenced in policy and practice in classrooms. It involves strengthening teaching plans, training manuals, and frameworks with an infusion of the benchmarks. This will help align the curriculum with current educational needs and trends. Plans are in place to field test the adopted benchmarks, singular phonics program for each

language, reading manual, and handwriting manual to assess its effectiveness before full implementation. The national curriculum official interviewed expressed that the progress should involve finalising the submission for field testing the developed documents, including benchmarks, and soliciting inputs from relevant stakeholders such as the Pan South African Language Board (PanSALB) to refine the literacy strategy.

Whole School Evaluation

The Whole-School Evaluation conducts school monitoring for the purpose of evaluating the overall effectiveness of a school as well as the quality of teaching and learning. This is done by evaluating overall performance in different areas to enable planning for appropriate programmes for improvement. One of the key focus areas assessed at schools is learner achievement, which includes aspects of reading proficiency. The WSE officials at the national and provincial are expected to assess this and provide schools with recommendations for improvement.

Standardised tools for measuring reading were a gap addressed by the research team through the benchmarking process. With the finalisation of the reading benchmarks, the WSE reading outcomes measurement tools will be updated for all languages.

National Foundation Phase Minimum Resource Package

In 2012 a national catalogue specifying approved classroom resources and textbooks was developed and instituted nationally. All provincial budget spending needed to be based on a selection from this catalogue. The main focus for resources started in Grade 4 until Grade 12. At the end of 2024 a call for publishers to submit a minimum package for the foundation phase across all 12 official languages was made. The reading benchmarks were specified for each language as guidance for reading books serving as an official way to guide material development. It is anticipated that the catalogue will be finalised in 2025 and procured by the end of 2025 for use in classrooms in 2026.

National Early Grade Reading Survey

The Funda Uphumelele National Survey referenced in table 2 is one of the key strategic uses for benchmarks at the national and provincial. The current design and aim of the survey is to measure learners' achievement in the foundational skills of reading. It will be the first survey of learners in grades 1, 2, 3, and 4 to track what percentages of children are achieving these benchmarks and what percentage of ten-year-old learners can read with meaning. This survey will be a supplementary data point that contextualises the measures reported in data sources like PIRLS and the Systemics by revealing which preceding skills intrinsic to achieving written comprehension competencies learners have mastered and are yet to master.

At the time of writing, the main data collection in a sample of 710 schools resulting in reading data of over 28000 unique learners across in all 9 provinces had just concluded. The data collected is nationally and linguistically representative. The first round of this survey (2025) will establish a national baseline set of statistics with the intention that government will institutionalise this survey and conduct it every three years to track progress in the country's reading outcomes. The results for this first cycle are expected to be launched by August 2025.

6.4 Diffused Uses

Research and Initial Teacher Education

A reflection from interviews with linguists collaborating on the development of benchmarks has been that the use of the benchmarks should be systematically and formally integrated into Initial Teacher Education Training. One of the linguists who worked on the benchmarking process has begun to incorporate the benchmarks in her training of pre-service teachers by integrating them into lesson plans, teacher training workshops, and classroom assessment frameworks. Another suggested that, a module be developed for reading pedagogy courses where assessment and remediation of reading in African languages is addressed. This module should include components for their micro-teaching. This will ensure that in addition to in-service teachers, being capacitated, the novice teachers are also getting the same orientation to the benchmarks and their use for improving reading.

The table below illustrates the use of benchmarks, these have not been specifically directed but have been taken up by different stakeholders. Two of the articles were written by a linguist and data analyst involved in the benchmarking, illustrating the benefits of collaborating across different institutions. The benchmarks were also used in Masters thesis by a student who was not part of the study, as well as for a South African journal article. Significantly, the benchmarks have also been integrated into the ongoing teacher guides commissioned by the Department of Higher Education for use in Initial Teacher training. So while the documents in the table below are a start, more could be done to further formalise this, and ensure take up and use in the content and training.

