Using standard setting in different large-scale assessments: An alternative to monitor SDG indicator 4.1.1

Results from a pilot using AMPL

Andres Sandoval-Hernandez

Context and background

With just a few years to go until 2030, we need now to produce information to report the level of SDG 4 and the direction of its trend.

This information will help countries understand where they are and set improvement targets in their sector plans.

There is some important progress:

- The Minimum Proficiency Levels (MPL)

- A Global Proficiency Framework

- A set of methods to align assessments to this common framework:
  - Assessments for Minimum Proficiency Level (AMPL),
  - Rosetta Stone,
  - Pairwise comparison,
  - Policy linking.

Indicators and challenges for SDG 4

The focus for the exercise was SDG **indicator 4.1.1(a)**

- Proportion of children and young people (a) in **grades 2/3**; (b) at the end of primary; and (c) at the end of lower secondary achieving at least a **minimum proficiency level** in (i) reading and (ii) **mathematics**, by sex

Challenges in measuring the indicator globally:

- Comparability of tests
- Curriculum differences
- Linguistic diversity
- Need for a standardized metric
- identifying what is a **minimum proficiency level** in any given proficiency assessment

Objectives

Purpose of the project:

- Develop **a step-by-step procedure for setting descriptive cut-scores for different assessments based on the MPL definition for SDG 4.1.1(a)**, allowing for the estimation of the proportions of students who achieve the indicator.

Purpose of this first exercise:

- Set a descriptive cut-score for **AMPL** based on the definition of the MPL for SDG 4.1.1(a) **mathematics**.

Methods

Adaptation of the Bookmark Method

- Eligibility criteria for reporting

- Psychometric Analysis: Assessing item difficulty and discrimination (IRT).

- Expert Panel Engagement: Structured workshops to review item maps and determine MPL cut-off scores.

- Standard-Setting: Adapting the Bookmark method to set cut-off scores that define the MPL.

- In short, experts received an ordered list of items, from least to most difficult.

- They individually placed a bookmark at the cut-point separating students who reach the MPL from those who do not.

- After discussion, they agreed on a single bookmark.

- This final bookmark determined the cut score.

Agenda for the standard-setting pilot workshop

Phase	Allocated time
Welcome and presentations	10 minutes
Method description and explanation	20 minutes
Individual judgement of MPL	70 minutes
Break	10 minute
Group discussion	60 minutes
Comparisons with previous MPL exercises	20 minutes
Closing remarks	10 minutes

* Experts received all materials one week in advance.

Assessment for Minimum Proficiency Level (AMPL)

- Use of AMPL on mathematics for grades 2/3

- Specifically designed to measure the MPL

- Already has set an MPL , which serves as a contrast to the results of this pilot

