



**unesco**

Institute for Statistics



TECHNICAL  
COOPERATION  
GROUP

**10<sup>th</sup> TCG MEETING**

# Reporting indicator 4.1.1

Workplan to address coverage Indicator 4.1.1a



**2024 CONFERENCE ON  
EDUCATION DATA  
AND STATISTICS**

- ▶ Coverage
- ▶ Reporting 4.1.1
- ▶ Criteria proposed to report 4.1.1a
- ▶ Parameters to reporting indicator 4.1.1.a – proposal
- ▶ TCG decisions

## SDG 4.1.1 indicators

Assessing learning  
**progression from  
foundational  
through early  
secondary years**



### SDG Indicator 4.1.1

Proportion of children and young people

(a) in grades 2/3; (b) at the end of primary; (c) at the end of lower secondary education

Using **globally  
agreed benchmarks  
as reference of  
what a child should  
know**



achieving at least a **minimum proficiency level** in

In **learning areas  
universally  
accepted as  
critical**



(i) reading and  
(ii) mathematics,  
by sex

## Aim: make some inferences from different tools

- “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.”

- Different countries
- Same domain
- Different framework
- Different assessment
- Similar inference



### **How to link or compare using same reference ?**

Need of an agreed definition or global reference

- a competency
- independent of a particular assessment framework
- independent of specific items or tests

# Why standards matter: linking through common definitions

## – MPL and GPF

---

- ▶ The minimum proficiency level (MPL) is the benchmark of basic knowledge in a domain (mathematics, reading) at a given age/grade measured through learning assessments
  - No need to use a single tool to be comparable, but they need to be able to identify the MPL
- ▶ The Global Proficiency Framework (GPF) provides internationally accepted definitions of reading and mathematics constructs
  - Defines up to four Proficiency Levels for each grade and domain to help identify a learning transition

## Take advantage of existing experience in cross national comparability

- ▶ Most of that at end of Primary and end of Lower secondary

## Reporting is based on

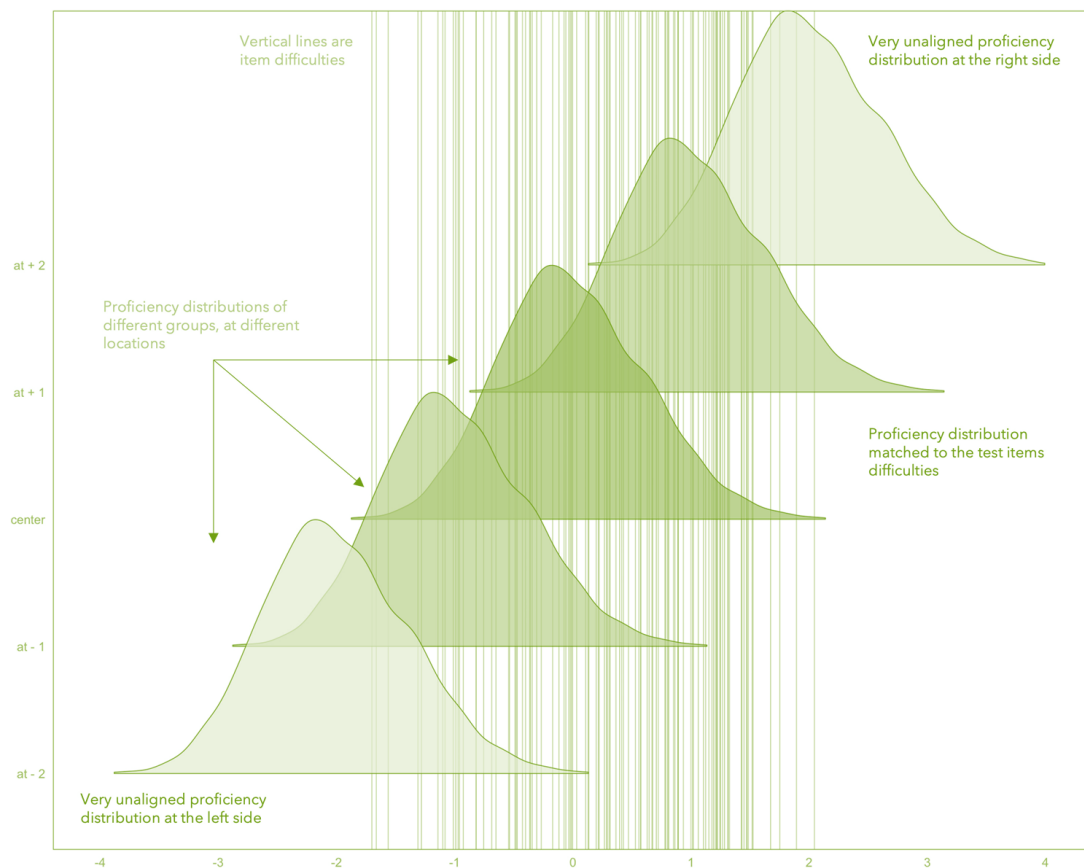
- ▶ Cross National
- ▶ Regional
- ▶ National assessments aligned through psychometric methods (a module aligned with MPL)
- ▶ Coverage growing over time

