



unesco

Institute for Statistics

Technical Advisory Group Indicator 4.1.1a

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 or 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

The complexities of early grade reporting

- ▶ Language of administration is critical and the challenges to establish comparability demands parameters in many more languages than the other 2 points of measurement
- ▶ Different types of administration co-exist (unlike the other levels):
 - individual
 - group
- ▶ Individual administration aims to be useful for policy making. However:
 - costlier for national representative samples if not needed
 - not justifiable if most of children in population are above precursory skills
- ▶ Group administered tests are more cost-effective and appropriate to some contexts. However, they do not allow the identification of precursor skills and then might have a more limited policy utility for some very low skills context.

Key issues to address for individual based assessments

- ▶ Find a minimum acceptable criteria for reporting that ensure comparability at the MPL for different types of assessments (individual/group administration)/different content coverage;
- ▶ Determine location in the learning progression continuum to facilitate comparability even when they are not capable to report at the MPL (define what are the skills that need scoring);
- ▶ Identify benchmarks per language for each of the precursor skills
 - ▷ These benchmarks per skills and languages are not only relevant for reporting but also serve as guidelines for countries to measure
- ▶ A scoring model for individual based assessment
- ▶ Comply with all the other aspects related to reporting

Status of reporting - 4.1.1a Early Grades

- ▶ 3 cross regional experiences with centralized quality assurance
 - ▶ ERCE (group administration in 3rd grade in Spanish and Portuguese)
 - ▶ PASEC (individual administration in 2nd grade)
 - ▶ AMPL (group administration at end of lower primary)

Individual based assessment challenges

- ▶ Measurement is not covering the sub domains needed and benchmarks for each of the skills are different precluding the comparability
- ▶ Have not been designed for cross country comparability and some of the administration has not been designed for comparability over time that makes a challenge to use data
- ▶ Issues with school-age population representativeness
- ▶ Data custodianship is heterogeneous and uncertain about processes in many cases with no central quality control and disperse documentation
- ▶ Tool are publicly available.
- ▶ Data is not available, not only publicly but also to the custodian agency
- ▶ Government ownership

Progress to date (1):

Eligibility criteria for reporting indicator 4.1.1a

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.

▶ **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**).

▷ Refinement of decoding

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). **Decoding skills can be demonstrated in a variety of ways, including through oral fluency.**

▶ **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**).

Progress to date (3): Scoring for SDG 4.1.1a

▶ 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

What is missing to make feasible 4.1.1.a MPL reporting ?

- ▶ Some parameters for individual administered test to be compliant with the content coverage and reliability:
 - ▶ define the set of precursors skills that with reading comprehension would be available to report for indicator 4.1.1 MPL
 - ▶ define under what conditions the content coverage for each of the sub skills would be enough to report
 - ▶ an identification of the benchmarks for the skills that are universal and tool-independent for each of the languages of major language family
 - ▶ This would allow to fill the table the UIS has proposed and are central to provide the standards needed for countries and the international community to act and that would allow the comparability between data point and over time today not feasible

UIS proposed table to fill based on available evidence

Sub-construct	Language group 1	Language group 2	Etc.	Etc.
Passage comprehension	X	X	X	X
Oral reading fluency	X	X	X	X
Reading comprehension	X	X	X	X
Etc.	X	X	X	X
Etc.	X	X	X	X

UIS proposed reporting of Indicator 4.1.1.a disaggregated

Sub-construct	Percentage
Oral reading fluency (ORF)	% above ORF benchmark
Reading comprehension (RC)	% above RC benchmark
Decoding (D)	% above D benchmark
Listening Comprehension (LC)	% above LC benchmark
MPL	% above the MPL t, s

- ▶ this is a necessary step that would allow the reporting of 4.1.1.a possible at the MPL
- ▶ limit the comparability to the cases where the comparability is possible
- ▶ allow the disaggregation by skills that could help to guide policy making
- ▶ make clear the assessment tools that are not fit-for-purpose of reporting indicator 4.1.1.a MPL guiding their progress

Thank you

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

 @UNESCO