



United Nations
Educational, Scientific and
Cultural Organization



UNESCO
INSTITUTE
FOR
STATISTICS



TECHNICAL
COOPERATION
GROUP



TCG4: SDG 4 Reporting: Linking to the UIS Reporting Scale through social moderation

TCG4/32

16-18 January 2018

Dusit Thani Dubai

133, Sheikh Zayed Road, Trade Centre,
Dubai, United Arab Emirates





The UIS' goal, as custodial agency for reporting against the SDGs, is to develop reporting scales that will support national governments to effectively measure and monitor student learning outcomes in mathematics and reading against SDG Indicator 4.1.1 over time, and to utilize the data for making informed policy decisions. It is a further goal to support the use of existing national assessments (NAs) and cross-national assessments (CNAs) to facilitate measurement and reporting of learning outcomes, rather than requiring a single assessment be used by all countries for SDG reporting purposes.

The Global Alliance to Monitor Learning (GAML) and its technical partner, Australian Council for Educational Research's Centre for Global Education Monitoring (ACER-GEM), are exploring the development of reporting scales – the UIS Reporting Scales, or UIS-RS – that would achieve these goals. Draft learning progressions have been developed and are currently undergoing broad review, and a proposal for validating the scales has been put forward for consideration. There is broad support for continuing the work to explore the development of these reporting scales, but also recognition that this is a longer-term effort.

A more immediate need in order for the UIS and countries to report against Indicator 4.1.1 is to define “minimum proficiency levels” (and ideally additional levels of performance) for reading and mathematics and produce a reporting metric and a mechanism for linking existing assessments and their performance levels to this metric. This paper has been prepared by Management Systems International (MSI) to provide the UIS and the GAML with a proposed approach for meeting this immediate need. The paper presents the steps involved in constructing a “UIS proficiency metric”, or UIS-PM, for each domain and education level in Indicator 4.1.1 and linking NAs and CNAs to them. The UIS-PM will describe the development of the following:

- 1) *Content standards*: What students are expected to **learn** in reading and mathematics at the three levels of education defined in Indicator 4.1.1 – grades 2/3, end of primary and end of lower secondary. For the purposes of this paper, and reflecting discussions at the 21–22 September 2017 Hamburg meeting, these levels will be referred to as: by the end of grade 3 (lower primary), 6 (end of primary) and 9 (lower secondary).
- 2) *Policy descriptors*: What students are expected to **perform**, in generic terms, without content.
- 3) *Performance standards*: What students are expected to **perform** in terms of content, with respect to knowledge, skills and abilities.
- 4) *Proficiency scale map(s)*: How proficiency scales (i.e. performance levels) of various NAs and CNAs are aligned with the UIS proficiency metric.
- 5) *Socially moderated performance standards*: What students should obtain (in terms of a score) on their NAs and CNAs to be classified into the “desired” performance level for SDG reporting.

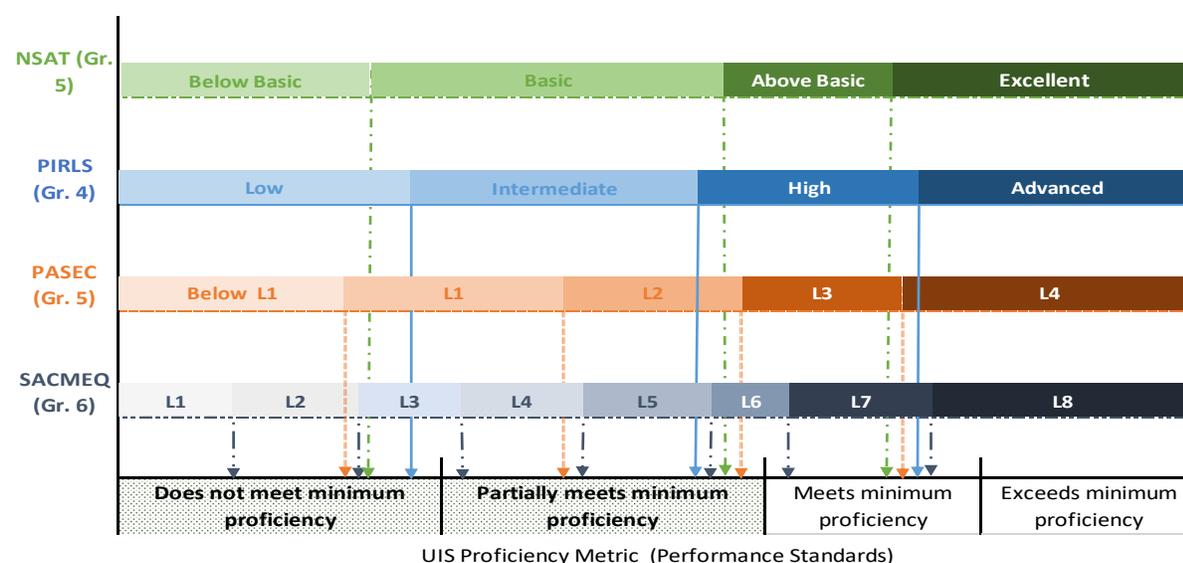
For linking the UIS proficiency metrics with other NAs and CNAs outputs, this phase requires the Global Framework contents for reference, the proficiency levels, the policy descriptors and the full proficiency description.

