



United Nations
Educational, Scientific and
Cultural Organization



UNESCO
INSTITUTE
FOR
STATISTICS



GLOBAL
ALLIANCE
TO MONITOR
LEARNING



Global Alliance to Monitor Learning (GAML): 2019 Progress Report

GAML6/WD/1

Acronyms and abbreviations

ACER	Australian Council for Educational Research
GAML	Global Alliance to Monitor Learning
IAEG-SDGs	United Nations Statistical Commission's Inter-agency and Expert Group on SDG Indicators
IBE	International Bureau of Education
IEA	International Association for the Evaluation of Educational Achievement
ILSAs	International Large-Scale Assessments
LAMP	Literacy Assessment and Monitoring Programme
LLECE	Latin American Laboratory for the Assessment for the Quality of Education
MPL	Minimum Proficiency Level
PASEC	Programme for the Analysis of Educational Systems
PILNA	Pacific Island Literacy and Numeracy Assessment
PIRLS	Progress in International Reading and Literacy Study
PLDs	Proficiency Level Descriptors
RFN	Regional Field Network
SACMEQ	Southern and Eastern Consortium for Monitoring Educational Quality
SDG	Sustainable Development Goal
SEA-PLM	Southeast Asia Primary Learning Metrics
TCG	Technical Co-operation Group
TIMSS	Trends in International Mathematics and Science Study
UIS	UNESCO Institute for Statistics



Contents

GAML WORKPLAN	4
COMMUNICATION AND OUTREACH	4
INDICATOR DEVELOPMENT	5
Indicator 4.1.1	5
Indicators 4.4.2	11
Indicator 4.6.1 - Mini LAMP	11
Indicators 4.7.4 and 4.7.5	12
New indicators on breadth of skills and love for learning	12



The Global Alliance to Monitor Learning (GAML) is an initiative to support national strategies for measuring learning and enable international reporting. Led by the UNESCO Institute for Statistics (UIS), GAML brings together UN member states, international technical expertise, and a full range of implementation partners — donors, civil society, UN agencies, and the private sector — to improve learning assessments globally. Through participation in GAML, all interested stakeholders are invited to help influence the monitoring of learning outcomes for Sustainable Development Goal (SDG) 4 and the Education 2030 goals.

This report presents the UIS work in GAML throughout the year 2019, in terms of knowledge production, communication and outreach, and finally, its role as a coordinator of GAML.

GAML Workplan

There are pending developments for indicators 4.6.1, 4.7.4 and 4.7.5, and there is a new proposed indicator about the breadth of skills. A new Task Force on Capacity Development has been formed in 2019. The Task Force has two objectives: 1. To better understand capacity development resources and practices, and 2. To support the organizational aspects of monitoring and reporting on the indicators.

In 2019, GAML aims to:

- Finalize the Global Frameworks of Reference for Indicators 4.7.4 and 4.7.5;
- Complete the definition of the minimum (or adequate) level of proficiency for the remaining indicators;
- Review and make recommendations on data collection and reporting procedures for indicators that currently do not have a designed survey;
- Develop the definition of the new indicator on breadth of skills; and
- Support countries through the UIS Regional Field Network (RFN).

Communication and Outreach

The communication of GAML work and development is a major responsibility at the core of the Secretariat's work. The UIS manages and regularly updates the GAML website, where information and resources are available on the GAML structure, Task Forces, and meetings.

Newsletter

Continuing the series of newsletters started in 2017, the UIS published the GAML newsletter in [EN](#), [FR](#), and [SP](#) in April 2019.

Blogs

Following the publication of the newsletter, the UIS released several key content blogs:

- [Calling All Assessment Experts! Online Consultation on Global Framework of Reference for Mathematics](#) introduces the Global Framework of Reference for Mathematics developed by UIS and the International Bureau of Education (IBE). The framework lists all contents and skills that can serve as a reference to teach, develop, and assess children, youth, and adults. It acts as a reference point, transparency tool, and normative instrument. The blog invited



stakeholders to an online consultation, and the feedback was later incorporated to the final version of the Global Framework of Reference for Mathematics.

- [The World is Off Track to Deliver on its Education Commitments by 2030](#): this blog introduces the first-ever projections on progress towards the SDG 4, discusses the main indicators that are out-of-track and proposes solutions such as the need for more financial investments and better data.
- [Digital Literacy Skills: From a Framework to a Measure](#): this blog highlights the challenges for operationalizing the framework in a way that is formative, informative and cost efficient.

