unesco
Institute for Statistics

25 YEARS

Education Data and Statistics Commission

Estimating Benchmarks for Precursor Reading Skills Using Reading Comprehension

February 25-26, 2025

Global Alliance to Monitor Learning

#25YearsOfDataInsights

unesco
Institute for Statistics

25 YEARS

Education Data and Statistics Commission

Background

Global Alliance to Monitor Learning

Background

- The Technical Advisory Group (TAG) recommended using only **reading comprehension (RC)** to report on indicator 4.1.1a and proposed setting the RC benchmark at 60% or 80% of the total RC score.
- TAG suggested that low-and middle-income (LMI) countries report on the percentages of students mastering the precursor skills to provide insights into early literacy development.
- These data will help LMI countries understand where students are on the path to achieving the RC benchmark.
- The purpose of this study was to explore a method for estimating precursor skill benchmarks that align with reading development theory and are psychometrically defensible.

unesco
Institute for Statistics

25 YEARS

Education Data and Statistics Commission

Methods

Global Alliance to Monitor Learning

Data

USAID: Early Grade Reading Assessment (EGRA) – Grade 2

	Country 4 Lang: Arabic	Country 6 Lang: Arabic	Country 13 Lang: Chitonga	Ghana Lang: English
Sample size	6867	2738	1024	3767
Subtasks	Listening comprehension (5) Letter sound (100) Syllable sound (100) Invented word (50) Oral reading fluency (42) Reading comprehension (5)	Listening comprehension (4) Syllable segmentation (10) Letter sound (100) Invented word (50) Oral reading fluency (76) Reading comprehension (7)	Listening comprehension (5) Letter sound (100) Syllable sound (100) Invented word (50) Oral reading fluency (56) Reading comprehension (5)	Listening comprehension (5) Letter sound (100) Invented word (50) Oral reading fluency (60) Reading comprehension (5)

unesco
Institute for Statistics

25 YEARS

Education Data and Statistics Commission

Reading Construct and Psychometric Properties

Global Alliance to Monitor Learning

Reading Construct and Psychometric Properties

Factor Analysis

	USAID Country 4 Arabic Grade 2	USAID Country 6 Arabic Grade 2	USAID Country 13 Chitonga Grade 2	USAID Ghana English Grade 2
Eigen Value	**4.41**	**3.38**	**3.96**	**3.22**
Listening Comprehension	0.48	0.45	0.32	0.63
Letter Sound	0.64		0.73	0.76
Syllable Sound	0.88		0.92	
Invented Word	0.87	0.85	0.94	0.84
Oral Reading Fluency	0.92	0.91	0.93	0.91
Reading Comprehension	0.85	0.83	0.85	0.84
Silent Reading Comprehension	0.81			

Reading Construct and Psychometric Properties

Item-Total Correlation

	USAID Country 4 Arabic Grade 2	USAID Country 6 Arabic Grade 2	USAID Country 13 Chitonga Grade 2	USAID Ghana English Grade 2
Oral Reading Fluency (timed task)	0.84	0.91	0.89	0.92
Reading Comprehension Item 1	0.41	0.47	0.48	0.55
Reading Comprehension Item 2	0.32	0.42	0.61	0.60
Reading Comprehension Item 3	0.12	0.19	0.46	0.43
Reading Comprehension Item 4	0.20	0.33	-0.03	0.39
Reading Comprehension Item 5	0.21	-0.03	-0.27	0.25
Reading Comprehension Item 6		0.27		
Reading Comprehension Item 7		0.20		

unesco
Institute for Statistics

25 YEARS

Education Data and Statistics Commission

Item Response Theory (IRT) Based Test Characteristic Curve (TCC) Mapping

Global Alliance to Monitor Learning

Item Response Theory (IRT) Based Test Characteristic Curve (TCC) Mapping

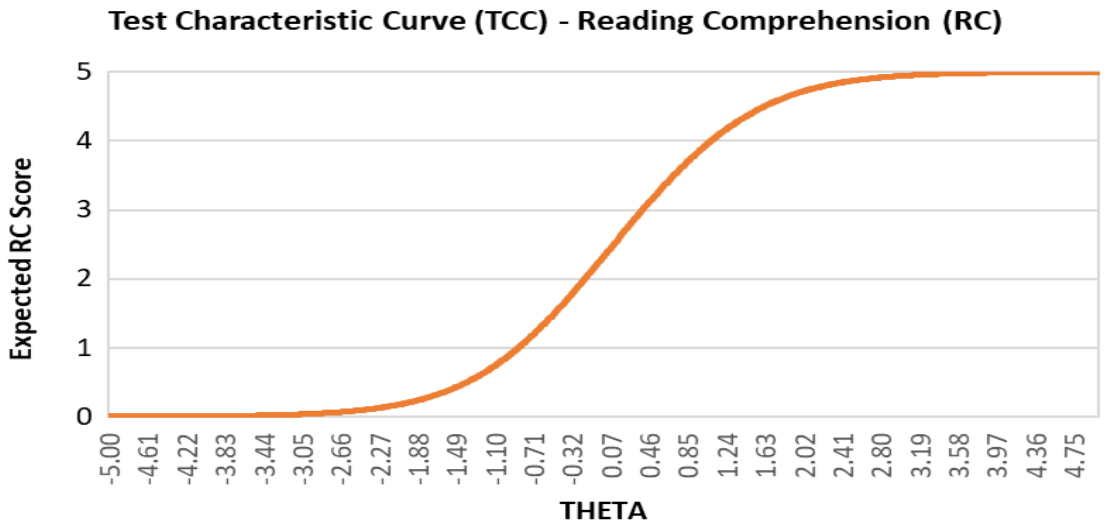
- **Step 1:** Develop a one-parameter item response theory (IRT)-based reading scale that integrates all subtasks assessing foundational reading skills, including phonological awareness, alphabetic principle, language comprehension, decoding, and reading comprehension.
- **Step 2:** Generate test characteristic curves (TCCs) for reading comprehension and other subtasks.
- **Step 3:** Estimate the test information function (TIF) and standard error (SE) curve for each subtask separately.
- **Step 4:** Set the reading comprehension benchmark at an 60% or 80% score point, equivalent to correctly answering 3 or 4 out of 5 reading comprehension items assessing the retrieval of explicit information.

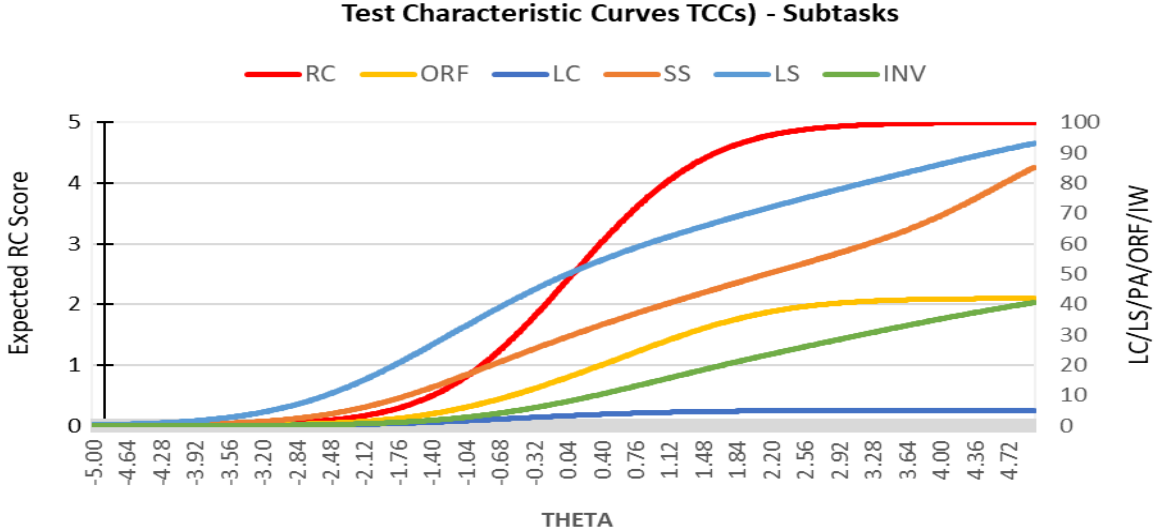
Item Response Theory (IRT) Based Test Characteristic Curve (TCC) Mapping

- **Step 5:** Convert the benchmarks of answering 3 and 4 out of 5 reading comprehension items into IRT-based theta reading scores (Θ_{RC}) and calculate standard errors of the corresponding theta value for each subtask.
- **Step 6:** Estimate benchmarks in expected score scale for other subtasks using the combined TCC mapping.

