



United Nations  
Educational, Scientific and  
Cultural Organization



UNESCO  
INSTITUTE  
FOR  
STATISTICS



GLOBAL  
ALLIANCE  
TO MONITOR  
LEARNING

# Aligning and reporting on indicator 4.1.1: UIS annotated workflow

17-18 October 2018  
Hamburg, Germany



### Acronyms

CS:	Coding Scheme
CAT:	Content Alignment Tool
EGRA:	Early Grade Reading Assessment
EGMA:	Early Grade Mathematics Assessment
GCF:	Global Content Framework
LaNA:	Literacy and Numeracy Assessment
LLECE:	Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación
MPL:	Minimum Proficiency Level
NAEP:	National Assessment of Educational Progress
PAT:	Procedural Alignment Tool
PASEC:	Programme d'Analyse des Systèmes Éducatifs de la CONFEMEN
PIRLS:	Progress in International Reading Literacy Study
PISA:	Programme for International Student Assessment
PLD:	Performance Level Descriptors (to define performance/tasks student could do)
RL:	Reference List
SACMEQ:	The Southern and Eastern Africa Consortium for Monitoring Educational Quality
SEA-PLM:	Southeast Asia Primary Learning Metric
TIMSS:	Trends in International Mathematics and Science Study

## Contents

Executive summary	4
1. Objectives and Structure	5
2. The challenges	5
3. Reporting Consistency: GAML work flow	6
4. Global Content Framework	8
5. Procedural alignment	12
6. Proficiency Framework and minimum level, Linking Strategies and Interim Reporting	14

## 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 rest of this note discusses the focus of UIS's work (second column), and the columns to the right.

**Table 1- Summary of Outputs and Status for GAML deliberations**

Phase/ Tool	Focus of UIS work	Expected Outputs	Status	Expected From GAML
(1)	(2)	(3)	(4)	(5)
Conceptual Framework	Global Content Framework	Global Content Framework (GCF) to serve as reference	Ready for adoption	Acknowledgment
		Content Alignment Tool (CAT)	Drafts for discussion	Adoption
		Online Platform for CAT	Development	Adoption
Methodological Framework	Procedural Alignment	Manual of Good Practices	Finalized	---
		Quick Guides to support implementation in countries	Development	Acknowledgment
		Procedural Alignment Tool Online platform	Finalized	Adoption
Reporting Framework	Proficiency Framework and Minimum Level Linking strategies Interim reporting	Scale and definition of minimum proficiency level	Drafts for discussion	Adoption
		A linking strategy portfolio	Drafts for discussion	Adoption of strategy
		An interim reporting strategy	Finalized	-----

## 1. Objectives and Structure

This annex aims to present the work of the UNESCO Institute of Statistics (UIS) related to reporting on indicator 4.1.1, to inform all members of task force 4.1, and guide the discussions of the 5th GAML meeting in October 17-18, 2018.

The document will explain the flow of work, the activities and the outputs in the context of GAML's 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 can be considered comparable to the percentage reported from another country.

Following **column 2 of the table above**, 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 quality of the data collected?; and
- A proficiency scale where all assessments could be informed (and its conversion function or the linking procedure), 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.

## 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, 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.

A study conducted by Trevino and Ordenes (2017) sets the stage by [exploring the commonalities and differences between regional and international assessments](#), with the objective of understanding the challenges and options in terms for reporting indicator 4.1.1.

The analysis suggests that:

- The different approaches to measuring indicator 4.1.1 all have advantages and shortcomings in relation to technical issues and feasibility.
- It is necessary to create political agreement and advance the technical sphere to define the minimum level of competency in reading and mathematics.
- It is also necessary to approach procedural consistency so a minimum level of data quality given the heterogeneity among assessment programs is attained.
- The paper lays out four strategies for reporting indicator 4.1.1, including a new unique SDG4 test.
- An alternative to developing a specific instrument with a clear definition of the minimal level of competency. This may ensure high levels of comparability of the results and avoids technical critiques, but loses flexibility.

### 3. *Reporting Consistency: GAML work flow*

The objective is to define the criteria and generate the tools that could serve as:

- **Reference points:**

The content, procedural and reporting alignments provide a common language and approach to the development of assessments contents (for Mathematics and Reading), minimum procedural practices and reporting ensuring comparable monitoring progress towards SDG4 indicator 4.1.1.

