



United Nations
Educational, Scientific and
Cultural Organization



UNESCO
INSTITUTE
FOR
STATISTICS



TECHNICAL
COOPERATION
GROUP



TCG4: Measurement Strategies

TCG4/23

16-18 January 2018

Dusit Thani Dubai

133, Sheikh Zayed Road, Trade Centre,

Dubai, United Arab Emirates





Contents

SDG Indicator 4.1.1	3
SDG Indicator 4.2.1	12
SDG Indicator 4.4.2	17
SDG Indicator 4.6.1	20
Proposal by GAML Task Force 4.6 (OECD proposal).....	20
Proposal by GAML Task Force 4.6 (UIL proposal)	27
Proposal by GAML Secretariat.....	32
SDG Target 4.7	38
Annex A: Concept Paper on the Short Literacy and Numeracy Survey (SLNS)	44
Annex B: Proposal to the IAEG-SDGs	53



SDG Indicator 4.1.1

Inputs to the Measurement and Reporting

Introduction

Sustainable Development Goal (SDG) Target 4.1 focuses on free, equitable, and quality primary and lower-secondary education. The Global Indicator (4.1.1) for Target 4.1 is the “*proportion of children and young people in Grade 2 or 3 (4.1.1a), at the end of primary education (4.1.1b), and at the end of lower secondary education (4.1.1c) who achieve at least a minimum proficiency level in reading and mathematics*”.

Task Force 4.1 of the Global Alliance to Monitor Learning (GAML) was convened to support UNESCO Institute for Statistics (UIS) in thinking through the measurement issues involved in reporting against this indicator and to help them come up with practical solutions. The Task Force’s deliberations have run parallel to other work on these issues by UIS and its technical partners and other stakeholders. **The objective of this Task Force 4.1 document is to serve as an input to this ongoing work program¹ by offering Task Force member insights and recommendations on some of the key measurement and reporting challenges.**

Key Challenges and Work Program for Indicator 4.1.1

There are several key challenges involved in measuring and reporting on reading and mathematics outcomes at the global level. These include mapping the content coverage of different assessments onto a common framework (in the absence of a common assessment instrument); developing a relevant learning scale; ensuring a certain level of data quality across assessments; establishing a coherent reporting metric; agreeing on the level of achievement that qualifies as “minimum proficiency” in different national contexts; and building country capacity to produce the needed data and manage financial and human resource allocation.

Task Force 4.1 addressed these challenges in relation to three key phases in an assessment work program:

Conceptual framework: Who and what to assess?

Methodological framework: How to assess?

Reporting framework: How to report?

A summary of Task Force discussions and conclusions for each phase is described in the rest of this report. In the course of these discussions, **Task Force members also arrived at some overall recommendations for next steps in the Indicator 4.1.1 work program. These included:**

The GAML Secretariat/UIS should convene a group of reading and mathematics content experts, developmental psychologists, assessment experts, and others who can bring the latest research,

¹ For example, see the August 2017 document, “*SDG Data Reporting: Proposal of a Protocol for Reporting Indicator 4.1.1*”, as well as the summary report for the technical expert meeting held in Hamburg, Germany from September 20-22, 2017. These documents outline interim and longer-term approaches to measurement and reporting for Indicator 4.1.1.

evidence, and data to bear on the drafting of the longer-term measurement strategy for Indicator 4.1.1, particularly Indicator 4.1.1a. This group of experts should be diverse in terms of regions, languages, and scripts.

Countries need to be more actively brought into the discussions on Indicator 4.1.1 to ensure that the proposed measurement and reporting approaches are sufficiently adaptive and responsive to their contexts. It's unclear, however, whether GAML is the context in which these country consultations should take place.

Conceptual Framework: Who and What to Assess?

The general view of Task Force 4.1 members was that existing national and cross-national assessments should provide the basis for determining who and what to assess at the three key measurement points. Table 1 provides an overview of some of these national and cross-national assessment options.

Table 1. Overview of some existing national and cross-national options for who and what to assess

National Assessments	Cross-National Assessments			Related GAML/ Activities	Expected UIS Outcomes / Product	Timeline
	Name (# countries)	Reading	Mathematics			
Mapping of national assessment frameworks for mathematics and reading (in progress)	LANA	Grade 4-6	Grade 4-6	Map national assessment frameworks	Content Reference Frameworks (Math, Reading)	2017
	LLECE (15)	Grade 3/6	Grade 3/6	Map assessment characteristics and use of assessment data	Catalogue of Learning Assessments	2017
	PASEC 2014 (10)	Grade 2/6	Grade 2/6			
	PILNA (13)	Grade 4/6	Grade 4/6			
	PISA 2015 (72)	Age 15	Age 15			
	PISA-D 2018 (8)	Age 15	Age 15			
	PIRLS 2016 (61)	Grade 4				
	SEA-PLM 2018 (11)	Grade 5	Grade 5			
	SACMEQ IV (14)	Grade 6	Grade 7			
	TIMSS 2015 (57)		Grade 4/8			

 Product was formally reviewed by Task Force 4.1 members.

Most of the national and cross-national assessments shown in Table 1 are designed to provide grade-based data on reading and mathematics performance that is relevant to measurement points 4.1.1b



and 4.1.1c. A Task Force subgroup was convened to discuss options for 4.1.1a in more depth. The subgroup determined that assessment at this level needed to focus on the precursor and early reading and mathematics skills required for future academic learning and that key facets of performance to be considered should include accuracy, comprehension, and automaticity/speed. It was recognized, however, that very few cross-national assessment programs currently measure these precursor and early skills (e.g., LLECE/TERCE in grade 3; PASEC in grade 2).

There was general agreement among Task Force members that ongoing **GAML/UIS activities to map assessment frameworks and capture the characteristics and uses of assessment data should continue, but with more focus on ensuring that these efforts include attention to grades 2/3 assessments in reading and math (4.1.1a)**. Given the relative dearth of national and cross-national assessment options for the early grades, the Task Force 4.1 subgroup also suggested that countries consider drawing on early-grades reading assessment (EGRA), early-grades mathematics assessment (EGMA), household-based (e.g., Multiple Indicator Cluster Survey [MICS]), and citizen-led (e.g., Annual Status of Education Report [ASER] and UWEZO) tools that measure key aspects of the constructs of early reading and math. **For the longer-term, the subgroup recommended developing a set of purpose-built tools that countries could draw on/adapt.**

One of the key outputs produced by UIS and its technical partners to support countries' efforts in reporting against Indicator 4.1.1 is a set of Content Reference Frameworks for mapping mathematics and reading assessment frameworks. Task Force 4.1 members were invited to submit feedback on the draft Content Reference Framework for mathematics. **Task Force 4.1 members signaled general agreement with the approach taken to developing the mathematics content reference framework, but also:**

Concern about the possible influence of a restricted number of language groups and cultures on the current version of the framework in terms of the relevance and adequacy of its domain coverage.

Requests for a more explicitly research-based approach to delineating key subdomains/levels and for the inclusion of more concrete examples for each sublevel.

Recommendations to apply the framework to a greater variety of national assessment frameworks in order to further refine and validate it. This might include determining how well the reference framework applies to a national assessment in a top-ranked PISA country and whether certain aspects of national assessment frameworks are deemed unsuitable for the Content Reference Framework (and why).

Requests to provide more information on how the framework might be adapted over time.

The draft Content Reference Framework for reading ("*Method for Developing an International Curriculum and Assessment Framework for Reading and Writing*") was circulated at a much later date. Task Force 4.1 members were not formally required to submit feedback on this framework, but were invited to do so if they had time. **Submissions by Task Force members on the draft reading framework methodology paper can be summarized as follows:**

Appreciation for the presentation of reading interlinked with writing as part of the broader construct of literacy.

Recommendation that given Indicator 4.1.1's focus on reading, that aspect of the literacy construct should be emphasized moving forward.

Recommendation to extend/test the framework against other languages, apart from alphabetic and European; at the very least, the framework should be tested against the remaining United Nations' languages of Arabic, Chinese, and Russian.

Concern that the framework is based on the perspective of one discipline (psychology) and one school of thought within that discipline (cognitive psychology) and that other perspectives and evidence bases (e.g., linguistic and sociolinguistic) should be incorporated.

Request for a more explicitly research-based approach to constructing the Content Reference Framework and for more specialist input.

Methodological Framework: How to Assess?

Task Force 4.1 members did not focus as much on the "how to assess" aspect of indicator 4.1.1, which seemed to end up falling more under the purview of the Assessment Implementation Task Force. Task Force 4.1 was not formally requested to review any of the technical outputs in this area. Table 2 provides an overview of some national and cross-national assessment approaches. Most emphasize sample-based and group-administered approaches, and also primarily focus on children and youth in school.

Table 2. Overview of some existing national and cross-national options for how to assess

National Assessments	Cross-National Assessments*			GAML/UIS Activities	Expected Outcomes/Products	Timeline
	Name	OOSC	Individual/group administration			
UIS Catalogue of Learning Assessments to provide information on methods used by national assessments, including sampling, administration, and quality checks	LANA	No	Group	Develop and compile good practices in learning assessment	Good Practices for Learning Assessment Manual	2017
	LLECE	No	Group			
	PASEC 2014	No	Both	Evaluate alignment in assessment content	Content Alignment Framework	2017-2018
	PILNA	No	Group			
	PISA 2015	No	Group	Evaluate data collection process	Data Quality Framework	2017-2018
	PISA-D 2018	Yes	Both			
	PIRLS 2016	No	Group			
	SEA-PLM 2018	No	Group			
	SACMEQ IV	No	Group			
	TIMSS 2015	No	Group			

* All sample-based



A Task Force 4.1 subgroup was convened to discuss 4.1.1a measurement approaches in more depth. Most early-years assessments are designed for one-on-one administration. EGRA, EGMA, and all household-based and citizen-led assessments use one-on-one approaches for this age/grade level, although school-based assessments employ a mix of one-on-one and group-administered approaches. In the short-term, it was felt that all of these should be viewed as options for countries to consider. **In the longer-term however, it was noted that there might be value in moving towards more school-based and group-administered approaches given the attendant savings in cost, time, and efficiency.**

Three key “how to” issues were addressed by Task Force 4.1 members during their virtual group discussions. These included:

How to include out-of-school children in measurement and reporting?

Task Force 4.1 members discussed whether and how to adjust school-based assessment results for countries with sizeable out-of-school populations as a way to reduce the bias produced by a non-representative sample. This included discussion on whether citizen-led assessments could be used to complement school-based assessment in such contexts given their coverage of both in- and out-of-school populations. Task Force 4.1 members felt that countries with this issue could be encouraged, but not required, to report data from their citizen-led assessments (if available) as an additional source of information on learning levels. In addition, countries could be encouraged to report the percentage of their student populations that are actually in school, but this statistic should not be used to adjust assessment results for (or otherwise ‘punish’) countries, at least not in the first phase of reporting under 4.1.1.

How to determine “minimum acceptable requirements” for assessment data?

Suggested requirements ranged from very specific technical and psychometric criteria (e.g., reliability and validity coefficients, sample size requirements) to more content-related requirements regarding the breadth and appropriateness of the content being assessed. Task Force members noted, however, that it would not be fair to prescribe very precise technical criteria that countries are unlikely to have been aware of ahead of time. Instead, “minimum acceptable requirements”, at least initially, would be more along the lines of ensuring that the submitted data are nationally representative and consistent with the national curriculum/standards. Evidence that the data are comparable over time also would be critical. More detailed technical and psychometric criteria could be used as a basis for country capacity building and system strengthening over time. It also was suggested that UIS request countries to submit their data sets, in addition to their assessment instruments, as part of the validation process. Reporting of results would then be accompanied by a “report card” of sorts on the quality of the underlying data. This would signal to the global community the extent to which the data could be “trusted” while at the same time providing a basis for countries and donor partners to determine capacity building needs.

How to decide which assessment data should be used for reporting?

This issue is likely to come up for countries that have participated in international and/or regional assessments in addition to their own national learning assessment. Task Force members considered whether countries should be given the freedom to choose which assessment data to report, or whether the decision should be made more centrally.

The sense among Task Force 4.1 members was that it would be important to be flexible on these and other decisions early on and focus more on encouraging countries to get in the habit of submitting data on learning. At the same time, efforts should be made to create incentives for countries to participate more systematically in international and regional assessments. From the UIS perspective, it would make sense to have a standardized protocol for making decisions about which of the data sources available for certain countries should be used for reporting against indicator 4.1.1. If all of the assessments meet basic technical quality requirements, then perhaps UIS could let countries choose which to use?

Reporting Framework: How to Report?

Table 3 provides an overview of some national and cross-national assessment reporting options. Most cross-national assessments convert raw scores to scaled scores using IRT approaches. In general, results are reported both in the form of scaled scores and/or as the percentage of students reaching specific proficiency levels or benchmarks on the scale. Each proficiency level tends to be accompanied by a description of what students at these levels are likely to know and be able to do. At the national level, the situation is more varied. Many national assessments, particularly in developing contexts, still report results as a mean raw score or percentage. Many do not have proficiency level descriptors or any benchmark for what constitutes “minimum proficiency”.

