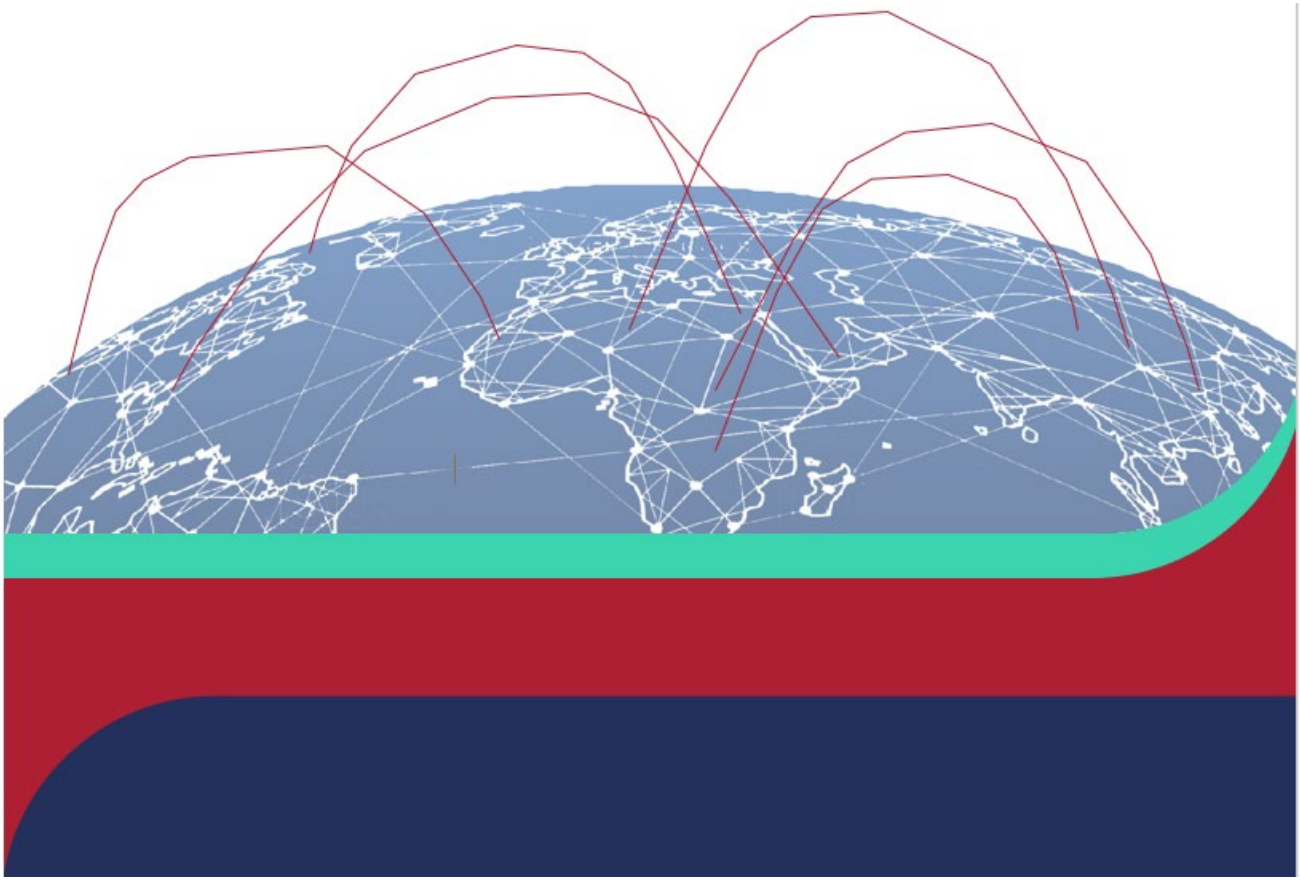
unesco
Institute for Statistics

Aligning and reporting on indicator 4.1.1: UIS annotated workflow

Last update: September 2023



Contents

Acronyms

ACER	Australian Council for Educational Research
AMPL	Assessments for Minimum Proficiency Levels
CS	Coding Scheme
CAT	Content Alignment Tool
EGRA	Early Grade Reading Assessment
EGMA	Early Grade Mathematics Assessment
GAML	Global Alliance to Monitor Learning
GCF	Global Content Framework
GPF	Global Proficiency Framework
LaNA	Literacy and Numeracy Assessment
LLECE	Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación
MILO	Monitoring Impacts on Learning Outcomes
MPL	Minimum Proficiency Level
NAEP	National Assessment of Educational Progress
PAT	Procedural Alignment Tool
PASEC	Programme d'Analyse des Systèmes Éducatifs de la CONFEMEN
PILNA	Pacific Islands Literacy and Numeracy Assessment
PIRLS	Progress in International Reading Literacy Study
PISA	Programme for International Student Assessment
PISA-D	Programme for International Student Assessment for Development
PLD	Performance Level Descriptors
RL	Reference List
SACMEQ	The Southern and Eastern Africa Consortium for Monitoring Educational Quality
SDG	Sustainable Development Goal
SEA-PLM	Southeast Asia Primary Learning Metric
TIMSS	Trends in International Mathematics and Science Study
UIS	UNESCO Institute for Statistics

Executive summary

This document summarizes the work carried out to develop a programme for reporting on SDG Indicator 4.1.1, or the '*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*'. A summary table at the start presents all the major outputs and resources produced. This is followed by a discussion of the main challenges faced and a detailed description of the UIS GAML workflow. The three subsequent sections cover all the activities, outputs and corresponding details in relation to the three phases, i.e., the conceptual, methodological and reporting frameworks. Acknowledging the importance of reporting on learning outcomes and the fact that much work had already been completed in this field, the UIS prioritized it and motivated others to carry out the work that has not been done yet.

Each of the activities and outputs help build the tools to generate a minimum level of consistency across the education systems' reporting against indicator 4.1.1, while maintaining a sufficient level of flexibility for these education systems to administer assessment programmes appropriate to their context and needs.

The reporting format of indicator 4.1.1 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 assessment programmes can be considered comparable, and the conditions (or conversion function) under which the percentage reported from an assessment programme can be considered comparable to the percentage reported from another country.

The following inputs are required to frame the indicator:

- What contents should be measured and what is the percentage of coverage needed by a given assessment to be comparable to others?
- What procedures are good enough to ensure quality of the data collected?
- What is the definition of the minimum level for each domain that would allow the estimation of the percentage of students achieving the minimum proficiency level (and what would be a well-defined proficiency scale)?
- What is the linking strategy or set of linking strategies that allow all assessments to be informed in a comparable way in the same scale?

An ideal programme for reporting on SDG Indicator 4.1.1 will have gone through the following three phases:

1. Conceptual Framework
2. Methodological Framework
3. Reporting Framework

Each of these phases contains several complex sub-steps. For various levels and types of assessments, the UIS had completed most of the work required before accepting the responsibility of being the custodian of reporting on SDG 4.1.1.