## Solution demanded a multi-stage process

- ▶ Definition of the MPL
- ▶ Identification of the MPL in the cross-national program
- ▶ Establish linking strategies
  - ▷ Rosetta Stone Concordance table expressed one scale into the other for IEA's (TIMSS/PIRLS) & ERCE & PASEC
  - ▷ Modules calibrated to the MPL (AMPL)

# The higher the heterogeneity of countries, the higher the challenges of the different tools and the suitability for reporting in a reliable and valid manner

## International assessment range ability distributions



## ERCE range of ability distributions

D. Carrasco et al.

International Journal of Educational Development 102 (2023) 102867

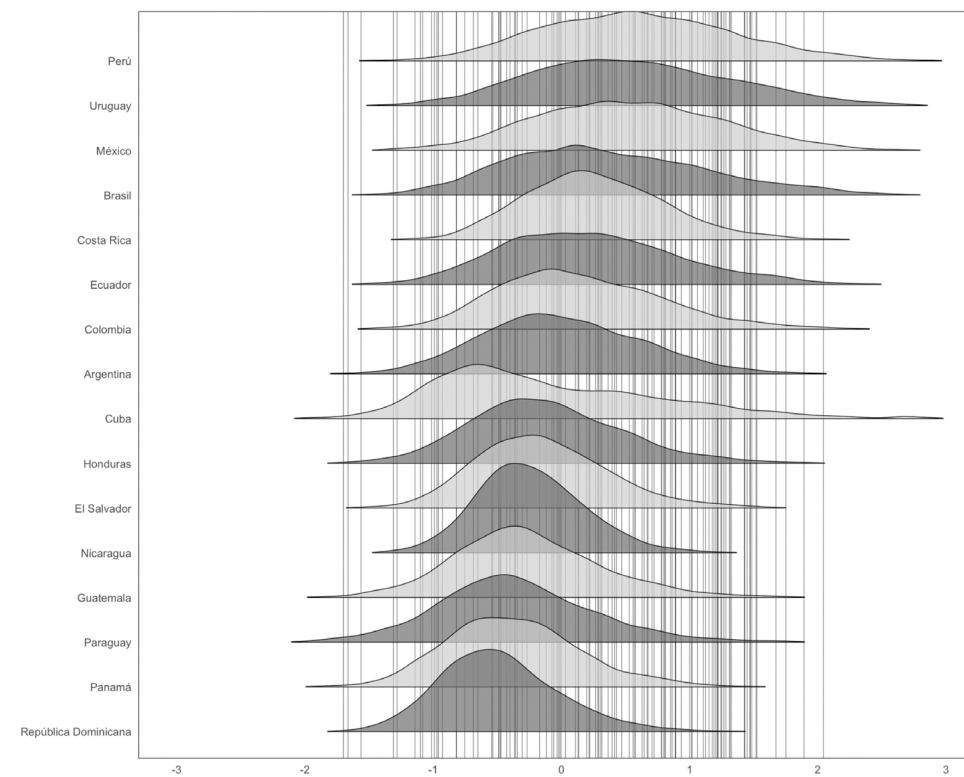
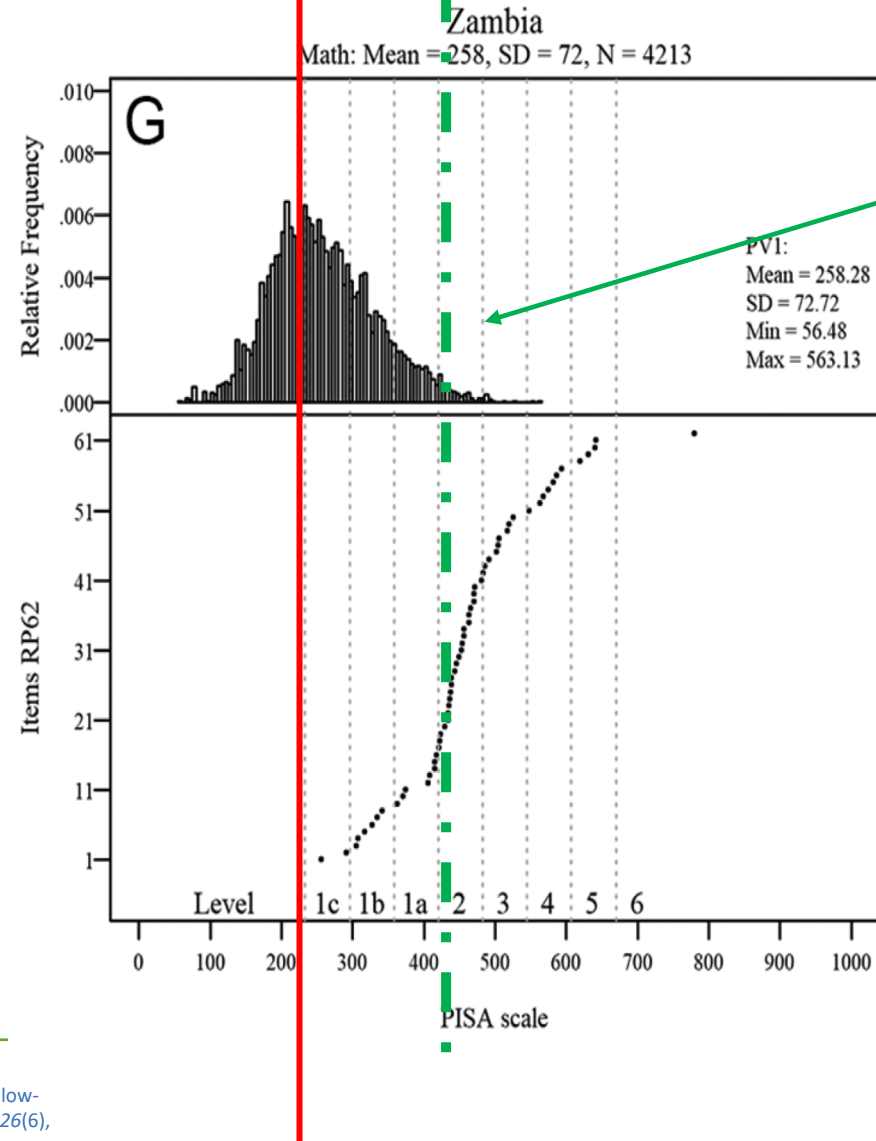


Fig. 2. Countries distribution of the person estimates, for sixth graders Math test, and item location estimates.

# Floor effects reduce the usability for policy and fitness for purpose in many countries

- ▶ The **further away is a participating country from a test mean difficulty the less reliable are its scores.**
- ▶ When the ability of a target population is lower than the assemble test, we can expect floor effects.
- ▶ These are scenarios when the test won't be as informative for the lower ability group, because we lack items in said location.
  - ▶ Zambia in PISA-D, where there were no items below the Zambian students means (Rutkowski et al., 2021).



%  
above  
MPL



## Reporting status - 4.1.1a Early Grades

---

- ▶ Less experience/interest on cross national comparability
- ▶ More complex
  - a. Differences of languages
  - b. Need individual administration to be useful for policy making
    - costlier for national representative samples so benefits must exceed costs
    - not justifiable if most of children population that are well above precursory skills
  - c. Group administration it is not allowing the identification of precursory skills distribution preventing policy utility
  - d. Needs to have solution for comparability (with the MPL) that accommodates group and individual administration assessments