Table 3. Overview of some existing national and cross-national options for how to report

National Assessments	Cross-National Assessments		GAML/UIS Activities	Expected Outcomes/Products	Timeline
	Name	Proficiency Levels (#)			
UIS Catalog of Learning Assessments to provide information on reporting methods used by national assessments, including use of scales, proficiency levels, and other benchmarks	LANA	To be determined	Define indicators and metadata	Glossary of Common Language and Terminology	2017-2018
	LLECE	5			
	PASEC 2014	4 (numeracy); 5 (literacy)	Develop reporting protocol	Interim reporting	2017
	PILNA	9			
	PISA 2015	6	Develop UIS reporting scale	Learning Progression Explorer and Reporting Scale	2017-2019
	PISA-D 2018	6			
	PIRLS 2016	4			
	SEA-PLM 2018	To be determined	Benchmark and define minimum proficiency level	Proficiency Level Definition	2018
	SACMEQ IV	8			
TIMSS 2015	4				



Product was formally reviewed by Task Force 4.1 members.



Task Force members noted that a key challenge in reporting, particularly in relation to Indicator 4.1.1a, was comparability across systems and languages. Early-years assessments tend to focus on precursor or early reading and math skills. If these instruments have to be translated into different languages, it can affect their relative difficulty and hence the comparability of results. Because of this, some early-years assessments (e.g., EGRA) avoid comparing results (e.g., precursors, fluency measured in words correct per minute) across languages and others (e.g., MICS) focus on skills that are less affected by differences across languages (e.g., accuracy, comprehension). Task Force member suggestions included: (i) possibly using a hybrid approach of translation and adaptation to balance the relative difficulty of instruments across languages and enhance comparability, and (ii) prioritizing comparisons within languages, at least to start with.

One of the key outputs produced by UIS and its technical partners to support countries in reporting against Indicator 4.1.1 are the UIS Reporting Scales for mathematics and reading. **Task Force 4.1 members were invited to submit feedback on the draft UIS Reporting Scale Concept Note (July 2017 version). Task Force member feedback can be summarized as follows:**

Recognition of the huge amount of work that had gone into developing the reporting scale, but, at the same time, noting some serious conceptual issues:

Whether such a scale is even required – Indicator 4.1.1 does not refer to a metric per se

Whether such a scale could ever truly allow for comparisons of student outcomes across countries

Whether such a scale might inadvertently dominate 4.1.1 discussions entirely, excluding a focus on the more important, broader learning agenda

Task Force members also voiced concern about the lack of clarity regarding the relationship between the UIS Reporting Scales and the Content Reference Frameworks and requested further clarification on how these would work in unison.

The Task Force was divided as to whether work on the scale should proceed or if alternatives should be sought.

Those in favor of continuing work on the scale suggested being clearer about the objective and target audience; e.g., is this primarily a “formative tool” for education systems to assist in monitoring and developing educational quality, or is it primarily a tool for international reporting?

Those in favor of alternatives to the current scale suggested:

Using a more traditional reporting scale that uses descriptors (such as below basic, basic, proficient, and advanced) to describe different achievement levels. This would involve first agreeing on the scale and proficiency levels/descriptors against which student performance should be measured, then identifying the instruments or items that fit the respective levels, and then dealing with the empirical part.

A methodology that allows for comparisons across assessments at each of the three points (4.1.1a, 4.1.1b, and 4.1.1c), but not necessarily spanning/connecting the three points.



Giving more attention to further development of existing cross-national assessments, in order to use these as a stepping stone for capacity strengthening and development of national assessments in countries.

Task Force members were generally supportive of the proposed empirical approach to validating the UIS Reporting Scale and offered the following additional suggestions:

Provide more detail on the country-level implementation workplan.

Ensure that in-country Task Teams include teacher union representatives and academics as well as specialists.

Carry out the proposed work in close cooperation with existing cross-national assessment programs and give a prominent role to regional assessment programs.

Ensure that key stakeholders (including international and regional assessment organizations) have an opportunity to review the scale once it has been prepared.

Consider how to address the potential risks incurred by using larger countries as “representative” of any particular region, perhaps by using a regional rather than a country-level approach in instances such as Oceania.

Conduct the assessments needed to validate the scale in countries where international and regional assessments have already taken place – this would reveal how the scales perform in different country contexts with the same assessments, and whether the performance levels match up across countries.

Explore a test-based linking exercise for each of measurement points 4.1.1a, 4.1.1b, and 4.1.1c, instead of an item-based linking exercise.

Another of the key outputs produced by UIS and its technical partners to support countries' efforts in reporting against Indicator 4.1.1 is guidance for the Setting Benchmarks on the UIS Reporting Scale. **Task Force 4.1 members were invited to submit feedback on the draft Concept Note on Setting Benchmarks on the UIS Reporting Scale (June 2017). Task Force member feedback on the draft Concept Note can be summarized as follows:**

Should there be global or national “minimum proficiency” benchmarks on the scale?

There was an even split among Task Force 4.1 members on this issue, with similar numbers in favor of each option.

Should there be 1 or 3 “minimum proficiency” benchmarks per domain (i.e., mathematics and reading)?

The overwhelming majority of Task Force members were in favor of 3 benchmarks per domain; i.e., a “minimum proficiency” benchmark for each of the three measurement points – 4.1.1a, 4.1.1b. and 4.1.1c.

Should existing “minimum proficiency” benchmarks be adopted or should new benchmarks be set?



Task Force members offered arguments in favor of both options. There were slightly more Task Force members in favor of adopting existing “minimum proficiency” benchmarks, although there was also recognition that over time there might be a need to set more customized benchmarks as a result of lessons learned from countries’ data and experiences.

Should there be global or national performance expectations for the percentage of students expected to reach “minimum proficiency”?

There was an even split among Task Force 4.1 members on this issue, with similar numbers in favor of each option.

Should there be status- or progress-based expectations for the percentage of students expected to reach “minimum proficiency”?

Task Force members offered arguments in favor of both options. However, more Task Force members were in favor of having status-based expectations for the percentage of students expected to reach “minimum proficiency”.

An overriding concern of Task Force Members was how to ensure that the benchmarking and other reporting strategies adopted for Indicator 4.1.1 would optimize the relevance and utility of the results for schools. In other words, how can we ensure that the results will have meaning for schools and that they will be able to take action on them?



SDG Indicator 4.2.1

Measurement/Reporting strategy. Proposal by GAML Task Force 4.2

Introduction

On 27 October 2018 an expert meeting hosted by GAML at the Brookings Institution, in Washington, D.C., took place. (See list of attendees present). Presentations were made and a rich discussion ensued. The 4.2.1. Task Force made recommendations after the workshop. The following are the conclusions and steps forward as seen from the Task Force, based on the discussion on 27th October and various report-backs and discussions post meeting.

This document, therefore, comprises recommendations for a strategy for interim *reporting* on 4.2.1 and at the same time outlines a strategy for medium-term improvement of both measurement *and* reporting. We define “interim” as the period starting essentially with the present moment, and extending until a sufficient improvement in *measurement* and data *collection* has been made to warrant a change in the *reporting* process. It is estimated that this could take two or three years. This does not mean that improvements in measurement and collection of data would await until the end of the interim period. Various partners will proceed to analyze, develop better measurement and data collection procedures, etc., starting at the present moment and going on until such a time as sufficient changes and evidence have accumulated so as to warrant a change in reporting policy.

There were two key points of agreement or “sense of the meeting” that are worth highlighting before moving on to the recommendations.

It was noted that given the definition of “interim” as starting at the present moment, interim reporting will use only existing data. Acknowledging that existing data may be flawed in important ways, because interim reporting begins now, it is not possible to wait for interim reporting until basic improvements in measurement and data collection are made.

It was agreed that for a child to be “developmentally on track” the child would have to be on track in all three domains not just in one or two of the three domains. However, what it means to be “on track” still needs further work as described below.

The following steps are provided as logically emanating from the meeting and the discussions.



Steps

Time period	Suggestions
Interim – the present moment	<p>Countries report according to the instrument or reporting method of their choice, subject to two caveats:</p> <p>Reporting is annotated according to a set of technical characteristics that measurement and data collection specific to early childhood should possess. Using this, the global community can judge how to use the reported information. These criteria (referred to as the “optimality” criteria) will be elaborated over the next few months, but a sketch for them was provided in the presentations discussed on 27 October. The sketched criteria are provided as an annex to this document for easy reference.</p> <p>In some cases, if there are serious technical issues with the underlying instruments or data, the data, or components thereof, might not be reported. An annotation to that effect will be made, to distinguish these cases from cases where there was no reporting. It would have to be accepted that in some cases there will be no reporting during the interim period.</p> <p>Given UNICEF’s mandated role in the SDG 4.2.1 reporting process it is expected that a common default reporting would be based on data from the ECDI. However, it is expected that as the ECDI improves, it too should fit the optimality criteria for all reporting. Furthermore, it is expected that some countries might not choose to participate in the ECDI or might choose to report using a different approach, even if they do participate.</p>
Interim – starting at present and with urgency, but not necessary in the next few months	<p>The following steps are listed in approximate order of urgency and sequentially.</p> <p>Analysis of existing databases including ECDI, but also others, plus an expert meeting, to define benchmarks (based on the empirical work plus expert opinion) for “developmentally on track.” This could be, but need not be, initially be used as part of the optimality criteria to annotate countries’ reporting (that is, do the country benchmarks resemble the GAML-recommended ones?) but, in the less-urgent interim period, to guide creation of country benchmarks and encourage a convergence to benchmarks that are as common as possible. A consultant or firm could be contracted to carry out the analysis and present the results to a specialized expert panel. The Secretariat or the Task Force would then comment, amend, and recommend the benchmarks. This step would “receive” and recognize analyses under way by various actors in the Task Force or outside it, but it is recommended that a person or team be designated to ensure that no information is “orphaned” and who can collate and “receive” these research inputs. The process would specifically help inform the ECDI and for that would need to be provided to the ECDI process by mid-year 2018.</p> <p>Clearer specification to be developed of the three domains (and sub-domains) covered, including the ECDI, but also providing domain definition useful for countries reporting using measures other than the ECDI. This would require a consultant to examine the main instruments that already exist and their background literature, collating and working out commonalities, in a manner similar to that done by UIS and IBE via a consultancy, for mathematics for 4.1.1.</p>



Time period	Suggestions
	<p>Finally, as the above work proceeds, or with those results in hand, engage a consultant or firm to develop equating procedures across various instruments, as well as guidelines for making country-based assessments “equitable” via the inclusion of suitable items. This procedure can include item-based equating or “social moderation” or “conceptual equating.” It would be similar to work described for 4.1.1 by ACER in the context of its work with UIS on a universal reporting scale, and by consultants appointed by UIS for that purpose who have provided some views on “social moderation.”</p> <p>It is recommended that at least steps 1 and 2 above be carried out in sufficient time to affect UNICEF’s process of ECDI improvement. But note that since it is assumed that many countries might not initially participate in the ECDI, or may participate but may not report based on the ECDI, the steps outlined here (and in the optimality criteria in the previous row) would feed a process that would:</p> <ul style="list-style-type: none"> Continue to improve annotation Guide processes whereby countries and assessment agencies that work with countries would improve their measurement tools, data collection, and reporting
<p>Longer term, within-interim and post-interim</p>	<p>Official and unofficial agencies and countries utilize the measurement and reporting guidelines to gradually improve the quality of measurement and its comparability. Measurement and reporting based on improved measurement. The optimality criteria for interim <i>reporting</i> also serve as a guide for ongoing <i>improved measurement</i> (i.e., measurement with known reliability and validity analyses, comparability analysis or comparable items, etc. as per the optimality criteria in the annex.)</p> <p>At fixed points in the future, such as the start of a new round of ECDI, reporting comes to include improved practices.</p> <p>At that point (and at various points in the process) the GAML 4.2.1 Task Force, as well as the Secretariat, take stock of how the various steps are proceeding.</p> <p>Sometime during 2020 a more intense process of scrutiny would take place to see how well the various measures are converging and how the quality of measurement is improving, and corrective action would take place.</p>



Annex 1: Participants in attendance on 27 October 2017

Name	Organization
Abbie Raikes*	University of Nebraska
Alvin Vista	Brookings Institution
Amanda Devercelli*	World Bank Group
Amber Gove	RTI
Amy Jo Dowd	Save the Children
Baela Raza Jamil*	ITA/PAL Network
Claudia Cappa	UNICEF
Dan Cloney*	ACER
Esther Care	Brookings Institution
Hiro Yoshikawa*	NYU Steinhardt
Kate Anderson	Brookings Institution
Luis Crouch	RTI
Magdalena Bendini	World Bank Group
Magdalena Janus	McMaster University
Manos Antoninis	Global Education Monitoring Report (GEMR)/UNESCO
Manuel Cardoso	UNICEF
Silvia Montoya	UNESCO Institute for Statistics

**denotes participation via videoconference*



Annex 2: Technical optimality criteria

The following list of technical optimality criteria, proposed during the 27 October 2017 meeting, with one or two which could be added, are to be used for:

Annotating in the interim period

Deciding on non-reporting in the interim period if a survey or instrument does not meet some crucial criteria. (There may be sub-criteria.)

