United Nations
Educational, Scientific and
Cultural Organization

UNESCO
INSTITUTE
FOR
STATISTICS

Comments on “Monitoring the SDGs: an IEA perspective”

Key issues identified

The documents present many things that we agree with and just a few that we do not.

Why measure outcomes? -- We agree that weighing a pig does not make it fatter – the broader context is that we must be advancing good use of the data. The purpose of the data collection should have implications for its design and if data is not used, the resources spent on its collection have almost certainly been wasted. Moving from a focus on monitoring outcomes to action based upon those outcomes is an agenda we support and it should not be lost in GAML discussions.

Quality -- We agree quality is important, but also recognise the realities of the context and suggest that we need to work within a framework of fitness for purpose. Quality is a variable thing. A Lexus, a Camry, a Corolla, and even a Yaris will all get you from point A to point B and even do so reliably and fairly safely, but with widely different comfort levels and power. However, do we need a Lexus, a Corolla, or a Yaris? Therefore, we need a definition of what is good enough, not just what is good. Furthermore, we think the determination of what is of good enough quality needs to be “official” and consensual, and to a significant degree based on client input and a neutral arbiter. The definition should not be driven (only) by the producers, valuable and important as those inputs may be.

Build on Existing Work -- We agree on the need to draw upon existing frameworks and existing learning outcomes studies as much as possible. The work on the UIS RS has done that. The drafting of the GP-LA has done that, and our discussion papers on equating and benchmarking draw heavily on the groundbreaking contributions of IEA and OECD.

Linking Regional and International Assessments -- We agree that linking of existing assessment is a great place to start. We acknowledge that this will be difficult to pull of for all kinds of technical and coordination reasons. It will also take some time to complete. There are, however, a number of examples of where both vertical and horizontal equating have worked well enough to meet the demands of the situation we face – reporting against the SDGs in a way that adds to our current information base.

There would be many benefits to the alignment of international and regional assessments that go beyond the demand of reporting progress toward the SDGs. They would include the improved quality of those assessments, increased uptake and the potential for greater impact.

An immediate constructive way forward is to encourage participation in a regional or international assessment, and also carry out a conceptual linkage using a common scale, whilst also exploring psychometrically rigorous linkages.

It also seems that there are historical experiences in linking assessments (SACMEQ and, if we recall correctly, TIMSS) but also some countries in Latin America appear to be creating links to the international assessments), whose eventual functioning and results we don’t yet fully understand. It would be good to recognize this and also learn more from it.

Use of National Assessments -- We do not accept that national assessments cannot be used. The IAEG_SDG and the Committee for the Chief Statisticians of the United Nations System (CCS-UNS) together with the Committee for the Coordination of Statistical Activities (CCSA) of the UN Statistical Division are recommending to use such data but to ensure quality through a quality assessment framework in addition to national frameworks. This is one of recommendations of the *Principles and practices of global data reporting and data sharing for the 2030 Sustainable Development Agenda* (see UNSYSTEM/2017/2). We do not think it is right and proper to require countries to use specific CNAs. An approach that supports countries in the development of their monitoring capacity, respects national sovereignty and results in fit for purpose reporting against SDG 4 is required.

There are a number of countries that run national monitoring whose quality is comparable to the best CNAs – two obvious examples are NAEP in the US and NAPLAN in Australia.

Whilst for some other countries NAs might not be an ideal solution for reporting, they may be an ideal solution for driving local improvement, and they may be a “good enough” solution for reporting, if certain quality standards are met, and certain comparability criteria are specified, by tying to a common or universal scale. There are also political and political-economic reasons why national assessments may need to be used, but one would have to note whether they are “good enough” for reporting to the SDGs.

A Single UIS RS -- We disagree with the view that multiple scales are to be preferred over a single scale. It is important to not confuse the idea of the UIS RS with a universal assessment. For instance, when the IEA note says: “In TIMSS 1995, IEA administered the same items to grade 3, grade 4, grade 7 and grade 8 students to test this, and our research showed that this was not viable.” And another example: “Reading passages developed for grade 4 students might not be suitable for grade 9 students.” We wonder if a misunderstanding of the UIS RS is exposed in these quotes; whether IEA thinks anyone is proposing a single assessment. We do not believe that the UIS RS requires such an approach.