- **Transparency tools:**

The adoption of common minimum coverage practices and reporting frameworks could make comparisons more transparent across countries and regions.

- **Normative references:**

The tools to be generated have the potential to become a standard against which countries, regions, institutions, international agencies and professionals 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 GAML's 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 2- Summary of Process and Focus of GAML**

Phase /Tools	What it addresses	Main Components	Focus of UIS Work
Conceptual Framework	<b>What to assess? - Concept</b> Who to assess? – Population: in and out of school? What contextual information to collect?	<b>Domain and subdomain: minimum coverage</b> Target population Background Questionnaire	<b>Global Content Framework (GCF)</b> <b>Content Alignment Tool</b>
Methodological Framework	<b>What are the procedures for data integrity</b>	Test design Sampling frame Operational design Data generation Data analysis	<b>Good practices guidance</b> <b>Procedural Alignment</b>
Reporting Framework	<b>What format to report?</b> <b>What is the minimum level?</b> <b>How to link or “harmonize”?</b>	Reporting model Scale or proficiency framework Linking Definition of an interim reporting strategy	<b>Proficiency Framework and minimum level</b> <b>Linking strategies</b> <b>Interim Reporting strategy</b>

### 3.1. Conceptual Framework

**What is covered:** Content (what is reading and what is mathematics?) and definition of population and contextual information to collect. Assuming countries are to take definitions based on their priorities on the target population (including only in school children) and the contextual information.

**Scope of work of UIS:** The focus is to define the content framework for each domain and point of measurement and to find a definition on the minimum contents that ensure comparability between tests. This leads to the Global Content Framework (GCF) shown in column 4 above.

### 3.2. Methodological Framework

**What is covered:** Assessment implementation faces many methodological decisions that are not identical between them. Examples of methodological decisions include the format of the test and sampling decisions.

**Scope of work of UIS:** The focus is to define minimum procedural practices that ensure integrity in the data generating process. This leads to the Procedural Alignment work shown in column 4 above.

### 3.3. Reporting framework

**What is covered:** Achieving statistics which are comparable over programmes and countries is perhaps more difficult than is assumed. This is due to the fact that different regions have different traditions concerning the stringency of proficiency benchmarks at different grades. Moreover, these realities further complicate comparisons across countries, which often involve comparing slightly different grades, even at the same educational level.

The only way to compare is under some criteria and related to a common scale built based on proficiency benchmarks including the definition of a minimum proficiency (that is what the indicator requires) with the accompanying definition of the alignment strategy.

**Scope of work of UIS:** The focus is to define a scale with the associated proficiency definitions, the definition of the minimum proficiency level and a set of linking strategies to the proficiency framework. This leads to the Proficiency Framework and minimum level, the linking strategies, and the interim reporting described in column 4 above.

#### 4. Global Content Framework

This section describes in more detail the work that needs to be done, or is being done, for row 1, column 4, in Table 1 above.

##### 4.1. Why and What

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. Furthermore, even domains can be defined differently. 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) have collaboratively developed a Global Content Framework (GCF) for the domains of Mathematics and Reading.

##### 4.2. Objective

To define the minimum common set of contents and skills that should be taught and assessed in each of the points (grade 2/3, end of primary, and end of lower secondary) of measurement the indicator requires.

##### 4.3. Expected Outputs

There are three final products:

- (1) Global Content Framework (GCF) of [Mathematics](#) and [Reading](#) to serve as reference (noted above and **Error! Reference source not found.**)
- (2) [Content Alignment Tool](#) (CAT) including alignment criteria (**Error! Reference source not found.**)
- (3) A platform to help countries self-assess (described further below)