Table 6: Publications citing or using South African Established Benchmarks

Publication type and link	Date	Brief description
University Article		
Benchmarking early-grade reading literacy skills in Nguni languages Author: Cally Ardington* *Data analyst involved in benchmarking for multiple languages	26 November 2020	University of Cape Town Newsletter Article with links to the Nguni summary and technical reports on DBE website and some diagrams
Reading Benchmarks for African Languages **Author:** Connie Makgabo* *Linguist involved in Sepedi benchmarks	8 May 2023	University of Pretoria Newsletter on all benchmarks with a focus on Sesotho-Setswana languages
Academic Qualifications		
Learning profiles and reading benchmarks: Trends from South African data on early grade reading **Author:** Jessica Qvist	August 2023	University of the Witwatersrand Masters Dissertation on Tracking learners reaching benchmarks from Grade 1 to 4
Initial Teacher Education		
PrimTEd Teaching Reading study guides Study guide 4: Reading Fluency **Authors:** Various consultants for Department of Higher Education and Training	2021	Department of Higher Education and Training Teacher guideline on teaching fluency as part of a series to enhance Foundation Phase literacy nationally
Donor Publication		
African Languages Reading Benchmarks Project: Early Grade Literacy **Author:** Zenex Foundation	18 October 2021	Zenex Foundation news update on benchmarking purpose and African language distinctions

Journal Article		
No-fee school consistently outperforms Progress in International Reading and Literacy benchmarks: Presenting early grade reading data from a case in Makhanda, Eastern Cape **Authors:** Kelly A. Long, Tracy N. Bowles	19 May 2023	South African peer-reviewed journal article on using English second language benchmarks

7. Reflections: Lessons, Gaps, and Next Steps

In this section, we consolidate the emerging themes, lessons, gaps and next steps. These are not limited to South Africa and in some cases may be applied to other developing country context and beyond including lessons for the international community.

7.1 Theme 1: Importance of benchmarks for literacy strategies

Measuring reading can be abstract conceptually, particularly in the broader education sector, which could hinder advocacy for early literacy skills. Measuring reading fluency addresses this by making measuring more tangible. The effective incorporation of reading benchmarks in literacy strategies enables the measurement of the impact of interventions and teaching practices more broadly. Typically, the Monitoring and Evaluation approach for literacy strategies may not adequately measure the impact of the literacy strategy on reading outcomes, focusing instead on outputs and expenditures.

Recommendation: once benchmarks have been developed, careful systematic incorporation into the relevant official documents should be done. This ensures thorough integration and enables potential use across the sector.

In South Africa, the declaration of the benchmarking reports as official documents is the start of this process, as well as incorporating the benchmarks into the latest teacher guides and most recently, the textbook catalogue. The next steps are to incorporate the benchmarks into the Annual Teaching Plan, and the strengthened curriculum. Documenting this more carefully would be important to create a more comprehensive mapping and ensure that there are no omissions. There is a need for longer-term planning and mapping to determine how and where the benchmarks will be integrated into various programs. There is a need for clear communication about the benchmarks and their integration timelines into ITE and other programs.

7.2 Theme 2: Funding

Funding innovation and development such as the benchmarks is not easy. Philanthropic organisations involved may not have a robust understanding of the strategic importance of the benchmarks. This requires a strategic approach for each respective board. An example of this was the submission of a concept note without requesting any funding to build rapport and buy-in, then only submitting a funded proposal in subsequent meetings. Mapping the work done by the directorate on the Early Grade Reading Studies series that the organisation had previously funded, and its interlinkage to the benchmarks was another lever. Lastly, the existing partnerships on the benchmarks that the research directorate had with the known researchers, other education NGOs, and other donors reduced the risk for the organisation leading to higher comfort levels for contributing financially.

Developing the benchmarks for the different languages in phases meant there was a different cohort of donors and contributors for each iteration. While there was the real burden of altering strategies to suit the different organisations to secure their partnership, the payoff was the goal of having benchmarks adopted by the whole sector because of the widespread involvement.

Recommendation: Multi-funder partnerships should be pursued to maximise efficiencies, co-ownership and reduced fragmentation in the sector. Further, often no single funder has all the resources to complete the full national scope of the benchmarks. The collaboration results in both meaningful and targeted funding, and increased the possible reach across all selected languages and grades. Upfront discussion on direction setting and leadership by the state and not the donors is core as a way to manage relationships and the direction followed.