Guiding technical developments during the interim period in order to create improved measurement, data collection, and reporting post-interim.

Include definition of “developmentally on-track”.

Criterion-referenced or, if not criterion-referenced, at least using clear, empirically well-based, and agreed norms

May start with a definition for each domain, but note that “on track” means that the child has to “on track” on each domain

Measure learning in a holistic way – that is, measure encompasses all three domains in the SDG.

Health, psychosocial well-being, learning

Population-based; that is, representative of the whole population in question, or, if not available for the whole population (e.g., not all age groups), then for representative sub-segments. It would be noted in particular if the measures are representative only of self-selected members or clearly non-representative parts of the population.

Conducted on a representative sample basis.

Useful to countries given national standards (or at least not be inconsistent with what countries are working towards for their own purposes).

Be globally comparable, or have items and definitions that allow one to determine its comparability with a determinable degree of accuracy.

Include background work that allows one to determine reliability and validity.

Administered at a variety of ages so that growth curves can be seen for the measures where it is relevant.

Have a well-defined reporting framework.

Follow the standards in the UIS Good Practices in Learning Assessment (GP-LA) and other “standard” codes of good measurement practice, incorporated by reference.



SDG Indicator 4.4.2

Measurement strategy and action plan. Proposal by GAML Task Force 4.4

According to its terms of reference, the Task Force will support the production of outputs that mirror those prepared by other task forces, notably 4.1, with respect to digital literacy skills, at least initially. It is proposed that the Task Force would engage toward two of these outputs **in 2017/18**.

1. There is currently no globally agreed definition of ICT and digital literacy skills, although there are:
 - examples of competence frameworks, at the national or cross-national level (e.g. European Commission's Digital Competence Framework for Citizens, DigComp 2.0)
 - examples of assessment frameworks of ICT and digital literacy skills at national (e.g. Chile SIMCETIC or France B2i) or cross-national level (e.g. IEA ICILS or OECD PIAAC problem solving domain); note that only PIAAC is targeted at the adult population, which is the focus of target 4.4

In that direction, the Task Force will:

- consult existing national and cross-national competence and assessment frameworks of ICT and digital literacy skills
- consider recommending to the GAML Secretariat the commissioning of a study that would synthesize the literature, identify any major gaps, and propose a definition as basis for a **global content framework of ICT and digital literacy skills**

The development of the content framework will need to address two major challenges:

- Existing cross-country school-level assessments of digital literacy skills show very large disparities between students in richer and poorer countries. This raises the question whether the gap reflects partly the degree of *access to computers at home and school*. However, the choice of themes in assessments appears also vulnerable to the criticism of cultural bias. To extend coverage to low and middle income countries, it is essential to develop context-appropriate items.
- A global tool to assess progress in digital literacy will need to address *rapid technological changes over time* and incorporate changes in patterns of ICT use. For example, the introduction of tablets and smartphones has resulted in new ICT applications and ways of working digitally. Assessment of skills needs to incorporate such developments without losing comparability over time.

2. At the same time, in order to further inform the development of a global content framework, the Task Force will use its networking and convening capacity to:

- identify more assessments of ICT and digital literacy skills;
- review the content of the UIS Catalogue of Learning Assessments tool to see how it could be amended or adapted for the case of ICT and digital literacy skills assessments;
- consider logistical issues that differentiate the administration of this tool, notably that:
 - not only ministries of education but also other government and non-government providers are responsible – and many tools are in fact proprietary;
 - there is less standardization in the skills assessed (compared with reading and maths);

- there is less standardization in the modes of assessment; and
- the age groups covered are very diverse (indeed there is more experience with assessments for those below age 15, which falls outside the scope of Target 4.4).
- consider recommending to the GAML Secretariat the commissioning of developing and rolling out at a pilot stage a **catalogue of assessments of ICT and digital literacy skills**

The following table suggests that once these two tasks have been completed:

- the mapping of assessments against the global content framework could begin **in 2018/19**
- the development of tools to evaluate the quality of assessments could begin **in 2019/20**

Finally, two other long-term issues specific to Target 4.4 that may affect the focus of the Task Force.

- Other potential skill domains for work (e.g. financial literacy, non-cognitive skills) could be considered and new members with expertise in these areas would be needed.
- Closer collaboration with the Task Force 4.6 on adult literacy and numeracy could be considered to be framed within an overall umbrella of skills for work related to youth and adults.

Proposed GAML Task Force 4.4 measurement strategy

	National	Cross-national	Global reporting Standard expected GAML outputs	TF activities		
				2017/18	2018/19	2019/20
Relevance	What is being assessed?					
Assessment frameworks	<p>Australia National Assessment Program – ICT Literacy Years 6 & 10 (NAP-ICT)</p> <p>Chile Habilidades TIC para el aprendizaje (SIMCETIC)</p> <p>France Cadre de référence des compétences numériques Brevet informatique et internet (B2i)</p>	<p>IEA International Computer and Information Literacy Study (ICILS)</p> <p>OECD Programme for the International Assessment of Adult Competencies: Problem solving in technology rich environments (PIAAC)</p> <p>ECDL Foundation International Computer Driving License (link)</p>	<p>Has a learning assessment taken place?</p> <p>► Catalogue of learning assessments</p>	X		
Competence frameworks	<p>France Platform to self-assess adult digital skills (Pix)</p>	<p>European Commission</p>	<p>What is the least common denominator?</p>	X		



	<p>Wales, United Kingdom</p> <p>Digital Competence Framework (DCE)</p>	<p>Digital Competence Framework for Citizens (DigComp 2.0)</p> <p>LSE / Twente / Oii</p> <p>Measuring digital skills (link)</p>	<p>► Global content framework</p> <p>How do different assessment frameworks map against the global content framework?</p> <p>► Content coding scheme</p> <p>► Evaluation of content alignment</p>		<p>X</p> <p>X</p>	
Implementation	Who is being assessed and how?					
<p>Technical standards</p> <ul style="list-style-type: none"> • sample, coverage etc. • modality, security etc. 			<p>Are the assessments technically robust?</p> <p>► Evaluation of data quality</p>			X
Interpretation	What do results mean?					
<ul style="list-style-type: none"> • reporting scale • performance levels • benchmarks 		<p>European Union Digital Economy and Society Index (DES)</p> <p>Dimension 2: Human capital / digital skills (note)</p>	<p>How does learning improve?</p> <p>► Learning progression</p> <p>A score that is attached to each learning level</p> <p>► Reporting scale</p> <p>What level should learners achieve on that scale?</p> <p>► Minimum proficiency level</p>			<p>X</p> <p>X</p> <p>X</p>

SDG Indicator 4.6.1

Proposal by GAML Task Force 4.6 (OECD proposal)

Measurement strategy and action plan

Introduction

Progress to the target will be measured by Indicator 4.6.1: Proportion 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.

This is further defined as: The proportion of youth (aged 15-24 years) and of adults (aged 15 years and above) have achieved or exceeded a given level of proficiency in (a) literacy and (b) numeracy. The minimum proficiency level will be measured relative to new common literacy and numeracy scales currently in development.

Key Issues

In developing a strategy to monitor progress towards Target 4.6, there are two main sets of issues. The first set are conceptual and the second, operational.

Conceptual issues

Definition of literacy and numeracy

The main issue at the conceptual level is that of agreement on the definitions and dimensions of the constructs of (adult) literacy and numeracy to be measured by indicator 4.6.1.

The UN's *Principles and Recommendations for Population and Housing Censuses Revision 3* define 'literacy' in the following way (UN, 2015, p236):

Literacy has historically been defined as the ability both to read and to write, distinguishing between "literate" and "illiterate" people. A literate person is one who can both read and write, with understanding, a short, simple statement on his or her everyday life. An illiterate person is one who cannot, with understanding, both read and write such a statement. Hence, a person capable of reading and writing only figures and his or her own name should be considered illiterate, as should a person who can read but not write as well as one who can read and write only a ritual phrase that has been memorized. However, a more modern understanding referring to literacy as a continuum of skills, levels, domains of application and functionality is now widely accepted.

No equivalent definition of 'numeracy' exists.

In terms of the conceptualisation of literacy and numeracy as a 'continuum', the situation in the field of adult assessments differs considerably from that of assessments of school age children. There are a number of international or cross-country comparative assessments of school age children reach with their own frameworks. However, there are only two cross-country programmes of comparative assessment of adult literacy and numeracy that are currently in place – the Programme for the International Assessment of Adult Competencies (PIAAC) and the World Bank's STEP assessment



which uses a version of the PIAAC literacy assessment. The PIAAC assessment frameworks draw on a theoretical tradition which has underpinned the conceptualisation of literacy and, subsequently, numeracy in the International Adult Literacy Survey (IALS), the Adult Literacy and Life Skills Survey (ALL) and UNESCO's Literacy Assessment and Monitoring Programme (LAMP). The conceptualisation of literacy and numeracy in PIAAC (and its predecessors) is closely related to that on which many international school-based assessments such as PISA, TIMMS, PIRLS and PASEC are based. The PIAAC literacy framework includes a reflection on the measurement of the skills that are preconditions for reading comprehension (described as reading components): print vocabulary knowledge, sentence processing and fluency.

In this context, there are strong reasons to consider the adoption or adaptation of the conceptual frameworks of the Programme for the International Assessment of Adult Competencies (PIAAC) as the basis for the development of a measurement framework for SDG target 4.6. The PIAAC frameworks represent well developed frameworks that have been validated in cross-national settings, including several middle and low income countries. In addition, given that they form the conceptual basis for the only existing international comparative assessments of adult literacy (PIAAC and STEP) and numeracy (PIAAC). If only for this reason, they must form the basis of any reflection on the conceptual framework for the measurement of target 4.6.

The PIAAC frameworks do not, however, cover 'writing'.

Alternatively, the development of new conceptual frameworks defining literacy and numeracy could be undertaken. However, any new frameworks would need to be compatible with the PIAAC frameworks. PIAAC is and will continue to be the main vehicle used to assess adult literacy and numeracy in the high income countries. To the extent that STEP continues, the PIAAC instruments will also continue to be the source of high quality information about literacy in several low and middle income countries.

Reporting thresholds

The extent of variation in the literacy and numeracy proficiency of the adult population in different countries represents a significant challenge for the establishment of benchmark levels that will make sense globally. The challenge is to set a benchmark that is far too high to be achieved by a large number of countries or alternatively one that is far too low to have any meaning for many countries.

Writing as part of literacy

Writing is included in the definition of 'literacy' cited above. However, there is no well-developed conceptual framework that could guide the assessment of writing in an international comparative setting and there are formidable practice challenges to assessing it. A position on the assessment of writing as part of literacy in the context of the SDGs needs to be developed.

Operational issues

Vehicles for assessment

The primary operational issue is that of determining the possible vehicle(s) for the collection of information on literacy and numeracy.



Four main options exist:

Existing international survey programmes such as PIAAC and STEP

New international comparative programmes

National literacy studies

Omnibus household surveys.

For younger cohorts, it may be possible to use the results from assessments of secondary school students (e.g. PISA or TIMSS) to estimate proficiency. This assumes that there is a reasonably close relationship between the proficiency of a cohort at the age of 14-15 and its proficiency at older ages². Given growing participation in studies such as PISA, this may be an option for middle and low income countries in which data about adults is lacking or of poor quality.

International assessment programmes

At the moment, PIAAC and STEP are the only two international comparative studies that collect information on literacy and numeracy. This situation is not likely to change in the foreseeable future. A new cycle of PIAAC is about to begin with data collection scheduled for 2021-22. It is possible that there may be additional rounds of PIAAC should additional countries wish to participate. In particular, a round of PIAAC for middle income countries may be possible with data collection in the period 2025-27 if there is sufficient interest from countries and donor organisations. At this point the STEP measurement study continues to be open to additional countries.

The cost and complexity of PIAAC and STEP makes it unlikely that more than a small number of low and middle income countries will participate in these programmes.

New international comparative assessment studies

The establishment of a new international comparative assessment of literacy and numeracy could be considered. This would represent an option for the longer-term. The feasibility, costs and benefits of such an option would need to be fully explored.

National studies

Dedicated national literacy and numeracy assessments exist. However, national assessments have been undertaken by a relatively small number of countries. They are also often based on country specific conceptual and assessment frameworks that make comparison of results with other surveys extremely difficult. In addition, the variation in the conditions under which studies are implemented (sampling, response rates, quality control) also has an impact on comparability. If comparability is a goal, countries planning national studies should be encouraged to join existing programmes or undertaking linking equating studies with existing international programmes.

Omnibus household surveys

² Obviously, the proportion of a cohort attending school would need to be taken into account.



