

REPORTING SDG INDICATOR 4.1.1A: DEVELOPMENTS TO DATE

Background document

DRAFT
18 June 2024

DRAFT

Update on TAG Development

1. Background

SDG 4.1.1.a focuses on the “Proportion of children and young people ... in grades 2/3 ... achieving at least a minimum proficiency level in (i) reading and (ii) mathematics.” The indicator thus measures learning at the foundational level while [SDG indicator 4.1.1](#) measures learning at several levels, namely early grades, end of primary, and end of lower secondary.

Up to late 2023, there had not been much [country measurement and reporting](#) on SDG 4.1.1a. As a result, the Inter-Agency and Expert Group on SDG Indicators ([IAEG-SDGs](#)) ‘demoted’ the indicator from Tier I to Tier II in [October 2023](#) due to low coverage, putting its status at risk during the 2025 framework revision. The community of interest concerned with foundational learning, including many important institutional stakeholders, and thought leaders, immediately expressed deep concern in blogs and at various meetings. Those interested in measurement quickly mobilized to increase the count by laying a better technical foundation for measurement and strengthening coordination of funding for country-level measurement efforts.

This blog, along with an upcoming companion blog, describes the actions taken by the UIS as the custodian agency for SDG 4.1.1a to reinforce its status, outlines current progress, and proposes next steps to be taken to further consolidate it. The road ahead, though, demands consistent collective action.

In the medium to long term, the vision is to improve countries’ own measurement capacity, using it not only for global reporting, but most importantly for institutional advancement.

The two objectives of quickly increasing coverage and developing better local capacity do not contradict each other; however, there is a risk of conflict if the process is not handled cautiously. “Dumping” or “parachuting” measurement initiatives onto countries without careful planning and collective coordination can strain countries’ limited resources which prevent them from pursuing multiple objectives simultaneously.

It is important to note that the measurement of SDG indicator 4.1.1a, especially using the newer instruments proposed, presents particular technical problems that are not as binding in the case of indicators 4.1.1b and 4.1.1c (end of primary and lower secondary) as

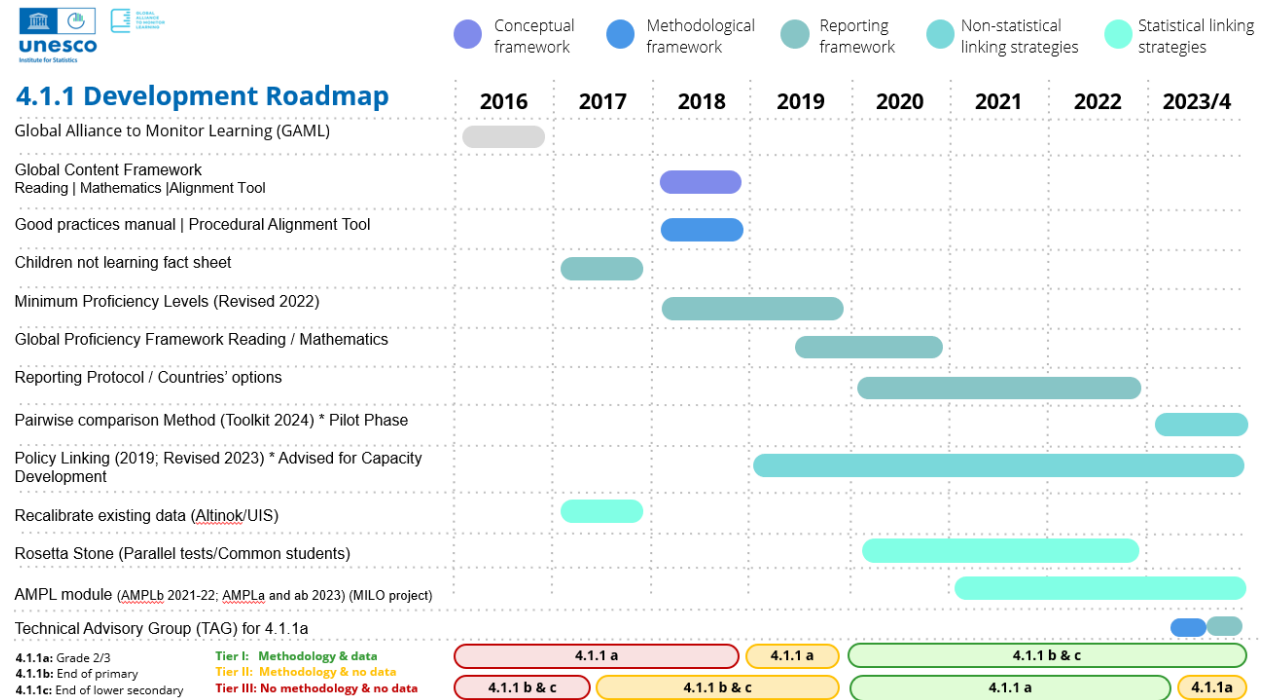
discussed [here](#). These constraints are relevant to understanding the necessary steps forward and addressing them is crucial for progress.

