ECDI2030
Early Childhood Development Index

unicef
for every child

THE EARLY CHILDHOOD DEVELOPMENT INDEX 2030
A NEW MEASURE OF EARLY CHILDHOOD DEVELOPMENT

Suggested citation: United Nations Children's Fund, *The Early Childhood Development Index 2030: A new measure of early childhood development*, UNICEF, New York, 2023.

Photo credit
Cover: © UNICEF/UNI333261/ Bhardwaj

© UNICEF/UN0312259/Sokol

Contents

Abbreviations

BSID-III	Bayley Scales of Infant and Toddler Development, Third Edition
CAPI	computer-assisted personal interview
CREDI	Caregiver-Reported Early Development Index
DHS	Demographic and Health Surveys
EAP-ECDS	East Asia-Pacific Early Child Development Scales
ECD	early childhood development
ECDI	Early Childhood Development Index
HKU	The University of Hong Kong
IAEG-ECD	Inter-agency Expert Group on Early Childhood Development Measurement
IAEG-SDGs	Inter-agency and Expert Group on Sustainable Development Goal Indicators
IDELA	International Development and Early Learning Assessment
IRT	item response theory
MDGs	Millennium Development Goals
MELQO	Measuring Early Learning Quality and Outcomes
MICS	Multiple Indicator Cluster Surveys
NLSCY	National Longitudinal Study of Children and Youth
OECD	Organisation for Economic Co-operation and Development
PRIDI	Regional Project on Child Development Indicators
TAG	Technical Advisory Group
SDGs	Sustainable Development Goals
SMEs	subject matter experts
SYC	Survey of Young Canadians
UNESCO	United Nations Educational, Scientific and Cultural Organization
UNICEF	United Nations Children's Fund
WHO	World Health Organization

Executive summary

In 2015, UNICEF initiated a process of methodological work that involved extensive consultations with experts, partner agencies and national statistical authorities. Over the following five years, a sequence of carefully planned technical steps were executed incorporating both qualitative and quantitative methods to identify the best items to measure SDG indicator 4.2.1: "Proportion of children aged 24-59 months who are developmentally on track in health, learning and psychosocial well-being, by sex". This process led to the development of the Early Childhood Development Index 2030 (ECDI2030).

The ECDI2030 addresses the need for nationally representative and internationally comparable data on early childhood development, collected in a standardized way. Data collected with the ECDI2030 can be used by countries to monitor progress against SDG Target 4.2.

The ECDI2030 can be integrated into existing national data collection efforts. This is a public good and freely accessible to all those countries interested in undertaking data collection on ECD outcomes at the population level. The ECDI2030 has been translated into a number of languages, including the six official UN languages. It is accompanied by standard guidance and a set of implementation tools that include interviewer guidelines, training materials, syntaxes, tabulation plans and templates for reporting. The ECDI2030 and accompanying implementation materials can be found on the dedicated resource page.[1]

4 QUALITY EDUCATION

TARGET 4.2:

By 2030, ensure that all girls and boys have access to quality early childhood development, care and pre-primary education so that they are ready for primary education

Introduction

Early childhood has been recognized by the international community – and proven by decades of research – to be a crucial time in a child's life. During these early years, children develop the skills and capacities they need for future learning and growth. Generally understood to apply to children aged 8 years and younger, early childhood development (ECD) encompasses motor, cognitive, language, socio-emotional and self-regulatory skills and capacities.

The importance of ECD is explicitly emphasized in the Convention on the Rights of the Child, which has a special set of recommendations for implementing children's rights during early childhood. That it is a central component of global and national development is further evident from its inclusion in the Sustainable Development Goals (SDGs). SDG Target 4.2 calls for countries to "Ensure that, by 2030, all girls and boys have access to quality early childhood development, care and pre-primary education so that they are ready for primary education". One of the indicators for tracking progress on Target 4.2 is indicator 4.2.1: "Proportion of children aged 24-59 months who are developmentally on track in health, learning and psychosocial well-being, by sex".

