



United Nations
Educational, Scientific and
Cultural Organization



UNESCO
INSTITUTE
FOR
STATISTICS



TECHNICAL
COOPERATION
GROUP



Benchmarking and minimum thresholds: concepts and feasibility of defining minimum levels

Silvia Montoya, Director, UNESCO Institute for Statistics

TCG 4

January 2018 – Dubai, United Arab Emirates



- ## Recommendations
- Indicators that already have a target set
 - Finance
 - Top levels
 - Indicators that have a benchmark definition in their framing

Summary of Strategies for Indicator 4.1.1

Strategy	Implications
<p>Strategy 1: use of national assessments to measure SDG4 with adjustments using international assessments. To be implemented in the short run</p>	<ul style="list-style-type: none"> • High levels of external validity for measuring the minimum level of competency established in official curriculum. • Low levels of international comparability
<p>Strategy 2: equating among international and regional assessments. To be implemented in the medium run</p>	<ul style="list-style-type: none"> • Apparent low cost by using existing assessments. • Entails performing one equating for each of the grades to be assessed in indicator 4.1.1 and defining new proficiency levels for each scale. • Technically questionable from a psychometric and substantive point of view. • Low levels of external validity for representing the national curriculum.

Summary of Strategies for Measuring SDG4 (cont'd)

Strategy	Implications
<p>Strategy 3: equating between different international evaluations aiming at similar school grades. To be implemented in the medium or long run</p>	<ul style="list-style-type: none"> • Requires the definition of anchor items that can be shared across the different evaluations and the creation of a consortium of different assessment projects. • Difficulties of comparison because of the differences in the domains assessed in the different assessments. • Psychometrically and substantively more robust. • Low levels of external validity for representing the national curriculum.
<p>Strategy 4: creating a Worldwide Proficiency Assessment on Numeracy and Literacy. To be implemented in the long run.</p>	<ul style="list-style-type: none"> • Psychometrically and substantively robust. • Politically difficult to convince countries to participate in this assessment. • Requires the participation of technical institutions in the design, implementation, and analysis of test results. • Low levels of external validity for representing the national curriculum.



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

What is the immediate need?

- Define “minimum proficiency levels” for reading and mathematics, and
- Produce a reporting metric and a mechanism for linking existing assessments and their performance levels to this metric.

