SEPTEMBER 2025

Data

Tracking Foundational Learning through a Zero-Word Reading Indicator

A feasibility study

Author:
Cally Ardington (AFLEARN, University of Cape Town)

AFLEARN
DATAHUB

Abstract

This feasibility study examines whether a zero-word reading indicator—the percentage of children in Grades 2 or 3 (or approximately age 8–9) who cannot read a single word—can serve as a practical tool for tracking foundational learning in African contexts. The motivation for this work stems from the limitations of the World Bank's "learning poverty" measure, which relies largely on Grade 6 assessments, excludes many countries in Sub-Saharan Africa, and uses proficiency thresholds that are difficult to communicate to non-specialists. A zero-word measure offers a simpler, more intuitive alternative that could provide an early warning signal of learning deficits, particularly in low-performing systems.

The study reviews the feasibility of constructing such an indicator using existing data sources (MICS, PASEC, EGRA, and national assessments) and considers the implications for comparability across different types of surveys. The findings suggest that while meaningful continental coverage is not yet possible, upcoming improvements to the MICS FLS module, the release of PAL's ICAR results, and the expansion of PASEC in 2026 will create a stronger basis for measurement. Challenges remain around age versus grade alignment, language of assessment, and data access, but with modest refinements, a zero-word indicator could become a feasible, interpretable, and impactful complement to existing SDG reporting on foundational learning.

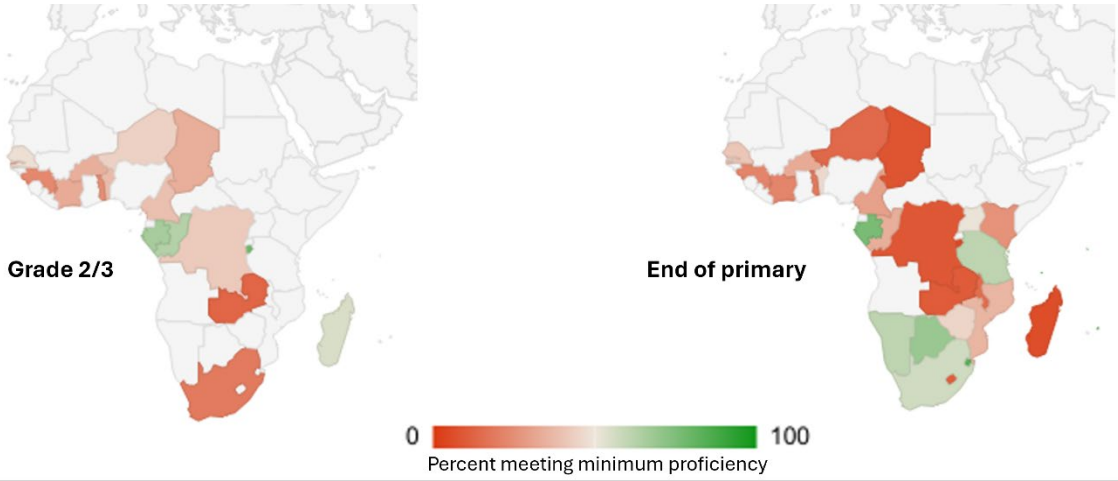
1. Background and Rationale

The World Bank's learning poverty indicator has played a pivotal role in raising global awareness of the learning crisis. Defined as the percentage of 10-year-olds who cannot read and understand a simple text, it captures both learning deprivation (children enrolled in school but not meeting minimum proficiency) and schooling deprivation (children out of school, assumed not to meet the minimum proficiency threshold).

Despite its global importance, efforts to track progress in foundational learning across Africa have been constrained by several challenges.

Assessment misalignment: In Sub-Saharan Africa, there is a mismatch between the target age (10-year-olds) and the actual grades levels of learners where data is collected. Because of data and comparability constraints, most learning poverty estimates in Africa come from Grade 6 assessments rather than from direct testing of 10-year-olds. Figure 1 shows the percentage of children at the grade 2/3 and end of primary level who are meeting the minimum proficiency level for Sustainable Development Goal (SDG) reporting. As shown, coverage is substantially higher for the end of primary school than at the grade 2/3 level.

Figure 1. Percent of grade 2/3 and end of primary children meeting minimum proficiency levels - SDG Report Card 2025



Data source: SDG Report Card 2025 accessed at https://dashboards.sdgindex.org/map/goals/sdg4/trends/

Limited country coverage: Fewer than half of children in Africa live in countries with suitable learning assessments for inclusion.

Opaque proficiency thresholds: The concept of "minimum proficiency" is technical and often difficult for non-specialists—such as parents, community members, and policymakers—to interpret or act upon.

Against this backdrop, the purpose of this document is to assess the feasibility of a zero-word reading indicator—the proportion of children in Grades 2 or 3 (or around age 8–9) who cannot read a single word. The study explores whether such an indicator can be derived from existing datasets, and whether forthcoming data sources will allow for meaningful coverage across the continent in the near

future. The ultimate aim is to determine whether a zero-word measure could serve as a simple, interpretable, and advocacy-friendly complement to existing indicators of foundational learning such as learning poverty.

2. Motivations for an Alternative Measure

There are several reasons why a zero-words indicator would be a valuable complement to the World Bank's learning poverty indicator:

Focus on foundational skills: The measure captures foundational skill failure within a critical developmental window when early interventions are still possible.

Intuitive: Unlike abstract proficiency thresholds derived through scaling or policy-linking, the inability to read a single word is immediately understandable to all stakeholders. It is a vivid and compelling marker of foundational literacy failure.

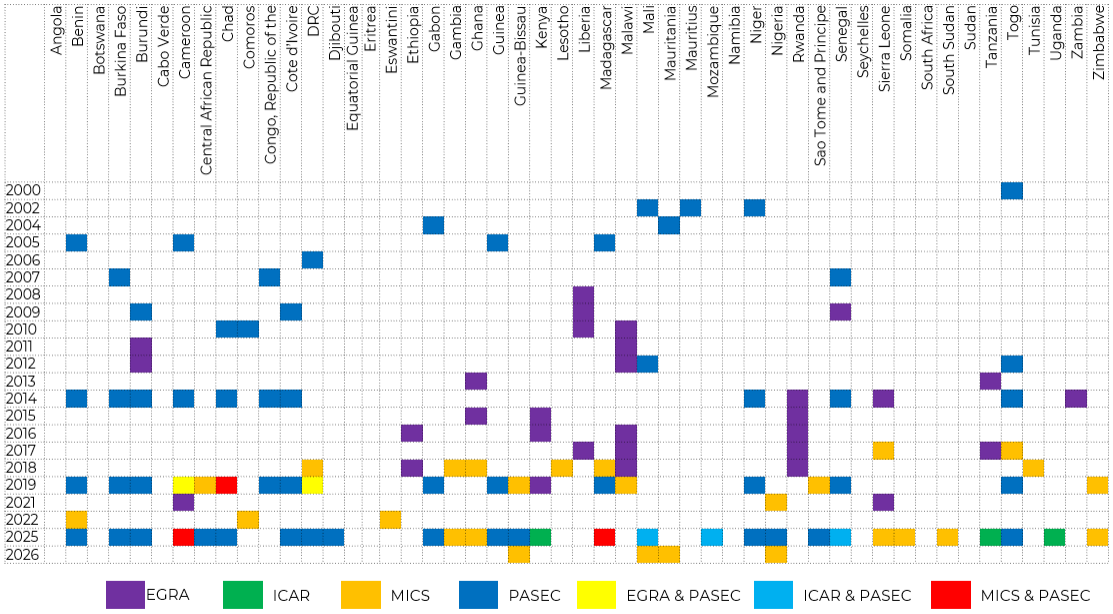
Better discrimination for low-performing systems: In many African contexts, very few children meet minimum proficiency benchmarks, limiting the usefulness of the learning poverty measure for tracking incremental progress. Measuring the proportion of children unable to read any words allows for the detection incremental gains in low-performing contexts

Methodological simplicity and comparability: The indicator is based on a binary outcome—zero correct words—avoiding the methodological challenges of complex scaling and mapping into proficiency levels. The task is basic enough to avoid language and curriculum distortions.