Linking the UIS Proficiency Metrics with NAs and CNAs

In order to explain the linking process and its relevant conditions effectively, the assumption can be made that the UIS proficiency metric for Grade 6 (end of lower secondary) reading has four performance levels (e.g. does not meet minimum proficiency; partially meets minimum proficiency; meets minimum proficiency; and exceeds minimum proficiency), as shown in Figure 1. The levels (without the lowest level) have been defined for both policymakers (i.e. policy descriptions) and other stakeholders (i.e. full descriptions). The performance levels and descriptions for Grade 6 reading have been defined based on the Steps 1 to 4 discussed below.

It is further assumed that the UIS has decided to count only the percentage of students who have mastered the required knowledge, skills and abilities defined in Partially Meets Minimum Proficiency and below of the UIS proficiency metrics for SDG 4.1.1 reporting – in other words, the percentage of students classified into the Meets Minimum Proficiency and Exceeds Minimum Proficiency levels of the UIS proficiency metric would be reported for Indicator 4.1.1.

Figure 1: Linking the UIS Proficiency Metric with National and Cross-National Assessments: An example



Social moderation or policy linking

Stage 1: Evaluate alignment of Performance-Level Descriptors (PLDs)

In this step, it is examined whether there is any alignment between the Partially Meets Minimum Proficiency description of the UIS proficiency metric and the performance-level descriptors (PLDs) of the NA and the CNAs. It is proposed that a three-point holistic scale (no or limited match, mostly matched and fully matched) is utilized to determine the degree of alignment. The rating rubrics include criteria related to matching strand level (i.e. specific content and competencies articulated in the PLD) and cognitive complexity level (i.e. low, medium or high) (Webb, 1997). For example,

- **No or limited match:** No or little match at strand level.
- **Mostly matched:** Matched fully at strand level, but somewhat matched at cognitive complexity level.
- **Fully matched:** Matched fully at strand and cognitive complexity levels.

A group of subject matter experts (SMEs) convene and rate by consensus the match between performance levels of the UIS proficiency metric, and the levels used in the NA and CNA. The performance levels of the NA and CNA that have holistic ratings of “mostly” or “fully matched” with the Partially Meets Minimal Proficiency level of the UIS proficiency metric will be used for reporting Indicator 4.1.1.

For example, it is seen in Figure 1 that Level 6 of SACMEQ, Level 2 of PASEC and the Basic level of NSAT (Namibia’s Census-based National Assessment) will most likely be rated as “mostly matched” with the Partially Meets Minimal Proficiency level of the UIS reporting scale. Therefore, the countries that participate in SACMEQ would report the percentage of students at Level 7 and Level 8, and countries that participate in PASEC would report the percentage of students at Level 3 and Level 4 for Indicator 4.1.1. Namibia would report the percentage Above Basic and Excellent.

The performance levels of the NAs and CNAs that have holistic ratings of “no or limited match” with the level of the UIS reporting scale (e.g. the Intermediate level of PIRLS in Figure 1) should go through the social moderation or policy-linking procedure described in step 2.