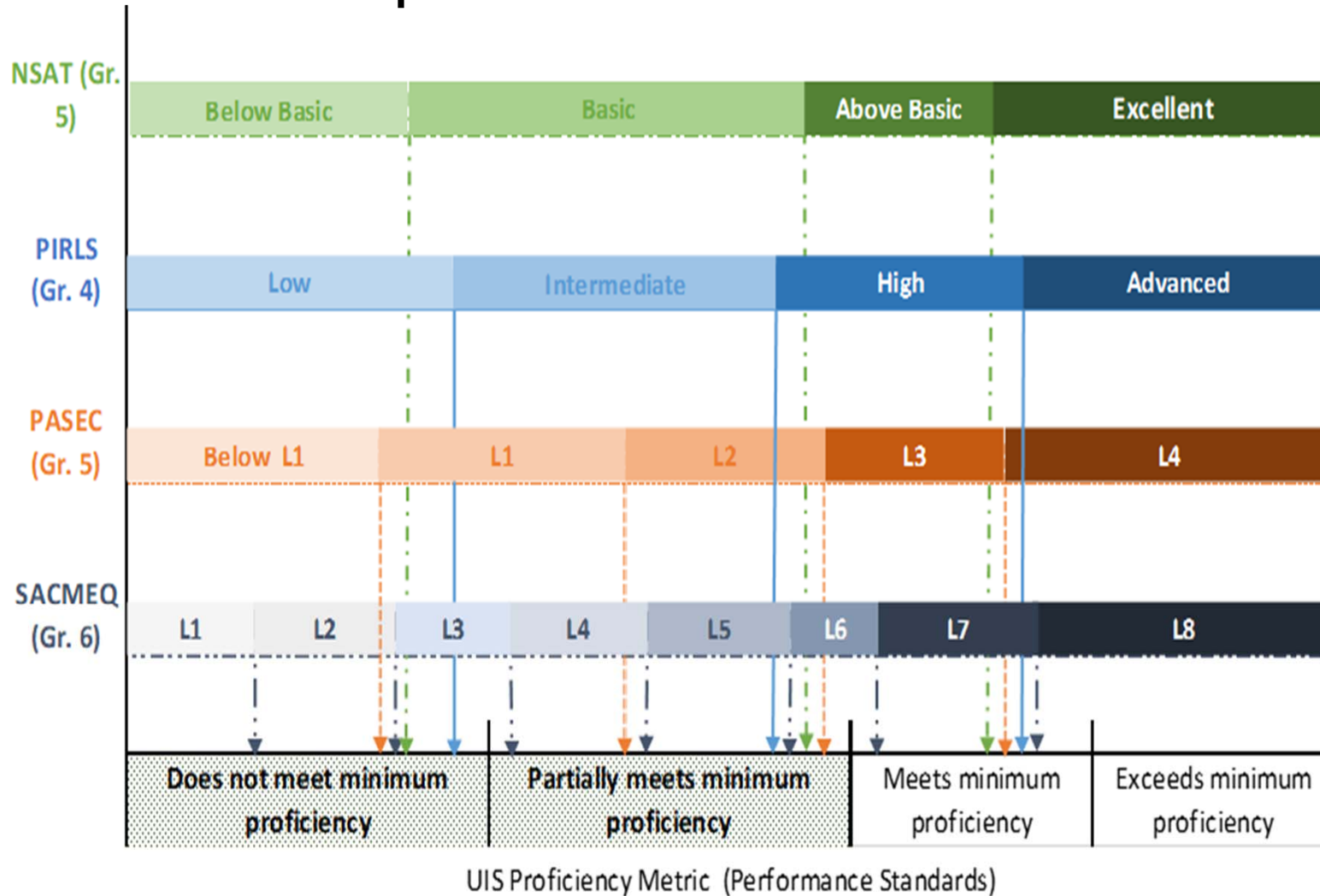


UIS Reporting Levels: the objective

Does not meet minimum proficiency	Partially meets minimum proficiency	Meets minimum proficiency	Exceeds minimum proficiency
--	--	----------------------------------	------------------------------------



Linking the UIS proficiency metric with national and cross-national assessments: An example





Construction of the UIS Reporting Scale

Step 1: Define common content standards

Steps 2 and 3: Define number of performance levels, determine labels, and write policy descriptions for the levels of the UIS-PM

Step 4: Develop full descriptions for the performance levels of the UIS-PM

Step 5: Evaluate alignment of Performance-Level Descriptors

Step 6: Set socially moderated performance standards for national and cross-national assessments

Step 7: Psychometrically link to a common scale

Advantages and Disadvantages of SDG 4 Benchmarking

Assessing international and national educational progress

Drivers of educational change

International comparisons

Means of achieving the set targets

Inconsistent definitions

Pressure on countries, particularly on developing nations

Creation of 'horse-race' mentality

Ambiguities

Choice of criteria, indicators and standards

Variations in commitment and effort

Narrowing efforts towards what is "achievable"

Other concerns

Pitfalls in international benchmarking

Cluster	Background conditions	Resulting in pitfall
Choice of benchmarking approach	International benchmarking can only be done on a consensual basis, no coercion	(1) Mismatch: Choice for hierarchical, disciplinary standards and/or results (functional) benchmarking without corresponding coercion mechanisms
Selection of criteria, indicators	Multitude of relevant criteria and objectives (inherent to complex policies and policy systems)	(2) Pick-and-mix approach to benchmarking
	Disagreement on criteria due to national diversity in preferences	(3) Construction of common objectives is disguised as benchmarking
	Choice of peers/partners is institutionally determined	(4) Inclusion of irrelevant benchmarking partners
	Data availability problems	(5a) Over-reliance on indicators that are easily available, but may not be relevant to the criteria at hand (5b) Over-reliance on quantitative data
Policy transfer	Complexity of policies and policy systems, limited amount of indicators taken into account	(6a) Uninformed transfers
	Complexity of policy systems, and diversity in national institutional contexts	(6b) Incomplete transfers
	Diversity of preferences	(6c) Inappropriate transfers

Guiding Principles

- Taking a Balanced Approach
- Using Universally Applicable and Consistent Definitions
- Differences in Starting Points and National Capabilities

To decide

- Global Vs. Other Levels
 - Difficult to find a global level of reference and politically difficult
 - Easier and probably more relevant at the regional level
 - National level
- All indicators or only a few of them
- Process for building benchmarks
- Absolute v.s. Relative
- Technical definitions

Recommendations

- Indicators that already have a target set
- Indicators that have a benchmark
- Global v.s. Other Levels
 - Overlapping
 - Encourage the setting of regional benchmarking
 - With care about a participatory processes
- Absolute v.s. Relative
- Differences in Starting Points and National Capabilities



United Nations
Educational, Scientific and
Cultural Organization



UNESCO
INSTITUTE
FOR
STATISTICS



TECHNICAL
COOPERATION
GROUP



Thank you!

Silvia Montoya, Director, UNESCO
Institute for Statistics

@Montoya_sil

Learn more: <http://uis.unesco.org/>

 @UNESCOstat