

# Policy Linking to Set Global Benchmarks and Track Student Progress Over Time

---

Abdullah Ferdous

Principal Psychometrician

American Institutes for Research, International Development Division

[aferdous@air.org](mailto:aferdous@air.org)

Global Alliance to Monitor Learning (GAML), Paris | December 2023

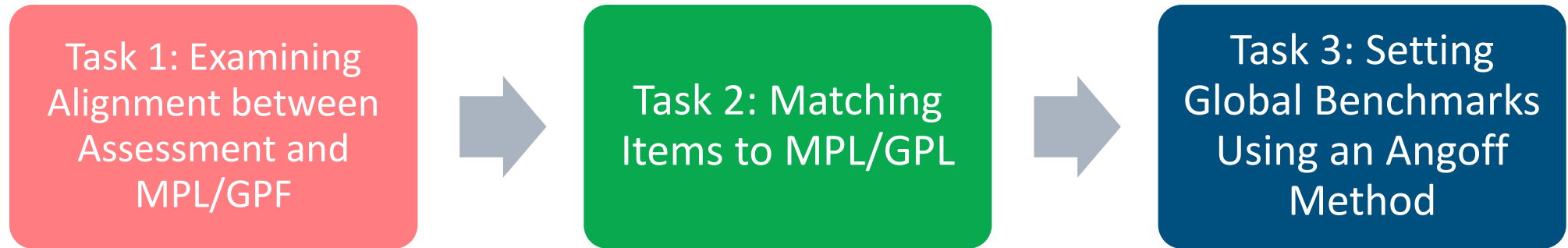
# INTRODUCTION

---

- The policy linking method is a valuable tool for establishing benchmarks on assessments.
- It employs the Angoff-based approach (1971), which has been successfully utilized for over 52 years.
- The Angoff method is one of the most extensively researched and widely adopted benchmarking methods globally. In fact, more than half of the states in the USA use Angoff-based method for their state-wide assessment programs.
- The policy linking method has been applied to at least 30 individually- and 20 group-administered assessments on a global scale.
- The implementation of this method has significantly improved since its initial pilot in 2019.
- This method sets robust and defensible benchmarks for assessments, supported by quality assurance statistics.

# A THREE-TASK PROCESS

---



[Policy Linking for Measuring Global Learning Outcomes Toolkit \(unesco.org\)](https://unesco.org)

# ROUND 1 FEEDBACK: ITEM RATINGS AND BENCHMARKS

---

## Review Round 1 results with the following feedback:

- Location statistics on panelist benchmarks.
- Impact data showing percentage of learners in each MPL/GPL.
- Item ratings in relation to actual item difficulty values (p-values).
- Rater consistency measures
  - Standard Error of Benchmark
  - Inter-Rater Consistency (Ferdous, 2004)
  - Intra-Rater Consistency (Chang, 1999).

# POLICY LINKING: SELF ASSESSMENT STANDARDS

---

- **Criterion 1** – Did all panelists meet the requirements for participation? (YES / NO)
- **Criterion 2** – Were the group of panelists sufficiently representative in terms of the characteristics agreed by the country? (YES / NO)
- **Criterion 3** – Were all outliers removed before calculating the final benchmarks? (YES / NO)
  - Conditional standard error (CSE) of the benchmark should be calculated (without suspected outliers).
  - If the outlier benchmark is higher/lower than the average benchmark plus/minus two times the CSE then the outlier should be removed.
- **Criterion 4** – Were benchmarks only set for GPLS that don't exhibit floor or ceiling effects? (YES / NO)
- **Criterion 5** – Is the inter-rater consistency statistic (Ferdous, 2004; Cohen, 1960) greater than or equal to 0.7? (YES / NO)
- **Criterion 6** – Has the Standard Error for each benchmark been calculated and reviewed to be determined as appropriate? (YES / NO)

