Setting and Using Benchmarks for Reading Performance

Why have benchmarks?

The focus in international development has shifted from increasing access to school to making sure that students are acquiring tangible skills and knowledge as a result of going to school. Most importantly, emphasis is being increasingly placed on ensuring that students learn to read in the first few years of primary school. It is clearly important for students to develop the ability to read with comprehension early in their schooling, so that they can use text as a vehicle for learning. They must learn to read in order to read to learn. A question that all education systems face then becomes, what is an appropriate standard for literacy? How good is good enough when it comes to students' ability to read?

To answer that question, education systems need to define a standard for reading skill. For example, many countries stipulate in their curricula that students should be proficient in reading by grade 2 or 3.[1] But was constitutes proficiency? What is a straightforward way to ascertain whether students are reaching the desired level of reading skill?

The spread of the Early Grade Reading Assessment (EGRA) has yielded a wealth of country-specific evidence on students' literacy skills. Those data were then used to help a dozen countries set meaningful benchmarks for student reading performance.

This brief summarizes the approach used, shares the results of the benchmarking workshops, and draws some lessons from our experience.

A benchmark is intended to provide a relatively easy measure of whether a learner has achieved a desired level of skill or ability. Researchers for the RTI-led Education Data for Decision Making (EdData II) project undertook to help develop benchmarks for reading, which entailed working with countries to apply data on student performance in reading in the relevant language(s). Our task was to determine the specific skills, and the desired level of performance of those skills, that each country's education leaders could agree would constitute a good measure or demonstration of the curricular goal of reading proficiency.

Language-specific measures of reading fluency

To be able to read with understanding, students need to combine a number of important cognitive and language processes and skills,[2] such as listening comprehension, print awareness, orthographic symbol knowledge, phonological awareness, morphological awareness, and orthographic awareness. Assessing all of these skills at the same time, for every child, can be impractical.

Being able to read fluently and with comprehension means that children are using the above skills to effectively process text. Oral reading fluency is the ability to accurately and rapidly read connected text with expression. It is strongly related to comprehension across languages. However, the exact nature of that relationship does vary depending on the properties of each language.

Graham and van Ginkel (2014) argued persuasively against using a single reading fluency standard across languages.[3] They pointed out how the path along which children develop fluency and comprehension does not follow a linear relationship between increased "speed" of reading and increased comprehension. That is, students reading at slower rates of fluency may indeed have higher than expected levels of comprehension.[4] And conversely, they showed that students with high rates of fluency could have low levels of comprehension, especially in languages with very transparent orthographies. Therefore, setting benchmarks in reading should be done for each language separately, using data on students' reading performance in each language.

In supporting the development of benchmarks, we carefully considered which measure could best serve the purpose of being a reliable indicator of students' ability to read and comprehend text, without being too onerous in terms of the data collection requirements associated with regular tracking of performance against those benchmarks. Fluency is an important skill in its own right, is a reasonably reliable proxy for comprehension, and is inherently easy for all stakeholders to understand. Reading fluency therefore is an easy-to-use indicator of students' levels of reading proficiency.

As discussed below, some countries that participated in a benchmarking exercise set a single benchmark for minimum proficiency—i.e., the level of reading fluency, measured in correct words per minute (cwpm), at which comprehension appears to be at an acceptable level. Others defined different degrees of proficiency. In all cases, the analysis used to align fluency and comprehension took into account the linguistic and orthographic characteristics of each language that would influence the relationship between those two skills.

Experience in 12 countries

USAID's EdData II project assisted 12 countries between 2014 and 2016 to develop language-specific benchmarks. We used data available from Early Grade Reading Assessments (EGRAs) in those languages in the respective countries. Table 1 shows the benchmarks developed for each language in each country. The percentages of students meeting the benchmarks reported in Table 1 for each country are based on available data. They do not represent an official record

