USAID
FROM THE AMERICAN PEOPLE

Setting Reading Benchmarks: Evidence from India

USAID All Children Reading-Asia
USAID/India

Jonathan Stern, RTI
April 2019

Important distinctions

Goal is an aspiration for the future, maybe without numerical value

Metric is a valid, reliable unit of measurement

Benchmark is numerical representation of the goal using the metric

Target is the number of children reaching the benchmark in a given time period

Goal: All children in Grade 2 should read fluently and with comprehension

Metric: "correct words per minute in passage reading"

Benchmark: 45 correct words per minute, understand 80% of what they read

Target: 60% of Grade 2 children achieving the benchmark in 3 years.

USAID
FROM THE AMERICAN PEOPLE

All Children Reading – Asia (ACR – Asia)

EGRA Benchmarks and Standards Research Report

Submission Date: December 21, 2017

AID-OAA-TO-16-00017 Number: REQ-ASIA-16-00017
Activity Start Date and End Date: September 30, 2016, to September 29, 2021
TOCOR: Mitch Kirby

Submitted by: **RTI International**
3040 Cornwallis Road
Research Triangle Park, NC 27709-0155
Tel: (919) 541-6000

The benchmarking process

1. Clearly define aims and scope
2. Obtain relevant data
3. Convene participatory workshop with representation from a range of stakeholder groups
4. Review data (in light of curriculum, context, and language)
5. Set appropriate and achievable benchmarks and targets
6. Institutionalization

1. Aims and Scope

First large-scale benchmarking activity across languages in India

1. Include the reading assessment results of USAID India's reading projects in the Global Count

2. Report Indicator E.S.1-1 for five projects: "Percent of learners who demonstrate reading fluency and comprehension of grade level text at the end of grade 2 with USG assistance."

2. Relevant Data

- **Assessment:** Combined EGRA/ASER instrument
- **Student sample:** ~14,467 Standard 2 students at baseline (across five projects)
- **Geographic locations:** Seven states
    1. Uttarakhand
    2. Chhattisgarh
    3. Maharashtra
    4. Uttar Pradesh
    5. Odisha
    6. Karnataka
    7. Rajasthan
- **Language(s)**: Intervention language of instruction was used for the assessment.
    - Hindi, Marathi, Oriya, Kannada, and English

3. Participatory Workshop

- 43 participants from 20 organizations
- Set reading benchmarks across all five languages
- Innovations based on EGRA Benchmarks and Standards Report
    1. Multiple data sources used (EGRA and ASER subtasks)
    2. No a priori assumption about which variable would be used for benchmarking
        a) Participants were provided with analyses and relationships across all administered subtasks (EGRA and ASER)
    3. Multiple methods were presented instead of reliance on traditional "mean" method