Table 1- Summary of outputs for reporting on SDG 4.1.1

Phase (1)	Focus (2)	Outputs (3)	Resources produced (4)
Overview	Process description	Aligning and reporting on indicator 4.1.1: UIS annotated workflow (Latest update: April 2023)	
Conceptual framework	Global Content Framework	Global Content Frameworks for reading and mathematics	Global Content Framework for reading (2018) Global Content Framework for mathematics (2018)
		Content Alignment Tool (CAT)	Content Alignment Tool (CAT) (2018)
		Online Content Alignment Tool (CAT) Platform	Online CAT Platform (2018)
Methodological Framework	Procedural Alignment	Manual of Good Practices	Principles of Good Practice in Learning Assessment (2017)
		Procedural Alignment Tool (PAT)	Procedural Alignment Tool (PAT) (2018)
		Online Procedural Alignment Tool (PAT) Platform	Online PAT Platform (2018)
Reporting Framework	Minimum Proficiency Level Global Proficiency Framework Linking strategies Official reporting *Metadata Protocol for reporting Country's options* Use of cross-national assessments for reporting	Minimum Proficiency Levels (MPL), definition.	Minimum Proficiency Levels Minimum Proficiency Levels (MPLs): Outcomes of the consensus building meeting (2018) Minimum Proficiency Levels: described, unpacked and illustrated (2019; 2022)
		Global Proficiency Framework (GPF)	GPF reading (English – Spanish) (2020) GPF mathematics (English – Spanish) (2020)
		Linking strategies	Exploring commonalities and differences in regional and international assessments (2017) Costs and benefits of different approaches to measuring the learning proficiency of students (SDG Indicator 4.1.1) (2019) The feasibility of harmonizing scores produced by Assessments for Minimum Proficiency Levels (AMPL) to the TIMSS and PIRLS test scores to measure and monitor SDG 4.1.1b (2023)
		Policy linking	Policy linking methodology (2017) Policy linking for measuring global

			learning outcomes toolkit: Linking assessments to the Global Proficiency Framework (2023)
		Ex-post Calibration	Mind the gap: Proposal for a standardised measure for SDG 4-Education 2030 Agenda (2017)
		Student based linking: concordance tables	Rosetta Stone project (2022) Rosetta Stone policy brief: Establishing a concordance between regional (ERCE/PASEC) and international (TIMSS/PIRLS) assessments (English, French, Spanish) (2022) Rosetta Stone Analysis Report: Establishing a Concordance between ERCE and TIMSS/PIRLS (2022) Rosetta Stone Analysis Report: Establishing a Concordance between PASEC and TIMSS/PIRLS (2022)
		MPL Calibrated Module	Monitoring Impacts on Learning Outcomes (MILO) (2021-2022) COVID-19 in Sub-Saharan Africa: Monitoring Impacts on Learning Outcomes (2022): - Main Report - Country Reports: Burundi - Burkina Faso - Côte d'Ivoire - Kenya - Senegal - Zambia Students reaching the Minimum Proficiency Levels reporting AMPL and PASEC (concept note) Assessments for Minimum Proficiency Levels (AMPLs): ground-breaking tools to produce internationally comparable data on SDG 4.1 indicators - Brochure (2022) (*English* - *French*)
		SDG Indicator 4.1.1: Reporting	Metadata Indicator 4.1.1 Protocol for Reporting (English - French) (2022) Reporting learning outcomes in basic education: country's options for indicator 4.1.1 (2022)

		Use of cross-national assessments for reporting	Monitoring of the Sustainable Development Goals using large-scale international assessments (2022)
Research and analysis			Learning divides: Using data to inform educational policy (2018) Costs and benefits of different approaches to measuring the learning proficiency of students (SDG Indicator 4.1.1) (2019) How Fast can Levels of Proficiency Improve? Examining Historical Trends to Inform SDG 4.1.1 Scenarios (2019) Evidence-based Projections and Benchmarks for SDG Indicator 4.1.1 (2020) Pandemic-related disruptions to schooling and impacts on learning proficiency indicators: a focus on the early grades (2021) - Tool for projecting the attainment of SDG 4.1.1 (Excel) (2021) Assessing Learning Proficiency Levels and Trends for Sustainable Development Goal 4.1: a focus on Africa (2021) Trends in learning proficiency in the last twenty years: How close are we to reliable regional and global SDG 4.1.1 trend statistics? (2022) Feasibility of using the data produced by the Early Grade Reading (EGRA) and Early Grade Mathematics (EGMA) to measure and monitor SDG 4.1.1, by complementing it with other banks of items (2023)
Dissemination		Platforms Microsite Easy-to-understand guides Data Digest	Global Alliance to Monitor Learning (GAML) Technical Cooperation Group on SDG4 Indicators (TCG) Learning Data Toolkit: measure what matters (Microsite) (2022)

			Quick Guide: Implementing a National Learning Assessment (2017) Quick Guide: Making the Case for a Learning Assessment (2018) SDG 4 Data Digest 2018: Data to nurture learning (2018)

1. Objectives and Structure

This document aims to present the work of the UNESCO Institute of Statistics (UIS) on the reporting of indicator 4.1.1 to inform Member States and guide them in the production and reporting of this indicator. It describes in detail the flow of work, the activities, and the outputs produced for reporting, and it is presented in a logical rather than chronological order. It was published for the first time in November 2017 for the 4th meeting of the Technical Cooperation Group; it was revised back in May 2021 and then in May 2023.

Each of the activities and outputs played an instrumental role in building the tools to generate a minimum level of consistency in reporting against indicator 4.1.1 across the various education systems, while maintaining a sufficient level of flexibility for these education systems to administer assessment programmes that are deemed appropriate to their context and needs.

The reporting format for indicator 4.1.1 aims to communicate two pieces of information:

1. the percentage of students meeting minimum proficiency standards for
    a. the relevant domains (reading and mathematics), and
    b. each point of measurement (grades 2/3; end of primary; and end of lower secondary education.
2. the conditions under which the percentage reported in a country can be considered comparable to the percentage reported in another country, and when different assessment programmes can be considered comparable.

The following inputs are required to frame the indicator (see column 2 of **Table 1**):

- What <u>contents or constructs</u> in each of the relevant domains (reading and mathematics) should be measured?
- What is the <u>percentage of coverage</u> of a given assessment to be comparable to other assessments?
- What procedures are needed to ensure that the <u>quality</u> of the data collected is adequate to report effectively?
- A definition of the <u>minimum proficiency level</u> for each domain that would allow the estimation of the percentage of students achieving the minimum proficiency level
- The <u>linking procedure/s</u> that allows to convert all assessments in the same scale.

The next section of the document presents the challenges that were faced, and the solutions developed to address these issues. Section 3 presents the logic of the GAML workflow. Sections 4, 5 and 6 go deeper into each of the stages of the workflow – i.e., the conceptual, methodological, and reporting frameworks, describing the objectives and activities and providing links to the main outputs and deliverables.

2. 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 programme 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. The approach of the UIS 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 programme that follows certain comparability guides ahead of time, and certain quality assurance and procedural guides, to report on 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 programmes as levers for system improvement.

