



# LEARNING ASSESSMENTS AND SKILLS SURVEY DATA

---

CHALLENGES AND SOLUTIONS FORWARD

FEBRUARY 2024

2024 CONFERENCE ON  
**EDUCATION DATA  
AND STATISTICS**

## INTRODUCTION

There are several decades of development and debate on measuring progress in learning outcomes and skills. This paper analyses and assesses the status of and gaps in frameworks and methodologies that help measure and monitor indicators on learning outcomes and skills related to SDG targets 4.1, 4.4, 4.6 and 4.7 (**Table 1**). It aims to identify areas that require further attention and improvement. More specifically, the paper aims to make a meaningful contribution to the ongoing discussions and initiatives in this field, ultimately aiming to establish an international community of practice that can collectively address the challenges ahead.

**Table 1. SDG 4 indicators related to learning outcomes and skills**

Indicator		Domain	Required definitions
4.1.1	Proportion of children and young people: (a) in grades 2/3; (b) at the end of primary; and (c) at the end of lower secondary achieving at least a minimum proficiency level in (i) reading and (ii) mathematics, by sex	Reading and mathematics	Minimum proficiency level Procedural quality minimum
4.2.1	Proportion of children under 5 years of age who are developmentally on track in health, learning and psychosocial well-being, by sex	Learning, socioemotional health	Developmentally on track
4.4.2	Percentage of youth/adults who have achieved at least a minimum level of proficiency in digital literacy skills	Digital literacy skills	Relevant digital skills for employment, decent jobs and entrepreneurship
4.6.1	Percentage of population in a given age group achieving at least a fixed level of proficiency in functional (a) literacy and (b) numeracy skills, by sex	Literacy and numeracy	Fixed level of functional numeracy and literacy
4.7.4	Percentage of students by age group (or education level) showing adequate understanding of issues relating to global citizenship and sustainability	Global citizenship and sustainability	Adequate understanding of global citizenship and sustainability
4.7.5	Percentage of 15-year-old students showing proficiency in knowledge of environmental science and geoscience	Environmental science and geoscience	Proficiency in environmental science and geoscience knowledge

## ASSESSMENT OF STATUS OF MEASURING SDG INDICATOR 4.1.1

SDG indicator 4.1.1 is 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 [MPL] in (i) reading and (ii) mathematics, by sex. It refers to three levels of schooling – lower primary, upper primary and lower secondary – and two subjects – reading and mathematics.

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

- The percentage of students meeting at least a minimum proficiency level (MPL) for the relevant domains (mathematics and reading) for each point of measurement (grades 2/3; end of primary and end of lower secondary).
- Whether a programme can be considered comparable, and the conditions under which the percentage of children at or above the minimum proficiency level can be considered comparable to the percentage reported from another country.

The indicator requires the following inputs:

- Domains: Reading and mathematics.
- Minimum proficiency level (MPL): Basic knowledge benchmark in the domains at a given age/grade.
- Sample: Representative of the relevant population.
- Procedures: Compliant with minimum standards of quality.

### Challenges

There are a few critical issues in reporting indicator 4.1.1.

#### *Comparability of grades and education levels*

Because primary schooling has a different duration in different countries, terms such as ‘end of primary’ can mean different things in different places, which complicates comparisons across countries and assessment programmes. However, 89% of countries end their primary cycle in grades 5–7 so the issue might be minor.

### *Comparability of assessment results across space and time*

While the comparability of statistics across countries influences comparability over time, the latter does not imply the former.

- Cross-country comparison through cross-national assessments helps comparability across countries, at one point in time. If each assessment programme produces statistics which are comparable over time, then statistics will be comparable across time and countries.
- National assessment programmes are not comparable to each other by design, but they can still provide relatively reliable trend data if measurement quality is good enough.

### *Timeliness and policy impact of the statistics*

Assessments produce national, and often subnational, statistics which can influence policymaking and policy implementation in positive ways. For these positive impacts to be felt, statistics must not only be accurate but also must be widely seen as credible, and the turnaround time between the assessment and the reporting of results should be as short as possible.

### *Procedural quality*

Robust, consistent operations and procedures are an essential part of any large-scale assessment, to maximize data quality and minimize 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 the regional level, where the goal is to establish procedural consistency across countries. 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 formats across assessments. However, a minimum set of procedures is needed – procedural alignment – so that data integrity is protected and the results are robust as well as reasonably comparable for any given country over time, as well as across countries at any given point in time.