SDG 4 Data Digest 2018

The UIS published the the [SDG 4 Data Digest 2018: Data to Nurture Learning](#). The Digest addresses the complex issue of learning outcomes. It presents a compilation of the most comprehensive and up-to-date work by international experts and relevant institutions to inform SDG 4's learning indicators.

The Digest discusses learning evidence concerning early child development, mathematics and reading skills for school-aged children, and digital and work-related skills for youth and adults. The discussion highlights the conceptual frameworks and tools developed by leading authors and institutions to understand, measure, monitor and support learning for all. It also reflects on the implications of informing SDG 4.

Indicator Development

Indicator 4.1.1

The UIS, through GAML, is working on an approach to monitor learning outcomes for Indicator 4.1.1 of the SDG 4: Quality Education:

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.

Thus, the methodological development referred to are:

1. Indicator 4.1.1.a. Propose a measurement strategy and methodological development plan to allow cross national comparability.
2. Indicator 4.1.1.b. Expand the comparability to express all assessments in the same reporting scale and minimum standards of quality.

Upgrading Indicator 4.1.1.a from Tier III to Tier II

In November 2018, the United Nations Statistical Commission's Inter-agency and Expert Group on SDG Indicators (IAEG-SDGs) upgraded indicator 4.1.1.a. following a proposal submitted by the UIS as the custodian agency.

In describing the process of indicator development, the proposal summarised the conceptual, methodological, and reporting frameworks, as well as the consensus building procedure that took place in order to submit the developed methodology for a Tier re-classification request. Each of the

frameworks included the concrete activities proposed, in addition to the expected outputs and deliverables.

Methodological Development for Indicators 4.1.1.a and 4.1.1.b

Alternative linking methodologies

1. Non-psychometric linking (pedagogical calibration or “social moderation”)

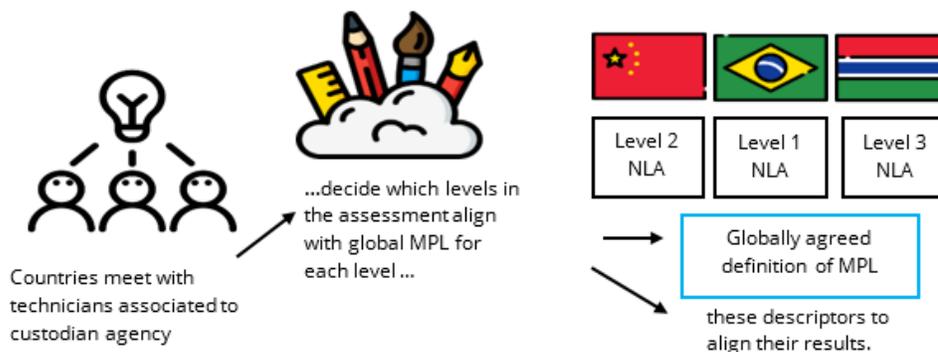
What? The approach involves using the definitions of the Proficiency Level Descriptors (PLDs) embedded in a given assessment, which are based on their own proficiency framework and items, to identify the PLD aligned with the Minimum Proficiency Level (MPL) global definition and sample items, sourced from various assessments, that have been judged by experts to be on either side of that global MPL. The alignment of PLDs (the one of the alternative assessment to be linked to the global MPL) to the global MPL is guided by experts’ judgement, but using a rigorous convergence process amongst the experts.

To increase accuracy, the majority of the items and the richness of the descriptors should be around the global MPL. Thus in a sense most of the analytical and empirical “supports” would be around only one break point (two PLDs), but more PLDs could be used, in order to provide context and increase utility to the countries (especially if some countries would bottom out or top out if only one break point is used).

How? Operationally, for each assessment (national or international) a group of 8 to 10 subject matter experts will convene for a socially moderated workshop. During the workshop, the experts provide individual and independent judgements about each item on the test to define alignment and set initial cut scores based on their understanding of the proficiency level descriptors and experience with the student populations and how they aligned and iterate until a desirable level of convergence is reached.

Which levels of education? Pedagogical calibration could be applied to all levels of education, but given the lower feasibility of using psychometric linking method in early grades and the absence of international and regional assessments at the end of primary education, it will be the preferred method for both levels.

Figure 1. Simplified illustration of the pedagogical calibration methodology



Country flags are used for illustrative purposes only and do not necessarily implicate the participation of these countries in the project.