Item Response Theory (IRT) Based Test Characteristic Curve (TCC) Mapping

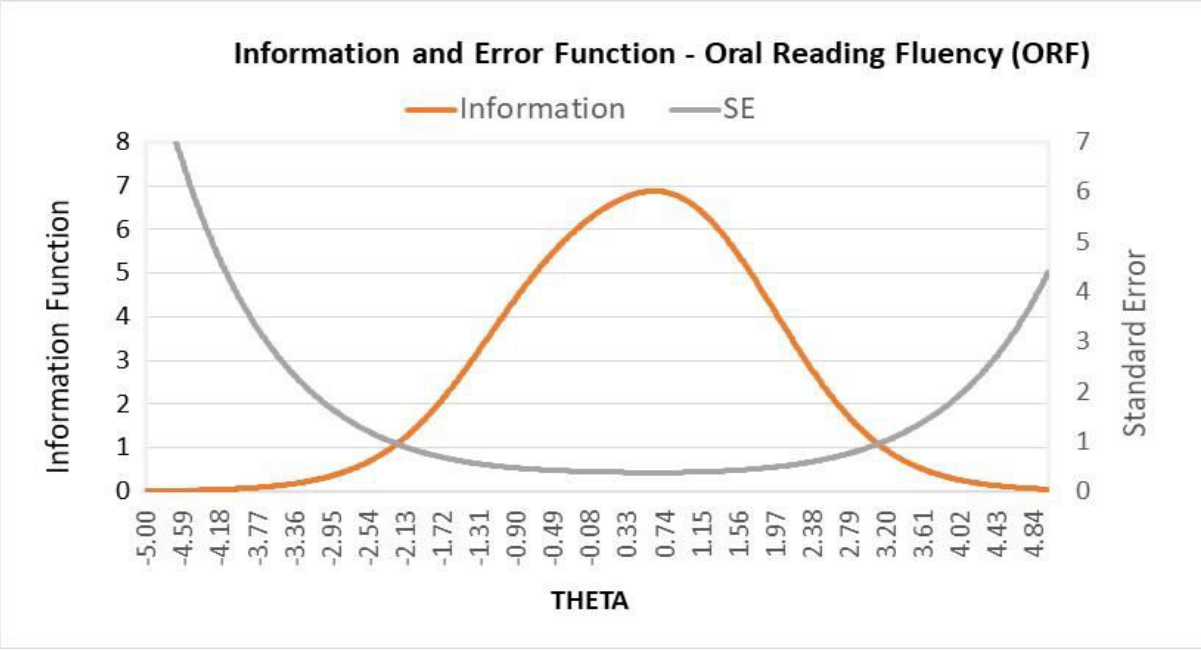
Steps 1-2: Develop a one-parameter IRT-based reading scale through concurrent calibration and generate test characteristics curves (TCCs)

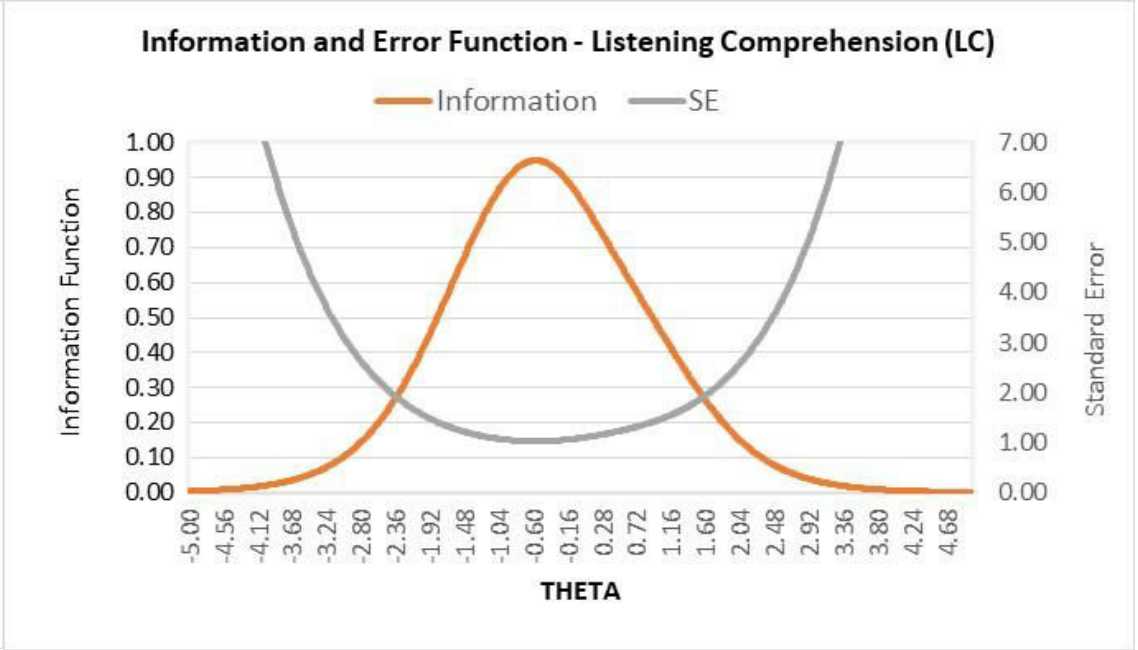




Item Response Theory (IRT) Based Test Characteristic Curve (TCC) Mapping

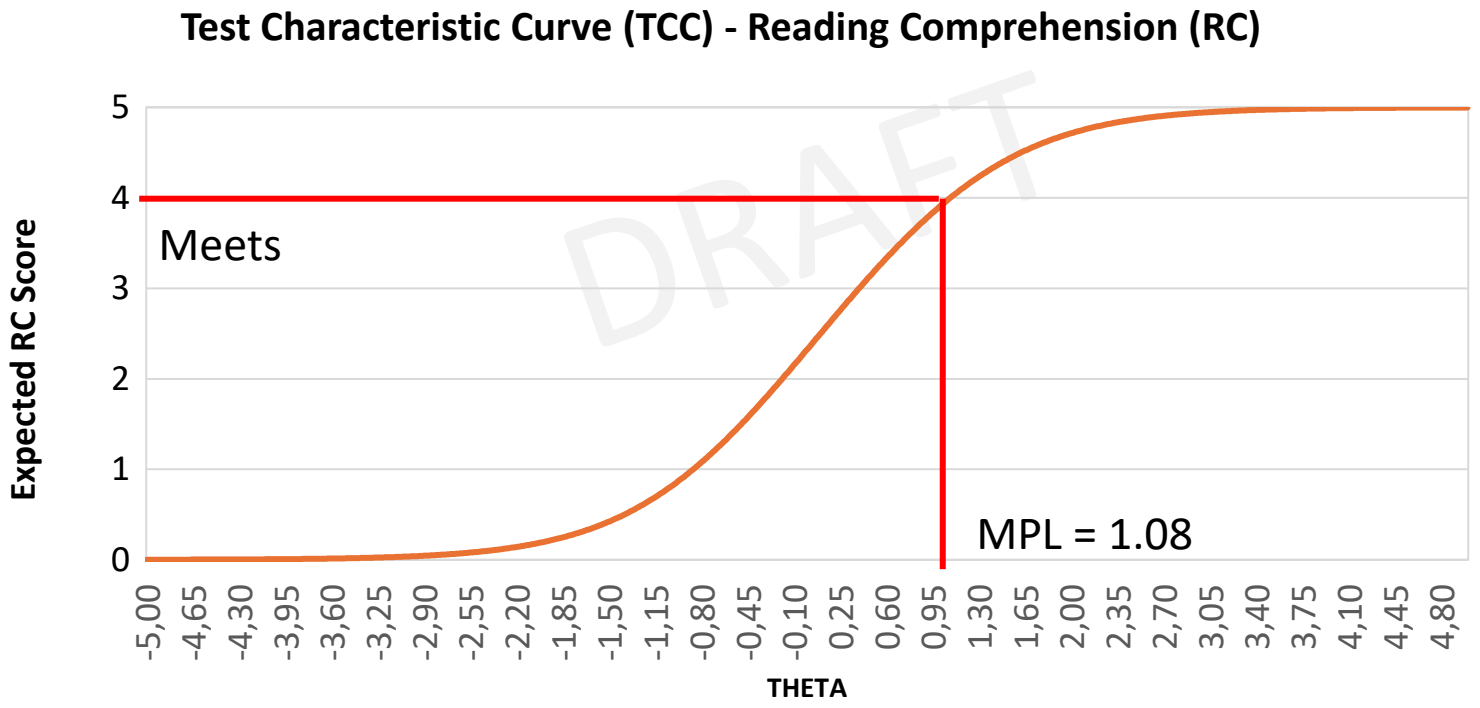
Step 3: Estimate test information function (TIF) and standard error (SE) curves for each subtask separately