### *Financial costs of assessments for countries*

Assessments are relatively costly. However, even for developing countries, the cost of systematically assessing outcomes is extremely low relative to the overall cost of providing schooling and of not measuring learning outcomes and skills.<sup>1</sup>

### *Low coverage of cross-national assessments in low- and lower-middle-income countries*

SDG indicator 4.1.1 is being reported using various cross-national studies that are international ([PIRLS](#), [TIMSS](#)) or regional ([PILNA](#), [SEA-PLM](#), [PASEC](#), [LLECE](#), [SACMEQ](#)) (**Table 2**). These tools have not been designed for SDG reporting but, in 2018, the Global Alliance to Monitor Learning ([GAML](#)) and the Technical Cooperation Group on SDG 4 indicators ([TCG](#)) agreed that these assessments could be used to report learning based on their proficiency levels that best ‘mapped’ to the global MPL.

**Table 2. Assessment programmes by grade or age and use for reporting on SDG indicator 4.1.1**

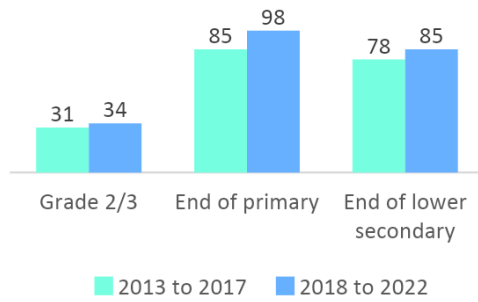
Grade	International assessment programme
SDG 4.1.1a: Early grades	
2	<a href="#">PASEC</a>
3	<a href="#">ERCE</a> , <a href="#">AMPLa</a>
SDG 4.1.1b: End of primary	
4	<a href="#">PILNA</a> , <a href="#">LaNA</a> , <a href="#">PIRLS</a> , <a href="#">TIMSS</a>
5	<a href="#">SEA-PLM</a>
6	<a href="#">LaNA</a> , <a href="#">PASEC</a> , <a href="#">PILNA</a> , <a href="#">SACMEQ</a> , <a href="#">ERCE</a> , <a href="#">AMPLb</a>
SDG 4.1.1c: End of lower secondary	
8	<a href="#">TIMSS</a>
Age: 15 years	<a href="#">PISA</a>

However, the production of comparable learning outcomes is not progressing fast and equally enough. Regardless of the coverage criterion (the number of countries or the population), coverage is much higher at the end of primary and end of lower secondary than for grades 2 or 3 (**Figure 1**).

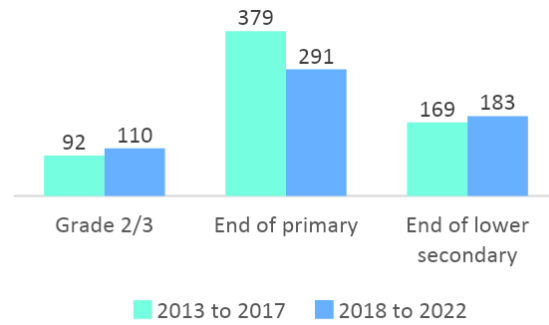
<sup>1</sup> For more information on costing, see UIS (2023). *Reporting learning outcomes in basic education: Country’s options for indicator 4.1.1*: [https://gaml.uis.unesco.org/wp-content/uploads/sites/2/2023/05/Countrys-reporting-option\\_Zambia\\_2023.05.15\\_FINAL.pdf](https://gaml.uis.unesco.org/wp-content/uploads/sites/2/2023/05/Countrys-reporting-option_Zambia_2023.05.15_FINAL.pdf).

**Figure 1. Coverage of learning assessments, by level of education**

a. Number of countries



b. School-age population in millions



#### *National assessments: alignment and procedural quality*

While data from many national learning assessments are readily available, every country sets its own standards, leading to inconsistent definitions of performance levels. 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 establish a common reference level for national assessments and the set of procedures used for sampling, data management and reporting also differ.

The second aspect that national assessments programmes should comply with is procedural quality. Reporting information on processes for national programmes should be publicly available and with enough level of detail. The UIS has provided guidance and self-assessment tools pertaining to this aspect in the [Aligning and reporting on indicator 4.1.1: UIS annotated workflow](#).

