Measuring development in children from birth to age 3 at population level

Dana Charles McCoy
Assistant Professor, Harvard Graduate School of Education, Cambridge, MA, USA

Maureen M. Black
Distinguished Fellow, RTI International and Professor, University of Maryland School of Medicine, Baltimore, MD, USA

Bernadette Daelmans
Coordinator, Department of Maternal, Newborn, Child and Adolescent Health, World Health Organization, Geneva, Switzerland

Tarun Dua
Medical Officer, Department of Mental Health and Substance Abuse, World Health Organization, Geneva, Switzerland

As early childhood development (ECD) is identified as an important target of the global development agenda, valid and reliable population-based indicators of young children's development are needed. The aim of this article is to briefly summarise the current landscape of initiatives to develop cross-culturally valid and easy-to-use indicators that can be used for global monitoring and assess ECD among children from birth to 3 years of age at the population level.

The foundations of adult health and well-being are built prenatally and during early childhood, when the brain develops most rapidly and both positive and negative experiences are most impactful (Shonkoff *et al.*, 2009). Scientific advances have shown that the consequences of early adversities, such as nutritional deficiencies and excessive stress, can last throughout the lifespan and into the subsequent generation, increasing risk for chronic illness, socio-emotional challenges, and economic difficulty (Hanson *et al.*, 2015; Johnson *et al.*, 2016). Evidence has also demonstrated that early intervention can mitigate the effects of such adversities, enabling young children to develop resilience and experience positive developmental outcomes despite difficult circumstances (Luby *et al.*, 2013).

In response to rapidly growing evidence of the importance of ECD for health, productivity and social stability, interventions to protect, promote and support early child development are high on the political agenda in the post-2015 Sustainable Development Goal (SDG) era (Lake and Chan, 2015). Target 4.2 of the SDGs describes 'access to quality early childhood development' (United Nations, 2015), while the derived Global Strategy for Women's, Children's, and Adolescents' Health (2016–2030) extends an emphasis on child survival to include children's ability to thrive (Every Woman, Every Child, 2015). To achieve the global target, population-level indicators are urgently needed to assess children's developmental status within and across countries. The aim is to enable policymakers and programme managers to make informed decisions regarding policies and resource allocation; to evaluate the effectiveness of ECD policies and intervention efforts; and to track progress in meeting global goals and targets (Raikes *et al.*, 2015). In this article, we outline methods that are available for the measurement of child development outcomes in young children from birth to 3 years of age individually, as well as three new approaches to developing population-level indicators for global monitoring of ECD in the same age group.

△ *Photo: © istock.com / Halfpoint*

Existing approaches and challenges

Historically, multiple approaches have been used to assess young children's developmental outcomes across diverse cultural contexts.

Individual-level assessments aim to quantify the developmental status of a particular child. The Bayley Scales of Infant and Toddler Development (BSID) (Bayley, 2006) and the Ages & Stages Questionnaire (Bricker *et al.*, 1999), for example, are two screening tools originally developed in the USA that have been widely used for research in low- and middle-income countries (LMICs) (for example, Kerstjens *et al.*, 2009; Sudfeld *et al.*, 2015). One of the major strengths of individual-level assessments is the precision with which they are able to quantify children's developmental status, as well as their ability to provide benchmarks for 'on-track' versus 'delayed' development. At the same time, individual-level assessments require a great deal of training and time to administer, are frequently subject to restrictive copyrights, and cannot be generalised outside of the cultural settings in which they were originally developed without careful adaptation (Fernald *et al.*, 2009; Peña, 2007; Sabanathan *et al.*, 2015). Even with the advent of scales designed specifically for LMIC settings, such as the Malawi Developmental Assessment Tool (Gladstone *et al.*, 2010) and the Kilifi Developmental Inventory (Abubakar *et al.*, 2008), the scalability of these assessments for use in global monitoring is limited.

'Indicators and measurement frameworks for population-level assessment are implemented at scale, and are therefore required to be easy to administer, while maintaining essential properties of reliability and validity.'

Population-level assessments differ from individual-level assessment in that they are meant to measure the developmental status of a population (of a region or country, for example), rather than an individual child. Indicators and measurement frameworks for population-level assessment are implemented at scale, and are therefore required to be easy to administer, while maintaining essential properties of reliability and validity. Universal indicators (those to be implemented globally) must also be cross-culturally and cross-linguistically comparable. Historically, population-level metrics for assessing children's development in the period from birth to age 3 have been extremely limited. As such, policymakers and researchers have largely relied on economic and health-related proxies such as poverty, mortality, stunting, and low birthweight to estimate the well-being of their children (Grantham-McGregor *et al.*, 2007). Although poverty and stunting meet many of the criteria for population-based indicators (Black *et al.*, 2016, forthcoming), they are distal to many developmental outcomes and minimally responsive to programmatic interventions.

'Historically, population-level metrics for assessing children's development in the period from birth to age 3 have been extremely limited.'

Ongoing efforts in population-level measurement from birth to age 3

At least three initiatives are working with the aim to develop population-level indicators of child development in the birth to 3 years age range. Although all began with a conceptual framework and a goal to develop indicators that are both scientifically rigorous and practically useful, each initiative has taken a different – but complementary – approach to achieving this aim.