7.3 Theme 3: Diversity and equity

The principles of equity, redress, and building the capacity (especially for black South Africans) remain a core for the development of South Africa. Establishing ways to take this agenda forward is a priority that should be engaged from the inception. This emerged across the discussions and interviews. It is important to use different people for productivity and the

diversity of views and approaches. However, diversity should be meaningful and not for the sake of compliance.

In South Africa the challenge is that the lack of exposure and capacity may result in the same people contracted in research and evaluation. There is work needed in the sector to attain diversification without compromising the quality of work. In the benchmarking process scaffolding through the development of guidelines, additional time allocation and co-creating with all new resources added to the process worked effectively. Workshops to ensure a shared understanding were also scheduled across the different milestones and deliverables. Future iterations would require expansion of the pool of resources to younger emerging researchers, linguists and data analysts.

Recommendations: the decision to work with diverse teams needs to be taken upfront and aspects such as timelines need to incorporate the iterative nature of onboarding new people and enabling them to receive multiple rounds of feedback as well as constantly checking that there is a shared understanding of processes and expectations.

7.4 Theme 4: Linguist and Language Specialist Role

When reflecting on their role as Linguist and Language Specialists for the development of the benchmarks, the linguists expressed how the process was pivotal in giving them a practical understanding of the theoretical knowledge of early literacy acquisition they already had. This included gaining experience in the use of tools such as the Early Grade Reading Assessments (EGRAs) for assessing early grade reading.

The cores skills the linguist brought to the work were knowledge of the language, and the context of Initial Teacher Education. A common reflection was the gap in quantitative data analysis and interpretation skills to further understand the meaning of the data that was used to refine the instruments in the piloting stage. A suggestion has been increasing field observations to more than the current one day after training, these would assist in broader instrument feedback and administration feedback.

Recommendation: While the role of linguists is largely on instrument development, interpretation of data is core. Efforts to equip the linguists on interpretation could be enhanced. Initial efforts may entail using prior data from other studies to build fluency in data interpretation followed by dedicated sessions on interpreting throughout the piloting process, not just at the start as in South Africa.

7.5 Theme 5: Development of Instruments

There is scarcity of EGRA-type tools developed in African languages that are also culturally responsive. The linguists reflected that most texts have English as their point of departure, noting how cultural nuances can be lost in translation, which complicates the understanding of the content. Citing the 2021 PIRLS passage one of the linguist reflected on how a core vocabulary word in one of the released items, “octopus” was hardly known by them. This cultural deafness in versioning means a lot especially in texts where the purpose is to assess reading proficiency and not broaden a learners’ understanding/vocabulary.

She found the bank of oral reading fluency texts that were provided for further development in the Sepedi benchmarking project to be sufficiently culturally reflective. However, the need for more instruments/texts developed specifically for and in African languages is still needed. CM continued to express her concern about the limited data available for these languages. She emphasized that having more data could have significantly aided the development and evaluation process.

In South Africa, efforts were taken in the benchmarking to originate passages from the African languages and version these across languages. English was not taken as the starting point. Language aspects such as idioms and proverbs were considered especially when versioning. Passage analysis in the different languages often showed different difficulty levels based on the internal logic and structure of the various languages. Efforts to document these, as well as what aspects make a particular language difficult or easy were documented in a dedicated linguistic features aspect of the benchmarking reports. Ensuring that these lessons are incorporated into future instrument development documents would be a next step.

Recommendation: While standardisation is important in national assessments, allowing for language-specific or language-directed instrument development is important in ensuring fair use and a lack of bias. This may mean agreeing to different languages origination, that is not the dominant language such as English, French, or Portuguese across developing countries. Specific attention should be paid to language specific features, as well as idioms and proverbs.

7.6 Theme 6: International and local expressions

Challenges in standardizing assessments across countries have been acknowledged across multiple forums, and reports, alongside the importance of focusing on learning outcomes rather than strict comparisons.

There is a significant gap in the existing literacy assessments in South Africa, as current methods do not adequately measure foundational skills or comprehension levels, which could lead to ineffective evaluations of student literacy. The annual national assessments conducted from 2011 to 2014 in grades one through six and grade nine primarily utilized conventional written tests, which may not accurately reflect students' literacy skills in the early grades or in low literacy contexts. While the national assessments may highlight the importance of reading benchmarks, that could also lead to an overemphasis on metrics at the expense of understanding the broader context of children's learning. So while specific benchmarks should be considered, this should ultimately be framed in the comprehensive literacy skills we want learners to have.

