



Background paper prepared for the 2016 Global Education Monitoring Report

Education for people and planet: Creating sustainable futures for all

Measuring of child development and learning

This paper was commissioned by the Global Education Monitoring Report as background information to assist in drafting the 2016 report. It has not been edited by the team. The views and opinions expressed in this paper are those of the author(s) and should not be attributed to the Global Education Monitoring Report or to UNESCO. The papers can be cited with the following reference: "Paper commissioned for the Global Education Monitoring Report 2016, Education for All 2000-2015: achievements and challenges" For further information, please contact gemreport@unesco.org

I. Abstract

The proposed education goal for the post-2015 sustainable development agenda includes a target focused on ensuring children begin formal schooling developmentally on track and “ready for primary school.” What does it mean for a child at school entry to be developmentally on track, and how should it be measured? This paper reviews issues and approaches to measurement of young children’s development in light of the education target on early childhood development. Basic principles of child development with implications for measurement are described, along with issues in accurately capturing the complex nature of early development with feasible, cost-effective measures. Analyses of some of the most commonly-used regional and global measures of young children’s development are described, with emphasis on identifying the policy relevance, feasibility, and methodologies that influence their potential usefulness for measuring national, regional or global progress towards proposed global education goals. Finally, directions for measuring early childhood development and learning in the proposed agenda are outlined.

II. Introduction

As part of the proposed agenda on sustainable development, early childhood development has been identified as a critical element of reaching proposed education goals, with the language of the proposed target placing emphasis on children’s development at the start of school. Early childhood development, which refers to children’s neurological and physical growth in the early years of life, has been shown to have lasting implications for learning, health, and well-being (Walker et al., 2007). While the start of school is the entry point for the education system, children’s development at the start of school is the manifestation of years of influence, reflecting health, nutrition status, and exposure to stimulation and emotional support from the first days of life.

As countries make investments in early childhood programs and policies, interest in measurement increases as well. Accurate, reliable measurement of early childhood development can help inform sound policy formation; identify where additional investments are needed; help inform curricula, instruction and teacher training; and perhaps most critically, identify children who are at risk of school achievement, either on an individual or group level. Measurement of early childhood development at the population level with representative samples of children, is also central to tracking equity: without knowing where children begin at the start of formal schooling, it is impossible to know whether education systems are successful in closing gaps between children as the school years progress. Measuring children’s development and learning early in their school careers is essential for knowing whether school systems are achieving the goal of learning for all, or whether children’s school experiences only mirror or exacerbate the inequities present before schooling begins.

The Education 2030 agenda includes emphasis on measurement of learning across all proposed targets, including early childhood development. There are unique questions on measuring learning for each target. Measuring young children’s development and learning has long triggered debate on both the likely reliability of such measurement, given the inability for young children to respond to written questions, and the ethics of involving young children in assessments of learning. It is sometimes questioned whether it is even possible to measure early childhood development and learning at a population level (e.g., GMR, 2015), especially when attempting to compare children’s development across countries or regions using one measure to determine the percentage of children “developmentally on track.”

Despite concerns, conceptual frameworks to guide measurement and several measures of early childhood development have been developed over the last decade for tracking at the population level, providing a strong starting point for considering how best to track progress towards proposed goals, be it on a national, regional or global scale. Measurement in early childhood has also been increasingly recognized as a critical piece of bringing early childhood programs to scale (Raikes, Dua & Britto, 2015; Frongillo, Tofail, Hamadani, Warren, & Mehrin, 2014). This paper discusses population-based measurement, or measures that are designed to be used groups of children to inform policy at a national, regional or global level, rather than to diagnose individual children as having developmental delays. Population-based measures differ from measures designed for research or program evaluation because they are designed for use at scale (Zill & Ziv, 2007), with an emphasis on feasible, cost-effective measurement, and therefore may be broader in scope than measures used for measuring program impacts or answering specific research questions. Experience from the several population-based measures used broadly to date demonstrate the strengths and challenges of measuring developmental status among groups of children.

There are several important questions to ask about large-scale measurement of early childhood development and learning, and how such measures can be used to determine which groups of children are “developmentally on track” for tracking of Target 4.2. While Target 4.2 is new, many of these issues have now been discussed and debated for several years (i.e., Myers 2006; Zill & Ziv, 2007). Issues relevant to measurement of child development and learning at a population level, and to tracking of Target 4.2 in particular, include the following: 1) what does it mean for a child to be developmentally on track in diverse contexts, and what contextual information is required to accurately interpret data; 2) technical issues and approaches to measuring child development; 3) examples of large-scale regional and global measures of child development and learning used to date; 4) policy implications, or how population-based measures of child development and learning should be interpreted and used to improve services for young children; and 5) next steps in large-scale, population-based measurement of child development and learning.

The sustainable development agenda places considerable emphasis on equity. A focus on early childhood development can be seen as a pathway to promoting equity, by ensuring that children begin formal schooling on an even plane. Based on existing research, it is likely that tracking early childhood development and learning will reveal inequities beginning very early in life, at birth or earlier, that grow over time. It is sometimes difficult to disentangle whether differences between groups are due to cultural variations in development, or are manifestations in inequity that arise due to insufficient support for young children’s development. These inequities can be apparent both within and between countries, and are equally relevant to all countries, regardless of overall wealth. While not easy to tease apart, a careful measurement agenda must also address this issue, which is explored throughout the paper.