##### 4.4. Expected Outcome

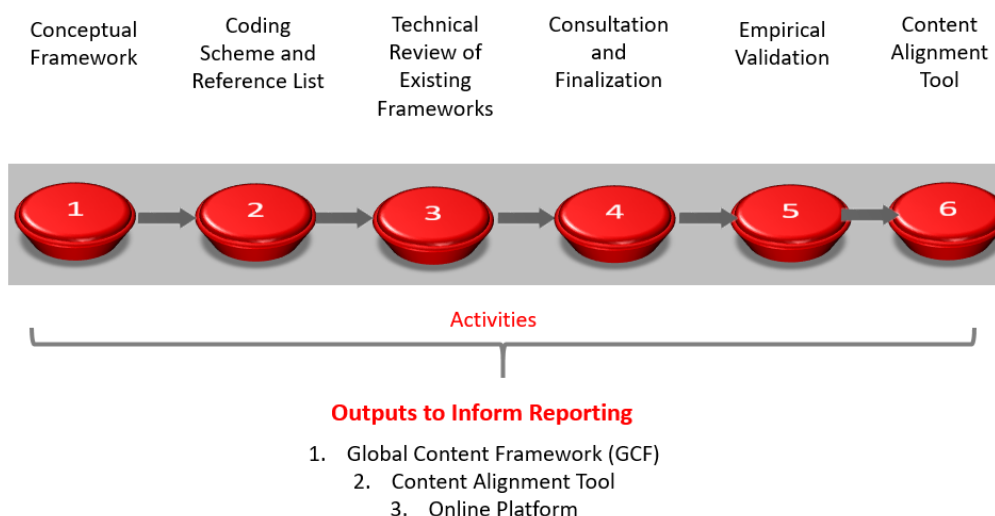
To ensure data integrity with respect to minimum comparability in the concepts each assessment program includes.

##### 4.5. Activities

In order to develop this **GCF**, the following activities were taken, summarized in Figure 1 and described more fully below the figure. Figure 1 helps to explain the process for creating the Global Content Framework (GCF).



**Figure 1. Process to develop the Global Content Framework**



### Activity 1: Conceptual model

**Definition of activity:** The conceptual development of a global framework based on cognitive learning theory and empirical inputs.

**Scope:** The first analysis used initial inputs from various national curricula, and, subsequently, national assessment frameworks to refine the coverage of frameworks.

Intermediate Products:

- [Method for developing an international curriculum and assessment framework for Mathematics;](#)
- [Method for developing an international curriculum and assessment framework for Reading](#)

### Activity 2: Development of coding scheme and initial reference list

**Definition of activity:** The **coding scheme and an initial reference list (CS-RL)** for mapping assessments was built based on theory and initial technical review. Qualitative information was used to help further improve the conceptual coverage of the GCF.

**Scope:** This CS-RL was then used to conduct a mapping exercise for 115 Mathematics national assessment frameworks and 73 Reading national assessment frameworks covering various languages and regional representativeness. This mapping shows considerable convergence in what is already assessed globally.

Intermediate Products:

- Monitoring Progress towards SDG 4.1: [Initial analysis of national assessment frameworks for Mathematics and](#)
- Monitoring Progress towards SDG 4.1: [Initial analysis of national assessment frameworks for Reading.](#)

Status: **Finalized - To inform GAML plenary**

### Activity 3: Technical review of existing frameworks

**Definition of activity:** (i) The technical review of mathematical and reading concepts and competencies assessed at the regional and international levels includes the initial review of existing assessment frameworks, identification of trends, differences, and commonalities using a coding scheme (CS). The CS grants that definitions of domains, sub-domains, constructs, and sub-constructs are comparable<sup>1</sup>. (ii) Analysis focused on assessment frameworks given their specificities. Curricula were used to “back fill” the mapping if needed.

**Scope:** The initial review was conducted by looking at all regional and international assessment, including the following tools: EGMA, EGRA, ePIRLS, LANA, LLECE, PASEC, PILNA, PIRLS, PISA 2015, PISA 2018, PISA for Development (PISA-D), SACMEQ, SEA-PLM, and TIMSS (assessments in alphabetical order).

Intermediate Products:

- A database ([International regional assessments](#)) presents a mapping of the contents of the assessment frameworks of the aforementioned assessments, following the coding scheme for national assessment frameworks. The information shows differences and commonalities in terms of both structure and content<sup>2</sup>.

Status: To inform GAML5 plenary

### Activity 4: Consultation and finalization

**Definition of activity:** The proposed global framework that incorporated a revision based on Activity 3, which includes an improved Coding scheme and Reference Lists (CS-RL), was sent for online consultations to receive feedback from diverse actors.

**Scope:** The consultation focused on the first two levels of the global framework: domain and sub-domain, and participants were asked to test the new framework by using it to map their country's national assessment frameworks at these two levels.