2. Mobilization of efforts, late 2023 and early 2024

Since 2016, the [UIS had developed frameworks](#) that created a base from which to mobilize efforts to improve SDG indicator 4.1.1a (see **figure 1** below).

Crucial to measurement were the establishment of Minimum Proficiency Levels ([MPL](#)), later related to the Global Proficiency Framework ([GPF](#)) that serves as guidance for the learning progressing in a given grade and domain.

Figure 1. Developments for SDG indicator 4.1.1 since 2016



It is important to note that part of the problem with accepting the newer measurements for reporting on SDG 4.1.1a (EGRA/UNICEF), mostly based on 1 on 1 measurement, is that they had not been explicitly linked to the MPL/GPF, despite the existence of these frameworks for several years, as the figure above illustrates.

3. Eligibility criteria for reporting indicator 4.1.1a: progress

The section below outlines the main steps taken since October 2023 aiming at reinvigorating SDG 4.1.1a (see **figure 2**).

1. GAML meeting (Paris, 6-7 December 2023)

During the UIS-coordinated [Global Alliance to Monitor Learning](#) (GAML) meeting of experts and stakeholders, it was noted that while substantial measurement activities were ongoing—primarily for advocacy, program design, program tracking, and evaluation—much of this data was not being reported, was not publicly available, and technical documents were scarce and scattered (see discussions [here](#) and [here](#)). All these points highlight existing measurement efforts were not good enough either for either global reporting or for tracking national progress over time.

Most importantly, the meeting provided a clear way to link to the MPL for these tools as described in a [document with a set of eligibility criteria](#) (both psychometric and procedural) that assessments would need to meet in order to report was shared and approved.