III. What does it mean to be developmentally on track? Defining normative development in diverse contexts

A. Theoretical definitions of healthy child development and implications for measurement

Target 4.2 was generated in response to policymakers’ questions about whether children are “ready for school,” or developmentally on track when school begins. In keeping with the intent behind the target,

measurement of Target 4.2 should begin with a shared understanding of what it means for children to be developmentally on track. The study of how children develop, which elements of development are universal vs. culturally-specific and how early development influences later development, is now decades old. The scientific and theoretical literature on child development have notable implications for population-based measurement, in some ways underscoring both the importance and the difficulty of building measures that are accurate, reliable and useful in influencing policy.

At the highest level, “developmentally on track” 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 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. Four themes with significance for measurement are outlined below: 1) development proceeds in trajectories; 2) environmental and cultural influences have a profound and reliable effect on all children’s development; 3) children’s development is non-linear and interconnected across domains; and 4) defining “on track” development should be based on both basic developmental progressions and cultural expectations for children’s development.

First, development proceeds in trajectories. Child development and learning begins at the point of conception. Young children’s development and learning at the start of school is the extension of a trajectory that began at conception. Inequities in children’s cognitive and language development are apparent as early as four months of age (Fernald, et al, 2012), and tend to widen, not decrease, over time. Interpretation of data on child development and learning at one point in time, such as the start of school, should be based on a view of children’s development that begins at birth, as science would suggest that inequities uncovered in measurement at the start of school begin several years earlier and are the result of cumulative risks and lack of environmental stimulation in the early phases of life.

Second, patterns of child development may be similar in all settings, but environmental influences also have a profound impact on development. Basic patterns of human development and expression of individual traits are governed by genetic information that leads to commonality in developmental patterns in all people, such as the acquisition of early language and communication and the first expression of cognitive problem-solving skills. Human development reflects unfolding of genetic potential in response to environmental cues: nearly all human traits, skills and competencies reflect an interaction of genetic information with environmental stimuli, with some traits being more strongly influenced by either environment or genes than others but all reflecting a complex set of interactions between genes and environment (e.g., National Research Council, 2000; van Ijzendoorn, Bakersman-Kranenburg, & Ebstein, 2011). In this way, both the expression of genes and the underlying genetic information itself, which in turn influence all of development (including cognitive development) are profoundly affected by emotional stress, exposure to environmental toxins, health status and nutrition (Shonkoff & Garner, 2012). The influence of these factors on development and learning continues throughout life but with a more profound influence in early childhood.

Drawing from existing research, the strongest and most reliable predictor of young children’s development in both high and low-income countries is the home environment, even when children attend high-quality preschools. This is in part because children’s development and learning begins at birth, and also because children spend most of their time in homes, even after schooling begins – the home environment provides years of influence on children long before access to schooling begins and continues throughout the school years. Access to stimulating, supportive home environments is one of the key ingredients for young children’s healthy development in all cultures (e.g., Bornstein et al.,

2015). Home environments can be measured by asking about the frequency of activities or observing caregiver-child interaction; it can also be assessed indirectly through proxies like maternal education or household wealth, both of which reliably predict how frequently caregivers interact with and stimulate their children. Access to quality preschool settings is also important, although tends to be less important overall than home environments in studies of children's school achievement over time.

At the same time that environmental influences are very powerful, there are also patterns in development that young children everywhere follow – the emergence of early language and cognition; motor development; and the strong inclination to form relationships with others are all markers of healthy development in every culture. Especially when considering the question of whether early childhood development can be measured in a way that is globally comparable, developmental science has provided a strong basis from which to build on skills and competencies that appear to be universal in nature. However, how those skills and competencies are manifested is likely to vary based on environment. Children who are given more opportunities to communicate; who have more room to explore; and who have a large degree of cognitive stimulation will develop skills faster and with greater surety than children who are lacking these inputs. For example, children will develop language faster when they are spoken to more frequently (Hart & Risley, 1995) whether the frequency of language interaction varies due to cultural preference (e.g., some cultures hold babies on the backs of adults, who then have less face-to-face interaction) or the inability to bond fully with a dedicated, emotionally available caregiver, regardless of the reason. The science is not yet strong enough to know whether children's failure to show competencies using measures developed in other settings is due to lack of cultural relevance, inadequacy of stimulation leading to delayed development, or another cause. Those designing and using measurement in cross-cultural contexts must acknowledge this lack of clarity, and continue to test core assumptions on the universality of development.

Third, children's development proceeds in a combination of linear and non-linear growth. Some areas of development proceed in a linear fashion, such as the accumulation of vocabulary words and early literacy skills, which then become quite straightforward to measure (Thompson, 2014). Other developments, such as the understanding that people have different perspectives (known as theory of mind), emerge in one step at about the same age in more than one culture, at about 4 years (Sabbagh et al, 2006). As well, children's development may show "sleeper" effects, where development at an early age doesn't seem to show much relation to learning and achievement over time at some ages, but will emerge later as a strong influence (Vandell et al, 2010).

Some skills are relevant across domains and encourage learning and development in many areas (National Research Council, 2000). Self-regulation, or the ability to focus attention and behavior, is hypothesized to be relevant across all domains because it is so central to what children learn and experience. Therefore in US samples, self-regulation is understood to play an especially central role in school readiness (Eisenberg, Valiente, & Eggum, 2010). Language development also has a strong influence on many domains of development, including mathematics, literacy and social interactions. Social interactions are especially sensitive to expressive language because children with better language skills can communicate better with peers. Expressive language plays a role across multiple domains of development, including social/emotional, literacy and cognitive. Finally, some areas of development, such as executive function, seem to facilitate acquisition of new skills and knowledge, and thus show strong relationships with many domains of learning from early childhood onward (Blair & Razza, 2007). This means that measurement of early childhood development should include all domains of development, rather than focusing on early academic skills alone, and for some areas, effects of early childhood development may become apparent several years into the future.