Table 1. Benchmarks for reading proficiency in selected countries

Country	Language	Fluency benchmark (in cwpm)[a]	Percentage of students meeting the benchmark[a]
Egypt	Arabic	50	11
Ethiopia	Afaan Oromo	48	5
	Af Somali	50	14
	Amharic	50	6
	Hadiyyisa	40	4
	Sidamu Afoo	45	1
	Tigrinya	55	<1
	Wolayttatto	43	8
Ghana	Ghanaian languages [b]	40	3
	English	45	7
Jordan	Arabic	46	3
Kenya	Kiswahili	45	NA
	English	65	NA
Liberia	English	35–40	4
Malawi	Chichewa	40	<1
Pakistan	Urdu	60–90	20
	Sindhi	50–80	24
Philippines	Ilokano	40	35
	Hiligaynon	45	34
	Cebuano	42	54
	Maguindinaoan	40	22
Tanzania	Kiswahili	50	5
West Bank	Arabic (w/ diacritics)	30	18
	Arabic (w/o diacritics)	35	27
Zambia	Zambian languages[c]	45	1

NA = Not available.

a Either grade 2 or grade 3, and from 2012 through 2014 depending on availability of EGRA data.

b Averaged results for Akuapem Twi, Asanti Twi, Dagaare, Dagbani, Dangme, Ewe, Fante, Ga, Gonja, Kasem, and Nzema.

c Averaged results for Chitonga, Cinyanja, Icibemba, Kiikaonde, Lunda, Luvale, and Silozi.

of reading performance as reported or endorsed by the education ministry in any country.

Table 1 shows the benchmark that was set for "proficient" reading (or reading with comprehension) in each language in either grade 2 or 3. Ghana and Zambia stand out as exceptions. In those countries, a single benchmark for reading fluency was set for multiple languages. Students' reading performance was extremely low across all the indigenous languages of instruction in those countries; thus, the numbers of students with high levels of fluency and comprehension were too few to allow statistically valid estimates of a reasonable standard of proficiency. Therefore, a cross-language sample was used to set the benchmark. Benchmarks in those two countries should be revisited and language-specific ones developed as additional data and more robust samples become available. Not shown in Table 1 are the other levels of reading ability for which benchmarks were set in some countries, in particular in Ethiopia, Ghana, Pakistan, and Zambia.

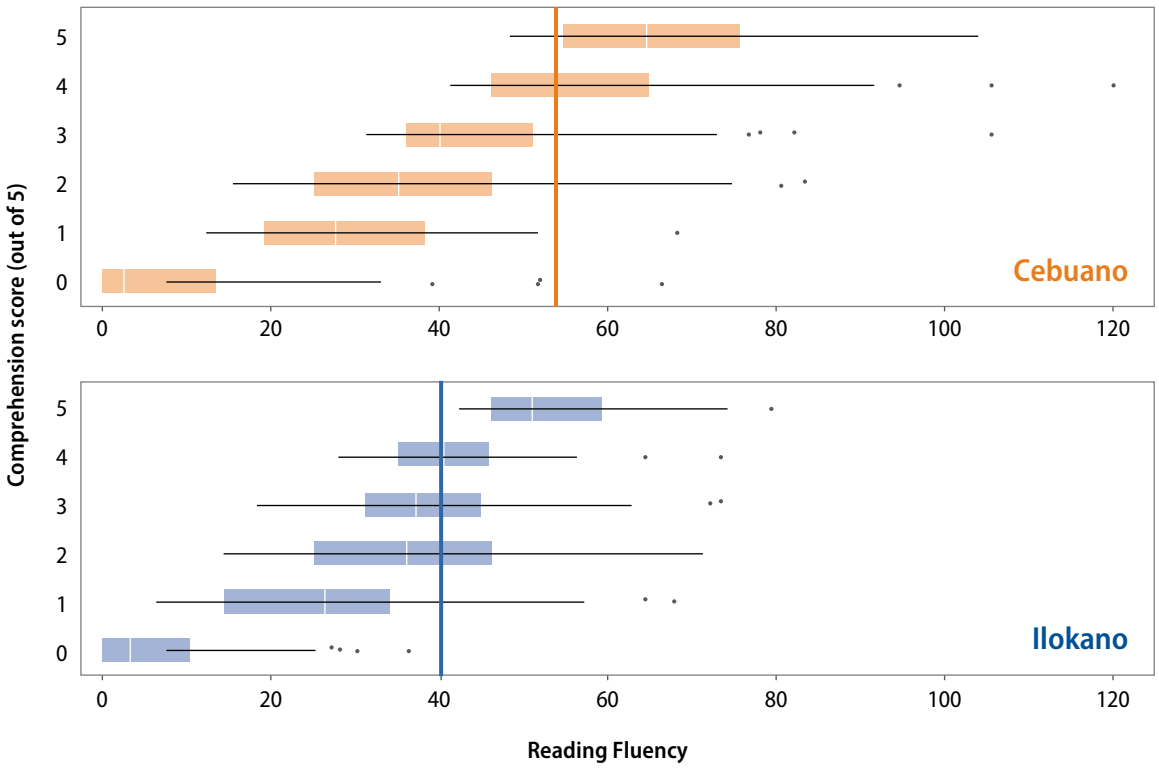
The steps taken to define the different reading proficiency levels are discussed below.

