

Key Questions on the Domains of Measurement for SDG 4.2.1

Recommendations from GAML Task Force 4.2

Kate Anderson and Abbie Raikes¹

Summary

SDG target 4.2.1 reads: “Proportion of children under 5 years of age who are developmentally on track in health, learning and psychosocial well-being, by sex.” The global indicator data source currently used is the UNICEF Multiple Indicator Cluster Survey (MICS) Early Child Development Index (ECDI). It is listed as a “Tier III” indicator by the Inter-Agency and Expert Group on the Sustainable Development Goal Indicators (IAEG-SDGs) as work is needed to develop other data sources and a common methodology required for international comparability that is applicable in all countries. The definition of “developmentally on track” has not yet been globally agreed, and there are no measures that are equally feasible for use in high- and low- income countries. This paper aims to unpack some of the issues associated with measurement of 4.2.1; to provide discussion questions to develop a common definition of “developmentally on track”; to propose measurement options for the overarching domains of health, learning and psychosocial well-being, and to summarize the feedback on these options from the 4.2 task force via a virtual meeting and written feedback.

Choosing domains

One advantage of internationally-comparable data is the ability to produce estimates of equity in early childhood development across countries. Yet because the domains of development are not equally feasible to measure in an internationally-comparable manner, the need for internationally-comparable data will likely make it easier to measure some domains over others. However, the task force members felt strongly that the domains of health, learning, and psychosocial well-being should all be considered in a global measurement framework. The group discussed the need to develop a common set of criteria for cross-country comparability, assess existing measures against these criteria, and choose sub-domains in each of the three domains, understanding that some sub-domains will be globally comparable and some will not.

Defining “developmentally on track”

To define developmentally on track, three routes were proposed to the task force: relying on national standards; exploring the feasibility of a global scale; or leaving it undefined and moving forward on measurement of specific domains to produce an empirically-based definition of “on track.” The task force members agreed that a hybrid approach of using national standards to develop a global scale would be a good way forward.

This paper also proposes two methods for estimating which children on track vs. not, an absolute definition that would outline a specific set of skills that all children should reach by a

¹ Authors listed in alphabetical order. This paper draws on text and ideas articulated in Raikes, 2016, <http://unesdoc.unesco.org/images/0024/002455/245579e.pdf>.

given age, or a relativistic definition that would designate some children “off track” in relation to an overall population (i.e., those children 1 or 2 standard deviations below the mean). The task force members felt that this was a topic that required further discussion, potentially at an in-person meeting to reach technical consensus.

Next Steps

Based on the discussion and feedback from task force members, the following next steps are proposed:

1. Draft a measurement framework for 4.2, which covers all three domains and proposed sub-domains.
2. Establish criteria for cross-country comparability and evaluate existing assessments against these criteria (potentially drawing criteria from other GAML task forces)
3. Map existing national early learning standards to determine if there is agreement on “developmentally on track,” drawing from UNICEF study if available.
4. Convening in-person meeting of task force and other key technical resource people to reach technical consensus and finalize the measurement framework.

The followings sections of this paper provide a background and current information on the key issues discussed by the GAML 4.2 task force.

Discussion 1: How do we determine domains that could be considered internationally comparable?

The SDG target language identifies health, learning, and psychosocial well-being as key domains in determining readiness for primary school. Within each of these broad domains, a smaller subset of domains can be selected for global monitoring based on feasibility and desirability. There are a few key considerations to keep in mind when evaluating the extent to which the domains may be considered globally comparable.

First, developmental science points strongly towards a holistic view of early childhood development because early development is interconnected, with many skills supporting development across domains. This means that multiple domains are necessary to describe children’s learning and development, regardless of comparability.

Second, some domains are more easily indexed than others. For some domains, internationally-comparable data may be easier to reliably achieve across countries because children typically follow a predictable pattern of progressively more complex development. However, the harder-to-index skills may be some of the most critical to measure (i.e., social/emotional development), and the least comparable across contexts.

Third, nearly all major assessments of child development include multiple domains, with different names and/or the same items, but often assigned to different domains. Assessment

of comparability thus should include careful examination of constructs and items as well as domains.

Finally, there has not yet been a systematic approach to determining standards for testing international comparability in early childhood. While there are certainly standards that can be applied from primary school learning measurement, the unique nature of early childhood development means that a specific set of standards should be developed and applied, before determining whether existing data point towards comparability or lack thereof in domains.

Despite this complexity, below we focus on aspects of child development measurement that help provide insight into what's feasible and desirable.