- A study was conducted by Treviño and Órdenes in 2017 setting the stage through Exploring Commonalities and Differences in Regional and International Assessments. The objective was to understand the challenges and options in terms of reporting indicator 4.1.1. It suggests: 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 programmes is attained.
- It lays out four strategies for reporting indicator 4.1.1, including the potential creation of unique SDG 4 test.
- An alternative to developing a specific instrument with a clear definition of the minimal level of competency to ensure high level of comparability (though risking flexibility), while avoiding technical critiques.

3. Reporting consistency: the UIS GAML workflow

The objective of this activity is to define the criteria and generate the tools that could serve as reference points, transparency tools and normative references as briefly described below.

Reference points

- The content, procedural and reporting alignments provide a common language and approach to the development of assessments contents (for reading and mathematics), minimum procedural practices and reporting ensuring comparable monitoring progress towards SDG 4 indicator 4.1.1.

Transparency tools

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

Normative references

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

The UIS GAML workflow was designed following the structure of the implementation of any learning assessment. **Table 1** above summarizes the relevant areas of GAML's work and contextualizes the work that has taken place and is still taking place, with regard to the three main steps in developing a means of reporting on SDG4 indicators: conceptual, methodological and reporting frameworks.

Table 2 below provides a more detailed context to the introductory points presented thus far and highlights the focus of the work accomplished by the UIS and its partners.

Table 2- Summary of the process and focus of the UIS

Phase	Objectives	Main components to research
Conceptual Framework	What to assess? Who to assess? What contextual information to collect?	Domain and subdomain: minimum coverage Target population Background questionnaire

Methodological Framework	What are the procedures for data integrity?	Test design Sampling frame Operational design Data generation Data analysis
Reporting Framework	What is the minimum level? How to link or "harmonize"? What format to report? What operational choices?	Minimum Proficiency Levels (MPL) Global Proficiency Framework (GPF) Linking Strategies Metadata Protocol for reporting

3.1 Conceptual Framework

What does it cover? It covers the content of what is meant by "reading" and "mathematics" and establishes a definition of the population and the contextual information which needs to be collected to report effectively. This assumes that countries are to use definitions based on their set priorities of the target population (including only in-school children) and the contextual information.

Scope of work of UIS: The focus during 2017/18 was to define the content framework for each domain (reading and mathematics) and specific points of measurement, and the minimum content that ensures comparability between tests. This resulted in the development of the Global Content Framework (GCF) defined for each domain, reading and mathematics as well as the Content Alignment Tool (CAT).

3.2 Methodological Framework

What does it cover? Assessment implementation faces many methodological decisions that are not identical from one assessment to another. Examples of such decisions include the format of the test and sampling decisions.

Scope of work of UIS: The focus was to define minimum procedural practices that ensure integrity in the data generating process. This resulted in the development of the Procedural Alignment Tool (PAT) and the Online PAT Platform.

3.3 Reporting framework

What does it cover? The only way to compare programmes across countries is to set criteria and related components of a programme or assessment to a common scale based on proficiency benchmarks, including the definition of a minimum proficiency level with a detailed description of the alignment strategy to express all assessments in the same scale.

What are the challenges? Producing statistics that are comparable across programmes and countries is perhaps more difficult than is assumed. This is because 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.

Scope of work of UIS: The focus of the reporting framework was the definition of a scale specifically associated with the proficiency definitions, and the definition of the minimum proficiency levels and a set of linking strategies to the proficiency framework.

This led to the development of the Global Proficiency Framework (reading and mathematics), and the Minimum Proficiency Levels. The work also focused on the Linking Strategy Portfolio, exploring different linking strategies

- including the Rosetta Stone project (concordance between international and regional assessments) and the MILO project (implementation of the Assessments for Minimum Proficiency Levels or AMPLs). The Protocol for Reporting was developed and a number of reference documents were produced to guide countries in reporting on 4.1.1.

4. Conceptual Framework: Global Content Framework

This section describes the work completed to define and establish the Global Content Framework for reading and mathematics.

4.1 Why and what?

Why is a Global Content Framework needed? Assessment programmes 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, programmes assess different skills, use different content to assess the same domain, and, eventually, do both differently, even for a same grade.

What is the objective of a Global Content Framework? The UIS and the International Bureau of Education (IBE-UNESCO) have collaboratively developed a Global Content Framework (GCF) for the domains of reading and mathematics with the objective to assess the degree of alignment among various assessments to common contents serving as the lay out of the basis for a global comparison.

4.2 Objective

To define the minimum common set of contents and skills that should be taught and assessed in each of the points of measurement of the indicator; in the case of indicator 4.1.1, the points of measurement are for grade 2 or 3, end of primary education, and end of lower secondary education.

4.3 Outputs

Three GCF-related outputs have been produced:

(1) Global Content Framework (GCF) for reading and mathematics, to serve as reference;

(2) Content Alignment Tool (CAT), that maps the coverage of the content of any assessment programme and includes an alignment criteria;

(3) An Online CAT Platform to help countries self-assess the coverage of their programmes.

4.4 Expected Outcome

To ensure data integrity with respect **to a minimum comparability of the domains and constructs included in each assessment programme.**

4.5 Activities

The elaboration of the content framework, which ultimately led to the GCF, consisted of the six activities presented in **Table 3** and described in the remaining of the section.

Activity 1: Conceptual framework

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

Scope: The approach was specifically intended to: (1) create a content and skills framework for mathematics from cognitive theory and various national curricula; and (2) develop a coding scheme to map various national assessment frameworks (NAF) onto the framework to subsequently refine the coverage of frameworks.

Outputs

- Method for developing an international curriculum and assessment framework for Reading (2018)

- Method for developing an international curriculum and assessment framework for Mathematics (2018)

Activity 2: Development of coding scheme and initial reference list

Definition of activity: The **coding scheme and 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 Global Content Framework.

Scope: The CS-RL was then used to conduct a mapping exercise of 115 national assessment frameworks in mathematics and 73 national assessment frameworks in reading, covering various languages and regional representativeness. This mapping shows considerable convergence in what is already assessed globally.

Intermediate products

- Coding scheme: Reading assessments (2018)

- Coding scheme: Mathematics assessments (2018)

Outputs

- Method for developing an international curriculum and assessment framework for Reading – Summary (2018)
- Method for developing an international curriculum and assessment framework for Mathematics – Summary (2018)

Activity 3: Technical review of existing frameworks

Definition of activity: The technical review of concepts and competencies of learners in reading and mathematics assessed at the regional and international levels includes:

1. An initial review of existing assessment frameworks, identification of trends, differences and commonalities using a coding scheme (CS).[1] The coding scheme grants that definitions of domains, sub-domains, constructs and sub-constructs are comparable.[2]
2. An analysis focused on assessment frameworks based their specificities. Curricula were used to complete the mapping of constructs if needed.