# POLICY LINKING: SELF ASSESSMENT STANDARDS

---

- Criterion 7 – Has the confidence interval for each benchmark been calculated and reviewed to be determined as appropriate? (YES / NO)
- Criterion 8 – Was the mean average score for each section of the evaluation greater than or equal to 4? (YES / NO)
- Criterion 9 – Was the mean average score for the overall evaluation greater than or equal to 3? (YES / NO)
- **Criterion 10** – Is the intra-rater consistency (Chang, 1999) statistic greater than or equal to 0.7? (YES / NO)
- **Criterion 11** – Is the actual classifications of test takers agree with those that would be made of their true scores (Livingston and Lewis, 1995) greater than or equal to 0.7? (YES / NO)

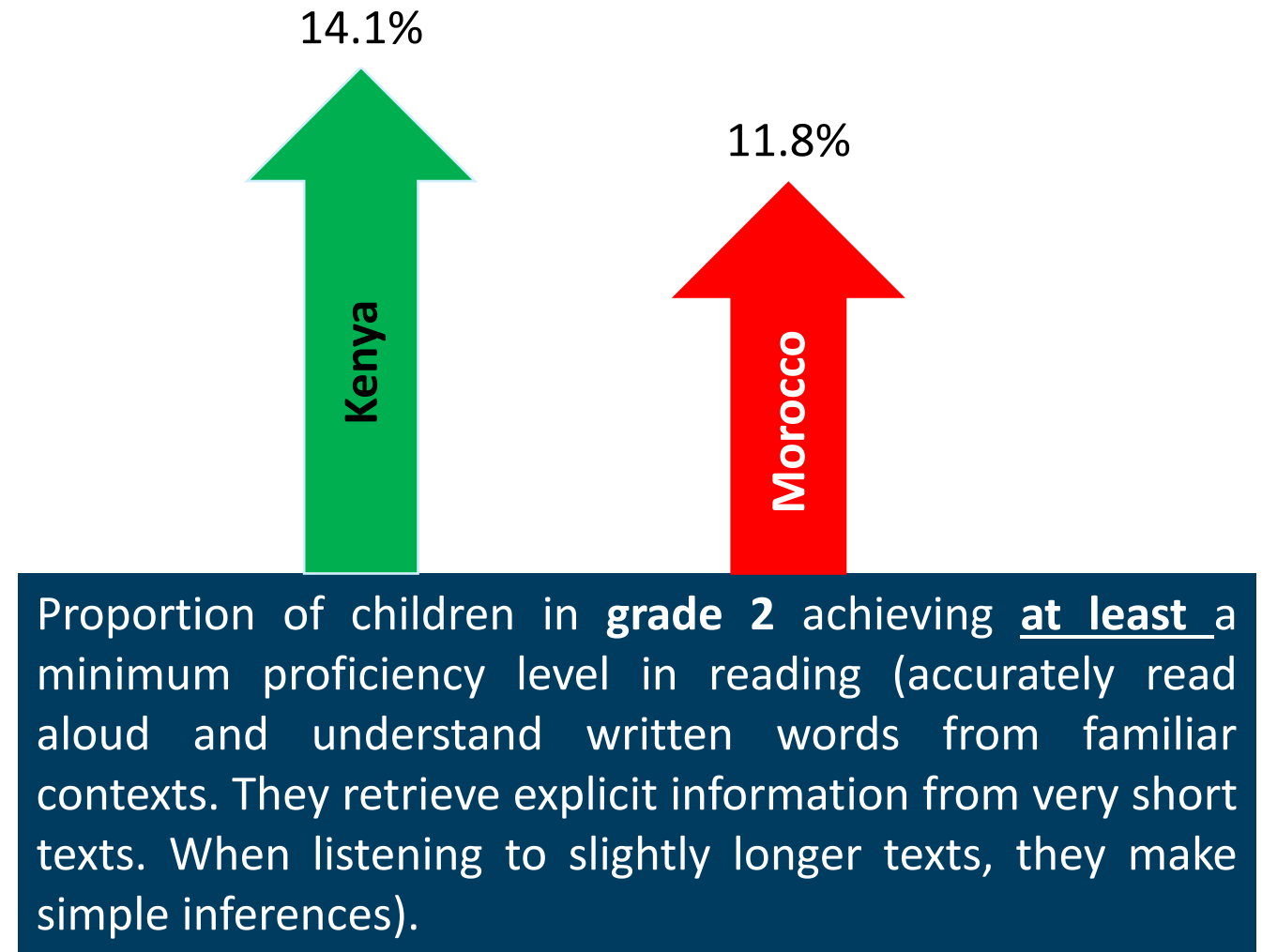
# POLICY LINKING FOR CROSS-COUNTRY COMPARISON

Kenya: Grade 2	Benchmark	% of Students	SDG 4.1.1