2. January/Feb 2024: feedback

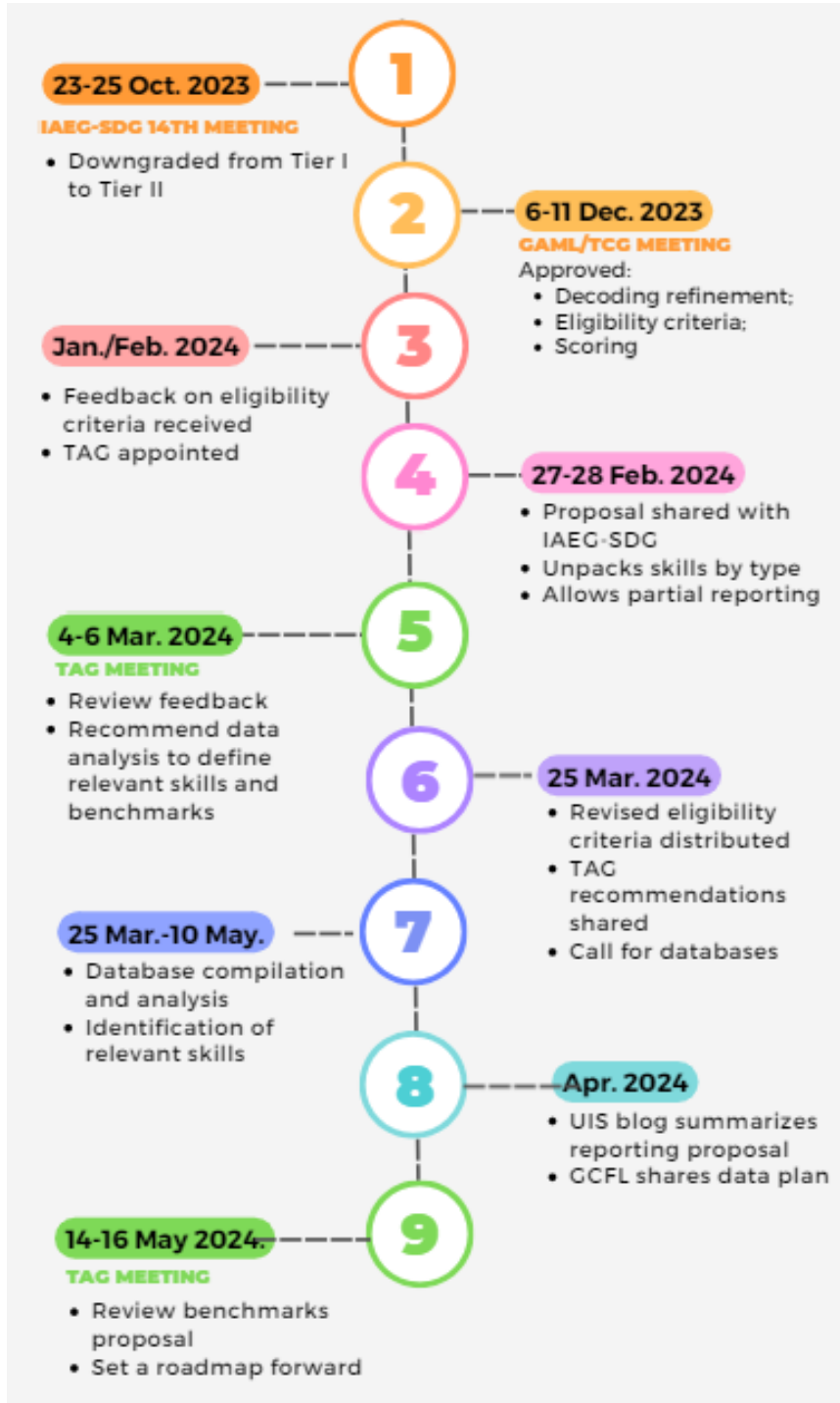
The UIS shared the eligibility criteria document with all stakeholders for their feedback. The engagement of the community of interest was reflected clearly in the hundreds of comments received on the document, to which UIS responded in writing.

3. TAG meeting (4-6 March 2024)

The UIS convened a Technical Advisory Group (TAG) of experts specialized in measuring foundational learning in early March to analyze feedback, further refine the criteria document and define the steps forward.

The UIS shared the revised version of the [eligibility criteria document](#) on 25 March along with the TAG recommendations which included further data analysis to define relevant skills and benchmarks; and a call to share databases for the analysis.

Figure 2. Timeline showing the main steps taken since October 2023



4. UIS reporting proposal: Reporting to the MPL, on the precursors to the MPL, or reporting both?

At the end of February, the UIS made a proposal to the IAEG-SDGs to unpack the reporting of SDG 4.1.1a.

As a background, the proposal aims to deal with two real issues with measurement in the foundational years:

- a. The fact that language matters much more at this level than at the end of primary or lower secondary; important that is reinforced by the different nature of languages.
- b. That many children have not mastered “reading to learn” and may or may not be making progress with the skills needed to master “learning to read.”; making it necessary to identify where they are in the learning progression.

Learning to read in the mother tongue is increasingly the policy for the foundational grades in most countries. Yet, some languages are written in ways that are inherently harder for children to learn to read. Later, children tend to master secondary languages, and most school systems do not teach in the home language by the end of primary school.