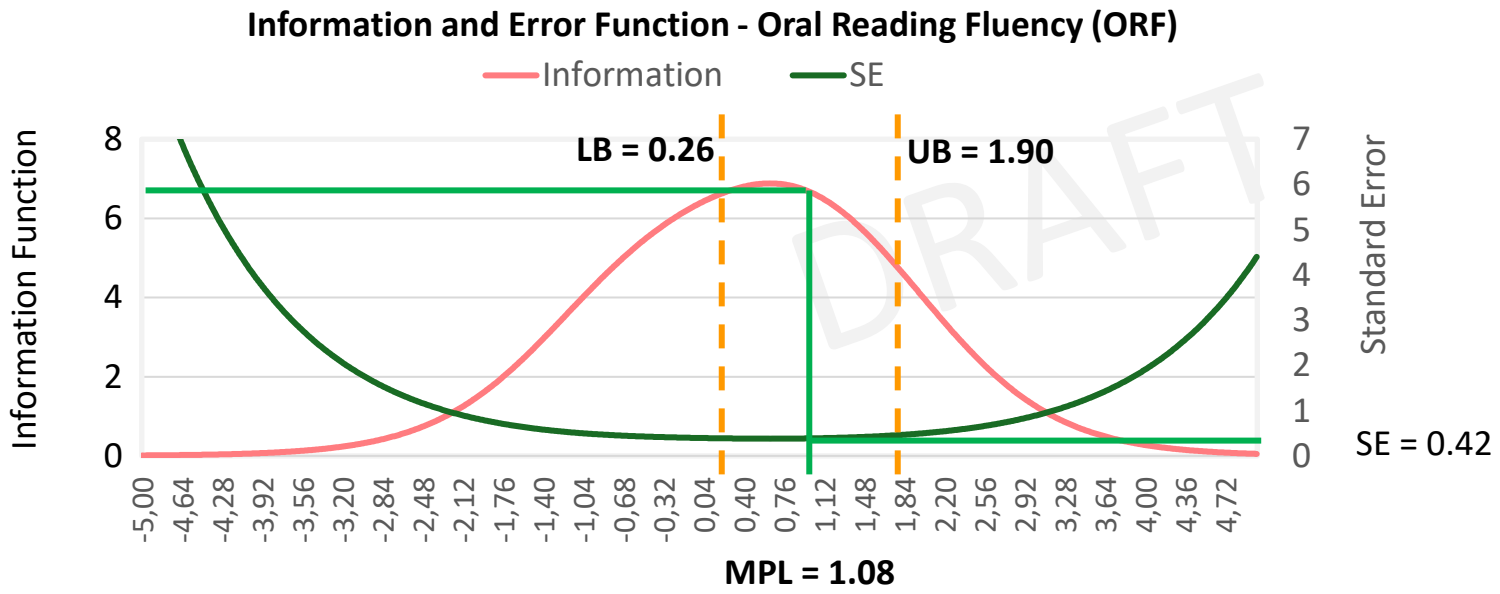
Item Response Theory (IRT) Based Test Characteristic Curve (TCC) Mapping

Steps 4-5: Convert the benchmarks of answering 4 out of 5 reading comprehension items into IRT-based theta reading scores



Item Response Theory (IRT) Based Test Characteristic Curve (TCC) Mapping

Steps 4-5: Convert the benchmarks of answering 4 out of 5 reading comprehension items into IRT-based theta reading scores

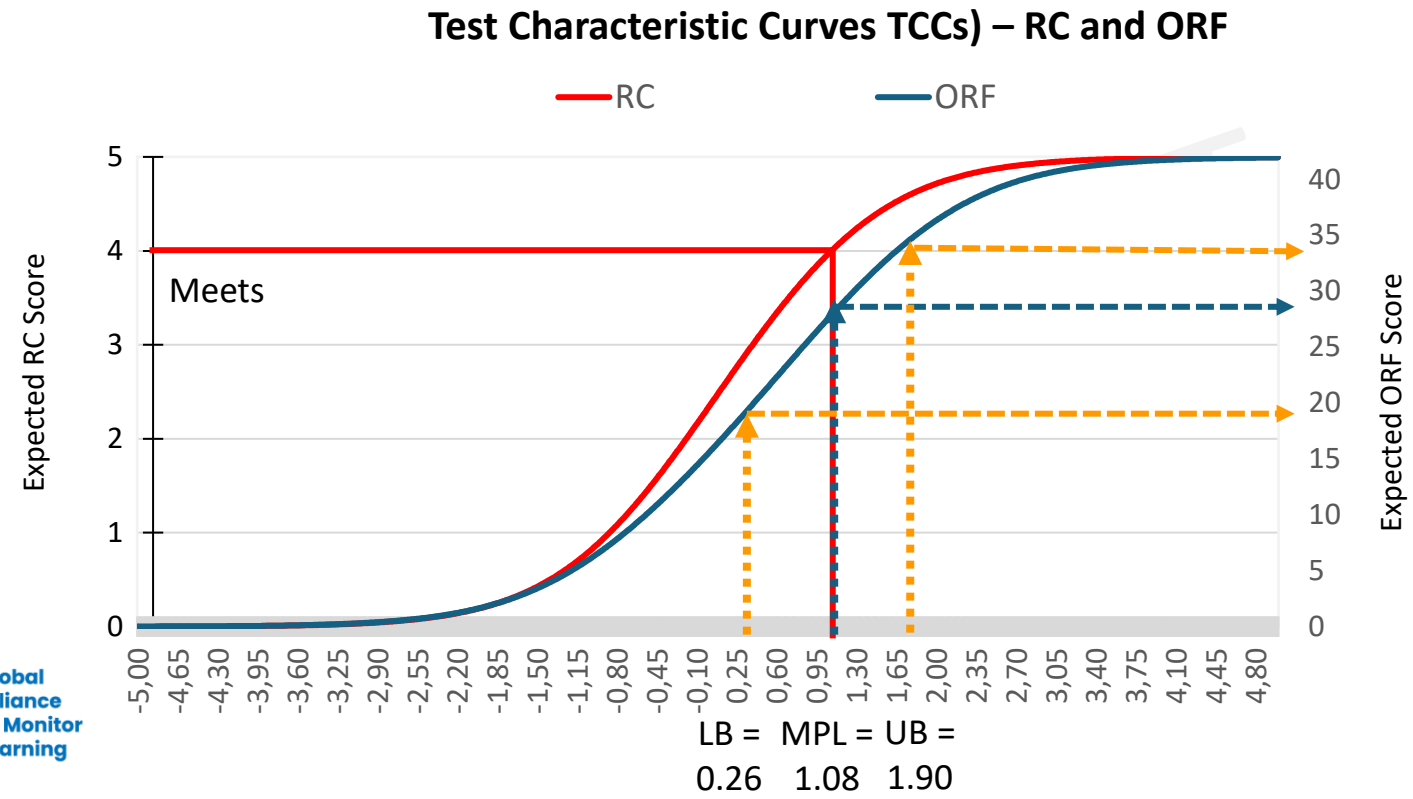


Lower Bound (LB) = 1.08 – 0.42 * 1.96 = 0.26

Upper Bound (UB) = 1.08 + 0.42 * 1.96 = 1.90

Item Response Theory (IRT) Based Test Characteristic Curve (TCC) Mapping

Step 6: Estimate benchmarks on raw score scale for other subtasks through TCC mapping.



ORF Benchmark

MPL (ORF) = 27 CWPM

Lower Bound (ORF) = 19 CWPM

Upper Bound (ORF) = 34 CWPM

Item Response Theory (IRT) Based Test Characteristic Curve (TCC) Mapping

Step 6: Estimate benchmarks on raw score scale for other subtasks through TCC mapping.