Fourth, while we have good science backing the basic developmental processes and understand some of the mechanisms by which environment affects them, the definition of “on track” development is not yet fully established for children in many cultures. Measurement of early childhood development and learning, like learning at other ages, is relativistic, meaning that standards for what children should be able to do is informed by knowing what’s typical, or what children around them can do. What is normative is ideally established through the creation of normal distributions of children’s development and skill acquisition in various parts of the world, with ages at which percentages of children are typically demonstrating specific skills. For example, a recent project by the World Health Organization to analyse developmental milestones in various parts of the world for children between the ages of birth and age three found that some developmental steps, like sitting and standing in infancy, seem to emerge on very similar timetables everywhere. But as children grow older, environmental influences including poor health and nutrition, home environments and exposure to school, have a greater effect, and the range of ages at which children will develop skills within the “normative” range widens, and seems to have a stronger effect on some domains of development than others. Specifically, motor development, such as the ages at which children can jump with two feet, may be less responsive to environmental influences than skills like holding a pencil, which would be more sensitive to exposure (to pencils) and the opportunity to practice.

At present, normative development is difficult to define in a globally-comparable manner, not only because culture influences development, but primarily because the extent to which typically-developing children are achieving milestones or demonstrating skills is largely not known for many parts of the world (Marfo, 2011). To accurately measure the idea of “developmentally on track,” we would ideally have normative data on the ages at which children are achieving a specific set of skills deemed to be culturally applicable everywhere, which would then serve as the basis for global measurement by providing a scale of “developmentally on track.” But this scale does not exist at present. The role of environmental influences on the timing and expression of specific developmental skills is not known for many cultures. There are several risks of basing measurement on culturally-inappropriate expectations for children’s development. These risks include masking the true competencies of a given population, inadvertently contributing to inequity rather than addressing it, and formulating policy and/or programmatic practice on concepts of child development that are mismatched to parental or cultural standards and therefore decrease rather than increase investment in early childhood over time. With careful research, it may be possible to show that some measures are relevant to all children and can be used to define global standards for “developmentally on track.” This should be considered a research priority for the new education agenda.

In sum, we have strong evidence outlining the basic developmental steps that children go through in various parts of the world. We also know that there will be cultural variation in how these steps are expressed, and that environmental influences, especially due to lack of adequate stimulation, poor health and nutrition, will affect how and when children acquire these skills. Finally, our knowledge of normative development in diverse contexts is limited at present. Yet because both family and country wealth have been shown to have a profound impact on child development and learning (Bornstein, 2010), even for very young children, despite the limitations, measurement is essential for tracking equity. What does this complex set of influences say about how to measure child development and learning at the population level, and in particular, how to define “on track” in a manner that will reveal inequities and lead to more effective action on the part of young children? Three conclusions emerge:

First, because child development and learning proceeds in trajectories, measurement of child development and learning should ideally begin much earlier than entry to school and should also be

conceptualized as continuing on throughout the primary school years. From a developmental standpoint, the start of formal schooling is not an especially notable transition in development – rather, most children begin school with uneven skills in many areas, with some children showing advanced social/emotional development and others moving quickly in early reading and math. It is to be expected that children at the start of school will not be at the same level in all areas, as each child follows his or her own pattern of development. Measurement should then expect that children will show uneven patterns, with some domains stronger than others, and that developmental status measured at the start of school is the result of environmental influences that began much earlier. Over time, methodologies for measurement should be created to track children’s progress over time rather than focusing on just one point in time.

Second, measurement of “on track” development at the start of formal schooling should be placed in the larger context of a measurement framework that includes measurement of children’s home environments, health and well-being in addition to child development and learning when children enter the school system. Measurement of children’s development and learning, especially when taken at one point in time and without the breadth of information on children’s health, nutrition, and family environments, provides limited information on why children’s development is either on track or not, and what can be done to better support healthy development. An ideal system of measurement for early childhood would include measurement of children’s development starting very early in life along with measures of health, nutrition and home environments.

Third, the possibility of using local or national-level measurement that allows more careful consideration of cultural and contextual influences on the timing and expression of child competencies and skills should be considered, along with a research agenda to track normative development across cultures and develop approaches to measurement based on findings.

These conclusions must be balanced with the need for feasible, cost-effective approaches to measurement that can be used at scale and over time. This creates a technical challenge – measurement systems that account for the complexity of child development could be more comprehensive than most countries, including high-income, have the ability to implement. Noting the gap between the ideal system and what countries are able to do in the near future, a series of compromises must be made that balance technical strength, protection of children’s rights, and feasibility.

B. Practical application of theories and science of child development and learning to large-scale measurement

What areas of development are most essential to measure? With the implementation of Education 2030, it is likely that governments will begin the process of measuring Target 4.2 with concrete measures focused on child development and learning for children between the ages of four and six years. At the start of school, child development and learning refers to skills and competencies that reflect normative development and acquisition of age- and culturally appropriate competencies for children roughly between the ages of 4 and 6 years across a range of domains. While different scientists may call these domains by different names, there is widespread agreement among scientists that children are developing in related but conceptually-distinct areas that represent different skills and competencies. These domains can be roughly divided into categories:

Executive function: self-regulation, approaches to learning and other skills that drive learning across a number of areas. This domain include some of the strongest drivers of children’s academic performance over time, including sustained attention, working memory and the ability to inhibit impulses.