The SDGs, adopted by the international community in 2015, supersede the Millennium Development Goals (MDGs) and significantly expand on them. They place more emphasis on previously overlooked issues, including ECD, and make all governments, not just those of low- and middle-income countries, accountable for achieving them.

The Inter-agency and Expert Group on SDG Indicators (IAEG-SDGs) initially classified indicator 4.2.1 as tier III (i.e., there was no internationally established way to measure it) and called for further methodological work to develop a measure that can be used across countries for the purpose of global monitoring and reporting. In order to assist governments in tracking progress on Target 4.2 – and as custodian agency for indicator 4.2.1 – the United Nations Children's Fund (UNICEF) led the technical research and design of an improved ECD measurement tool that is universally relevant for children aged 24 to 59 months, taking into account cultural and social differences in their development. To oversee the work, UNICEF established the Inter-agency Expert Group on ECD Measurement (IAEG-ECD), a taskforce that brought together representatives of national statistical offices, experts from development organizations and academics.

This report describes how UNICEF, along with the IAEG-ECD, developed a measurement tool that can be used to report against SDG indicator 4.2.1 and inform the development of evidence-based ECD policies and programmes. The work on the ECDI2030 began in 2015 and was completed in 2020.

The report is divided into four parts. Part One provides an overview of ECD, and its importance in the SDGs and in informing evidence-based policymaking, and briefly looks at existing ECD measures including an introduction to the ECDI2030. Part Two discusses the work process for the development of the ECDI2030, while Part Three explains in detail the conceptual framework, domains, subdomains and constructs of the new measure. The methodological work in establishing its reliability and validity is also presented, along with an overview of the testing conducted. Finally, Part Four introduces the final version of the ECDI2030 and its properties and explains how to calculate the related indicator.

Early childhood development measurement

What is early childhood development (ECD)?

Early childhood development (ECD) is a maturational and iterative process involving an ordered progression of motor, cognitive, language, socio-emotional and self-regulatory skills and capacities across the first few years of life. Generally understood to apply to children up to the age of 8, it is viewed as holistic and encompasses health, nutrition, social protection and stimulation, as articulated in the Nurturing Care Framework developed jointly by UNICEF, the World Bank Group and the World Health Organization (WHO).[2]

Why is this period of development so important?

Science has demonstrated that experiences in early childhood have lifelong implications for learning and well-being. During these early years, from conception to infancy and early childhood, a child's newly developing brain is highly plastic and responsive to change. This is evidenced by the billions of integrated neural circuits established through the interaction of genetics, environment and experience. Therefore, ECD is a vital period for children that sets the stage for lifelong thriving and is a critical window for human capital investment.

A child's experiences in the early years are linked to educational achievement later in life as well as the development of skills, capabilities and productivity in adulthood. This makes investing in ECD one of the most critical and cost-effective investments a country can make. In fact, economic analyses have found that investing in the early years of a child's life yields some of the highest rates of return to families, societies and countries.[3]

Do all children develop the same skills and at the same rate regardless of where they live?

While the overall developmental process is similar across cultures, children develop at different speeds and may reach developmental milestones at different times. What is considered 'normal' child development also varies across cultures and environments, since expectations and parenting strategies may differ between countries as well as among cultural, ethnic or religious groups within the same country. This makes measuring ECD in a way that allows for cross-cultural comparisons especially challenging.

Why is the measurement of ECD so complicated?

Approaches to measuring ECD at scale have traditionally been based on screening and/or diagnostic tests that depend on highly trained professionals and substantial administration time to generate valid information, making them inadequate for large-scale population monitoring.

The multi-dimensional nature of ECD further complicates the task of generating accurate data at population level through household surveys since specifically designed and validated instruments that can be implemented in a standardized way are required to generate robust and comparable data.

Additional constraints of population-level measurement in the context of multi-topic household surveys imply that instruments need to be short and that enumerators can effectively be trained in their administration.

ECD measurement tools: Past and present

Interest in ECD surged at the turn of the twenty-first century. The MDGs, established in 2000, called for governments of low- and middle-income countries to achieve and report on progress on eight goals related to human development and quality of life. While the MDGs did not have any specific goals or targets on ECD, they did include aspects that contribute to early child development, such as health (i.e., reducing child mortality) and investing in education.

