Technical Cooperation Group
9th meeting
22-24 November 2022

WG/GAML/19

FOUNDATIONAL LEARNING METRICS: REPORTING TOWARDS THE SDG

UNESCO Institute for Statistics

Global Alliance to Monitor Learning (GAML)

23 November 2022

unesco
Institute for Statistics

Foundational Learning metrics: reporting towards the SDG

Since the declaration of the SDGs and the ensuing processes, the UIS has been the custodian agency of most of the indicators under SDG4. It is the case, arguably, that most international agencies and countries are logically interested in target SDG4.1: it is the one most held in common and as a high priority by countries themselves and agencies, if measured by the level of funding and attention it receives, explicitly or received. Neither of this target's two indicators, namely 1) learning and 2) completion, are easy to measure, but one could make the case that learning is harder to measure, as it is divided into three levels and requires two subjects: grades 2/3, end of primary, and end of lower secondary on one hand; and reading and mathematics, as subjects, on the other. Completion is a one-off, binary event, is imputable through standard school and Education Management Information System (EMIS) record-keeping and demographic data, and is therefore much easier to measure, in a way that is comparable across nations, for those three reasons[1]. Learning achievement is another matter: it is inherently much harder (both technically and financially) to measure, especially in an internationally comparable manner, as demonstrated by the fact that relatively few low-income countries do it, and by the fact that it typically requires teams of PhD-level specialists with many years of experience to do it well.

To deal with the measurement difficulties, the UIS, through the GAML and TCG processes has been leading and funding technical development. These have focused on many measurement aspects, but the issue of comparability has been the most prominent. Much has been gained through processes such as MILO, the Rosetta Stone project, AMPL, and others. To gauge progress at the specific point in time of late 2022, one can think in terms of a ratio of need relative to existing advancement. Of the three 4.1.1. sub-goals, it is arguably the case that 4.1.1.b, namely end of primary, and 4.1.1.c. namely end of lower secondary, have the lowest ratio of need to existing advancement because relatively so much progress has been made here, and because 4.1.1.a., in addition to being much further behind technically, is also in great need from a programmatic point of view, as so many institutions have focused on learning poverty, defined approximately (and variously, but with the same emphasis on foundational skills) as the inability to read simple texts, and perform the equivalent level of arithmetic, by age 10 or so.

[1] And comparability is inherent in point 30 of "Report of the Inter Agency and Expert Group on Sustainable Development Goal Indicators," E/CN.3/2016/2/Rev.1*. https://unstats.un.org/unsd/statcom/47th-session/documents/2016-2-IAEG-SDGs-Rev1-E.pdf.

That said, there is movement on 4.1.1.a. Individual institutions and countries, including multilateral agencies, bilateral agencies, and others, have created (or are planning) instruments such as the FLM (Foundational Learning Module) module of UNICEF's MICS surveys, the EGRA/EGMA class of instruments, the PAL class of instruments ("citizen-led" such as ASER and Uwezo). UIS has been in discussion with the ACER of Australia to create an AMPL (Assessments for Minimum Proficiency Levels) suitable for 4.1.1.a, called tentatively AMPL-a. There are many important technical challenges to overcome in this regard. One of them is that, as is well known from years of application of instruments such as the PAL class of instruments or the MICS FLM, many (in some cases, most) children in grades 2 or 3 in the poorest countries are unable to read the instructions of the assessments or, worse, are simply unable to decode text in the most elementary sense, e.g., cannot even decode or sound out simple individual words such as "cat" or "dog" or even individual letters. This means that one must measure at this simplest level if one is to get a sense of what children can do if they are way below the minimum level required for reporting under 4.1.1. It is important to note that instruments such as the MICS FLM, the PAL, and EGRA/EGMA classes of instruments already have an enormous user base of countries and international agencies (e.g., UNICEF, USAID, South Africa) that know and appreciate these instruments but mostly use them for tracking of improvement projects—**and it is an improvement *towards* the SDGs** that one ought to really be concerned about, not measurement for its own sake, no matter how good are the qualities as metrics.

As a result, the UIS would like to appoint an entity to carry out the following tasks, ideally performed in approximate sequential order. Tasks 1-5 are largely desk work, involving reading, key informant interviews, and possibly some re-analysis of existing data. Task 6 is much more expensive and would involve field work.

1. Study the characteristics of the most used truly *foundational* skills measurement tools, specifically EGRA/EGMA, PAL, and the MICS FLM and possibly the "components" section of LAMP. Also, study the proposed characteristics of AMPL-a as it will be necessary to "bridge" to these in the sense of a conceptual and pedagogical continuum or learning progression. The issues to study should include:
    a. the validity and useability in terms of the MLP framework and GPF in terms of what these really assess;
    b. their statistical and psychometric properties, especially in terms of sample size needed with the objective of generating a minimally acceptable confidence interval around the ratio that meets the MPL with statistical and psychometric properties;
    c. use the technical write-ups and/or data archives on these instruments that already exist, as these do exist and there is no need for more field measurement.

2. Analyze implementation and cost issues such as the pros and cons of various approaches to field deployment.

3. Analyze the pros and cons of different approaches to data gathering and storage (paper versus electronic).

4. Analyze and propose links (progression) to the proposed AMPL-a, including ideas around concordance and prediction in either direction. Possibly predictive validity to AMPL-b or the MILO.

5. Propose a best-of-all, and highest value-for-money hybrid of the various approaches that are as easy to implement as possible while keeping to the minimum statistical, sampling, and psychometric requirements needed for tracking any one country's progress, and reporting globally.

6. Test this approach in three countries, first as small pilots and then with an "at-scale" effort, and report the learning from the field deployment, and the technical results as well as, of course, the learning levels measured. Include the desk work of linking to AMPL-a or AMPL-b.
    a. Report on the results of points 1-5 as one separate report.
    b. Report on results of point 6 as a separate report.

7. Disseminate the results of point 6 (both a and b) to at least 3 key sets of stakeholders: international donor agencies, a selection of 5 NGOs who represent intensive users of these sorts of assessments in the past, and 5 representative governments who might already be familiar with these instruments and would like to consider "officializing" them more, and 5 who are new to the process.