How were benchmarks set?

Through participatory workshops, we helped the aforementioned 12 countries set benchmarks for several skill areas, sometimes including letter sound recognition, decoding (nonword reading), word reading, fluency, and comprehension. However, when governments are monitoring *system* performance (as opposed to diagnosing students' specific areas of skill development), they can use reading fluency as a proxy indicator of reading ability. But what level of fluency can be said to indicate proficient reading—i.e., reading with understanding?

The approach we took to answer that question relied on the relationship between fluency and comprehension in each language. Figure 1 shows how this relationship was represented for two of the languages in the Philippines. The "box-and-whisker plots" summarize grade 2 student performance in Cebuano and Ilokano. Each row plotted vertically (along the *y*-axis) for each language corresponds to a level of reading comprehension, ranging from 0 to 5 questions answered correctly (out of 5). The horizontal axis shows reading fluency. Each box is based on the reading fluency scores at the 25th, 50th, and 75th percentile for each level of reading comprehension.

Figure 1. Box-and-whisker plots summarizing the distributions of reading fluency scores corresponding to different levels of comprehension in grade 2 for two Philippine languages



In every country where EdData II supported work on setting benchmarks, 4 answers correct out of 5, or 80% comprehension, was taken as the standard for students demonstrating an acceptable level of understanding what they read.

For example, the 25th percentile of students who answered 4 out 5 correctly in Cebuano had a fluency rate of 47 cwpm (the left edge of the box). The 50th percentile (the orange line) had a rate of 54 cwpm and the 75th (right edge of the box) had a rate of 65 cwpm. The figure shows that for Ilokano, students achieving 4 out of 5 correct in comprehension were reading at fluency levels lower than their peers reading in Cebuano: the 50th percentile of students scoring 4 out of 5 is at 40 cwpm in Ilokano (the blue line).

The workshop attendees discussed the differences in the point at which reading fluency was associated with at least 80% comprehension, and offered explanations based on differences in the characteristics of the two languages. For example, Department of Education experts explained that Ilokano is an agglutinating language in which "words" can contain several units of meaning (e.g., subject, predicate, and object). So students can glean more meaning from text, even though they may be technically reading fewer words correctly per minute as measured by an EGRA subtask.

For every language for which we had reading data, there was a range of fluency scores (the boxes in the box-and-whisker plots) that corresponded to the desired level of comprehension. Therefore, the stakeholders at the workshops had to discuss what level of fluency should be set as a benchmark to show that students were reading with comprehension. In almost all cases, this range of possibilities led to vigorous discussion among the stakeholders as to what the benchmark should be—and ultimately resulted in greater ownership of the benchmark they finally decided to put forward.

In addition to determining a language-specific benchmark for reading proficiency, some countries and some other USAID programs were interested in tracking performance across levels of reading skill development. Three countries took different approaches to defining levels of reading ability. All cases shared one common classification: *Nonreaders* were defined as students scoring zero on the oral reading fluency subtask of the EGRA.

In **Ethiopia**, data were used to establish grade-specific cutoff points (distinct for each language) for three classifications of reading ability above zero:

- **Reading fluently with full comprehension** – students achieving the level of reading fluency that corresponds with full or almost full comprehension;
- **Reading with increasing fluency and comprehension** – students who have some reading fluency, but have not yet reached the abovementioned levels of fluency and comprehension;
- **Reading slowly and with limited comprehension** – students scoring above zero, but at the lower end of the reading fluency score distribution.

