TCG4:
Progress on LO/Skills indicators

TCG4/22

GAML SDG4 Measurement Strategy – Overview Framework

Objective

This document aims to inform UNESCO Institute of Statistics (UIS) reporting strategy for Sustainable Development Goal (SDG) Indicators

4.1.1: *Proportion of children and young people: (a) in grades 2/3; (b) at the end of primary; and (c) at the end of lower secondary achieving at least a minimum proficiency level in (i) reading and (ii) mathematics, by sex*

4.2.1.P roportion of children under 5 years of age who are developmentally on track in health, learning and psychosocial well-being, by sex

4.4.2 Percentage of youth/adults who have achieved at least a minimum level of proficiency in digital literacy skills

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

4.7.4 Percentage of students by age group (or education level) showing adequate understanding of issues relating to global citizenship and sustainability

4.7.5. Percentage of 15-year old students showing proficiency in knowledge of environmental science and geoscience

Reporting on SDG4 global and thematic indicators constitute a substantial program of work by UIS and its partners, to work on comparability across assessments and to implement the various processes through which it may be applied, coverage could be expanded and capacity is built in countries. This work in progress is a challenges that currently prevent harmonised global reporting and learning assessment. Key considerations in interim reporting

The challenges of achieving consistency in global reporting go far beyond the definition of the indicators themselves. In many cases, there is no "one-stop shop" or single source of information for a specific indicator that is consistent across international contexts. Even when there is agreement on the metric to be used in reporting, a harmonising process may still be necessary to ensure that coverage of the data is consistent. Education expenditure is an example of an agreed metric for which international harmonisation of data coverage has dramatically changed how education systems view their results.[1]

The measurement of learning and skills poses particular challenges. Learning is typically assessed through complex processes that require definition of what learning is and/or what skills are under consideration, and how to measure it. Learning, functional literacy and numeracy and skills itself are complex construct, involving cognitive and non-cognitive processes accumulated over a sustained period of time. The measurement of may be seen as the last stage in this long process, meaning that indicators must be guided by deep understanding of what processes underpin the data.

---

[1] See *Who Pays for What in Education?* http://uis.unesco.org/sites/default/files/documents/who-pays-for-what-in-education-national-revealed-through-accounts-2016-en_0.pdf

Broadly, the development and implementation of any learning assessment/skills surveys follows four key phases, illustrated in Table 1. These may be applied at an international, regional or national level, depending on the scope of the assessment/survey program.

Table 1: Key phases in an assessment /skills measurement

Phase	What it addresses	Main components
Conceptual Framework	What and who to assess	• Assessment framework (cognitive, non-cognitive, and contextual) • Target population
Methodological Framework	How to assess	• Instrument design • Sampling frame • Operational design • Data generation • Data analysis (e.g. classical or Item Response Theory) • Contextual information to be collected/ disaggregation/ indicators?
Reporting Framework	How to report	• Defining scales • Benchmarking (type, level) • Defining progress (longitudinal equating)

Each of these phases needs specific sets of activities depending on the indicator to address not only international consistency, but also the overall quality of the program, and its utility to country's education/skills/social systems themselves.

Conceptual framework

• *Assessment framework*

Assessment programs differ in the conceptual frameworks that are used to develop their overall assessment framework or survey.

In the case of learning assessment programs may be either age or grade based; and may also vary in the point within a grade that is assessed. For example, some programs assess at the middle of an education level, some at the end of an education level, while others assess at both mid- and end points of an education level. Furthermore, the number of years of schooling (or duration of schooling) represented by a particular education level may vary across education systems. For example, some systems have six years of primary education, so testing at the mid- and end point of an education level may represent Grades 3 and 6 respectively. Others have four years of primary school, so the mid- and end point of the same education level may be Grades 2 and 4 respectively.

A second limitation is that when assessments are school-system based - usually referred as school-based learning assessments - the indicators cover only those in school. The proportion of in-school target populations varies from country to country due to differences in out-of-school children and populations of young people in the country. Assessing competencies of children and young people who are out-of-school would require household-based surveys. Assessing children and young people in households is under consideration, but may be very costly and difficult to administer.

