



Africa Foundational Learning
Assessment Initiative

Learning Assessment
Benchmarks in Africa

Lessons from South Africa's journey

Dr. Nompumelelo Mohohlwane,
Director: Reading, Department of Basic Education

Prof. Cally Ardington,
AFLEARN, University of Cape Town



Preamble

Join the conversation on foundational learning assessment in Africa

In September 2025, education leaders from 13 African ministries gathered in Nairobi, Kenya, for a co-creation workshop on the Africa Foundational Learning Assessment Initiative (AFLAI). Together, participants explored how data and assessments can strengthen foundational learning across the continent.

The September convening led to the launch of the AFLAI Knowledge Exchange Series in December 2025 with our [first blog examining AI's role in education assessments](#). We are now diving deeper with a four-part series on benchmarks — and this is part one.

What's next? Watch for upcoming blogs and webinars exploring language in assessments. We'd love to hear your perspectives and experiences as we build this knowledge exchange together.

Around the world, children are spending more years in school than ever before. Yet for too many, this increase in schooling is not translating into meaningful learning. South Africa is no exception: repeated assessments have shown that large numbers of children complete the early grades unable to read with adequate comprehension in any language.

The recent launch of South Africa's Funda Uphumelele National Study (FUNS) marks a major milestone in responding to this challenge. FUNS is the culmination of several years of work by the Department of Basic Education (DBE) and its partners to establish empirically grounded early-grade reading benchmarks across all 11 official languages. For the first time, South Africa can report nationally representative data on early reading outcomes in Grades 1–3, benchmarked against standards that are both linguistically valid and instructionally meaningful.

In this blog, we share South Africa's journey in identifying and establishing benchmarks for the foundational reading skills that children need to acquire by the end of Grades 1, 2, and 3 to be on track for reading comprehension in Grade 4. We highlight why benchmarks were needed, how they were developed, and how they are now being used to strengthen early grade reading and foundational learning more broadly.

Why did South Africa need benchmarks for reading, anyway?

South Africa's Systemic Evaluation results and repeated findings from international assessments such as PIRLS show that only about 20 to 30 percent of children complete the Foundation Phase able to read with adequate comprehension in any language. Yet comprehension is the only visible outcome of the complex process of learning to read. Beneath it lies a set of foundational literacy skills that must be learned, automated, and integrated before meaningful understanding is possible.

Even where children's foundational skills such as decoding are assessed, data from assessments are only useful when we know what "good enough" looks like. Without clearly articulated benchmarks, standards of reading proficiency in the early grades remained vague. The curriculum specifies what should be taught but does not have a critical element: empirically-based targets on what constitutes proficiency in comprehension or its pre-requisite skills.

Given South Africa's linguistic diversity, proficiency needed to be defined for each language. Without benchmarks for proficiency that represented the variety of South Africa's classrooms and languages, teachers were left to guess when children had sufficient skills in reading. In early grades, where teachers lacked clear goalposts, they were unable to judge whether children were on track to be proficient readers. As a result, learning gaps compounded by Grade 3 – the point at which children are expected to shift from "learning to read" to "reading to learn".

South Africa's leaders realized that early grade reading benchmarks were essential to provide teachers (and the system) with the critical information they needed. Benchmarks provide clear, measurable targets for foundational skills in Grades 1 to 3 which can then be used to guide teaching and support from the outset. Over time, they would help reduce the significant gaps in foundational skills that had been compromising learning in later grades. The DBE plans to use the resulting benchmarks to mobilise change at three levels of the system – national and provincial, mid-tier and in classrooms.

National and Provincial Administration	Mid-tier support (Districts and School Management)	Classrooms
Defines reading proficiency (benchmarks)	Use standardised tools to assess reading across schools	Standard against which to measure reading skills
Communicates benchmarks and targets	Identify schools and classrooms needing support	Identify early on learners at risk of not learning to read
Monitor progress	Provide targeted support to schools and teachers	Adapt instruction to learner needs

Why set benchmarks by language?