#### **Standards**

Given these challenges, the UIS has focused on defining (i) the minimum proficiency level, aligning it to a competency concept that is independent of a particular assessment framework, specific items or tests, to allow reporting and (ii) a set of linking strategies to the proficiency framework.

### *Minimum proficiency level*

The minimum proficiency level (MPL) is the benchmark of basic knowledge in a domain (e.g. mathematics, reading) at a given age/grade measured through learning assessments. The MPL is a reference point for reporting on minimum competencies at each schooling level, without requiring a single test to solve comparability.

The MPL proficiency level descriptor is the key standard for each grade and domain that allows the use of multiple assessments to report for the indicator (Table 3).<sup>2</sup> In 2018, an agreement was reached with the cross-national assessment programmes on which of their descriptors was better aligned to the MPL descriptor, although further validation is needed, through a standard-setting exercise, as the assessment programmes have not been designed to measure SDG 4.

**Table 3. Minimum proficiency levels for reading and mathematics for indicator 4.1.1**

Educational level	Descriptor	
	Reading	Mathematics
Grade 2	Students read and comprehend most of written words, particularly familiar ones, and extract explicit information from sentences.	Students demonstrate skills in number sense and computation, shape recognition and spatial orientation.
Grade 3	Students read aloud written words accurately and fluently. They understand the overall meaning of sentences and short texts.	
Grade 4 to 6	Students interpret and give some explanations about the main and secondary ideas in different types of texts. They establish connections between main ideas on a text and their personal experiences as well as general knowledge.	Students demonstrate skills in number sense and computation, basic measurement, reading, interpreting, and constructing graphs, spatial orientation, and number patterns.
Grade 8 to 9	Students establish connections between main ideas on different text types and the author's intentions. They reflect and draw conclusions based on the text.	Students demonstrate skills in computation, application problems, matching tables and graphs, and making use of algebraic representations.

Source: UNESCO Institute for Statistics, 2021, p. 4.

As part of this process, it was necessary to:

<sup>2</sup> It was agreed to report according to the textual definition of the minimum proficiency level for each domain and level in the cross-national assessments. This was established by conducting an analysis of the performance-level descriptors of these assessments in reading and mathematics.

- Approximate the grade levels of interest for reporting
- Use assessment programme existing proficiency levels
- Identify the proficiency level descriptor that is better aligned to the minimum proficiency levels
- Use that level to report until the standard setting exercise is finalized
- Build national technical capacity but do not directly address national assessment development.

### *Global proficiency framework*

The global proficiency framework (GPF) is a useful global reference, which defines a common reference for proficiency levels for reading and mathematics that learners are expected to demonstrate at the end of each grade level, from grades 1 to 9. The four levels outlined in the GPF – ‘below partially meets’, ‘partially meets’, ‘meets’ and ‘exceeds global minimum proficiency’ – form a common scale from low to high achievement. The GPF and its related proficiency levels give guidance on the minimum set of skills that students should acquire on the pathway to mastery of reading and mathematics.

### *Linking assessment programmes to the minimum proficiency level*

Linking a national, regional or international assessment to the global MPL definition requires a methodology to identify the same concepts and definitions in assessment programmes built for completely different purposes to allow some degree of comparability and fair inferences about the countries compared.

The process of making different assessments comparable (moderation) can be statistical or non-statistical. Three processes are the Rosetta Stone project, the policy linking methodology, and the Assessments for Minimum Proficiency Levels (AMPLs).<sup>3</sup>

---

<sup>3</sup> For more information on linking strategies, their costs, benefits, execution status, executed and pending milestones, and time frames see UIS (2023) *Reporting learning outcomes in basic education: Country's options for indicator 4.1.1*  
[gaml.uis.unesco.org/wp-content/uploads/sites/2/2023/05/Countrys-reporting-option\\_Zambia\\_2023.05.15\\_FINAL.pdf](https://gaml.uis.unesco.org/wp-content/uploads/sites/2/2023/05/Countrys-reporting-option_Zambia_2023.05.15_FINAL.pdf).