The **Saving Brains Early Childhood Development Scale** (SBECDS) project is led by academic researchers from Harvard University and funded by Grand Challenges Canada. The SBECDS initiative began by developing a conceptual framework based on the core ECD constructs identified to be most predictive of later-life outcomes, and then filling in this framework with a set of caregiver-report items that were either inspired by those tested in earlier individual-level assessments, or wholly original (McCoy *et al.*, 2016, forthcoming). The resulting set of motor, cognitive, language, and socio-emotional items was then tested using both qualitative and quantitative data collection in an iterative fashion across nine high-, middle-, and low-income countries, with revisions to the items conducted on a regular basis to improve precision and ensure conceptual rigour. Testing criteria include:
1 clarity and alignment with the conceptual framework
2 criterion validity relative to direct individual-level assessments (such as the BSID)
3 test–retest reliability
4 internal consistency and factor structure, and
5 metric invariance and cross-cultural validity.

After a final round of pilot testing, the most robust set of items will be selected for public dissemination based on the above criteria, the results of item response theory analysis, and relevance for policy and practice.

The World Health Organization (WHO) commissioned a systematic review and located 14 datasets from ten countries that used one or more of seven developmental tools, with more than 22,000 children from birth to 3 years old. Using matrix mapping done by expert consensus, individual items that measured seemingly related developmental constructs were identified across tools. A logistic regression analysis was carried out on all sets of related items to identify those that were most developmentally meaningful. All sets of logistic regression curves were examined by an expert panel for their discrimination between older and younger children, between-country agreement regarding age of milestone achievement, equivalence across methods of data collection (parent report and direct observation) and countries, as well as their representativeness of developmental constructs identified by the systematic review. The final prototype includes items that:

1   had good fit on logistic regression between countries and between data collection methods
2   had low variability on age range of attainment, and
3   were considered feasible to undertake by caregiver report.

The prototype covers fine motor skills and perception; gross motor skills; receptive and expressive language; and socio-emotional domains. Pilot testing has been intiated in three countries, to be followed by factor analysis to determine the best grouping of items per sub-scale. The project is being funded by Grand Challenges Canada and the Bernard van Leer Foundation.

The **Global Child Development Group** assembled data from over 16,000 children representing 15 cohorts from 11 LMICs with the objective of developing a Developmental 'D-score' Growth Chart. Items were harmonised across scales for children under 3 years old. Many cohorts include longitudinal data after age 5 years, thus facilitating measurement of predictive validity. The process of developing D-score trajectories is based on a two-stage estimation procedure that uses the Rasch model and calculates change scores to form a continuous latent variable (D-score) (van Buuren, 2014). Using item response theory (IRT), the trajectory of D-scores has interval scale properties that can be used globally to calculate differences within and across ages and countries (Jacobusse *et al.*, 2006), much as height-for-age growth charts are used to indicate rates of stunting. Additional steps include:

1   Remove or adjust items when necessary and test for differential item functioning to examine whether individual items are comparable across cohorts.
2   Examine the predictive validity of the D-score with longitudinal data.
3   Construct descriptive age-conditional reference charts of normal development using data from a South African birth cohort of approximately 3000 children.
4   Determine a cut-off point for 'off-track' development and age-related indicators (akin to the definitions for stunting and wasting) based on psychometric properties.

The project is funded by the Bill and Melinda Gates Foundation's All Children Thriving initiative.

'At least three initiatives (each with a different – but complementary – approach) are working with the aim to develop population-level indicators of child development in the birth to 3 years age range.'

Points of convergence and future directions

Although distinct in origin, the approaches used by the SBECDS team, the Global Child Development Group and the WHO have converged to suggest a common framework for assessing population-level development in the birth to 3 years age range (see box). Moving forward, these teams will build on these points of commonality and work together to develop a common set of conceptually and empirically rigorous indicators for public dissemination. By pooling not only the data collected as part of these individual initiatives, but also the collective expertise of multiple stakeholders and institutions, the goal of this joint effort is to produce a common set of items for assessing children's early development on a global scale. In doing so, this collaboration will provide a cohesive approach for ensuring the successful monitoring of global progress in the post-2015 international development agenda.

Points of consensus for population-level measurement of ECD

1 **Conceptual framework** Global monitoring efforts focus on multiple domains of ECD relevant to early and later functioning, including:
   - fine and gross motor skills
   - receptive and expressive language skills
   - cognition and problem-solving skills
   - social and emotional skills.

2 **Psychometrics** Tools for measuring population-level ECD show clear evidence for reliability and validity, including a particular emphasis on:
   - replication over time and setting (test–retest reliability)
   - correspondence between direct assessment and caregiver report formats (criterion validity)
   - sensitivity to maturation, biological and environmental inputs, and intervention

   - prediction of later-life outcomes (predictive validity)
   - correspondence in meaning and relevance across socio-economic, linguistic and cultural contexts (cross-cultural validity).