South Africa participates in PIRLS, an international assessment, as well as SEAQMEC, a regional assessment. Then more recently, the reintroduction of the national systemic evaluations. Carefully communicating the complementariness and differences of these is important to ensure ongoing support and integration in interpreting the results and ensuring that both national and international objectives are addressed.

Recommendation: Establishing the clarity of purpose of each assessment and a common framework to categorise the skills, measures and advantages and disadvantages of each method may be a good way to systemically articulate a meaningful comparison across

assessments, purposes and audiences. This would be useful to evaluate if the overall assessment regime within a country addresses the different but complementary national and international objectives.

7.7 Theme 7: Key success factors

There are several factors necessary for success. There may be summarised as 1) the necessity of government involvement to avoid issues of recognition and implementation, indicating a strategic approach to policy influence, 2) government ownership of the benchmarks and process, 3) integration of the benchmarks into educational materials and teacher training programs, 4) a focus on long-term planning for implementation.

This list could be elaborated but this report overall is a way to document and interrogate this with core factors identified and replicated in other subjects in South Africa or in other contexts.

7.8 Theme 8: International considerations

In considering the replicability of the process undertaken by South Africa, any given country has its own context and the development of the educational assessment model like the South African benchmarks was achieved because of its existing opportunities and limitations. A critical catalyst for South Africa's benchmarking process was its failure to reach the PIRLS benchmarks that are of written comprehension, establishing benchmarks for precursor skills, like fluency, that would be applicable for all languages taught in the country was therefore identified as most useful.

Developing benchmarks for this skill for a country that is already exceeding the written comprehension benchmarks or setting fluency benchmarks for linguistic contexts where fluency even more difficult to measure because of the orthography, may be imprudent. So, while the establishment of comparable benchmarks is important, the focus should be on communicating the significance of learning outcomes rather than on strict comparisons between countries. Similarly, international donors should be mindful of the context they are partnering with and be open to being flexible on the types of evaluation tools and measures

they use especially when working in different country contexts- even when they are neighbouring countries.

8. Conclusion

This report has attempted to document a core development process in the benchmarking of early reading skills in South Africa across 11 official spoken languages. The report documents the rationale, methods, data sets and the resulting benchmarks. It also incorporated the practical, educational and methodological considerations. Funding stakeholder, engagement and usage were also documented, concluding with reflection and next steps for South Africa and the broader context.

9. References

- Ardington, C., Wills, G., Pretorius, E., Deghaye, N., Menendez, A., Mohohlwane, N., Mtsatse, N. and Van der Berg, S. 2020. Technical Report: Benchmarking early grade reading skills in Nguni languages. Stellenbosch: ReSEP, Stellenbosch University.
- Ardington, C., Wills, G., Pretorius, E., Deghaye, N., Mohohlwane, N., Menendez, A., Mtsatse, N. and van der Berg, S., 2020. Benchmarking early grade reading skills in Nguni languages. International Journal of Educational Development, 84, p.102433.
- Ardington, C., Mohohlwane, N., Sebaeng, L., Jodar, P., Beggs, C., Makgabo, C., Maledu, A., Zwane, Z. 2022. Sepedi Early Grade Reading Skills Benchmarks. 2022. Technical Report.
- Department of Basic Education (2024) *Review of Progress in the Basic Education Sector to 2024: Analysis of Key Statistics*. Department of Basic Education, p. 84.
- Department of Basic Education (2023) *PIRLS 2021: South African Preliminary Highlights Report*. Pretoria: Department of Basic Education.
- Department of Basic Education (DBE). (2011). National Curriculum Statement (NCS), Curriculum and Assessment Policy Statements (CAPS), Grade R- 3, English Home Language. Pretoria: DBE
- Gustafsson, M. (2020) *A Revised PIRLS 2011 to 2016 Trend for South Africa and the Importance of Analysing the Underlying Microdata*. Working Paper WP02/2020. Stellenbosch University: Research on Socio-Economic Policy (ReSEP), pp. 1–22.
- Jukes, M., C. Cummiskey, M. Gargano and P. Dubeck. 2018a. Guidance Note: Setting Data-Driven Oral Reading Fluency Benchmarks. RTI International, Research Triangle Office.
- Jukes, M., C. Cummiskey, M. Gargano and P. Dubeck. 2018b. "Data-Driven Methods for Setting Reading Proficiency Benchmarks." RTI International, Research Triangle Office. Accessed from https://www.roomtoread.org/media/984466/room-to-read_fluency-benchmarking-analysis_report_may-2018.pdf
- Khulisa Management Services. 2020. Setting Reading Benchmarks in South Africa. USAID Publication. Johannesburg: Department of Basic Education.
- Lobelo, A. *et al.* (2023) *Teacher Benchmarks Usage Pilot Summary Report*. Research Finding. Pretoria: Department of Basic Education, pp. 1–47.