4. Review Data

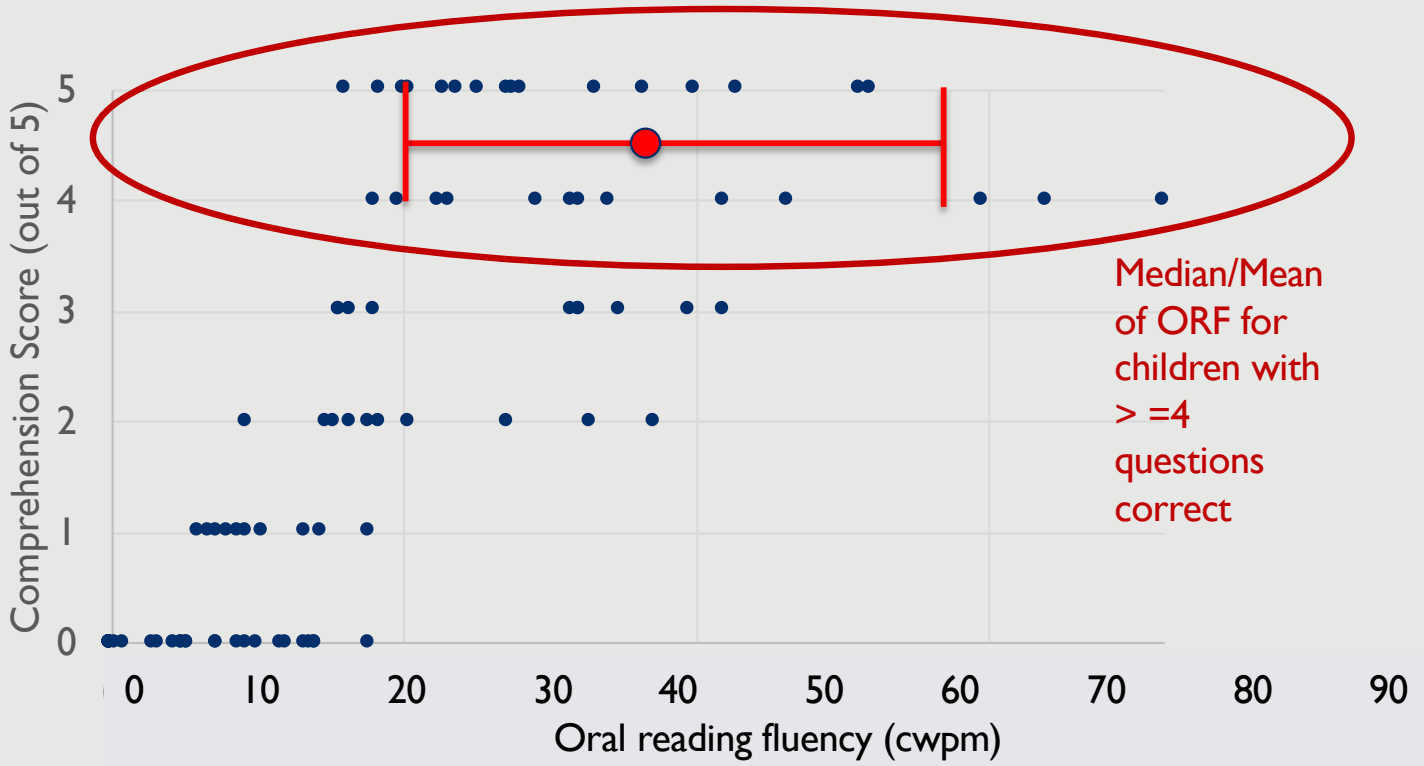
- Which skill(s) should be benchmarked?
  - Reading comprehension
  - Oral reading fluency
  - ASER Std 2 level
  - ASER Std 1 level
- Direct or indirect measurement?