- [Global Content Framework of Reference for Mathematics: Global Consultation Results](#)
- [Global Content Framework of Reference for Reading: Global Consultation Results](#)

**Output 1:** The consultation feedbacks have been used as input to review and update the content reference list and further **improve the GCF descriptors. The GCF descriptors** present the ‘preferred’ learning into groups and they are further classified into in four categories: Domain, Sub-domain, Construct and Sub-construct, from the most global (Domain level) to the most detailed (Sub-construct level). The presentation is to help conceptualize the grouping of learnings which may happen at different stages of learning development or build on other learnings. The descriptors are grouped by concept and not by development stage.<sup>3</sup>

- [Global Content Framework of Reference for Mathematics](#)
- [Global Content Framework of Reference for Reading](#)

<sup>1</sup> Information on sub-constructs is present only in four assessments for both subjects, due to the different categorisations each assessment framework followed.

<sup>2</sup> This could serve to further analysis in later stages.

<sup>3</sup> The feedbacks from the global consultation suggested that the mapping should be done at least at construct level with inputs of sub-construct as references. This also helped the UIS conceptualize the interactive platform for data collection that would be accessible to country.

Status: **Finalized - To inform GAML plenary**

### Activity 5: Empirical validation

**Definition of activity:** To analyze how the emerging GCF compares to the international assessment frameworks. Improves the mapping of the international assessments frameworks onto the GCF.

**Scope:** (i) The International assessment framework includes IEA's TIMSS, PIRLS and OECD's PISA. Given that these are the most known by countries and have well established conceptual and analytical frameworks with rigorous psychometric properties in assessment, they are used as initial comparison to the global framework to validate the comprehensiveness of global content framework. (ii) Looks at how national frameworks (Assessment) align to the GCF for a selected group of 20 countries.

Intermediate Products:

- **International:** several short papers show mapping of the respective assessment frameworks from each of the international assessment to the GCF and found that in most cases the global frameworks for reading and math are more comprehensive. The GCF have a wider range of coverage than TIMSS and PISA.
  - [GCF\\_TIMSS Alignment paper](#)
  - [GCF\\_PIRLS Alignment paper](#)
  - [GCF\\_PISA\\_Math Alignment paper](#),
  - [GCF\\_PISA\\_Reading Alignment paper](#).
- **National:**
  - [Comparative Analysis of Curriculum National Assessment Frameworks for Mathematics](#)
  - [Comparative Analysis of Curriculum National Assessment Frameworks for Reading](#)

Status: **Finalized - To inform GAML plenary**

### Activity 6: Content Alignment Tool

**Definition of activity:** Since countries' assessment programs do not need to cover all contents in the GCF but should cover in a proportion of the framework, it is necessary to generate a mechanism for countries to assess their alignment to the GCF.

**Scope:** Generate the tools that, in a simplified way, allows one to map assessment frameworks, against the GCF, in order to:

- Generate a content alignment questionnaire using the GCF as a reference point.
- Define preliminary criteria about minimum alignment to help countries evaluate whether their assessments have met minimum content coverage to ensure reporting. This will be discussed at the GAML plenary.
- Generate a tool to map and assess the level of alignment (coverage) of national assessment frameworks to the GCF.

**Outputs 2 and 3:**

- [Content Alignment Tool](#) for assessment programs not studied by IBE.

- *A platform* that would generate a database with the countries alignment to GCF

The multilingual website would display geographic heat-map and charts and invite users to complete an online survey designed to capture the data needed to complete the 4.1.1 Global Content Framework and will allow, afterwards, to compare a given country against another country, a region, or the world.

Respondents will enter data via a series of questions that form a dialogue between the respondent and the UIS. The respondent's answers will be stored in a database

When the questionnaire is completed, the system should provide to the user a scorecard that measures the level of compliance of the national against the global framework in reading and/or mathematics.

Status: **Finalized - Waiting for GAML plenary adoption**

## 5. Procedural alignment

This section describes in more detail the work that needs to be done, or is being done, for row 2, column 4, in Table 1 above.