Omnibus or multi-purpose household surveys such as DHS, MICS, living standards surveys and censuses usually collect some information regarding literacy. The validity of this information (based on very simple reading tests or respondent reports) is largely unknown. It may, however, be possible, to develop a short assessment module that could be administered as part of these studies to gain better information on the literacy proficiency of the target populations of these studies.

Summary

A realistic and pragmatic strategy for the collection of data on literacy and numeracy needs to recognise the fact that for many countries, the chance that direct assessments of literacy and numeracy proficiency that provide comparable information will be implemented between now and 2030 is low. More countries will participate in such assessments, but for reasons of cost, complexity and capacity coverage, particularly of low income countries is likely to be limited.

In addition, data collection in large scale literacy studies is usually carried out over relatively long cycles because of high costs and slow rates of change in the proficiency of the adult population. PIAAC collects data on a 10 year cycle for example. Low frequency of collection means that it is unlikely that there will be more than one observation for any single country over the period 2018-2030 and that this observation may not be very close to the end of the SDG reporting period (2030). For example, for the countries that participate in PIAAC, there will only be one observation available for any single country between now and 2030. In the case of countries participating in the upcoming 2nd cycle of PIAAC (the majority of high income countries), the information available will relate to 2021-22.

It may be possible to develop a short literacy assessment that was linked to the PIAAC scales could be administered in conjunction with other household surveys in low income countries such as the MICS and DHS programmes. The OECD is currently proceeding with the development of such a test. However, the challenges of this should not be ignored, primarily that of empirically establishing the link between each of the languages in which the instruments would be administered and the PIAAC scales. In addition take-up would be dependent upon the interest of the sponsor of the survey programme.

Assuming the above, it would also be important to review the information on literacy and numeracy collected in household survey programmes such as national censuses, DHS, MICS and other studies such as living standards surveys. This would cover issues such as the type of measures used and their validity, reliability and comparability. It is possible that greater harmonisation of information collected in such programmes as well as improvement in the design of the measures could considerably improve the quality of data from such studies.

At least for young cohorts (e.g. 15-24 year olds), it may be possible to use results from school-based assessments of lower secondary students (e.g. PISA, TIMSS) as the basis for estimations of proficiency in reading and mathematics in the event that good quality literacy and numeracy data from other sources is not available. However, the utility of such a strategy would be lower in those countries with relatively low rates of participation in school of 15 year olds.

A strategy to improve the quality and coverage of information on the proficiency of the adult population in literacy and numeracy available globally would need to have several components:

Encourage countries to participate in projects such as PIAAC and STEP



Review information on literacy and numeracy collected in census collections and household surveys with a view to improving data quality and comparability

Explore the use of results from assessments of secondary school students (e.g. TIMSS and PISA) for the estimation of proficiency among youth cohorts.

Work programme

A work programme for Taskforce 4.6 could be based around the following pieces tasks:

Develop a position paper on the definition and description of the constructs of literacy and numeracy; options and issues

Develop Reporting thresholds. Options and issues, analysis of PIAAC and STEP data.

Review of literacy and numeracy information collected in multi-purpose household surveys – possibilities for improving data quality and comparability

Prepare a paper identifying vehicles for the collection of information on literacy and numeracy – periodicity, costs, other constraints

Explore the possibility of using results from assessments of secondary-school students for estimation of literacy and numeracy levels for youth cohorts.

Possible Projects

Defining Literacy and Numeracy

The objective of this paper would be to discuss the options for developing a conceptual framework for the measurement of literacy and numeracy for the purposes of indicator 4.6.1. This would include identifying and evaluating existing frameworks used in national and international assessments. It would, in particular, explore, the advantages and disadvantages of adopting or adapting the PIAAC frameworks as the basis for defining and describing literacy and numeracy for the purposes of.

The paper could draw on document analysis, the inventory of literacy assessments and the results of an expert meeting on this topic proposed for early November 2017.

Developing reporting thresholds

The purpose of this paper would be to provide a theoretical and empirical background to the discussion of reporting thresholds. In particular, it would:

Discuss the possible approaches (e.g. expert judgement, statistical analysis) that could be used to establish and justify a benchmark level or levels on a scale such as that of PIAAC³ as well as the consequences of establishing a normative benchmark.

³ This would be used as an example due to the fact that empirical data are available. It would imply no judgement about what reporting scale should be used.



Use data from PIAAC and STEP to examine what benchmark levels could make sense across the range of countries in the world given the variation in proficiency in literacy and numeracy proficiency among adults

Assess the likely change in literacy levels among the adult population that is likely to be observed over the next 12 years (i.e. to 2030) under reasonable assumptions and the subsequent consequences for countries of the choice of different thresholds. .

Review the quality of literacy data collected in household surveys on literacy and numeracy

The purpose of this paper would be to review the information on literacy and numeracy collected in household surveys in terms of its validity, reliability and comparability. This would include examining the validity of the single sentence reading test used in the MICS and DHS programmes as well as the validity and comparability of respondent reports on their own or others reading and mathematics skills as well as their reading practices. The analysis would examine validity in relationship to the construct of literacy (particularly the threshold separating readers from pre-readers) as well as the empirical evidence regarding the relationship of simple and more developed direct assessments and respondent reports and direct assessments. The outcome would be recommendations regarding steps to take to improve the quality of literacy and numeracy data in household surveys.

Using school-based assessments to estimate the literacy and numeracy proficiency of youth cohorts

The growing coverage of assessments such as PISA provides information on the literacy and numeracy proficiency of 15 year olds on a common scale that could be used to estimate proficiency of youth cohorts (e.g. 15-24 year olds). As can be seen from the table below, in 2020, information will be available regarding the proficiency in reading and mathematics at the age 15 of cohorts aged 17, 20 and 23 in those countries that participated in PISA 2012, 2015 and 2018. This information could be used as the basis for estimating the proficiency (on the PISA scales) of the 15-24 year old cohort in those countries. The objective of this paper would be to examine the kinds of assumptions that would need to be made to do this (for example regarding the out of school population) and the likely robustness of estimates developed,. To the extent that data from national assessments are available for countries that do not participate in cross-national assessments, consideration could be given to using these data as well.

Table 1. PISA cohorts: age in 2020 and 2030

Data collection	Age in 2020	Age in 2030
2000	36	46
2003	33	43
2006	30	40
2009	27	36
2012	23	33
2015	20	30
2018	17	27
2021		24
2024		21
2027		18
2030		15

References

United Nations (UN) (2015), *Principles and Recommendations for Population and Housing Censuses Revision 3*, Department of Economic and Social Affairs Statistics Division, T/ESA/STAT/SER.M/67/Rev.3



Proposal by GAML Task Force 4.6 (UIL proposal)

Introduction

In developing a strategy to monitor progress towards Target 4.6., there are conceptual, methodological, and reporting issues.

Current measurement and reporting issues

The main issue at the conceptual level is the agreement on the definitions and dimensions of the constructs of adult literacy and numeracy to be measured by indicator 4.6.1.

UNESCO definition of literacy is a common one that many countries use to develop their policies and measure their progress. UNESCO defines literacy as following:

“Literacy is the ability to identify, understand, interpret, create, communicate and compute, using printed and written materials associated with varying contexts. Literacy involves a continuum of learning in enabling individuals to achieve their goals, to develop their knowledge and potential, and to participate fully in their community and wider society.” (UNESCO, 2017a; UNESCO, 2017b)

The above definition subsumes the definition of numeracy. As such there is no separate definition of numeracy.

In terms of the conceptualisation of literacy and numeracy as a ‘continuum’, the situation in the field of adult assessments differs considerably from that of assessments of school age children. Currently there are two cross-country programmes of comparative assessment of adult literacy and numeracy that are currently in place – the Programme for the International Assessment of Adult Competencies (PIAAC) and the World Bank’s STEP assessment which uses a version of the PIAAC literacy assessment. The following is the definition of literacy of OECD that the PIAAC is based on:

“Literacy is understanding, evaluating, using, and engaging with written text to participate in the society to achieve one’s goals and to develop one’s knowledge and potential.” (OECD, 2009; 2016).

As such the definitions of PIAAC and UNESCO do not necessarily contradict one another. There are also national direct literacy assessments that are often based on the above UNESCO definition.

The PIAAC assessment frameworks draw on a theoretical tradition which has underpinned the conceptualisation of literacy and, subsequently, numeracy in the International Adult Literacy Survey (IALS), the Adult Literacy and Life Skills Survey (ALL) and UNESCO’s Literacy Assessment and Monitoring Programme (LAMP). The conceptualisation of literacy and numeracy in PIAAC (and its predecessors) is closely related to that on which many international school-based assessments such as PISA, TIMSS, PIRLS and PASEC are based. The PIAAC literacy framework includes a reflection on the measurement of the skills that are preconditions for reading comprehension (described as reading components): print vocabulary knowledge, sentence processing and fluency.

Writing is included in the definition of ‘literacy’ cited above. However, there is no well-developed conceptual framework that could guide the assessment of writing in an international comparative setting and there are formidable practice challenges to assessing it. A position on the assessment of writing as part of literacy in the context of the SDGs needs to be developed and agreed.



Current literacy data is measured by omnibus or multi-purpose household surveys such as DHS, MICS, living standards surveys and censuses usually collect some information regarding literacy. Often a head of the household answers a single question: Can you or others in your household read and write a sentence? Some of such surveys also test adults' ability to write a sentence. In other surveys, literacy is assessed by a proxy measure on educational attainment.

This validity of this information (based on very simple reading tests or respondent reports) is largely unknown.

Dedicated national literacy and numeracy assessments exist. However, national assessments have been undertaken by a relatively small number of countries. They are also often based on country specific conceptual and assessment frameworks that make comparison of results with other surveys extremely difficult. In addition, the variation in the conditions under which studies are implemented (sampling, response rates, quality control) also has an impact on comparability. If comparability is a goal, countries planning national studies should be encouraged to join existing programmes or undertaking linking equating studies with existing international programmes.

At the moment, PIAAC and STEP are the only two international comparative studies that collect information on literacy and numeracy. This situation is not likely to change in the foreseeable future. A new cycle of PIAAC is about to begin with data collection scheduled for 2021-22. It is possible that there may be additional rounds of PIAAC should additional countries wish to participate. In particular, a round of PIAAC for middle income countries may be possible with data collection in the period 2025-27 if there is sufficient interest from countries and donor organisations. At this point the STEP measurement study continues to be open to additional countries, however, it is not based on representative sample but it focuses on urban population. The cost and complexity of PIAAC and STEP makes it unlikely that more than a small number of low and middle income countries will participate in these programmes.

Unlike the indicator 4.1.1., there is a low coverage on the indicator 4.6.1 that is globally comparable and methodologically rigorous. In this context, the extent of variation in the literacy and numeracy proficiency of the adult population in different countries represents a significant challenge for the establishment of benchmark levels that will make sense globally. The challenge is to set a benchmark that is far too high to be achieved by a large number of countries or alternatively one that is far too low to have any meaning for many countries.

Ensuing from this context, the following is the proposal for measuring 4.6.1 in the short and long run.

Proposal for measurement strategy 4.6.1

This proposal identifies steps for conceptual, methodological and reporting frameworks.

It is proposed to adopt UNESCO's (2017) definition of literacy, which is generally and globally accepted. This definition includes reading and numeracy. As earlier stated, there are a few examples of national assessment that includes writing, however, the international comparative assessments do not measure writing.

Moreover, it is also proposed to reach consensus on the methodological comparability especially when the literacy continuum is taken into consideration in the interim and long term reporting contexts.



In the immediate context, there are strong reasons to consider to start with the conceptual frameworks of the PIAAC as the basis for the development of a measurement framework for SDG target 4.6. The reason is that the PIAAC frameworks represent well developed frameworks that have been validated in cross-national settings, including in lower middle and middle income countries. There is however a need to extend the current PIAAC assessment framework to include the lower end of literacy continuum.

To bring the lower end of the continuum into the measurement framework for the indicator 4.6.1, it is necessary to examine whether the existing PIAAC can fit for non-OECD countries. The validity and relevance of the existing cross-national adult literacy and numeracy conceptual frameworks should be critically investigated to measure indicator 4.6.1 and the mapping of other national assessment tools would be necessary for a roadmap to extend the PIAAC assessment framework.

In the long term perspective, there is a need to develop a global common framework for reference that defines the constructs to be evaluated across all contexts. This can be achieved by mapping of assessment frameworks of existing surveys as well as linking these to national core competencies for adults and national qualification frameworks. The proposed framework for reference should be extensively consulted with regional and national stakeholders and experts before it is agreed to become as a global framework for reference.

This does not diminish the possibilities for measuring different domains, sub-domains and constructs that are deemed to be purposeful and relevant for different national contexts. This also includes the possibility of assessing writing skills in their national and other local languages.

Moreover, a global catalogue of learning assessment for adult literacy should be created and proposed to Member States. UIS could further extend their current Global Catalogue for Learning Assessment, including list of assessments for 4.6.1 to collect more information on the assessments in adult literacy and numeracy. Member States should be encouraged to submit their information in this regard.

