unesco
Institute for Statistics

GLOBAL ALLIANCE TO MONITOR LEARNING

PROTOCOL FOR REPORTING:

SDG INDICATOR 4.1.1

(2023)

Acronyms and abbreviations

ACER	Australian Council for Educational Research
AMPL	Assessment for Minimum Proficiency Level
CNA	Cross-national assessments
CAT	Content Alignment Tool
EGMA	Early Grade Mathematics Assessment
EGRA	Early Grade Reading Assessment
ERCE	Estudio Regional Comparativo y Explicativo
GAML	Global Alliance for Monitor Learning
GCF	Global Content Framework
IBE	International Bureau of Education
IEA	International Association for the Evaluation of Educational Achievement
IRT	Item response theory
ISCED	International Standard Classification of Education
LLECE	Latin American Laboratory for Assessment of the Quality of Education
MICS	Multiple Indicator Cluster Survey
MILO	Monitoring Impacts on Learning Outcomes
MPL	Minimum Proficiency Level
NAEP	National Assessment of Educational Progress
NLA	National learning assessments
OECD	Organization for Economic Co-operation and Development
PAL	Network People's Action for Learning Network
PASEC	Programme d'analyse des systèmes éducatifs de la confemen
PAT	Procedural Alignment Tool
PILNA	Pacific Island Literacy and Numeracy Assessment
PIRLS	*Progress in International Reading Literacy Study*
PISA	*Programme for International Student Assessment*
PISA D	PISA for development
PLD	Proficiency Level Descriptors
SACMEQ	Southern and Eastern Africa Consortium for Monitoring Educational Quality
SDG	Sustainable Development Goal
SEA-PLM	Southeast Asia Primary Learning Metrics
SEAMEO	Southeast Asian Ministers of Education Organization
TIMSS	Trends in International Mathematics and Science Study
UIS	UNESCO Institute for Statistics

Table of Contents

Acronyms and abbreviations **1**
Executive summary **3**
1. Objectives and Structure 4
2. The Challenges 4
3. Reporting Consistency 5
4. Global Content Framework 6
4.1. Why? 6
4.2. Outputs 6
4.3. Expected Outcome 6
5. Procedural alignment 6
5.1. Why? 6
5.2. How? 7
5.3. Outputs 7
5.4. Expected Outcome 7
6. Proficiency Framework and Minimum Level, Linking Strategies and Interim Reporting 7
6.1. Why? 7
6.2. How? 8
6.3. Outputs 8
6.4. Expected Outputs and Outcome 8
7. Protocol for reporting SDG indicator 4.1.1 in the UIS data release 8
7.1. Alignment to the global MPL: 9
7.2. Selection of reporting source when various sources are available 9
7.3. Mapping of grade to measurement point: 10
7.4. Mapping of domains to reading or mathematics: 10
7.5. Parity index indicators (SDG indicator 4.5.1) 11
7.6. Summary of the reporting protocol 11

List of tables

Table 1. Summary of process and Focus of GAML 5
Table 2. 2019 Indicator 4.1.1 reporting 11

List of figures

Figure 1. Indicator 4.1.1 9

Executive summary

4.1.1 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

The reporting format of the indicator aims to communicate two pieces of information:

I. the percentage of students meeting minimum proficiency standards for the relevant domains (mathematics and reading) for each point of measurement (grades 2/3; end of primary and end of lower secondary); and

II. when different programs can be considered comparable, and the conditions under which the percentage can be considered comparable to the percentage reported from another country.

This requires the following inputs to frame the indicator:

- What contents should be measured and what is the percentage of coverage to be covered by a given assessment to be comparable to others?
- What procedures are good enough to ensure quality of the data collected? and
- A proficiency scale where all assessments could be informed (and its conversion function or the linking procedure), and a definition of the minimum level for each domain that would allow the estimation of the percentage of students achieving the minimum proficiency level.

An ideal program for reporting on SDG4.1.1 will have gone through three steps: Conceptual Framework, Methodological Framework, and a Reporting Framework. Each of these contains several complex sub-steps. For various levels and types of assessment, UIS had completed most of this work before accepting the responsibility of being custodian of reporting on SDG4.1.1.

Acknowledging that much work had already been done, UIS has prioritized and motivated others to carry out work that had not yet been done. The table below, and this document in general, summarize the work to date. This is represented in the second column of the table. The note discusses the focus of UNESCO Institute for Statistics' (UIS) work (second column), and the columns to the right; and presents a protocol that guides how reporting for Indicator 4.1.1 is now implemented.