## Status of reporting - 4.1.1a Early Grades

---

- ▶ 2 cross regional experiences
  - ▷ ERCE (group administration in 3rd grade in SP and Portuguese)
  - ▷ PASEC (individual administration on 2nd grade)
- ▶ Other tools
  - ▷ Have not been designed for cross comparability and some not even reinforced the overtime comparability that makes a challenges to use data
  - ▷ Representativeness
  - ▷ Government ownership
  - ▷ Data custodianship is heterogeneous and uncertain in many cases

## Progress to date and next steps (early grades)

---

### Agreement that needs

- ▶ Content coverage
- ▶ Benchmark by language and skills and other clarifications needed regarding reporting process
- ▶ Documentation and technical procedures
- ▶ Sampling

### Solutions

- ▶ MPL agreed in 2018
- ▶ Identification of MPL in the 2 Cross National program
  - GPF and criteria to align defined
- ▶ Policy linking (and pairwise) allows benchmarks setting (cut-off points)

# Parameters to reporting indicator 4.1.1.a - proposal

The criteria were selected to ensure consideration of the quality of the assessment instrument and its implementation and have been agreed by partners:

- ▶ **Criterion 1** – is the assessment sufficiently aligned to the MPL?
- ▶ **Criterion 2** – is there evidence that the items in the assessment have been reviewed qualitatively and quantitatively to determine their suitability for inclusion in the assessment?
- ▶ **Criterion 3** – is the sample of learners that took the assessment representative of the population against which the results will be reporting?
- ▶ **Criterion 4** – is there evidence that the assessment was administered in a standardised way?
- ▶ **Criterion 5** – are the outcomes of the assessment sufficiently reliable?

All assessments must contain a minimum of 20 items

- ▶ **Option 1** : require measurement of all precursor skills such as decoding or fluency.
  - higher- and upper-middle income countries, and countries with relatively easy languages, see this as highly unnecessary as their children are almost always advanced beyond this.
  - Assessments that have already been accepted as valid do NOT measure these precursor skills.
  - Reading science suggests (but does not prove) that children who comprehend must have the precursor skills.
  
- ▶ **Option 2** : requires ONLY reading comprehension (as described in the MPL and GPF) be used as a sine qua non requirement for assessments,
  - The precursor skills are an needed and useful addition in assessments aimed at LICs and LMICs, since they help countries understand, if their children are NOT understanding, why they are not understanding, in that they may lack the precursor skills.

All assessments must contain a minimum of 20 items

- ▶ **Reading** – minimum 10 score-points assessing reading comprehension and the assessment must cover both reading comprehension subconstructs at grade 2 in the GPF. The remaining items can be drawn from any of the domains (**decoding, listening comprehension or reading comprehension**).
  
- ▶ **Mathematics** –minimum 10 score-points assessing number and operations and the assessment must cover all four number and operations subconstructs at grade 2 in the GPF. The remaining items can be drawn from any of the domains (**number and operations, measurement, geometry, statistics and probability or algebra**).

## Parameter 2: scoring model for individual-based assessments that use different assessment approaches

---

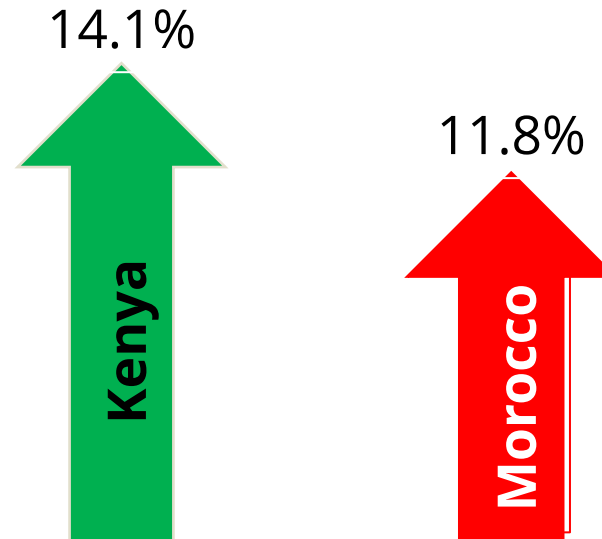
For instance, standard items in reading comprehension but ‘read a passage’ in decoding, such that combining sections into **an overall score is not meaningful**

- ▶ Students meeting global minimum proficiency level should be determined either using a compensatory or a conjunctive scoring method
  - In a **compensatory model**, it is assumed *that strong performance in one skill can make up for weak performance in another skill ( e.g. students achieve at least 31 out of 78 marks)*
  - In a **conjunctive model**, students must achieve *a specified level of performance on each skill to be classified at that proficiency level (Students obtained at least 4 in LC, 24 in OR, and 3 in RC)*



## Parameter 2: scoring model for individual-based assessments that use different assessment approaches (2)

- ▶ A conjunctive scoring model ensures that results from diverse assessments aligned with MPL/GPF can be compared effectively.
- ▶ Classified students with the same minimum knowledge and skills across countries into the meeting global MPL, irrespective of the differences in their learning assessments and conditions of their learning.