There is a range of literacy assessment tools, including those that cover the lower end of literacy continuum. The proposal is that the countries using PIAAC assessment frameworks should be further encouraged to conduct their literacy surveys and report on indicator 4.6.1. A mapping of the existing assessment tools should be done to examine the feasibility to use these datasets to report on indicator 4.6.1. A set of criteria for data and measures should be determined and agreed upon. In this regard, the criteria for data should include the following components:

- It defines literacy as a continuum, not as a dichotomy i.e. literate and illiterate
- It assesses a full range of literacy proficiency levels
- It uses statistical models to confirm psychometric stability
- It uses statistical methods to support comparability

There are countries that do not collect literacy data at all. In this context, there is a need for proposing to develop and pilot a short literacy assessment that is linked to the proposed global assessment framework, methodologically rigorous and operationally pragmatic. The above criteria for data and measures could be strictly followed by this new assessment survey.

As a pragmatic step for interim reporting, methodologies of linking need to use of existing databases and collection of new data with existing or new instruments. For this, bottom-up approach is essential to use national data as well as utilize national benchmarks. Also, methodological solutions would be worked out with governments for relevant alignment between national and international reference frameworks. Practical steps are recommended for advancing the development of the measurement for the indicator 4.6.1:

- Define common domains and subdomains, continuum of skills
- Define number of skills levels
- Determine labels and write policy descriptors for the levels
- Develop full descriptions for the skills levels of the UIS reporting continuum. Choose a global or regional reference level of functional literacy and numeracy.

UIS reporting scale should be established and consultations should be carried out for Member States.

The following table summarizes the proposal for conceptual, methodological, and reporting frameworks to develop indicator 4.6.1 within from 2018 to 2020.

Conceptual framework

Activities	Expected outputs	Tentative timeline
Mapping the existing assessment frameworks for adult literacy and numeracy skills (national/cross-national levels)	Definition of adult literacy and numeracy to be measured Main constructed measured Harmonized assessment framework agreed	2017-2018

Methodological framework

Activities	Expected outputs	Tentative timeline
Mapping the existing assessment characteristics and use of assessment data	Catalogue of learning assessment for adult literacy and numeracy prepared	2017-2018
Mapping the existing content frameworks for Target 4.6	Basic content reference framework of adult literacy and numeracy prepared A paper and database to identify the commonality and difference of assessment domains for literacy and numeracy prepared	2018

Reporting framework

Activities	Expected outputs	Tentative timeline
Compiling good practices in adult literacy learning assessment	Good Practices for Learning Assessment manual prepared	2018
Alignment	Technical paper on the process to link between UIS reporting scale and the basic reference framework prepared	2018
Defining fixed level of proficiency for Target 4.6.1	Technical paper on the benchmarking process in defining fixed level of proficiency prepared	2019
Developing UIS reporting scale	Methodological paper on the reporting of fixed level of proficiency in Target 4.6.1 prepared	2020



References

OECD (2009) PIAAC Literacy: A Conceptual Framework, OECD Education Working Papers, No. 34, OECD Publishing, Paris

[http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=EDU/WKP\(2009\)14&doclanguage=en](http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=EDU/WKP(2009)14&doclanguage=en)

OECD (2016), Survey of Adult Skills (PIAAC)

https://www.oecd.org/skills/piaac/The_Survey%20of_Adult_Skills_Reader%27s_companion_Second_Edition.pdf

UNESCO (2017a). Reading the Past, Writing the Future: Fifty Years of Promoting Literacy.

Paris: UNESCO. <http://unesdoc.unesco.org/images/0024/002475/247563e.pdf>

UNESCO Institute for Statistics (2017b). UIS Literacy Assessment and Monitoring.

Programme (LAMP) <http://uis.unesco.org/sites/default/files/documents/literacy-assessment-and-monitoring-programme-lamp-information-brochure-en.pdf>



Proposal by GAML Secretariat

Overall Principles

Universal defining criteria not using a unique approach or tool unless agree globally

Desirably Long term given that many countries will choose either their own tools to report

Have a definition of constructs desirable in a framework\Be globally comparable, or have “hooks” that allow comparability

Guide the best possible, cost effective measurement not only reporting to SDGs

Produce easy to read guidance to countries in each target about options, costing and capacity development tools

About Process

A group of experts should advise in terms of content, with a diverse composition that in terms of regions, languages, and scripts.

Countries should be brought into the discussions in order to ensure that the proposed measurement approaches are sufficiently adaptive and responsive to their contexts. It’s unclear, however, whether GAML should be the context in which these consultations take place.

Connection to the TCG

Challenges (and opportunities)

Frameworks for skills/competencies/qualifications

Definition of functional Literacy and Numeracy

There is currently no globally agreed definition of functional literacy and numeracy although there are according to the Secretariat mapping (See Annex I)

there are not various frameworks both national and cross national (see Annex I)

Measurement and reporting

There are heterogeneities in measurement and reporting strategies that could be briefly summarized as (see Annex II for more details)

National and cross-national level surveys some of them measuring only dichotomous literacy

Tools have different scope and coverage

There are different way of reporting (direct or not)

There is different coverage in terms of domain

There are different modes of assessment (paper/computer based;

The main issue in reporting at this level seems to be comparability across systems and languages.

Focus on skills that are less affected by differences across languages (accuracy, comprehension).

No agree standards with respect to contents and data quality



Strategy: the way forward

The Secretariat aims to solve the three main areas

Conceptual framework: Who and what to assess?

Longer-term approach:

Agree on An assessment Framework that includes a definition functional literacy and numeracy

Contemplates different tools

Confirms on periodicity ten years cycle

Activities

Map existing national and cross-national survey assessment frameworks skills that include definition and measurement of functional literacy and numeracy

Skills Levels: map descriptors of Continuum of Skills and use this mapping for conceptual moderation.

Advance connection with basic education and digital literacy definition.

Interim approach:

Draw on the assessment frameworks and tools and report on that with the appropriate footnoting.

Methodological framework: How to assess?

Longer-term approach:

Define a framework for skills/qualifications

Data robustness and alignment: define the criteria that are required for global reporting and the Good Practices

Evaluate to develop a set of purpose-built tools that countries can draw on/adapt.

Activities

Map ools: Map the tools in terms of Coverage of Domain and in terms of skills/proficiency levels.

Define Global Framework

Asses conceptual alignment for each tool to the Global Framework

List the set tools that could serve to inform the target

Evaluate strategies to implement functional literacy and numeracy measurement taking into account the financial, human resources and logistic implication

Propos a standalone set of questions/module as global public good to be incorporated by countries to their HHS

Interim approach:

Define criteria about quality standards to be used as footnoting.

Reporting framework: How to report?

Longer-term vision:

Report comparable data with reference to good quality standards and referred to skills levels agreed by social moderation

Figure out how to make the reporting of results relevant for policy making including skills policies

Activities

The main issue in reporting at this level seems to be comparability across context, level of development, systems and languages. Are less affected by differences across languages (accuracy, comprehension).

Interim approach:

Report according to their own frameworks, tools, scope and coverage with the proper footnoting.



Annex 1

International organization	Defining literacy	Note
European Literacy Policy Network: European Declaration of the Right to Literacy	Literacy refers to the ability to read and write at a level whereby individuals can effectively understand and use written communication in all media (print or electronic), including digital literacy.	A multi-layered definition of literacy, from baseline literacy to functional and multiple literacy.
OECD: Survey of Adult Skills (PIAAC)	Literacy is understanding, evaluating, using and engaging with written text to participate in the society, to achieve one's goals and to develop one's knowledge and potential.	It measures adults' proficiency in key information-processing skills - literacy, numeracy and problem solving in technology-rich environments
World Bank: Skills Towards Employability and Productivity (STEP)	Cognitive skills are defined as the "ability to understand complex ideas, to adapt effectively to the environment, to learn from experience, to engage in various forms of reasoning, to overcome obstacles by taking thought"	<ol style="list-style-type: none"> 1. It assess the skills (cognitive, technical, and non-cognitive) of adults in urban areas. 2. The scales of the test are the same as those used in the PIAAC.

Annex 2

Four criteria are used to describe assessment/survey on literacy

Definition invokes continuum

Assessment covers full range of skills

Statistical methods confirm psychometric stability

Statistical methods support comparison

A glance of existing assessments grouped into four:

Direct assessments: International v.s. national

Indirect assessment: International v.s. national

	Survey/census	Defining literacy	Mode	Definition invokes continuum	Assessment covers full range of skill	Statistical methods confirm psychometric stability	Statistical methods support comparison	Example of countries
International household survey programmes	Multiple Indicator Cluster Survey (MICS)	Can read part of the sentence	Literacy test	N	N	N	N	Barbados and other 15 countries
		Can read and write	Self-declaration	N	N	N	N	Algeria and other 3 countries
	Demographic and Health Surveys (DHS)	Able to read parts of or a whole sentence*	Self-declaration	N	N	N	N	Benin and other 20 countries
National census or survey (selected examples)	Labour Force Survey	Can read and write a simple text in ...(given languages) or any languages	Literacy test	N	N	N	N	Colombia and other 25 countries
		Ability to read and write in any language	Self-declaration	N	N	N	N	
	Population Census	Can read, understand and write a short story regarding one's routine life	Literacy test, or self-declaration	N	N	N	N	Azerbaijan 2009
	Living Standard Measurement Survey	Ability to write a personal letter or read a newspaper.	Self-declaration	N	N	N	N	Albania 2012
	Continuous Household Survey	Can read and write	Self-declaration	N	N	N	N	Uruguay 2015
	Core Welfare Indicators Questionnaire	Can read and write in any language	Self-declaration	N	N	N	N	Liberia 2010
	Functional Literacy, Education and Mass Media Survey	Can read and write a simple message in any language or dialect	Self-declaration	N	N	N	N	Philippines 2013
	General House Survey	The literate are those who have no difficulty or some difficulty in reading and writing	Self-declaration	N	N	N	N	South Africa 2015
	National Socio-Economic Survey	Can read and write one of the following languages: Latin, Arabic, or other alphabet	Self-declaration	N	N	N	N	Indonesia 2016
The Reading of Population Survey	Can read and write	Self-declaration	N	N	N	N	Thailand 2015	



Country	Assessment	Definition invokes continuum	Assessment covers full range of skill	Statistical methods confirm psychometric stability	Statistical methods support comparison	Adaptive
Bangladesh	Education Watch 2016	Y	N	N	N	N
Botswana	National literacy survey	Y	N	N	N	N
Canada	Test of Workplace Essential Skills	Y	Y	Y	Y	Y
France	The Information and Everyday Life Survey	Y	N	Y	N	N
Germany	Level One Study (LEO)	Y	N	Y	Y	N
India	National Literacy Mission.	Y	Y	Y	N	N
Kenya	The Kenya National Adult Literacy Survey	Y	Y	Y	N	N
Lao PDR	Lao National Literacy Survey	Y	N	N	N	N
New Zealand	The Literacy and Numeracy for Adults Assessment Tool	Y	Y	Y	N	Y
Papua New Guinea	Education Experience Survey and Literacy Assessment	Y	N	N	N	N
USA	Adult Literacy Supplemental Assessment	Y	Y	Y	Y	N



SDG Target 4.7

Proposal by GAML Task Force 4.7

SDG Target 4.7:

By 2030, ensure that all learners acquire the knowledge and skills needed to promote sustainable development, including, among others, through education for sustainable development and sustainable lifestyles, human rights, gender equality, promotion of a culture of peace and non-violence, global citizenship and appreciation of cultural diversity and of culture's contribution to sustainable development.

Target 4.7 is one of the most important targets in terms of linkages with other SDGs and it is important to align measurement for target 4.7 related targets such as 12.8: "By 2030, ensure that people everywhere have the relevant information and awareness for sustainable development and lifestyles in harmony with nature" and 13.3: "Improve education, awareness-raising and human and institutional capacity on climate change mitigation, adaptation, impact reduction and early warning".

The current global indicator for this target is: "4.7.1 Extent to which (i) global citizenship education and (ii) education for sustainable development, including gender equality and human rights, are mainstreamed at all levels in: (a) national education policies; (b) curricula; (c) teacher education; and (d) student assessment." The existing reporting for the global indicator solely depends on the mechanism of the UNESCO 1974 Recommendation⁴ concerning Education for International Understanding, Co-operation and Peace and Education relating to Human Rights and Fundamental Freedoms. UNESCO is currently in the process of finalizing the measurement methodology for the global indicator 4.7.1 using the most recent round of 1974 recommendation data collection in 2016.

Target 4.7 includes five thematic indicators (including the Global Indicator). However, considering the fact that the current document will only focus on outcome indicators (learning assessment), the ongoing efforts for the measurement and monitoring of the 4.7.1 global indicator and 4.7.2⁵, 4.7.3⁶ thematic indicators are not the topic of this document, as they are based on the concept of "provision". The current document is specifically looking into the measurement strategy for the two thematic indicators which are broadly based around learning outcomes;

- 4.7.4: Percentage of students by age group (or education level) showing adequate understanding of issues relating to global citizenship and sustainability; and
- 4.7.5: Percentage of 15-year-old students showing proficiency in knowledge of environmental science and geoscience.