Scope: The initial review to check on validity was conducted by looking at all of the following regional and international assessment:

- EGMA Early Grade Mathematics Assessment
- EGRA Early Grade Reading Assessment
- LaNA Literacy and Numeracy Assessment for Developing Countries
- LLECE El Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación
- PASEC Programme d'analyse des systèmes éducatifs de la confemen
- PILNA Pacific Islands Literacy and Numeracy Assessment
- PIRLS Progress in International Reading Literacy Study
- ePIRLS Innovative PIRLS assessment of online reading (2016)
- PISA Programme for International Student Assessment (PISA 2015; PISA 2018)
- PISA-D Programme for International Student Assessment for Development
- SACMEQ Southern and Eastern Africa Consortium for Monitoring Education Quality
- SEA-PLM Southeast Asia Primary Learning Metrics
- TIMSS Trends in International Mathematics and Science Study

Intermediate Products

[1] A **coding scheme** is a set of codes, defined by the words and phrases that researchers assign to categorize a segment of the data by topic.
[2] Information on sub-constructs is present only in four assessments for both subjects, due to the different categorisations each assessment framework followed.

- Monitoring Progress towards SDG 4.1: Comparative Analysis of Curriculum and Assessment National Frameworks for Mathematics – Summary; Paper (2018)

- Monitoring Progress towards SDG 4.1: Comparative Analysis of Curriculum and Assessment National Frameworks for Reading – Summary; Paper (2018)

Outputs

- Monitoring progress towards SDG 4.1: Initial analysis of national assessment frameworks for Reading (2018)

- Monitoring progress towards SDG 4.1: Initial analysis of national assessment frameworks for Mathematics (2017)

- Global frameworks of reference present a mapping of the contents of the following international and regional assessments, showing differences and commonalities in terms of both structure and content (domains, sub-domains, constructs and sub-constructs):

    - **Reading:** EGRA, ePIRLS, LaNA, LLECE, PASEC, PILNA, PIRLS, PISA, PISA-D, SACMEQ, and SEA-PLM

    - **Mathematics:** EGMA, LaNA, LLECE, PASEC, PILNA, PISA, PISA-D, SACMEQ, SEA-PLM, and TIMSS

Activity 4: Consultation and finalization

Definition of activity: The proposed global framework that incorporated a revision based on Activity 3 (technical review of existing frameworks), which includes an improved Coding Scheme and Reference Lists with the feedback received from diverse actors during consultation.

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.

Intermediate Products

- Global Content Framework of Reference for Reading: Global consultation results (2018)
- Global Content Framework of Reference for Mathematics: Global consultation results (2018)

The consultation feedbacks were used as inputs to review and update the content reference list and further improve the **Global Content Framework (GCF) descriptors. The GCF descriptors** present the 'preferred' learning into groups, and they are further classified into four categories:

- domain,
- sub-domain,
- construct,
- sub-construct.

These range 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**. 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 countries.

Outputs

- Global Content Framework for reading

- Global Content Framework for mathematics

Activity 5: Empirical validation

Definition of activity: Empirical validation was conducted to analyse how the emerging GCF compares to international assessment frameworks. It also improves the mapping of international assessments frameworks onto the GCF.

Scope: There are two aspects of the empirical validity scope:

(i) **International:** Includes the International Association for the Evaluation of Educational Achievement (IEA) 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) **National:** Explores the alignment of national frameworks (assessments) to the GCF for a selected group of 20 countries.

Outputs:

- **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.
  - UIS-TIMSS Framework Alignment: Methodology and Results (2018)
  - UIS-PIRLS Framework Alignment: Methodology and Results (2018)
  - UIS-PISA Framework Alignment: Methodology and Results (mathematics) (2018)
  - UIS-PISA Framework Alignment: Methodology and Results (reading) (2018)

Activity 6: Content Alignment Tool

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

Scope: Generate the tools that, in a simplified way, allow 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 for minimum alignment. This will help countries evaluate whether their assessments have met the minimum content coverage to ensure adequate reporting.
- generate a tool to map and assess the level of alignment (coverage) of national assessment frameworks to the GCF.

Outputs

- Content Alignment Tool for assessment programmes that aim to report for SDG4 in order to ensure minimum compliance with the minimum content
- An Online CAT Platform which provides the user with a scorecard that measures the level of compliance of the national against the global framework in reading and/or mathematics.

Respondents enter data via a series of questions, forming a dialogue between the country and the UIS.

Table 3. Process to develop the Global Content Framework

Activities		Intermediate products	Outputs to inform reporting
1	Conceptual Framework		Method for developing an international curriculum and assessment framework for Reading (2018) Method for developing an international curriculum and assessment framework for Mathematics (2018)
2	Coding Scheme and Reference List	Coding scheme: Reading assessments (2018) Coding scheme: Mathematics assessments (2018)	Method for developing an international curriculum and assessment framework for Reading – Summary (2018) Method for developing an international curriculum and assessment framework for Mathematics – Summary (2018)
3	Technical Review of Existing Framework	Monitoring Progress towards SDG 4.1: Comparative Analysis of Curriculum and Assessment National Frameworks for Mathematics – Summary; Paper (2018) Monitoring Progress towards SDG 4.1: Comparative Analysis of Curriculum and Assessment National Frameworks for Reading – Summary; Paper (2018)	Monitoring progress towards SDG 4.1: initial analysis of curriculum and assessment national frameworks for reading and for mathematics (2018) Global frameworks of reference (International and regional assessments)
4	Consultation and Finalization	Global Content Framework of Reference for Reading: Global consultation results (2018) Global Content Framework of Reference for Mathematics: Global consultation results (2018)	Global Content Framework for Reading (2018) Global Content Framework for Mathematics (2018)

5	Empirical Validation		**International level** Alignment papers - UIS-TIMSS Framework Alignment: Methodology and Results (2018) - UIS-PIRLS Framework Alignment: Methodology and Results (2018) - UIS-PISA Framework Alignment: Methodology and Results (mathematics) (2018) - UIS-PISA Framework Alignment: Methodology and Results (reading) (2018)
6	Content Alignment Tool (CAT)		Content Alignment Tool (CAT) (2018) Online CAT Platform (2018)

5. Methodological Framework: Procedural Alignment

This section describes the procedural alignment as part of the development and establishment of a methodological framework.

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 (procedural alignment) 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

Define the minimum procedures to ensure data integrity that grants comparability and compliance with minimum standards.

5.3 Outputs

1. Manual: Principles of Good Practice in Learning Assessment (2017)
2. Procedural Alignment Tool (2018)
3. Online PAT Platform (2018)

5.4 Expected Outcome

Ensure a minimum level of data integrity that is good enough to compare results from different assessment programmes procedural-wise.

5.5 Activities

The workflow process of activities to develop the procedural alignment tool are described in **Table 4**.

Activity 1: Conceptual development

Goal of activity: Describe good practices in an assessment implementation cycle to ensure the production of good quality data.

Outputs

- Principles of Good Practice in Learning Assessment (Manual - 2017)

Activity 2: Procedural alignment tool

Goal of activity: Generate a **tool and scoring guide** to assess compliance with the minimum set of procedures (or standards) of an assessment and to ensure reported data for indicator 4.1.1 is of acceptable quality.

Scope: The scope of this activity includes the following:

(i) Questionnaire
(ii) Scoring guide
(iii) Online platform

Outputs

- Procedural Alignment Tool (2018)
- Online Platform (2018)

Table 4. Process to develop the Procedural Alignment Tool

Methodological Framework		
Activities		**Outputs**
1	Conceptual Development	Principles of Good Practice in Learning Assessment - Manual (2017)
2	Procedural Alignment Tool (PAT)	Procedural Alignment Tool (2018) Online PAT Platform (2018)

6. Reporting Framework

6.1 What and why?

This section describes in more detail the work that is being done or that needs to be done to support reporting framework for 4.1.1.