Listening Comprehension (5)	4	27.6%	14.1% Students obtained at least 4 in LC, 24 in OR, and 3 in RC
Oral Reading (68)	24	60.5%	
Reading Comprehension (5)	3	28.7%	
TOTAL (78)	31		

Morocco: Grade 2	Benchmark	% of Students	SDG 4.1.1
Listening Comprehension (6)	3	29.2%	11.8% Students obtained at least 3 in LC, 28 in OR, and 3 in RC
Oral Reading (55)	28	36.7%	
Reading Comprehension (6)	3	28.3%	
TOTAL (67)	34		

## POLICY LINKING FOR CROSS-COUNTRY COMPARISON

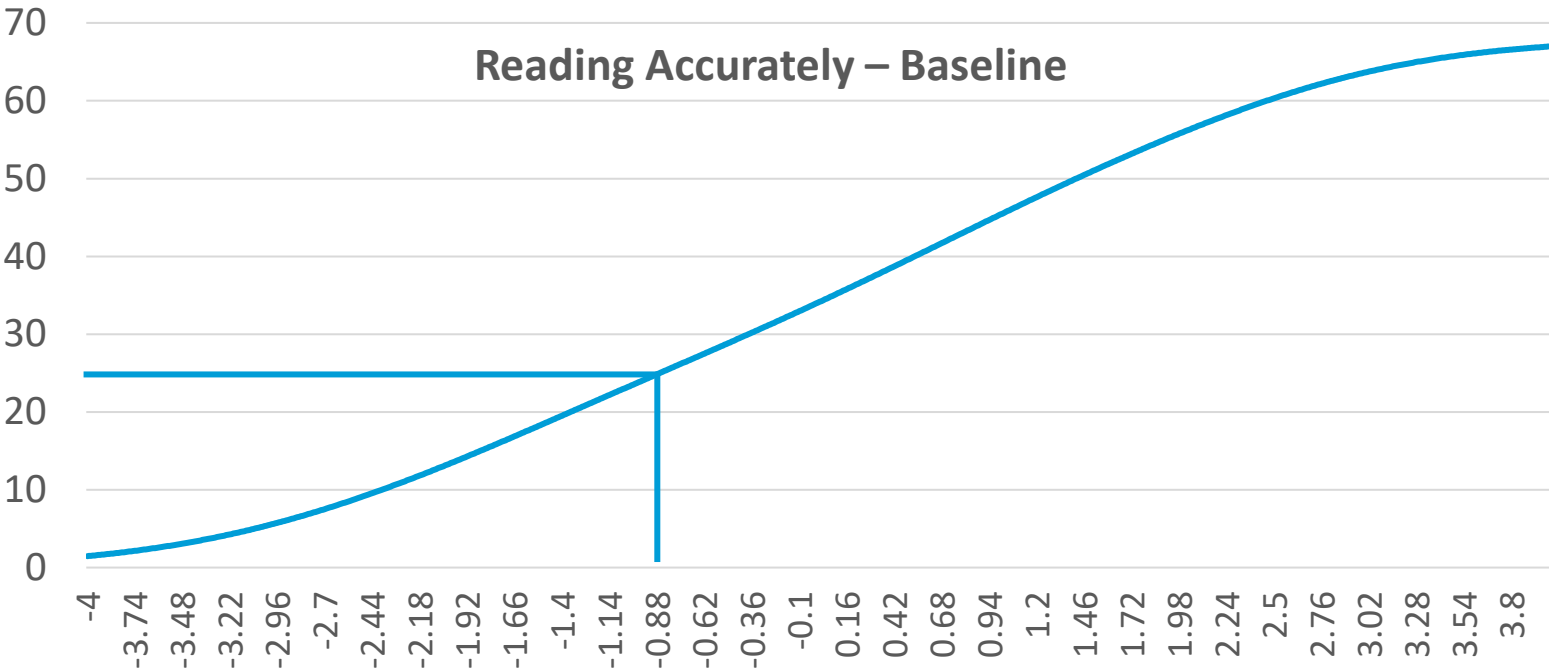
- If assessments are aligned with MPL/GPF, the policy linking method ensure that results from diverse assessments can be compared effectively.
- It will classify students with the same minimum knowledge and skills across countries into the meeting global MPL/GPF, irrespective of the differences in their learning assessments and conditions of their learning.