### 5.1. What and 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. Objective

To define the minimum procedures that ensure data integrity

### 5.3. Expected 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](#)

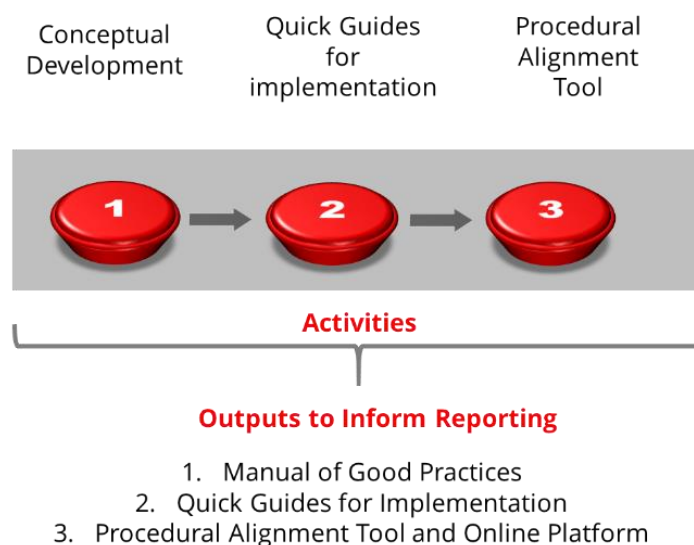
### 5.4. Expected Outcome

Generating process with a minimum level of data integrity, sufficient enough to report and compare results from different assessment programs.

### 5.5. Activities

The workflow process of activities to develop the procedural alignment tool are described in Figure 2 and described more fully below the figure.

**Figure 2. Process to develop the Procedural Alignment Tool**



### Activity 1: Conceptual development

**Definition of activity:** To define a set of good practices in an assessment cycle that could ensure the production of good quality data

**Scope:** (i) To define based on existing literature and documents a set of good practices to guide implementation; (ii) To define quick guides to the implementation of SDG4; (iii) To generate a tool and scoring guide to assess compliance with minimum standards.

#### Output 1

- [Manual of Good Practices in Learning Assessment](#) (GP-LA): a guideline of good practices, and
- Two quick guides on learning assessment for reference: [Making the Case for a Learning Assessment](#) and [Implementing a National Learning Assessment](#)

Status: Endorsed

### Activity 2: Quick guides for implementation

**Definition of activity:** To provide countries an abridged and handy tool to implement assessments

**Scope:** cover broader guidance on key issues

- Why carry out a learning assessment and main procedural decisions to take
- How to implement a learning assessment
- How to maximize the data collection to report on SDG4 using learning assessment

#### Output 2

- Measuring SDG4 using Learning Assessment (**under development**)
- [Making a case for a Learning Assessment.](#)

- [Implementing a National Learning Assessment](#).

Status: In development – To inform GAML plenary

### Activity 3: Procedural alignment tool

**Definition of activity:** to ensure reported data for indicator 4.1.1 have an acceptable quality,

**Scope:** (i) Questionnaire, (ii) Scoring guide, (iii) Online platform

#### Outputs 3 and 4:

- [Procedural Alignment Tool and](#)
- [Online procedural alignment tool platform](#)

Status: Finalized - Waiting for GAML plenary adoption

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

### 6.1. What and why?

This section describes in more detail the work that needs to be done, or is being done, for **row 3, column 4, 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.

Currently, there are no common standards as a global benchmark. While data from many national learning assessments are readily available, every country sets its own standards, leading to inconsistent definitions of performance levels. This is also true with cross-national learning assessments, including international and regional learning assessments. For education systems who participated in the same cross-national learning assessments, results are comparable, but not across different cross-national learning assessments, and certainly not across national assessments.

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<sup>4</sup>.

### 6.2. Objective

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

- A metric that is arbitrary

<sup>4</sup> Taking from the [NAEP](#) on policy statement: “Policy definitions are general statements to give meaning to the levels.”