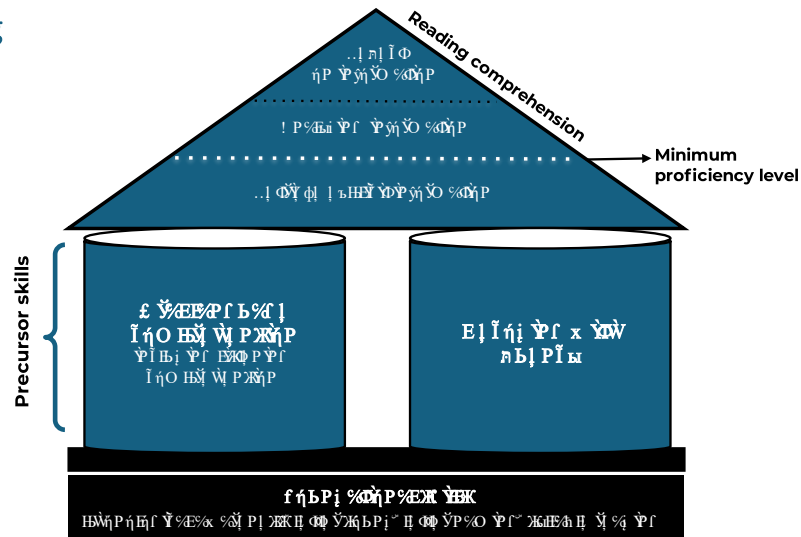
This creates an inherent comparability problem: progress in a transparently written (“phonetic”) language and/or one with an easy script (such as Spanish, some other European languages, many African languages among others) looks very different from progress in languages where the correspondence of sound to print is complex, or where the scripts are more complex or not even phonetic (e.g. Chinese, Arabic, Hindi). So, if one is to talk meaningfully about children’s progress in an internationally comparable manner, this issue must be tackled head on.

In addition, in many countries, especially the low-income and lower-middle-income one, where there is relatively little measurement, and yet where we know children are falling seriously behind, children are realistically only starting to master the most basic or “precursor” elements of reading by grades 2 or 3.

The UIS proposal is illustrated in **figure 3** below which presents a graphical description of these skills and the idea of setting benchmarks is described [here](#).

Figure 3. Unpacking reading in early grades

..M E wò g



Fortunately, many of the newer assessments address either or both language and early or precursor skills issues. Aside from the fact that presumably there is a good bit of data or data effort under these assessments, these two features are an important advantage. Yet the disadvantages noted above (e.g. lack of standards for cross and temporal comparability, lack of documentation) are also real.

To take advantage of the efforts, the UIS proposed a table for reporting where children who meet certain benchmarks for the “precursor” skills of reading, by specific language or language groups (that hold language difficulty empirically and approximately constant), **might possibly be counted, or partially counted, using some sort of weighted scheme, but where the eventual goal was reading comprehension.**

4. Second TAG meeting (14-16 May 2024)

The TAG discussed data analysis and benchmarks-setting. The conclusions of this meeting will be shared with stakeholders via a separate document and can be summarized as follows for the moment:

- a. If the data are of high quality and plentiful, it is possible to establish benchmarks by language or by language group. Experts from South Africa and Kenya informally shared their national methods. This is good news.