2. The statistical approach

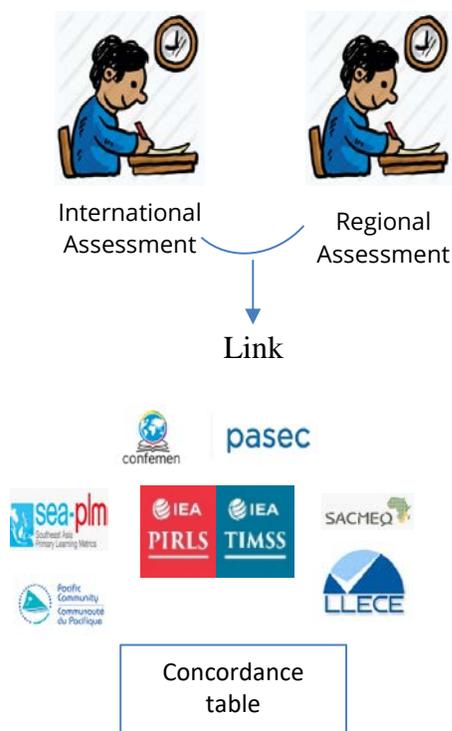
A. Recalibration by running a parallel test on a representative sample of students

What? The test-based linking approach relies on the participation of countries to cross-national assessments (both at the international and regional levels). The methodology allows two assessments, one international and the other regional, to be expressed on the same scale (and that scale to be aligned to the global one, or becoming, de facto, the key to the global one). The alignment will be possible through the creation of concordance tables which will provide a translation of the countries' regional assessment results to the Trends in International Mathematics and Science Study (TIMSS) and Progress in International Reading Literacy Study (PIRLS) achievement scales.

How? Students will take two assessments, an international and a regional one, and their results on both tests will be aligned in such a way that a link is established between the results from regional assessments conducted at the primary level and the TIMSS and PIRLS International Benchmarks for numeracy and literacy.

Which levels of education? It could be done at the end of primary education where regional and global assessments coexist and there is global capacity and interest. It would take advantage of the proficiency levels set forth by the International Association for the Evaluation of Educational Achievement (IEA) for its two assessments: TIMSS and PIRLS. TIMSS has been measuring trends in mathematics and science at four-year intervals since 1995. PIRLS has measured trends in reading literacy at five-year intervals since 2001. With 50 to 70 countries participating in each assessment cycle, the TIMSS and PIRLS achievement scales and their International Benchmarks are well established and used by countries all around the world.

Figure 2. Simplified illustration of the test-based linking methodology



There are five regional assessments conducting reading and mathematics assessments, and these represent the potential of the application of the methodology:

- SACMEQ – Southern and Eastern Consortium for Monitoring Educational Quality
- PASEC – Programme for the Analysis of Educational Systems
- LLECE – Latin American Laboratory for the Assessment for the Quality of Education
- SEA-PLM – Southeast Asia Primary Learning Metrics
- PILNA – Pacific Island Literacy and Numeracy Assessment

UIS and IEA, in collaboration with LLECE, will be implementing for the first time this methodology in the Latin American region. Because the concordance table will provide a projected TIMSS or PIRLS score for all possible regional assessment scores, it will be possible to determine the regional assessment scores equivalent to each of the TIMSS and PIRLS International Benchmarks.

Figure 3. Status of the implementation of social moderation and test-based linking approaches

Intermediate outcomes	Main milestones				
	2 nd Half 2018	1 st Half 2019	2 nd Half 2019	2020	2021
4.1.1a Methodologies for measuring and comparing learning in grades 2/3 have been developed by UIS, are embraced by stakeholders, and have secured 4.1.1a as a Tier II indicator.	Social moderation approach for grades 2/3 is tested and developed using LLECE as a test case, linking it to the UIS-RS directly. Technical reports/toolkits will be produced for applying the social moderation approach in grades 2/3. A <i>national</i> assessment test case will need to be found.				
	Application to IAEG-SDG to elevate indicator 4.1.1a to Tier II status is submitted and ideally approved.	Stakeholder engagement strategy for 4.1.1(a) is developed for general outreach on the importance of measuring indicator 4.1.1.a.	A bank of items for 4.1.1.a is developed with common anchor elements for integration into assessments or impact evaluation. An investment case is developed for measuring learning under 4.1.1.a.		
			UIS Reporting Scale created for indicator 4.1.1a		
4.1.1b Methodologies for measuring and comparing learning at the end of primary education have been developed and are embraced by stakeholders.	Toolkit, training materials, and technical reports for the Rosetta Stone approach are produced.				
		Stakeholder engagement strategy for 4.1.1(b) is developed for general outreach on the importance of measuring indicator 4.1.1.b.	Social moderation approach for end of primary is tested and developed using LLECE and Zambia as a test cases, linking them to the UIS-RS directly. Technical reports/toolkits will be produced for applying the social moderation approach at the end of primary. A bank of items for 4.1.1.b is developed with common anchor elements for integration into assessments or impact evaluation. Rosetta Stone approach carried out in 3 LLECE countries, open source concordance tables are generated.	IEA / Rosetta Stone scale is linked to UIS-PS through social moderation. All LLECE and PASEC (once Rosetta Stone completed) countries can be reported on UIS-PS.	
			An investment case is developed for measuring learning under 4.1.1.b. UIS Reporting Scale created for indicator 4.1.1b		

