

Concept Note

Measurement of Target 4.2: Technical issues and next steps

Measuring early childhood development globally and with validity is doable but complex. This note makes the case that global technical convening and operational support to countries will be required to generate reliable measurement of Target 4.2. Specifically, work is needed to clarify what “developmentally on track” means in diverse contexts, beginning at birth and extending through the early childhood years; generate options for global measurement, including technical solutions for integrating national and regional data into global monitoring; and provide opportunities for sharing information among designers and implementers of early childhood development assessments. Above all, measurement should be viewed as a means to an end, with priority placed on how the data will be used to affect policy and programmatic changes on behalf of young children. This note offers reflections on the present status of early childhood measurement and next steps.

The inclusion of early childhood development (ECD) in the SDGs under the education goal signals growing understanding that children’s early development has a profound impact on learning throughout the rest of life, and thus has notable implications for the design and implementation of effective school systems. Target 4.2 is intended to capture children’s development prior to the start of formal schooling, with global measurement indexing the percent of children “developmentally on track.” At present, there are major challenges to meeting this target. Access to pre-primary education varies dramatically across and within countries. Rates of malnutrition remain very high in some low-income countries, and risks to development are also apparent in high-income countries with uneven quality of care and high rates of child poverty. It is impossible to know the full scale of this challenge as few countries consistently measure early childhood development and learning at the start of school, despite the importance for informing curricula; building stronger partnerships between health, nutrition, and education; and tracking progress in reducing inequity throughout the school years.

Global monitoring of 4.2 is intended to include both high and low-income countries. It is important to note that the targets were designed through an inclusive, political process, and the ability to measure the targets was part of the process but did not drive decision-making on the final target language. The MICS ECDI has been identified as the primary indicator of Target 4.2 for global monitoring, and has been used in multiple low and middle-income countries to date. With the increased attention on ECD, many national governments and regional entities have also invested in ECD measurement. The challenge now is to create a workable global strategy for ECD measurement that

balances the need for globally-comparable data with national relevance, while adequately handling the complexity of cultural and contextual influences on child development. The nature of early development presents unique challenges for globally comparable measurement, primarily due to the cultural and contextual variation in child development and the lack of clarity on which aspects of development are most critical to measure. There are several ways that global measurement of ECD could move forward, including 1) reliance on one existing tool (the MICS ECDI); 2) integration of data from multiple tools, especially national and regional assessments in use now; 3) identifying constructs rather than items to serve as the basis for comparison; or 4) most ambitious, working to gain consensus on normed, global scale of child development for children birth to age 8 that would serve as a common scale for all children.

This note covers four key areas of discussion on measurement of Target 4.2: 1) what does developmental science tell us about measuring 4.2; 2) what tools and approaches are available now; 3) what are unique challenges related to measurement of early childhood development; and 4) strategies for moving forward, outlining the advantages and disadvantages to the four options above.

1) What does developmental science tell us about measuring 4.2?

A short overview of the nature of development helps lay the groundwork for understanding measurement options. Children's development is driven by a combination of unfolding biological processes and responsiveness to environmental and cultural cues. Starting at conception, environmental influences interact with genetic information to design the brain's architecture. This architecture sets children on trajectories of optimal or sub-optimal development that become increasingly difficult to modify as children grow. Scientific evidence has clearly documented several influences on children's early neurological development, including health status (e.g., exposure to infectious disease); nutrition; and early emotional support and stimulation. These influences work together in early childhood to influence learning and development.

In young children, normative development is perhaps most easily defined by the extent to which children are developing skills and competencies that are on par with other young children. Some skills are more strongly influenced by biological and neurological patterns than others – for example, in a range of cultures, children learn to stand at about the same time. Others are more strongly influenced by environments, such as language development. Cultural expectations and practices affect children's development by dictating the types of interactions children have with caregivers, the reinforcements children receive for their behaviors, and their exposure to environmental stimulation.

At the highest level, "developmentally on track" at the start of school means that children are developing the skills and competencies that will allow them to participate successfully in their environments and reach their developmental potential, both at present and by building the groundwork for lifelong development. While the idea of

developmentally on track is intuitive to many parents, teachers, caregivers, and policymakers, the nature of child development also presents complexities that are critical to understand when deciding how best to measure.

Five themes with significance for measurement include the following:

Development proceeds in trajectories that begin at the moment of conception.