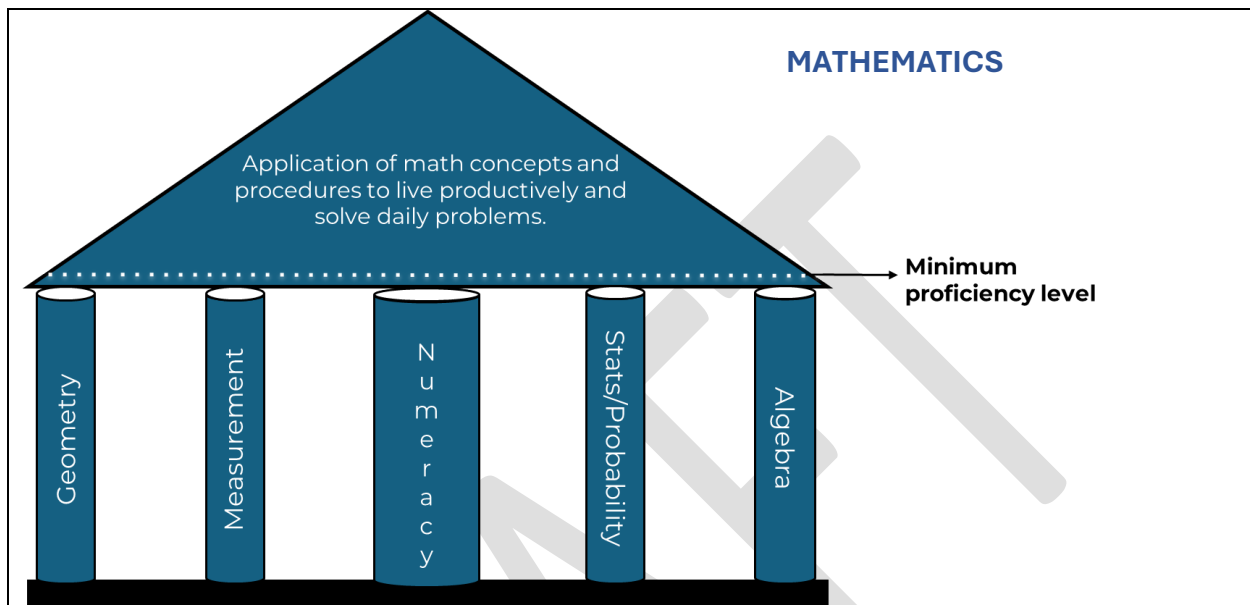
- b. Some **more steps** are needed before one could claim to have benchmarks for a significant number of languages/countries in reading, and move on to reporting:
- i. more **statistical and psychometric analysis** to comply with a minimum rigor that could be defensible.
 - ii. **more precision in criteria and standards** for reporting **would help** to delimit the difficulty of the assessments done so far. For example, in mathematics, it is not enough to call for “10 numeracy items” as it would be one thing to have all ten be for single-digit addition without carry-over, and another thing to have 2-digit multiplication.
 - iii. **more data to set benchmarks** is needed, as for some important languages there were only one or two unidentified countries, and it was unclear how representative of the language the countries could be.
 - iv. **more examples** of national experiences in setting benchmarks. Even though the methods may not be IRT-based, if rigorous and well-supported by local stakeholders and experts would help the task.

The MPL will report the % of students achieving the reading comprehension benchmark. However, reporting of the precursor skills would be asked for, and shown against benchmarks, to help countries move towards that MPL level assessing progress towards reading comprehension.

Mathematics

A similar idea was agreed upon for numeracy (see accompanying graphic), using it (at least for now) as the skill that counts for reporting, though the reasoning is not parallel, because numeracy skills are not precursor skills to mathematics in the same way that reading accuracy is a precursor for reading comprehension. Numeracy and mathematics are important and correlated skills, but they are not to be thought of as “precursor” skills, and the reason to focus on numeracy, as opposed to the reason to focus on comprehension in reading, is more a matter of expedience than a matter of fundamental priority. A call for databases on numeracy/mathematics results from the newer assessments will follow.

Figure 4. Unpacking reading in early grades



5. Where we stand now and the road ahead

Even though indicator 4.1.1a continues to be part of the SDG framework, relieving some of the pressure, hard work and improved collective coordination are still needed and should aim at more than keeping the indicator in the framework. The goal is to ensure that by 2030, all countries can produce regular and comparable data on learning outcomes through more transparent, coordinated, and targeted support to learning assessments.

A few important issues seem to come up repeatedly among interested stakeholders and it seems that it is necessary to clarify where we believe the community stands or needs to stand.

1. **The plan is that all assessments that meet the criteria should be acceptable.** At least in the medium term, agencies that support assessments should all ideally stop funding or worrying about whether a given assessment is declared acceptable or is preferable. What matters is not whether assessment X or Y should be promoted or is acceptable; what matters is whether an assessment meets the criteria of optimality, that is, whether it meets the technical criteria and is optimal for countries' needs. This includes the specific technical or psychometric criteria

the process described in the previous blog, but also the “meta” criteria of whether an assessment is part of a sustainable approach to country-based measurement.