B. Psychometrically informed recalibration based on common items

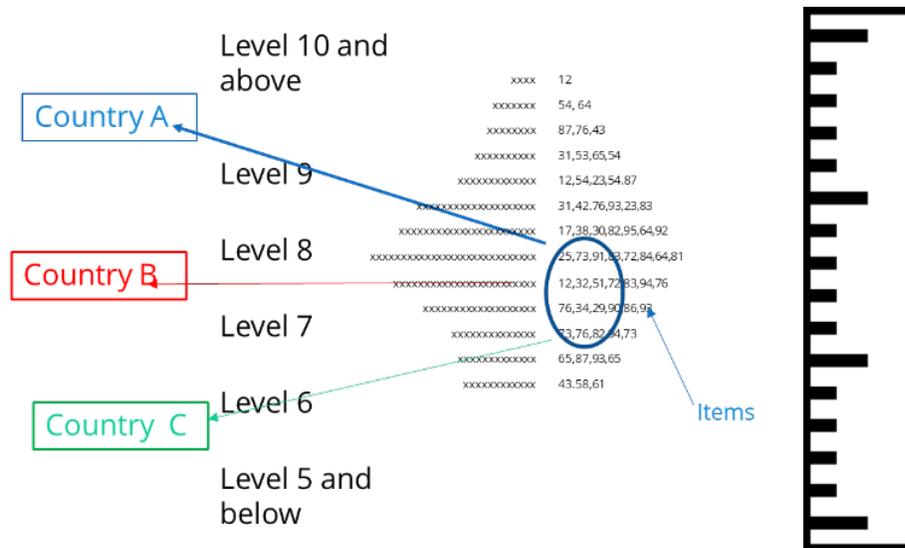
What? Item-based linking is a statistical method that relies on common assessment items given to different students and uses them to calibrate the level of difficulty of items that are not in common. It relies on psychometric calibration of these common items to achieve an alignment of standards and levels of difficulties. This method implies the use of common items in different assessment programmes. 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 exhausted.¹

How? The methodology is based on the collection of hundreds of test questions from a wide range of assessment programs and then by ordering them by level of difficulty using statistical methods and experts’ judgment to compare data from the different tests to construct a single scale for each domain. The test questions are analyzed to identify the kind of skills required to answer each of

¹ Note that the reference scale is built from items coming from various assessments.

the questions correctly. Each level of the reporting scales describes the skills based on test data. The reporting scales stem from these descriptions and are based on empirical evidence of skills demonstrated at similar points in the development of reading and mathematics, combined with expert judgments. All the items, or a useful subset of them, would then be calibrated using a set of common items at each “education level.” That being said, it is important to highlight that the methodology does not measure achievement. It is rather an approach to report learning outcomes in a meaningful and consistent way.

Figure 4. Simplified illustration of the item-based linking methodology



Tools to support measurement

Bank of Items

In addition to UIS’ plans mentioned above, UIS has set a proposal to develop a global bank of items or constructs with common anchor elements. From a capacity development perspective, these can be integrated into national learning assessments in order to strengthen them and allow for comparability and reporting. From the reporting perspective, a global bank of items could be useful for the item based-linking. The bank of items will be hosted by UIS, and as it is the case with all UIS products, will be provided free of charge.

Country Hub

In order to better guide countries in the process of implementation of the different indicators, the UIS presents all tools of the GAML under the Country hub section².

² <http://gaml.uis.unesco.org/country-hub/>



Indicators 4.4.2

Recommendations on assessment tools for monitoring digital literacy within the Global Framework for Reference on Digital Literacy Skills

The UIS published a report mapping existing assessments to the global framework of digital literacy skills, evaluating the advantages and disadvantages of selected assessments that cover a large part of the framework, and recommending next steps on an assessment tool. The paper recommends a suitable solution for monitoring the 4.4.2 Indicator.

