UNESCO Institute for Statistics
FAQ for the application to upgrade SDG indicator 4.1.1(a)
October, 2018

Background

Sustainable Development Goal (SDG) 4 aims to ensure that, by 2030, "all girls and boys complete free, equitable and quality primary and secondary education leading to relevant and effective learning outcomes."

One of the indicators chosen for monitoring progress in achieving this target is indicator 4.1.1, the proficiency indicator referring to three levels of schooling: lower primary, upper primary, and lower secondary. The indicator reads as follows:

"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".

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

i. the percentage of students meeting minimum proficiency standards for the relevant domains (mathematics and reading) for each point of measurement (grades 2/3; end of primary and end of lower secondary); and

ii. whether different learning assessments can be considered comparable, and the conditions under which the percentage reported by one country can be considered comparable to the percentage reported by other countries.

The core challenge addressed by the UIS application to upgrade indicator 4.1.1(a)

Indicator 4.1.1(a) is currently classified as a Tier 3 indicator, meaning it lacks an internationally established methodology. In practice this means that there is no established method for reporting learning data which are drawn from different learning assessments and of varying difficulty on a common performance scale. There are many assessments already in use that produce more learning data for grades 2/3 than ever before; what's needed now is a way to compare and report them on a common scale.

In September 2018, the UNESCO Institute for Statistics (UIS) submitted a proposal to the United Nations Statistical Commission's Inter-agency Expert Group on SDG Indicators (IAEG-SDGs) to upgrade indicator 4.1.1(a) from Tier III to Tier II. The UIS application proposes a new reporting methodology to addresses the central challenge of comparability.

What follows below are a set of common questions about the UIS application.

UNESCO Institute for Statistics
FAQ for the application to upgrade SDG indicator 4.1.1(a)
October, 2018

Frequently asked questions

What are the foundational elements that need to be in place for indicator 4.1.1(a) to report data which are drawn from different learning assessments?

a) First, to include an assessment in UIS/SDG reporting, we would need to know that the assessment covers a minimum quantity of reading or math content/concepts at the given grade level. Fortunately, the UIS, along with UNESCO's IBE, have mapped a wide range of countries' curriculums and assessments and found significant overlap in what countries around the world expect their children to know and be able to do at different ages. Nevertheless, an assessment would need include some minimum level of content coverage to be considered a valid instrument for measuring the skills reported in SDG 4.1.1(a).

b) Second, we would need to know that an assessment was carried out with sufficiently rigorous procedures (e.g., sampling) and that it produced reliable data. The UIS have a variety of tools and manuals of best practice for countries to refer to on this question.

c) Third, we need a global definition of what all students should know and be able to do in reading and math by the time they complete grades 2 or 3, regardless what country they live in or what assessment they take. Fortunately, UIS and a range of stakeholders have already formalized this definition.

d) The last piece of the puzzle, and the focus of the UIS application to the IAEG-SDGs, is developing a method for translating an achievement score on any given assessment into the percentage of students in that country who meet the global definition of minimum proficiency. For example, if one country's reading assessment is conceptually very "difficult", a low score may be all that's needed to meet the global minimum proficiency standard. However, another country's assessment might be "easy" and therefore students would have to achieve a higher score to meet the global minimum.

It is this last point that is the central challenge at issue in the application to the IAEG. Fortunately, the UIS have proposed a new methodology for doing just that, which is discussed below.

Before discussing the method for comparing assessments, how do individual learning assessments define their own performance scales?

Every learning assessment starts by defining 3 to 4 different levels of performance. For example, an assessment may define student performance according to:

1. *Basic*: partial competency of prerequisite knowledge and skills for a given grade
2. *Proficient*: demonstrated competency in the prerequisite knowledge and skills for a given grade
3. *Advanced:* superior performance beyond proficient

From there, an assessment will describe the content knowledge and skills a student must demonstrate in the given subject matter. These are called "performance level descriptors" (PLDs). For example, an assessment may require that by the end of primary, in order to be *Proficient* in reading, a student must be able to "trace the development of an argument and evaluate the author's claims and evidence in a text", among other tasks. There may be multiple PLDs for each performance level.

For illustration purposes only..