2. **Country ownership and sustainability are central** to the UIS position of promoting a more efficient marketplace of assessments, as well as less tied aid , and this has been discussed in the past in blogs such as [here](#) and [here](#). Some might see this idea of de-linking assessment support from requiring specific assessments as a long-term goal and an ideal. But we believe that with the urgency of late 2023 around SDG 4.1.1a reduced, and with plans increasingly in place, now is the time to push hard towards a more transparent and equitable ecosystem for assessments. The next two points are related to that.
3. UIS will continue to promote the notion of a [virtual fund or a coordinating mechanism](#) whereby donors can make funding available to countries that choose to carry out an assessment that meets criteria, without forced preference for a given tool¹. The UIS has consistently sought for such a mechanism not only through blogs, such as [here](#), but by the introduction of a [virtual register](#) of investment in learning data to record, coordinate and ‘broker’ existing donor support. The UIS will prepare a Concept Note and insist on this issue. Such an approach is among the key points of the success of [the Vaccine Alliance](#) (GAVI).
4. To assist in this process, UIS will also commission a “buyer’s guide” that implicitly adopts the technical criteria but also looks beyond these, at issues such as the track record or planned attention of an assessment approach to issues of local capacity, sustainability, and cost among other criteria. This buyer’s guide can be part of how the virtual fund helps allocate particular assessments to particular countries, based on countries own analysis, according to the buyer’s guide, of which assessments suit them best.
5. UIS will continue to coordinate the remaining technical tasks emerging from both the March Technical Advisory Group (TAG) on the [Eligibility Criteria document](#), and the May TAG on benchmarking. This includes, specifically:
 - a. Increasing the rigor of the Item Response Theory (IRT) based benchmarking process.
 - b. Analyzing item difficulty in existing databases for both reading and numeracy/mathematics with a view to better delimit the level of difficulty of the newer assessments, in linking to the Minimum Proficiency Level (MPL), to make them more comparable.

¹ For more information, please refer to Annex I.

- c. Calling for more data to improve the number of languages, as well as numeracy and mathematics skills, for which one can craft benchmarks.
 - d. Calling for examples of country-based benchmarks that have been set for reading, especially precursor skills, and for numeracy / mathematics (if any).
 - e. Calling for examples, or getting them from existing assessments with permission, of items matching the level of difficulty at the median of low-income and lower-middle-income countries, that can be provided as guidance. These can be extracted from examples from the newer assessments.
 - f. Setting up a mechanism for applying the various criteria, on an ongoing basis, as interested parties apply to have certain assessments in particular countries to be used for reporting specifically on SDG 4.1.1.a, though the mechanism could also be used for the indicators at the end of primary and lower secondary (4.1.1.b and 4.1.1.c). The work would basically apply all the Criteria – both existing and emerging. This mechanism will be coordinated by UIS but will be overseen by an independent steering committee of country experts to vouch for the impartiality and quality of the work done by a team of mid-level experts. This mechanism can be linked to, or administered as part of, the virtual fund mentioned above. How much to link the two is still under consideration.
6. UIS will continue to coordinate and support the plans to increase the number of countries reporting, and communicate these to the IAEG, the Statistical Commission of the UN, along the appropriate channels, and as time goes by. UIS will continue with the practice of keeping all parties informed at key nodes in the process.
 7. Going forward, it is essential that all interested parties collaborate openly and respond to calls for data or documentation. The support of those who have funded these processes so far is appreciated, while noting that more may be required, as the analytical and consensus-generating tasks that are needed are an unexpected strain on UIS resources.

6. Next steps

The next steps include the following:

1. The TAG has to complete the definitions as clarified throughout the paper.

2. The TAG has to finetune the eligibility criteria.
3. Once the definition of the reference for reporting reading comprehension and mathematics is finalized, it will be shared with the Education Data and Statistics Commission (formerly known as the Technical Cooperation Group on SDG 4 indicators) for its consideration.

7. HSCL decision on indicator 4.1.1a

The SDG4 High-Level Steering Committee (HLSC) met on 17 June 2024. The first decision of the HLSC with respect to accelerating and monitoring progress towards SDG 4 was about SDG 4.1.1a.