Assessment programmes typically report using different scales. Analysis of results therefore remains contained to one 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 and to have a clear definition of what is the minimum level for each as requested by indicator 4.1.1.

Data from many national learning assessments are readily available, but every country sets its own standards, leading to non-comparable definitions of performance levels that makes comparisons difficult.

Comparability between different international and regional programs 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. **Appendix 1** provides the main characteristics of the various existing cross-national assessments such as the domain, grade or age, frequency, and the fees associated with implementation of each assessment.

Given those difficulties in the comparability, the use of similar proficiency level descriptors. A proficiency level descriptor is an overarching policy statement or policy definition of what a student could do at each point of measurement. Most importantly, this is a very useful tool for defining what constitutes a *minimum* (which is what the SDG4.1.1 indicators call for) proficiency level.[3]

In 2018, there are standard definitions for Minimum Proficiency Levels (MPL) at the global level for each of the domains. This has been completed after a careful process and mapping of proficiency levels and through a Consensus Building Meeting on Proficiency Levels that has agreed on the global benchmark definition for minimum proficiency levels for each point of measurement of the indicator.

The MPL are part in each assessment program of a scale that is a set of proficiency benchmarks or levels embedded within a numerical scale and their cut points on that numerical scale. A scale gives meaning to the succession of cut scores and the proficiency levels associated. These benchmarks are 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.

An immediate step is to link the different assessments to this common definition of MPL that goes beyond harmonization of the proficiency level descriptors but resort to psychometric methods in order to achieve this.

A final important point is the guidance to reporting provided to Member States and other stakeholders which is conveyed through the metadata and reporting guidelines (protocol for reporting; country choices).

6.2 Objective

To define a scale where all learning assessment programmes can be located and a linking strategy for the programmes to the scale. The scale includes:

- A definition of the minimum level and the policy statements associated to a set of benchmarks.
- A strategy to link assessments to express them in the same scale at minimum linked to the MPL definition;
- A definition of a reporting strategy for indicator 4.1.1, including metadata and a protocol for reporting that embeds the operational decisions taken to report in the presence of various assessments.

[3] Taking from the NAEP on policy statement: "*Policy definitions are general statements to give meaning to the levels*."

6.3 Outputs

The outputs resulting from the reporting framework phase are:

- A definition of the minimum proficiency level or each domain (reading and mathematics) and point of measurement (grade 2 or 3, end of primary, and end of lower secondary education).
- A reporting scale for each domain and point of measurement where it is possible to locate the definition of the minimum proficiency level.
- A portfolio of linking strategies and the tools that allow to locate assessments proficiency levels in a scale.
- Metadata
- A protocol for reporting that makes transparent the operational choices and definitions that define the numbers published.

6.4 Expected Outcome

Ensure consistency in the reporting and granting comparability through a common definition of Minimum Proficiency Levels (MPL) for each domain (reading and mathematics) and point of measurement (grades 2 or 3, end of lower, end of secondary education) and a linking strategy between assessment programmes to report at the global level that paired with the metadata document and the protocol for reporting the conceptual definition reach an operational clarity.

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 in **Table 5** and described afterwards.

Table 5. Process related to the reporting framework

Reporting Framework			
Activities		**Intermediate products**	**Outputs**
1	Minimum Proficiency Level (MPL)	Minimum Proficiency Levels (MPLs): outcomes of the consensus building meeting – Background papers Constructing UIS proficiency scales and linking to assessments to support SDG Indicator 4.1.1 reporting (2017)	Minimum Proficiency Levels used to report for indicator 4.1.1 (last updated in 2022) Minimum Proficiency Levels (MPLs): outcomes of the consensus building meeting (2018) Minimum Proficiency Levels: described, unpacked and illustrated (2019; 2022)
2	Global Proficiency Framework		UIS Reporting Scales (UIS-RS): Definition of a reporting scale (2017) Global Proficiency Framework (2020) GPF reading (English – Spanish) GPF mathematics (English – Spanish)
3	Linking Strategies	Linking strategies documents	Exploring Commonalities and Differences in Regional and International Assessments (2017) Costs and benefits of different approaches to measuring the learning proficiency of students (SDG Indicator 4.1.1) (2019) The feasibility of harmonizing scores produced by Assessments for Minimum Proficiency Levels (AMPL) to the TIMSS and PIRLS test scores to measure and monitor SDG 4.1.1b (2023)
		Policy linking	Policy linking methodology (2017) Policy Linking for Measuring Global Learning Outcomes Toolkit: Linking Assessments to the Global Proficiency Framework (2023)
		Ex-post Calibration	Mind the Gap: Proposal for a Standardized Measure for SDG4 - Education 2030 Agenda (2017)

		Subject based linking: concordance tables	Rosetta Stone Project (2022) Rosetta Stone policy brief: Establishing a concordance between regional (ERCE/PASEC) and international (TIMSS/PIRLS) assessments (English, French, Spanish) (2022) Rosetta Stone Analysis Report: Establishing a Concordance between ERCE and TIMSS/PIRLS (2022) Rosetta Stone Analysis Report: Establishing a Concordance between PASEC and TIMSS/PIRLS (2022)
		MPL Calibrated Module	Monitoring Impacts on Learning Outcomes (MILO) (2021-2022) - COVID-19 in Sub-Saharan Africa: Monitoring Impacts on Learning Outcomes - Main Report (2022) - COVID-19 in Sub-Saharan Africa: Monitoring Impacts on Learning Outcomes - Country Reports (2022) Burundi - Burkina Faso - Côte d'Ivoire - Kenya - Senegal - Zambia - Students reaching the Minimum Proficiency Levels reporting AMPL and PASEC (concept note) - Assessments for Minimum Proficiency Levels (AMPLs): ground-breaking tools to produce internationally comparable data on SDG 4.1 indicators - Brochure (English - French) (2022)
4	Reporting of indicator 4.1.1	Metadata	Metadata for 4.1.1
		Protocol for Reporting	Protocol for Reporting (English - French) (2022)
		Country's options	Reporting learning outcomes in basic education: country's options for indicator 4.1.1 (2022)
5	Use of cross-national assessments for reporting	Cognitive indicators; non-cognitive indicators; equity	Monitoring of the Sustainable Development Goals using large-scale international assessments (2022)

Activity 1: Minimum Proficiency Level (MPL)

Goal of activity: To define a minimum **global proficiency level** for each point of measurement and domain including the **Performance Level Descriptors (PLD[4])**.

[4] To define performance/tasks student could do at each grade/level.

The MPL is related to a proficiency scale which includes: the definition of performance levels required of students to be considered proficient, and the definition of the number of performance levels, determining the labels and writing descriptions for the levels of the proficiency metrics.[5] The MPL is based on the policy statement definition and can be used to identify roughly comparable proficiency benchmarks within national assessment programmes and even examinations.