⁴ Adopted in 1974 by the 18th UNESCO General Conference. Member States have the obligation to report their progress towards 1974 recommendation every four years.

⁵ Thematic Indicator 28 (4.7.2): "Percentage of schools that provide life skills-based HIV and sexuality education."

⁶ Thematic Indicator 29 (4.7.3): "Extent to which the framework on the World Programme on Human Rights Education is implemented nationally".



Both thematic indicators cover learning outcomes achieved as a result of the educational inputs presented under the global indicator. This strategy elaborates on measurement solutions to address the challenges of monitoring indicators 4.7.4 and 4.7.5.

It should be noted that both indicators were originally inspired by existing data sources and international large-scale assessments, the IEA's International Civic and Citizenship Education Study (ICCS) in the case of 4.7.4 and the OECD's Programme for International Student Assessment (PISA) for 4.7.5, in particular the aspect of environmental science included in the 2006 cycle.

***Note1: While this strategy remains as a living document to be updated if and when necessary, at this point of time and in absence of any mapping exercises to identify available data sources with reasonable conceptual framework and coverage on the key topics of geoscience and environmental science, the current version of the strategy is limited to the thematic indicator 4.7.4.

Key challenges

Conceptual Issues

The main issue at the conceptual level is that of agreement on the definitions and dimensions of the constructs of Global Citizenship Education (GCED) and Education for Sustainable Development (ESD) to be considered in the measurement of indicator 4.7.4.

With the recognition that the strategy will remain a living document and flexible for future revisions and in consideration of all the existing limitations to clearly define GCED and ESD, for the purpose of this measurement strategy, GCED and ESD measurement components will be considered as below:

GCED is tentatively defined as any educational effort that aims to encourage the acquisition of skills, values, attitudes and behaviors empower learners to assume active roles to face and resolve global challenges and to become proactive contributors to a more peaceful, tolerant, inclusive and secure world. GCED nurtures the following three core dimensions of learning:

- The cognitive – to acquire knowledge, understanding and critical thinking about global issues and the interconnectedness/inter-dependency of countries and different populations.
- The socio-emotional – to have a sense of belonging to a common humanity, sharing values and responsibilities, sharing empathy, solidarity and respect for differences and diversity.
- The behavioral – to act responsibly at local, national and global levels for a more peaceful and sustainable world.

ESD is tentatively defined as any educational efforts that equip learners with the key learning components of:

- Knowledge (on ESD topics of sustainable lifestyles/sustainable ways of life, climate change, biodiversity, and the greening of the economy),
- Skills,
- Values ,
- Engagement,
- Attitudes and,
- Experiences

to address social, environmental and economic challenges of the 21st century through integrating critical issues such as climate change, biodiversity, disaster risk reduction (DRR), and sustainable consumption and production (SCP).

To facilitate the measurement efforts, an initial breakdown of measurement domains (knowledge, value, skills, engagement, and attitude) together with their respective content dimensions covering operationalized aspects under each of the ESD and GCED topics is provided in Annex1 (To be developed). This list is an inclusive effort to cover as many terms, understandings and interpretations on GCED and ESD as possible, to moderate the challenge towards country level understandings based on cultural, traditional or other contextual lenses affecting the perceptions.

Considering the wide range of established and some relatively new concepts covered under the thematic indicator 4.7.4 combined with the absence of specific processes to collect and analyze related data for the indicator, it is certainly one of the most challenging targets to measure and monitor progress on. Measuring real life skills and competencies such as empathy and creativity that are needed to promote sustainable development is vastly difficult and there are some debates on whether we should assess “non-cognitive”⁷ attributes/achievements in a standardized way at all. In addition, the interpretation and understanding of the concepts under target 4.7.4 are highly influenced by different cultural understanding across countries.

Other challenges relate to: the process of establishing mechanisms for mapping diverse content domain coverage, developing a relevant learning scale, streamlining varied data quality, establishing a coherent reporting metric, building country capacity to produce needed data and managing financial and human resource allocation.

Summarizing, the key questions to ask are:

- How can the “adequate understanding”, “proficiency” and performance levels be defined in the context of 4.7?
- Can heterogeneity in data collection and processing be effectively managed, quality controlled and evaluated for global monitoring and reporting?
- Is comparability for global reporting relevant in the two learning-related indicators?
- How can the best method of reporting be defined?
- What wide-range learning scale can be used for diverse levels of learning and for mapping skills?
- What kinds of guidelines are needed for data analysis and policymaking?
- For indicator 4.7.4, who should be assessed?

***NOTE2: This question not only refers to age group vs. education level, but also to considering at which levels the data should be collected? Grade 8 like ICCS? Or at multiple grade levels?

⁷ The usage of the term “non-cognitive” is discouraged by TF members due to the ambiguities associated with this term.



*** NOTE3: This question remains valid for further review. However, as an approach to match grade (or ISCED) level based curriculum and school organization and in view of the cyclic IEA ICCS reference data source for measurement, the target group will be school-based youth. In addition, through this approach, a wider data collection can include information on home, community, peer and school contexts, the teacher context and classroom-based opportunities to learn, which provides the opportunity to connect different aspects for an in-depth analysis to what students report). This does not preclude collection in older age cohorts, the general public or out-of-school populations. However, formal school based education provides for the most obvious leverage for policy.

- How often should the data be collected and how can we harmonize information from school-based and household based assessments? What are the costs of data collection? And what is the acceptable level of error and bias in reporting?
- How can the cultural differences and various understandings at country level on ESD and GCED be tackled in the process of measurement.

Reporting Thresholds

The relevant content of ESD and GCED focus on both covering and other list of skills, values and attitudes aspects of learning for which measuring “adequate understanding” may not be a relevant measurement criterion. Even for the cognitive component, the extent of variation in the definition of “showing adequate understanding” in different countries represents a significant challenge for the establishment of benchmark levels and cut scores that can be communicated well globally. For the non-cognitive component, the challenge is to set a benchmark that identifies the levels (on some continuum, from low to high) in which a high level might be associated with social justice and transformation orientations while a low level refers to basic understanding or engagement in a more limited way.

Operational Framework

The primary operational issue is the identification of the most relevant and already operational data collection tool with accepted definitions and reasonable coverage for the regular collection and analysis of information on GCED and ESD.

This identification is an ongoing process for which an initial step has been taken through a review of the available data collection/analysis efforts, mapping of definitions and comparison on data coverage [Measuring Global Citizenship Education a Collection of Practices and Tools⁸].

Noting that the toolkit collated by Brookings deliberately excluded large-scale assessments and instead focused on tools for schools, classrooms, communities and individuals, also outside of schools, UNESCO is in the process of launching additional studies to enrich the references for mapping exercises concerning ESD and GCED.

As a result and at this point of time, for the purpose of this strategy document and with the recognition of limitations in terms of definition and coverage, IEA ICCS has been selected as the relevant program and platform for the measurement and monitoring of thematic indicator 4.7.4. Exchanges with the

⁸ Download link: https://www.brookings.edu/wp-content/uploads/2017/04/global_20170411_measuring-global-citizenship.pdf



IEA confirmed that the cognitive test as well as the areas of attitudes and engagement are planned to be augmented with measures relating to global aspects of citizenship.

Mapping the threshold topic to the selected data source of IEA ICCS, it should be noted that ICCS has levels on the civic knowledge scale (established in 2009, extended for 2016). Level 2 (2009, called level B in 2016) could be seen as a possible minimum level, where students demonstrate a broader familiarity with concepts and notions. In each case, these levels are the second highest (of 3 in 2009, four in 2016). However, a standard setting beyond describing the hierarchical continuum/scale has not yet been completed for ICCS in light of the 4.7 target. A key issue will be to unpack whether 4.7 is meant to be taken as a transformational agenda and hence, levels of understanding and engagement towards e.g. social justice are meant, in which case a higher level could be the target. Pragmatically, and reviewing proportions of students at or above level 2 (2009) or B (2016), these could be a starting point.

Once released (assumed for end of 2019), the PISA 2018 cross-curricular domain of “global competency” might add additional insights but it is assumed that the country coverage will be limited and this domain will not be repeated in future cycles.

Work program for the measurement of indicators 4.7.4

Since global reporting is envisaged, school or subnational measurements are possible but these cannot be standardized or equated without additional efforts.

- Existing international survey programmes such as ICCS and PISA
- New International comparative programmes
- National studies
- Equating and projecting regional assessments on international metrics using bridging exercises (such as those discussed under the “Rosetta stone” strategy in the GAML cross-national assessment working groups).

Also, in order to further inform the development of work program for indicator 4.7.4, the Task Force will use its networking and convening capacity to:

- identify what have been collected in term of content coverage in national and cross-national assessment;
- clarify the definition of “issues related to sustainable development and global citizenship” and what it means “to show adequate understanding”;
- review the content of the IEA ICCS frameworks, instruments and reports, GCED assessment tools and ESD assessment tools to see how it could be improved or combined to collect relevant and target information;
- consider accessibility options in case self-reported survey modes are confounding reading ability and domain related aspects in locations where the former is low (e.g. through interviewers, computer/tablet-based collection or other means of voice-over modes such as CDs);
- issues on the alignment, linking of cross-national assessment for reporting;

- discuss with the larger GAML network how country coverage can be extended in terms of key obstacles such as funding and capacity;
- set-up criteria for monitoring and reporting due to difference in quality in the national and cross-national assessment data;
- develop interim reporting strategy while the reporting scale, definition of adequate understanding, tools and processes are still under development.
- consider recommending to the GAML Secretariat the commissioning of paper and developing of tools when deemed necessary to forward the agenda of indicator development



Annex A: Concept Paper on the Short Literacy and Numeracy Survey (SLNS)

Draft for discussion

Background

For decades, countries have been sampling and testing adults on skills at the international level. Large-scale, international adult assessment programmes such as the International Adult Literacy Survey (IALS), the Adult Literacy and Lifeskills (ALL) survey, Literacy Assessment and Monitoring Programme (LAMP), the Programme for International Assessment of Adult Competency (PIAAC), and the Skills Toward Employment and Productivity (STEP) study allow countries to compare the skills of their adult population. These assessment programmes are technically rigorous. However, except for LAMP, the instruments are developed for OECD countries and consist of tasks that do not reflect the reality of large segments of low-income countries' least literate population. If these countries administer the current international assessments, they would not identify whether the low-skilled population possess any literacy skills. LAMP seeks to address this needs through an assessment that consists of tasks with wider range of difficulty levels.

LAMP experiences provide information to the development of indicator 4.6.1 in Sustainable Development Goal 4. LAMP produces relatively detailed information on participants' reading and numeracy profiles from three low-middle income countries (Jordan, Mongolia and Palestine) and one upper-middle income country (Paraguay). Like most international assessments, LAMP requires substantial resources to administer and it is challenging- though not impossible - that a LAMP-like study could be implemented globally. More thoughts will need to be put in the assessment design and implementation strategy that take into account the reality of low- and middle-income countries. Eventually other more cost effective and pragmatic options could arise.

Options – framework and implementation

Based on discussion in indicator 4.1.1 and drawing the parallel, there are three options for indicator 4.6.1 to establish a global reporting metric:

- Option 1: Establish a global framework and develop a global instrument. However, this could be costly if it is to be started from scratch.
- Option 2: Tap on an existing established assessment and use the existing framework and instrument. There is currently no such assessment that has a wide framework and a diverse set of instruments that could adapt to all countries' needs.
- Option 3: Build on an existing framework, identify gaps in the framework and fill the gaps to develop a comprehensive framework. Develop a reporting metric based on the framework, develop instruments adapted to countries' needs that could cover different continuum on the global framework. As long as the instruments are statistically linked, using different instruments does not affect comparability.



With three implementation strategies:

- Option A: Conduct a full literacy assessment on a random sample.
- Option B: Add a stand-alone literacy module to existing household surveys. A short literacy practices background module and an adaptive literacy and numeracy assessment module are administered on a random sample.
- Option C: Conduct purposeful sampling to existing household surveys. An adaptive test and skills-relevant background information module are administered on a purposeful sample. The relationship of background information and skills is applied to the country's census to generate literacy estimates for the population.

Each option has its own merits and limitations and is beyond the discussion of this concept paper.

Based on the outcomes of the Task Force 4.6 expert meeting in November 2017 in Paris, option 3 builds on existing framework and is recommended. The PIAAC framework was recommended as a starting point, gaps in the framework to be identified and the PIAAC framework further developed into a common framework. PIAAC instruments could continue to be used for OECD countries or countries with similar education levels who are interested in full analyses of the literacy situation in their country. Other instruments like the Short Literacy Survey (which only covers literacy and not numeracy that is currently under development by OECD) could be used for countries interested in only literacy. Given that indicator 4.6.1 requires the reporting of adult population on both literacy and numeracy globally, an alternate instrument that covers both literacy and numeracy could be developed to cover the lower end of continuum. In order to report for indicator 4.6.1, all instruments developed from the conceptual framework should be statistically linked.