“The SDG4 High-Level Steering Committee:

1. Recognizes the importance of the SDG global indicator on early grade learning and the mandated role UNESCO plays in achieving consensus on the minimum proficiency level; and welcomes the consultative and transparent process of the Technical Advisory Group to enable countries to report on building blocks to minimum proficiency in reading and mathematics, the HLSC:

i Requests bilateral and multilateral agencies to make available resources to allow countries to choose from a menu of assessments, including national assessments, that meet the agreed upon eligibility criteria for 4.1.1 (a)

ii Calls on countries to support the formative purpose of national assessment systems; and

iii Urges countries with existing plans for reporting on SDG indicator 4.1.1a to share them at the 2024 Global Education Meeting, to be held in Fortaleza, Brazil on 31 October and 1 November and calls on countries that have no such plans to develop them by February 2025.

ANNEX I. TERMS OF REFERENCE - SCOPING STUDY ON A VIRTUAL FUND AND VETTING FUNCTION FOR LEARNING ASSESSMENTS

Background

It has become common cause among international organizations, NGOs, and think tanks that low-income and lower-middle income countries (in particular – but the problem sometimes extends to upper-middle income) do not sufficiently assess their children’s learning levels. This makes it difficult to make progress: if one does not know where one is, it is hard to decide how much effort to make. Since learning levels are now known to be predictive of economic growth, and, much more importantly, human well-being and citizenship skills, the lack of measurement is not just an education problem of concern to measurement institutions such as the UNESCO Institute for Statistics, but is, instead, a serious development issue. Countries themselves might not perceive the need to measure as much as global actors do. This suggests that subsidizing such measurement as both a local merit good and a global public good, might make sense. While countries may not always realize the value of measurement, some do, and UIS constantly receives requests for help and coordination, and other multilateral and bilateral agencies and philanthropies receive such requests.

But assessments continue to be under-funded. More importantly, what funding there is, is often chaotic in many respects. Funding may be tied to a particular assessment X that a country must use, if it is to get funding from donor Y. This deprives countries of choice or drives them into sub-optimal choices if assessment X was not the ideal one for the country. In the past few years there have been several blogs written by UIS staff or UIS associated thought leaders about the inefficient nature of the assessment market. Any potential consultant is urged to peruse these, available [here](#), [here](#), and [here](#). The issues with the market include lack of price transparency, high barriers to entry due to economies of scale, un-pairing of the basic assessment as a good, from a more sustained capacity development service. UIS has proposed many solutions to these issues, and some of them are on track to being implemented or at least scoped out. For instance, UIS has worked to create item banks, or mini-assessments that can be used by countries or incorporated into their assessments, via the AMPL effort, described [here](#). In addition, UIS is planning a “buyer’s guide” that would help countries choose an assessment.

One of the issues that plagues assessment development in many countries is the uncoordinated and somewhat chaotic way that countries are provided with assistance. Some countries have had three or four EGRAs within one or two years for instance. In other cases, there has been no assessment of any kind in years.

A more recent development is that UIS has been asked, by the community of partners interested in the measurement of SDG indicator 4.1.1.a, to develop criteria for accepting a set of newer, non-traditional assessments as being reportable for international comparability purposes. UIS is preparing a guide or checklist of criteria for the eligibility of particular assessments in particular countries. Importantly, this could include national assessments. However, UIS does not have the staff to provide the approval of particular assessments; that is, to apply the checklist or guidelines. One aspect that one would envision could become extremely labor intensive is if a country or assessment provider would like to report using assessment X, but the application misses significant numbers of the criteria. In that case, UIS would like to be able to provide suggestions on how to improve, how the country could get technical assistance to improve, and so on.

Product desired, questions to be addressed

UIS would like to appoint a consultant to draft a Scoping Study to create some form of function, ideally within an existing institution, to deal with the two problems above.

The concept note or scoping study should not exceed 30 pages in length. Annexes can be of any length.

