

Concept Note

Measurement of Target 4.1: Technical issues and next steps

5th May 2016

Introduction

The international education community has pledged to “ensure that all girls and boys complete free, equitable and quality primary and secondary education leading to relevant and effective learning outcomes”, as part of Target 4.1 of the new Sustainable Development Goal for Education (SDG 4). The focus on quality of education has led to an emphasis on the measurement of learning outcomes at all levels of education. However, measuring learning is not as straight-forward as learning is a cognitive construct and does not have any physical traits.

There are many similarities in how children develop cognitively across cultures, but bigger environment affects the timing and manifestation of skills. Overall, even though cultural and contextual influences matter, children everywhere show similarities in how they learn to communicate with others, solve math problems, read and write. There are strong points of similarity that can form the basis for global measurement, with some important caveats. National standards for primary and secondary education can be a source of information on local goals for children’s learning development.

As an UN statistical agency focus on comparable data and evidence-based decision making, UIS is providing evidence for developing targeted approaches towards better data collection, and working with global community in defining robust indicators to report progress in learning outcomes towards the 2030 goals.

The priority is to generate data that are comparable across time and disaggregated by age, sex, disability, socioeconomic status, geographical location (urban/rural areas) and other relevant factors¹. However, there are technical challenges in measuring across countries with diverse cultural and education system differences.

Globally-comparable measurement of children learning would be able to track children’s learning over time, so that the learning development can be indexed. This measurement should include contextualize information: home environment and family background, school environment which include school leadership,

¹ See: http://www.un.org/ga/search/view_doc.asp?symbol=A/C.3/69/L.9/Rev.1

teachers' training and competencies, curriculum and learning activities, and opportunity to learn, and these should be sensitive to cultural influences on learning. Furthermore, the global measure should include a range of skills and should place emphasis on the importance of basic skills.

How has learning been measured to-date?

According to UIS estimates, 80% of countries have conducted a national learning assessment or participated in a cross-national initiative in the last five years. Table 1 describes the number of countries per type of assessment.

Table 1: Number of countries by type of assessment

Region	# of countries with assessments			# of countries with no assessments	# of countries with at least one assessment	Total #	% of region
	National	Regional	Cross - national				
East Asia and Pacific	22	14	14	8	32	40	90
Europe and Central Asia	41	0	46	9	28	57	84
Latin America and Caribbean	26	18	17	14	27	41	66
Middle East and North Africa	3	0	2	1	3	4	75
South Asia	8	0	1	0	8	8	100
Sub-Saharan Africa	30	27	3	11	37	48	77
TOTAL	146	59	101	47	173	220	79

Source: Based on UIS Catalogue of Learning Outcomes and World Development Report.

Household-based learning assessments have been used to target populations that may or may not be enrolled in or attend school. This includes citizen-led assessments and any household surveys that include an assessment component in their data collection. Together school-based assessment and household-based learning assessments are providing a snapshot picture of how the children around the world are learning. However, due to difference in construct and framework all these assessed learning are not on comparable level and it is hard to have a global picture of how children around the world are learning. Further to that there are technical challenges in creating an invariant construct that can be used across cultures and regions.

There are common technical challenges in measuring learning which include:

- Establishing cultural invariance, or the applicability of the items across all contexts;
- Establishing predictive validity, or the relevance of the items to measure children skills development;
- Creating feasible tools that can be administered at a reasonable cost while still adequately capturing children’s learning; and
- Ensuring adequate alignment between global tools and specific policy contexts, such as the connection between national curricula standards and items on assessments. As well, data on learning can more easily be interpreted when accompanied by contextual information on family and school environments, which requires additional data collection.

Creating and testing measurement tools requires considerable investment, and also has several tensions:

- First, tools require ongoing modification in response to data from countries on which items work well, which is not only time-intensive but also can prevent use of scales for tracking trends over time, if too many items change.
- Second, establishing cultural invariance, or the relevance of items across contexts, is ideally based on representative samples from several countries and is informed by use of statistical modeling that demonstrates how well the factor structure applies to various settings.
- Third, evidence of predictive validity, or the usefulness of the measure in predicting children’s learning, is critical and requires further research and studies. The precise mix of items that predict learning in one country may not be the same as those in another, ideally leading to multiple sets of items relevant to cultural context from a larger item pool to establish predictive validity. This means to invest in the design, psychometric analyses and tool development.

Beyond these issues, the idea of creating a learning scale for all countries necessitates a new level of technical development. Because development is strongly affected by context, a large range of items will be required for accurate measurement of children in all settings – children may differ in developmental skills by a few years, meaning that floor and ceiling effects will be problematic when scales are used across contexts. By creating learning scales with narrow ranges, data will be more likely to demonstrate that almost all children in some countries are doing well or poorly, which will undercut the overall goal of promoting equity within and between countries. New tools or approaches with wider ranges may be required to capture variation in learning development across contexts.

Challenges in creating a 4.1 measurement strategy

The most central challenge of global measurement of Target 4.1 is deciding what counts as “learning on track” in different contexts, and generating globally comparable tools to describe it. Broadly speaking, “learning on track” is perhaps best understood as the extent to which children are able to master skills and competencies as defined by their local communities. At present, there is no “absolute” definition of on-track learning that applies across contexts.