The activities put in place are

1. Cognitive Framework: defining a global framework
   a. Conceptual model
      i. Takes an existing framework that could be adapted/extended or
      ii. Build based on existing surveys and assessments
   b. Build a reference list and coding scheme per area or subject
   c. Maps as many frameworks as possible
   d. Draft a global reference framework
   e. Consult experts and countries with regional representation
   f. Propose final versions
1. Contextual questionnaire
2. Target population

Outputs are

a. A global framework for reference
b. A standalone module or set of questions in some targets
c. A proposed background questionnaire addressing the reporting needs of SDG4 in general attending disaggregation and indicators in both global and thematic framework

Methodological framework

The main aspects highlighted in the table

- *Test design*

Assessments can be built in different formats, from multiple choice questions only to a combination of multiple choice and constructed response items. Over the years, technology has enabled more dynamic assessment design. With improved psychometric modelling, with which reasonable estimations can be done using a smaller number of items and target populations, different implementation platforms and operational procedures can be used. These possibilities have led to more complex test design, which must be carefully examined to ensure that it provides appropriate coverage of the learning domains under assessment.

The levels of learning progress represented in the test design is another important issue.

- *Sampling frame*

The nature of the sample is critical to the robustness of the assessment program as a measure of student learning progress, independent of any considerations of international consistency.

Likewise, sample survey data must be reported along with standard errors, so inference is allowed and proper confidence intervals can be inferred. The only exception is where an assessment program includes all students at the relevant age or grade.

- *Operational design and data generation procedures*

Robust, consistent operations and procedures are an essential part of any large-scale assessment or survey, to maximise data quality and minimise the impact of procedural variation on results. Examples of procedural standards may be found in all large-scale international assessments, in household or population surveys where the goal is to establish procedural consistency across international contexts. Many national assessments and surveys also set out clear procedural guidelines, to support consistency in their operationalisation.

- *Data analysis*

In terms of data analysis, some countries may use more sophisticated modelling and reporting methods to simpler ones. Depending on the model used, reporting scores will differ in their scales and metrics.

Data analysis in some indicators typically includes disaggregation by student demographic characteristics such as age or age-group of students, sex, location, socio-economic status, migrant status and ethnicity. This supports confirmation of the representativeness of the sample for cohorts for whom learning outcomes may differ, and also assists education systems to understand which student cohorts are best served by their schools. Disability status is not currently available in most national and cross-national learning assessments.

The activities put in place are
1. Definition on a data alignment strategy
2. Mapping of Learning Assessment and Skills Surveys

Outputs are
1. Data alignment Concept note and guidelines
2. Catalogue of Learning Assessments
3. Data Alignment Reporting Tool

Reporting framework

- *Defining scales*

Reports on learning indicators are usually using different scales. Analysis of results therefore remains contained to their particular test, linked to one methodology and one scale. While there are some convergence in methodologies through time it is still difficult to situate an individual student's learning progress on an indicative pathway

- *Benchmarking*

Currently, there are no common standards validated by the international community as a global benchmark in no indicator of the group. While data from many national learning assessments is

available now, every country sets its own standards so the performance levels defined in these assessments may not always be consistent.

Over time, benchmarks will be identified and linking will be established to facilitate comparison.

- *Linking to a common scale*

The process of linking for comparability needs to be established.

The activities put in place are

1. Linking concept note and protocols for social moderation and psychometric linking
2. Finalization on benchmarking

Outputs are

1. UIS reporting Scale with benchmarks

Format for reporting: long terms and interim reporting

In general, reporting format aims to communicate two pieces of information:

1. the percentage of students meeting minimum proficiency standards for the relevant domain and measurement point; and
2. the conditions under which the percentage can be considered comparable to the percentage reported from another country, including any caveats that may affect comparability.

In the first round of reporting, the number of caveats on comparability (limitations) is likely to outweigh the number of conditions under which cross-country comparability can be considered (possibilities). This does not detract from the value of interim reporting, recalling that the primary goal of SDG reporting is not to compare results across countries, but to inform system improvement within individual countries or country groups. Over time, possibilities for international comparability may increase, but this primary purpose will remain until the criteria for

The activities put in place are

1. Definition of a protocol for interim reporting that incorporates the long term views
2. Validation process Concept Note

Outputs are

1. Protocol for interim reporting
2. Guide for countries on how to report for LO/Skills indicators and the validation process

An example: activities and outputs for reporting indicator   4.1.1.