# POLICY LINKING FOR WITHIN COUNTRY TRACKING

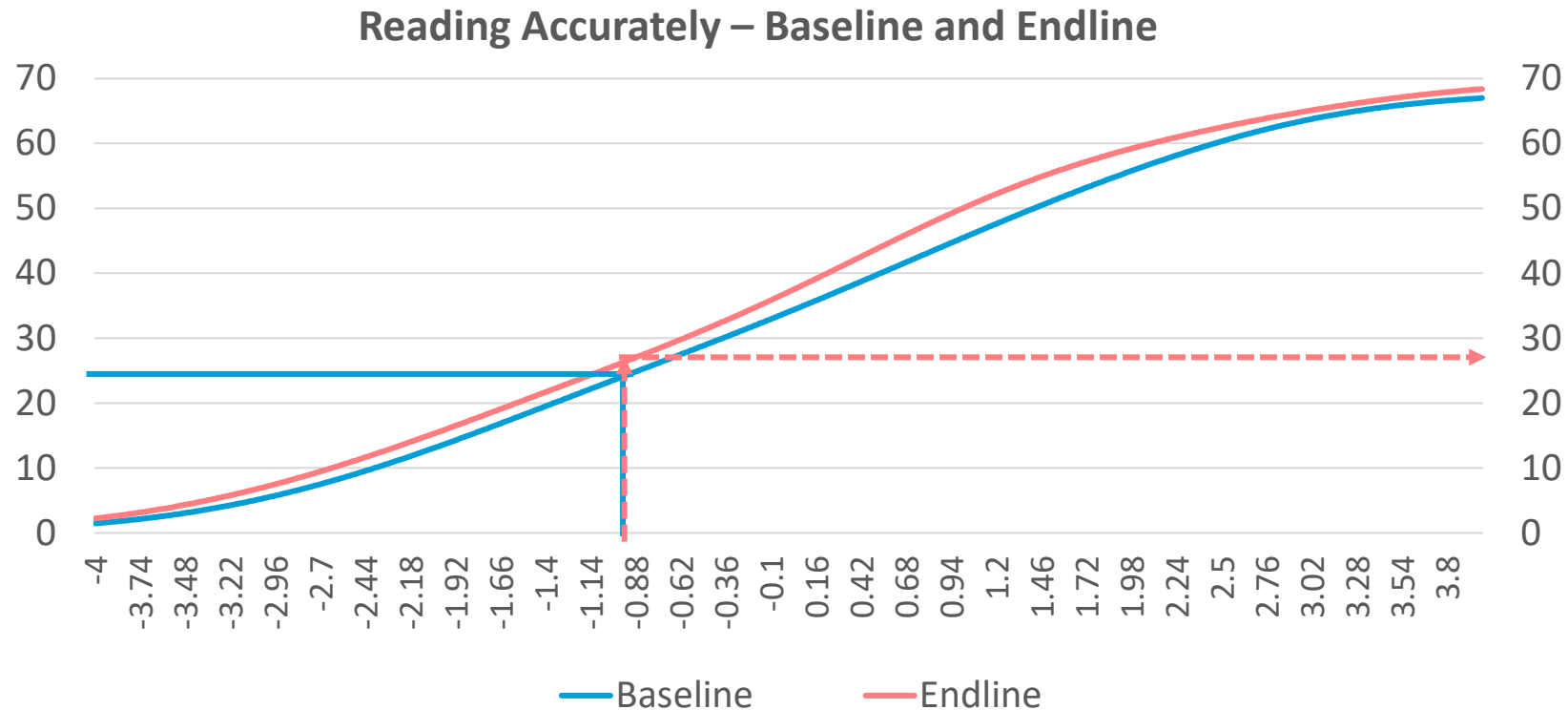
- The baseline reading accuracy benchmark was set at 24 correct words. This benchmark was converted into IRT-based scaled score “theta”.