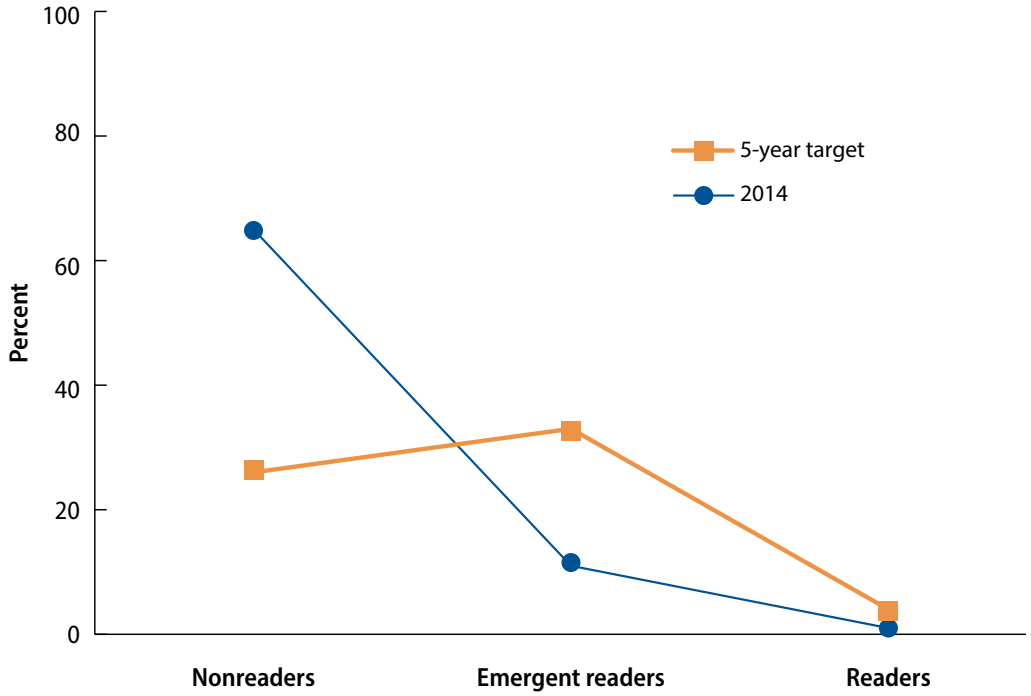
In **Pakistan**, three levels of reading ability were also defined for each grade level for the two languages studied. Based on their level of reading fluency, students would fall into one of three categories, as shown in Table 2. Oral reading fluency cutoff points for each category (measured in cwpm) are also shown for the two languages.

Zambia employed a different approach, defining only two classifications other than nonreading: emergent readers and readers. Students achieving oral reading fluency of at least 20 cwpm and at least 40% comprehension (2 out of 5 questions correct) were defined as emergent readers. Those reaching at least 45 cwpm and 80% comprehension were defined as readers.

Table 2. Reading ability levels in Pakistan for two languages, grades 1–3 (in cwpm)

	Grade 1		Grade 2		Grade 3	
Level of reading ability	**Urdu**	**Sindhi**	**Urdu**	**Sindhi**	**Urdu**	**Sindhi**
Does not meet expectation	< 30	< 30	< 60	< 50	< 70	< 60
Meets expectation	30 to 60	30 to 50	60 to 90	50 to 80	70 to 100	60 to 90
Exceeds expectation	> 60	> 50	> 90	> 80	> 100	> 90

Figure 2: Distribution of students in reading fluency categories, Zambia example



In each of these cases, benchmarks defined in relationship to different levels of reading ability will allow interested parties to track improvement below proficient reading. For example, if in Ethiopia many, if not most, students are not reaching the benchmark for reading fluently with full comprehension, over time it will be possible to see whether diminishing percentages of students are, say, scoring zero, while increasing percentages are falling into the ranges defined as reading slowly with limited comprehension or reading with increasing fluency and comprehension.

In Zambia when the benchmark-setting workshop took place, a 2014 EGRA had revealed the distribution of students depicted by the blue (dark) circles in Figure 2. Targets that the attendees set for 5 years later are shown by the orange (light) squares.