The whole process for indicator 4.1.1. is summarized in Table and Figure below that includes the UIS capacity development activities.

Steps	Actions	Tools/Process	Output	Responsible	Progress to Date	Expected date of Completion
Objectives						
Global	Reporting	UIS reporting protocol	Databases	UIS	Doc Lit	2017
National	Improvement Plan if needed	Catalogue of Learning Assessment (modules 1 and 2)	Mapping characteristics	UIS/Partners	Pilot Modules 1 and 2 undergoing	2018
		Data alignment recording tool (DART)	Mapping results	UIS/technical Partners	CN finished Tool in development	2018
		System-Wide Analysis of Assessment Practices (SWAAP)	Concept Note Tool	ACER	Concept Note for discussion	2018
UIS Reporting Scale						
1. Construction of UIS Reporting Scale	1. Contents standards	Mapping of CNA		UIS/IBE mapping of CNAFs IEA/Regional Work would round up process	Mapping for Math NAFs and CNA finalized Reading Undergoing	2018
		UIS proposed definition on Concepts	UIS Draft GFFR	UIS/IBE mapping of NAFs	Math almost finalized Reading in Progress	2018
		Mapping pf CNA PLs and PLDs	UIS Draft PL PLD	UIS/IBE mapping of NAFs	CN proposal Guidelines and protocols	2018
	2. Number and name of Prof Levels	Mapping and experts judgment	UIS Draft levels	UIS led with experts/countries	CN proposal Guidelines and protocols	2018
	3. Policy Proficiency level descriptors	UIS Proposal	UIS Draft PLDs	UIS led with experts/countries	CN proposal Guidelines and protocols	2018
	4. Full definition of the Performance Standards	UIS proposal	UIS Complete description of PLDs	UIS led with experts/countries	CN proposal Guidelines and protocols	2018
Data alignment for reporting						
	1. Mapping alignment of	Protocol concept note	Degree of Alignment	UIS/Experts/technical partners	CN proposal	2018

2. Socially-Moderated Alignment[2]	Policy Level Descriptors				Guidelines and protocols	
	2. Set socially moderated Performance levels	Protocol concept note	Protocol for mapping Experts/Countries	UIS/Experts/technical partners	CN proposal Guidelines and protocols	2018
3. Psychometric Alignment[3]	1. Alignment	Protocol / Meeting	Experts	UIS/Partners	CN proposal Guidelines and protocols	2018
	2. Concordance	Protocol / Field Work	Degree of Alignment	UIS/Partners	Guidelines	2020
	3. Test and/or item based linking 4. Pair Wise Comparison?	Protocol / Field Work	Degree of Alignment	UIS/Partners	Guidelines and protocols	2020
4. Procedural Alignment	1. Define minimum set of processes that grant quality	Sampling	Protocol	UIS/Partners	Guidelines and protocols	2018
		Rate of response	Protocol	UIS/Partners		
		Translation	Protocol	UIS/Partners		