Questions to be investigated include:

1. The name of such a function, to make it attractive, and to differentiate it from existing facilities.
2. Its legal status or affiliation.
 - a. For instance, could it be something UIS outsources, with donor funds, to a consultancy or think tank, ideally one that could be objective? What would be the ideal nature of the procurement in that case? How competitive?
 - b. Or could it be an existing official agency such as Statistics Canada (just to pick more or less at random) with whom one would then have an MOU, rather than an outsourcing contract. Could they do it as an in-kind donation from the relevant government to the UIS, or would a transfer of funds be needed from UIS?
 - c. Most likely, it would not be efficient to put such a function or facility “under” an existing donor such as GPE or the World Bank, as it is hard to see what is gained in efficiency by doing so. However, the consultant can entertain and discuss this option.

3. Its governance. UIS would propose keeping managerial or coordination control, for the sake of operational efficiency. However, the governance, as distinct from managerial control, is something that needs to be scoped. Does one need a Board? A steering committee? Who should be on it? Official donors only? Recipient countries only? Can philanthropies be included? Can one offer models? (E.g., GPE Board?)
4. Size. What size would one be talking about? How much funding would be needed per year, assuming that only the operational function is to be funded, and that the actual payment for assessments would be contracted between funding partner X and country Y would pass through an assessment service provider or would flow to the country itself. Or should UIS in fact insist on intermediating the funds, in a fiscal or fiduciary sense? What would be gained or lost by doing so?
5. What would be a good process for countries to “apply” for funding? How formalistic? What is the tradeoff between making it fair and transparent and ensuring good use of the funding versus requiring bureaucratically elaborate plans and reviews by large committees, etc.?
6. How one the funding be targeted so as to help those who truly need it, rather than those who would be perfectly able to afford it on their own, but simply have not done so? That is, how does one avoid the perverse incentive to not pay for one’s own assessment, if one can afford it? Target by income level? But what if a lower-middle income country can more or less afford it, but simply needs a one-time nudge?
7. What would be the most efficient way to deal with the fact that this function or facility has to perform two quite separate tasks: a) vet proposals to consider an assessment as reportable, and b) act as a coordinating agent to arrange funding for countries to carry out assessments, make sure that donors are not duplicating efforts, etc. Does it make sense to put these two important functions in one single unit? If so, is there an advantage to tackling one function first? Which would be deemed most urgent? Or which would be deemed most easy to tackle first?

Suggested method of working

- Review of documentation of previous attempts to carry out a similar function, as close to the assessment world as possible, e.g., fundings by the World Bank, USAID, etc.
- Review of successful case studies of a similar function in other areas that are technically demanding and have required institutional innovations outside the mainstream donors, but are not learning assessment, such as agricultural research,

the providing of public goods such as data, institutional innovations in credit provision for small enterprises, etc.

- Key informant interviews of some 30-50 actors in:
 - Leading countries that have moved forward with assessments.
 - Countries one could expect to have taken a lead but have not: what blockages do they face?
 - Countries that have been recipients of past funding but still seem unable to make progress
 - Donor agencies
 - Agencies that have been successfully contracted, outsource, or launched by sister UN institutions such as WHO, FAO, ILO, to provide similar services
 - Specialized agencies that have been successful at solving similar problems such as GAVI, the Global Fund, etc.
 - Think-tanks, particularly in developing countries
 - “Traditional” assessment agencies/providers
 - Innovators in the assessment area
 - Philanthropies

The consultant is expected to lead, or at least be a resource person at one or more policy dialogue sessions where the findings are discussed with:

- The UIS Board
- Important UNESCO member states
- Members of current UNESCO or UN initiatives such as IAEG, TCG (provide links)
- UIS partners in this area such as FCDO, the World Bank, etc.

The consultant can expect to have to draft and re-draft at least two or three times based on these sessions.

It may be an advantage of the consultant is part of, or has access to, an institution or a colleague who is very good at the logistics of such meetings, so as to free both UIS and the consultant from having to arrange such logistics.

Estimated level of effort

It is estimated that the task will require some 30 person days of the senior consultant, and some 10-20 person days of a logistician-assistant.

DRAFT