Scope: The following inputs have been used to define the output:

- the mapping of cut-points in each cross-national assessment with the respective policy descriptors including the one that defines 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 to respect the fact that for some countries, the chosen global MPLs as global reference might be too high a value while for others, they will be too low.
- The set of cut-off points and their descriptors are convenient to set a framework that can contextualize the minimum level and could serve to track progress in the distribution of the skills. The cut-off points are not necessary for global reporting—only the minimum level is.

Intermediate Products:

- Minimum Proficiency Levels (MPLs): outcomes of the consensus building meeting – Background papers:
  - Paper 1: Mathematics – Methodology for Ordering Performance Level Descriptors
  - Paper 2: Mathematics – Methodology for PLD Compilation and Cross-Functional Alignment
  - Paper 3: Reading – Compilation of Performance Level Descriptors Across Regional and International Assessments
  - Paper 4: Reading – Cross-National Assessments Alignment with the Global Framework for Reading and MPL Analysis

  These papers with a proposed proficiency framework empirical scale, preliminary performance level descriptors 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 previous step, define a proficiency framework including proposed preliminary performance level descriptors (PLDs).

  - Alignment with the Global Content Framework (GCF)

- Constructing UIS proficiency scales and linking to assessments to support SDG Indicator 4.1.1 reporting (2017)

Outputs

- Minimum Proficiency Levels used to report for indicator 4.1.1 (updated in 2022)
- Minimum proficiency levels (MPLs): outcomes of the consensus building meeting (2018)
- Minimum Proficiency Levels: described, unpacked and illustrated (2019 - 2022)

Activity 2: Global Proficiency Framework

Goal 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 writing descriptions for the levels of the proficiency metric[6] along with set of agreed-upon policy statements about the abilities of students.

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

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

Scope: All cross-national assessment programmes and their reporting scale in initial mapping

Outputs:

- UIS Reporting Scales (UIS-RS): Concept note July 2017, a document with the definitions of a reporting scale.
- Global Proficiency Framework (GPF) for reading and mathematics–2020
- GPF reading (English – Spanish)
- GPF mathematics (English – Spanish)

Activity 3: Linking strategies

Goal of activity: Define a portfolio with the different possible linking strategies to link one assessment programme to another and locate them on a common scale, and a mapping of what exactly can be linked and when to link the assessment programmes.

Scope: Define methodological approaches for two main linking strategies established for reporting on SDG 4 (see **Table 6**).

Table 6: Non-statistical and statistical approaches to linking strategies

Strategy 1 Non-statistical approaches	Strategy 2 Statistical approaches
Policy Linking Pedagogically informed recalibration of existing data Pairwise comparison method Evaluation of assessment items by group of experts in an independent way	**1) Recalibration through collection of new data** a) Item- based linking psychometrically informed recalibration based on common items. b) Student-based linking: Administration of parallel tests to build concordance tables (Rosetta-Stone) or based on an MPL calibrated module (AMPL)
	2) Recalibration of existing data

Strategy 1 - Non-statistical approach

Policy linking**: Pedagogically informed recalibration of existing data**
The policy linking approach involves using the proposed framework, which describes the range of competencies that children or youth have at each level of education, to locate proficiency levels of different assessment programmes based on the proficiency level descriptors (PLDs) and guided by experts' judgement. It expands the coverage in terms of educational systems reporting for SDG 4 using national data sources. For instance, coverage at the primary level could double, in terms of the population-weighted world, if national assessments were included.

The Policy Linking Toolkit offers the methodological steps and guidelines to standardize assessment programmes. It is designed for project teams, more specifically workshop facilitators, and resource persons - i.e., government officials, assessment agency officers, donor representatives, and partners - who will be organizing, funding, and/or implementing the policy linking methodology in their country or region.

The Pairwise comparison method consists of a group of experts in pedagogy and psychometrics doing the same evaluation but in an independent way. It relies on Learning Progression Scales for reading and mathematics. The toolkit to implement this method is still under development.

Strategy 2 - Statistical approaches

The statistical approaches propose three options for linking assessment programmes by gathering new data or using existing data.

2.1 Linking assessment programmes through the collection of (new) data

2.1.a Psychometrically informed recalibration based on common items

It consists of linking different assessment programmes by embedding the same (common) items in the programmes which then act as anchors and allows the selected assessment programmes to be calibrated on a common scale.

This anchoring item approach can be used cross-sectionally, i.e. across different assessment programmes typically countries, or longitudinally, i.e. for one country but over time (at different years).

One version has been proposed by the Australian Council for Educational Research (ACER) as part of an overall proposal of progression in learning, but options are not exhaustive there.[7]

2.1.b Student-based linking: Recalibration through the administration of parallel test to build a concordance table (Rosetta Stone)

IEA outlined the Rosetta Stone option[8] which links the results of regional assessments conducted at the primary level and international assessments together, such as TIMSS or PIRLS, to be expressed on a common scale. Concretely, IEA proposes that sub-samples of students in three to five countries per programme write both regional and the IEA tests to produce a 'concordance table' across the countries, putting the outcomes on a common scale.[9]

The objective of the Rosetta Stone is to link together assessments, which have been administered in the recent past and, to build concordance tables to compare the assessments outcomes and benchmark national results to those of the regional assessments. Concordance tables provide a link between regional assessments and the TIMSS and PIRLS achievement scales. The countries participating in the regional assessments can use the translations to determine what percent of their students could be expected to reach the TIMSS and PIRLS International Benchmarks of Achievement.

2.1.c Student based linking: Recalibration through the administration module calibrated to the Minimum Proficiency Level (AMPL**/**MILO)

The project 'COVID-19: Monitoring the Impacts on Learning Outcomes'(MILO), aimed at measuring learning outcomes and analyzing the impact of the pandemic on learning. More specifically, the project administered 'Assessments for Minimum Proficiency Level modules to report against SDG 4.1.1 at the end of primary (AMPL-b) in 2021.

AMPLs are ground-breaking and robust tools developed to produce internationally comparable data on SDG 4.1 indicators as they allow the identification of the proportion of children in each level of education who are achieving at least the Minimum Proficiency Level. AMPL assessments were administered in 2021 alongside national or regional assessments and aligned to the Global Proficiency Framework.

2.2 Existing data - Recalibration of existing data

This approach relies largely on statistical adjustments,[10] taking advantage of the fact that some countries, referred to as 'doubloon countries', participate in more than one cross-national programme. Using the 'doubloon countries' has allowed for the identification of roughly comparable proficiency thresholds. This option can serve to review outcomes but it is foreseen that it is an option unlikely politically bought in.

Proposed by Altinok in 2017, the approach develops a methodology to create indices of comparison between two assessments where enough countries participate in both assessments. It enables efficient comparison, since no additional instruments or costs are incurred in the anchoring process. Altinok noted that it is a second-best approach, and the ideal is comparison of micro or individual learner data, ideally using standard data

[7] Note that the reference scale is built using items from various assessments.

[8] IEA's Rosetta Stone: Measuring global progress towards the UN Sustainable Development Goal for quality education by linking regional assessment results to TIMSS and PIRLS International Benchmarks of Achievement (2018)

[9] For countries, the option is to either participate in a regional programme or in a global programme (something that might be difficult or not possible if the region does not have any regional initiative).