Feasibility and desirability of measurement in three domains

The map of ECD measurement tools (Annex A) presents a selection of tools administered in more than one country, with an emphasis on tools used at the population level and those developed for low- and middle-income countries (LMICs), organized by domain. The feasibility and desirability of measuring health, learning, and psychosocial well-being are described below.

Health

Health status is clearly an important part of child development, especially in countries with high rates of undernutrition.

Feasibility: Data on malnutrition, HIV status, and other infectious diseases is available now through UNICEF and WHO. Child health is covered in other SDG targets, namely in target 2.2 on prevalence of stunting, wasting, and overweight among children under 5, and in target 3.3 on under 5 mortality. ECD measurement tools that cover health status may include general health questions such as children's overall health status or immunization history, and also have included anthropometric data on height-for-weight.

It is important to note that child physical development can be considered part of health status, but also should be considered part of the cognition/learning domain since some aspects of physical development, such as fine motor development, are also indicators of executive function and overall neurocognition.

Desirability: Because indicators on malnutrition and disease prevalence are covered elsewhere in the SDG indicator framework, the focus on health in 4.2 is in the context of readiness to enter primary school. Children's undernutrition and evidence of stunting is likely to have a profound influence on learning, and therefore should be included in overall assessment of health. Likewise, HIV status and exposure to other infectious diseases like malaria and diarrhea are also associated with delayed neurocognitive development, so may be important to index as part of an overall picture of school readiness at a population level. Therefore, a brief measure of health status may be supplemented with the indicators found in other SDG targets. Beyond health status, fine and gross motor development could also be considered in this domain.

Learning

Learning outcomes for children under age 5 are generally clustered in domain-specific areas such as language/pre-literacy and pre-numeracy, and domain-general areas such as approaches to learning and executive function.

Feasibility: Nearly all ECD measurement tools include items on early literacy and numeracy. In addition to these two domains, executive function tasks are being included increasingly in assessments of young children. Results from four countries using the MELQO tools indicated that early math development and executive function may be more readily measured in an internationally-comparable manner than literacy or social/emotional development.² These developments could be linked the UIS Reporting Scale developed for primary and lower secondary, which could be expanded to ECD.

Desirability: Academic subjects such as literacy and numeracy are assessed beginning in primary school, including in SDG target 4.1.1 and 4.1.2, so extending measurement down to include pre-literacy and pre-numeracy in early childhood may make sense, although early literacy skills may not be as easily compared across countries as early mathematics skills. Executive function also predicts later learning across academic subjects, and should be considered here.

Psychosocial well-being

Psychosocial well-being includes social and emotional development, as well as the absence of mental health disorders (which requires a combination of measurement to capture the range of normative and non-normative development).

Feasibility: Nearly all ECD measurement tools include items on social-emotional development. However, social/emotional development emerges in a manner that is culturally sensitive, and therefore measurement in this domain may not lend itself to international comparability. Parent and caregiver reporting captures a wider range of social/emotional development than direct child assessment in this domain. Work is needed to arrive upon one set of items with equal relevance across countries.

Desirability: Many countries have prioritized social/emotional development as a critical piece of children's readiness for school. Children's development is interconnected, meaning that while some domains may be easier to measure across countries, the risk of leaving out social-emotional development is an incomplete view of child development. The overall desirability of internationally-comparable data is perhaps a question if it means that only some domains of development would be measured. This

² Raikes et al., 2017, under review, "Measuring early childhood development cross-nationally." Questions regarding comparability of domains outlined in UNESCO, UNICEF, Brookings Institution, World Bank (2017). *MELQO Overview*.

could send a political message that some domains of child development are more critical than others, leading to unintended consequences of SDG monitoring.

Issues in globally comparable measurement

There are many elements of child development that follow a neurobiologically-driven, universal pathway. General domains like cognitive development/learning, social/emotional development and health all have elements that are universally relevant. An immediate step is to decide upon standards for international comparability in early childhood data, and to assess existing data sources against these standards.

There are potential tensions between feasibility and precision. For population-wide measurement of children who have not yet started formal schooling, household surveys are the most viable form of data collection for capturing all children, if the overall goal is international comparability. Direct, oral assessment of children will yield the most accurate information on specific aspects of children's skills and knowledge, but requires trained observers.

Cost is another consideration. Household surveys are typically more expensive than center- or school-based assessments because it's necessary to sample and visit individual households, and less travel time when a group of children are in one location. Direct assessments of learning and development typically involve a longer process of reliability training and can take more time to administer than parent or caregiver surveys. Finally, some assessments used internationally have licensing fees or require countries to pay for training by the test developers, which can make some internationally-developed assessments more expensive than locally-developed ones. However, there are an increasing number of freely available assessments available to countries for the early years. At the same time, it is critical to note that even "free" assessments require ongoing investments in technical development and staffing to maintain the quality of the assessment materials. These costs must be covered, if not by countries then by other organizations or sources.