Improved data availability at the right age/grade: Unlike learning poverty, which relies mostly on end-of-primary assessments, this indicator draws on Grade 2 and 3 data, when foundational skills are still being established. Relevant large-scale sources already capture or can easily capture this information, including:

- Early Grade Reading Assessments (EGRA)
- UNICEF's MICS Foundational Learning Skills (FLS) module
- PAL Network's ICAR tool
- PASEC (Programme d'Analyse des Systèmes Éducatifs de la CONFEMEN)
- National one-on-one assessments (e.g. Madagascar, South Africa)
- Nationally representative one-on-one assessments that have been conducted since 2000 or are planned in the near future are summarised in Figure 2.

Figure 2. Nationally representative one-on-one assessments by country, year and assessment.



The zero-words indicator would not replace learning poverty but would serve as an earlier, more visible, and more accessible signal of whether education systems are equipping children with the most basic literacy foundations.

3. Feasibility of the Measure

To establish the feasibility of using existing data for this indicator, two key requirements must be met:

1. The ability to reliably identify children who cannot read a single word.
2. The ability to construct comparable samples across countries and assessments.

The following sections assess these requirements in the context of major large-scale assessments.

3.1. PASEC Grade 2

The PASEC Grade 2 instrument includes a one-on-one assessment of oral reading fluency. In the publicly released data, oral reading fluency is reported in five categories rather than as a continuous measure of correct words per minute. Children who cannot read a single word are included in the lowest category, which groups all children reading 0–5 correct words per minute.

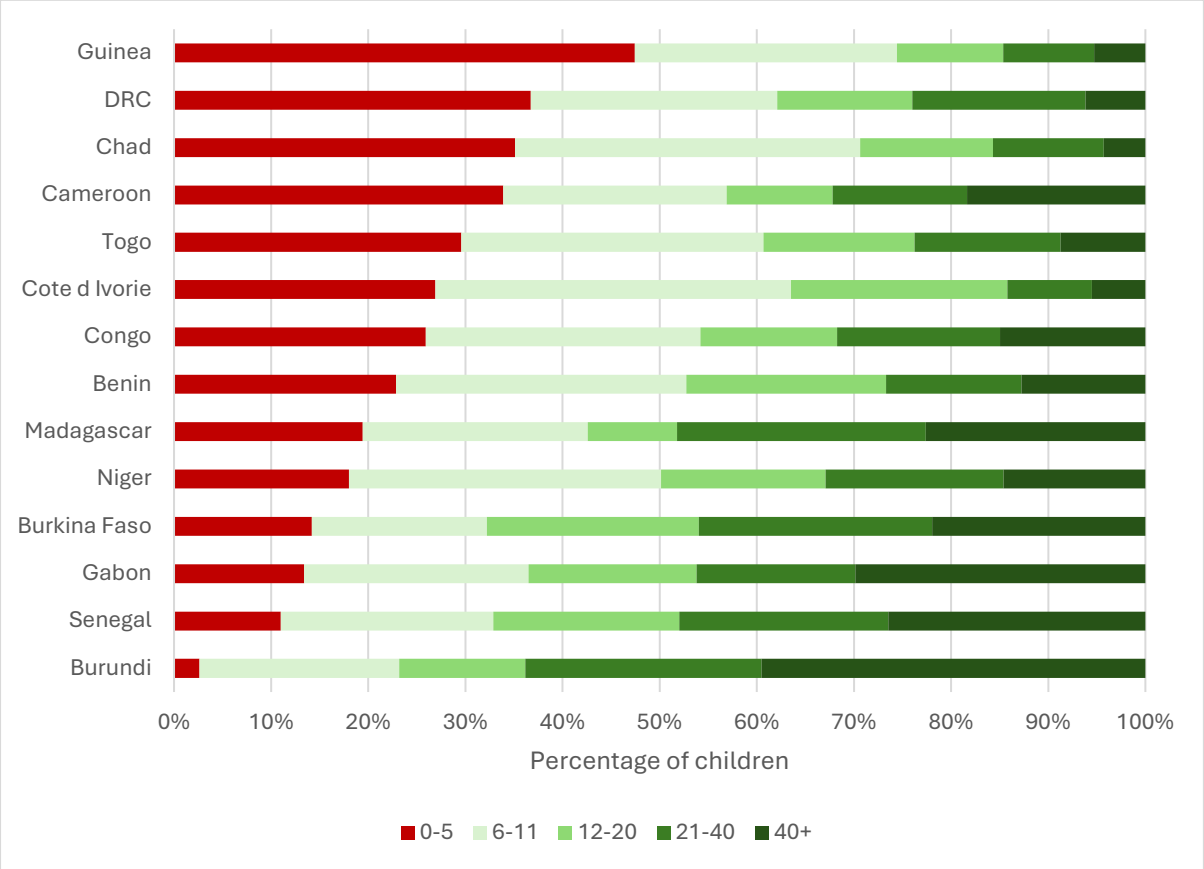
For the purposes of constructing a zero-score indicator, two options are feasible:

Short-term: Use the proportion of children reading five or fewer words per minute as a proxy for the zero-score measure.

Medium-term: PASEC could either release the raw oral reading fluency data or add a dedicated category for children with zero words correct in future rounds.

Figure 3 illustrates the distribution of oral reading fluency across PASEC countries. The percentage of children reading five or fewer words per minute ranges, for example, from 3% in Burundi to 47% in Guinea.