[10] See Altinok, N. (2017). Mind the Gap: Proposal for a Standardised Measure for SDG 4-Education 2030 Agenda. Information paper No. 46. Montreal: UNESCO Institute for Statistics

collection instruments. Treviño and Órdenes[11] proposed the utility of this statistical recalibration approach in its ability to provide a reality check against which to compare statistics based on national assessments.

In other words, the basic idea behind the methodology presented by Altinok is that some countries took part in several assessments: by using the results obtained in these assessments, we obtain anchored achievement tests. This is a quick and efficient method since it does not require any additional assessment with linking items and is based on a clear and basic idea according to which similar participation of several countries in different assessments may be used as anchoring countries.

Choosing a strategy

In choosing a strategy to link assessment programmes, using more than one linking strategy should be considered more as complementary routes than as alternative options to minimise risk if some of the approaches prove to be too costly, the margin of error too high, politically unfeasible or a combination of all these. **Table 7** shows the relationships between the different linking strategies and the coverage of various types of assessment. **Appendix 2** provides a detailed comparison of all available statistical and non-statistical linking options.

Table 7 Relationships between linking strategies and coverage of assessment types

	Strategy 1 Non-Statistical approach (Policy Linking)	**Strategy 2 Statistical approaches**			
		2.1 New data			**2.2 Existing data**
		2.1.b	**2.1.a**	**2.1.c**	
	Pedagogically informed recalibration	**Psychometrically informed recalibration based on common items**	**Recalibration through parallel tests (Rosetta Stone)**	**AMPLs**	**Statistical recalibration of existing data or ex-post calibration (Altinok)**
PISA, TIMSS and PIRLS	Yes	Could be used	Will be used	Yes	Yes
Regional cross-national assessments	Yes	Could be used	Will be used	Yes	Yes
National assessments	Yes	Could be used	Could be used	Yes	Not clear how
National examinations	To be used	-	-	Yes	Not clear how

Source: UIS adapted from Gustafsson (2019). Costs and benefits of different approaches to measuring the learning proficiency of students (SDG Indicator 4.1.1). Information paper No. 53. Montreal: UNESCO Institute for Statistics

The strategies help each other to build a sustainable reporting strategy. There are stepping stones between strategy 1 and strategy 2.1.a, and there is a complementarity between strategy 1 and strategy 2.1.b, such as the Rosetta Stone which needs to be expressed in a proficiency framework. Strategy 2.2 has a potential use as a check to compare statistics based on national assessments (Treviño and Ordenes, 2017). Finally, a triangulation of the various strategies is possible, including Strategy 2.2 by recalibrating existing data, as proposed in Mind the Gap: Proposal for a Standardised Measure for SDG 4 - Education 2030 Agenda by Altinok (2017), by creating comparable estimates across various international and regional assessments using adjusted scores, thereby obtaining the proportion of students reaching the MPL.

Output:

Strategy 1 – non-statistical approach

- **Pedagogically informed recalibration of existing data**

[11] See Treviño E. and M. Órdenes (2017). Exploring Commonalities and Differences in Regional and International Assessments. Information paper No. 48. Montreal: UNESCO Institute for Statistics

- Policy Linking for Measuring Global Learning Outcomes Toolkit: Linking Assessments to the Global Proficiency Framework (2020)

Strategy 2 – statistical approach

- **2.1 Linking that demands gathering new data**

  - Rosetta Stone: Measuring global progress towards SDG 4 by linking assessments results to TIMSS and PIRLS International Benchmarks of Achievements (2018)
  - Rosetta Stone policy brief: Establishing a concordance between regional (ERCE/PASEC) and international (TIMSS/PIRLS) assessments (English, French, Spanish) (2022)
  - Rosetta Stone Analysis Report: Establishing a Concordance between ERCE and TIMSS/PIRLS (2022)
  - Rosetta Stone Analysis Report: Establishing a Concordance between PASEC and TIMSS/PIRLS (2022)
  - COVID-19: Monitoring Impacts on Learning Outcomes (MILO): Study Design
  - COVID-19 in Sub-Saharan Africa: Monitoring Impacts on Learning Outcomes - Main Report (2022)
  - COVID-19 in Sub-Saharan Africa: Monitoring Impacts on Learning Outcomes - Country Reports (2022): Burundi - Burkina Faso - Côte d'Ivoire - Kenya - Senegal - Zambia
  - Students reaching the Minimum Proficiency Levels reporting AMPL and PASEC (concept note)
  - Assessments for Minimum Proficiency Levels (AMPLs): ground-breaking tools to produce internationally comparable data on SDG 4.1 indicators. Brochure (2022) (English - French)

2.2 Recalibration using existing data or ex-post calibration

  - Mind the Gap: Proposal for a standardized measure for SDG4 - Education 2030 Agenda (2017)

Activity 4: Metadata, Protocol for Reporting and countries' reporting options

Metadata: To provide countries with a summary of the methodological decisions and basic information used to publish data reported by UIS. The UIS develops metadata for SDG Indicator 4.1.1 to present the standardized steps to collect, process and produce statistical data.

Output:

- Metadata for SDG Indicator 4.1.1

Protocol for reporting: To provide countries with clarity regarding the decision taken to define the numbers in the tables.

Output:

- Protocol for Reporting on SDG Global Indicator 4.1.1

Countries' options to report on 4.1.1: To present countries with a menu of options they can choose from to report on indicator 4.1.1.

- Reporting learning outcomes in basic education: country's options for indicator 4.1.1 (2022)

Activity 5: Use of cross-national assessments to report on cognitive and non-cognitive SDG 4 indicators

Goal of activity: Provide guidance on how to use cross-national assessments (CNAs) to report on SDG 4 indicators

Scope:

- Highlight the contribution of CNAs in reporting on SDG 4 indicators by examining key issues including:
  - *How to use CNAs to measure cognitive SDG 4 indicators, i.e. learning outcomes?*

- *How to use CNAs to measure non-cognitive SDG 4 indicators in relation to students, teachers and schools?*
- *How to use CNAs to measure equity in learning and identify who is left behind? Equity dimensions include gender, location and socio-economic status among others.*

- Present a step-by-step guide of the methodology to be followed.to produce SDG 4 indicators based on CNAs, including Stata codes used to define the SDG indicator variables.

Outputs

- Monitoring of the Sustainable Development Goals using large-scale international assessments (2022)

7. Research and analysis

As part of the developments in relation to SDG indicator 4.1.1, the UIS has completed a great deal of additional research and analysed 4.1.1 data to explore trends and data characteristics over time.

