Fluency Benchmarking at Room to Read

2008
Began piloting reading instruction intervention

2011
Began rigorously evaluating reading instruction intervention using EGRA adaptation

2015-16
Launched small-scale research studies across countries to determine language-specific fluency benchmarks

2016-17
Received grant from Pearson Education to further explore fluency benchmarking

- Desire to set targets for learning and accountability.
- Belief that 45-60 correct words per minute may not be universal.
Framing Research Questions for Fluency Benchmarking

• To what degree are fluency benchmarks useful indicators of comprehension? If so, at which stages of reading development?
• How do different methods of estimating a benchmark compare?
• What factors determine the level and precision of a benchmark estimate?
• Does the recommended approach to benchmarking differ by language?
• What language characteristics affect fluency rates or the fluency-comprehension relationship?
• How do the properties of the assessment instrument influence this work?
• What data collection and data analysis approaches best facilitate accurate benchmark estimates?
Matthew presents
Fluency Benchmarking Next Steps at Room to Read

- Continue to collect **EGRA data to support fluency benchmarking**
- Improve and evolve our traditional passage/questions **comprehension measures** across languages and expand use of **sentence choice** comprehension measure.
- Work with our Country Offices to set **language-specific benchmarks for Room to Read** goal setting.
- Develop **reading measurement research agenda**, frameworks for **collaboration**.
- **Collaborate with partners** at country and global level on methods and benchmark setting.
Call for more language-focused analyses:

Further explore the relationship between fluency and comprehension across languages and language characteristics (e.g., strength of the relationship across languages, variation based on linguistic characteristics).

Examine other linguistic characteristics beyond depth (such as visual form, parametric complexity, visual complexity, etc.) to understand if those factors influence a) the strength of the fluency/comprehension relationship and b) benchmark levels.

Further explore the relationship between L1 and L2 and benchmarking, including the role oral language skills play in fluency/comprehension results.
Call for more longitudinal studies:
Track children as they progress from Grade 1 to Grade 3 or 4 to understand how fluency progresses in children over time.
Identify children who have high comprehension in Grades 3 and 4 and trace their fluency levels in Grade 1 and 2.

Comprehension measurement research:
Administer the EGRA ORF sub-task with 1 minute and 3 minute limits with the same students to see how fluency rates differ based on time allowed to read the passage.
Assess reliability, leveling and adaptation of the sentence choice test.
### Checklist of data-driven methods to set fluency benchmarks.

#### STEP 1: Define the goals of benchmarking
- Define “reading proficiency” for the given context
- Identify how benchmarks will be used

#### STEP 2: Define the population
- Set separate benchmarks for each language of assessment

#### STEP 3: Select the Instruments
- Assess comprehension independent of fluency
- Consider the sentence choice test as an alternative to five questions about a passage
- Pilot comprehension questions, measure reliability, and set high threshold for reliability
- Allow students 3 minutes to read to the passage when assessing comprehension
- Use more than one passage to set benchmarks
- Assess the difficulty of passages with a readability formula

#### STEP 4: Select a sample
- Select a sample with the same mother tongue and language of instruction as the target population (sample does NOT need to be representative of the target population in other respects)
- Select a sample that is large enough (at least 150 students) for increased precision
- Select a sample with a mix of high- and low-achieving students (roughly 20%–50% who can read with 80% comprehension)
- Use more than one sample (i.e., more than one EGRA administration) to set benchmarks