Once the case for benchmarks was conceptually clear, operationalising them posed its own challenge in a linguistically diverse country as South Africa. While reading benchmarks for Grades 2 and 3 are well documented for English, there was very little research to guide the development of benchmarks for African languages.

Benchmarks cannot simply be transferred between languages. Benchmarks that apply to English cannot easily be applied to languages such as isiZulu or Sepedi because each language fundamentally differs in its phonological (speech sounds), morphological (how words relate to one another), and orthographic structures (written). Benchmarks are sensitive to all of these factors.

For example, Sesotho–Setswana languages are written disjunctively which means they have short word segments with prefixes and infixes written separately. By comparison, Nguni languages are conjunctive which means they can be written down in fewer but longer words. The following example, drawn from an early-grade reader, illustrates these differences across three languages.

- English: “There was a stranger who was very hungry.”
- isiZulu (Nguni language): “Kunesihambi esasilambile kakhulu.”
- Sepedi (Sesotho–Setswana language): “Go be go na le mosepedi yo a bego a na le tlala.”

These sentences contain very different numbers of words, average letters per word, and frequency of short words (i.e., words of one or two letters). A benchmark based purely on the number of words read per minute in English would therefore be inappropriate for a conjunctive language like isiZulu, and vice versa.

Once the decision was made to establish benchmarks across all South African languages, we embarked on a process to set benchmarks for each language. The process could not rely on existing English benchmarks as stepping stones; instead, it needed to account for language-specific relationships between accuracy, speed and comprehension so that the resulting benchmarks reflected how children actually learn to read in each language.

A similar consideration applied to English itself. Existing benchmarks for English as a second language are from settings where English is the dominant language and could not automatically be applied to South Africa’s multilingual context. In addition to benchmarks for African languages, it was therefore necessary to develop benchmarks for English in South Africa’s specific context.

The data behind the benchmarks

When the benchmarking work began in 2019, South Africa did not have nationally representative early-grade reading data that could be used to establish empirically grounded benchmarks. Rather than waiting for a single purpose-built national study, the DBE and its partners drew strategically on a rich body of existing and ongoing evaluations of early-grade reading interventions. While these data were not nationally representative, they were collected in typical public schools and, critically, included learners across the full spectrum of reading proficiency—from non-readers to fluent readers—making them well-suited for identifying developmental thresholds.

For the Nguni language benchmarks, learner assessment data from five different studies were combined to produce multiple assessment points for nearly 16,400 unique learners across approximately 660 schools. For the Sesotho–Setswana languages, a combination of existing datasets and targeted top-ups to ongoing data collection produced a sample of more than 19,000 learners from Grades 1 to 7 across 359 no-fee schools. In almost all these studies, learners were assessed multiple times, with between two and five assessment points per learner, strengthening the analysis of accuracy, fluency, and comprehension trajectories.