Key research documents:

Learning divides: Using data to inform educational policy (2018)

Costs and benefits of different approaches to measuring the learning proficiency of students (SDG Indicator 4.1.1) (2019)

How Fast can Levels of Proficiency Improve? Examining Historical Trends to Inform SDG 4.1.1 Scenarios (2019)

Evidence-based Projections and Benchmarks for SDG Indicator 4.1.1 (2020)

Pandemic-related disruptions to schooling and impacts on learning proficiency indicators: a focus on the early grades (2021) - Tool for projecting the attainment of SDG 4.1.1 (Excel) (2021)

Assessing Learning Proficiency Levels and Trends for Sustainable Development Goal 4.1: a focus on Africa (2021)

Trends in learning proficiency in the last twenty years: How close are we to reliable regional and global SDG 4.1.1 trend statistics? (2022)

Feasibility of using the data produced by the Early Grade Reading (EGRA) and Early Grade Mathematics (EGMA) to measure and monitor SDG 4.1.1, by complementing it with other banks of items (2023)

8. Dissemination

The UIS has several platforms that are continuously updated with the most recent resources and publications on learning. The UIS has also developed a microsite on learning including relevant information and resources, in addition to easy-to-understand guides and a data digest on learning.

Platforms

Global Alliance to Monitor Learning (GAML)
Technical Cooperation Group on SDG4 Indicators (TCG)

Microsite

Learning Data Toolkit: measure what matters (Microsite) (video)

Easy-to-understand guides

Quick Guide: Making the Case for a Learning Assessment (2018)
Quick Guide: Implementing a National Learning Assessment (2017)

Data Digest

SDG 4 Data Digest 2018: Data to nurture learning (2018)

Appendix 1: Main characteristics of Cross-National Learning Assessments

Assessment	Domain, Area	Grade/Age	Cycle every ... Years	Estimated fees per round (in thousand USD)	Capacity development		Number of countries
					Test Related	Country report	
GLOBAL							
Progress in International Reading Literacy Study (PIRLS)	Reading	Fourth	4	227	included	not included	60
Trends in International Mathematics and Science Study (TIMSS)	Mathematics and Science	Fourth and eighth	4	222	included	not included	60
Literacy and Numeracy Assessment (LaNA)	Reading and Math	End of primary	on demand	100 to 150	not included	not included	
Programme for International Student Assessment (PISA)	Reading and Math	15-year-olds	3	199	with extra costs	with extra costs	79
REGIONAL							
ERCE	Language (reading and writing) and Mathematics.	Third and sixth	6	300	included	included	16
Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ)	Literacy and numeracy	Sixth grade	6	150	included	not included	14
Programme d'Analyse des Systèmes Educatifs de la CONFEMEN (PASEC)	French and mathematics	two and sixth	5	630	included	not included but supported	15
The Southeast Asia Primary Learning Metrics (SEA-PLM)	Reading, Mathematics, Writing, Global Citizenship	Fifth	4	119	included	not included	6
Pacific Islands Literacy and Numeracy Assessment (PILNA)	Literacy and Numeracy	4th and 6th	3	97	included	not included	15
CALIBRATED MODULE							
AMPL	Reading and Math	Upper Primary	on demand	80*	included	analysis and short report only AMPL	8

Note: * on average; PILNA: Secretariat Costs paid by Australia and New Zealand; Department of Foreign Trade and Affairs (DFAT) Australia pay the technical partners costs; Country costs are estimative.

SEA-PLM: UNICEF- EAPRO and UNICEF Country offices paid for the SEA PLM Expenses of participating countries and co-shared in regional expenses (regional workshops and field trial and main survey expenses) and staff support.

Source: *UIS based on assessment program information.*

Appendix 2: Comparing linking options

	Statistical				Non-Statistical	
	Ex-post calibration (Altinok)	**Common Students**	**Common items**	**AMPL module**	**Policy Linking**	**Pairwise comparison**
Data collection	Ex-post	Ex-ante	Ex-ante	Ex-ante	Ex-post	Ex-post
Students	Different	Same	Different	Different	Different	Different
What	Set different assessments on a common scale.	Concordance table of one scale into other.	Common items are inserted in the assessment	A module calibrated to the MPL is inserted either as an additional booklet or by running parallel assessments	Matches up definitions of the MPL descriptor using subjective judgement and, under certain conditions, allow those assessments to be aligned across countries.	
Items/Test	Different assessments	Different assessments	Common items in different assessments	Same module across different assessment program	Different assessments.	
Calibration	Puts all information in the same scale	Calibration needs a various step and builds a concordance table	joint calibration of assessment forms	accurate to report on the MPL	Depends on assessment program	
Alignment with Global MPL	No	Yes, but needs standards setting to define accurate alignment	depends on alignment and sufficiency	Yes	Depends on alignment and sufficiency of items	
Sufficient # of items	n/a	yes	depends on choice	Yes	Depends on each assessment tool	
Measurement skills continuums	No	yes	depends on the assessment programs	Not now but possible with current and future developments r	Depends on each assessment tool	
Track progress over time	Unclear	yes	yes	Yes	Not clear depends on quality of tools and the longitudinal equating	
Frequency	n/a	Cycle depending on each assessment	On demand	On demand	n/a	n/a
Output	Common scale using a modelling strategy	Concordance table	Allows to report on selected cut off points for both scales (e.g. MPL)	Calibrated to the MPL	Identifies the MPL cut-off points	Identifies the MPL cut-off points
How	-	Relies on the participation of countries in two assessments. Students take the two	Construction of a single reporting scale for each domain with items	Insert the booklets either as a standalone running parallel	Experts judge each item and set initial cut scores based on their understanding	Group of experts provide judgement about difficulty of each item on the

Institute for Statistics		assessments to help link between the results of both assessments.	from assessment programs.	assessment or as rotating booklet.	of the levels and the population.	assessment relative to items that have already been calibrated to MPL.
Country ownership	None	Very low	Medium to low	High	High	Medium
Needs	Tests measure the same latent construct	Tests have enough quantity of items that could identify linking	A common subset of calibrated items to be piloted to proof utility.	A tool built with items that are aligned and sufficient to measure the MPL	Good quality cognitive tools and procedures. Strong alignment of assessment tools to GPF.	
Pros	Inexpensive	Technically rigorous	Technically rigorous	Technically rigorous	Cost-effectiveness	
Cons	Unless there are equivalent tools not accurate for higher stakes uses, may be suitable for group and approximate uses	Costly. Efficient if done between a regional and a global assessment	Costlier financially and operationally.	Does not allow deep investigation of the construct.	Relatively subjective (less for pairwise). Depends on the quality of the assessment tool and implementation of the linking process.	
Achieved so far	Many attempts explored but most notably all the work of Altinok (2017))	Rosetta Stone: ERCE (LAC) and PASEC (SSA) participated with idea in the Rosetta Stone exercise.	--	AMPL-b administered AMPL-c ready to be administered (PISA) AMPL-a under preparation	First phase of Pilots around 10 countries run	Standard setting exercise for MILO (ACER, 2022)
Next/remaining steps	--	Potentially expansion to other regions and national assessments	--	Scale-up depends on country's interest and development partners support	Revision of toolkit	Methodology guidance and analysis
National Cost	None	Between US$ 250,000 and 400,000	--	Printing cost of a booklet. Extra administration costs depends on modality.	Between US$30,000 to 50,000 for national workshop	none
International Cost	100,000 to 250,000	International US$ 1 million per region. Regional – US$ 500,000	--	US$ 100,000 on average for technical assistance	Between US$ 50,000 and 75,000 per country	US$ 40,000

Source: UNESCO Institute for Statistics