1. Objectives and Structure

The document will explain the flow of work, the activities and the outputs in the context of Global Alliance for Monitor Learning's (GAML) broader work program for Indicator 4.1.1. We present them in a logical rather than chronological order.

Each of the activities and outputs help build the tools to generate a minimum level of consistency of education systems' reporting against Indicator 4.1.1, while retaining sufficient flexibility for education systems to pursue assessment programs appropriate to their context and needs.

The reporting format aims to communicate two pieces of information:

I. the percentage of students meeting minimum proficiency standards for the relevant domains (mathematics and reading) for each point of measurement (grades 2/3; end of primary and end of lower secondary); and

II. when different programs can be considered comparable and the conditions under which the percentage of programs can be considered comparable to the percentage reported from another country.

This requires the following inputs to frame the indicator:

- What contents should be measured and what is the percentage of coverage covered by a given assessment to be comparable to others? ;
- What procedures are adequate to ensure the quality of the data collected? ; and
- A proficiency scale where all assessments could be informed (and its conversion function or the linking procedure), and a definition of the minimum level for each domain that would allow the estimation of the percentage of students achieving the minimum proficiency level.

Next section defines challenges and Section 3 provides deeper context and sets the logic of workflow. Sections 4, 5 and 6 go deeper in each of the stages of process following same logic and format. Sections 7 and 8 describes the protocol for reporting the indicator to the UIS.

2. The Challenges

The challenges of achieving consistency in global reporting go far beyond the definition of the indicators themselves. In many cases, there is no "one-stop shop" or single source of information for a specific indicator consistent across international contexts. Even when there is agreement on the metric to be used in reporting, a harmonising process may still be necessary to ensure that coverage of the data is consistent.

There are two extremes: at least in theory, the greatest confidence would arise by reporting using a perfectly equated assessment program while, again in theory, the greatest flexibility would arise

if reporting could happen with minimal alignment. Both extremes are unsatisfactory for reasons too complex for this document. UIS's approach is a middle one: allow flexibility of reporting, but with growing alignment and comparability over time, without ever necessarily reaching the extreme of a perfectly equivalent assessment or set of assessments. This would allow any assessment program that follows certain comparability guides ahead of time, as well as certain quality assurance and procedural guides, to report in the relevant domains. This flexible approach implies developing tools to guide countries' work that, if complemented by capacity development activities, will ensure that Indicator 4.1.1 reporting drives knowledge sharing, and growth in global capacity to use assessment programs as levers for system improvement.

3. Reporting Consistency

The objective is to define the criteria and generate the tools that could serve as reference points, transparency tools and normative references.

The tools to be generated have the potential to become a standard against which countries, regions, institutions, international agencies and professionals can benchmark their programs and certificates, and make international comparisons, if they choose to do so. This process already takes place informally in many ways and/or it is now de facto embedded into the various international (and national) assessments.

The workflow is designed following the structure of the implementation of any learning assessment. **Table 1** summarizes the relevant areas of work and contextualizes the work that has taken place and is taking place, with regard to the three main steps in developing a means of reporting on SDG4. This table provides deeper and more detailed context to the introductory materials presented thus far, and highlights the focus of the current work of the UIS and its partners in the last column. It is exactly the same as column 2 in the introductory table above.

Table 1. Summary of process and Focus of GAML

Phase /Tools	What It Addresses	Main Components	Focus of UIS Work
Conceptual Framework	**What to assess? - Concept** Who to assess? – Population: in and out of school? What contextual information to collect?	• **Domain and subdomain: Minimum coverage** • Target population • Background Questionnaire	• **Global Content Framework (GCF)** • **Content Alignment Tool (CAT)**
Methodological Framework	**What are the procedures for data integrity**	• Test design • Sampling frame • Operational design • Data generation • Data analysis	• **Good practices guidance** • **Procedural Alignment Tool**
Reporting Framework	**What format to report?** **What is the minimum level?** **How to link or "harmonize"?**	• Reporting model • Scale or proficiency framework • Linking	• **Proficiency Framework and minimum level** • **Linking strategies**

		• Definition of an interim reporting strategy	• **Interim Reporting strategy**

Source: UNESCO Institute for Statistics (UIS).

4. Global Content Framework

This section describes in more detail the work that needs to be done, or is being done, for **row 1 of column 3, in Table 1** above.

4.1. Why?

Assessment programs differ in their conceptual frameworks. For example, depending on the curriculum in a country, national assessments usually have different content coverage for a given grade. Domains can be defined differently and, in some cases, programs assess different skills, use different content to assess the same domain, and do both differently, even for the same grade.