Despite consensus on the importance of inputs affecting ECD in the MDGs, a population-based measure was not readily available in 2000, particularly for low- and middle-income countries. This was probably due, in part, to the lack of a specific mention of ECD in the goals and hence no globally coordinated drive, or investment, to develop such a measure.

The lack of an ECD measurement tool was problematic since population-based data provide one of the most effective ways to draw policy attention to the situation of children, propel action and ensure progress in creating equity in outcomes for all children.

In recognition of this gap, UNICEF initiated a technical process to create a global measure of developmental status in early childhood for use at population level, the Early Childhood Development Index (ECDI), within the context of the Multiple Indicator Cluster Survey (MICS) programme. MICS surveys produce nationally representative and comparable data on a number of key indicators of the health and well-being of children and their families.

Global monitoring requires comparable and standardized statistics. To achieve this, countries need a common set of data collection tools, including a core set of questions and a standard protocol to minimize the variability of key aspects related to implementation (such as measurement error due to varying levels of capacity in implementing agencies/ partners). These tools need to be suitable for population-level measurement of ECD outcomes and include questions that can be easily integrated into existing national data collection efforts.

When UNICEF started the process of creating a global, survey-based measurement tool for ECD in 2006, there were only a few population-based measures that could be used as reference. However, the landscape has changed since that time, with a number of groups working to develop, test and validate tools and measures of ECD for various purposes. Most notably and of greatest relevance, these efforts include:

- Early Development Instrument (EDI), developed by McMaster University

- International Development and Early Learning Assessment (IDELA), developed by Save the Children

- The Regional Project on Child Development Indicators (PRIDI), developed by the Inter-American Development Bank

- East Asia-Pacific Early Child Development Scales (EAP-ECDS), developed by The University of Hong Kong (HKU) and UNICEF East Asia and the Pacific

- The Measure of Development of Early Learning (MODEL), developed by the Measuring Early Learning Quality and Outcomes (MELQO) initiative led by UNICEF, the United Nations Educational, Scientific and Cultural Organization (UNESCO), World Bank and Brookings Institute

- Caregiver-Reported Early Development Index (CREDI), developed by Harvard University

- Global Scales for Early Development (GSED), developed by WHO

- International Early Learning and Child Well-being Study from the Organisation for Economic Co-operation and Development (OECD)

While all of the tools listed here are designed to capture children's development, they vary greatly in terms of the domains covered, methodologies employed and applicable age groups.[4] Most of these are also limited in one or more of the following ways:

- relying on direct assessment of children and/or teachers' reports

- not designed to produce population-level estimates

- lacking sufficient testing and implementation at scale across a large cross-section of countries

- designed for individual assessment or programme evaluation rather than population-based measurement and global monitoring

MICS ECDI

The construction of the MICS ECDI in 2006 involved work with countries and partners using a multi-method approach. To establish the reliability and validity of the survey items, psychometric properties were examined at each stage in the development process, which also included an extensive literature review and pilot testing.[5]

The MICS ECDI comprises four domains: literacy-numeracy, physical development, social-emotional development and learning. It contains 10 items to measure whether a child is developmentally on track in each domain:

Literacy-numeracy: The child can do at least two of the following: identify/name at least 10 letters of the alphabet; read at least 4 simple, popular words; and/or know the name and recognize the symbols of all numbers from 1 to 10.

Physical: The child can pick up a small object with two fingers, such as a stick or rock from the ground, and/or the mother/primary caregiver does not indicate that the child is sometimes too sick to play.

Social-emotional: The child demonstrates at least two of the following: gets along well with other children; does not kick, bite or hit other children; does not get distracted easily.

Learning: The child follows simple directions on how to do something correctly and/or when given something to do is able to do it independently.