- The baseline passage had a total of 68 words.
- Benchmark was set at 24 correct words.
- Equivalent theta benchmark is -0.96

# POLICY LINKING FOR WITHIN COUNTRY TRACKING

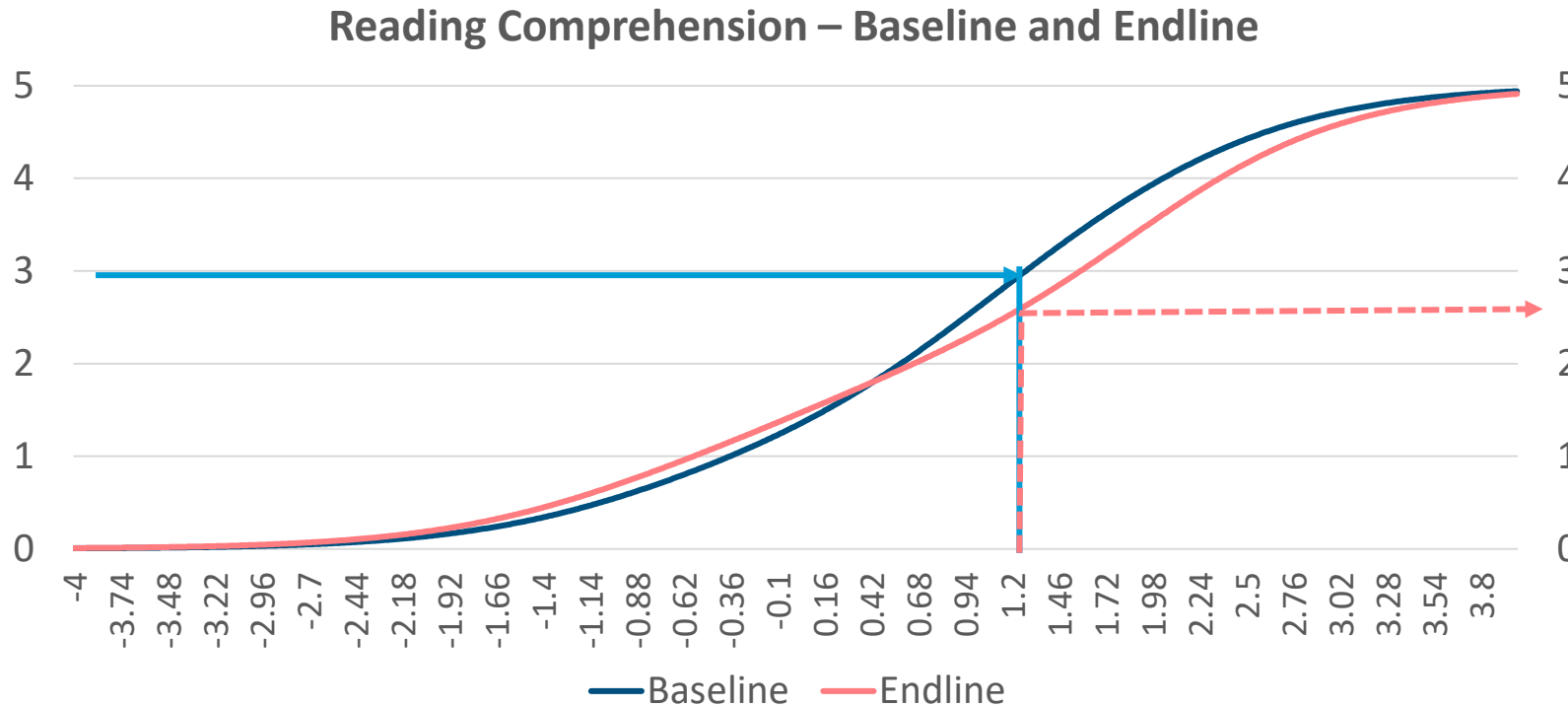
- Common item or common person equating design will help bring baseline and endline scores onto the same scale.