Proportion of children in **grade 2** achieving **at least** a minimum proficiency level in reading (accurately read aloud and understand written words from familiar contexts. They retrieve explicit information from very short texts. When listening to slightly longer texts, they make simple inferences).

## Parameter 2: scoring model for individual-based assessments that use different assessment approaches(3)

- ▶ Assessments measure three distinct skills in reading (**listening comprehension, decoding, and reading comprehension**) that cannot be combined into an overall score,
  - benchmarks should be set for each skill separately
- ▶ **Once benchmarks are set**, student learning **progress is tracked over time** and **standard setting process does not need to be repeated for this purpose.**
- ▶ For countries **with multiple official languages of instruction**, a three-step process :
  1. Benchmarks should be set for each subtask and for each language separately
  2. Calculate the percentage of students meeting the MPL for each language
  3. Aggregate the percentages of meeting the MPL across the languages (through sampling weights) for SDG 4.1.1a reporting.
- ▶ Cross-country comparisons assessments' **validity, reliability and consistent** score interpretations, **are integral part of the reporting**

## Snapshot of guidance for reporting SDG 4.1.1a

### ▶ Eligibility criteria

- ▶ Adopt the five criteria checklist
- ▶ The **Standards book** offers comprehensive guidelines, specifying the detailed requirements for assessments to ensure their reliability, validity, and fairness.
- ▶ Documentation should be public

### ▶ Scoring

- ▶ A conjunctive scoring model should be adopted to compare effectively.
- ▶ Benchmarks should be set for each skill separately
- ▶ Once benchmarks are set, student learning progress is tracked over time and no standard setting process needs repetition
- ▶ For countries **with multiple official languages of instruction**, a three-step process is recommended score

### ▶ Guidelines to be produced

- ▶ Consumer's guide to alternative methodologies for linking with costing
- ▶ Protocol will confirm the hierarchy of methodologies for linking
- ▶ A users' guide to the conjunctive scoring model should be prepared and disseminated
- ▶ A comparison between Policy Linking and Pairwise Comparison should be elaborated to guide the selection of either of the 2 methodologies
- ▶ Assessment blueprint recommended

## Use of past data : challenge

- ▶ Data that has been collected in foundational level that they should run the criteria checklist and proof suitability for reporting
- ▶ If the aspect regarding past data must run criteria check and exploration is a case-by-case due to the lack of standardization in tools that in some cases are not aimed at comparability

| Issues                         | Explore                                                                                                                                                                                                                       |
|--------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Expansion to population</b> | Reporting and estimation of sample size, CI or relevant age population, CI size accepted                                                                                                                                      |
| <b>Content coverage</b>        | Partial coverage present a thorough analysis with research                                                                                                                                                                    |
| <b>Documentation</b>           | Countries undertaking these assessments are required to thoroughly document the processes and procedures involved in test development or adaptation, as well as sampling and the standardized test administration procedures. |
| <b>Microdata not available</b> | Make microdata available and public                                                                                                                                                                                           |
| <b>Stop Rule</b>               | Support evidence on reliability and impact                                                                                                                                                                                    |