To assess the degree of alignment among various assessments and to begin to lay out the basis for a global comparison, UIS and the International Bureau of Education (IBE-UNESCO) collaboratively developed a Global Content Framework (GCF) for the domains of Mathematics and Reading.

4.2. Outputs

There are three final products:

1. Global Content Framework (GCF) of Mathematics and Reading to serve as reference;
2. Content Alignment Tool (CAT) including alignment criteria;
3. A platform to help countries self-assess.

4.3. Expected Outcome

To ensure data integrity with respect to minimum comparability.

5. Procedural alignment

This section describes in more detail the work that needs to be done, or is being done, for **row 2 of column 3, in Table 1** above.

5.1. Why?

Robust, consistent operations and procedures are an essential part of any large-scale assessment, to maximise data quality and minimise the impact of procedural variation on results. Examples of

procedural standards may be found in all large-scale international assessments, and for many large-scale assessments at regional level, where the goal is to establish procedural consistency across international contexts. Many national assessments also set out clear procedural guidelines, to support consistency in their operationalization.

Assessment implementation faces many methodological decisions including test formats and sampling decisions. There is no need for identical procedures and format across assessments. However, there is a need for a minimum set of procedures so data integrity is protected, and results are robust as well as reasonably comparable for any given country over time, but also across countries at any given point in time.

5.2. How?

By defining the minimum procedures that ensure data integrity sufficient enough to report and compare results from different assessment programs.

5.3. Outputs

1. Manual of Good Practices in Learning Assessment
2. Quick Guide: Making the Case for a Learning Assessment
3. Quick Guide: Implementing a National Learning Assessment
4. Procedural Alignment Tool
5. Online procedural alignment tool platform

5.4. Expected Outcome

Comparable data from the point of view of procedures.

6. Proficiency Framework and Minimum Level, Linking Strategies and Interim Reporting

6.1. Why?

This section describes in more detail the work that needs to be done, or is being done, for **row 3 of column 3, in Table 1** above**.**

Assessment programs typically report using different scales. Analysis of results therefore remains contained to one particular test, methodology and scale. While methodologies tend to converge between international and regional assessments, it is still difficult to situate assessments in a common reference continuum of learning outcomes for each level and domain.

The most important issue in the definition of the scales are the proficiency benchmarks or levels embedded within the numerical scale and their cut points on that numerical scale. These benchmarks are typically associated with Proficiency Level Descriptors, which describe in some detail the skills that are typical of students at any given cut point in the scale. Typically, an

overarching policy statement or policy definition gives meaning to the succession of cut scores and the proficiency levels but most importantly for defining what constitutes a *minimum* (which is what the SDG4.1.1. indicators call for) proficiency level that has reference to the content.[1]

After the Consensus Building Meeting in 2018, an interim reporting period of 3 years was initiated to allow the use of all assessments to report against 4.1.1, regardless of whether they were school-based or aligned to the Global Minimum proficiency levels. Upon the conclusion of the interim reporting period, the UIS ceased publishing learning outcomes from assessments not aligned to the global minimum proficiency levels.

6.2. How?

A scale where all the learning assessment programs could be located and the definition of a linking strategy to that scale. The definition of a scale implies:

- A metric that is arbitrary
- The definition of a set of proficiency levels or benchmark including the minimum level
- The policy statements associated to the sets of benchmarks

6.3. Outputs

The final products are:

1. A scale for each domain and point of measurement (benchmarks and definition of the minimum proficiency level or each domain and point of measurement)
2. A portfolio of linking strategies and the tools that allow to locate assessments proficiency *levels in a scale*

6.4. Expected Outputs and Outcome

A proficiency scale that involves the definition of performance levels that are required of students to be proficient, the definition of the number of performance levels, determining the labels and writing descriptions for the levels of the proficiency metric.[2]