- The endline passage had a total of 70 words.
- Equivalent theta benchmark at baseline was -0.96
- Equivalent benchmark at endline is 26 correct words.
- Learners read 26 correct words at endline will be classified as meeting MPL in decoding.

# POLICY LINKING FOR WITHIN COUNTRY TRACKING

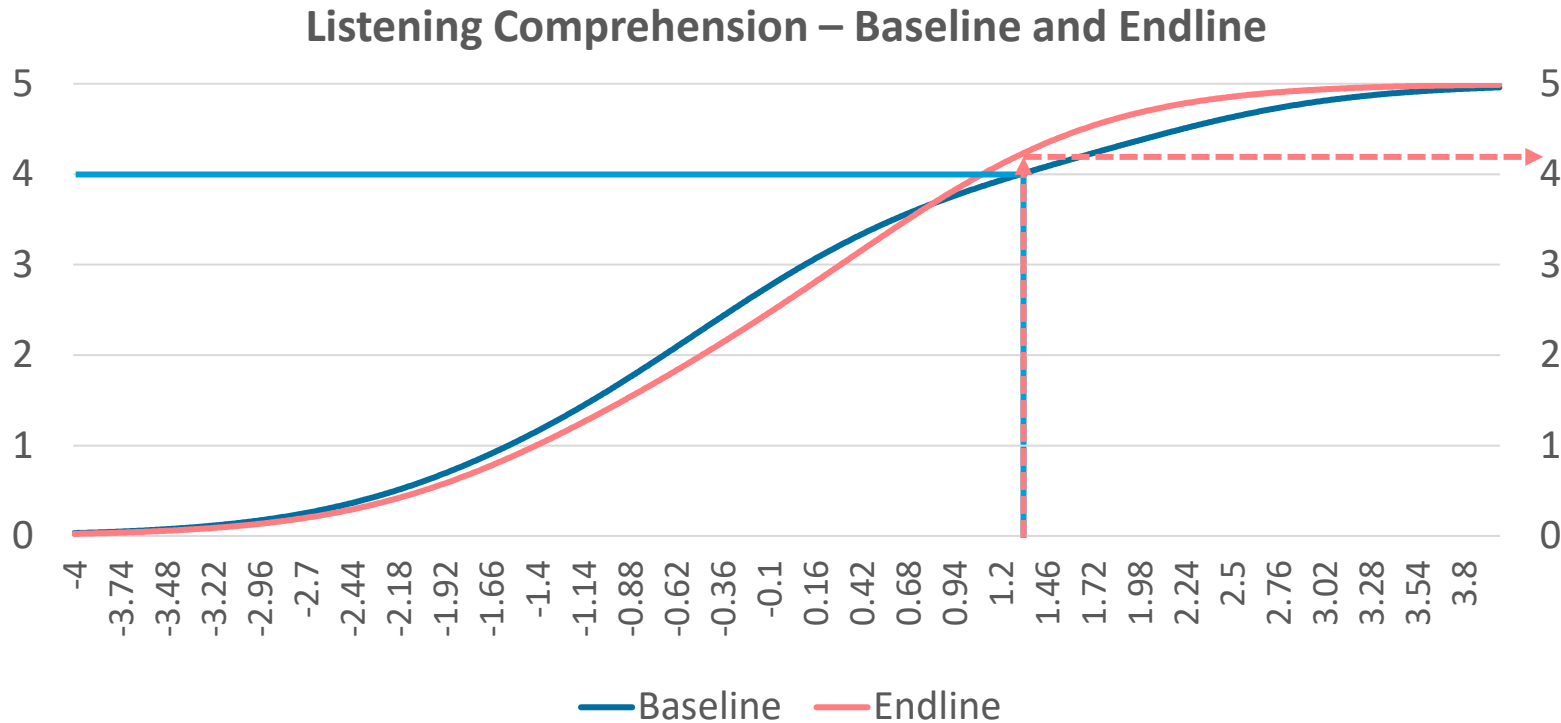
- Common item or common person equating design will help bring baseline and endline scores onto the same scale.