Questions on what it means to be learning on track include the following:

- By definition, if some children are on track, are others not on track? Are children who are not on track children with disabilities, living in rural area, at low socio-economic status or are girls? If so, how are present efforts to measure children’s learning linked with efforts to measure children in the disadvantage groups?
- The differences in stage of learning skills may be vastly different from one country to the next. Children in high-income countries may develop skills years ahead of those in low-income countries. Because the SDGs are intended to apply to all countries, should measures be equally appropriate for children in all countries, and if so, how can such scales be created?
- Because the overall intent behind Target 4.1 is to index the percentage of children globally who have achieved some minimum standards of learning, questions on predictive validity is central to measurement. Policymakers will likely assume that children who are learning on track will not drop-out and can continue to learn well, but there is a bigger environment to consider. How critical is it to establish predictive validity and collect relevant contextual information?
- Cultural expectations likely have an influence on learning – for example, knowing letters and numbers by a certain age may be more reflective of cultural standards and expectations. How should global measurement appropriately account for cultural influences?

Finally, the political context of measurement means that countries will choose different paths forward for measurement. It may be more acceptable to begin generating technical approaches that allow countries to participate in global monitoring using a tool that is adaptable to different development level of the countries. Some countries may choose to participate in global assessments, while others in development their own national assessments and embed the relevant tool in their assessments to bring their assessment onto the global scale. If the overarching goal of SDG measurement is to encourage collection of data on children learning that can lead to better policies and practices, there is a high priority on generating politically feasible solutions to global monitoring. Therefore, it may be appropriate to generate methods of integrating data across a range of sources that meet international standards.

Next step in measuring Target 4.1

In sum, key points on children's learning must be taken into account when devising strategies for global learning measurement:

- Children's learning proceeds in patterns that are in many ways similar across all populations of children. There may be efficiencies in measurement that are gained through reliance on a common set of items or constructs, based on the degree of similarity in existing measures.
- Aim for construct equivalence, which could provide some degree of population-based tracking using a similar set of constructs, but with adapted items that vary based on culture and context; or develop a common learning scale.
- Data on children learning at grade 2/3, end of primary and end of lower secondary should be accompanied by contextual information, home and school environments, as a more complete picture emerges when taking multiple pieces of information into account.
- A range of analytic strategies can be generated to measure Target 4.1, and a next step is to examine these strategies in greater detail while acknowledging the nature of learning development.

A key policy question is how critical it is to have directly comparable data, and at what cost, conceptually and practically. Reliance on one tool is most efficient for global monitoring, but will also require additional development to ensure that it is useful across all countries, and in particular, engaging stakeholders and organizations, especially the regional assessment bodies, in the dialogue is necessary. Integrating data from multiple sources, school-based and household-based assessments, and administrative data, may be able to efficiently take advantage of existing data and provide a greater degree of cultural relevance.

The UIS Learning Outcomes monitoring strategy² has taken the following approaches by suggesting:

- a global common content framework for reference (GCCFR) for each of the learning domains
- a data quality assurance framework (DQAF) applied to learning outcome data and its implementation tool, the assessment of data process (ADP)

² Implementation strategy for measuring SDG 4 learning outcomes targets concept note in the following link:
<http://www.uis.unesco.org/Education/Documents/UIS%20LO%20Monitoring%20Strategy%20concept%20note.pdf>

- a reporting package
 - Linkage package - Global Linking Package (GLP)
 - Universal Basic Test (UBT) which consist of common set of items adaptable to cultural difference for each domain and point of measurement to be used to link across assessments
 - Universal Background Questionnaire (UBQ) that contain basic contextual information to help put learning into context
 - Global Reporting Metric (GRM) and related benchmark to put data onto the same platform for reporting
 - Guidelines for data analysis and data use

The tools could serve as a starting point for national assessments for countries to put their data onto the global metric. More work is needed to complete technical development. The work of UIS also highlights the need to build community around measurement and provide opportunities for sharing ideas, tools, and ultimately data that will lead to more comprehensive and innovative approaches to measurement. This will require strong commitment of how data apply to policy and practice to improve conditions for children.

Finally, the idea of creating a global scale of learning and reporting metric is ambitious. The creation of such a scale and metric would underscore children learning on trajectories. To create scales that would work across populations, we need to track the learning of children in a range of countries, and used that point of comparison to generate a global curve.

It is clear that the dialogue should continue among stakeholders and experts on approaches to measurement for 4.1. Next steps could include the following:

- 1) Create technical and institutional homes for global technical convening assisting countries in implementation and use of data to improve policy and practice.
- 2) Convene stakeholders and experts to define “learning on track;” and evaluate the alignment of existing tools with the desired definition and scope of learning.
- 3) Define a common set of technical standards for tools used for population-based measurement to inform global monitoring.
- 4) Explore the process to the creation of a learning scale as a benchmark for children learning across cultures.
- 5) Develop technical approaches to integrate contextual information to create projections of the percent of children learning on track, and compare to estimates from existing tools.

- 6) Develop psychometric methods for integrating national and regional data to generate estimates of children's learning.

Overall, a challenge in “learning on track” is the comparative and contextual nature of the term and the difficulty in defining “on track” globally, for which there is no internationally-agreed upon definition. The dialogue on measurement thus has implications for the implementation of strategies to reach Target 4.1, and therefore is an important part of the overall SDG agenda for children learning.