- 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. Expected Outputs

The final products are:

- *Scale for each domain and point of measurement (benchmarks and definition of the minimum proficiency level or each domain and point of measurement).*
- *A portfolio of linking strategies and the tools that allow to locate assessments proficiency levels in a scale*
- [An Interim reporting](#) strategy protocol

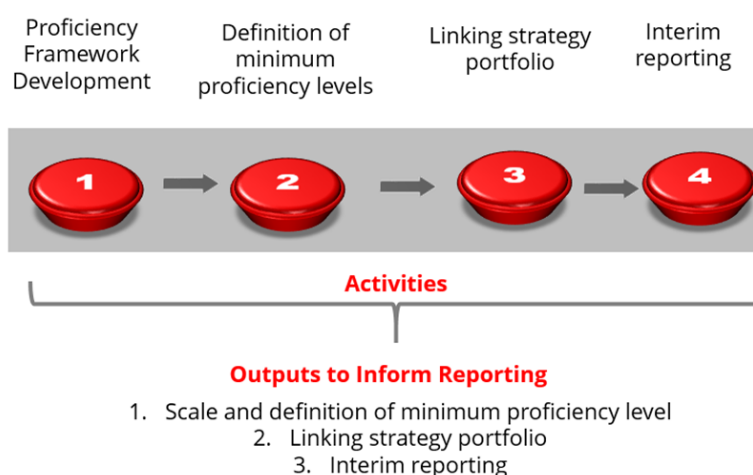
### 6.4. Expected Outcome

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<sup>5</sup>. Once completed, it could be used to identify roughly comparable proficiency benchmarks within national assessment programmes and even examinations.

### 6.5. Activities

There are several proposals from different international organizations on how to link assessments to a common scale using different approaches and methodologies in a process summarized by Figure 3 and described below.

**Figure 3. Process to develop the Procedural Alignment Tool**



#### Activity 1: Proficiency Framework

**Definition of activity:** A proficiency scale that involves the definition of common content standards, the definition of the number of performance levels, determining the labels and

<sup>5</sup> The initial development of the reporting proficiency scale would draw from both expert opinion and analysis of existing data and policy level descriptors.

writing descriptions for the levels of the proficiency metric<sup>6</sup> along with set of agreed-upon policy statements about the abilities of students

**Scope:** All cross national assessment programs and their reporting scale in initial mapping

Intermediate Products:

- Document with a proposed proficiency framework empirical scale, preliminary performance level descriptors (PLDs) and the set of minimum proficiency level (MPLs) based on these descriptors:
  - The mapping of all proficiency levels of existent cross-national assessments with their descriptors, put into a standardized language, and building a continuum based on PLDs from lower to higher levels of proficiency for each domain regardless of grade.
  - Based on this prior step, define a proficiency framework including proposed preliminary performance level descriptors (PLDs).
  - Alignment with the GCF

Status: Discussed in September. The UIS, through a consensus building meeting with cross-national agencies and country representatives discussed and refined this proficiency framework.

### Activity 2: Minimum Proficiency Level (MPL)

**Definition of activity:** To define a minimum **global proficiency level** for each point of measurement and domain including the **performance level descriptors (PLD)**.

**Scope:** The following inputs will be used to define the output

- the mapping of cut-points in each cross-national assessment that define the MPL
- the analysis of experts about the number of cuts needed (to accommodate countries at different socio- and economic-development stages) for this framework at each of the three educational levels knowing that for some countries the MPLs chosen as global reference might be too high a value while for others it will be too low.
- The set of cutoff points and their descriptors are convenient to set a framework that can contextualize the minimum level, but are not necessary for global reporting—only the minimum level is.

Intermediate Products:

- *Document* with a proposal of the global minimum proficiency level for each point of measurement and domain in SDG4 4.1.1 including the PLDs. ([link to summary paper by content experts – to come](#))

Status: [Under development to be discussed in September and in GAML plenary](#)

### Activity 3: Linking strategies

**Definition of activity:** (i) A linking strategy portfolio to link assessments and locate them in the scale; (ii) A mapping of what and when to link

<sup>6</sup> The initial development of the reporting proficiency scale would draw from both expert opinion and analysis of existing data and policy level descriptors.



**Scope :**

**Strategy 1 - Non-statistical approach:** Pedagogically informed recalibration of existing data – policy linking.

- Policy linking approach involves using the proposed framework that describes the range of competencies that children/youth have at each level to locate proficiency levels from alternative assessment programs based on the PLDs and guided by experts' judgement.
- This proposal would allow one to expand coverage in terms of educational systems reporting for SDG 4. For instance, coverage at the primary level would double, in terms of the population-weighted world, if national assessments were included.