The extremely fast rate of neurological development in the early years of life designs the brain's architecture, which becomes increasingly set over time. To identify and support children in reaching developmental goals by school entry, measurement should ideally start at birth, with attention to children's rate of growth as well as their starting and ending points. With age, children will acquire more skills – even in the most difficult of circumstances. There is a common set of cognitive characteristics that all humans eventually attain, such as the ability to think abstractly, that may be best understood as biologically driven. But the timing and the depth of these skills will depend on environmental influences.

Children's development is inter-connected across domains, and non-linear.

Social/emotional, language and cognitive development all work together to affect young children's learning. Some studies have found that early mathematics, for example, is a better predictor of later reading than early literacy. This means that measures should include all domains of development rather than focusing on one area or another, and should test how the domains fit together, rather than assuming there are unique domains that are not inter-related. This mix of skills and the non-linear nature of development means that some areas of development, such as math, may be accurately indexed using traditional methods of assessment development like IRT, and others, like executive function, may require use of normed scales that are focused on describing the average age at which certain skills are achieved. *This is an important difference between ECD measurement and later academic skills*, which has enormous implications for measurement. For example, early math skills may unfold in a more or less linear fashion, with graduated steps that proceed in a similar order among children in many places. But the developmental progression of executive function is likely not as linear and may move in stages, meaning that children acquire new insights not gradually but as one large step forward (e.g., the ability to understand that people do not necessarily share perspectives, which typically emerges between three and four years of age). These diverse patterns of development mean that using traditional psychometric approaches may be useful in some, but not all domains of development, and that the age at which children acquire skills may be central to defining what is on track vs. not.

There are many similarities in how children develop across cultures, and cultural expectations also affect the timing and manifestation of skills, likely in some areas of measurement more than others. Overall, even though cultural and contextual influences matter, children everywhere show similarities in how they learn to control their behavior, 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. Measurement of social/emotional skills, for example, is probably more likely to be influenced by cultural expectations on how best to express emotions and handle conflict than mathematics, which has elements that appear to be culturally-neutral. National standards for pre-primary education can also be considered a source of information on local goals for children's development. Areas that are culturally-influenced are just as important to measure, but may require more locally-relevant approaches to measurement to ensure that children's skills are being indexed as accurately as possible.

Family, health and nutrition have a profound influence on development.

Health status, nutrition history (especially whether a child was stunted prior to age two years), and family environments are extremely critical to understand when assessing whether groups of children are "on track." Measurement of early childhood development thus has connections to measurement for several other SDG goals. Children who are stunted, for example, may acquire fewer skills at the start of formal schooling but may also experience a sharper rate of growth when given the opportunity to attend preschool. These children may be more "on track" than children whose absolute scores are higher, but whose rate of growth has slowed. *Nutrition status also deserves a special note:* Some policy makers have suggested using stunting as a proxy for developmental status. While stunting provides very important information, it does not measure developmental status fully. Some children who are not stunted will not be developmentally on track, and some children who are stunted will – these two indicators should be used together, but are not interchangeable.

Looking across these themes, globally-comparable measurement of early childhood development would ideally be able to track children's growth over time, so that the rate of development can be indexed; should include early health and nutrition history; should be sensitive to cultural influences on development; should include a range of skills; and should place emphasis on the importance of early skills for later development. This is a long-term agenda for early childhood measurement. The next section addresses measures used to date, and how these issues have been addressed.

2) How has early childhood development been measured to date?

The idea of using one globally-comparable approach to measure ECD in all countries, rather than focusing on a region or grouping of countries (such as high or low-income) is new. But this new question can be informed by a long history of ECD measurement. For decades, researchers and clinicians in a range of countries have developed and used measures of early childhood development, leading to a strong literature of the psychometric properties and validity of ECD measurement, typically using standardized scales that were normed for use in high-income samples.