The experience that the UIS gained in modifying the original LAMP instruments built from IALS/ALL items and operational procedures through field testing in low-middle income countries:

- Acknowledging the data collection challenges and data capturing procedures in the household;
- Conducting data cleaning and analysis with inputs from the countries;
- Identifying background questions and cognitive items that work; and
- Completing the five country assessment surveys are all elements that will enable the UIS to devise a more viable plan and sustainable design for an alternate version of the adult assessment.

This concept paper represents an effort to provide:

- The definition of literacy in Short Literacy and Numeracy Survey (SLNS)
- The SLNS: framework, design, and implementation options
- Indicative cost for selected implementation options
- The governance structure of SLNS

The main focus is to implement this SLNS onto an existing household survey that the country is planning to conduct in the coming years, through MICS or DHS, either through random sample or purposeful sample.

The definition of literacy in the SLNS



Based on the outcomes of the Task Force 4.6 expert meeting, it was agreed to adopt the UNESCO definition of literacy (2004) as the conceptual framework since it does not contradict with other definitions and is currently in use by many countries.

UNESCO's definition of literacy looks beyond the ability to read and write a simple sentence. UNESCO maintains that functional literacy is the ability to identify, understand, interpret, create, communicate and compute, using printed and written materials associated with varying contexts. It involves a continuum of learning that empowers people to achieve their goals, to develop their knowledge and potentials, and to participate fully in their community and wider society. (UNESCO, 2005)

This definition is similar to the one used for PIAAC and STEP, which defines literacy as: "understanding, evaluating, using and engaging with written texts to participate in society, to achieve one's goals, and to develop one's knowledge and potential" (OECD, 2013).

All these definitions acknowledge that reading and numeracy literacy encompass a range of activities in real life contexts. The tasks could range from simple to complex based on the cognitive process requests to answer the questions.

The SLNS: objectives, design, implementation strategy

The UIS is initiating a new survey SLNS, provided that funding is available. It is designed to gather literacy and numeracy data for individuals aged fifteen years and older. The programme will include (a) a set of person characteristics and literacy-relevant background questions that characterize the contexts which promote or impede literacy development of individuals in a country, and (b) reading and numeracy assessment tasks that provide data on the early stages to mid-level literacy/numeracy skills.

The data will be gathered by presenting the instruments through a tablet to randomly selected or purposefully selected individuals in households via a one-on-one interview. This includes a short background questionnaire plus one to two literacy assessment modules that are described later in this paper.

The interviews will be conducted by trained interviewers. If the literacy/numeracy module is to be included in a household survey (as mentioned in option B and possibly option C, the operational plan might differ), alternate arrangement of the selected respondent should be made and a specially trained interviewer should follow-up. The total estimated length of the interview, including short relevant literacy background questions and cognitive tasks with fully adaptive test in a tablet, should keep below 60 minutes per person assessed.

a. The SLNS objectives

SLNS will address the need for literacy assessments that can provide useful information to countries with a large proportion of low literate population. The three main objectives of SLNS are to:

- Develop a methodology for assessing literacy (including numeracy) in low- and low-middle income countries, who might not have similar financial and technical capacity to conduct a full literacy assessment;



- Enable all countries - irrespective of development stage - to collect literacy and numeracy data along with relevant literacy practices on a household survey platform (that countries are familiar with) and that will inform national policy and provide information for SDG 4 monitoring; and
- Build statistical capacity in countries with different financial and technical capacity.

b. The SLNS design

SLNS will consist of two major parts: a short background questionnaire and an adaptive cognitive test. Both modules will be administered to a randomly selected individual in the household (for option B) or purposeful sampled individual in the household (for option C). The target population for the survey is people fifteen years and older.

SLNS will include a short questionnaire with questions related to individual literacy practices. The basic individual characteristics like sex, location, household wealth and household education level are collected in the main household survey, the questionnaire will concentrate on literacy practices that will inform policy and could be used to characterize basic relations between literacy practices and skills to generate literacy estimates.

The decision at the Task Force 4.6 expert meeting to build on PIAAC framework means that we could take advantage of the existing reporting metric. However, indicator 4.6.1 could have a different reporting metric so as not to contradict PIAAC and the later developed common framework. This reporting metric, building on the expanded framework and item pool, could extend to provide more precision in the lower skills continuum. The SLNS instrument will be statistically linked to the common framework (so is PIAAC instrument) to provide the international education community a monitoring mechanism on adult literacy.

Giving countries the implementation choices, PIAAC instruments could be used by countries with higher education levels, and SLNS instruments could be used by countries with lower education levels. Based on countries capacities, they will be presented with three implementation options (A to C).

In order to keep total assessment within 60 minutes, SLNS will assess respondents in a small number of reading and numeracy items but yet produce reliable estimate of their skills through an adaptive test platform. The modules will be administered with a computer tablet and their responses will be captured directly in the tablet. In conjunction with the cognitive items, contextual information from the background characteristics and literacy practices module will be used in the statistical model to produce skills estimates for subgroups.

There is information to gain by separating reading literacy into Prose (reading of continuous text) and Document (reading of non-continuous text). These two types of reading literacy require different skills; reading Document texts is usually learnt through school while reading Prose texts could be gained through daily reading.

SLNS instruments will extend below Level 1 (the lowest level). Below Level 1 is to allow for greater differentiation of performance at this level. The assessment will address reading and numeracy literacy at levels lower than those included in PIAAC. It will assess components or enabling skills such as number and word recognition.



Through adding learning-to-read component items and easier cognitive items - that capture the skills of low literate population,-the performance in the lower continuum of the framework will be better described. The performance descriptors could be expanded based on skills and types of texts covered in RAMAA, LEO and country-level adult literacy assessments, in addition to the component items developed in LAMP, STEP and PIAAC.

c. The implementation options

As mentioned in section 1 there are three options to implement adult literacy assessments:

- Option A: Conduct a full literacy assessment to a random sample

The country could choose to participate in the OECD's PIAAC assessment in 2021. This option suits the needs and capabilities of educationally and economically developed nations, generally those countries who have achieved universal secondary education and significant post-secondary participation.

- Option B: Add a stand-alone literacy module (background and cognitive) to an existing household survey

Administer a custom assessment using a fully adaptive test administered to a nationally representative sample of either 3,000 or 5,000 cases. This option is tailored to meet the needs for less-educationally advanced countries who possess the operational capacity and financial resources to field an adult literacy assessment using a fully adaptive test that yields reliable proficiency estimates. The choice will be between 3,000 and 5,000 completed assessments and will be dictated by the number of key population subgroups for which the country would need estimates..

- Option C: Conduct purposeful sampling to obtain the relationship of skills and characteristics and apply the relationship to census population.

Administer a custom assessment using a fully adaptive test administered to a purposeful sample. In this case, population level approximate proficiency estimates are generated indirectly by using the relationship between skill and background characteristics observed in the purposeful sample onto the target sample. The purposeful sample is selected from an existing representative household sample such as MICS, DHS or a Census of population. The characteristics are selected to capture key sources of variation in skill, normally education, urban/rural, age and gender. This option is tailored to the subset of countries who may lack the technical/operational capacity and/or the financial resources to administer an adult literacy assessment.

Costing and timeline

Given the three implementation options discussed earlier, below are the indicative costs for the last two options. The costs are separated into:

- Variable national costs
- Fixed national costs
- Variable international costs
- Fixed international costs

**a. Variable national costs**

Based on a sample size of 3000 with varied local cost, the total data collection costs could range from USD 158,200 to USD 334,600.

Based on a sample size of 5000 with varied local cost, the total data collection costs could range from USD 242,100 to USD 526,300.

Based on a sample size of 1500 with varied local cost, the total data collection costs could range from USD 79,100 to USD 167,300.

b. Fixed national costs

The second cost reflects the cost of the national team that will assume responsibility for all aspects of design, implementation and basic analysis. These also vary with local cost. The indicative cost could range from USD 40,250 to 72,450.

c. Variable international costs

The third cost associates with the implementation of international quality assurance processes. As fielding any assessment is technically and operationally demanding, containing risks to an acceptable level requires the imposition of a series of external quality assurance processes. The indicative cost is at approximately USD 100,000 per country.

d. Fixed international costs

The final cost associates with funding the international team that assumes responsibility for overseeing the design, implementation, and national and comparative analysis of the results. This team could be hosted at UIS or within the UNESCO, with strong technical capacity. The indicative cost is approximately USD 740,000. The cost could be borne by countries who participate in the assessment, therefore the more countries, the lower per country contribution.

e. Indicative costs

The total implementation cost therefore ranges from a low of \$234,150 per country to \$713,550 per country. This cost includes the fixed and variable national and international – quality assurance and coordinating (divided by 50 countries) - costs. Based on experience, there is a need for each country to develop a definitive estimate of cost based on the design documented in their National Planning Report as costs can vary in idiosyncratic and unpredictable ways. Similarly, collection costs can vary significantly by country, in response to how much of the sample is located in rural areas where travel costs are higher.

In addition to these implementation, operational and coordination costs, there are the initial development costs including finalizing a common framework, developing and/or collecting and validating cognitive items that could be included in the item pool for the adaptive test module, developing the computer adaptive platform, creating the database, and developing operational manuals. These costs could range from USD 500,000 to USD 750,000 depending on the extent of work involved. These costs could be negotiated through cost sharing among involved international institutes.



f. Timeline

The processes include finalizing the common framework, collecting and/or developing new items to enrich the existing item pool based on the common framework, developing assessment design, operational guidelines, data collection platform, field administration guides, and communicating to countries the different options for the production of indicator 4.6.1.

Most importantly, an initial communication should take place with major household based survey implementers, like UNICEF and USAID to communicate the possibility to include the skills module in their countries' data collection. The main points include expectation, operational development and cost sharing.

The conceptual framework should be finalized before any further development work could take place. The indicative milestones are presented in Annex B.

The governance structure of SLNS

The development of SLNS will be led by the UNESCO Institute for Statistics (UIS) in partnership with the UNESCO Institute for Lifelong Learning (UIL) and UNESCO headquarter (HQ) within the broader framework of UNESCO's efforts to gather data to promote Sustainable Development Goals (SDG) in Education.

a. Secretariat

The secretariat will be responsible for communicating, coordinating and overseeing all aspects of the SLNS implementation.

b. Steering Committee (SC)

The SC will be responsible for the strategic direction and the policy interests of the programme.

c. Technical Advisory Group (TAG)

The TAG will address design, content, technical and implementation issues.

d. Implementing organization

A group of contractors specialized in various aspects of the assessment process should be hired to conduct the cognitive and non-cognitive item development, write the operational manual, produce the scoring and coding guidelines, perform data analysis and implement statistical linking. Although the contractors will be responsible for the final forms of the assessment and questionnaire, the various committees will have inputs throughout the development process.

The Steering Committee will have to provide the final approval of all aspects of the programme. The coordination of various technical and operational activities will be housed at UIS through the GAML Secretariat.



Annex 2: Milestones to generate data for indicator 4.6.1

Key Steps	2017			2018			2019			
	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
Finalize measurement strategy to generate data for indicator 4.6.1	X									
Organize expert meeting to define conceptual framework		X								
Develop conceptual framework (build on existing PIAAC framework as agreed in expert meeting)			X							
Finalize conceptual framework through global consultation			X							
Develop performance descriptors and define fixed proficiency level				X						
Raise fund to form international coordination team and manage communication on operational issues			X	X						
Work with implementing agencies (like UNICEF and USAID) who are conducting household based survey in countries to develop pragmatic operational plan				X						
Develop procedures, tools, instruments and guidelines based on agreed assessment design				X	X					
Field test and data collection in selected pilot countries						X				
Revise and refine the instruments based on Field Test experiences.							X			
Data collection								X	X	X



Annex B: Proposal to the IAEG-SDGs

Upgrade the global indicator 4.1.1.i) Grades 2 and 3 to Tier II

Tier re-classification Material Submission

Additional information/documentation accompanying the summary: - Draft metadata
 Full methodology development narrative (including list of pilot countries, data and other results from pilot studies)
 Excel File with Detailed Data Availability
 GAML Overview

Goal 4	Ensure inclusive and equitable quality education and promote lifelong learning opportunities for all
Target 4.1	By 2030, ensure that all girls and boys complete free, equitable and quality primary and secondary education leading to relevant and effective learning outcomes.
Global indicator	4.1.1. Percentage of children/young people (i) in Grade 2/3 achieving minimum proficiency level in reading and mathematics
Custodian Agency	UNESCO Institute for Statistics
Current Tier	Tier III
Proposed Tier	Tier II
Submission	UIS based on the work in the Global Alliance to Monitor Learning

1. Background and rationale for indicator re-classification

There are existing, clear methodologies. For example, at least two regional comparative assessments, LLECE (Latin America) and PASEC (largely Francophone Africa but also including some Francophone non-African countries), already assess at Grades 3 and 2 respectively. In addition, a regional assessment, PILNA, in the Pacific, assesses at Grade 4, and this could be modified. These assessments include reading and mathematics domains. For the interpretation and communication of results, these assessments use the methodology based on the definition of proficiency levels that describe knowledge and skills that students must demonstrate. As students progress to the advanced level, they will show advanced knowledge. In general, these assessments have a minimum achievement level that indicates what percentage of children achieved at least the minimum competencies established for a grade.