|               | Future                                                                                                                                                                                                                                                                                                                                                        | Past                                                                                                                                                                                                             |
|---------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Content       | <p>Five criteria presented</p> <p>The development and implementation of EGRA, FLM-MICS, PAL-ICARE, and PASEC are poised to meet all the requirements outlined in the policy linking toolkit for SDG 4.1.1a.</p>                                                                                                                                               | <p>Check on compliance with 5 criteria</p> <p>Collect documentation.</p> <p>Estimate CI, Parameters<br/>(assessments' validity and reliability, along with consistent score interpretations, become crucial)</p> |
| Sampling      | Sampling frame and post sampling treatment needs to prove relevant population expansion will happen                                                                                                                                                                                                                                                           |                                                                                                                                                                                                                  |
| Benchmark     | <ul style="list-style-type: none"> <li>○ Content coverage</li> <li>○ Benchmark by language and skills</li> <li>○ Aggregation method conjunctive</li> <li>○ Guidance on tracking over time</li> </ul>                                                                                                                                                          | <p>Do the analysis of past data and against 5 criteria</p> <p>Estimate size of the CI of the lack of compliance</p>                                                                                              |
| Documentation | <p>Countries are required to thoroughly document the processes and procedures involved in test development or adaptation, sampling and standardized test administration procedures.</p> <p>Failure to meet the criteria (e.g., insufficient documentation or non-nationally representative sample) then that administration cannot be used for reporting.</p> |                                                                                                                                                                                                                  |
| Test security | To enhance the security of test instruments, PAL and MICS may consider developing multiple pre-                                                                                                                                                                                                                                                               |                                                                                                                                                                                                                  |

# TCG decisions

# Decision point L1

## Adopt the eligibility criteria for reporting indicator 4.1.1

|                          | Option 1                                                                                                                                                                                                                                                                                                                          | Option 2                                                                                                                                                                                                                       |
|--------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Description</b>       | Approve the eligibility criteria present in <a href="#">“Eligibility criteria for reporting SDG 4.1.1”</a>                                                                                                                                                                                                                        | No action                                                                                                                                                                                                                      |
| <b>Pros</b>              | <ul style="list-style-type: none"><li>- Improve monitoring of SDG target 4.11 in general and indicator 4.1.1a by ensuring objective criteria for the quality of the assessment instrument and its implementation]</li><li>- Guides countries and assessment programs in the implementation of their assessment programs</li></ul> | Favors heterogeneity in the quality of the data produced in terms of content coverage, procedural quality, sampling and post-sampling procedures hindering the production of good quality data and conditioning data coverage. |
| <b>Cons</b>              | None                                                                                                                                                                                                                                                                                                                              | Approach is less rigorous in terms of comparability                                                                                                                                                                            |
| <b>Proposed decision</b> | Option 1 - Approve the eligibility criteria proposed.                                                                                                                                                                                                                                                                             |                                                                                                                                                                                                                                |
| <b>Document</b>          | <a href="#">Eligibility criteria for reporting SDG 4.1.1</a> that complements the hierarchy of reporting criteria for indicator 4.1.1 describe in the protocol for <a href="#">reporting indicator 4.1.1</a>                                                                                                                      |                                                                                                                                                                                                                                |

# Decision point L2

## Content Coverage to report indicator 4.1.1a

|                          | Option 1                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         | Option 2                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |
|--------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Description</b>       | <p>All assessments must contain a minimum of 20 items</p> <ul style="list-style-type: none"> <li>• <b>Reading</b> – minimum of 10 score-points assessing <i>decoding</i>, 5 score-points assessing <i>listening comprehension/comprehension of spoken or signed language</i> and 5 score-points assessing <i>reading comprehension</i>. The assessment must also cover 5 of the 9 subconstructs at grade 2 in the GPF.</li> <li>• <b>Mathematics</b> – minimum of 10 score-points assessing <i>number and operations</i>, 5 score points assessing <i>measurement and geometry</i> and 2 score-points assessing <i>statistics and probability</i> and <i>algebra</i>. The assessment must also cover 7 of the 14 subconstructs at grade 2 in the GPF.</li> </ul> | <p>All assessments must contain a minimum of 20 items</p> <ul style="list-style-type: none"> <li>• <b>Reading</b> – minimum 10 score-points assessing <i>reading comprehension</i> and the assessment must cover both reading comprehension subconstructs at grade 2 in the GPF. The remaining items can be drawn from any of the domains (<i>decoding, listening comprehension or reading comprehension</i>).</li> <li>• <b>Mathematics</b> – minimum 10 score-points assessing <i>number and operations</i> and the assessment must cover all four <i>number and operations</i> subconstructs at grade 2 in the GPF. The remaining items can be drawn from any of the domains (<i>number and operations, measurement, geometry, statistics and probability or algebra</i>).</li> </ul> |
| <b>Pros</b>              | <ul style="list-style-type: none"> <li>• The domain requirements align with the description in the MPL</li> <li>• Encourages countries where learners are still focused on foundational skills in reading to include all domains in their assessments</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 | <ul style="list-style-type: none"> <li>• <i>Maximises the number of assessments that can be used for reporting</i></li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |
| <b>Cons</b>              | <ul style="list-style-type: none"> <li>• Countries <i>that assess with an assessment of reading comprehension only would not be able to report against SDG 4.1.1a</i></li> <li>• Countries <i>that are focused only on foundational skills in mathematics (number and operations) would not be able to report against SDG 4.1.1a</i></li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                  | <ul style="list-style-type: none"> <li>• Assessments being used for reporting are likely to contain different domains outside the core elements required but the global reporting does NOT intend to compare whole assessment but % over the MPL</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |
| <b>Proposed decision</b> | Approve Option 2 (both options rely on a reliability of 0.7)                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |
| <b>Document</b>          | <a href="#">Eligibility criteria for reporting SDG 4.1.1</a>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |



# Decision point L3

## Enable a broader conceptualisation of assessing decoding

|                          | Option 1                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   | Option 2                                                                                                                                                                                                                                                                                                           |
|--------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Description</b>       | <p>Enable a broader conceptualisation of how decoding could be assessed.</p> <p>GPF <b>precision construct</b> can be assessed in other ways than oral fluency.</p> <p><b>Proposed revision</b></p> <p>In a short simple connected text of one or two sentences, students <b>decode</b> most words – including some unfamiliar words.</p> <p>In a short and simple connected text of one or two sentences, students decode most words, including some unfamiliar words with familiar sound–symbol patterns (applies to alphabetic and alpha-syllabic languages only). <b>Decoding skills can be demonstrated in a variety of ways, including through oral fluency.</b></p> | <p>Do not change the unpacked document definition of decoding.</p> <p><b>Unpacked document expanded statement</b></p> <p><i>In a short simple text of one or two sentences, students <b>read aloud</b> most words – including some unfamiliar words – <b>accurately but slowly and often word by word.</b></i></p> |
| <b>Pros</b>              | <p>Current wording is unnecessarily restrictive of ways decoding might be assessed</p> <p>Changes give more flexibility</p> <p>Reading with comprehension implies fluency</p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              | <p>Reliable large-scale assessments of fluency may not be feasible</p>                                                                                                                                                                                                                                             |
| <b>Cons</b>              | <p><i>Asks for understanding of the population to be measured and choice of a fit for purpose tool</i></p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 | <p><i>Reduce the number of assessment to be used</i></p>                                                                                                                                                                                                                                                           |
| <b>Proposed decision</b> | <p>Approve Option 1.</p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |                                                                                                                                                                                                                                                                                                                    |
| <b>Document</b>          | <p><a href="#">MPL</a> and the measurement of decoding.</p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |                                                                                                                                                                                                                                                                                                                    |

## Decision point L4

Adoption of scoring process scoring model for individual-based assessments that use different assessment approaches

|                          | Option 1                                                                                                                                                           | Option 2                                                                                                                                               |
|--------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Description</b>       | In a <b>compensatory model</b> , it is assumed <i>that strong performance in one skill can make up for weak performance in another skill</i>                       | In a <b>conjunctive model</b> , students must achieve <i>a specified level of performance on each skill to be classified at that proficiency level</i> |
| <b>Pros</b>              |                                                                                                                                                                    | The interpretation of scores across countries will be aligned with MPL, consistent, and comparable.<br>Calculation is easy and simple.                 |
| <b>Cons</b>              | Simple compensatory model cannot be used it has to be weighted and weighting can be confusing<br>Interpretations of scores will not be the same across assessments | --                                                                                                                                                     |
| <b>Proposed decision</b> | Recommend the Option 2 conjunctive model to scoring for individual-based assessments that use very different assessment approaches                                 |                                                                                                                                                        |
| <b>Document</b>          | <a href="#"><u>Quality measures of individual administered assessments</u></a>                                                                                     |                                                                                                                                                        |

# Thank you

Learn more: <https://tcg.uis.unesco.org/10th-meeting-of-the-tcg/>  
<https://ces.uis.unesco.org/>

 @UNESCO