- Oral reading fluency as indirect measure of comprehension
- Reading comprehension is difficult to assess reliably
- Oral reading fluency is often referred to as bridge between decoding and comprehension

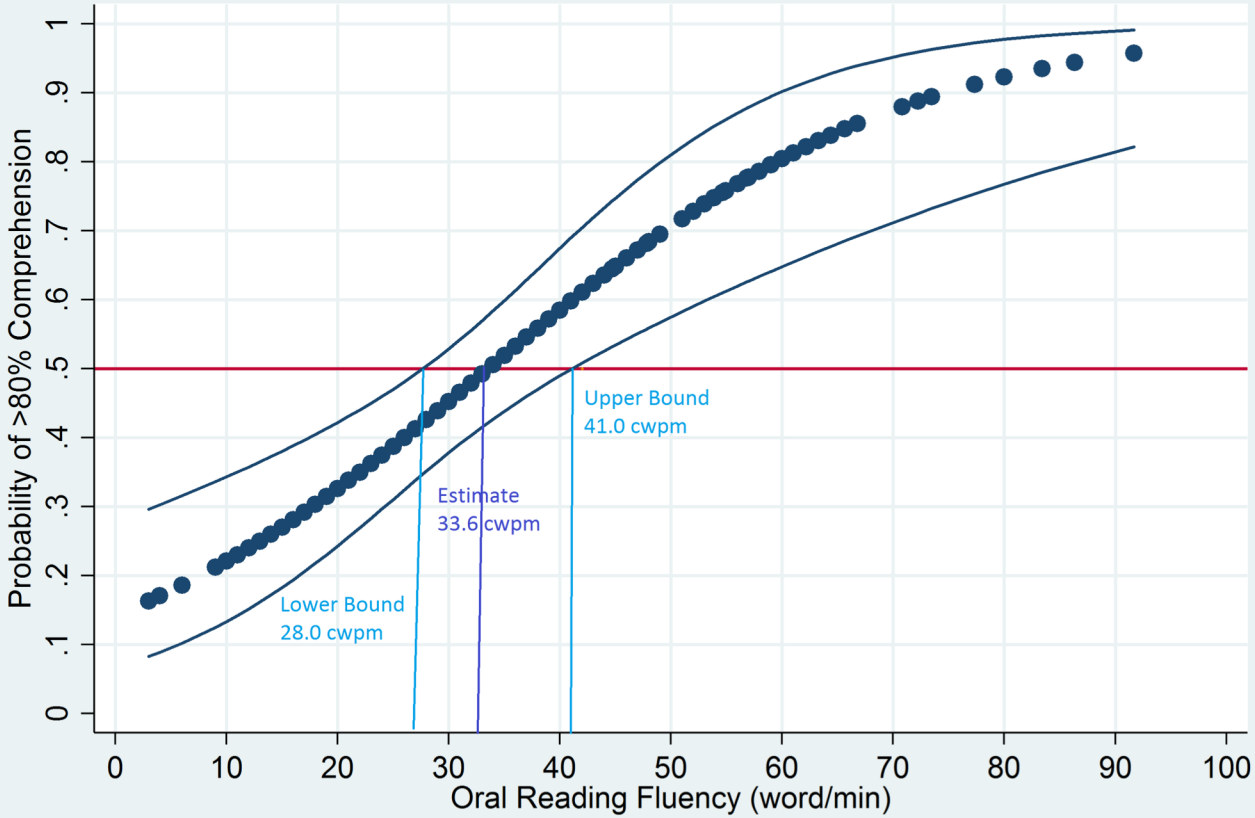
4. Review Data (continued)