It is important to note that what is feasible to measure in an internationally-comparable manner across all domains is likely quite general, and may not be specific enough to drive national-level policymaking. However, this can still yield useful data on the overall status of children's development. Internationally-comparable data has the advantage of spurring global action in ways that country or regional data is not able to do. At the same time, to achieve an internationally-comparable tool, a certain degree of generality is likely required, meaning that only a very rough overview of child development may be possible. The value of those data for policymaking should be evaluated.

Country and regional data, on the other hand, may be able to provide a more nuanced look at child development, by allowing the inclusion of nationally and regionally relevant constructs and items. However, to achieve a global picture of equity in child development, it will be necessary to create a method for integrating the measures to generate estimates of child

development across countries – and high-income countries may end up on a different scale than low-income countries.

Options for choosing domains

Theoretically, all domains of child development could be measured in an internationally-comparable way. Human development has much in common across cultural settings, so there is no reason why one domain should be excluded from consideration. However, some domains have been assessed more thoroughly to date than others for this age group, resulting in differences in experience and feasibility in measuring in some domains vs. others. Based on existing data and the balance with feasibility and desirability, the task force discussed three options for consideration:

Option 1: Measure all three. Agree to examine all three domains, and ask for input from those who have developed regional or international assessments to more closely examine the domains and constructs that may be most workable. Most of the assessments in Annex A contain some items on health, learning, and psychosocial well-being.

Option 2: Measure what is most globally comparable. While more research is needed, there is some evidence³ that early math and cognition may show the greatest degree of comparability. Most of the assessments in Annex A contain items on numeracy and early mathematics. Executive function is included in a small number of cross-national assessments.

Option 3: Measure what will be measured in subsequent levels of education. Focus on a downward extension from Target 4.1.1f., which would place focus on language and literacy, and begin working on a universal approach for language/literacy. Language is included in nearly all of the assessments in Annex A, and pre-literacy is included in some.

After discussion on a virtual call and through written feedback submitted by task force members, it became clear that the first option, measuring all three domains, had the most support. Given the interconnected nature of children’s development and the availability of data in all domains, the task force members felt strongly that none of the domains be excluded from measurement, even if some are not yet globally comparable.

Discussion 2: How should we define “developmentally on track”?

There are presently no agreed-upon definitions of “developmentally on track” that are specific enough to guide internationally-comparable, regional or national measurement. Conceptually, identifying some children as “developmentally on track” implies that other children are not “developmentally on track” simply by the nature of the statement, which is articulated as a binary option (either “on” or “off” track). One question that has not been addressed is whether “off track” means that children have disabilities, as there may be a range of reasons why children are not on track, including but not limited to disabilities or special needs. Therefore, it may be necessary to view this population of children as having some overlap with population of

³ See Raikes et al., 2017

children with disabilities. In either case, children who are on track could be defined relatively, meaning by designating children on track if they are no more than two standard deviations below the population mean; or in absolute terms, meaning that children need to achieve a specific set of skills that define “on track.”

For either conceptualizations, a first step is to generate a way of defining the content of “on track” – meaning the mix of skills and competencies that then could be measured in either an absolute or relativistic manner. Because this construct is at the heart of all measurement of 4.2.1., below we outline options that the task force discussed to address “developmentally on track”:

Option 1: Rely on national standards. Many countries have gone through the process of developing early learning development standards (ELDS) or other types of standards that include children’s development. These standards are holistic in nature, and are intended to inform measurement by outlining consensus on what children should be able to do at certain ages. To date, we are not aware of a careful review of the content of these standards, and how they could be used to generate an international child development measure. It is possible that a thorough review of national standards can be done to determine whether there are common definitions of developmentally on-track within the selected domains, and use a global definition that is most common across countries.

It is also possible to *not* define “developmentally on track” at the global level, and instead, allow each country to generate its own definition based on standards. This approach runs the risk of perpetuating inequity because the quality of the standards, and the extent to which the standards are developmentally appropriate, may vary considerably by country. Another option is to have both an international standard for reporting and country-defined standards, similar to the international and national standards for the poverty line. Poverty estimates at the national level reflect the line at which individuals’ needs for housing, nutrition and clothing cannot be met within that country. To generate a globally-comparable estimate, the purchasing power parity estimate is calculated, based on a common global currency scale⁴. The applicability of this approach to early childhood development could be explored as a path towards synchronizing national-level and globally-relevant data.