The MICS ECDI was introduced in 2009 in the fourth round of MICS (MICS4) and has since also been collected in MICS5 and MICS6 as part of the questionnaire for children under 5, as well as in a number of Demographic and Health Surveys (DHS). Data on the MICS ECDI are now available for more than 80 (mostly low- and middle-income) countries. The process behind the development of the MICS ECDI is documented elsewhere.[6]

Inclusion of ECD in the SDG monitoring framework

In 2011, a series of published papers on early childhood development[7] and its influence on human development indicators later in life augmented the global discussion around the need to include specific targets related to ECD in the MDGs' successor, the Sustainable Development Goals (SDGs).

In anticipation of the adoption of the SDGs, UNICEF convened a technical consultation on the measurement of ECD with some 35 key experts and partners in January 2015. The main objectives of the meeting, among others, were to discuss the need and possibilities for creating new measures of ECD, to reflect on opportunities for future methodological work and to discuss and generate consensus on a set of ECD indicators to be recommended for inclusion in the post-2015 agenda.

Following the creation of the Inter-agency and Expert Group on SDG Indicators (IAEG-SDGs) in March 2015 during the 46th session of the UN Statistical Commission, work began on developing and implementing a global indicator framework to accompany the SDGs, which were formally adopted by the UN General Assembly in September 2015. As part of the global consultation process, UNICEF submitted a proposal, with support and endorsement from other UN agencies and partners in the global ECD community, for inclusion of a dedicated indicator on early childhood development.

These advocacy efforts resulted in the successful inclusion of an ECD indicator (4.2.1) in the global indicator framework developed by the IAEG-SDGs, agreed on at the 48th session of the UN Statistical Commission in March 2017 and subsequently adopted by the UN General Assembly in July 2017. In its original formulation, 4.2.1 referred to the "Proportion of children under 5 who are developmentally on track in health, learning and psychosocial well-being, by sex".

To support implementation of the indicator framework, all SDG indicators were classified into three tiers on the basis of their level of methodological development and the availability of data at the global level. SDG indicator 4.2.1 was initially classified as tier III, meaning the IAEG-SDGs deemed in 2017 that there was no internationally established

methodology available for its measurement. A new measurement tool therefore needed to be developed and validated. As the custodian agency for indicator 4.2.1, UNICEF had the mandate to lead the technical and methodological work on the development of an improved measure of ECD outcomes.

Although the MICS ECDI was significant in that it was one of the first international population-based measures of early childhood development, it was not sufficient to measure the new dimensions and breadth of ECD outlined in the SDG indicator. Beyond measuring a narrower scope of ECD, it was limited in that it was designed only for children aged 3 to 4 years. However, in the absence of a measurement tool fully aligned with SDG indicator 4.2.1, the MICS ECDI was recommended as an interim proxy measure for global monitoring and reporting.

As part of the process of developing and refining the global indicator framework, the IAEG-SDGs has met regularly since 2015. At its 9th meeting in Beirut in March 2019, it reviewed workplans and tier reclassification requests for tier III indicators. Indicator 4.2.1 was at risk of being dropped from the SDG global monitoring framework because it is not possible to have an indicator without an available measure. During this meeting, and in recognition of the efforts and significant amount of work accomplished by UNICEF and partners to develop such a measure, the IAEG-SDGs approved the reclassification of indicator 4.2.1 as a multi-tier indicator,[8] with the tier II portion referring to children aged 24-59 months while the tier III portion would continue to apply for children aged 0-23 months.

This decision was subsequently re-evaluated at the IAEG-SDGs' 11th meeting in November 2019. The Expert Group then recommended, as part of the 2020 comprehensive review process, to retain only the portion of indicator 4.2.1 referring to children aged 24-59 months since methodological work being led by WHO to develop a population-level measure of ECD outcomes for children aged 0-23 months had not been finalized.

© UNICEF/UNI211322/Rusanganwa

This recommendation by the IAEG-SDGs to delete the portion of indicator 4.2.1 referring to children aged 0-23 months (which was still classified as tier III) was approved by the UN Statistical Commission at its 51st session in March 2020. The Expert Group subsequently approved refinement of the indicator name for 4.2.1 as: "Proportion of children aged 24-59 months who are developmentally on track in health, learning and psychosocial well-being, by sex". Approval of this revised indicator name came from the UN Statistical Commission during its 52nd session in March 2021.