**Strategy 2 - Statistical approach**

- **2.a. Psychometrically informed recalibration based on common items.** One version has been proposed by ACER as part of an overall proposal of progression in learning but options are not exhausting there<sup>7</sup>.
- **2.b. Recalibration through the running of parallel test on representative sample of students.** IEA outlines the '[Rosetta Stone](#)' solution that deals only with the primary level and allows two assessment, one international other regional to be expressed on the same scale. Concretely, the proposal states that sub-samples of students in three to five countries per programme would write not just the regional tests, but also IEA's test. This would produce a 'concordance table' with all countries, participating and not participating in the same scale<sup>8</sup>.
- **2.c. Recalibration of existing data.** This approach relies largely on statistical adjustments<sup>9</sup> taking advantage of the fact that some countries, referred to as 'doubloon countries', participate in more than one cross-national programme. Using several such overlaps has allowed for the identification of roughly comparable proficiency thresholds. It could serve to double check but there is foreseen unlikely political buy-in.

**Strategy 3- New test:** a third strategy could be to develop a new test that all countries take for reporting under common comparable tool but this is neither politically feasible nor cost-efficient so it has not been followed..

Weighing on options

These efforts should be taken more as complementary routes than as alternative options in order to minimize risk. The strategies help each other to build a sustainable reporting strategy. It is easy to see

- **Stepping stones between strategy 1 and 2a**
- **Complementarity between 1 and 2b** (as the Rosetta Stone needs to be expressed in a proficiency framework).
- And checking for **2 c as proposed by Trevino and Ordenes.**

## Intermediate Products

<sup>7</sup> Note that the reference scale is built from items coming from various assessments.

<sup>8</sup> For countries the option is to either participate in a regional program or in a global program (something that might be difficult or not possible if the region does not have any regional initiative).

<sup>9</sup> See Altinok, N. (2017). "Mind the Gap: Proposal for a Standardised Measure for SDG 4-Education 2030 Agenda". UNESCO Institute for Statistics (UIS).

## Reporting Scale:

- Document with a proposal of the minimum proficiency level for each point of measurement and domain in SDG4 4.1.1 containing the PLDs.
- Linking options
- **Strategy 1 - Non-statistical** approach there is a paper to be discussed on [Social Moderation](#) (SM)
- *Toolkit to align (will be developed)*
- **2.a. Psychometrically informed recalibration based on common items**
- **ACER proposal**
- **2.b. Recalibration through the running of parallel test on representative sample of students**
- Rosetta Stone
- Concordance Table
- **Weighing on options: Costs benefit analysis of linking strategies**
- The UIS has commissioned a [paper](#) that summarize the various alternatives (except ACER's item based linking approach) with its costs and benefits. It is hoped that this paper will provide an overview to help the plenary think through the best way forward on linking.

Status: Under development to be discussed in September

#### Activity 4: Interim reporting

**Definition of activity:** To provide a reporting strategy until the content and procedural alignment are finished

**Scope:** The UIS has defined an interim reporting strategy that lies within the long-term vision of the UIS reporting strategy.

Currently, the UIS is accepting all national and international assessment data with footnotes and qualifiers to explain where the data come from and to help the users in understanding the limitations of these data.

**Table 3 – Interim reporting in a nutshell**

	In school- Based		Population Based	What Grade
	Cross National	National		
2/3 Grade	LLECE PASEC TIMSS PIRLS	Yes	MIC6 EGRA/EGMA PAL Network	2/3
End of Primary	LLECE PASEC SACMEQ PILNA SIMEAO TIMSS PIRLS	Yes	PAL Network	Plus one minus one of Last Year of Primary according to ISCED level in your country
End of Lower Secondary	TIMSS PISA PISA4D	Yes	Young Lives	Plus <b>two</b> minus one of Last Year of Lower Secondary according to ISCED level in your country
Definition of Minimum Level	The ones defined by each assessment by point of measurement and domain			
Grade for End of Primary and End of Lower Secondary	As defined by the ISCED level of each country			
Validation	Send from UIS for countries approval			

*Note: TIMSS/PIRLS Grade 4: these results are allocated to the end of primary when, according to the ISCED levels in a given country, there are 4 grades in primary. When primary has more than 4 grades, they are allocated to grade 2/3.*

**Output 4:**

- [Interim reporting](#) strategy protocol

Status: Developed and published.