Implementation of AMPLs
By tool and language

Tool
O AMPL-a
□ AMPL-ab
+ AMPL-b

Language
■ English
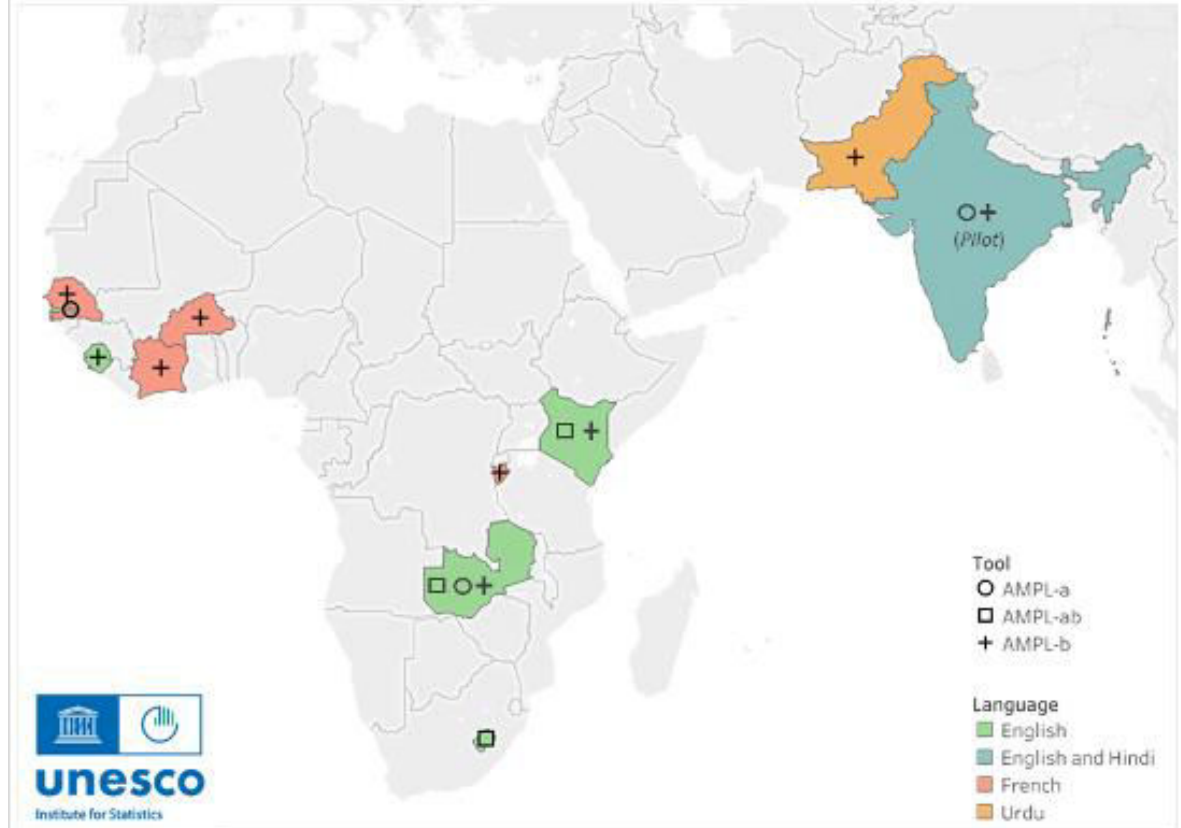■ English and Hindi
■ French
■ Urdu

Figure 1. AMPL coverage

Results

AMPL cut score set in item 23 with a difficulty of -1.76

Cut score of the pilot set on item 27 with a difficulty of -1.4

Item #	Item Key	Difficulty	E1	E2	E3	E4	E5	E6	E7	E8	E9
1	AM002	-4.11									
2	AM009	-3.56									
3	AM006	-3.43									
4	AM010	-3.27									
5	AM001	-3.22				Cut 1					
6	AM012	-3.08									
7	AM022	-3.02									
8	AM011	-2.97									
9	AM019	-2.54									
10	AM007	-2.33									
11	AM003	-2.27									
12	AM017	-2.26									
13	AM027	-2.19									
14	MM011	-2.17									
15	AM020	-2.11									
16	AM030	-2.09									
17	PM459	-2.07								Cut 1	
18	AM029	-2.01									
19	AM008	-1.97									
20	AM014	-1.82									
21	AM004	-1.78							Cut 1-2	Cut 2	
22	AM013	-1.77									
23	AM025	-1.76			Official AMPL Cutscore						
24	AM005	-1.6									
25	AM026	-1.52					Cut 2				
26	AM028	-1.42									
27	AM021	-1.4	Cut 2		Cut 1-2						
28	PM462	-1.4									Cut 1-2
29	AM023	-1.38									
30	AM016	-1.35									
31	AM024	-1.13									
32	MM209	-1.06	Cut 1					Cut 1			
33	AM018	-1.02									
34	MM029	-0.9		Cut 1-2				Cut 1-2			
35	MM125	-0.85									
36	MM004	-0.76									
37	MM208	-0.74									
38	MM022	-0.63									