The first reason for our current preference for a single scale lies in the wide range of performance that we are likely to observe across countries. A scale suitable for the end of primary or the end of secondary would have to span a very wide range to be of value to countries. So it seems to us that developing a single "long" scale is far more efficient than developing three, not quite so long, overlapping (but independent?) scales. Three different scales with substantial overlap would seem to be a conceptual challenge.

The second reason, related to the first, is a view that we should be focusing on progress in student learning and we will only be able to talk about progress across the three SDG reporting points if we have a common scale.

The third reason is that developing a single scale would support some of the points made later in the IEA document (particularly about assessing learning at all levels, not just minimum standards).

Tools for Measuring Progress as a Global Good -- Whilst the tools for measuring progress towards the SDGs cannot be in the public domain in a way that makes it possible to invalidate the monitoring, the product and approaches cannot be a closed shop – GAML is working toward a global good not a proprietary international test. We would argue that the “constructs” (the UIS RS) should be in the public domain as a shared agreed statement of what is valued and a source of meaning for statements like x% of students are proficient.

There must be a potential for, and value in, being open-sourced about *some* things. There is a huge spectrum here, and pushing for being as open as reasonable is both feasible and useful. A combination of the UIS RS and a large item pool may very well be adequate.

Age vs Grade -- The discussion of the pros and cons of grade-based versus age-based is an important one in the TIMSS versus PISA debate, but we do wonder if it is directly relevant here. The SDGs have given us three reporting points; it is their operationalisation that is crucial to moving forward.

Out-of-school Children -- The rallying cry of not measuring out-of-school populations is disappointing. It comes across as elitist and not grounded in reality. While the sentiment of getting all kids into school and getting them learning is a noble intention, it simply is not happening anytime soon. Equity is central to the SDGs and they are there for marginalised populations too!

Threats for the future – We agree with the points made here.

The Core of the IEA Position

The core conclusion of the IEA commentary is that countries must take up an existing program. This is a conclusion that we believe cannot be accepted on a point of principle and that is ratified to some extent by the position of the CCS-UNS/CCSA specially regarding interim reporting. We cannot require countries to take up specific assessment programs -- no matter how good they are. Not all countries will adopt international or regional assessments in the short or medium term, however desirable this might be. In fact, there may be no appropriate CNA available. As an alternative, the UIS is adopting a pragmatic position with an approach that supports countries in the development of their monitoring capacity, respects national sovereignty and results in fit for purpose reporting against SDG 4.

The overall argument of the paper hinges on an assertion that 'the matrices used for measuring these different grade levels cannot be linked on one common scale'. Further on, the IEA suggests, perhaps as a consequence, that a single common scale is not necessary. Is this not an empirical question? There are good examples of scales that have been constructed and used across year levels (we have recent evidence from Afghanistan). Rather than asserting that it cannot be done based upon one attempt, we think we should be asking whether it can be done and, very importantly, done well enough to be useful in the SDG 4 context.

Phase 2 of the process we have proposed would allow testing of their contention that it is not possible to develop a scale across multiple grades. A single or multiple scale option could be pursued contingent on the outcomes.

Finally, as mentioned earlier we would agree that measurement needs to be of high quality, but it does not mean that we must require all students at the same point in their schooling to undertake the same secure tests. First, a single test is not likely to be appropriate because of the range of student performance levels -- we know that, even within countries like Australia, the most advanced students are 5 to 6 years ahead of the least advanced students (even by the end of primary school). Second, this is not in the GAML spirit of developing tools that are for the public good -- the learning progression, global module, item bank and so on.

What we need is a quality assurance process that:

- accommodates the variability in performance levels across (and within) countries,
- respects national sovereignty, meets national needs and is sensitive to cultural values,
- permits a focus on progress in literacy and mathematics outcomes, and
- yields data that are fit for the purpose of SDG reporting