Meets : RC

4 RC → MPL = **1.08** → **27** Correct Words Per Minute (ORF)

→ **19** Correct Words Per Minute (LB-ORF)

→ **34** Correct Words Per Minute (UB-ORF)

Item Response Theory (IRT) Based Test Characteristic Curve (TCC) Mapping

Results

Assessment (Grade 2)	Subtask	Item	LCI	MPL	UCI	CI
80% Reading Comprehension						
USAID Country 4 Arabic	Oral Reading Fluency	42	19	27	34	15
USAID Country 6 Arabic	Oral Reading Fluency	76	38	46	46	8
USAID Country 13 Chitonga	Oral Reading Fluency	56	19	26	29	10
USAID Ghana English	Oral Reading Fluency	60	45	54	54	10
60% Reading Comprehension						
USAID Country 4 Arabic	Oral Reading Fluency	42	12	20	28	16
USAID Country 6 Arabic	Oral Reading Fluency	76	29	36	44	15
USAID Country 13 Chitonga	Oral Reading Fluency	56	9	16	23	14
USAID Ghana English	Oral Reading Fluency	60	39	49	54	15

unesco
Institute for Statistics

25 YEARS

Education Data and Statistics Commission

Generalized Linear Mixed Model (GLMM)

Global Alliance to Monitor Learning

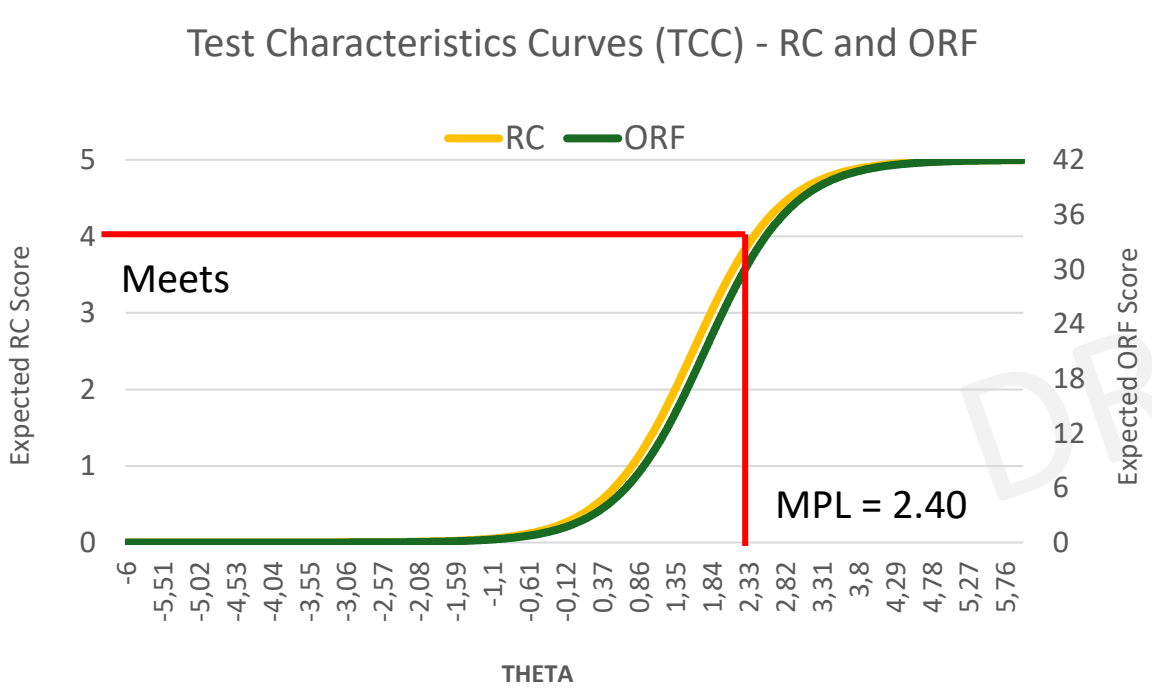
Generalized Linear Mixed Model (GLMM)

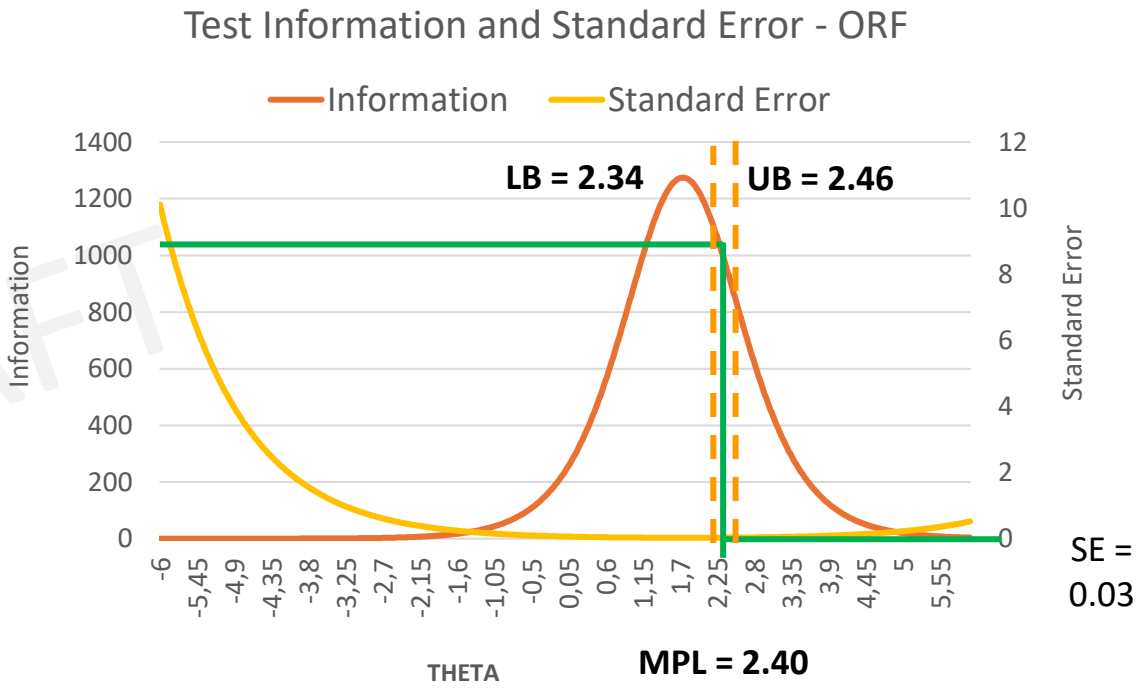
- GLMM approach used subtask score (number of words read correctly) as opposed to item level (correct/incorrect) data to estimate benchmarks.
- Expected accuracy score for each subtask is calculated by $\boldsymbol{E[U_{ij}] = n_i \cdot P(U_{ij})}$
- $\boldsymbol{U_{ij}}$ is a proportion of correct responses on subtask i for person j.

  and $\boldsymbol{n_i}$ is the total possible score on subtask i.
- $\boldsymbol{P(U_{ij})}$ is calculated based on the Rasch model:

$$\boldsymbol{P(U_{ij}) = \frac{e^{1.7(\theta_{ij} - b_i)}}{1 + e^{1.7(\theta_{ij} - b_i)}}}$$