Option 2: Invest in the creation of a global scale. The World Health Organization invested in the development of growth scales that have had a profound impact on attention to malnutrition. WHO tracked the growth of the highest-income children in a set of countries, based on the assumption that these children would experience optimal conditions for growth. WHO is now exploring this possibility to develop a measure for children birth to age three years. A similar approach could be explored for older children as well. A first step would be a

⁴ <http://www.worldbank.org/en/topic/poverty/brief/global-poverty-line-faq>

careful examination of the pros and cons of the feasibility and desirability of this approach, including costs and expected benefits.

Option 3: Leave undefined. Assume that “developmentally on track” is useful as a conceptual model, but that it cannot be precisely quantified and therefore will not be measured anytime soon. This would require detailed and nuanced communication on any measurement for Target 4.2.1, but before a strategy can be developed that multiple stakeholders find workable, one option is to move forward on measurement and leave the definition of “developmentally on track” unaddressed for the time being. Over time, it could be informed empirically, by using existing data to more fully define a cross-nationally relevant definition.

Table 1. Options for defining developmentally on track

Method of Comparison	National Standards	Creation of Global Scale	Leave undefined
Absolute	% children reaching agreed-upon set of skills/competencies, using national standards as starting point	Set of skills defined by experts, but no “absolute” threshold because would be structured as a relative scale	Up to countries to define standard set of skills to measure against, could look across countries over time to identify points in common
Relative	Using national, regional or global population, define “not on track” as 1 or 2 SDs below the mean. With this option, the definition of “on track” would be empirically defined according to a set of items agreed upon using national standards.	% of children reaching development on par with most advantaged children in the world, with set of skills defined by scientists/experts	Create integrated scale that would combine data from range of country-level measures; could then use to create global profile

The task force members proposed a hybrid approach between options 1 and 2, where national standards are reviewed and used to develop a global definition of developmentally on track and a possible global scale. One task force member suggested conducting a review of national Early Learning and Development Standards (ELDS) across all countries where they are available, and

determining where there is agreement to see if setting global standards is possible. The task force learned that UNICEF has done a content analysis of the ELDS in 35 countries which will be made available in May 2017.

To determine whether the method of comparison should be relative or absolute, the task force members felt that this should come later, potentially through an in-person meeting of task force members and other technical experts with the goal of reaching technical consensus.

Annex A. ECD Measurement Tools

Tools in this table measure child development and learning in children age 5 years and have been tested in more than one country. For the purposes of this table, gross and fine motor skills are classified under “health” but could also be considered in “learning.”

Tool	Type of administration	Health	Learning	Psychosocial well-being	Tested in high-middle-, or low-income countries?
East Asia Pacific Child Development Scales (EAP-CDS)	Direct assessment	Motor development, Health, hygiene, and safety	Cognitive development, Language and emergent literacy, Cultural Knowledge and Participation, Approaches to Learning	Socio-emotional development	Middle-income
Early Development Instrument	Teacher survey	Physical health and well-being (includes gross and fine motor skills)	Language and Cognitive Development, Communication skills and general knowledge	Social competence, Emotional maturity	High- and middle-income
Early Human Capability Index (eHCI)	Direct child assessment	Physical health and well-being	Approaches to learning, Language and Cognitive, Communication Skills, Perseverance	Social competence, Cultural identity/spirituality, Emotional maturity	Middle-income
International Development and Early Learning Assessment (IDELA)	Direct child assessment and caregiver survey	Gross and fine motor development	Emergent literacy and language, Emergent numeracy, Approaches to learning, Executive functioning	Socio-emotional development	Low- and middle-income

MICS Early Child Development Index (ECDI)	Parent survey	Health status Fine motor	Learning approaches, Literacy, Numeracy, Fine motor	Social/emotional	Low- and middle-income
MELQO Measure of Development and Early Learning (MODEL)	Direct assessment or parent or caregiver survey	Health status	Language/literacy, Numeracy/math, Executive function	Social/emotional	Low- and middle-income
Regional Project on Child Development Indicators (PRIDI)	Direct child assessment	Motor skills	Cognition, Language and communication	Social-emotional development	Middle-income
UNICEF WCARO Early Learning Assessment (ELA) of Primary Education Entrants	Direct child assessment and group assessment		Cognitive development, Language	Socialization	Low- and middle-income
Strengths and Difficulties Questionnaire	Parent survey			Social/emotional and behavior problems	Low-, middle- and high-income