Figure 3. PASEC Grade 2 Oral Reading Fluency Categories

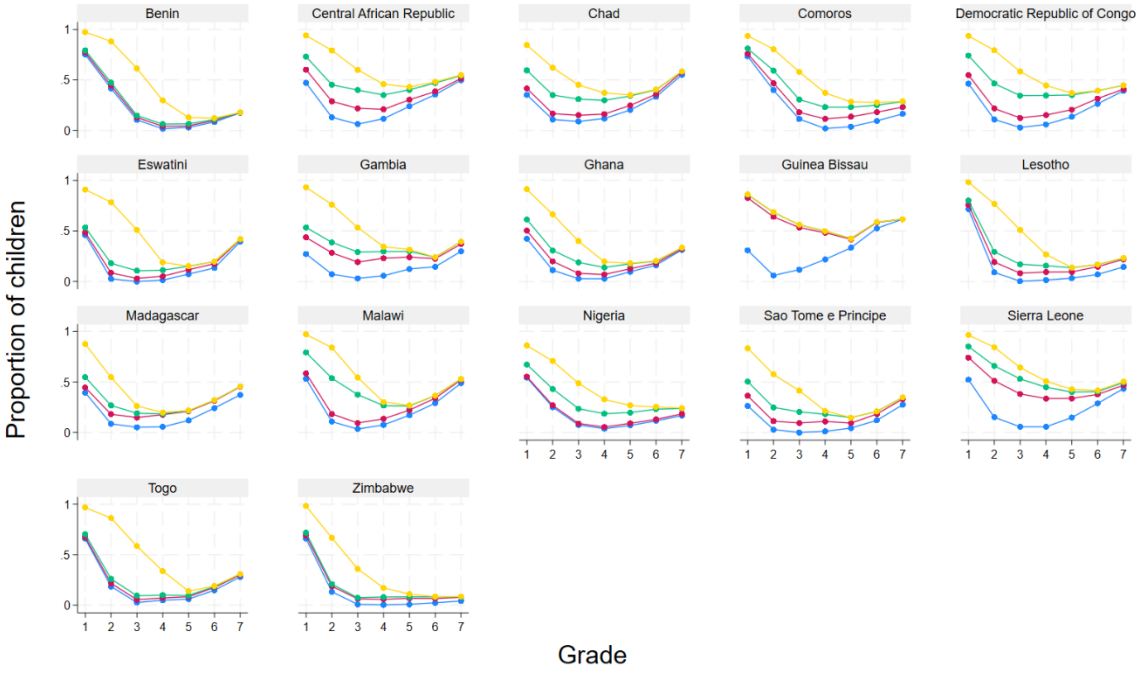


3.2. MICS Foundational Learning Skills (FLS) Module

The MICS Foundational Learning Skills (FLS) module includes a reading passage and records the number of words correctly read, enabling the identification of children who cannot read a single word. Conducted at the household level, MICS offers flexibility in constructing samples of children by grade or age and, importantly, includes out-of-school children. However, for the purposes of a zero-score indicator, it is necessary to construct a sub-sample that is broadly comparable to at least the PASEC Grade 2 sample and, ideally, also to nationally representative EGRA samples, which are school-based and generally focus on Grades 2 or 3.

Focusing on an early grade sample presents several challenges for constructing a zero-score indicator from MICS data.

Figure 4. Sources of missing data for zero-score indicator in MICS by grade



Note: The blue line indicates the proportion of children in each grade who are excluded from the MICS FLS module due to age-ineligibility. The red line also includes children who do not attempt the reading passage due to a language mismatch (i.e. the red line indicates both age-ineligible and language mismatch). The green line additionally includes children who refuse to read the story. The yellow line includes all the previous categories together with the children who failed the practice screener. The yellow line therefore indicates the proportion of children in the grade who are missing data for the zero words indicator.

The FLS module is administered only to children aged 7–14. This restriction creates problems when analyzing early-grade learners. The blue line in Figure 4 shows the proportion of children in each grade who are not age-eligible to complete the FLS module across the MICS countries. On average, 15% of Grade 2 learners across MICS countries are excluded on the basis of age, though this varies substantially—from 3% in Eswatini to 42% in Benin. By Grade 3, the average exclusion rate falls to 5%. Expanding the eligibility criteria to include six-year-olds would reduce the proportion of Grade 2 learners excluded to just 3%. While this change is modest, it would have a significant impact on the indicator's coverage. Interestingly, these age cut-offs also pose challenges for estimating outcomes at the end of lower primary (Grades 6 and 7).

A second barrier is the high proportion of children who do not attempt the reading passage due to a language mismatch. In most MICS countries, if neither the child's home language nor their school's language of instruction matches one of the assessment languages, the reading task is terminated. The red line in Figure 4 indicates the proportion of children who are missing reading data because they are age-ineligible to complete the FLS module, or because of a language mismatch i.e. the gap between

the red and blue lines is due to the language mismatch[1]. The proportion of excluded children varies greatly depending on linguistic diversity, language-of-instruction policies, and the availability of assessments in local languages. For example, in The Gambia, 7% of Grade 2 learners are excluded for being under- or over-aged, while a further 19% are excluded due to language mismatch. This issue disproportionately affects the early grades, as many African countries transition completely to an international language (often the colonial language) in later grades.

Refusals present an additional challenge for MICS. Although caregiver or child refusals to complete the overall FLS module are rare, refusals are higher for the reading passage. Enumerators explicitly ask whether the child wants to read a story, and a substantial proportion decline—even though the same children go on to complete the numeracy assessment. This problem is concentrated in the early grades. In The Gambia, for instance, an additional 10% of Grade 2 learners do not attempt the reading passage for this reason. The gap between the green line and the red line in Figure 4 indicates the proportion of children who declined to read the story. The prevalence of refusals varies considerably across countries.

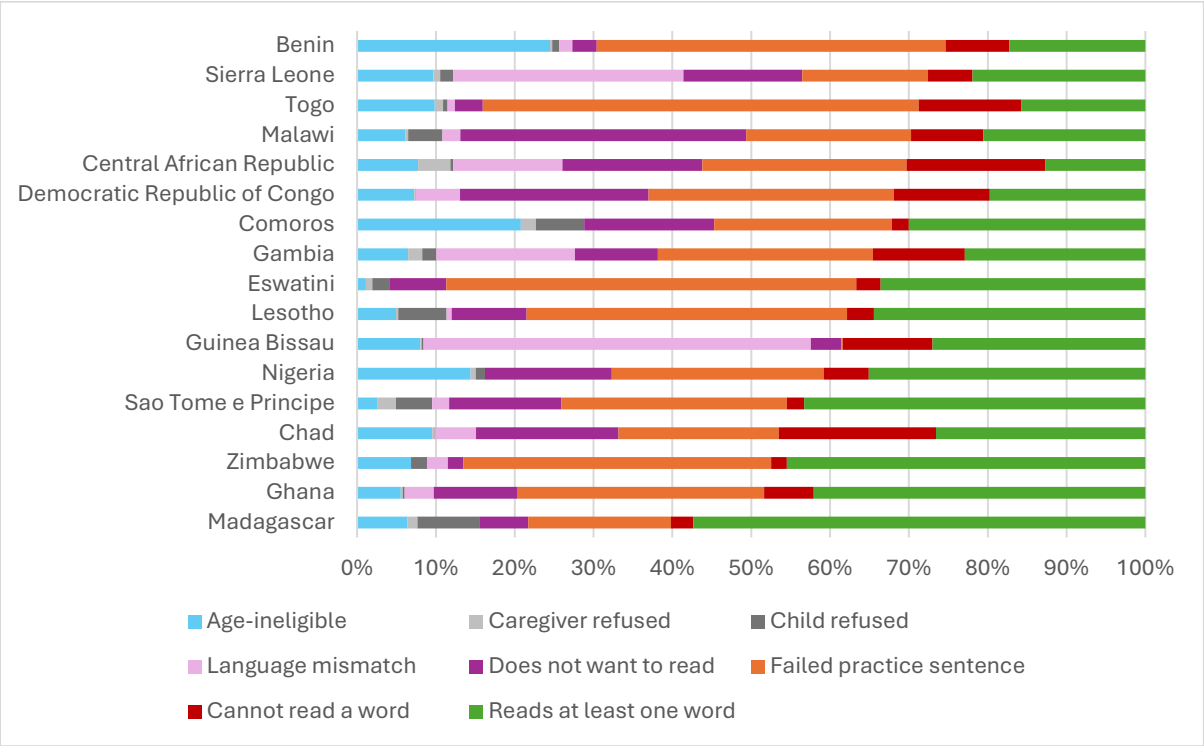
The final barrier results from the screening of children into the reading task. Children aged 7–9 and those who have never attended school must first read three short practice sentences (14 words in the English version). Only those who read all the words correctly proceed to the full passage. For children who fail this screening, we do not know whether they were able to read at least one word. This creates substantial missing data for early-grade learners. In The Gambia, for example, an additional 37% of Grade 2 learners fall into this category. The yellow line in Figure 4 includes children who zero-word status is unknown as they failed the screening. A simple solution would be to adjust the response coding for practice sentences from the current binary (“all correct” vs “not all correct”) to three categories: “zero words correct,” “some words correct,” and “all words correct.”