In recent years, attention has turned to development of regionally or globally-comparable population-based measures of early childhood. Many of the items and

constructs documented previously are now used to generate population-based estimates, and there are considerable points of commonality between many tools of ECD. There are now several measures of early childhood development used across more than one country and at the population level. A list of some of these tools and their basic characteristics appears below. All of the tools listed here are designed to capture children’s development in the late preschool years using a combination of math, literacy, language, social/emotional and motor development items:

Tool	Region	Purpose	Method of Administration
Early Development Instrument (McMaster University)	Originally developed for use in Canada; has been adapted and used in representative samples in several countries	Population-level measurement of children’s development at start of school, for children 4 to 6 years	Teacher report
East Asia Pacific Child Development Scales (UNICEF)	East Asia region; used in mostly representative samples in 9 countries to date	National level and regionally-comparable data on children’s development between 3 and 6 years	Direct assessment; short form of scale now developed and ready for use
IDELA (Save the Children)	Global tool; used in at least 30 countries to date	Program and national-level data on children’s development between 3 and 6 years	Direct assessment
MICS Early Childhood Development Index (UNICEF)	Global tool; used in representative samples in at least 50 countries to date	Globally-comparable and national-level data on children’s development between 3 and 4 years, 11 months	Parent report through household survey
Measuring Early Learning Quality & Outcomes	Designed for use as a global “core” to integrate into existing tools and national-level assessments	Globally-comparable and national-level data on children’s development between 4 and 6 years	Direct assessment, teacher or parent survey
PRIDI (Inter-	Latin America	Regional and	Direct assessment;

American Development Bank)	region; used in four countries to date	national-level data on early childhood development and household contexts	parent survey
West and Central Africa Regional Office Regional Prototype (UNICEF)	West Africa; used in representative samples in 8 countries to date	National-level and regionally-comparable data on children's development in the first year of school (average age of 6 years)	Direct assessment of children through groups and individual assessments in schools

Common technical challenges in measuring early childhood development include 1) establishing cultural invariance, or the applicability of the items across all contexts; 2) establishing predictive validity, or the relevance of the items to children's long-term development; 3) creating feasible tools that can be administered at a reasonable cost while still adequately capturing children's development; and 4) 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 early childhood development can more easily be interpreted when accompanied by information on family environments, which requires data collection from families.

There are pros and cons to each type of assessment. Direct assessment is sometimes considered the most objective way to capture information on children's development, but may not be feasible unless children are enrolled in schools unless it is within a household survey, and isn't able to capture as much information children's social/emotional development. Parents may not be as accurate on specific details of children's development as direct observers, but have the most depth and breadth of knowledge, and therefore offer different information than can be captured in a direct assessment; teachers are good reporters of children's behavior in schools and therefore may be well-suited to predict which children will succeed over time, but only if they have the chance to get to know each child.

Creating and testing these tools requires considerable investment, and also has several tensions worth noting. 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. Beyond the need for considerable and ongoing technical development, some instruments are designed to be short and therefore may not have enough items to conduct full modeling. Third, evidence of predictive validity, or the usefulness of the measure in predicting children's development in later years, is critical for good

measurement of ECD and requires longitudinal studies. Predictive validity has not yet been generated for many of the tools listed above. This lack of evidence doesn't mean that the tools are not predictive – for example, the Early Development Instrument, one of the longer-established tools, has strong evidence of predictive validity. It is likely that many of the other tools would demonstrate validity as well, since the content is largely based on existing research on children's development, but investment is required to generate the evidence. Moreover, the precise mix of items that predict later development in one country may not be the same as those in another, ideally leading to multiple longitudinal studies to establish predictive validity. In sum, the time and resources required to create tools means that many are developed using limited budgets and within short time-frames, leading to limited ability to invest in psychometric analyses and tool development.

Beyond these issues, the idea of creating a workable scale for children in 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. Some children in all countries face risks to development; by creating 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 development across contexts.

Finally, some of the tools have been designed to be more closely adapted to national standards than others. The question of how much adaptation is required is closely related to the question of how the data will be used: if the data are intended to inform national policy, it is ideal if items are very tightly connected to national standards. But if data are intended to provide a regional or global snapshot of children's learning and not necessarily to be integrated into national systems, less adaptation may be required, as long as the items show cultural relevance to children's development.

3) Challenges in creating a 4.2 measurement strategy

Perhaps the most central challenge of global measurement of Target 4.2 is deciding what counts as “developmentally on track” in different contexts, and generating globally comparable tools to describe it. Broadly speaking, “developmentally 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 early childhood development that applies across contexts and has been agreed upon by stakeholders.

Questions on what it means to be developmentally 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? If so, how are present efforts to

measure children's development linked with efforts to measure children with disabilities?