Such a change would reflect the percentage of nonreader students decreasing substantially, with a growing proportion of students moving out of that category and some even meeting the cutoff for emergent reading. The shifted distribution line therefore would show progress even if only a very limited percentage of students were meeting the cutoff defined for readers.

What did we learn?

In each country, it was evident that all stakeholders wanted the benchmarks to represent actual reading ability. When we started this work, especially in countries where performance in reading was shown to be extremely low, we anticipated that in-country stakeholders, and even USAID, would push to have low standards. In almost every case, however, the opposite was true. The refrain "reading is reading" was often invoked, meaning that people wanted to know whether children were actually able to read and understand grade-level text, and not just whether they were meeting some arbitrary low benchmark. Another issue that consistently arose in almost every case was that stakeholders wanted benchmarks to be based on students reading grade-level text.[5]

The experience also raised the issue of whether it is prudent to set benchmarks for every skill area that is tested. There is a tendency to want to formulate benchmarks for everything just because data are available for them, and because people are interested in knowing whether students are acquiring all the skills. In future work, it will be important to resist this tendency. Distinctions need to be made among what the curriculum may need to stipulate, what teachers should be able to assess regularly in their classrooms, and what is prudent for system-level monitoring. Having a single measure that

represents a higher order skill—one that is based on students assembling and applying the subskills—is best for system monitoring. Reading fluency is still a good option for that kind of measure. Comprehension may be equally good, but it is more difficult to measure well. And for situations where performance in general is very low, then the approach taken in Ethiopia, Pakistan, and Zambia is a useful way to capture changes in performance below the threshold for proficiency.

Conclusion

Our experience revealed a strong appetite in every participating country for setting benchmarks. Stakeholders willingly engaged the process, worked through the data, and crafted meaningful benchmarks. What was less evident was how those benchmarks could then become official. Only in Kenya, and to some extent in Pakistan, were the benchmarks adopted as official policy for measuring performance in the education system.

And finally, setting benchmarks involves both science and art: science in terms of using data, but art in terms of a process that stakeholders can buy into. The pursuit of sound science (e.g., ever more precise measures or more precisely determined benchmark cutoff points) should not come at the expense of the art of building stakeholder ownership and commitment to monitoring system performance based on tangible learning outcomes.

Endnotes

1 Sustainable Development Goal 4, as adopted by the United Nations General Assembly, focuses on education. It stipulates that boys and girls should complete high-quality primary and secondary education that leads to effective learning. The indicators for tracking progress toward Goal 4 include the percentage of children in grades 2 and 3 achieving at least a minimum proficiency level in reading (and mathematics).

2 A superb summary of the research on reading skill development is now available in Kim, Boyle, Zuilkowski, and Nakamura (2017).

3 Numerous citations for the relationship between oral reading fluency and reading comprehension are available in the USAID *Landscape Report*, Kim et al. (2017).

4 For EGRA implementers, using both timed and untimed versions of the text-reading subtask would reveal more nuances regarding the relationship between reading fluency and reading comprehension. This would be an especially prudent approach in orthographically complex languages.

5 The emphasis on grade-level text is significant, because standard practice in developing and implementing most EGRAs is to use the same text across grades in a given country and language, primarily for easily comparing performance. When it comes to setting standards for what students should be able to read and how well they should read it, however, using text appropriate for each grade becomes more important.

References

Graham, B., & van Ginkel, A. J. (2014). Assessing early grade reading: The value and limits of 'words per minute'. *Language, Culture and Curriculum* [Online journal], 244–259. https://dx.doi/10.1080/07908318.2014.946043

Kim, Y.-S. G., Boyle, H. N., Zuilkowski, S. S., & Nakamura, P. (2017). *Landscape report on early grade literacy.* Washington, DC: USAID. https://globalreadingnetwork.net/publications-and-research/landscape-report-early-grade-literacy-skills

More Information

Joseph DeStefano
+1 216 371 3834
jdestefano@rti.org

Follow @RTI_EdWork
Follow @RTI_Intl_Dev
http://SharEd.rti.org

RTI International
3040 E. Cornwallis Road, PO Box 12194
Research Triangle Park, NC 27709-2194 USA

RTI 10746 0217