Figure 1

Dark green border = process led by individual education systems

---
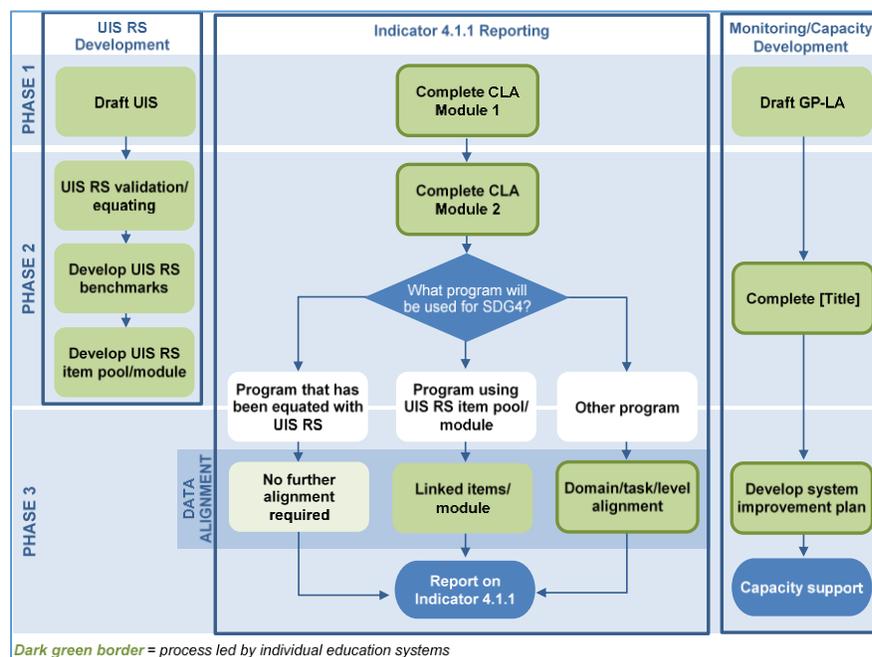
[2] Key for countries either not participating in a global or regional assessment, or who may be participating in those, but do not wish to report based on them, and wish to report based on a national assessment. It is necessary because psychometric alignment may not be possible as there may not be enough shared items between the UIS-RS and the national assessments, especially at first.
[3] Only possible when there are enough shared items.

GAML – Indicator Development

Indicator	Theme	Issues
4.1.1.b/c	Reading and Mathematics at the End of Primary and end of lower secondary	Work plan addressing - Global comparability with Global Framework for Reference - Definition of proficiency levels and description of levels - Benchmarks, one on each measurement point - Long and interim reporting strategy - Possibility of stand-alone module
4.1.1. a	Reading and Mathematics at the early grades	Work plan addressing - Global comparability with Global Framework for Reference - Specificities of the indicator - Definition of proficiency levels and description of levels - Benchmark - Long and interim reporting strategy - Possibility of stand-alone module
4.2.1	ECD developmentally-on-track	Support to UNICEF - Definition on three domains (learning, health, psychosocial) - Definition of developmentally-on-track on the composite of three domains - Development of reporting scale - Benchmarks - Long and interim reporting strategy - Possibility of stand-alone module
4.4.2	Funding formulas	Work plan addressing - Global comparability with Global Framework for Reference - Definition of proficiency levels - Benchmarks - Long and Interim Reporting strategy - Possibility of stand-alone module
4.6.1	Country participation and coverage issues	Work plan addressing - Comparability with expanded Framework for Reference - Definition of proficiency levels - Benchmarks - Implementation options to encourage country participation - Long and interim reporting strategy - Possibility of stand-alone module
4.7.4	GCED	Work plan addressing more fundamental issues - Relevant of global comparability in GCED - Possibility of common definitions and proficiency levels - Target group – age/grade and in/out of school youth - Defining benchmarks - Reporting strategy, use existing ICCS or 'link' across existing national assessments - Commission studies to examine the fundamental issues
4.7.5	ESD	Work plan addressing more fundamental issues - Relevant of global comparability in ESD - Possibility of common definitions and proficiency levels - Target group – age/grade and in/out of school youth - Defining benchmarks - No existing cross-national assessment except PISA but the assessment might not repeat - Reporting strategy - Commission studies to examine the fundamental issues

GAML Work Plan

		2017				2018			
		Jan - Mar	April - June	July - Sep	Oct - Dec	Jan - Mar	April - June	July - Sep	Oct - Dec
Task Force 4.1									
Indicator	*Proportion of children and young people: (a) in Grade 2 or 3; (b) at the end of primary education; (c) at the end of lower secondary achieving at least a minimum proficiency level in (i) reading and (ii) mathematics, by sex*								
Activities	**Expected output 1:** Conceptual framework - Global Framework for Reference							Math	Reading
	Expected outcome: Reference framework of contents and competencies to guide teaching, learning and assessment								
	Reference list and coding scheme to help systematically map national mathematics assessment frameworks								
	Database of the mapping of national mathematics assessment frameworks					Math	Reading		
	Reference list and coding scheme to help systematically map national reading assessment frameworks								
	Database of the mapping of national reading assessment frameworks								
	Database of the cross-national assessments' proficiency descriptors								
	Database of the cross-national assessment mathematics and reading frameworks								
	Mapping of the mathematics and reading reference list and the respective reporting scales								
	Global consultation of the mathematics and reading reference frameworks							Math	Reading
	Expected output 2: Methodological framework - Reporting Scale and Proficiency Benchmarking								
	Expected outcome 2: A developed reporting scale and consensus reached proficiency benchmark for each measurement point and the performance descriptors								
	Learning progression explorer that describe mathematics and reading reporting scale								
	Performance descriptors writing workshop								
	Proposed proficiency descriptors								
	UIS-PM performance levels' descriptors and labels								
	Expected output 3: Reporting framework - Quality assurance and data alignment to reporting scale								
	Expected outcome 3: Clear protocol for reporting against indicator 4.1.1								
	Data Alignment Reporting Tool (DART)								
	Quality assurance content and data alignment workshops								
	Standard-setting (social moderation) workshops to define benchmarks for each measurement point								
	Concordance table to establish psychometric linking between selected regional and international assessments								
	Interim Reporting Strategy								