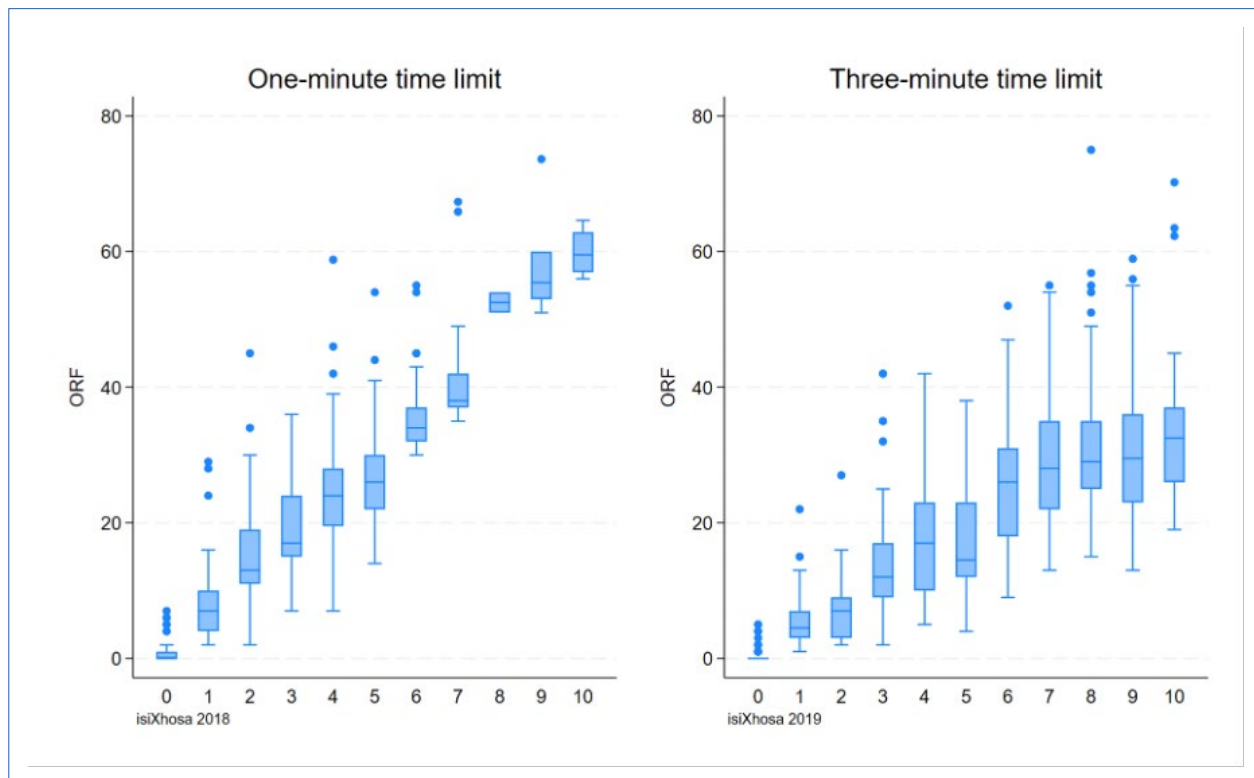
Data for the Afrikaans benchmarks were collected through targeted additions to an ongoing evaluation, while bespoke data collection was undertaken specifically for Xitsonga and Tshivenda, where no suitable existing datasets were available. Taken together, this approach ensured that benchmarks were grounded in extensive empirical evidence across languages, grades, and stages of reading development, while making efficient use of available data and existing research investments.

Why not rely on existing approaches to set benchmarks?

As South Africa’s government-led team embarked on the journey to set benchmarks, they first had to agree on a methodology that was acceptable for the country’s linguistic diversity and system. Initial investigations revealed that existing approaches to setting benchmarks had important technical limitations and were not well suited to low-literacy contexts. Most existing benchmarks were based on data collected using the original Early Grade Reading Assessment (EGRA), where a child typically only attempts the reading comprehension questions related to the portions of the passage that they read within a minute.

Using such data to set benchmarks represented two technical limitations. First, benchmarks are based on higher-performing learners who were able to read the full passage within one minute, limiting their usefulness for monitoring progress across a full distribution of learners. Second, the one-minute limit artificially inflates the observed relationship between fluency and comprehension, because learners with slower fluency rates attempt fewer comprehension questions.

The EGRA has since been adapted and refined to be better suited for benchmark setting. In EGRAs administered in South Africa, learners were allowed an additional two minutes to complete the passage. As shown in the benchmarking analyses, these extended-time data are more appropriate for establishing benchmarks because they allow independent measurements of comprehension from fluency rates. When children are given a total of three minutes to read the same passage, the relationship between oral reading fluency (ORF) and comprehension remains strong but tends to flatten, providing a more accurate picture of how fluency supports understanding.



As a result, benchmarks based on a fixed comprehension-level approach tend to be inherently arbitrary and at risk of being distorted by variation in item difficulty across languages and passages.