- Both baseline and endline had 5 RC items.
- Baseline benchmark was set at 3 score point.
- Equivalent theta baseline benchmark was 1.28
- Equivalent benchmark at endline is 2.64 (=3) score point.
- Learners who get 3 of the 5 RC items correct at endline will be classified as meeting MPL in RC.

# POLICY LINKING FOR WITHIN COUNTRY TRACKING

- Common item or common person equating design will help bring baseline and endline scores onto the same scale.



- Both baseline and endline had 5 LC items.
- Baseline benchmark was set at 4 score point.
- Equivalent theta baseline benchmark was 1.32
- Equivalent benchmark at endline is 4 score point.
- Learners who get 4 of the 5 LC items correct at endline will be classified as meeting MPL in LC.

# POLICY LINKING FOR WITHIN COUNTRY TRACKING

---

At **baseline**, percentage of students who got at least 4 listening comprehension items correctly, read at least **24** words accurately, and answered at least 3 reading comprehension items correctly.



At **endline**, percentage of students who got at least 4 listening comprehension items correctly, read at least **26** words accurately, and answered at least 3 reading comprehension items correctly.

Proportion of children in **grade 2** achieving at least a minimum proficiency level in reading (accurately read aloud and understand written words from familiar contexts. They retrieve explicit information from very short texts. When listening to slightly longer texts, they make simple inferences).

# CONCLUSION

---

- If assessments meet all five criteria listed in the policy linking toolkit, policy linking method can establish global benchmark(s).
- Once benchmarks are set, student learning progress is tracked over time through a test equating procedure and the policy linking process does not need to be repeated for this purpose.
- For countries with multiple official languages of instruction, a three-step process could be proposed for SDG reporting.
  1. Benchmarks should be set for each subtask and for each language separately
  2. Calculate the percentage of students meeting the MPL for each language
  3. Aggregate the percentages of meeting the MPL across the languages (through sampling weights) for SDG 4.1.1a reporting.

# CONCLUSION

---

- Various policy linking technical resources have been produced for public goods.
  - Policy linking toolkit 1.0 (USAID, 2019)
  - Policy linking toolkit 2.0 (UIS, 2023)
  - Quality assurance policy for policy linking report (UIS, 2020).
  - Policy linking e-learning courses (USAID, 2024)
  - Numerous papers and presentations at international conferences (CIES, 2018-2023; AEAA 2023).

## **ABDULLAH FERDOUS, PH.D.**

---

Principal Researcher/Psychometrician

WhatsApp: +1 508 969 7033

Email: [aferdous@air.org](mailto:aferdous@air.org)

LinkedIn: [LinkedIn.com/in/abdullah-ferdous-9719a124](https://www.linkedin.com/in/abdullah-ferdous-9719a124)