	2017				2018			
	Jan - Mar	April - June	July - Sep	Oct - Dec	Jan - Mar	April - June	July - Sep	Oct - Dec
Catalogue of learning assessments 2.0 (CLA 2.0) to collect national assessment meta-information and performance level data							▓	
Background questionnaire module					▓	▓	▓	
Expected output 4: Research, comparative analysis, policy and concept papers addressing specific methodological development								
Expected outcome 4: Sound methodological decisions based on research								
The commonality and difference of regional and international assessments						▓		
The Value of Learning Data: A case for Investing in cross-national Assessment						▓		
A Review of the use of cross-national assessments data in educational practices						▓		
Mind the Gap: Proposal for a Standardised Measure for SDG 4 – Education 2030 Agenda					▓	▓		
The Methodology for a Global Composite Indicator for Education: Counting the Number of Children Not Learning					▓	▓		
More Than One-Half of Children and Adolescents Are Not Learning Worldwide					▓	▓		
Analysis of results of 2017 Survey of Cross-national Assessments						▓		
Expected output 5: Special measurement strategy for early grades including short-term strategy								
Expected outcome 5: Specificities of early grades and cultural heterogeneity properly addressed								
Convene a group of experts who can bring the latest research, evidence, and data to bear on the drafting of a longer-term measurement strategy for Indicator 4.1.1a							▓	▓
Countries brought into the discussions on Indicator 4.1.1a in order to ensure that the proposed measurement approaches are sufficiently adaptive and responsive to their contexts.							▓	▓
Stand-alone module as a global public good								

Task Force 4.2

Indicator	*Proportion of children under 5 years of age who are developmentally on track in health, learning and psychosocial well-being, by sex*

Activities		2017 Jan-Mar	2017 April-June	2017 July-Sep	2017 Oct-Dec	2018 Jan-Mar	2018 April-June	2018 July-Sep	2018 Oct-Dec
	Expected output 1: Conceptual framework - Mapping of early childhood initiatives								
	Expected outcome 1: Knowledge of methodologies and approaches								
	Convene researchers/holders of large-scale data sets to map out methodology and approach								
	Expected output 2: Methodology framework - reporting scale and developmentally-on-track benchmark		░						
	Expected outcome 2: Develop report scale for indicator 4.2.1		░						
	Methodology work to develop reporting scale for ECD		░						

		2017				2018			
		Jan - Mar	April - June	July - Sep	Oct - Dec	Jan - Mar	April - June	July - Sep	Oct - Dec
	Convene experts including member state countries to define what it means by developmentally-on-track								
	Expected output 3: Reporting framework - preliminary work in collecting information from countries								
	Expected outcome 3: Meta information to inform the development of reporting framework								
	Catalogue of learning assessments 2.0 (CLA 2.0) to collect meta-information and performance level data from countries								
	Expected output 4: Final interim reporting protocol								
	Expected outcome 4: Use of existing initiatives to report against indicator 4.2.1								
	Identify psychometricians from each team to work in partnership with the expert group to complete the analyses and make recommendations for using the findings to inform the interim reporting strategy								▨
	Stand-alone Module as a global public good								

Task Force 4.4

Indicator	**4.4.2:** Proportion of youth and adults with information and communications technology (ICT) skills, by type of skill
	4.4.2: Percentage of youth/adults who have achieved at least a minimum level of proficiency in digital literacy skills