### **Rosetta Stone**

The Rosetta Stone project, led by the International Association for the Evaluation of Educational Achievement (IEA), tried to harmonize data from different assessments. Named after the famous archaeological discovery that enabled translation between different written languages, it is designed to relate regional to international learning assessments. The goal is to provide countries that participated in regional (or national) but not in international assessments with information about the proportion of students who achieved MPL in reading and mathematics. In a first effort to implement this approach and establish concordance tables, the regional assessments Regional Comparative and Explanatory Study (ERCE) and Programme d'analyse des systèmes éducatifs de la CONFEMEN ([PASEC](#)) were linked to the Trends in International Mathematics and Science Study ([TIMSS](#)) for mathematics and the Progress in International Reading Literacy Study ([PIRLS](#)) for reading.

### **Policy linking**

Another approach to harmonize assessments is the policy linking methodology, which is a non-statistical method that uses judgment to align and match items from the national assessment with the GPF. This process establishes internationally comparable global benchmarks based on the descriptors of each GPF benchmark. Three major tasks – alignment, matching and setting benchmarks – are to be completed in a workshop of 5 to 6 days with 15 to 20 panellists (teachers), curriculum and assessment experts of each grade/subject, to identify and set, if feasible, the required benchmarks for international reporting on SDG indicator 4.1.1.

To produce reliable benchmarks for international reporting, the Policy Linking Toolkit specifies five criteria including that a sufficient number of national items are aligned with the GPF; samples are nationally representative; and national assessments are administered according to minimum quality standards. The policy linking methodology was proposed in 2017, piloted in 2019, revised in 2020 and piloted again in 2021–22. The toolkit was then revised in 2023, and is now in a pilot phase.

### **Assessments for Minimum Proficiency Levels**

Assessments for Minimum Proficiency Levels (AMPLs) have also been proposed as a way to obtain data that can be used to measure and monitor SDG indicator 4.1.1. AMPLs are robust

tools targeted at measuring the attainment of MPL in reading and mathematics at a given level of the education cycle. This allows countries to produce internationally comparable learning outcomes data to report on SDG global indicator 4.1.1.

AMPL-a, which measures proficiency in early grades, is under development and will be piloted and administered in 2023, in both English and French. AMPL-b, which measures proficiency at the end of primary, was developed in 2021 in both English and French and was implemented in six African countries as part of the Monitoring Impacts on Learning Outcomes project: Burkina Faso, Burundi, Côte d'Ivoire, Kenya, Senegal and Zambia. AMPL-b was administered as a stand-alone module in Sierra Leone in 2022 and is scheduled for implementation in Jordan and Pakistan.

#### *Countries' alternatives for reporting*

To guide the choice of learning measurement and to ensure assessment data are consistent with long-term strategic goals of effective decision-making, the UIS, UNESCO, World Bank and UNICEF have developed a set of principles based known as the Learning Data Compact. These principles are important not just for designing assessments or deciding which assessment to use 'off the shelf,' but also for developing national assessment systems. Those principles build on existing ones; allow the flexibility to ensure alignment with country needs (they are not 'one size fits all'); foster country ownership through a demand-driven approach; and ensure data are relevant for decision-making.

A national assessment system should be able to do reporting, manage improvement at all levels of education, guide decision making and link system-level assessments to formative assessments and classroom practices. To ensure assessments can accurately monitor progress for decision making, data must be internationally comparable. Every country should have an assessment that is designed for, or can be used for, international comparability – a commitment in the SDG process. Countries' options to report are various (**Table 4**) but the choice should be guided by what assessment(s) are fit-for-purpose and most cost-effective, taking into account the country's initial situation and the objective to have comparability over time and representativeness of results at the national level.

Two special cases should be noted. National assessments could be used to report, subject to the use of statistical linking that could be implemented using calibrated modules such as AMPL. Tools such as the MPL and the GPF serve to understand and benchmark to global standards, while policy linking can engage national stakeholders to analyse the assessment vis-à-vis those standards. The UIS considers countries that want to report globally on indicator 4.1.1. The UIS adds a calibrated module to the national assessment, such as AMPL, and supplements this with policy linking for capacity-building purposes, given the methodology is still in a pilot phase.