Social-emotional development: social and emotional skills that facilitate children’s successful interactions with others, including peers, teachers and family members. These skills and competencies facilitate children’s inclusion and the development of healthy, positive relationships with peers and adults, and may play an important role in helping children engage in school over time.

Pre-academic skills: early mathematics and literacy skills (sub-domains). This area is perhaps the best established in existing literature and includes early skills like letter/sound identification and counting that are considered fundamental to developing math and literacy skills later in life.

Motor development: fine and gross motor skills, including children’s coordination with walking, balancing, jumping and throwing balls; and ability to write, pick up small objects, and use their fingers successfully. Motor development, particularly fine motor development, has been shown to predict children’s cognition, presumably because it provides an indication of the degree to which children’s neurological development is on track (e.g., Grissmer, Grimm, Aiyer, Murrah & Steele, 2010).

All of these areas are important for children’s well-being and learning in school. Measurement of each domain is possible, but can require different methodology: for example, social/emotional development is sometimes more accurately measured through reports by parents or teachers who know children well, while pre-academic skills are likely most accurately measured through direct assessment of children. Some countries have also expressed desire to measure areas of knowledge or competencies such as moral development, religious knowledge, or knowledge of cultural practices and country traditions. These areas are undoubtedly important for many countries, and likely express one or more of the developmental constructs listed above.

There is also more agreement on how to measure some domains than others. There is a great deal of concordance in measures of early academic skills, as evidenced by many similar items in existing measures of children’s early math and literacy skills. There is less agreement on how to measure social/emotional development and executive function, both because the constructs are less clearly defined in existing literature, and also because the manifestation of these skills is less direct and concrete, and thus may vary more by culture and context. However, these are preliminary conclusions that deserve more research as assessments are developed and used on a wider scale.

Protecting children’s rights through measurement. At the start of the process to either develop or evaluate tools, it is essential to define measurement and its uses in a manner that protects children’s rights. For population-level tracking, population-based measures of child development and learning, or measures that are intended to provide an overview of children’s development and learning at a group level and not intended to diagnose individual children for developmental delays or specific learning needs¹, are most appropriate. There are notable risks for children when measures are used to measure the learning of specific children at the start of school, as children could be excluded from school on the grounds that they are “not yet ready.” For measurement to accurately track equity, all children, including those at risk for disabilities, must be included in samples, which means that even though the measures are not

¹ Identification of children with learning disabilities, especially for the purposes of referring children to special

intended to identify individual children, the range of items must be broad enough to capture all children's development, not just those that are typically developing. Items then would ideally be developed to capture a range of development, to ensure that children's competencies and skills are accurately reflected. As well, efforts to reach all children, which may necessitate household surveys rather than direct assessments of children who are enrolled in school, is also important.

Applying psychometric standards of technical strength to measures of child development and learning. Because measuring child development and learning is complex, technical standards for test construction that are often applied for education assessments can be difficult to meet but are also extremely valuable in helping to determine whether measures are robust enough for use in policy-making. Methodologies to evaluate the rigor of assessments are well-developed within the field of psychometrics and test construction and can be applied to population-based measures of child development and learning as well. Measures of child development and learning can be based on reports from caregivers or teachers, or can be administered through direct assessment of children in school or home settings; the technical issues outlined here largely apply to both types of measures. This section outlines some of the approaches to developing tools to measure child development and learning, as well as technical approaches for determining whether measures reach acceptable standards of technical strength.

Test construction begins with designation of underlying constructs upon which the instrument is based. For school readiness, constructs are theoretical and therefore not necessarily directly observable in their entirety – for example, children's social-emotional development refers to a group of behaviors or capacities, some of which can be observed and measured, and some of which are not very easy to measure. No test, especially not one that can be conducted in 20 minutes, will fully describe children's competencies or the quality of learning environments. Items are selected to reflect underlying constructs, but there is always a gap between what the items can measure and the underlying construct – the goal of good measurement is to reduce the gap as much as possible, but it is never fully addressed.

Many measures of child development and learning developed for use in low- and middle-income countries began with a review of items used in other assessments, especially tests developed for use in high-income countries, where investments in test design and validity are highest. Commonly-used large-scale assessments used to generate items for population-based instruments used in low and middle-income countries include the Bayley Scales of Infant Development, Revised Edition (birth through age 4 years) (Bayley, 2006); the Peabody Picture Vocabulary Test (age two through adulthood) (Dunn & Dunn, 2007); and the Kaufmann ABC for direct assessment (Kaufman & Kaufman, 2004); and the Early Development Instrument for teacher or parent report (Janus & Offord, 2007). Items from these assessments can serve as a starting point for adapting and testing items in other settings. However, there is an implicit assumption when using items across cultures that the items represent constructs that are equally applicable across settings, meaning that the constructs and the items will work in a similar manner in all countries. This is a difficult standard to achieve. Existing data suggest that scores on assessments of child learning do vary by country in ways that are not explained by family context.

Establishing technical strength requires a focus on validity and reliability. Reliability refers the degree to which an assessment tool produces stable and consistent results, while validity is the extent to which the test measures what it says it measures. Tests can be reliable without being valid – reliability is necessary but not sufficient for establishing validity. Types of validity vary based on the discipline, but all are focused on establishing that the assessment measures what it says it measures. Basic requirements for validity and reliability include the following:

- Reliability:
 - Test-retest: The test produces the same results when administered at different times
 - Inter-rater reliability: Different administrators of the assessment are able to use the assessment to gain similar scores
 - Internal reliability: The items that measure the same construct are related to one another (inter-item correlations or Cronbach's alpha).
- Validity:
 - Predictive: There is an implicit assumption in measures of school readiness that they are valuable because they will indicate how well children will perform in school over time.
 - Construct: The assessments measure what they say they measure and not other constructs – for example, items on school readiness reflect a range of underlying skills or competencies that are valued in each culture, rather than a limited set of skills that drive scores on all items.
 - Cross-cultural, or cultural invariance: The extent to which items work in similar ways when used in larger populations and across different cultures.
 - Concurrent: Scores on the assessment are correlated with scores on more established and fully validated measures, indicating that the new measure is valid in measuring the underlying construct of interest.