Focusing on Grade 2 and 3 learners, Figure 5 illustrates the detailed reading status by country. Overall, missing data are extensive, with significant cross-country variation ranging from 42% in Madagascar to 76% in Benin[2]. The reasons also differ: in some countries, missingness is driven mainly by language mismatches, while in others it is due to refusals or screening procedures.

[1] The red line also includes children who refused to complete the FLS module and children whose caregivers refused on their behalf. However, refusal rates are very low. In the case of the Gambia, 1% of caregivers refused and 1% of children refused.

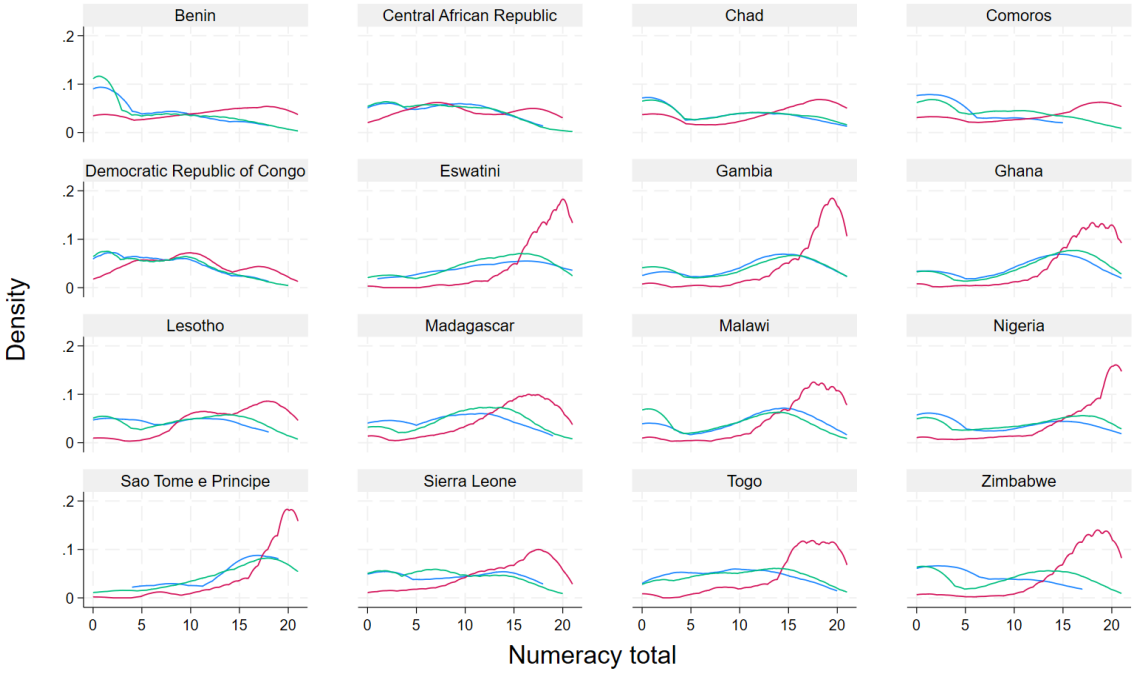
[2] Note that for the purposes of determining minimum reading proficiency, children who fail the practice screener would be considered not proficient and rates of missing data therefore range from 14% in Eswatini and Zimbabwe to 64% in Guinea Bissau.

Figure 5. Reading status of children enrolled in grade 2 or 3 – MICS FLS



In its current form, MICS data contain too much missing information to reliably estimate the percentage of Grade 2/3 children unable to read a single word. One possible strategy is to use outcomes from the numeracy assessment to impute values for children with missing reading data. Figure 6 illustrates the distribution of numeracy scores by reading status. Children who can read at least one word (red line) cluster heavily to the right of the 50% mark, indicating higher numeracy performance. By contrast, children who score zero on word reading (blue line) display a wider spread, with relatively few achieving scores above 15 out of 20. The green line shows the distribution of numeracy scores for children who failed the practice sentence and whose zero-word status cannot be determined. Across countries, the distributions of zero-word readers and children with unknown status are fairly similar to each other, and both differ clearly from the distribution of children who have started to read. While this might suggest that children with unknown status are closer to zero-word readers, the degree of missing data and the substantial overlap across the three groups make imputation inappropriate.