- The differences in timing of developmental 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 4.2 is to index the percentage of children globally who are well-prepared for schooling, questions on predictive validity becomes central to measurement. Policymakers will likely assume that children who are on track will develop more optimally over time. How critical is it to establish predictive validity, and what is the most efficient path for doing so?
- Cultural expectations likely have a stronger influence on some elements of children's development than others – for example, because of the tight link to neurological development, executive functions might be more relevant to all children, while knowing letters and numbers by a certain age may be more reflective of cultural standards and expectations, but are nonetheless important indicators. How should global measurement appropriately account for cultural influences?
- Given that science clearly indicates that nutrition status is integral to development, should data on stunting be integrated into measurement of developmentally on track, and if so, how?

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 while using a wide range of tools. Some countries may choose to participate in global surveys, while others are investing in development of national assessments and may prefer to use more locally-relevant tools. Regional efforts like WCARO, PRIDI and EAP-CDS have also gained traction in some regions, and now are interested in working with other regional assessments to expand their work. The OECD is also in the process of developing an early childhood development assessment. If the overarching goal of SDG measurement is to encourage collection of data on early childhood development that can lead to better policies and practices, there is a high priority on generating politically feasible solutions to global monitoring. Therefore, rather than suggesting one scale, it may be most appropriate to generate methods of integrating data across a range of sources, when countries can use their own technically-sound and contextualized approaches that still meet international standards.

4) Where to go from here? Next steps in ECD measurement

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

- Children's development proceeds in patterns that are in many ways relevant 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.
- While these patterns are consistent, it is not yet clear if there is enough commonality in development to propose the same set of items used for children everywhere; aim for construct equivalence, which could provide some degree of population-based tracking using a similar set of constructs, but with items that vary based on culture and context; or develop a common child development scale.
- Data on child development at the start of school should be accompanied by information on children's health and nutrition status, home environments, and access to early childhood development programs, 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 4.2, and a next step is to examine these strategies in greater detail while acknowledging the nature of early development.

A key policy question is how critical it is to have directly comparable data on ECD, and at what cost, conceptually and practically. Reliance on one tool may be most efficient for global monitoring, but will also require additional development to ensure that it is useful across all countries, and in particular, engaging organizations working with high-income countries in the dialogue is necessary. Integrating data from multiple sources may be able to efficiently take advantage of existing data and provide a greater degree of cultural relevance, and has the added advantage of potentially being able to integrate data on health and nutrition. Comparing at the construct level will also provide advantages for ensuring cultural and contextual alignment.

The Measuring Early Learning and Quality Outcomes project (MELQO) has taken these two latter approaches by suggesting a common set of items that could serve as a starting point for national assessments and/or integration into other tools, and would also be globally comparable. This approach appears promising and more work is needed to complete technical development. The first phase of tools will be shared this summer in an open-source format. The work of MELQO also underscores 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. Reaching scale with ECD measurement will require strong commitment to creating open-source tools with support for training, documentation of countries' experiences with the tools, and examples of how data have been applied to policy and practice to improve conditions for young children.

Finally, the idea of creating a global scale of child development and learning is ambitious, but may be worth exploring. The creation of such a scale would link children's development starting at birth with development through the early school years, and perhaps most critically, would underscore that children develop on trajectories. One notable example is the WHO growth curves, which were developed to identify children with stunting. To create scales that would workable across populations, WHO tracked the growth of children in high-income households in a range of countries, and used that point of comparison to generate a global curve. The World Health Organization would be an ideal partner for pursuing this path.

Across all of the options, it is clear that the dialogue should continue among ECD stakeholders and experts on approaches to measurement for 4.2. 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 “developmentally on track;” analyze the various options for measurement listed above; and evaluate the alignment of existing tools with the desired definition and scope of developmentally on track.
- 3) Define a common set of technical standards for tools used for population-based measurement of ECD, including standards for national and regional assessments that could be used to inform global monitoring.
- 4) Explore the process, advantages and disadvantages to the creation of a normative ECD scale as a benchmark for healthy child development and learning across cultures.
- 5) Develop technical approaches to integrate stunting, health and family background information to create projections of the percent of children developmentally 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 development.

Overall, a challenge in the phrasing “developmentally on track” is the comparative and contextual nature of the term and the difficulty in defining “on track” development globally, for which there is no internationally-agreed upon definition. Beyond measurement, clear guidance to countries on the nature of early development will have important implications for policy and programmatic development. The dialogue on measurement thus has implications for the implementation of strategies to reach Target 4.2, and therefore is an important part of the overall SDG agenda for early childhood development.