- Which method(s) should be used?
  - Normative
  - Mean/median
  - Logistic regression
  - All of the above
- What outcome is most appropriate?
  - Reading comprehension (overall)
  - Reading comprehension (of attempted)
  - ASER Std 2 level
  - ASER Std 1 level
- Should there be one benchmark for all languages or a separate benchmark for each?

The Mean/Median Method



Logistic Regression Method

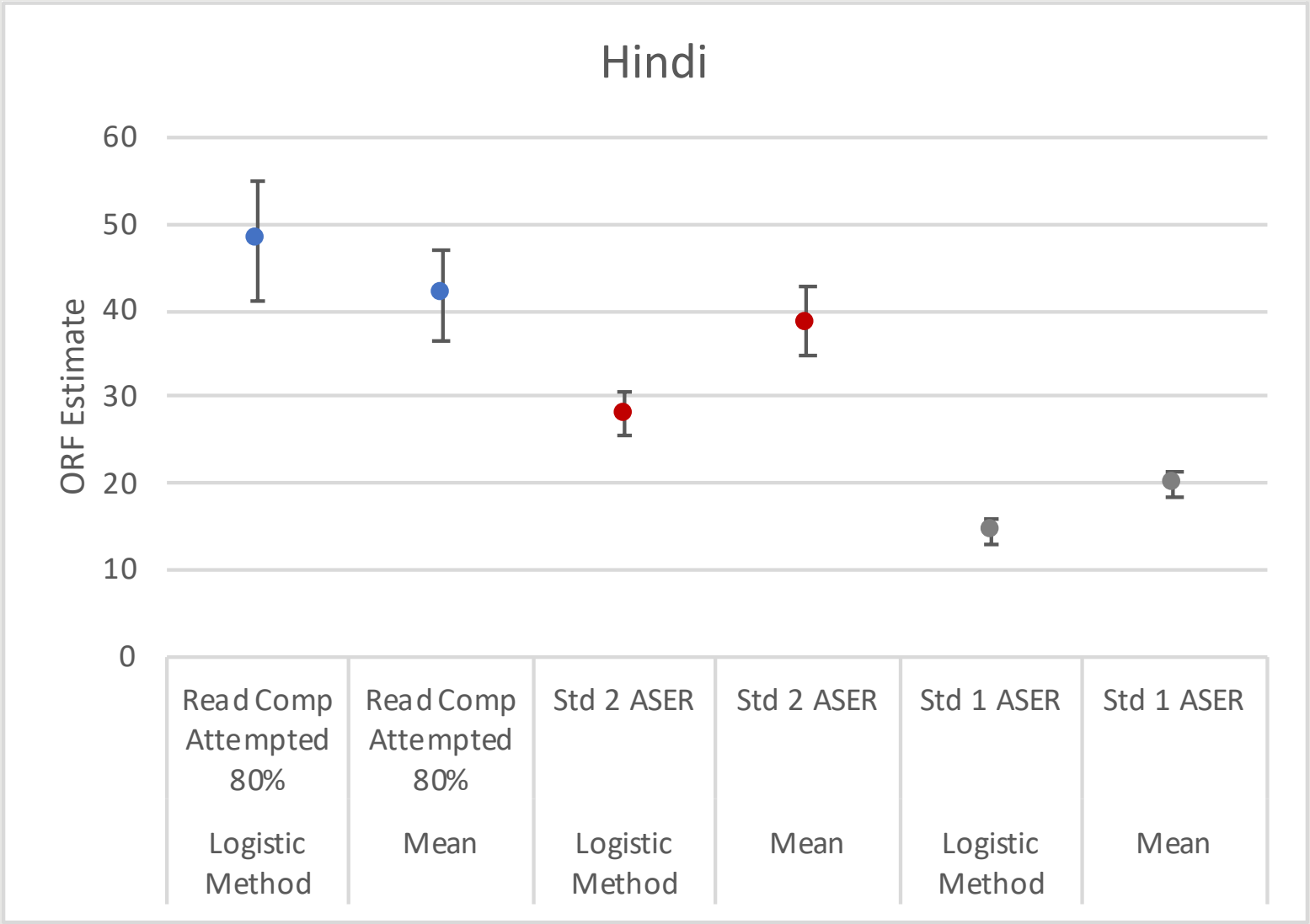


Data from Nepal, Grade 2

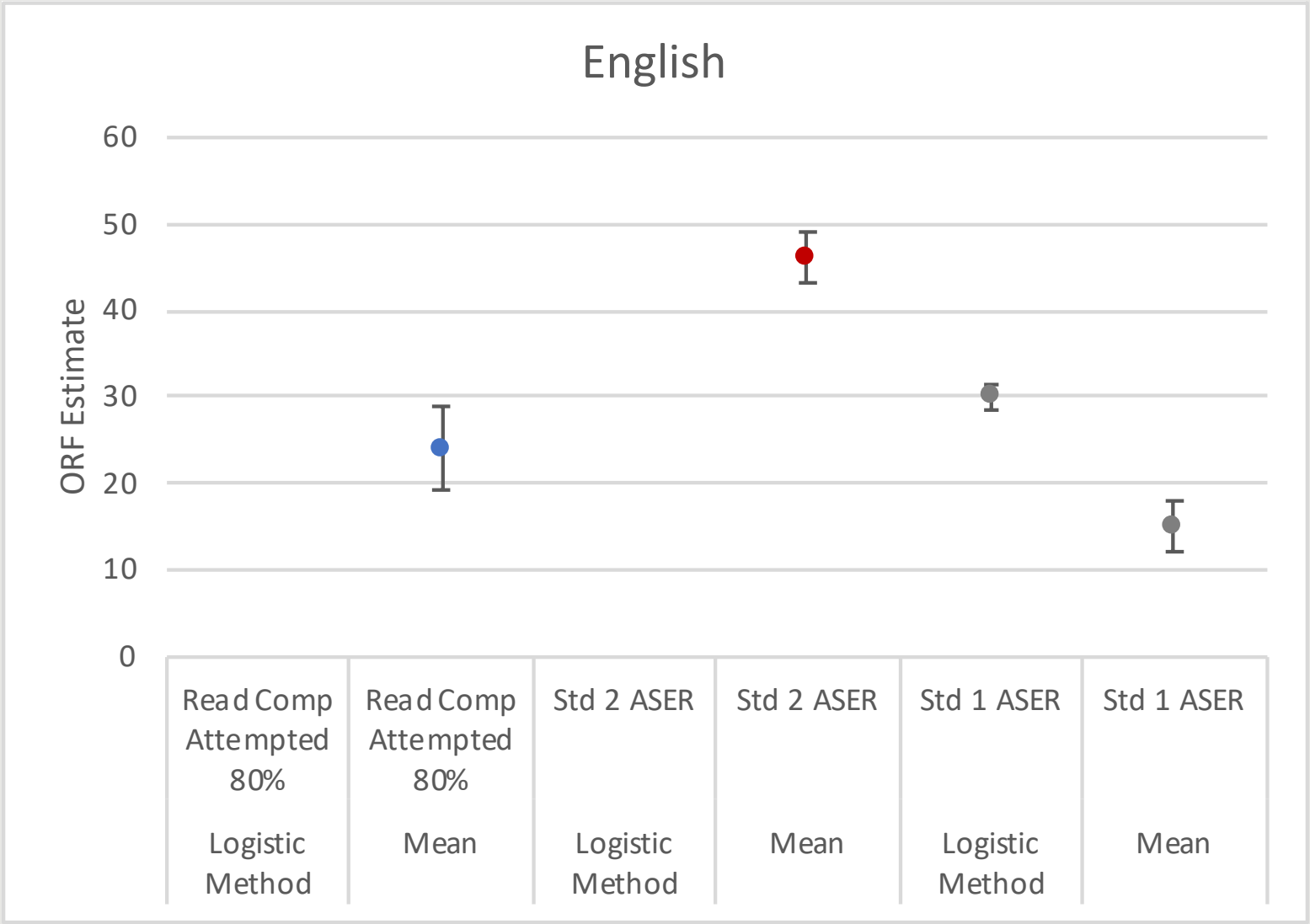
Fluency at probability of 0.5

Upper Bound – 41.0 cwpm
Lower Bound – 28.0 cwpm

Confidence interval – 13 cwpm


Hindi
ORF Estimate
60
50
40
30
20
10
0
Read Comp Attempted 80%
Logistic Method
Read Comp Attempted 80%
Mean
Std 2 ASER
Logistic Method
Std 2 ASER
Mean
Std 1 ASER
Logistic Method
Std 1 ASER
Mean

HINDI						
Outcome	Method	Oral Reading Fluency Level	Lower Bound Estimate (95% Confidence)	Upper Bound Estimate (95% Confidence)	R2	Number of Students Reaching Standard
ASER Standard 1 Level	Logistic	15	13	16	0.53	1462
	Mean	20	18	21		
ASER Standard 2 Level	Logistic	28	26	31	0.406	659
	Mean	39	35	43		
Reading Comprehension 80% Overall	Logistic	64	56	76	0.396	138
	Mean	51	42	60		
Reading Comprehension 80% of Attempted	Logistic	48	42	56	0.328	347
	Mean	42	36	47		


English
ORF Estimate
60
50
40
30
20
10
0
Read Comp Attempted 80%
Logistic Method
Read Comp Attempted 80%
Mean
Std 2 ASER
Logistic Method
Std 2 ASER
Mean
Std 1 ASER
Logistic Method
Std 1 ASER
Mean

ENGLISH						