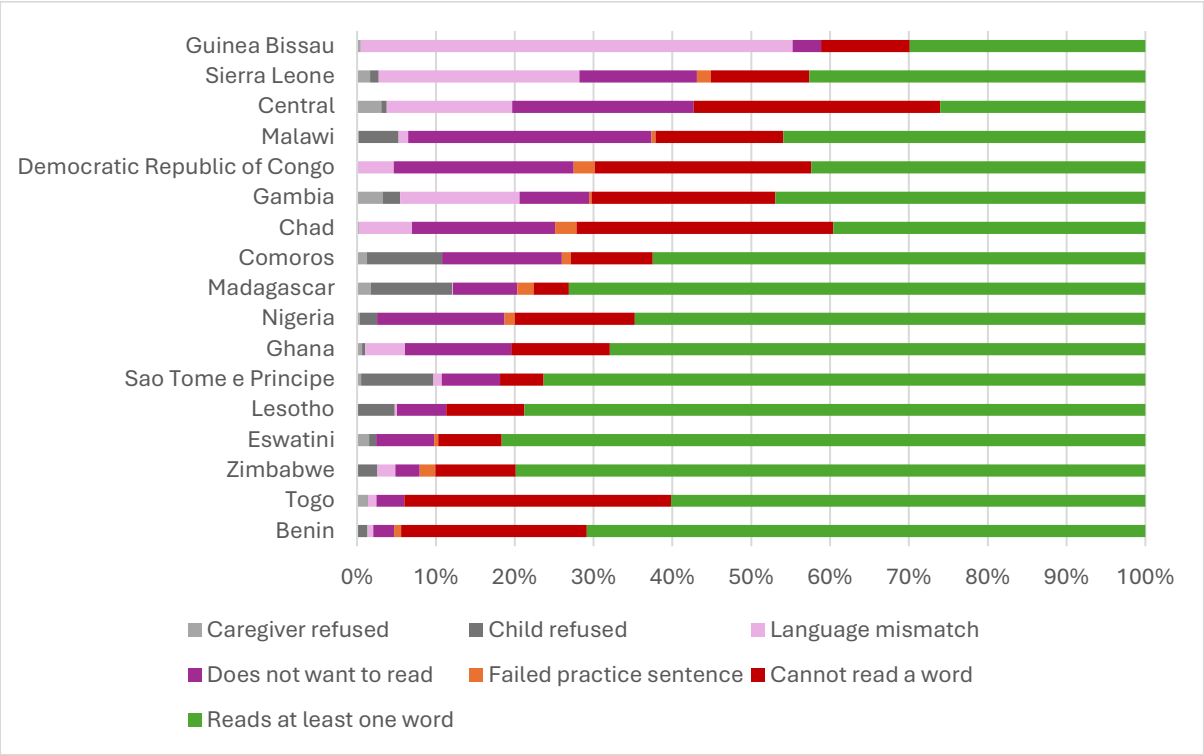
Figure 6. Distribution of numeracy total scores by reading status of children enrolled in grade 2 or 3 – MICS FLS



Note: The red line indicates children who can read at least one word. The blue line indicates children who are unable to read one word. The green line represents children who failed the practice sentence and whose zero-word status we cannot determine.

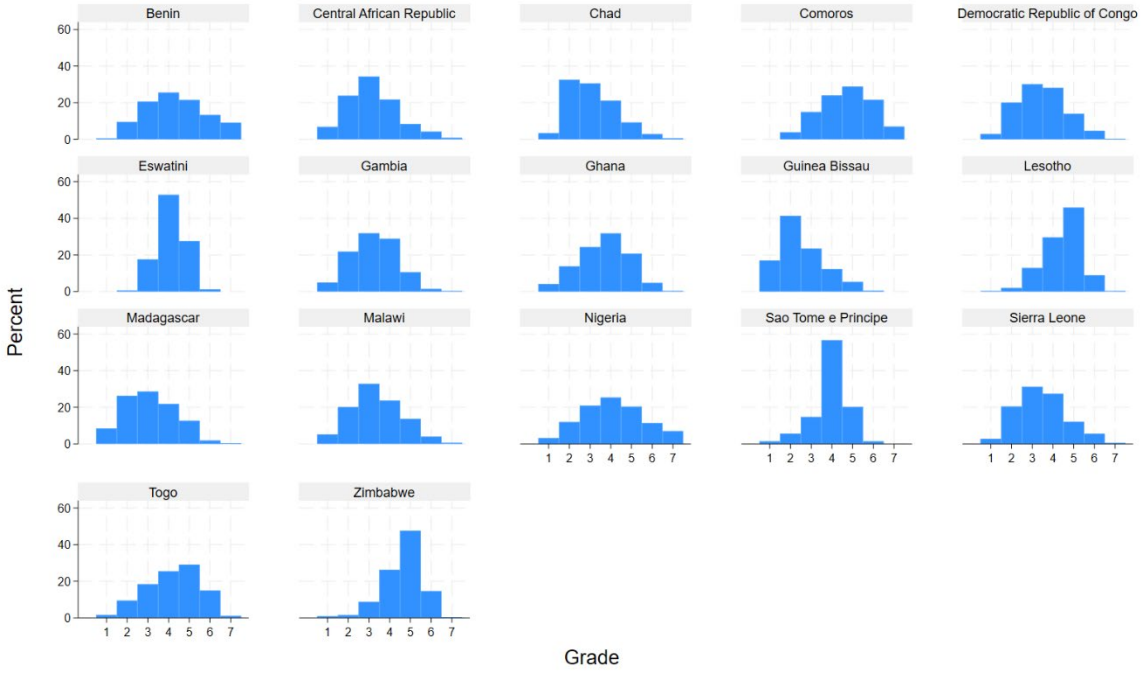
Another possibility would be to look at the sample of enrolled 10-year-olds, who would not be excluded from the FLS by age and who also would not attempt the practice sentence before attempting the reading passage. Figure 7 shows the reading status of 10-year-old children. The degree of missing data is substantially lower than for the grade 2/3 sample, ranging from 6% in Benin and Togo to 59% in Guinea Bissau. The change in status of Benin for the highest level of missing data in the grade 2/3 sample to the lowest level of missing data in the 10-year-old sample, highlights just how much the drivers of missing information differ between countries. Nevertheless, language mismatches, particularly in Guinea Bissau and Sierra Leone remain problematic and in eight countries more than 10%e of children say that they do not want to read the story.

Figure 7. Reading status of 10-year-old children – MICS FLS



The main challenge with using a sample of 10-year-olds is how to align it with datasets that are structured by grade, typically Grades 2 and 3. Figure 8 illustrates the grade distribution of 10-year-old children. Although Grade 4 is the modal category in most countries, there is wide variation across both grades and countries. By contrast, Figures A1 and A2 in the appendix show the age distribution of children enrolled in Grades 2 and 3, confirming that the vast majority are younger than 10 years old. Comparing a MICS sample of 10-year-olds with Grade 2 and 3 samples from school-based assessments would therefore require adjustments to account for these age differences. While such adjustments are technically possible using MICS data, they would once again be constrained by the high levels of missing information for children under the age of 10, and by the limitations inherent in working with a select subsample.