1. *Basic*: partial competency of prerequisite knowledge and skills for a given grade
    a. PLD Example: demonstrates limited comprehension of literary and informational texts
    b. PLD 2
    c. …
2. *Proficient*: demonstrated competency in the prerequisite knowledge and skills for a given grade
    a. Example: traces the development of an argument and evaluate the author's claims and evidence in a text
    b. PLD 2
    c. …
3. *Advanced*: superior performance beyond proficient
    a. Example: The student thoroughly compares and contrasts texts in different forms or genres
    b. PLD 2
    c. …

PLDs are developed by panelists of content experts and practitioners. There are well established procedures for how to conduct these workshops to ensure a high degree of quality and validity.

Assessment items (i.e., individual assessment questions) are then developed to align with each of the PLDs and test for student ability in these skills.

Finally, numeric achievement scores are set to distinguish between each performance level. These are often called "cut scores" and indicate for example, that below a certain numeric score a student would be considered *Basic* and above that score they would be considered *Proficient.*

Why can't we just use the reporting scales from each country's national assessment?

As discussed above, while national and cross-national assessments do determine proficiency levels and report these for their students, the performance level descriptors that describe the content knowledge and skills a student must demonstrate in the given subject matter vary between assessments. As a result, the current reported outcomes from different assessments are not immediately comparable.

The additional process described below is required to determine a common standard across all assessments. Countries will continue reporting using their own scale, but the below process will allow UIS to determine the score on each assessment that represents the global minimum proficiency standard to enable comparison across assessments.

If each assessment has its own definition of student ability, how are we able to determine which score on each assessment represents the global definition of minimum proficiency?

The easiest way to ensure we are measuring student ability in a comparable way would have been to use a single global assessment. However, that is not logistically or politically feasible, so UIS have found an alternative method.

UNESCO Institute for Statistics
FAQ for the application to upgrade SDG indicator 4.1.1(a)
October, 2018

Remember that the first two steps in creating any new assessment is to define the performance levels and then to describe the skills and knowledge that students should be able to demonstrate (PLDs) within each of those levels.

In that sense, the SDGs have already defined the performance level as "minimum proficiency". The UIS took the next step of describing the skills and knowledge students must demonstrate in reading and math by the end of grade 2 and 3 to meet that "minimum" standard. That is, the UIS in collaboration with a wide range of stakeholders (assessment agencies, GAML, the TCG), have described what it means to be minimally proficient in reading and math at this age. The grade 3 descriptors for reading and math for SDG 4.1.1(a) are:

- Reading: Students read aloud written words accurately and fluently. They understand the overall meaning of sentences and short texts. Students identify the texts' topic.
- Math: Students demonstrate skills in number sense and computation, shape recognition and spatial orientation.

If the world were going to create a single global test (and to clarify there was never an intention to create a single test, this is only to illustrate the process), we would now have to develop assessment items that would demonstrate whether students are capable of these skills. Instead, where existing assessments are already in use by countries, UIS are proposing to produce an equivalency with the SDG 4.1.1(a) performance level descriptors (above).

The process for determining this equivalency is called "policy linking" and involves panels of subject matter experts (in literacy or math), reviewing the individual assessment items from each assessments in relation to the global definition of minimum proficiency. Using an internationally recognized standard setting methodology (for example, Angoff or Bookmarking), the panelists use their expert judgment to evaluate whether a correct answer on individual items would satisfy the global minimum standard. In this way, and through multiple rounds of reflection, calibration, and scoring, the group comes to a consensus about what score a student would need to achieve on a given assessment to meet the global definition of minimum proficiency.

There are slightly different approaches to this process that depend on the format of the assessment (e.g., multiple choice, short answer, portfolio, etc.). But in all cases, the process is well established in the field of educational measurement with best practices for selecting panelists, conducting the workshops, and calibrating experts' judgments about item difficulty.

Has "policy linking" been done before across countries? How will countries carry this out going forward?

To date, policy linking has been done mostly within countries or at sub-national levels to draw equivalencies between two different assessments. It has not yet been carried out internationally across multiple assessments.

The UIS is in the process of planning for pilot implementations in a few countries in January-April 2019. From these pilots, toolkits and best practice protocols will be created to help countries carry out this process on their own or with the support of technical advisors.

How was consensus achieved among stakeholders and countries for the definition of minimum proficiency and for the policy linking process?