Activities									
	Expected output 1: Conceptual framework - Global Competency Framework of Reference on digital literacy skills								
	Expected outcome 1: A definition of digital literacy skills								
	Paper with the review existing cross-national and national competence, curriculum and assessment frameworks of ICT and digital literacy skills						▨	▨	
	Mapping of competency frameworks for digital literacy						▨	▨	
	Global consultation of the competency framework							▨	
	Expected output 2: Catalogue of assessments of ICT and digital literacy skills								
	Expected outcome 2: Identify assessments of ICT and digital literacy skills								
	Commission the development and rolling out at a pilot stage a catalogue of assessments of ICT and digital literacy skills							▨	▨
	Stand-alone Module as a global public good								

Task Force 4.6

Indicator	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.

Activities	**Expected output 1:** Conceptual framework - Define functional literacy and numeracy and produce options for measurement the Sustainable Development Goal 4.6
	Expected outcome 1: A global framework of functional literacy and numeracy skills and competencies to measure

		2017				2018			
		Jan - Mar	April - June	July - Sep	Oct - Dec	Jan - Mar	April - June	July - Sep	Oct - Dec
	Concept paper on the definition of functional literacy and numeracy						▓		
	Paper identifies measurement options						▓		
	Expected output 2: Determine common content								
	Expected outcome 2: Reference Framework to guide policy making and evidence based decisions								
	Mapping of national assessment frameworks and performance level descriptors						▓	▓	
	Mapping of cross national assessment frameworks and performance level descriptors						▓	▓	
	Mapping performance level descriptors						▓	▓	
	Expected output 3: Final interim reporting protocol								
	Expected outcome 3: Clear protocol for reporting against indicator 4.6.1								
	Define a common framework								▓
	UIS-Skills Levels Descriptor for each learning/skills indicator								▓
	Write full descriptions of UIS-skill levels								▓
	Linking UIS-Performance Metrics with cross-national and national assessments								
	Evaluate alignment of proficiency level descriptors workshops								
	Standard setting (social moderation) workshops								
	Psychometric Linking								
	Stand-alone module as a global public good								
	Expected output 4: Data alignment to defined standards								
	Expected outcome 4: Better Quality Data								
	Catalogue of learning assessments 2.0 (CLA 2.0) to collect meta-information and performance level data from countries							▓	▓
	Code of Good Practices - 'How-to' guide								▓
Task Force 4.7									
Indicator	**4.7.4:** Percentage of students by age group (or education level) showing adequate understanding of issues relating to global citizenship and sustainability.								
	4.7.5: Percentage of 15-year-old students showing proficiency in knowledge of environmental science and geoscience.								
Activities	**Expected output 1:** Final interim reporting protocol								
	Expected outcome 1: Clear protocol for reporting against indicators 4.7.4 and 4.7.5								
	Commission paper on the mapping of existing GCE		░			▓			
	Concept paper on the results of ICCS 2016 & 1974 6th consultation					▓			
	Conduct a study to map ESD measurement tool						▓		
	Create a platform among identified key sources of regional data							▓	▓

			2017				2018			
			Jan - Mar	April - June	July - Sep	Oct - Dec	Jan - Mar	April - June	July - Sep	Oct - Dec
Secretariat / Guidelines to countries										
Activities	**Expected output 1:** Capacity development for countries by providing guidance on key decisions									
	Expected outcome 1: Better quality data for country and monitoring									
	Principles of Good Practice in Learning Assessment						�			
	Quick Start Guide for Implementing a National Learning Assessment							grey		
	Expected output 2: Guidelines to countries			blue						
	Quick Guide	**Expected outcome 2:** Ensuring that the maximum number of countries report against SDG 4 indicators	blue							
		For Monitoring Learning Globally. How does it work? What should my country do?					grey	grey		
		To Assess or not? How and how much does it cost? Strategic decisions in Learning Assessments						grey	grey	grey
		Implementing a National Learning Assessment						grey	grey	grey
		What and how to report? Countries options for reporting						grey	grey	grey
		How do learning assessments integrate with other data sources that inform education						grey	grey	grey