7. Protocol for reporting SDG indicator 4.1.1 in the UIS data release

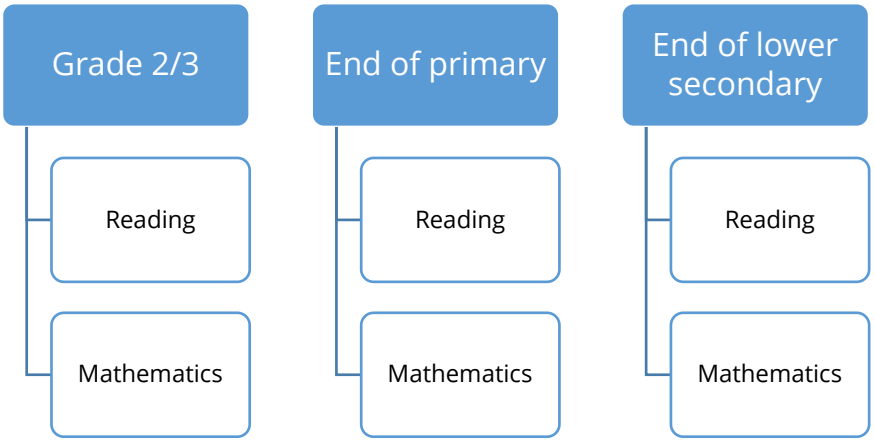
Indicator 4.1.1 has three points of measurement and two domains resulting in 6 sub-indicators as reflected in the figure below.

4.1.1 Indicator comprises the following 6 indicators:

[1] Taking from the National Assessment of Educational Progress (NAEP) on policy statement: "*Policy definitions are general statements to give meaning to the levels*."

[2] The initial development of the reporting proficiency scale would draw from both experts' opinion and analysis of existing data and policy level descriptors.

Figure 1. Indicator 4.1.1



Source: UNESCO Institute for Statistics (UIS)

7.1. Alignment to the global MPL:

1. For each assessment, programme alignment should be completed using:
    a. CNA: according to agreed alignment from the consensus meeting and its revision
    b. National learning assessment (NLA): provided country does not have any CNA and use statistical linking.
    c. Other assessments (e.g., MICS, EGRA/EGMA, Young Lives, etc.): proficient level is pending and subject to harmonization.

7.2. Selection of reporting source when various sources are available

For each of the indicators listed above for global reporting, the sources of data selected should be prioritized according to the following order of assessments, providing that a mapping of grade has guided a first selection of sources:

i. International assessments
ii. Regional assessments

The assessment which maps the best to the grade will always be the preferred choice.

Here is a concrete example to illustrate the selection process: Honduras has learning assessments data for Indicator SDG4.1.1a in different years- Reading. These data have been generated from regional and international learning assessments programs in different years:

Year	Source of data
2011	PIRLS 2011 (Grade 4)
2013	ERCE 2013 (Grade 3)

The assessment program to use for reporting on SDG 4.1.1.a - Reading for Honduras will be ERCE 2013. The alternatives PIRLS targets grade. Grade 4 is one grade above the intended grade of the indicator –grade 2/3. Thus, ERCE is assessing grade 3 students and is a better option than PIRLS.

Until the process of equating international and regional assessments will be finalized, it is important to use only one source of information so that progress can be tracked on a comparable basis. The restriction is one level and domain. The country could have different reporting sources for a given year for different levels (for instance ERCE for 2/3 grade and PISA/TIMSS for end of lower secondary).

7.3. Mapping of grade to measurement point:

1. Grade 2 or 3: Plus one year when primary lasts more than 4 years according to ISCED levels in the country, except for TIMSS/PIRLS grade 4, which are mapped to the end of primary education when primary education lasts six years or less.
2. End of primary education: Plus or minus one year from the last year of primary education, except for TIMSS/PIRLS grade 4, which are mapped to the end of primary education when primary education lasts six years or less, according to the ISCED level mapping in the country .
3. End of lower secondary education: Plus two or minus one of last year of lower secondary according to ISCED level mapping in the country.

7.4. Mapping of domains to reading or mathematics:

1. Reading:
    a. The country has no assessment in reading alternative domains, like language or writing, should be used if reading is not available for reporting;
    b. Different languages could be used for reporting? when results are available in different languages, the official or most relevant language in the country should be used.
2. Mathematics: alternative domains can also be considered and used for reporting.

7.5. Parity index indicators (SDG indicator 4.5.1)

Parity index indicators are calculated using the adjusted parity index formula.

7.6. Summary of the reporting protocol

Table 2. 2019 Indicator 4.1.1 reporting

Assessment	Sources of data	Alignment to the Global MPL	Priority for reporting by education level	Footnotes
International assessments: PISA, PISA4D, TIMSS/PIRLS	Yes	According to Consensus Meeting	1	[Assessment name and year]
Regional assessments: LLECE, PASEC, SACMEQ, PILNA	Yes	According to Consensus Meeting	2	[Assessment name and year]

The reporing of national assessments could be used for reporting only if the alingment methodologies have concluded the alingment phase and only for results previous to 2020.

Source: UNESCO Institute for Statistics (UIS)