3 **Usability** Measures for global ECD monitoring show evidence for usability, including the ability to:
   - be easily and quickly implemented and analysed using standardised protocols, at low cost
   - align with other global measurement initiatives
   - show clear implications for policy and practice
   - be freely accessed and in the public domain.

Abubakar, A., Holding, P., Van Baar, A., Newton, C.R.J.C. and van de Vijver, F.J. (2008). Monitoring psychomotor development in a resource limited setting: an evaluation of the Kilifi Developmental Inventory. *Annals of Tropical Paediatrics: International Child Health* 28(3): 217–26.

Bayley, N. (2006). *Bayley Scales of Infant and Toddler Development: Third Edition*. San Antonio, TX: The Psychological Corporation.

Black, M.M., Walker, S.P., Fernald, L.C.H., Andersen, C.T., DiGirolamo, A.M., Lu, C., […] Grantham-McGregor, S. (2016, forthcoming) Early child development coming of age: science through the life-course. Manuscript under review.

Bricker, D.D., Squires, J. and Mounts, L. (1999). *Ages & Stages Questionnaires: A parent-completed, child-monitoring system*. Baltimore, MD: Paul H. Brookes.

Every Woman, Every Child. (2015). Global Strategy for Women's, Children's, and Adolescents' Health, available at: http://globalstrategy.everywomaneverychild.org/ (accessed April 2016).

Fernald, L.C., Kariger, P., Engle, P. and Raikes, A. (2009). *Examining Early Child Development in Low-income Countries*. Washington DC: World Bank.

Gladstone, M., Lancaster, G.A., Umar, E., Nyirenda, M., Kayira, E., van den Broek, N.R. and Smyth, R.L. (2010). The Malawi Developmental Assessment Tool (MDAT): the creation, validation, and reliability of a tool to assess child development in rural African settings. *PLoS Medicine* 7(5). Available at: http://dx.doi.org/10.1371/journal.pmed.1000273 (accessed April 2016).

Grantham-McGregor, S., Cheung, Y.B., Cueto, S., Glewwe, P., Richter, L., Strupp, B. and International Child Development Steering Group. (2007). Developmental potential in the first 5 years for children in developing countries. *The Lancet* 369(9555): 60–70.

Hanson, J.L., Nacewicz, B.M., Sutterer, M.J., Cayo, A.A., Schaefer, S.M., Rudolph, K.D. *et al.* (2015). Behavioral problems after early life stress: contributions of the hippocampus and amygdala. *Biological Psychiatry* 77(4): 314–23.

Jacobusse, G., Van Buuren, S. and Verkerk, P.H. (2006). An interval scale for development of children aged 0–2 years. *Statistics in Medicine* 25(13): 2272–83.

Johnson, S.B., Riis, L. and Noble, K.G. (2016). State of the art review: poverty and the developing brain. *Pediatrics* 2016 Mar 7. pii: peds.2015–3075.

Kerstjens, J.M., Bos, A.F., ten Vergert, E.M., de Meer, G., Butcher, P.R. and Reijneveld, S.A. (2009). Support for the global feasibility of the Ages and Stages Questionnaire as developmental screener. *Early Human Development,* 85(7): 443–7.

Lake, A. and Chan, M. (2015). Putting science into practice for early child development. *The Lancet* 385(9980): 1816–17.

Luby, J., Belden, A., Botteron, K., Marrus, N., Harms, M.P., Babb, C. *et al.* (2013). The effects of poverty on childhood brain development: the mediating effect of caregiving and stressful life events. *Journal of the American Medical Association: Pediatrics* 167(12): 1135–42.

McCoy, D.C., Sudfeld, C., Bellinger, D.C., Muhihi, A., Ashery, G., Weary, T.E. *et al.* (2016, forthcoming). Development and validation of a population-level early childhood development scale for children 18 to 36 months. Manuscript under review.

Peña, E.D. (2007). Lost in translation: methodological considerations in cross-cultural research. *Child Development* 78(4): 1255–64.

Raikes, A., Dua, T. and Britto, P.R. (2015). Measuring early childhood development: priorities for post-2015. *Early Childhood Matters* 124: 74–7.

Sabanathan, S., Wills, B. and Gladstone, M. (2015). Child development assessment tools in low-income and middle-income countries: how can we use them more appropriately? *Archives of Disease in Childhood* 10(5): 482–8.

Shonkoff, J.P., Boyce, W.T. and McEwen, B.S. (2009). Neuroscience, molecular biology, and the childhood roots of health disparities: building a new framework for health promotion and disease prevention. *Journal of the American Medical Association* 301(21): 2252–9.

Sudfeld, C.R., McCoy, D.C., Fink, G., Muhihi, A., Bellinger, D.C., Masanja, H. *et al.* (2015). Malnutrition and its determinants are associated with suboptimal cognitive, communication, and motor development in Tanzanian children. *Journal of Nutrition* 145(12): 2705–14.

United Nations. (2015). *Report of the Open Working Group of the General Assembly on Sustainable Development Goals*. A/68/970. Available at: http://undocs.org/A/68/970 (accessed April 2016).

Van Buuren, S. (2014). Growth charts of human development. *Statistical Methods in Medical Research* 23(4): 346–68.