How we approached benchmarking: Exploratory and empirical methods



Identifying upper and lower thresholds by language

Reading development theory suggests that the relationship between fluency and comprehension may be non-linear. We applied this theory to identify critical decoding thresholds in learners' reading development by examining the full range of data rather than focusing only on a fixed comprehension level. Specifically, we looked for lower and upper thresholds for each language.

The lower threshold represents the fluency point below which comprehension is unlikely to develop because decoding is still too laboured. The upper threshold represents the point at which limited comprehension skills become the binding constraint, such that further increases in fluency yield no substantial gains in understanding.

Because African languages are under-studied, we adopted an exploratory and empirical approach. We were careful not to impose external assumptions on the accuracy–speed or fluency–comprehension relationships. Instead, we used non-parametric analytical methods to identify patterns for each language.



Identifying upper and lower thresholds by language

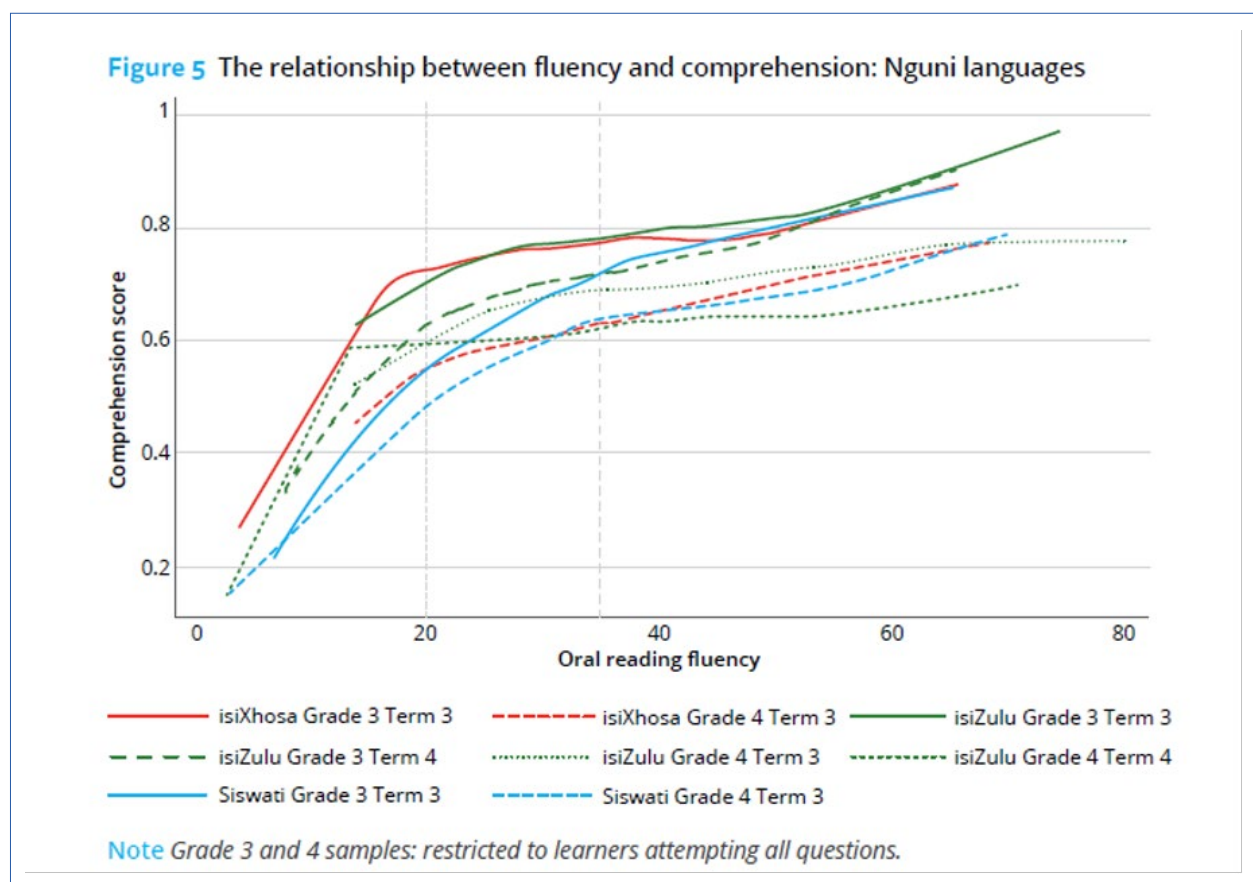
For learners reading below the instructional level of accuracy, decoding is typically laboured and hinders meaning-making. Across samples, passages and grades, we examined the relationship between accuracy and speed, focusing on the speed associated with instructional-level accuracy, typically 95 percent (Betts, 1946). The lower threshold was then set around the speed level below which the average learner fell below this instructional level of accuracy.

