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

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Implications</th>
</tr>
</thead>
</table>
| **Strategy 1: use of national assessments to measure SDG4 with adjustments using international assessments. To be implemented in the short run** | • High levels of external validity for measuring the minimum level of competency established in official curriculum.  
• Low levels of international comparability |
| **Strategy 2: equating among international and regional assessments. To be implemented in the medium run** | • 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)

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Implications</th>
</tr>
</thead>
</table>
| **Strategy 3: equating between different international evaluations aiming at similar school grades. To be implemented in the medium or long run** | - 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. |
| **Strategy 4: creating a Worldwide Proficiency Assessment on Numeracy and Literacy. To be implemented in the long run.** | - 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

<table>
<thead>
<tr>
<th>Does not meet minimum proficiency</th>
<th>Partially meets minimum proficiency</th>
<th>Meets minimum proficiency</th>
<th>Exceeds minimum proficiency</th>
</tr>
</thead>
</table>

Benchmarking and minimum thresholds
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**
- Inconsistent definitions
- Pressure on countries, particularly on developing nations
- Creation of ‘horse-race’ mentality

**Drivers of educational change**
- Ambiguities
- Choice of criteria, indicators and standards
- Variations in commitment and effort

**International comparisons**
- Narrowing efforts towards what is “achievable”

**Means of achieving the set targets**
- Other concerns
### Pitfalls in international benchmarking

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

Silvia Montoya, Director, UNESCO Institute for Statistics

@Montoya_sil

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

@UNESCOStat