The ECDI2030

While SDG monitoring and reporting was the motivation for developing a new measurement tool, this work was also undertaken in response to the growth in the field of ECD measurement and recent advances in research on child development. There was also heightened interest among governments, donors and the wider international community in collecting better data in order to inform ECD policies and programmes. The ECDI2030 serves both purposes: to monitor and report on the SDGs; and to assist governments and other bodies in informing policy and programmes.

The ECDI2030 measures the three domains covered by SDG indicator 4.2.1: health, learning and psychosocial well-being. Some items are domain-specific, while others tap into multiple constructs and cross domains or subdomains. Rather than producing estimates of individual domains, the ECDI2030 generates a single overall prevalence estimate for all three domains. This is what is needed by countries to estimate the percentage of children developmentally on track (i.e., to report on SDG indicator 4.2.1).

The ECDI2030 was created to measure progress against a specific SDG indicator (i.e., 4.2.1) and therefore had to align with the indicator in a number of ways.

First, because the SDG indicator outlines three domains of ECD, UNICEF and its partners had to develop a new conceptual framework built around these, with health being a domain that was insufficiently addressed in the MICS ECDI. Constructing a new conceptual framework also required determining new constructs, subdomains and items.

Second, the SDGs call for all governments to monitor and report on the indicators, meaning that the ECDI2030 needed to be universally relevant for application and use in all countries, whereas the MICS ECDI was developed primarily for use in low- and middle-income countries. Therefore, the methodological work paid special attention to issues of cultural and social relevancy.

The ECDI2030 was also designed, as was the MICS ECDI, to rely only on caregiver reports, not direct assessment. This decision responded to the expressed need from countries for a relatively simple and easy-to-administer tool that could be integrated into existing data collection efforts and would not create an added burden in terms of time, resources and capacity.

Further, a key part of the methodological work has also been the establishment of formal collaborative partnerships, including the creation of an Inter-agency Expert Group on ECD Measurement (IAEG-ECD) as well as a Technical Advisory Group (TAG). Annex A lists the membership of these two groups. See Part Two for more details on these partnerships.

The ECDI2030 is a population-level measure that can be integrated into international household survey programmes (such as the MICS and DHS) or other national data collection efforts. Therefore, similar to the MICS ECDI, the ECDI2030 has a comparative advantage in that it will not require the implementation of separate, dedicated survey efforts, which are often time and resource intensive.

Moreover, by embedding the ECDI2030 in existing data collection efforts, countries will be able to produce disaggregated data on the ECDI2030 by key demographic and socio-economic variables. This is not always possible with other ECD measurement tools and is key to addressing the 'leave no one behind' agenda in the SDGs.

While the ECDI2030 provides governments with a tool to report on SDG indicator 4.2.1 and will produce comparable and representative prevalence estimates at the national level, it is not intended to evaluate programmes or interventions or to conduct clinical assessments of individual children.

PART TWO

Work process behind the ECDI2030

The ECDI2030 is the result of a carefully planned methodological process that involved extensive consultations with experts, partner agencies and national statistical authorities to generate the shortest possible set of items to measure the core domains and subdomains of early childhood development (ECD) in children aged 24 to 59 months.

In light of UNICEF's mandate to undertake the methodological work of developing a new measure of ECD outcomes, a two-day technical consultation was held in September 2016 with academic and technical experts as well as key partners. Participants at the meeting formed a Technical Advisory Group (TAG) composed of 11 academics and researchers from selected institutes, agencies and universities. Members of the TAG have technical expertise in the measurement of ECD and former experience with tool development and/or technical expertise in tool/instrument testing and validation. An important outcome of the meeting was a well-defined plan with regards to next steps in the process and a timeline for developing the ECDI2030.