Taken together, reliability is easier to establish, while validity takes longer and a broader range of data and statistical tests to establish. Predictive validity is especially central to measurement of child development and learning. It is assumed that measures administered at the start of school are useful because they will provide insight into children's success in school over time. To date, there is less research establishing predictive validity of population-based instruments than is ideal. While there is good evidence showing that children who have higher levels of development at the start of school also show more learning and development as school progresses, many of the tests administered to children across countries have not yet been used in longitudinal studies that provide clear evidence of how scores at the start of school are related to children's achievement over time. Of the measures outlined below, none have yet documented evidence of predictive validity extending into the school years for the specific instrument, although several are comprised of items that have been shown to predict children's school achievement in other samples or settings.

Finally, language of administration is critical for accurate assessments of children's skills and competencies. For direct assessments, substantial debate takes place on what language children should be assessed in, the mother tongue or the language of instruction in the classroom. While many countries have mother-tongue policies mandating instruction for children in their mother tongue, in reality, children are often exposed to another language when attending preschool. Children who learn information in one language are likely to recall it better in that language, regardless of mother tongue. To gain insight into what children know from school settings, such as early academics and literacy, the language of instruction may be most accurate, while the mother tongue may be more accurate for expressive language. Such challenges must be thoroughly discussed with local stakeholders and alternatives piloted with groups of children before the administration of the assessment.

IV. Selected approaches to measurement

Several population-based measures of early childhood development have been tested in recent years. Below please find brief descriptions of six of these measures, with information on their methodology; relevance for policy; and technical feasibility.

A. East Asia Pacific Child Development Scales (EAP-CDS)

Background. Developed by UNICEF, the EAP-CDS was initiated following the implementation of the Early Learning Development Standards (ELDS), which helped countries outline developmentally-appropriate expectations for young children’s development (Rao et al., 2015). While the ELDS was implemented in several countries and regions, the East Asia region was especially engaged in the process of developing the standards. Under the leadership of the Asia Region Network for Early Childhood (ARNEC) and the UNICEF EAPRO, the scales were designed to “equip stakeholders across East Asia and the Pacific with a common measurement tool to assess the holistic development of children ranging in age from three to five years” (Rao et al, 2015). As there were no normative tests of child development specifically designed for the East Asia region, the test was intended to provide culturally-relevant measurement for children in the region, with a specific focus on informing policy for child development.

Methodology. In the first phase of instrument development, the items for the EAP-CDS were selected to correspond to the Early Learning Development Standards, which included several domains (approaches to learning; social/emotional development; cognitive development; cultural participation and knowledge; language and emergent literacy; motor development; and health, hygiene and safety). These items were then tested in several countries in a phased process, with a high degree of emphasis on validation and identifying which items were most critical to retain. The final instrument has 85 items with relatively equal distribution across domains. The EAP-CDS also collected information on home environments, including maternal education and the frequency of early-learning related activities at home. Due to some issues with sampling, the results should not be considered representative.

A six-country validation study was conducted in 2013 to 2014 with a sample of over 7,000 children in Cambodia, China, Mongolia, Timor-Leste, Vanuatu, and Vietnam. The results showed expected patterns across participating countries, with scores higher for children who were older; female; with more educated mothers; and who had participated in pre-primary education. Overall, the general similarity in the pattern of results suggests that the scales functioned roughly in the same manner across several countries. However, patterns of results varied by country, with some countries showing especially high scores in some areas (e.g., children in China has math scores that were significantly higher than children in other countries). As well, effect size estimates indicated that maternal education; age; and gender had different impacts on child development and learning in different places.

Policy relevance. The EAP-CDS was designed to have close concordance to the ELDS, which were in turn intended to provide a backbone for identifying culturally-relevant standards for children’s development that could be used to build better policies to support young children’s development. The team administering the EAP-CDS recently produced a report outlining differences in child development between countries. Placed against the context of basic demographic indicators from the participating countries, an indication of why such differences might arise can be surmised – for example, attendance in an early childhood programme had a significant impact on early child development, but the size of the effect varied by country, with the biggest impact (Cambodia) nearly three times the size of the smallest impact (Timor-Leste), which may arise due to the low rates of participation in Timor-Leste.

Given that results were recently released, it may be too soon to tell how and whether the results were used to influence policy in each country. It is also not clear whether the use of a regional measure with comparisons to other countries in the region has a stronger impact on policy than the use of national measures.

Feasibility. The EAP-CDS grew from the adoption of the ELDS; its items were designed to be aligned with national priorities for early childhood. To achieve this level of alignment in other parts of the world may require substantial background work. It is not clear whether the EAP-CDS would have the same relevance everywhere, although many of the items have been integrated into a new assessment in Africa. It is also important to note that the EAP-CDS requires trained assessors who are skilled and practiced in assessing children; the total length of the assessment is close to an hour for each child. The organizations supporting the EAP-CDS possess a high degree of technical skill and expertise, which was manifested in the strong design and careful attention to training. This level of training and expertise may not be possible to achieve across all administrations, although it is critical to note that such training is central for accurate direct assessments of child development and learning.