A second case is related to early grades, or 4.1.1a, where tools such as the Early Grade Reading/Mathematics Assessment (EGRA/EGMA), the PAL Network citizen-led assessments and UNICEF’s Foundational Learning Module of its MICS household survey could be relevant and be used for reporting. Even though these assessments have been applied globally, they cannot currently be used for global reporting, mostly because they were not intended to generate comparable data. Nevertheless, they have the potential to be used for global reporting and the UIS is looking into how to make the best use of such assessments.<sup>4</sup>

**Table 4. Alternatives for country reporting on SDG indicator 4.1.1**

	4.1.1.a	4.1.1.b	4.1.1.c	Coverage
National assessments – statistical linking through calibrated modules				
AMPL	●	●		
PISA module			●	
Cross-national assessment participation				
PILNA		●		Pacific islands
PASEC	●	●	●	Mainly Africa (Francophone)
SACMEQ		●		Africa (Southern and Eastern)
SEA-PLM		●		Southeast Asia
LLECE	●	●		Latin America
TIMSS	●	●	●	Global
PIRLS	●	●		Global
PISA			●	Global

<sup>4</sup> See <https://world-education-blog.org/2023/09/13/compare-align-track-the-foundational-learning-data-challenge/>.

## ASSESSMENT OF STATUS OF MEASURING INDICATORS 4.6.1, 4.7.4 AND 4.7.5

Significant progress has been made in the establishment of conceptual, methodological and reporting frameworks for indicators 4.6.1, 4.7.4 and 4.7.5. The frameworks provide a structured approach and offer guidelines and principles for data collection and analysis. However, despite the progress in defining frameworks since 2015, a substantial problem is low data coverage. Addressing this challenge requires a concerted effort to establish common definitions and metrics, ensuring a standardized approach across assessments (Table 5).

**Table 5. Availability of assessments to measure SDG indicators 4.6.1, 4.7.4 and 4.7.5**

Indicators	Metadata	Data sources	Last administered	Cycle length	Coverage countries (N)	Coverage population (%)
4.6.1	<a href="#">Yes</a>	Programme for the International Assessment of Adult Competencies (PIAAC)	2017	10 years	37	21
4.7.4	<a href="#">Yes</a>	International Civic and Citizenship Study (ICCS)	2016	6/7 years	23	10
4.7.5	<a href="#">Yes</a>	TIMSS, PISA	2019/2022	4/5 years	38	16

Low- and middle-income countries often lack the resources and infrastructure needed to develop and implement tools to measure these indicators effectively. Addressing this issue necessitates not only the development of appropriate measurement tools but also targeted capacity-building initiatives in low- and middle-income countries to ensure that these indicators are comprehensively and accurately measured.

The UIS has proposed a solution to increase reporting of SDG global indicator 4.6.1. Building on its experience with the Literacy Assessment and Monitoring Programme (LAMP), it has developed a reduced version of the survey to lower the operational, financial and technical burden. In particular, this mini-LAMP survey would, among other adjustments, reduce the number of skill domains assessed, adapt the sampling procedures, and offer the possibility of

introducing the survey as a module to an existing household survey, rather than as a standalone survey.<sup>5</sup>

There is currently no alternative to increase reporting on SDG indicators 4.7.4 and 4.7.5, other than more countries joining the cross-national assessments that measure them. However, a similar approach of a module that is focused on measuring the global agreed standard of what is an adequate understanding of issues relating to global citizenship and sustainability (indicator 4.7.4) and, similarly, the proficiency in knowledge of environmental science and geoscience (indicator 4.7.5) could be explored.

## CONTEXT QUESTIONNAIRES

Beyond harmonizing data on learning outcomes and skills, harmonizing context questionnaires from different large-scale assessments is critically important for enabling robust comparative analyses of trends, patterns and determinants of educational inequality across countries and over time.

While the primary objective of cross-national assessments is to estimate learning outcomes, they also collect a rich set of background information:

- From students, they collect information about their school experiences, their attitudes towards subjects being taught, and the characteristics of their parents and households in addition to core demographic information, such as age and sex (Table 6).
- From teachers, they collect information about their attitudes towards teaching, their opinions about teaching resources, their educational background and ongoing professional development.
- From schools, they collect information about infrastructure, location and opinions from the school directors about the availability of resources at school and how they interact with parents. There is some variation in the objectivity of the data collected.

---

<sup>5</sup> See <https://uis.unesco.org/sites/default/files/documents/mini-lamp-monitoring-progress-sdg4.6.1-2018-en.pdf>.