Figure 8. Grade distribution of enrolled 10-year-olds – MICS FLS



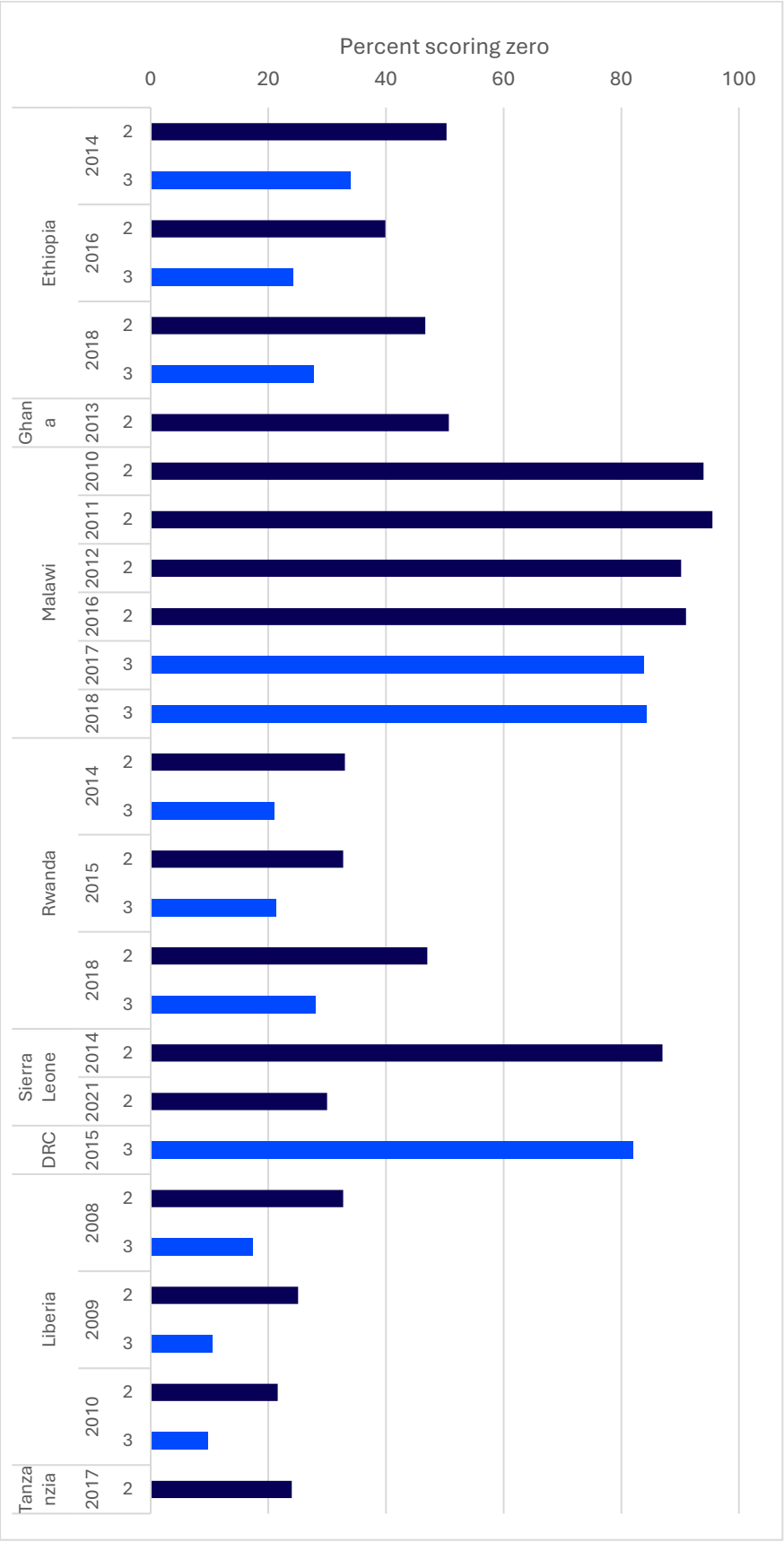
The FLS module was originally designed to measure the percentage of children proficient in reading and mathematics, rather than to capture the proportion who cannot read a single word. In its current form, MICS data contain too much missing information to reliably estimate the share of Grade 2/3 children unable to read a word. However, the MICS FLS module has recently been substantially redesigned to align with SDG reporting criteria, and several of these changes improve the measurement of zero-word readers. Children will no longer be asked whether they want to read a story, and the practice sentence for younger children has been removed. Instead, all children will attempt a simpler word-reading task before progressing to the passage, which will allow for the direct identification of those who cannot read a single word. The main remaining limitation is the continued age exclusions for younger grades.

3.3. Early Grade Reading Assessment

Early Grade Reading Assessment (EGRA) data typically include the number of words correctly read from oral reading fluency tasks, which makes it possible to directly identify children who score zero. While a large number of EGRAs have been conducted across Africa, relatively few have been nationally representative; most have been implemented as part of localised impact evaluations.

Figure 9 shows the percentage of Grade 2 or 3 children scoring zero on the oral reading fluency subtask, by country and year, for nationally representative EGRAs. These figures draw on a combination of hard copy reports and newly released datasets published on DataLumos at ICPSR. With the exception of Kenya and Sierra Leone in 2021, these EGRA assessments were conducted in 2018 or earlier.

Figure 9. Percentage scoring zero on nationally representative EGRAs



Looking ahead, data availability may prove challenging. Unlike regional programs such as PASEC, EGRA is not a standing initiative but rather a tool used in specific projects. Access to data will therefore depend on negotiations with individual countries. In addition, the future of EGRA itself is uncertain following the closure of USAID's global reading programs, which historically provided much of the funding and technical support.

3.4. International Common Assessment of Reading

The PAL Network has recently conducted nationally representative assessments using the International Common Assessment of Reading (ICAR) tool in Kenya, Tanzania, Uganda, Mali, Mozambique, and Senegal, with results expected in November 2025.

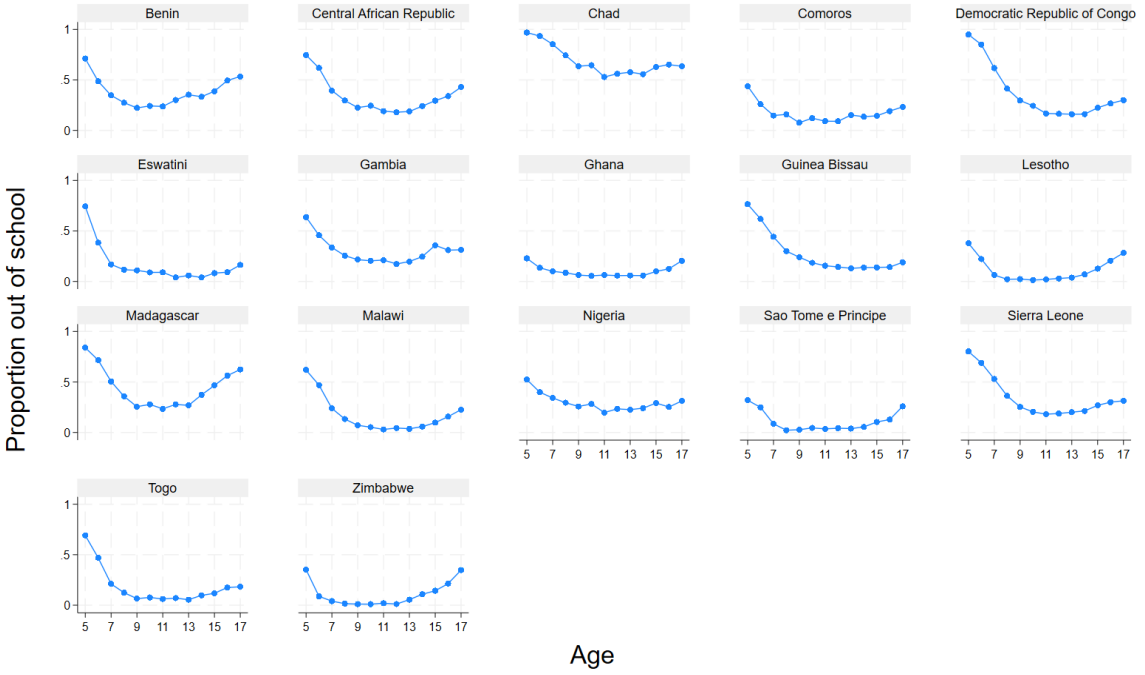
It is conducted as a household survey, covering children aged 5 to 16 years. The assessment begins with a letter-identification task: children are shown five letters and must correctly identify at least three to move on to the word-reading task. Every item is recorded individually (e.g. each letter, each word), not just whether a child passed the level (such as getting three of five correct). This detailed item-level data will facilitate identifying children who cannot read a single word. The only assumption required is that children unable to identify at least three letters would also not be able to read a word.