To oversee the work, UNICEF formally established the Inter-agency Expert Group on ECD Measurement (IAEG-ECD) in March 2017 at a side event to the UN Statistical Commission. The IAEG-ECD was a global inter-agency advisory and coordination body whose overarching purpose was to oversee the revision, testing and validation of the ECDI2030. The IAEG-ECD was chaired by UNICEF and included partner agencies (OECD, UNESCO Institute for Statistics, WHO and World Bank) as well as regional organizations (Inter-American Development Bank), international non-governmental organizations (Save the Children), national research institutes (National Institute of Public Health of Mexico) and national statistical offices (Palestinian Central Bureau of Statistics and Statistics Canada). The IAEG-ECD was advised by and collaborated with the TAG.

Over the course of five years, UNICEF hosted a series of technical consultations with the IAEG-ECD and TAG, as outlined in Figure 1. Between the in-person meetings, UNICEF also hosted virtual webinars to regularly update the IAEG-ECD and TAG members on the status of the work and gain their inputs at key moments of decision-making.

Figure 1

In-person technical consultations of the TAG and IAEG-ECD between 2016 and 2019

SEPTEMBER 2016	JANUARY 2018	NOVEMBER 2018	JUNE 2019
• Review the MICS ECDI and discuss areas for possible improvement or revision • Explore the possible replacement of certain domains or items in the MICS ECDI, particularly with respect to the physical and social-emotional domains • Explore the possibility of lowering the age of eligible children and related implications on questionaire content/format • Explore the possibility of revising the analytical plan for construction of the total score, specifically with respect to the weight assigned to individual items within each domain as well as the contribution of number of items within each domain • Identify next steps and plans for validation, including cognitive and field testing	• Review results of cognitive testing in four countries • Discuss implications for modifying item set prior to field testing	• Review results from field test • Gain consensus on next steps and timeline for constructing the final measure	• Review and discuss final analyses and results on item selection and come to a consensus on the final draft set • Review and discuss results from the pilot of the standard-setting exercise • Discuss scoring of the new measure and development of the overall 'performance profile' • Discuss plans for the global panel of the standard-setting exercise

The methodological work involved a selection process to identify the best items across 12 core subdomains within the three general domains of health, learning and psychosocial well-being (Figure 2). Conceptual and psychometric criteria used for item selection included the following:

- The initial pool of questions was selected from previously validated instruments that produced reliable results across different cultural and socio-economic contexts and had relevance for policymaking.

- All questions were further tested, improved or discarded on the basis of results from cognitive testing in four countries, results from dedicated field testing in population-based household surveys carried out in three countries and harmonized data points from an additional 30 countries.

- Successive rounds of item selection based on psychometric testing were undertaken. Each round was carried out respecting content coverage and age coverage within each domain. In other words, all efforts were made to ensure that, after dropping any item, there were still items allowing measurement of all the conceptually identified subdomains for a given domain. In addition, age-coverage was achieved by ensuring that selected items resulted in a good distribution in terms of item difficulty across all months of age.

- Psychometric criteria included item difficulty, discrimination, overall reliability and specificity in terms of maternal report versus direct assessment of the child's development.

- Once the final set of 20 questions was identified, a global panel of experts on ECD participated in a standard-setting exercise to define the number of milestones expected for each age group. This led to the definition of the ECDI2030 cut scores, which identify children who are developmentally on track in health, learning and psychosocial well-being.

Figure 2
Work process behind development of the ECDI2030

LITERATURE REVIEW

Comprehensive review and expert rating on 500+ items from 20+ instruments (March 2017)

DRAFT SET OF QUESTIONS

Evaluating items in terms of comprehension, interpretation and cultural adequacy through cognitive testing in Bulgaria, Mexico, Uganda and the United States (June-October 2017)

REVISED DRAFT SET OF QUESTIONS

Testing the administration of items in population-based surveys in Belize, Mexico and State of Palestine (June 2018-February 2019)

Harmonizing a global dataset with data from the ECDI2030 field tests in three countries and from an additional 30 countries

Identifying the best items using discrimination and difficulty properties

Undertaking further analyses to identify the minimum set of items fulfilling content and age coverage criteria

FINAL SET OF QUESTIONS

Defining cut scores through a standard-setting exercise (August 2019)

CONSTRUCTION OF THE INDEX

Translating the module into the six official UN languages

Developing manuals, training material, syntaxes for data analysis, tabulation plans and templates for reporting

COUNTRY IMPLEMENTATION

PART THREE

Building the ECDI2030

Conceptual framework

Early childhood development (ECD) is considered a holistic process in that capabilities within different domains are highly correlated and overlapping, particularly among very young children. Therefore, the model selected for the conceptual framework assumes one underlying latent construct of development.