UNESCO Institute for Statistics
FAQ for the application to upgrade SDG indicator 4.1.1(a)
October, 2018

The UIS first brought together all the major international, regional, and citizen led assessment agencies to review the performance level descriptors already in use across the world. There was little need to create something new when many assessments were already evaluating student abilities and with substantial overlap in content and performance level.

With this group, the UIS put all of the performance levels from each of these assessments onto a single ordinal scale and then undertook a consensus building process to set the minimum proficiency level for grades 2/3, the end of primary, and the end of lower secondary.

The UIS is also seeking the endorsement of the Global Alliance for Monitoring Learning (GAML) and the Technical Cooperation Group on the Indicators for SDG 4 (TCG) on the minimum proficiency definition and the policy linking methodology. The UIS anticipates the full endorsement of these groups by November 20th 2018.

The GAML is an advisory group to the UIS on matters related to SDG 4 and has been meeting since 2016. It is comprised of a range of stakeholders, including experts and decisionmakers involved in national and cross-national learning assessment initiatives, donors, and civil society organizations advocating for education.

The Technical Cooperation Group provides more formal guidance to UIS on its approaches and is composed of 38 regionally-representative members from Member States, international partners, civil society and the Co-Chair of Education 2030 Steering Committee, with the UNESCO Institute for Statistics hosting its Secretariat.

Between ongoing meetings with GAML, the TCG, and the close collaboration with all of the major international and regional assessment agencies, the UIS is ensuring wide agreement on its policy linking approach.

Will this process account for differences in the difficulty of languages?

Policy linking is intended to be flexible and driven by experts in each country or region. In this way, the meaning of minimum proficiency according to SDG 4.1.1(a) will have to be interpreted appropriately for each language and its characteristics and alphabet/script. While there are meaningful differences in language difficulty, there is also significant overlap in what countries expect their children's reading abilities to be by the end of grade 2 or 3. The fact of this overlap will make the policy linking process easier and more accurate, in spite of inherent language differences.

The entrance age in primary and the provision of pre-primary education varies across countries. How will these considerations be dealt with to ensure appropriate comparisons?

The aim of SDG 4.1.1(a) is not simply to create comparisons for their own sake, but to understand why some countries perform better than others. If the measurement of student learning in grades 2/3 highlights that countries without pre-primary do worse in lower primary than similar countries, then that will be an intended and beneficial outcome.

Further, while age of entrance might have some bearing, there is only a trivial number of countries in the world where entrance to primary school is at an age one could claim is developmentally inappropriate. For nearly all countries, students enter at age 6 or 7, and the elementary skills being assessed in most existing tools (such as ASER, UNICEF, EGRA, Save-the-Children's tool, etc.) for Grades 2/3 are easily mastered by children who are 7 or 8 (Grade 2) or 8 or 9 (Grade 3), with even modestly adequate instruction.

UNESCO Institute for Statistics
FAQ for the application to upgrade SDG indicator 4.1.1(a)
October, 2018

Additional resources and blogs on the topic of early grade learning and measurement more generally

Harry Patrinos from the World Bank makes the economic case for early grade assessment and describes the World Bank's new Human Capital Index, which maps data from early grade assessments onto a harmonized learning outcomes scale.

Hilaire Hounkpodote, coordinator of the PASEC regional assessment for West and Central Africa, describes how PASEC captures reading and mathematics at Grade 2 in a regionally comparable way, and how this contributes to SDG monitoring.

Rukmini Banerji of Pratham describes the life-changing benefits of early grade assessments from the viewpoint of Meera, a student in Rajasthan who was attending school for years but not learning.

The World Bank, UNICEF and Pratham share their perspectives on why assessing learning early on is so important for children around the world. [link forthcoming]

Hannah-May Wilson of the PAL Network describes the citizen-led assessment approach, which provides data on early reading and mathematics in countries where often there is no national data produced by the government.

Silvia Montoya, Director of UIS, has written several recent blogs, including one describing the social moderation approach and consensus on definitions for minimum proficiency, and one on the tools developed by UIS for monitoring SDG 4.1.1.

Immediately after the SDGs were adopted in 2015, Pauline Rose of the University of Cambridge argued that to meet the ambitious goals for education, we need to start monitoring children's learning early.