We then examined the fluency–comprehension relationship to determine whether there was an upper threshold beyond which no further comprehension gains were achieved with increases in fluency. Beyond this point, comprehension skills become the limiting factor and instructional focus should shift from improving fluency to developing comprehension skills. Where comprehension continued to improve with increasing fluency, the upper threshold was defined as the fluency point above which a reasonable level of comprehension was supported.

An example from Nguni languages

The thresholds are illustrated for samples of Grade 3 and 4 learners in Nguni languages. Across grades and passages, children reading slower than around 20 words per minute tended to be below the instructional level of accuracy. This became the lower threshold: the relationship between fluency and comprehension is initially steep and then tended to flatten, suggesting that comprehension skills became the limiting factor.

The fluency level at which the fluency-comprehension relationship flattened differed across passages, highlighting the limitations of focusing on a fixed comprehension level (for example, 80 percent) that does not allow for variation in question difficulty across passages and languages. Across grades and passages, around 35 correct words per minute emerged as the upper threshold: beyond this point, further gains in fluency yielded limited improvements in comprehension.



From thresholds to meaningful benchmarks

Establishing concurrent and predictive validity

Once thresholds were identified for each language, they needed to be contextually relevant and practically meaningful to function as benchmarks. To turn thresholds into benchmarks, the numerical values had to represent levels of achievement that were meaningful at the grade-level of interest (concurrent validity) and predictive of performance in later grades (predictive validity). They also had to be connected to the South African curriculum – its content, pace and sequence – to enable monitoring of progress within and across school years.

We used data on related reading skills in Grades 2 and 3 to explore concurrent validity: did the potential benchmarks align with meaningful distinctions between learners and stages of reading development? We then assessed concurrent validity by examining how the fluency thresholds aligned with learners' performance on written comprehension assessments.

To explore predictive validity, we used longitudinal data to examine the relationship between a learner's benchmark status in Grades 2 to 4 and their performance on written comprehension tasks in Grades 5 to 7. Children classified as having met the benchmark in the earlier grades were much more likely to successfully tackle the written comprehension task, confirming that the benchmarks were useful predictors of future performance.

Aligning benchmarks with curriculum and context

Concurrent and predictive validity make benchmarks useful markers of learner progress. However, benchmarks must also be linked to the curriculum and be contextually realistic. A benchmark with strong predictive validity but far beyond reach for most learners in a given context has limited utility. It does not support tracking incremental improvements or guiding remediation and instruction.

At the same time, benchmarks must be ambitious. Ambitious yet realistic benchmarks can be powerful tools to encourage system improvement toward the levels of learning expected in the curriculum. In South Africa, we balanced these considerations through a data-driven approach combined with expert judgement. We assessed the attainability of proposed benchmarks and engaged curriculum experts and language specialists to ensure that benchmarks were both meaningful and feasible.