The formation of the domains involved determining the most salient subdomains and constructs within each domain based on a review of the research literature and existing evidence. The process was also informed by substantial input from the Technical Advisory Group (TAG) through a desk review and expert rating of many existing tools and measures of ECD to identify a pool of potential items to capture the subdomains and their constructs. The initial conceptual framework is depicted in Figure 3.

Figure 3

Initial conceptual framework of the ECDI2030

📖 Learning		
Early numeracy*	Language, literacy and communication*	Fine motor development*
• Counting (verbal and set production) • Number identification • Number/size discrimination	• Letter/alphabet knowledge • Phonological awareness • Expressive language • Receptive language	
Executive function*		
Mental flexibility Working memory Inhibition		
Approaches to learning/play		

Note: The executive function and approaches to learning/play subdomains are cross-cutting areas impacting each of the other subdomains of learning.

🤝 Psychosocial well-being	
Social development*	Emotional development*
• Social competence • Relationship skills • Interpersonal conflict resolution • Prosocial behaviour • Social cognition • Theory of mind	• Empathy • Emotion knowledge
Self-regulation	

Note: Self-regulation is a cross-cutting subdomain impacting the other subdomains of psychosocial well-being.

⊕ Health	
Self-care*	Gross motor development*

Note: Subdomains marked with an asterisk (*) were retained in the final ECDI2030.

A general description of each domain, subdomain and associated constructs follows, while additional details on the importance of the constructs to ECD and supporting research evidence in the learning and psychosocial domains are included in Annex B.

Learning

Learning has been broadly conceptualized as comprising a number of subdomains that reflect early pre-academic skills and competencies critical to the later acquisition of more complex skills as well as academic success. Research literature and broad agreement among technical experts has identified the following five core subdomains of learning: (1) early numeracy, (2) language, literacy and communication, (3) fine motor development, (4) executive function and (5) approaches to learning/play.

The executive function and approaches to learning/play subdomains are considered cross-cutting underlying areas impacting each of the other subdomains.

Early numeracy

The subdomain of early numeracy covers 'numeric concept', 'number knowledge' and 'identification, classification and sorting'[9] as well as 'comparison and measurement', 'patterns' and 'geometry'.[10] The constructs within this subdomain considered on the ECDI2030 include: counting (both verbal and set production), number identification and number/size discrimination.

Measuring early numeracy during early childhood is especially important because number sense in infancy predicts the math abilities of preschoolers.[11] Early math skills are also the basis for readiness for more formal and complex math instruction in later primary schooling[12] – an especially important point since early math skills are the strongest predictors of later school achievement, stronger even than early reading or early attention skills.[13] Moreover, early math skills are important to measure since they have been shown to be a more powerful predictor of later reading skills than early reading in predicting later math skills.[14]

Language, literacy and communication

The area of language on its own includes the "knowledge and use of words, both in print and in oral form",[15] while language and literacy together can include letter identification, reading, receptive and expressive language/communication, pre-writing and fine motor skills such as holding pens/chopsticks and writing one's name, among other things.[16] Communication is defined as "the child's ability to verbally and non-verbally express needs, preferences, as well as emotions, and to listen and respond to the communications of others".[17]

The language, literacy and communication subdomain is comprised of the following four constructs: letter/alphabet knowledge, phonological awareness, expressive language and receptive language.