**Table 6. Typical questionnaires and data collected in cross-national assessments**

Questionnaire	Typical question areas
Student	Demographic information (sex, age) Household and socioeconomic background School-related experiences (including exposure to bullying) Learning-related experiences (classroom activities) Self-perceptions, interests and aspirations related to different subjects Use and proficiency of information and communication technology (ICT)
Teacher	Demographic and background information (sex, age, years teaching, subjects taught) Qualifications and training Types of teaching practices used and challenges faced
School (director)	School characteristics Opinions about availability and adequacy of resources Management and governance Interaction with parents and school communities

Some of these questionnaires provide information for selected SDG 4 indicators. These questionnaires also determine for which subpopulations these indicators can be estimated based on characteristics such as the urban or rural location of schools, student socioeconomic status (relative to other students, not to the population), and the sex of students and teachers, all of which can inform discussions on equity.<sup>6</sup>

There are significant challenges to developing policy advice, due to differences in questionnaire definitions across various dimensions, such as rural/urban classifications, socioeconomic status (or wealth), period of reference, and more. Differences in the definitions hinder assessment outcome comparability and data interpretation. For instance, what constitutes ‘rural’ in one country might be different in another. Similarly, varying definitions of socioeconomic status can impact the analysis of disparities in educational outcomes by different social or economic strata. Bridging these definitional gaps requires international collaboration and the development of standardized frameworks that ensure uniformity in definitions, enabling accurate assessments and facilitating meaningful comparisons across regions and socioeconomic contexts.

<sup>6</sup> For more information, refer to UIS (2022). *Monitoring of the Sustainable Development Goals using large-scale international assessments*  
<https://tcg.uis.unesco.org/wp-content/uploads/sites/4/2022/04/Monitoring-of-the-SDGs-Using-Large-Scale-International-Assessments-April-2022.pdf>.

## **AGENDA FORWARD**

The following solutions are proposed for the issues identified in this paper.

### **Assessment harmonization and reporting handbook**

With the progress made in recent years, it is time to compile and regularly update a handbook with all the information on eligibility criteria for reporting. This handbook will serve as a technical guidance for learning assessments programs and will feature and update the menu of assessments suitable for reporting. The aspects to be included are:

- Alignment to learning standards and frameworks (assessments must adequately measure the intended curriculum and skills)
- Psychometric properties (evidence of reliability, validity, appropriate difficulty, and discrimination among others)
- Representativeness (samples must reflect target populations.)
- Comparability of administrations (consistent, standardized administration procedures affect comparability)
- Transparency of processes (assessment design, sampling, analysis should be well documented)
- Capacity for linking (enough equivalent items/proficiency levels to enable linking)
- Stakeholder involvement (consultative process throughout design and implementation)
- Feasibility of participation (reasonable costs, schedules, capacity-building and burdens for countries)

### **Accreditation system**

Accordingly, it is also the right time to introduce a clear and transparent accreditation system. Assessment providers, including government organizations, will be able to apply to having assessments vetted for their fitness of purpose to report on SDG indicator 4.1.1. Based on the handbook, a checklist will contain the standards and eligibility criteria with which applicants need to comply.

### **Increase investment and approach to fund learning data**

Data coverage remains uneven, especially for developing countries. More investment to expand learning assessment data globally is critical with a holistic approach that looks at building infrastructure for data production at the national level transferring knowledge and skills but also purchasing power and decision-making.

#### **Promote the standardization of context questionnaires in learning assessments.**

Context questionnaires is an important next step to support comparability. Agreement on key definitions and agreement of standard items and format to capturing individual student, teacher, and school characteristics would enable better comparability of equity and drivers of learning dimensions.

#### **Scale up the mini-LAMP approach to adult literacy measurement to increase coverage of SDG indicator 4.6.1**

The low coverage of SDG global indicator 4.6.1 means that it will be deleted from the list of global indicators during the 2025 Revision by the Inter-agency and Expert Group on SDG indicators. Yet adult skills will remain on the global education agenda and a cost-effective solution remains a priority. Member States are invited to pilot the mini-LAMP tool, which can be added to existing household surveys.

#### **Innovative Methodologies on indicators with low coverage**

SDGs 4.7.4 and 4.7.5 deserve the exploration of approaches mirroring the AMPL approach of 4.1.1 by enabling a module that measures the minimum agreed standards taking advantage of existing assessments, such as the IEA International Civic and Citizenship Study (ICCS).