B. MICS Early Childhood Development Index

Background. The Multiple Index Cluster Survey is administered by UNICEF and is designed to provide country-level information on children's health, nutrition, social protection and education. Collected through parent report on household surveys, MICS is designed to produce nationally-representative data on children's well-being across a number of areas in low- and middle-income countries. MICS modules cover health, nutrition, education and social protection, and are relevant throughout childhood. Efforts to create a global indicator of holistic early childhood development (subsequently named the "ECDI," or early childhood development index) began in 2005, with the purpose of generating a succinct set of questions that could be integrated into parent surveys to index children's development between the ages of three years and four years, 11 months (Zill & Ziv, 2005). The goal of the MICS ECDI effort was to identify a set of simple, practical, holistic and commonly-accepted indicators that could be used to track children's development globally (Janus & Duku, 2013).

Methodology. After reviewing several large-scale surveys of child development conducted both at the national and international level, the MICS team selected a set of items for field-testing, drawing heavily from the Early Development Index (Janus & Offord, 2007), a teacher report instrument developed in Canada. A 48-item survey was developed, which was then reduced to 18 items for pilot testing, and finally reduced to 10 items, which were then used as the MICS ECDI module with a total of around 40,000 children. The final list of 10 items includes 3 items on numeracy/literacy; 2 items on social/emotional development; 2 items on approaches to learning; and 2 items on physical development. The MICS also includes a family environment module, with information on caregiver-child activities and access to playthings.

From the first round of the 14 countries participating in MICS ECDI (MICS4), two middle-income and two low-income countries were selected as "case study" countries to examine how the items were working, by examining sensitive to age (younger children should score lower than older children); household wealth; and gender. Results indicated that MICS items showed predicted associations with family wealth in most but not all countries, and that internal consistency of the scales was good for some areas (math and literacy) and good in some countries, but insufficient in others, for other areas (e.g., social-emotional development). Overall, it was concluded that the scale had adequate psychometric properties, although the evidence also suggested that the results were uneven and fluctuated based on country, signalling at least some degree of cultural variance. The reliance on one

set of items for children over a relatively large developmental span of two years, coupled with the desire to reduce the total number of items down to 10, creates tension within the MICS scale to fully capture children's development across a number of domains while remaining feasible for collection at scale.

Policy relevance. The MICS ECDI has now been administered in more than 50 countries (www.unicef.org/statistics). The MICS survey has the considerable strength of including information on home environments, children's health and nutrition as well as the ECDI, which improves its relevance to policy as the information on child development and learning can be interpreted within the context of more complete information on children's early experiences. However, reliance on a global set of items does not allow countries to align the item with local expectations for children's development, thus decreasing the relevance to policies on preschool curricula or teacher training. It is also difficult to know why children are not achieving specific items on the MICS, as the scale is very short and also shows different patterns of results in different countries that are not easily interpreted. Finally, the MICS ECDI is only available for children up to age 4 years, 11 months, which is not at the start of formal schooling in all countries.

Feasibility. Because it's a parent report instrument, the MICS ECDI does not require substantial training for data collectors and can be administered through household surveys in a short period of time. Accordingly, it is very feasible to collect the items on the scale. The sampling requirements for MICS are substantial, which decreases the overall feasibility of the MICS ECDI. Still, it remains one of the most feasible approaches for collecting information on child development and learning at scale.

C. PRIDI

Background. The PRIDI was created to generate regionally-comparable evidence on early childhood development for children between the ages of two years and four years, 11 months (Verdisco, et al., 2014). The PRIDI was initiated in 2009 to follow on models for regional and global assessments of learning in primary and secondary school, such as LLECE, TIMMS and PIRLS, given their notable success and influence on policy in the Latin American region. The two goals of the PRIDI were to 1) generate high quality, population-based and regionally comparable and relevant data on child development in nationally representative samples, and 2) identify gaps in child development between different groups of children (PRIDI Technical Annex, no date). The PRIDI was designed to capture an integrated view of young children's development by focusing on the contributors to children's development as well as describing developmental milestones. A central priority for PRIDI was to include indigenous and other marginalized populations, reflected in the process of designing the items and the resulting decisions on survey methodology. Four countries participated in the PRIDI: Costa Rica, Nicaragua, Paraguay, and Peru.

Methodology. The PRIDI had joint goals of cultural relevance across a range of groups, policy relevance, and the generation of globally-comparable data. To generate appropriate items, teams representing each country were assembled to map existing efforts to measure child development; agree upon a basic conceptual framework which then could guide the development of the survey; and ensure cultural alignment and political acceptability of the proposed items. After obtaining input from countries and reviewing literature, the PRIDI team agreed upon four domains for measurement: cognition; language and communication; social-emotional; and motor. Development of the PRIDI proceeded in three phases, beginning with a formative phase designed to test new items, followed by a validation phase with samples of 200 children in each country to determine the psychometric properties and construct validity; and a final phase to test the measure within nationally-representative samples of about 2,000 children in each country. Child development and learning as well as child, family and community characteristics and children's participation in early childhood development programs were

included in the survey. Items are collected through a combination of direct observation and maternal report.

Results from the validation study and nationally-representative samples indicated that the items showed expected sensitivity to age and cultural background, including maternal education. Correlations between social/emotional development and age were very small but stronger for household characteristics, while age was more strongly associated with cognition and language/communication than household characteristics. Like results from the EAP-CDS, the size of the association between maternal education and children's development varied by country. Within the nationally-representative samples, scores showed generally normal distributions, with higher frequency of children on cognition and motor scales showing very low scores, or children who were able to complete very few or no items on the scale, and a higher density of children with very high scores on social/emotional development.