Table 1. Cut scores and item difficulty

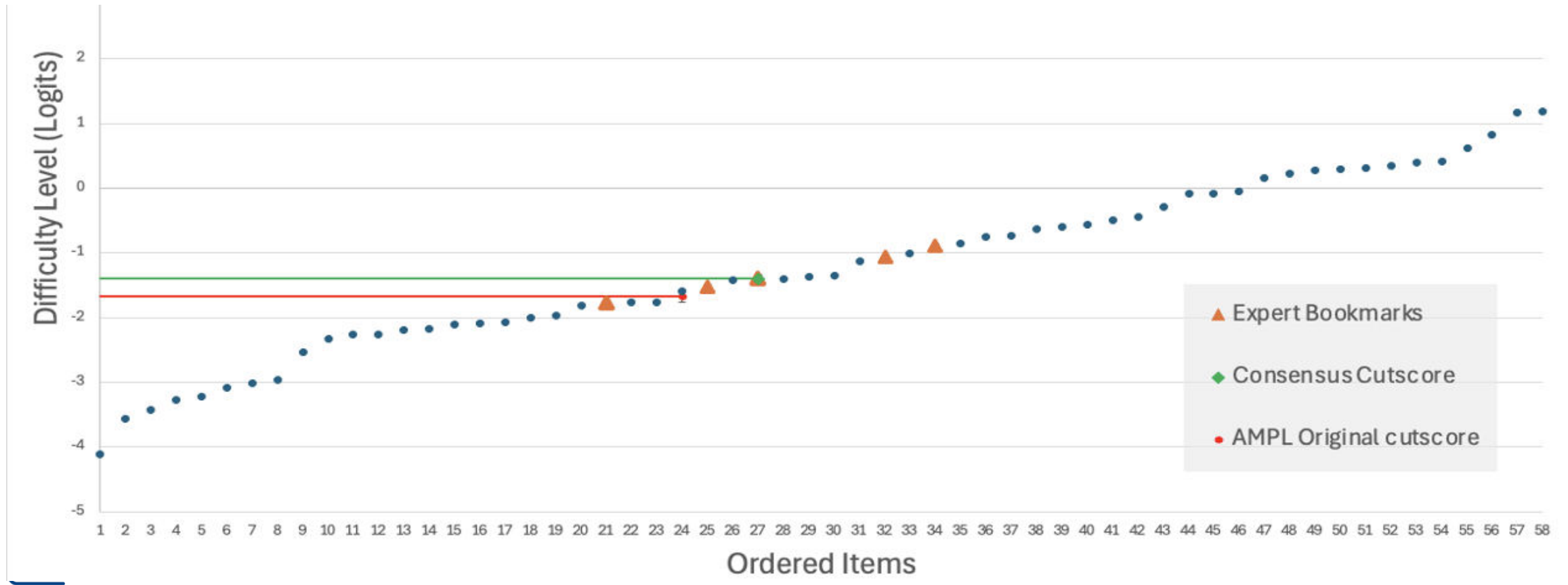# Ordered Item Map with corrected bookmarks and final cuts core

Figure 2. Final ordered item map with corrected bookmarks and final cuts core

Experts cut scores and item difficulties of the items

Expert	Relative Ordered Position of Bookmarked Item		Item Difficulty (Logits) of Bookmarked Item	
	Cut score 1	Cut score 2	Cut score 1	Cut score 2
E1	32	27	-1.06	-1.40
E2	34	34	-0.90	-0.90
E3	27	27	-1.40	-1.40
E4				
E5	32	25	-1.06	-1.52
E6	34	34	-0.90	-0.90
E7	21	21	-1.78	-1.78
E8	17	21	-2.07	-1.78
E9	28	28	-1.40	-1.40
Mean	28.13	27.13	-1.32	-1.39
Median	30	27	-1.23	-1.40
SD	6.26	4.99	0.43	0.34

Table 2. Summary of experts cut scores for the individual judgment exercise and after the discussion session by the relative ordered position of the bookmarked item and item difficulties of the items expressed in logits

Proportion of students meeting or exceeding the MPL (Pilot vs AMPL)

Country	Proportion of students meeting MPL (original)			Proportion of students meeting MPL (pilot exercise)		
	Estimate	SE (lower limit)	SE (upper limit)	Estimate	SE (lower limit)	SE (upper limit)
Gambia	0.28	0.24	0.31	0.20	0.17	0.23
Kenya	0.88	0.86	0.89	0.77	0.75	0.80
Lesotho	0.83	0.80	0.85	0.70	0.66	0.73
Zambia	0.49	0.46	0.51	0.36	0.34	0.39

Table 3. Proportion of students meeting or exceeding the MPL in the original AMPL assessment and the proportion of students meeting or exceeding the MPL as set in the pilot exercise

Concluding remarks

Results from the pilot exercise
- Cut scores for MPL
  - AMPL original cut score was set at item 23 with difficulty of -1.4
  - Pilot cut score using AMPL was set at item 27 with difficulty of -1.76
- Proportion of students reaching MPL in mathematics.
  - Statistically significant differences of nearly 10% less students reaching MPL in pilot exercise compared to original AMPL estimates.

Measuring the proportion of students meeting MPL
- Standard setting with a stable panel of experts is a promising alternative for reporting the proportion of students reaching the MPL using different assessments.
- Standard setting requires both familiarization with the MPL definitions and the Bookmark Method.
- Further adjustments should be considered.

THANK YOU

Learn more:

uis.unesco.org

databrowser.uis.unesco.org

@UNESCOstat

#25YearsOfDataInsights