4. Comparability across assessments

A key strength of the zero-word indicator is its methodological simplicity, which avoids the challenges of scaling or mapping into proficiency thresholds. However, the greater challenge for comparability lies in differences across assessment samples.

Household surveys such as MICS and ICAR allow for analysis by both age and grade and include out-of-school children. In contrast, school-based assessments (e.g. EGRA, PASEC) only capture the enrolled population. Aligning across these data sources therefore requires focusing on grade-based samples. Nonetheless, household surveys offer unique insights into disparities between enrolled and out-of-school children that would otherwise remain hidden. Figure 10 shows the proportion of children at each age who are out of school across MICS countries. There is substantial variation, with the highest out-of-school rates occurring at ages associated with the early grades.

Figure 10. Proportion of children out of school by age and country - MICS



Another issue relates to whether school and household samples are nationally representative. For example, Gustafsson (2025) highlights discrepancies between school-based assessments and household surveys on indicators such as electricity in the home, suggesting possible biases in school sampling frames. Figures 11 and 12 compare the age distribution of Grade 2 learners in PASEC and MICS samples for five countries where both assessments were conducted. The distributions appear broadly similar, which is reassuring for cross-assessment comparability.

While methodological simplicity makes the zero-word measure attractive, ensuring comparability across household- and school-based surveys remains a central challenge. Addressing differences in sampling frames, age–grade profiles, and representativeness will be critical if the indicator is to provide robust and comparable insights across countries.

Figure 11. Grade 2 age distribution for selected countries - PASEC

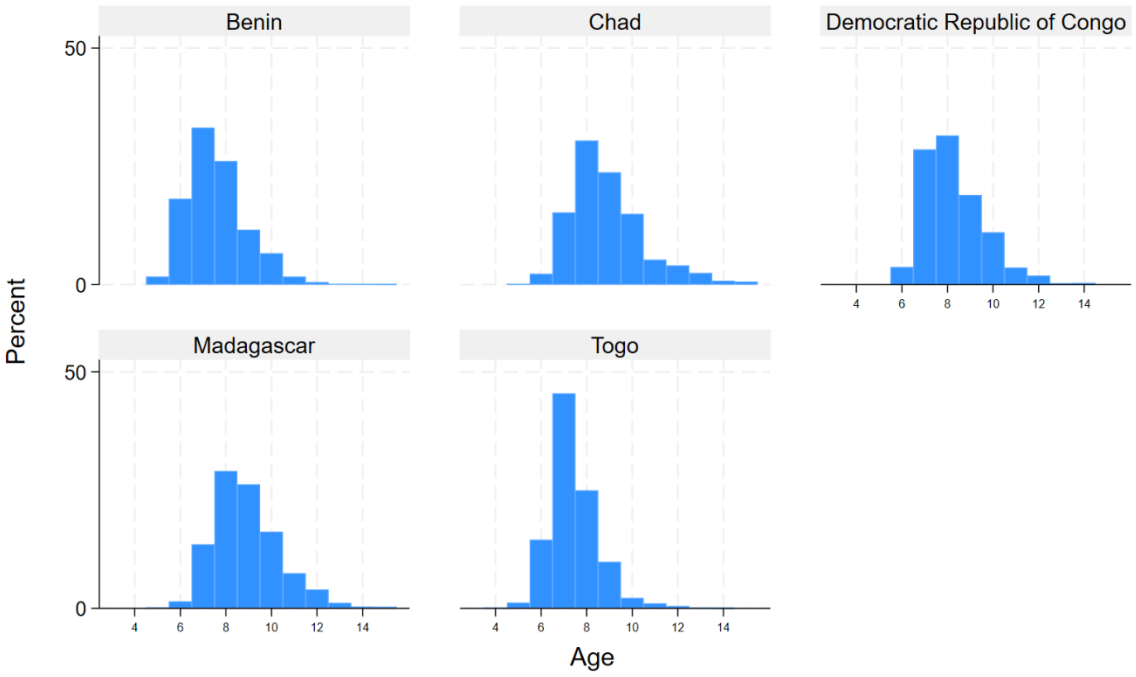
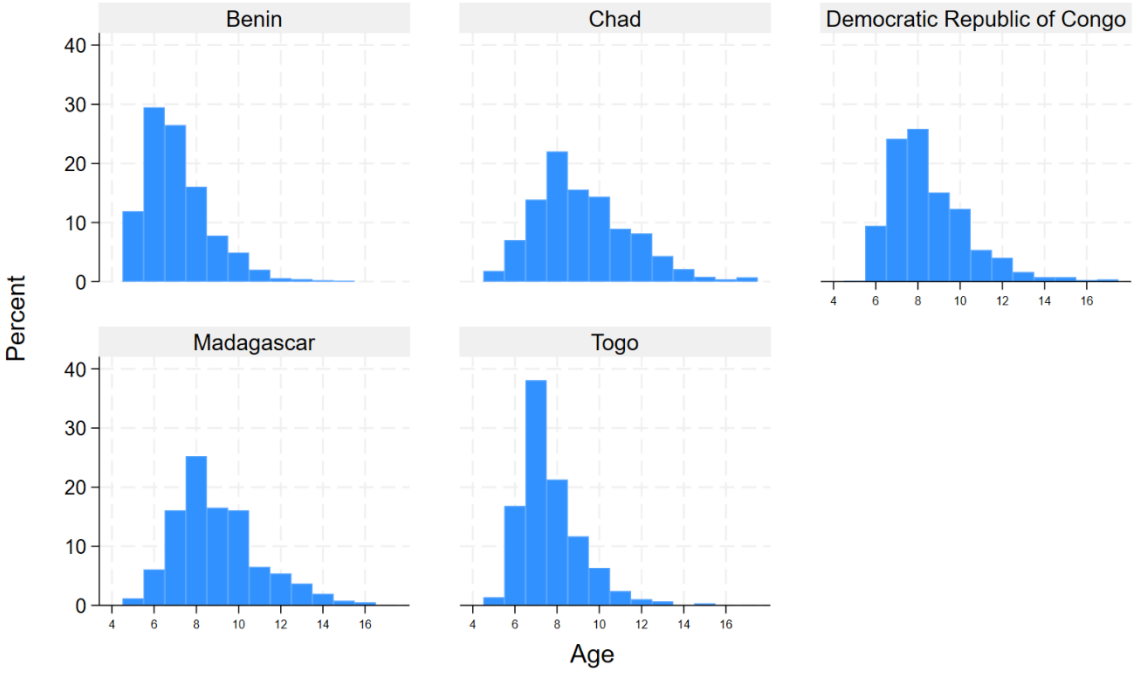


Figure 12. Grade 2 age distribution for selected countries - MICS



5. Conclusion

A zero-word reading indicator has strong potential to serve as a clear, interpretable, and impactful tool for tracking foundational learning in African contexts. Although meaningful continental coverage is not yet possible, the use of the updated FLM2.0 module in new rounds of MICS, the release of ICAR results towards the end of the year, and an expanded set of PASEC countries coming online in 2026 will significantly enhance the feasibility of this indicator.