Where we are now: common, grade-specific benchmarks across languages

As a result of the empirical analyses, expert consultations and further testing, South Africa was ready to set meaningful and usable benchmarks for early-grade reading proficiency. Our analyses showed sufficient similarities within language families to allow common benchmarks in some cases. We were able to set oral reading fluency benchmarks for all 11 official languages in Grade 2 and Grade 3, as summarised below (correct words per minute):

Language	Oral reading fluency benchmark (correct words per minute)	
	Grade 2	Grade 3
Nguni language family (isiZulu, isiXhosa, isiNdebele, SiSwati)	20	35
Sesotho-Setswana language family (Sepedi, Setswana, Sesotho)	40	60
English First Additional Language	30	50
Afrikaans	45	60
Xitsonga	35	55
Xitsonga	35	55
English Home Language	50	70

The upper thresholds from our exploratory work came to represent the end-of-Grade 3 benchmarks. For Grade 2, benchmarks were set closer to the lower thresholds, based on the initial analyses, additional analytics and expert inputs. For Grade 1, where most South African children are not yet reading words by year-end, a benchmark of 40 correct letter sounds per minute was established.

At this stage, we view these benchmarks not as aspirational targets but as minimum standards – points that every child should reach by the end of the year to be on track for comprehension in subsequent grades.

How we did it and why it worked: Collaboration and Capacity Building

The success of South Africa's benchmarking work lies as much in its process as in its methodology. From the outset, it was a collaborative effort — uniting the Department of Basic Education, universities, and local research institutions in a shared national project.

This was not an externally driven exercise but a South African-led, government-owned initiative, designed to strengthen long-term institutional capacity. True collaboration and engagement takes time – especially when it involves multiple grades, languages, organizations, institutional interests and policy. Over several years, partners worked together to refine methods, test assumptions, interpret results and embed the findings in policy and practice.



A key lesson in the ongoing journey to develop African language benchmarks has been that strong collaboration between funders, stakeholders, research organisations, and African language specialists is vital to taking the benchmarking agenda forward. Collaboration has supported innovation and excellence in the benchmarking process while promoting capacity building.” Mohohlwane et al. 2023

What Next: From Data to System Change

Through FUNS, South Africa has committed to collecting nationally representative data on foundational learning levels across the country every three years. Having set meaningful and usable benchmarks is only the first step; the next challenge is integration and use of the data to drive system change.

The DBE has already begun embedding the benchmarks into teacher training, learning materials, and monitoring frameworks. Provincial and district teams are being equipped to use benchmark-linked indicators to identify struggling schools, tailor support, and track progress. At the classroom level, the benchmarks are being communicated as foundational milestones: reaching them is a first step, after which instructional focus should shift toward developing learners' comprehension skills.

At the national level, benchmarks provide a coherent foundation for aligning curriculum, pedagogy, and assessment. They also offer a platform for peer learning across Africa. South Africa's experience provides a practical roadmap for how to establish benchmarks at scale in multilingual settings and how to link them to broader efforts to improve foundational learning.



Conclusion: measuring what matters, in every language

The journey to establish early-grade reading benchmarks in South Africa has been both technically-demanding and transformative. It has required navigating complex linguistic realities, building national capacity, and rethinking how our system defines and measures learning.

The FUNS study is not just the culmination of that work; it is a symbol of what is possible when evidence, collaboration, and commitment align. As more countries across Africa invest in measuring foundational learning, South Africa's experience offers an encouraging message: with the right methods, partnerships, and purpose, we can measure what matters, in every language, for every child.

Resources:

Ardington, C., Wills, G., Pretorius, E., Mohohlwane, N. and Menendez, A. (2021) "Benchmarking oral reading fluency in the early grades in Nguni languages" International Journal of Educational Development 84

Mohohlwane, M., Wills, G. and Ardington, C. (2022) A review of recent efforts to benchmark early reading skills in South African languages. In Spaul, N. & Pretorius, E. (Eds) Early Grade Reading in South Africa. Cape Town: Oxford University Press.

Benchmarking reports for each language and the Funda Uphumelele National Study reports can be found here:

<https://www.education.gov.za/Research,MonitoringEvaluationReports.aspx>



Africa Foundational Learning
Assessment Initiative