Generalized Linear Mixed Model (GLMM) Results

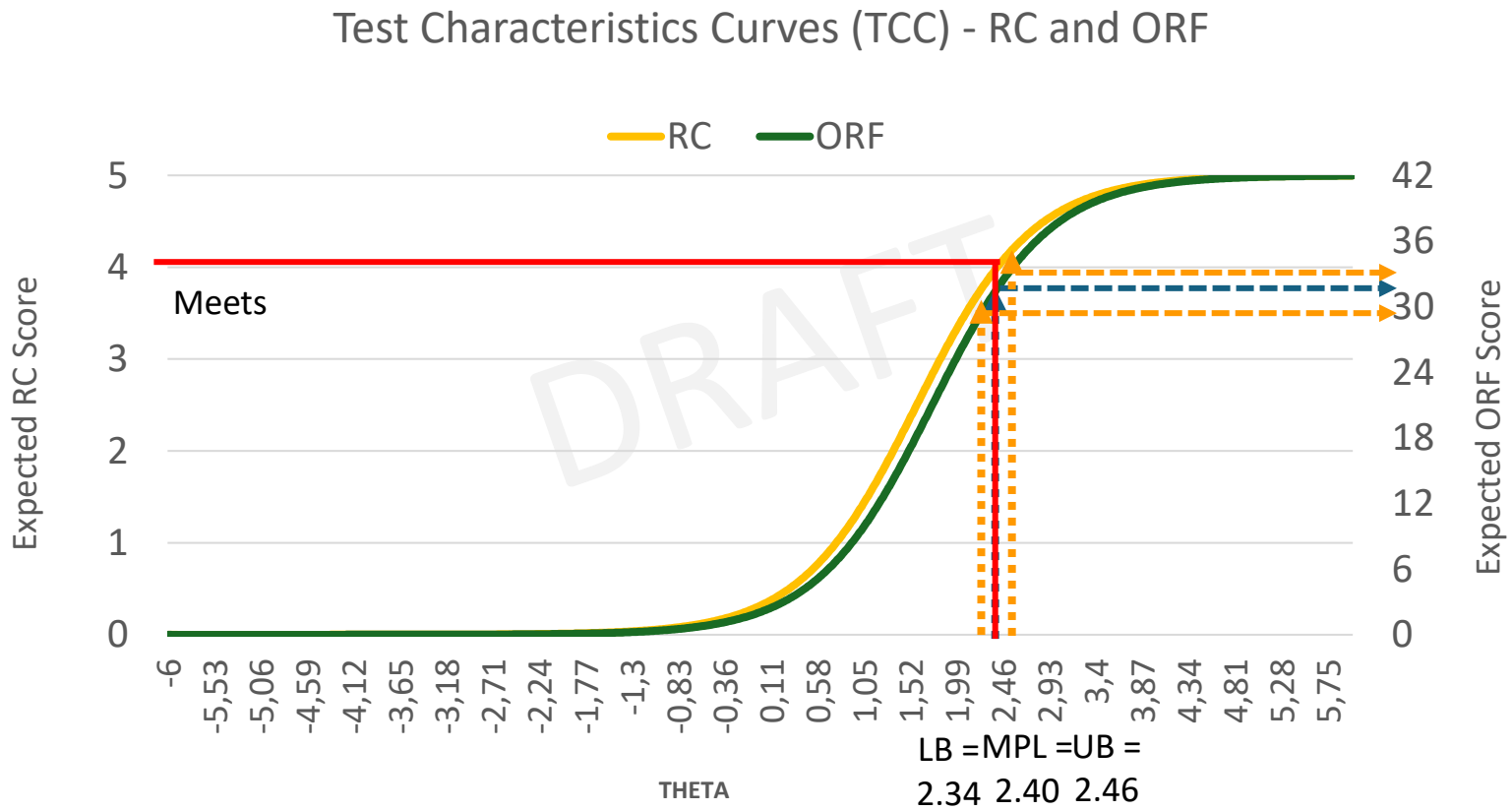




Lower Bound (LB) = 2.40 – 0.029 * 1.96 = 2.34

Upper Bound (UB) = 2.40 + 0.029 * 1.96 = 2.46

Generalized Linear Mixed Model (GLMM) Results



Generalized Linear Mixed Model (GLMM) Results

Meets : RC

4 RC → MPL = **2.40** → **32** Correct Words Per Minute (ORF)

→ **31** Correct Words Per Minute (LB-ORF)

→ **32** Correct Words Per Minute (UB-ORF)

Generalized Linear Mixed Model (GLMM) Results

Assessment (Grade 2)	Subtask	Item	LCI	MPL	UCI	CI
80% Reading Comprehension						
USAID Country 4 Arabic	Oral Reading Fluency	42	31	32	32	1
USAID Country 6 Arabic	Oral Reading Fluency	76	55	56	57	2
USAID Country 13 Chitonga	Oral Reading Fluency	56	19	19	20	1
USAID Ghana English	Oral Reading Fluency	60	54	54	55	1
60% Reading Comprehension						
USAID Country 4 Arabic	Oral Reading Fluency	42	21	22	23	2
USAID Country 6 Arabic	Oral Reading Fluency	76	29	29	30	1
USAID Country 13 Chitonga	Oral Reading Fluency	56	9	9	10	1
USAID Ghana English	Oral Reading Fluency	60	46	47	48	2

unesco
Institute for Statistics

25 YEARS

Education Data and Statistics Commission

Classical Test Theory (CTT) Based Raw Score Approach

Global Alliance to Monitor Learning

Classical Test Theory (CTT) Based Raw Score Approach

- Benchmark for each subtask was calculated based on average performance among students with reading comprehension scores equal to the target or threshold.
- Standard error for each subtask was calculated as:
    - SE (Subtask) = Standard Deviation of Subtask Score / SQRT (Number of Students – 1)

Classical Test Theory (CTT) Based Raw Score Approach

Assessment (Grade 2)	Subtask	Item	LCI	MPL	UCI	CI
80% Reading Comprehension						
USAID Country 4 Arabic	Oral Reading Fluency	42	33	34	36	3
USAID Country 6 Arabic	Oral Reading Fluency	76	67	77	87	20
USAID Country 13 Chitonga	Oral Reading Fluency	56	31	33	35	4
USAID Ghana English	Oral Reading Fluency	60	68	77	87	19
60% Reading Comprehension						
USAID Country 4 Arabic	Oral Reading Fluency	42	26	27	28	2
USAID Country 6 Arabic	Oral Reading Fluency	76	52	57	62	10
USAID Country 13 Chitonga	Oral Reading Fluency	56	24	26	27	3
USAID Ghana English	Oral Reading Fluency	60	45	52	59	14

unesco
Institute for Statistics

25 YEARS

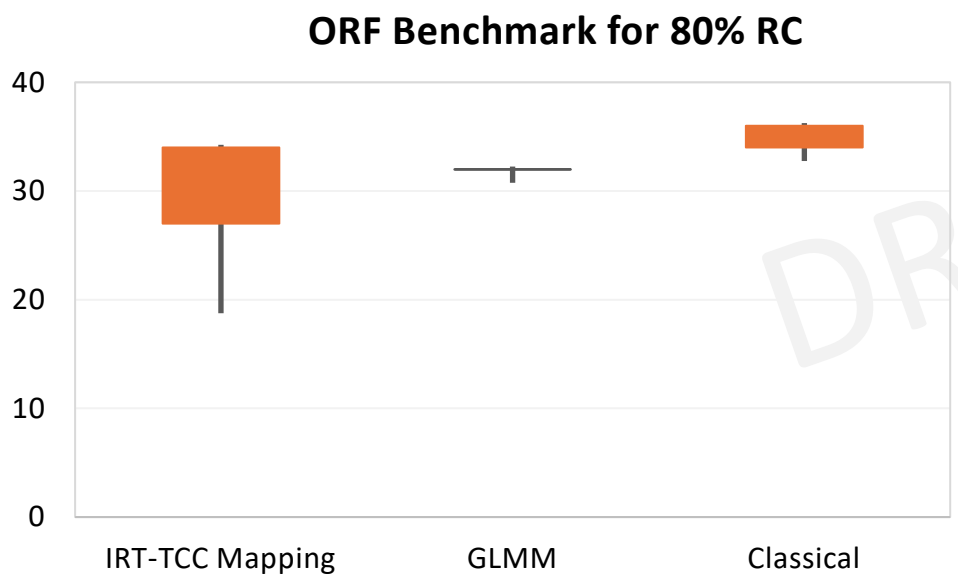
Education Data and Statistics Commission