Various global assessment programs exist, and they use a simpler standard methodology but do not claim to produce exactly comparable data. The instruments (oral and written evaluation) of these programs require local adaptations and their methodology allows to obtain percentage of correct



answers of tasks or contents of the national program of specific grade. These include EGRA and EGMA that examine basic skills in reading and numeracy respectively. Also the family of “Citizen-Led Assessments” that are descendants of India’s ASER and assess basic literacy and numeracy skills, and others. This year, in an effort to support the monitoring of indicator SDG 4.1.1, the agencies promoting these assessments began to publish non-comparable data from some countries as a percentage of students or children achieving basic tasks defined in the instruments. Finally, UNICEF’s MICS household surveys methodologists are developing and pilot-testing a brief module of foundational reading and numeracy skills related to grade 2 of primary for children of 7-14 years. Although the results and MIC methodology of the pilot tests were not published, their informative documents indicate that their instruments follow the EGRA methodology.

Countries also use their own methodologies. Some have extensive experience in assessments in grade 2 or 3 grade (Argentina, Australia, Chile, India, Peru or United Kingdom, e.g.). Other countries have recently begun evaluating this grade (Brazil, Colombia, Ecuador or South Africa, e.g.). In some countries, their regions began to evaluate grade 2 or 3 independently (United State, e.g.). In page 4 see the list of countries that reporting national learning outcomes in this grade (16 countries). As an example, the following table shows the national sources of historical data and framework.

Countries	Source of historical data
Australia	https://www.nap.edu.au/results-and-reports/national-reports
Argentina	http://portales.educacion.gov.ar/diniece/2014/05/22/evaluacion-de-la-calidad-educativadocumentos/
Peru	http://umc.minedu.gob.pe/evaluaciones-censales/
South Africa	https://www.education.gov.za/Curriculum/AnnualNationalAssessments(ANA)/tabid/569/Default.aspx
United Kingdom	https://www.compare-school-performance.service.gov.uk/historical-information

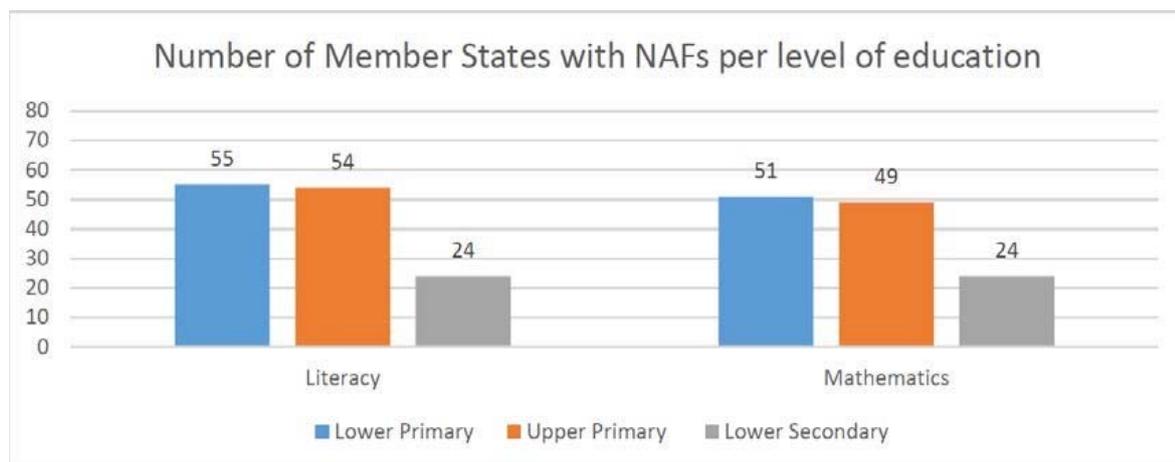


Figure 6: Number of Member States with NAFs per level of education

According to IBE mapping of national assessment frameworks per grade 55 countries have assessment frameworks for 2/3 grade. However, only a few countries publish learning outcomes or successfully complete the administration of their assessments.

2. Information on how and when the methodology has become an international standard and who is the governing body that approves it

The tasks needed for making the methodologies comparable and compatible are known. The science is clear but not everyone has the same knowledge of the science. Thus, to arrive at sufficient consensus for comparable measures would require committee work and some more piloting. The procedural or process-managed methods (technical and practitioner consultations, facilitation by neutral experts, etc.) for doing this are also clear. But, this work is still to be done. And, there are not, yet, 50 countries *already* reporting in a standard fashion. However, a large number of countries already measure, sometimes in a collective fashion, that would not be difficult to generalize.

3. Development and testing of the methodology (please also include information on how NSSs, and in particular NSOs, are involved in methodology development, data collection and data validation)/ Result of the pilot studies and list of countries consulted that are regionally representative

There is established methodology. However, some consensus and compilation still needs to take place, to enable comparability across countries, as noted above. The UNESCO Institute for Statistics is working with various agencies who have the capacity to do this work, and work is ongoing. Specifically, UIS is working with the Australian Council for Educational Research (ACER) on

There are agreed defined standards by type of Organization administering the assessment or surveys that in some cases are limited to region (LLECE, PASEC, PILNA) in others are global IEA's TIMSS, PIRLS, MICS and ASER having a global definition of minimum proficiency level .

(For example, see

<http://www.unesco.org/fileadmin/MULTIMEDIA/FIELD/Santiago/pdf/Reportetecnico-TERCE.pdf> for the methodology used by a regional assessment, and <https://globalreadingnetwork.net/resources/early-grade-reading-assessment-egra-toolkit-secondedition> and



https://iercpublicfiles.s3.amazonaws.com/public/resources/EGMA%20Toolkit_March2014.pdf for the methodology used by EGRA or EGMA)

UNICEF MICS, Learning Module of Foundational Skills,

<http://mics.unicef.org/files?job=W1siZiZlsljwMTcvMDYvMTUvMTYvMjcwMDAvNzIxL01JQ1NfTWV0aG9kb2xvZ2ljYWxfUGFwZXJfNS5wZGYiXV0&sha=39f5c31dbb91df26>

In addition, there is a network of institutions collaborating together to establish methodology under the UIS-led Global Alliance to Monitor Learning (GAML), with participation from scientific institutes with implementation experience (e.g., RTI International), NGOs (e.g., the Annual Status of Education Report in India, the cross-national PAL network that deals with Citizen-led Assessments), development agencies (e.g., UNICEF and USAID), and others, working on tools for assessment in these grades. (<http://www.norrag.org/news-hamburg-big-steps-forward-towards-reliable-metrics-harmonise-learning-assessment-data-globally-silvia-montoya-dirk-hastedt/>)

Not yet, as noted above. However, as also noted, the scientific basis for establishing standards exists, and a process for getting the science popularized and discussed among key actors, in well-structured ways, is also clear and under way).

4. Conclusion

On these grounds one could claim that this indicator is far above Tier III. There is already considerable standardized measurement, as discussed above, and there are established methodologies. In reality, it is somewhere between 2 and 3 as there is comparability among the countries that participated in a given assessment but no global scale **Thus, Tier 2 is the most logical placement.** In any case this is the same characteristic of point of measurement B and C.

5. Data Availability

MICS is not included in the Tables but per information from UNICEF 40 countries will be included in the wave of MIC6. Some of those countries currently do not have or are not set to participate in any assessment so MICS Learning Foundational skills modules will be the only one sources. Some of the countries are ; DPRK, Sierra Leone, Ghana, Gambia, Lesotho, Tunisia, Togo, DRC, The Pakistani province of Punjab. By the end of 2017, early 2018 three countries DPRK, Sierra Leone y Togo will be publishing their results.



Below a summary table but it is detailed in EXCEL.

	Grade 2 or 3							Grade 4			
	National Assessments (2011-2016)	ASER (2015/2016)	LLECE III (2013)	PASEC III (2014)	EGRA (2010/2014)	EGMA (2013/2014)	Citizen Led Assessment (2015-2016)	MICS (2016)	PILNA (2015)	TIMSS (2011 and 2015)	PIRLS (2011)
Africa (Sub-Saharan)	1	0	0	10	6	3	6	8	0	1	1
Asia (Central and Southern)	1	2	0	0	1	0	2	0	0	2	1
Asia (Eastern and South-Eastern)	2	0	0	0	1	0	0	1	0	5	3
Latin America and the Caribbean	8	0	15	0	0	0	1	2	0	1	2
Northern America and Europe	0	0	0	0	0	0	0	0	0	26	25
Oceania	1	0	0	0	1	0	0	0	12	2	2
Western Asia and Northern Africa	0	0	0	0	2	1	0	1	0	14	7

Table 1: List of countries with learning assessments (CNA and NA) in grades 2, 3 or 4. Administered between 2010 and 2016.

Countries or territories	Region	LLECE III (2013)	PASEC III (2014)	EGRA (2010-2014)	EGMA (2013-2014)	Citizen Led (2011) PILNA (2015) Assessment Grade 4 (2015-2016)	TIMSS and 15) Grade 4	National PIRLS (2011) Assessments Grade 4 (2010-2016)**
Argentina	Latin America and the Caribbean	x						x
Armenia	Central Asia						x	
Australia	East Asia and the Pacific					x		x x
Austria	North America and Western Europe						x	x
Azerbaijan	Central Asia					x		x
Bahrain	Arab States						x	
Benin	Sub-Saharan Africa		x					
Brazil	Latin America and the Caribbean	x						x
Bulgaria	Central and Eastern Europe					x		x
Burkina Faso	Sub-Saharan Africa		x					
Burundi	Sub-Saharan Africa		x					
Cameroun	Sub-Saharan Africa		x					
Canada	North America and Western Europe					x		x
Chad	Sub-Saharan Africa		x					
Chile	Latin America and the Caribbean	x				x		x
Chinese Taipei	East Asia and the Pacific						x	x
Colombia	Latin America and the Caribbean	x						x x
Congo	Sub-Saharan Africa		x					
Cook Islands	East Asia and the Pacific					x		
Costa Rica	Latin America and the Caribbean	x						
Côte d'Ivoire	Sub-Saharan Africa		x					
Croatia	Central and Eastern Europe						x	x
Cyprus	North America and Western Europe					x		
Czech Republic	Central and Eastern Europe						x	x
Denmark	North America and Western Europe					x		x
Dominican Republic	Latin America and the Caribbean	x						
Ecuador	Latin America and the Caribbean	x						x
Egypt	Arab States			x				
Ethiopia	Sub-Saharan Africa			x				
Finland	North America and Western Europe						x	x
France	North America and Western Europe					x		x
FSM (Micronesia)	East Asia and the Pacific					x		
Ghana	Sub-Saharan Africa			x	x			
Georgia	Central Asia						x	x
Germany	North America and Western Europe					x		x
Guatemala	Latin America and the Caribbean	x						x



Countries or territories	Region	LLECE III (2013)	PASEC III (2014)	EGRA (2010-2014)	EGMA (2013-2014)	Citizen Led (2011) PILNA (2015) Assessment Grade 4 (2015-2016)	TIMSS and 2015) Grade 4	National PIRLS (2011) Assessments Grade 4 (2010-2016)**
Honduras	Latin America and the Caribbean	x						x
Hong Kong, China	East Asia and the Pacific						x	x
Hungary	Central and Eastern Europe					x		x
India	South and West Asia					x		x
Indonesia	East Asia and the Pacific					x		x
Iran Islamic Republic of	South and West Asia						x	x
Ireland	North America and Western Europe					x		x
Israel	North America and Western Europe							x
Italy	North America and Western Europe					x		x
Japan	East Asia and the Pacific						x	
Jordan	Arab States			x	x	x		
Kazakhstan	Central Asia						x	
Kenya	Sub-Saharan Africa					x		
Kiribati	East Asia and the Pacific					x		
Korea Republic of	East Asia and the Pacific					x		
Kuwait	Arab States						x	
Lao	East Asia and the Pacific							x
Liberia	Sub-Saharan Africa			x				
Lithuania	Central and Eastern Europe					x		x
Malawi	Sub-Saharan Africa			x				
Malta	North America and Western Europe					x		x
Marshall Islands	East Asia and the Pacific						x	

Note: (*) EGRA 2010. The data are not comparable across countries. **= some countries report only the average score. **Source: Citizen Led Assessment, Early Grade Mathematics Assessment (EGMA), Early Grade Reading Assessment (EGRA), Latinoamericano de Evaluación de la Calidad de la Educación (LLECE), Programme d'analyse des systèmes éducatifs de la Confemen (PASEC), Pacific Island Literacy and Numeracy Assessment (PILNA), Progress in International Reading Literacy Study (PIRLS), Trends in International Mathematics and Science Study (TIMSS) and Ministries of Education or national agency responsible for national assessments.**