Policy relevance. The PRIDI provides insight into both the developmental status of young children and the quality of their home environments and access to early childhood development programs. Results indicated inequities between groups of high and low-income children in every country, with some children especially at risk for inequity (indigenous and rural children).

Feasibility. The PRIDI requires household visits to administer, and takes about 30 minutes per child. Like other direct assessments, the PRIDI requires two weeks of training for proper administration, ideally with assessors who have experience with small children. The PRIDI has not been integrated into on-going country monitoring to date.

D. IDELA

Background. The International Development Learning Assessment was developed by Save the Children, a non-governmental organization engaging in program implementation and advocacy for children around the world (Pisani, Borisova & Dowd, 2015). The IDELA was originally developed for use as a program evaluation tool to build evidence regarding the impacts of Save the Children's programs on children's development and learning, in the absence of other feasible, low-cost and available tools to use for program evaluation. As the tool has been further developed, Save the Children is now promoting the tool as a measure for population-based monitoring in low and middle-income countries. The tool is designed for children between the ages of three years and six years and covers four domains: physical development, language, math/cognitive development, and social/emotional development.

Methodology. Save the Children began by identifying 65 items from other assessments, which were then tested in 11 countries where Save the Children had operational programs. The assessment was updated or modified based on several criteria, including the complexity of the item and its adaptation; feasibility of item administration; children's understanding of the task; ability to standardize training and administration; and relevance to early childhood standards where applicable (Pisani, Borisova, & Dowd, 2015). Items were also evaluated by their sensitivity to children's ages when possible, and the amount of time spent in pre-primary education, on the grounds that older children and those who have spent more time in preschool should have higher scores. The instrument has been used in more than 30 countries, with varied sample sizes and no representative samples to date. Results reported in 2015 indicate that the items demonstrate different relations with one another and with family characteristics in different countries. The instrument is most frequently used among children who are attending preschool programs, and it is not clear if IDELA has been used to date among representative samples.

Policy relevance. The IDELA includes four domains of development to communicate to policy makers and school administrators that all domains of development are important, not only pre-academic skills. The use of similar items across all countries should enable comparisons over time if Save the Children chooses to use the data in this way; to date, it is not clear how IDELA has been adapted to local policies or used to influence policy within specific country contexts. A randomized trial in partnership with the World Bank in Mozambique demonstrating that children who attended preschool had higher scores at the end of preschool, and more engagement in the early years of school (Martinez, Naudeau & Pereira, 2012) generated much attention and may have led to policy changes.

Feasibility. The items on the IDELA were specifically selected on the basis of feasibility for use across various settings. Like other direct assessment tools, the IDELA is adapted to country context, with some room for adaptation of items to local languages and adjustment of pictures and other aspects of tasks to better align with cultural expectations for young children's development. Also like other assessments, IDELA requires a week of training to ensure assessors are able to correctly and reliably administer the assessments. To date, IDELA has not been integrated into on-going system monitoring.

E. MELQO

Background. The Measuring Early Learning and Outcomes project (MELQO) was initiated to promote feasible measurement of child development and learning and quality of learning environments within low- and middle-income countries. MELQO was designed to be a process to help create and scale workable tools, with joint leadership from Brookings Institution, World Bank, UNICEF and UNESCO, and partnership from many independent experts and non-profit organizations. MELQO was initiated in 2014 and began working in countries in 2015 to test and validate tools, with the goal of creating open-source tools and guidance to be used at the population level across countries.

Methodology. The MELQO tools are designed to propose a "common core" that comes from existing assessments and can be used as a starting point for further country adaptation, as well as a possible means for comparisons between countries. The tools are comprised almost entirely of items taken from other instruments that have been tested within at least two low- or middle-income countries. Tool development began with articulation of common constructs by groups of experts and stakeholders, and then proceeded to select items from existing assessments of use in pilot testing. Two tools, one focused on parent/teacher and direct assessment of child development and learning for children between the ages of four and six years, and one focused on measurement of quality in formal, pre-primary learning environments, have been developed.

Policy relevance. The MELQO tools are intended to provide specific guidance for policy makers and others on how the data can be used to improve children's learning, through changes at the classroom, school and policy levels. The tools undergo an adaptation process at the country level to ensure alignment with government standards. For quality measurement, experience to date suggests that some countries require such substantial modification of the constructs and core to align with cultural and government expectations that the idea of a common core may be difficult to attain, whereas for child development and learning, the adaptation process is relatively easier. The intention of linking quality and child development and learning is not yet fully manifested in the tools, as the design and testing is now underway. The question of how the tools and data are used by policy makers for improvement is a key research question to be explored in present and future validation studies. For maximum use, it is anticipated that the tools should be integrated into monitoring systems, especially for quality of learning environments.

Feasibility. The MELQO tools require the same amount of training for child development and learning as other tools; the quality instrument also requires substantial training. Whether the tools can eventually be integrated into government systems is not clear; to date, all validation trials have relied on outside experts and data collection firms to train data collectors and analyse data.

V. Priorities and next steps for a global measure

Taken overall, it is clear that child development and learning can be measured at the population level, and in fact, such measurement may be essential for revealing inequities in development and learning that begin early and persist throughout schooling and beyond. Population-based measurement is also doable, with four important conclusions. First, 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. Second, 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, or if it is more appropriate at present to aim for construct equivalence (e.g., Frongillo et al., 2014), 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. Third, measurement of children's development and learning should be placed within a larger measurement framework that acknowledges the holistic and ongoing nature of young children's development. Data on child development at the start of school should be accompanied by information on children's characteristics, home environments, and access to early childhood development programs, as a more complete picture emerges when taking multiple pieces of information into account. Finally, to accurately measure the proposed Education 2030 agenda for early childhood, more research on normative development in diverse contexts is needed, ideally beginning with a common set of items and proceeding carefully through stages of testing to determine if these sets of items achieve reasonable standards for construct and item equivalence across settings.