From the report, it has emerged that PIX is a suitable tool to report on Indicator 4.4.2. The UIS will present in Armenia the implementation by Pix of the surveys on "Global Indicator on Digital Skills", as requested by UNESCO, in the largest possible number of countries, within the framework of data collection necessary for the verification of Indicator 4.4.2 of the 2030 Sustainable Development Goals. To go beyond the single measure of SDG Indicator 4.4.2, the proposal also considers the strengthening of the capacity of countries to achieve the goal related to digital skills. During the 6th Meeting of GAML, a proposal to implement this project, including its costs and standards, will be presented.

Indicator 4.6.1 - Mini LAMP

The **Literacy Assessment and Monitoring Programme (LAMP)** was developed by the UIS to respond to the pressing need to measure literacy and numeracy. It provides a sound methodology and tools to help countries, especially low- and middle-income countries, to monitor and improve literacy skills.

LAMP was field-tested in 10 countries: Afghanistan, El Salvador, Jordan, Lao PDR, Mongolia, Morocco, Niger, Palestine, Paraguay and Viet Nam. While the testing demonstrated the value of the LAMP methodological approach, it also highlighted the challenges of implementing assessments across a range of diverse locations and linguistic settings.

Based on this experience, the UIS has created Mini-LAMP, which can be adapted to meet the specific needs of countries.

Mini-LAMP sheds light on:

- the real needs of countries;
- factors that influence the distribution of literacy skills, including literacy practices; and
- data needed to produce SDG Indicator 4.6.1 and others, including:
 - Indicator 4.3.1 on youth and adult participation in education;
 - Indicator 4.3.3 on the participation rate in technical and vocational programmes; and
 - Indicator 4.6.3 on the participation of youth and adults in literacy programmes.

Mini-LAMP provides the data to help governments and their partners to:

- develop policies and allocate resources where they are needed the most based on sound evidence;

- improve the overall quality of education;
- reduce inequalities; and
- promote economic development.

Indicators 4.7.4 and 4.7.5

There are two main actions related to Indicators 4.7.4 and 4.7.5.

The first one pertains to Indicator 4.7.5 and was the object of a consultation during GAML 5 and TCG 5. The indicator was rephrased to include lower secondary; the indicator currently reads “**4.7.5** *Percentage of students showing proficiency in knowledge of environmental science and geoscience at the end of lower secondary*”.

When it comes to more concrete developments, the UIS has commissioned research with the objective of supporting the development of a measurement strategy for SDG Global Indicator 4.7.1 and Thematic Indicators 4.7.4 and 4.7.5 using International Large-Scale Assessments (ILSAs) in Education. The study has identified a global content framework for the three indicators based on existing mapping exercises. Then, it has evaluated the extent to which the different concepts contained in the content framework can be measured with the instruments and procedures of existing ILSAs. Finally, the document presents a proposal to define proficiency levels for each of the indicators based on definitions from the same ILSAs; and concludes with an overview of the three sections described above, its limitations and suggestions for developing a measurement strategy for the three indicators. The Secretariat aims to adopt the framework for Indicator 4.7.4 and to agree on the definition of “adequate understanding of issues relating to global citizenship and sustainability” and “proficiency in knowledge of environmental science and geoscience” as proposed in Indicators 4.7.4 and 4.7.5 respectively.

New indicators on breadth of skills and love for learning

Breadth of skills

The 21st century demands a broad set of skills that ranges from respect to diversity, emotional skills to creativity, ingenuity, and ‘thinking outside the box’, all important skills and determinants of future skills for the young generations. They are equivalent – and sometimes arguably superior – to knowledge. Hence, a proposal will be made during the 6th Meeting of the Technical Co-operation Group (TCG) to stimulate discussion on the feasibility and relevance of a new indicator on a broader set of skills. Throughout this document these skills will be referred to as “**breadth of skills**” or “a broad set of skills” and should be understood to refer to 21st century skills or transversal competencies. Definitional issues will be addressed in the proposed conceptual framework.

Love for Learning

It’s often said, what’s measured is what matters. It is what inspires leaders to take action by recognizing challenges and setting priorities. Extensive evidence has illustrated the relationship between intrinsically motivated children and learning outcomes, which is critically important to measure core indicators that define a child who loves learning. A proposal for the measurement of **love for learning** through behavioral indicators of intrinsic motivation (engagement, trust, safety, curiosity, etc.) is to be discussed.