Stage 2: Set socially moderated performance standards for NAs and CNAs

In this step, socially moderated performance standards for the NAs and CNAs will be set using a standard-setting method in order to link these NAs and CNAs with the UIS proficiency metric. A yes–no variation of Angoff (Plake, Ferdous and Buckendahl, 2005), Bookmark (Lewis, Green, Mitzel and Patz, 1999), or Body of Work (Kahl, Crockett, DePascale and Rindfleisch, 1995) methods could be used for estimating three cut scores for Does Not Meet Minimum Proficiency/Partially Meets Minimum Proficiency, Partially Meets Minimum Proficiency/Meets Minimum Proficiency, and Meets Minimum Proficiency/Exceeds Minimum Proficiency on the NAs and CNAs for which there was not a “fully” or “mostly matched” rating (in the example, this would be PIRLS). However, the selection of the standard-setting method will be based on item and test formats of the NAs or CNA. If the test contains only multiple-choice items then the yes–no variation of Angoff method would be used. If the test contains both multiple and open-ended items then the Bookmark or the Body of Work method would be used.

For each NA and CNA, a group of 8 to 10 SMEs per domain convenes for a socially moderated standard-setting workshop. During the workshop, the SMEs are provided a thorough orientation to the standard-setting method and the UIS-PM PLDs. Then, SMEs are asked to provide individual and independent judgements about each item on the test to set their initial (also called Round 1) cut scores based on their understanding of the PLDs and experience with the student populations. For example, when using the Angoff yes–no method, the experts would say whether they thought a student meeting minimum proficiency would answer the item correctly. Other methods may require a different type of judgement to be made.



After collecting each SME's independent and initial judgements, the judgments will be aggregated to estimate the panel-recommended cut scores. The panel is then provided feedback on their judgments and information about the implications of the proposed cut scores (e.g. item difficulty, student classification). The feedback is provided so the SMEs have an opportunity to reconsider their initial judgements and to identify errors or any misconceptions about the process of setting the cut scores and the use of the cut score to classify student scores by category.

The cut scores estimated for Partially Meets Minimum Proficiency/Meets Minimum Proficiency and Meets Minimum Proficiency/Exceeds Minimum Proficiency and percentages of students classified into Meets Minimum Proficiency and Exceeds Minimum Proficiency levels (due to the Partially Meets Minimum Proficiency/Meets Minimum Proficiency and Meets Minimum Proficiency/Exceeds Minimum Proficiency cut scores on the NAs and the CNAs) would be used for Indicator 4.1.1 reporting. In other words, the students classified into the Meets Minimum Proficiency and Exceeds Minimum Proficiency levels of the UIS proficiency metric would demonstrate required knowledge and skills (as defined in the Partially Meets Minimum Proficiency level of the UIS proficiency metric) assessed on the NAs and the CNAs.

Construction of the UIS-PMs

Step 1: Define common content standards

- This task has already been undertaken by IBE-UNESCO. It will be necessary to review the work to date and determine if additional information on synthesis is required.

Steps 2 and 3: Define number of performance levels, determine labels, and write policy descriptions for the levels of the UIS-PM

- These tasks could be carried out by the GAML, perhaps during the next GAML meeting in November. It would require one full day with the GAML.

Step 4: Develop full descriptions for the performance levels of the UIS-PM

- One three-day workshop per domain would be required to define the detailed performance level descriptors. For each domain, a total of 15 subject matter experts (5 lower primary, 5 end of primary, and 5 lower secondary) from different countries would attend the workshop.

Linking UIS-PMs with NAs and CNAs

Step 1 is mandatory for examining the alignment between the performance-level descriptors of the UIS proficiency metrics and NAs and CNAs. However, step 2 will be planned based on the findings of step 1.

Step 1: Evaluation of PLDs

- A one-day workshop would be required for each NA or CNA to examine the alignment between PLDs of the UIS-PM. Five experts per domain/grade would participate.

Step 2: Set socially moderated standards



- One two-day standard-setting workshop would be required for each NA and CNA. A total of 8 to 10 subject matter experts for each domain/grade would attend.

Tentative timeline

The following schedule would yield UIS proficiency metrics by the end of February 2018 and linked CNAs (and potentially NAs) by July 2018.

	2018						
	Jan	Feb	Mar	Apr	May	Jun	Jul
Linking UIS-PM with CNAs (and NAs)							
Evaluate alignment of PLDs (workshops)							
Set socially moderated standards (workshops)							