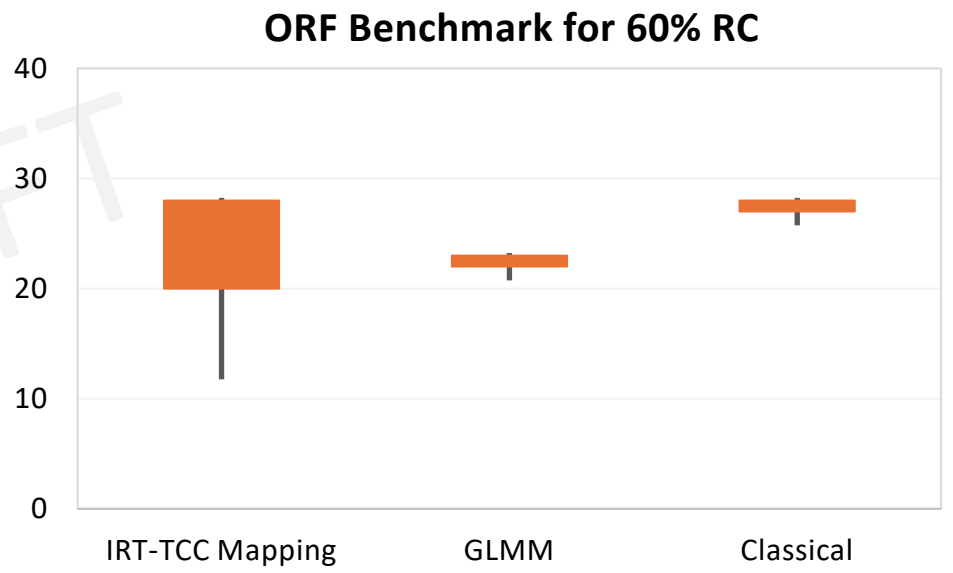
Comparison of Three Methods

Global Alliance to Monitor Learning

Comparison of Three Methods

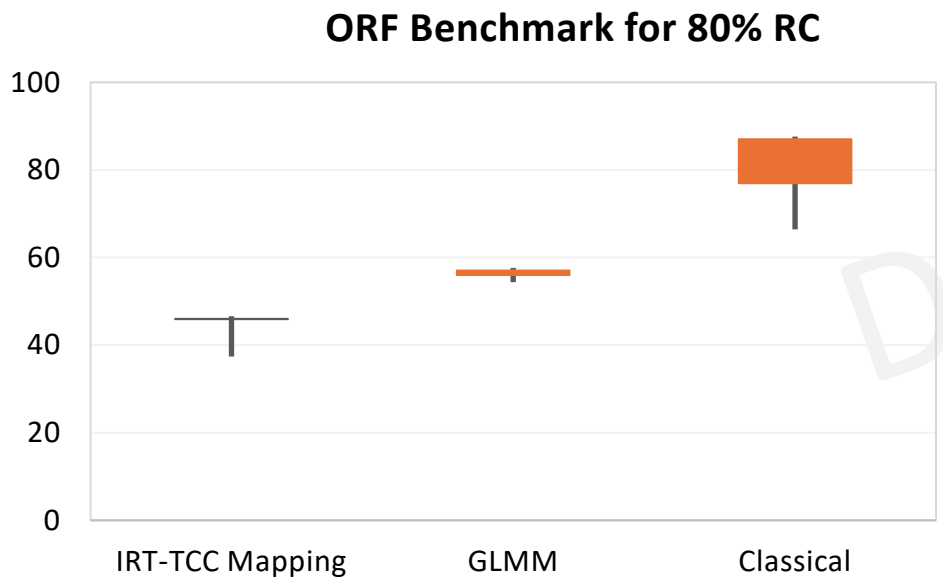
Country 4 Arabic Language

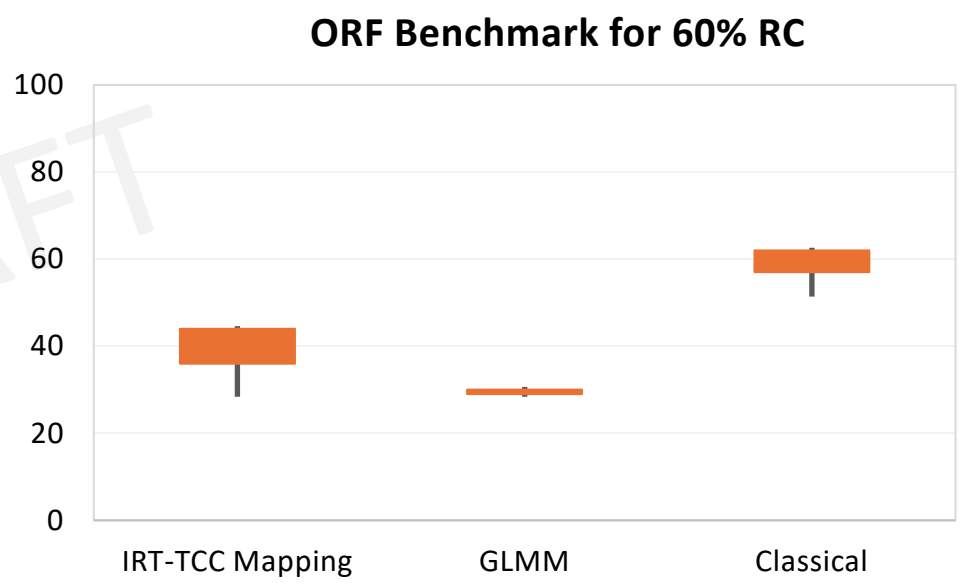




Comparison of Three Methods

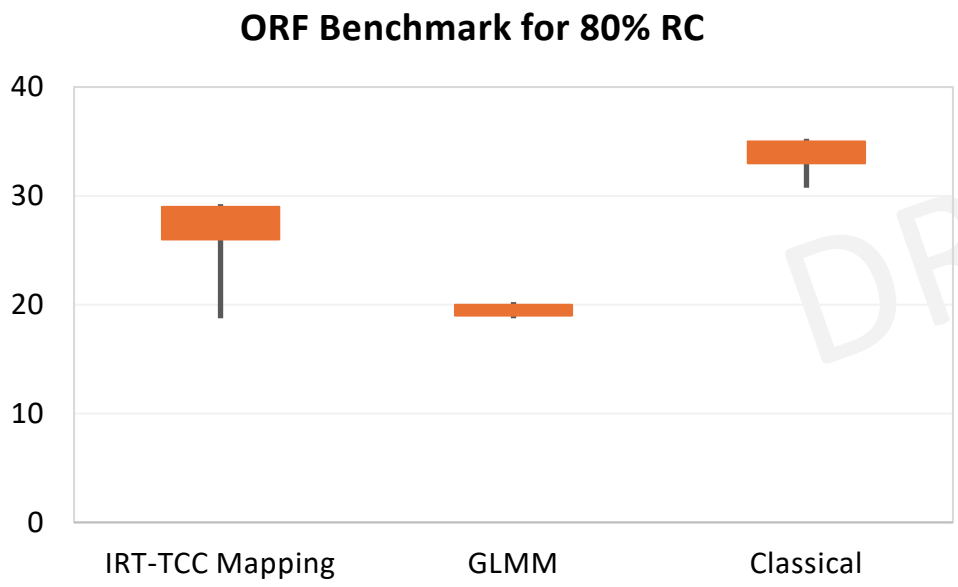
Country 6 Arabic Language

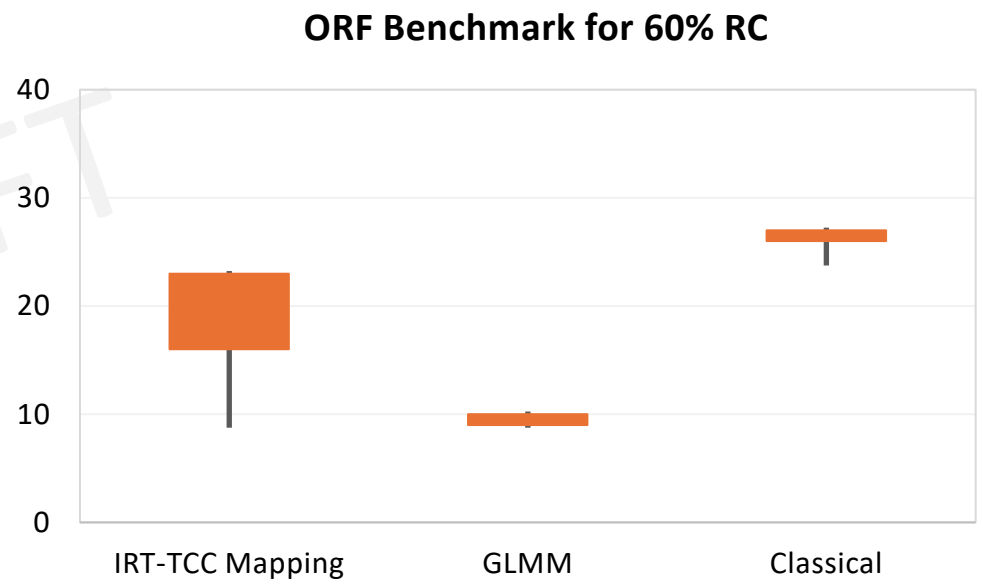




Comparison of Three Methods

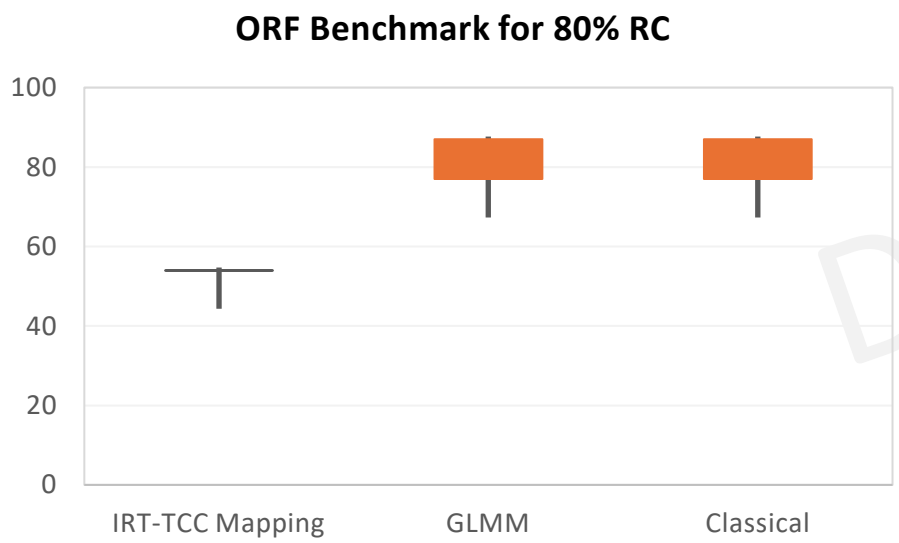
Country 13 Chitonga Language

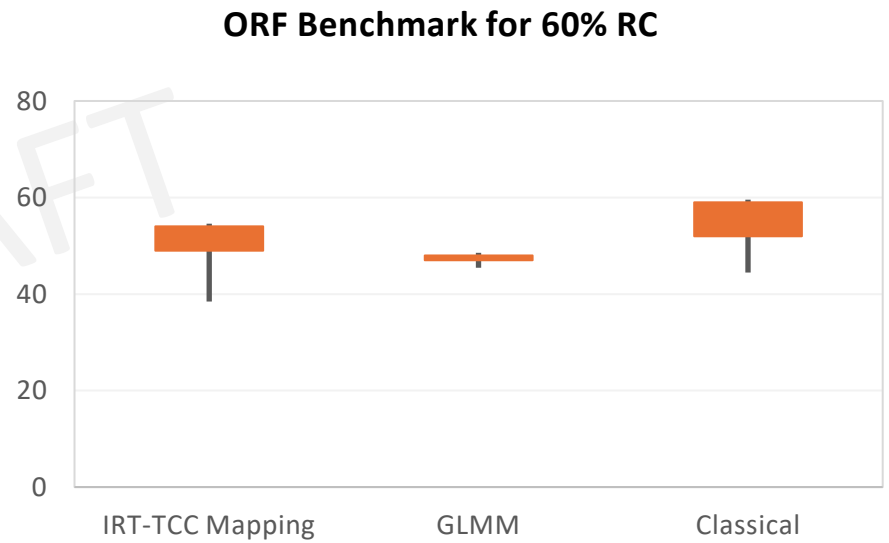




Comparison of Three Methods

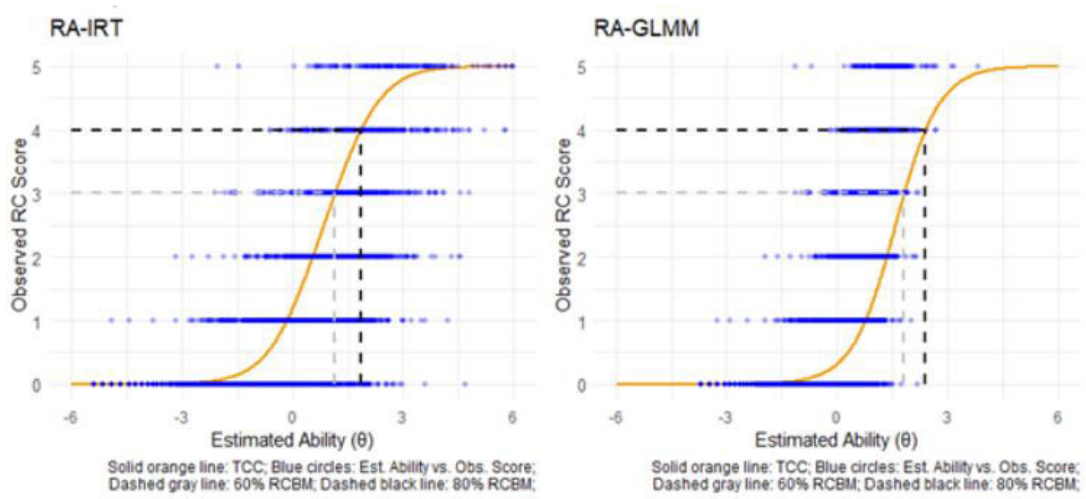
Ghana English Language

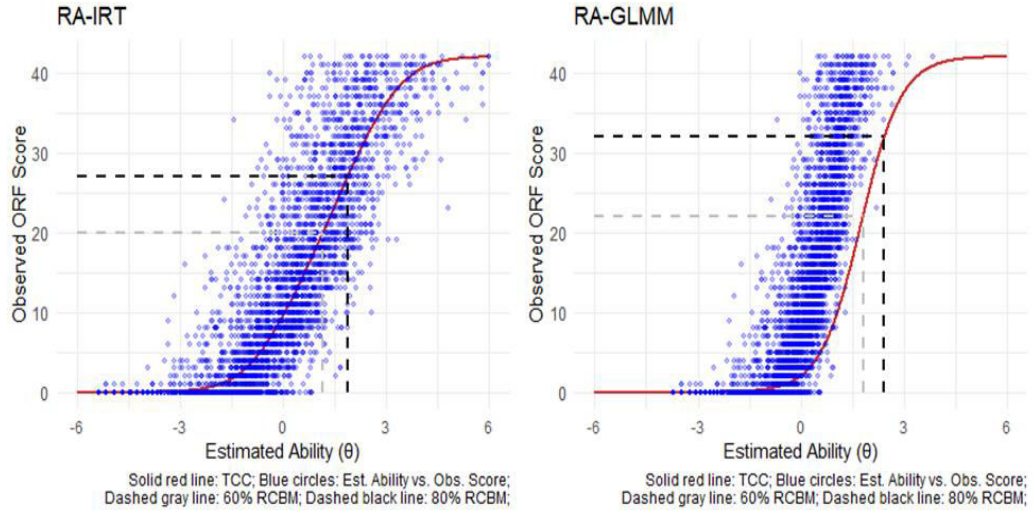




Comparison of Three Methods

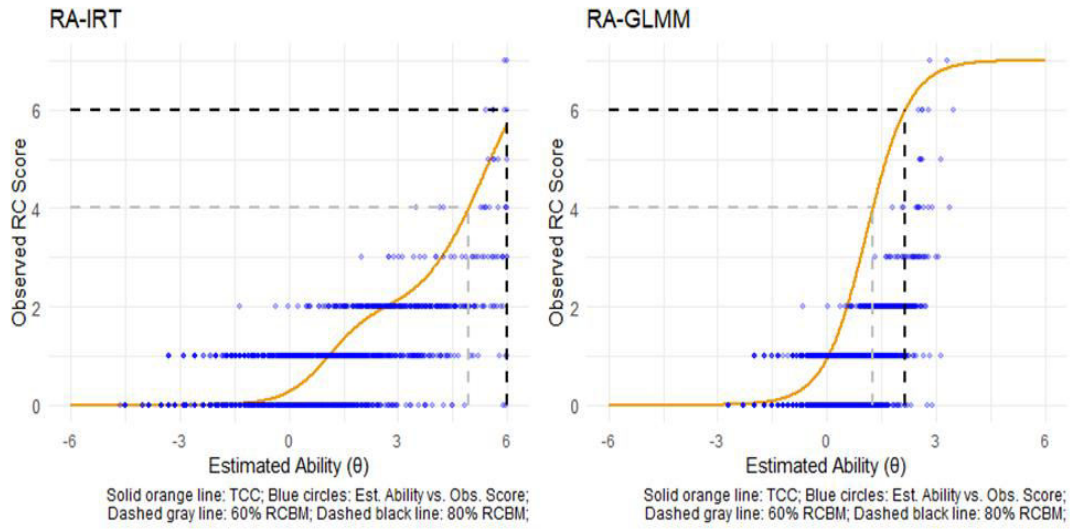
Country 4 Arabic Language

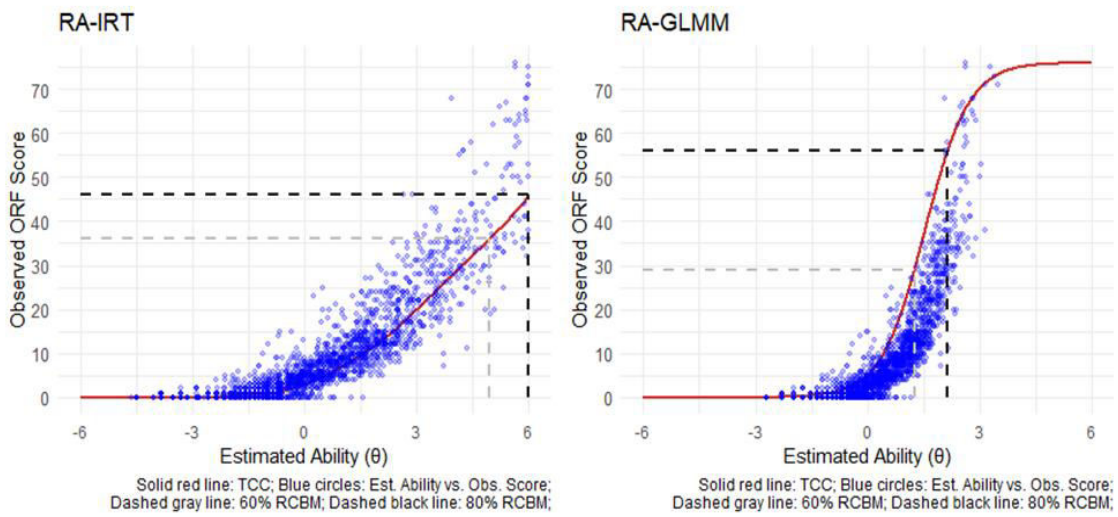




Comparison of Three Methods

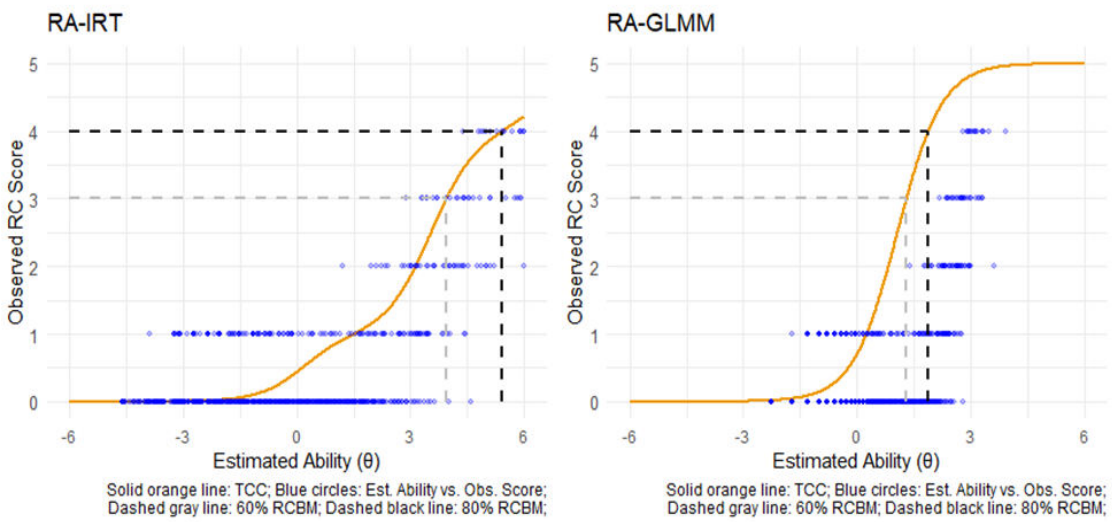
Country 6 Arabic Language

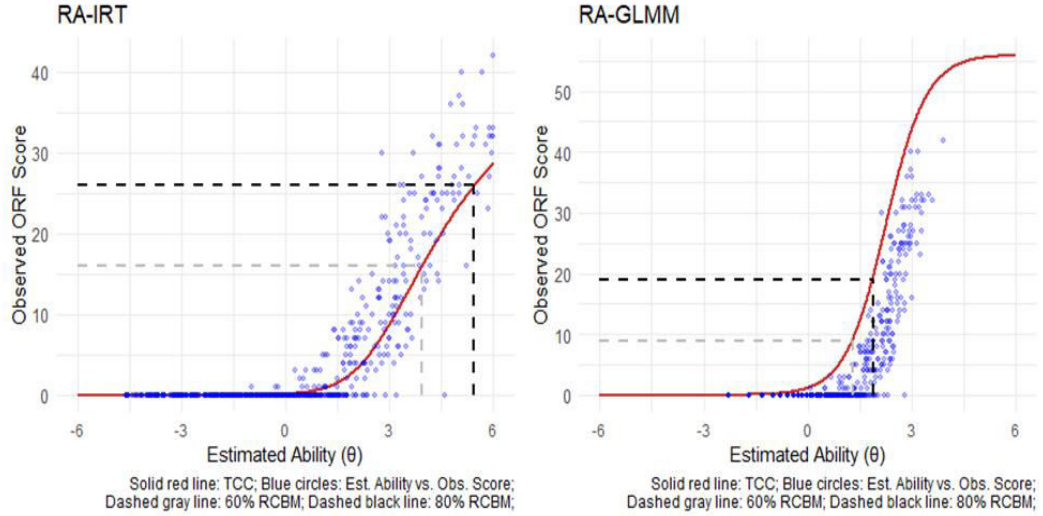