However, several additional challenges will need to be addressed:

Alignment with SDG reporting. Ongoing efforts to harmonise reporting standards may reduce the distinction between data available for minimum proficiency measures and for zero-score indicators in the early grades.

Language of assessment: While it is technically feasible to compare zero scores across languages, this does not resolve the more fundamental challenge of ensuring that children are assessed in the correct language of instruction.

Data access: Updating databases from regional organisations such as MICS, PASEC, and the PAL Network is relatively straightforward. However, accessing EGRA datasets remains more complex, as they are typically managed at the country level. Similarly, stand-alone national assessments will need to be sourced on a case-by-case basis.

To move this work forward, two immediate priorities are clear:

Integrate upcoming assessments: Prepare to incorporate the ICAR results expected in late 2025 and plan for the expanded set of PASEC countries from 2026, ensuring these datasets report zero-word categories.

Strengthen data access agreements: Explore mechanisms for securing systematic access to EGRA and national one-on-one assessments, to avoid case-by-case negotiations that could delay indicator development.

Taken together, these actions would create the foundation for a robust, scalable zero-word indicator that complements existing measures of learning poverty and provides policymakers with an early, easy-to-interpret signal of progress in foundational learning.

References

Gustafsson, M. (2025). National primary-level learning proficiency statistics from Africa. A stocktaking and hybrid harmonisation

UNESCO Institute for Statistics (2021). Assessing learning proficiency levels and trends for Sustainable Development Goal 4.1: A focus on Africa. Montreal.

Appendix

Figure A1. Age distribution of children enrolled in grade 2 – MICS FLS

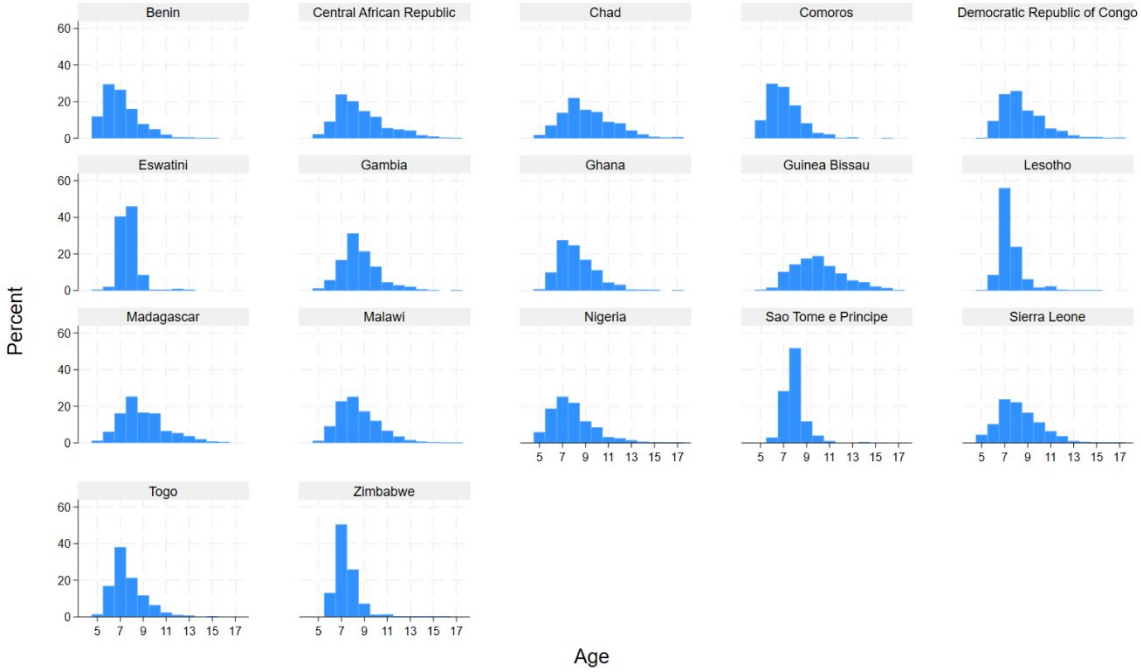
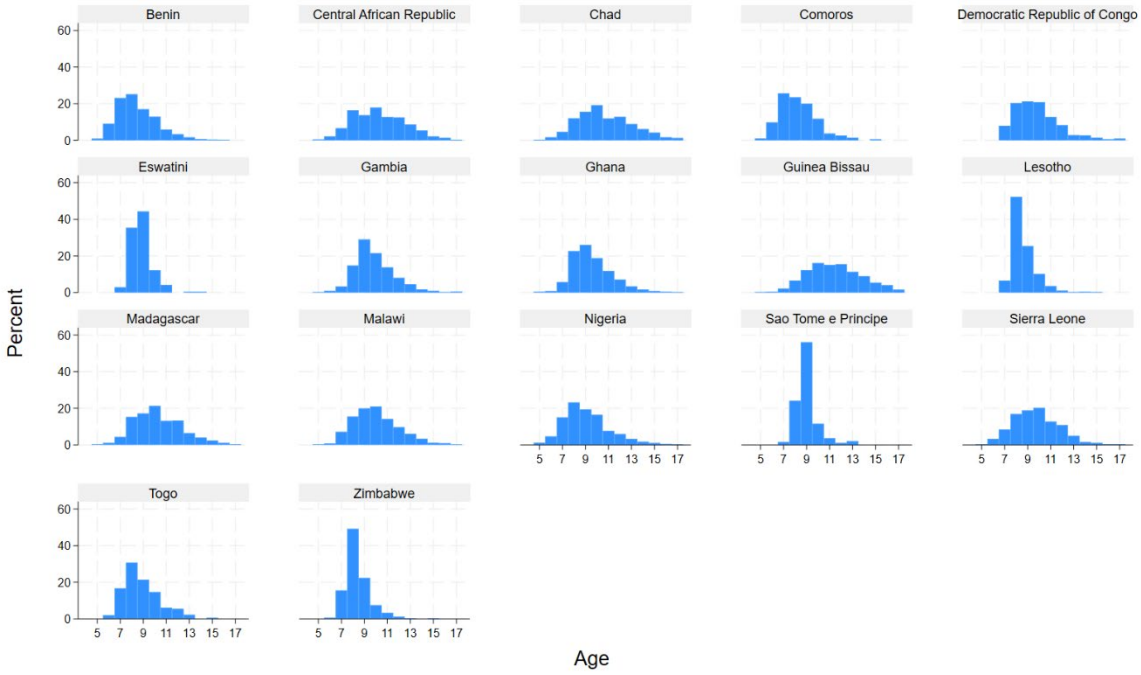


Figure A2. Age distribution of children enrolled in grade 3 – MICS FLS



African Foundational Learning Data Hub
DataFirst
University of Cape Town
www.datafirst.uct.ac.za/aflearn