- Mohohlwane, N., Wills, G. and Ardington, C. (2022) 'A review of recent efforts to benchmark early reading skills in south African languages', in Early Grade Reading in South Africa. Cape Town, South Africa: Oxford University Press, pp. 83–108.
- Mohohlwane, N.L., Mtsatse, N. and Courtney, P., 2023. Bridging the 20-Year Literacy Divide. Tracking Changes in South African Reading Literacy Achievement, p.151.
- Mullis, I.V.S. and Martin, M.O. (2019) *PIRLS 2021 Assessment Frameworks*. International Association for the Evaluation of Educational Achievement (IEA), pp. 1–88.
- Nuga Deliwe, C.O.N. (2017) An Analysis of the Measurement of the Progress in Learning Outcomes at Country Level: The Case of South Africa. University of Witwatersrand
- Sapire, I. and Roberts, G. (2017) The Status of the Language of Learning and Teaching (LoLT) in Schools: A Quantitative Overview: 2008-2016. Pretoria: Department of Basic Education. Available at: https://www.education.gov.za/Portals/0/Documents/Reports/The%20Status%20of%20LoLT%20In%20Schools%20final%20report.pdf?ver=2023-07-09-210745-167.
- Spaull, N. and Pretorius, E. (eds) (2022) *Early Grade Reading in South Africa*. Cape Town, South Africa: Oxford University Press Southern Africa (Pty) Limited.
- Statistics South Africa (2023) *Census 2022 Statistical Release*. P0301.4. Pretoria: Statistics South Africa, pp. 1–113.
- Wang, Z., Sabatini, J., & O'Reilly, T. (2020). When slower is faster: time spent decoding novel words predicts better decoding and faster growth. Scientific Studies of Reading, 24(5), 397-410.
- Wills, G., Ardington, C., Pretorius, E., Pooe, E., Ramagoshi, R. (2022) Benchmarking Early Grade Reading Skills: Setswana and English First additional Language. Technical Report. Khulisa Management Services. Johannesburg.

Annexure A: Interviewee List

Organisation Type	Role in Benchmarking Process
National Department of Basic Education: Curriculum Implementation and Quality Improvement (GET)	DBE official in the Curriculum Directorate responsible for curriculum development, monitoring and reporting for languages
National Department of Basic Education: Research Coordination, Monitoring, and Evaluation	DBE official in the Research and Evaluation Directorate. Part of the principal researchers in the conceptualisation and development of the South African early grade reading benchmarks.
Provincial Education Department: Curriculum Development: General Education and Training (GET)	Two officials responsible for curriculum implementation, support, training and monitoring in languages in the Foundation Phase.
University: Development Economics and Statistical Science	A researcher leading the approach to the data analysis methods and analysis for several reports and languages
University: Linguistics and Initial Teacher Education	Two lecturers in Language in Education and Initial Teacher Education (ITE) at a local university. They were contracted to be language specialist in the development, piloting, and training for data collection aspects for the benchmarking of their respective languages.
University: Applied Linguistics	A linguist and reading development specialised involved in conceptualizing reading benchmarks and providing technical support across various languages
Local Education Funder	A representative from a local education foundation and grant-maker seeking to support programmes that improve South Africa's language and mathematics education. The foundation supported the benchmarks through funding and through provision of materials used for instrument development.
Local NGO	Two representatives from two organisations working in literacy in no-fee public schools across several provinces over several years.
International NGO	A representative working in literacy internationally with a South Africa presence working in several provinces in literacy over several years

uis.unesco.org

© UNESCO-UIS 2025