Outcome	Method	Oral Reading Fluency Level	Lower Bound Estimate (95% Confidence)	Upper Bound Estimate (95% Confidence)	R2	Number of Students Reaching Standard
ASER Standard 1 Level	Logistic	30	20	46	0.364	18
	Mean	15	10	20		
ASER Standard 2 Level	Logistic	n/a	n/a	n/a	n/a	1
	Mean	46	46	46		
Reading Comprehension 80% Overall	Logistic	n/a	n/a	n/a	n/a	0
	Mean	n/a	n/a	n/a		
Reading Comprehension 80% of Attempted	Logistic	n/a	n/a	n/a	n/a	1
	Mean	24	24	24		

5. Set Benchmarks

- Which method(s) should be used?
  - Use all available information, as opposed to reliance on a single method
- What outcome is most appropriate?
  - Preferred outcome was reading comprehension (of attempted)
- Should there be one benchmark for all languages or a separate benchmark for each?
  - Create separate benchmarks for each language
- Consensus was reached on benchmarks for all five languages

Limitation and Future Research

- Concerns about Indian language orthography and visual complexity—and relationship between fluency and comprehension.
- Discussion on the impact of mother tongue on benchmarking method
- Limited comprehension measures
  - ASER does not directly assess comprehension
  - EGRA could be improved (e.g. using separate passages to assess fluency and comprehension).

Conclusions

- Overall, participants viewed the activity as a valuable exercise (first of its kind).
- More is better: more data, more methods, more engagement.
- Expected to serve as the starting point for a larger conversation regarding methods and use for early grade reading benchmarks in India.