How data from these measures is used to inform policy and practice is also deserving of attention. Despite agreement that early childhood is important for future development, there is less evidence to date that policy makers and practitioners have made changes based on such data. More work is needed to understand how and why such measures lead to change, so that investments in measurement are as high-leverage as possible.

References

- Bayley, N. (2006). *Bayley Scales of Infant and Toddler Development— Third Edition*. San Antonio, TX: Harcourt Assessment.
- Blair, C., & Razza, R.P. (2007). Relating Effortful Control, Executive Function, and False Belief Understanding to Emerging Math and Literacy Ability in Kindergarten. *Child Development*, 78(2), 647-663.
- Dunn, L., & Dunn, D. (2007). *Peabody Picture Vocabulary Test-IV*. Pearson International.
- Education for All Global Monitoring Report (2015). *EDUCATION FOR ALL 2000-2015: achievements and challenges*. UNESCO, Paris.
- Fernald, L., Kariger, P., Hidrobo, M., & Gertler, P. (2012). Socioeconomic gradients in child development in very young children: Evidence from India, Indonesia, Peru, and Senegal. *Proceedings of the National Academies of Science*, Oct 16; 109(Suppl 2): 17273–17280.
- Frongillo, E.A., Tofail, F., Hamadani, J., Warren, A., & Mehrin, S. (2014). Measures and indicators for assessing impact of interventions integrating nutrition, health, and early childhood development. *Annals of the New York Academy of Sciences*, 1306, 68-88.
- Grissmer, D., Grimm, K., Aiyer, S., Murrah, W., & Steele, J. (2010). Fine motor skills and early comprehension of the world: Two new school readiness indicators. *Developmental Psychology*, Vol 46(5), 1008-1017.
- Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children*. Baltimore, MD: Brookes.
- Janus, M., & Offord, D. (2007). Development and psychometric properties of the Early Development Instrument (EDI): A measure of children's school readiness. *Canadian Journal of Behavioural Science / Revue canadienne des sciences du comportement*, Vol 39(1), Jan 2007, 1-22.
- Marfo, K. (2011). Envisioning an African **child development** field. **Child Development Perspectives**, 5: 140–147
- Myers, R. (2006). Background paper for Education for All Global Monitoring Report 2007 Strong foundations: early childhood care and education. Quality in program of early childhood care and education (ECCE). UNESCO, Paris.
- National Research Council and Institute of Medicine (2000). *From Neurons to Neighborhoods: The Science of Early Childhood Development*. Committee on Integrating the Science of Early Childhood Development. Jack P. Shonkoff and Deborah A. Phillips, eds. Board on Children, Youth, and Families, Commission on Behavioral and Social Sciences and Education. Washington, D.C.: National Academy Press.
- Raikes, A., Dua, T., & Britto, P. (2014). A measurement framework for early childhood: Birth to 8 years of age. Discussion paper for the Institutes of Medicine, National Academies of Science, Washington, DC.

Rao, N., Sun, J., Ng, M., Becher, Y., Lee, D., Ip, P., & Bacon-Shone, J. (2014). Validation, Finalization and Adoption of the East Asia-Pacific Early Child Development Scales (EAP-ECDS). UNICEF, East and Pacific Regional Office.

Shonkoff, J., & Garner, P. (2012). The Lifelong Effects of Early Childhood Adversity and Toxic Stress. *Pediatrics* (Vol. 129 No. 1, pp. e232 -e246 (doi: 10.1542/peds.2011-2663).

Van Ijzendoorn, M., Bakersman-Kranenburg, M., & Ebstein, R. (2011). Methylation matters in child development: Toward developmental behavioral epigenetics. *Child Development Perspectives*, 5(4), 305-310.

Vandell, D., Belsky, J., Burchinal, M., Steinberg, L., Vandergrift, N., & NICHD Early Child Care Research Network (2010). Do effects of early child care extend to age 15 years? Results from the NICHD Study of early child care and youth development. *Child Development*, 81(3), 756-783.

Verdisco A., Cueto, S., Thompson, J., Engle, P., Neuschmidt, O., Meyer, S., González, G., Oré, B., Hepworth, K., Miranda, A. (no date given, retrieved in 2015). Urgency and Possibility Results of PRIDI A First Initiative to Create Regionally Comparative Data on Child Development in Four Latin American Countries. Retrieved online July 2015; http://www.iadb.org%2feducation%2fpridi%2finstrumentos%2fPRIDI_Technical_Annex.pdf/RK=0/RS=xHhu2R0ngacvr5laP9dVCdRZw5Q-

Walker, S. P., Wachs, T. D., Gardner, J. M., Lozoff, B., Wasserman, G. A., Pollitt, E., Carter, J. A. and the International Child Development Steering Group. 2007*b*. Child development: risk factors for adverse outcomes in developing countries. *Lancet*, Vol. 369, pp. 145–57.

<http://download.thelancet.com/pdfs/journals/lancet/PIIS0140673607600762.pdf?id=baat9JAFbm14CdiSXipsu>

Zill, N., & Ziv, Y (2007). Toward A Global Indicator of Early Child Development: Final Report. UNICEF.