



United Nations
Educational, Scientific and
Cultural Organization



UNESCO
INSTITUTE
FOR
STATISTICS

Expanding Coverage in Indicator 4.1

Linking

Silvia Montoya, UIS Director

Junio 2017



1. Main Issues in addressing linking
2. Why linking is still unsolved
3. The implication of the 2030 agenda for linking
4. The UIS Reporting Scale
5. Phases I and II
6. Options to facilitate linking
7. Issues for discussion

Existing national learning assessment might not be able to use for global reporting because data are not comparable in various dimensions:

- Varied in construct and content learning domain coverage
- Varied in quality assurance undertaken in country
- Varied in test design, item format
- Varied in operational and implementation that might affect quality
- Different in scaling methodology that affect reporting metric
- Different data analysis undertaken that might affect outcomes

Why comparable issues do not solve?

- It has been difficult to find agreement on sharing items which make it hard to develop a scale that can span across wide ability continuum.
- Need to have a pool of items contributes from different cross-national especially from regional assessments since it will cover a wider regional and cultural contexts.
- Need to cover some aspects that international assessment might not have covered and this information could eventually only come from Low Income Countries (LIC)
- Need of a systematic way to harmonize the differences.

- Reporting is in one cut-point (achieve or not achieve minimum level) and not reporting on range of scores. Implies
 - Relax the reliability and accept bigger error of measurement?
- Reporting is group (or sub-group) performance and not individual performance. Implies
 - On linking methodology that is not that rigid and strict?
 - A bigger sampling error?

- No matter what methodology that is used there are assumptions need to be met.
 - Learning domains assess and target population needs to be similar to have valid outcomes.
- Needs good design, systematic implementation to ensure data collected met minimum quality before analysis and linking.
 - Compliance with a minimum standard of quality

Depends on data availability and strength of comparability

- **Item Response Theory (IRT) calibration (ACER's UIS Reporting Scale)**
 - Collect item responses to a wide set of items from a wide range of students and from diverse group of countries, calibrate the item responses and empirically establish a reporting scale.
 - Once items are calibrated if these set of items are used in assessments, bringing the score of those assessments onto this reporting scale.
- **Equi-percentile**
 - Two sets of metric, convert both assessment scores into percentile and identify the score of Assessment 1 on Assessment 2 with the same percentile.
 - Needs a reference scale. Could be done if two international assessment have different IRT reporting scale.

- Identify key elements of knowledge, competence, growth ('frameworks' for mathematics, reading)
- Locate a wide range of diverse items that embody the ability of interest
- Analyse and describe cognitive demand of items
- Use item difficulty estimates to order the items along the metric
- Define regions (levels), and describe typical accomplishments in each region; items to illustrate

Phase I Inputs

- Existing assessment datasets -
 - relative difficulties of items within assessment programs
 - information about how particular items behave in particular linguistic, cultural and curriculum contexts.
- Views of expert test developers –
 - judgement on relative difficulties of items across assessment programs
 - thoughts on how particular items might behave differently across different linguistic, cultural and curriculum contexts

- Phase I not enough: For a robust reporting scale test, check, and further explore the test developer judgements and thoughts.
- Need for a new data collection to obtain empirical evidence from children's responses.
- Two approaches:
 - **Test-based:** administer assessments in their intact forms, either to the same group of children or equivalent groups of children
 - **Item-based:** draw on items from all assessments to create new test forms with items in common; administer new test forms to non-equivalent groups
- **Implications: Time is needed for Phase II**

Backward-linking cross-national assessments (CNAs)

- connect results using items and responses from past assessment to build the common denominator.”

Forward-linking CNAs:

- link future cycles of CNAs redesigning the assessment in a way that makes them comparable. Implies a core common framework to allow to link assessments and have new reporting metrics.”

Enhance an assessment:

- expand or enhance the framework of an existing assessment at the cross national level taking advantage of coverage in order to include countries or regions currently not covered.

New test- Reference benchmark

- develop a new assessment specific to a given target population, limited to a given number of domains to make it available.

Option	Cost	Technical Difficulties	Rank
New Test - Reference benchmark	5	5	5
Backward-linking cross-national assessments (CNAs)	4	4	4
Forward-linking CNAs	2	3	2
Forward-linking national assessments	3	4	3
Expand (enhance) Assessment	1	3	1

- What are the issues and challenges in the approach proposed?
- Start by agreeing on linking at least conceptually?
- Why not to agree on some common ground? Alignment in contents/constructs?
- To what extent does failure to meet assumptions underlying the method render the outcomes invalid/unusable?
- Can we discuss a preliminary strategy?



United Nations
Educational, Scientific and
Cultural Organization



UNESCO
INSTITUTE
FOR
STATISTICS

Linking

Silvia Montoya

UIS Director

s.montoya@unesco.org

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



@UNESCOstat