Comparison of Three Methods

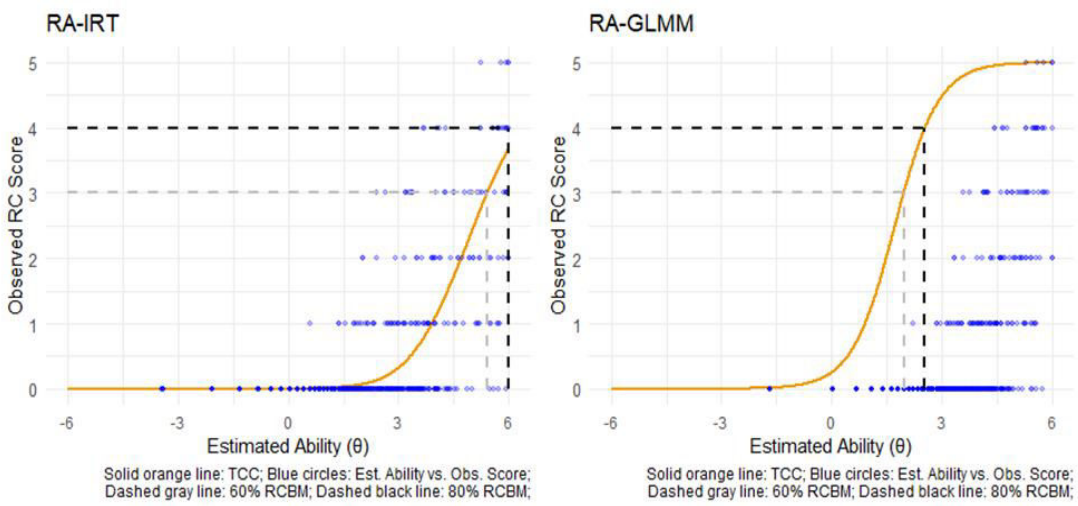
Country 13 Chitonga Language

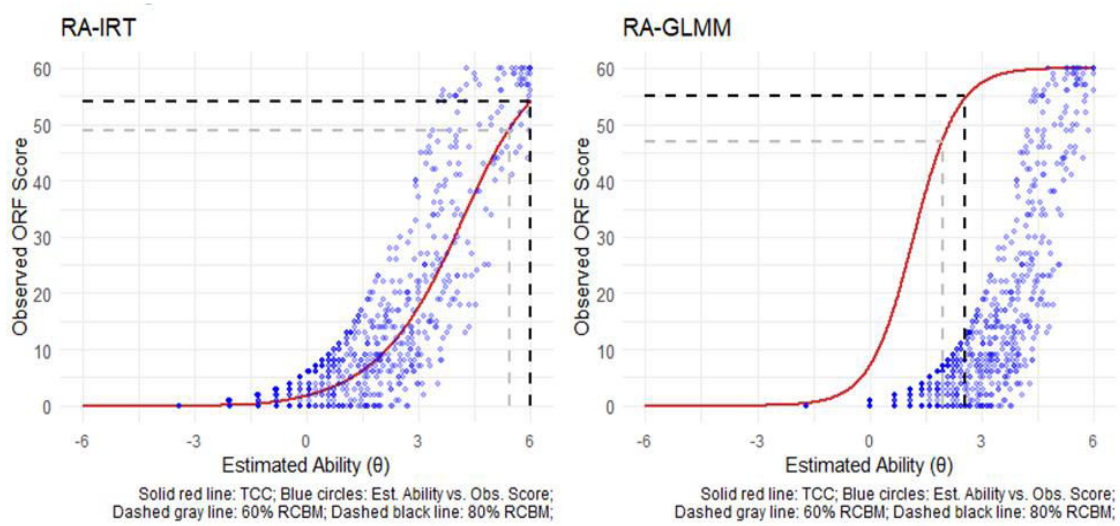




Comparison of Three Methods

Ghana English Language





unesco
Institute for Statistics

25 YEARS

Education Data and Statistics Commission

Conclusions

Global Alliance to Monitor Learning

Conclusions

- The IRT-based TCC mapping is an effective and reliable method for setting benchmarks.
- It provides a more precise understanding of student ability by accounting for item difficult while offering a unified measure of ability across various skills, including comprehension and precursor skills, on the same latent scale.
- It consistently demonstrates a stronger model fit to the data, effectively capturing patterns.
- GLMM and classical approaches offer computational efficiency but fail to capture differences between test items.
- These approaches exhibit a lack of model fit to the data, a pattern observed consistently across all subtasks.

Conclusions

- Despite its advantages, the IRT based TCC mapping method has a wider confidence interval compared to GLMM and the classical approach.
- Further refinements are needed to improve the precision of estimated benchmarks.
- The IRT based TCC mapping method shows strong potential for assessments measuring both fluency and accuracy (e.g., India FLS) or those required by UIS for 4.1.1a reporting.
- To further assess its robustness and generalizability, the analysis should be replicated with additional datasets measuring both accuracy and fluency.
- Facilitate a stronger dialogue or discussion between psychometricians and reading science experts to validate benchmarks.

unesco
Institute for Statistics

25 YEARS

Education Data and Statistics Commission

THANK YOU

Learn more:

uis.unesco.org

databrowser.uis.unesco.org

@UNESCOstat

#25YearsOfDataInsights

Global Alliance to Monitor Learning