It is important to measure language, literacy and communication because the early ability to read is directly linked with later reading ability and reading achievement.[18] Vocabulary, knowledge of letters, words and beginning and end sounds in particular have been found to predict later learning.[19] These skills are also associated with varied and rich verbal interactions with parents, teachers and peers and with the availability of books in the home environment.[20]

Further, it is important to note that this subdomain and its constructs can be categorized and differentiated by age. For example, it is especially important to measure vocabulary for children under 3 years of age, given it is a meaningful proxy for overall language development.[21] For older children aged 4 to 5 years, it has been shown in high-income countries that this age group's language scores are linked with higher school achievement between the ages of 6 and 15 years.[22]

Fine motor development

The development of fine motor skills represents mastery of and control over small, precise movements, primarily through the use of the hands and fingers.[23] Typical skills developed in the early years include the ability to hold writing utensils (such as pencils or pens), using utensils for eating (such as spoons or chopsticks), writing one's name, drawing and picking up/manipulating small objects.

The TAG initially decided to include this subdomain as a component of learning instead of placing it under the health domain (which includes gross motor development, as outlined below) because of its conceptual relevance to learning and the documented associations that it has with later academic and learning success.[24] In addition to

demonstrating general physical development, fine motor development is also an indicator of executive function and overall neurocognition.[25]

Executive function
The subdomain of executive function refers to the ability to concentrate, focus, follow directions and inhibit reactions.[26] This definition has been elaborated on to also include goal formation, planning, carrying out goal-directed plans and effective performance.[27] While the roots of rudimentary executive function are discernable in the infant and toddler years,[28] it predominately develops during the preschool years.[29] Within the executive function subdomain, the conceptual framework considers the following three constructs: mental flexibility, working memory and inhibition.

Executive function is considered a cross-cutting subdomain, posited by some to overlap with self-regulation.[30] It is important to measure executive function because, together with self-regulation, it is associated with many skills required for kindergarten, including the ability to attend selectively, show appropriate social responses and stay engaged in academic tasks.[31] Executive function also predicts later school achievement, particularly in reading and math,[32] and it may be especially important for keeping children engaged in school over time.[33] Finally, executive function may be one of the most critical areas to measure due to its cross-cultural relevance.

Approaches to learning/play
This subdomain is considered cross-cutting and includes skills and behaviour linked to the processes of learning; in other words, how children learn. The literature on this subdomain indicates that it includes interest in learning, curiosity, creativity and initiative[34] as well as emotion regulation, attention and persistence.[35]

Measuring approaches to learning in ECD is especially important because the development of skills in its subcomponents is predictive of academic achievement in the later elementary grades and strengthens children's ability to take advantage of the learning opportunities that present themselves inside and outside of school.

Psychosocial well-being
While the literature offers diverse interpretations and definitions, this domain has been conceptualized as referring broadly to aspects of children's social and emotional development as well as the absence of mental health disorders. It captures skills and competencies related to forming and maintaining healthy interpersonal relationships with adults and peers, as well as regulating and expressing emotions in socially and culturally appropriate ways.

Significant associations have been found between psychosocial well-being and "outcomes in education, employment, criminal activity, substance use, and mental health […], even after controlling for important child, family, and contextual characteristics".[36]

Pursuant to the literature review and technical consultations, three broad subdomains were identified: (1) social development, (2) emotional development and (3) self-regulation, which is proposed as a cross-cutting subdomain.

Social development
The definition of social development refers to "the ability of young children to interact and sustain relationships with others".[37] Six constructs were chosen to capture this subdomain: social competence, relationship skills, interpersonal conflict resolution, prosocial behaviour, social cognition and theory of mind.

The differences between the constructs of social cognition, theory of mind and empathy are nuanced. All three can refer to the ability to take the perspective of other individuals, understand their emotional reactions and coordinate that into socially desirable interactions.[38] However, there are challenges in comparability due to differing frameworks and terminologies. Therefore, these three constructs were considered separately (with empathy categorized under the emotional development subdomain).

Emotional development
Emotional development refers to children's feelings about themselves and others and includes characteristics such as self-control, self-efficacy (i.e., the sense of being able to affect events) and the ability to properly interpret the emotions of others.[39] The two constructs that compose this subdomain on the ECDI2030 are empathy and emotion knowledge.

Self-regulation

Self-regulation is defined as "a child's ability to recognize and control impulses, manage stress and emotions, and exert self-control".[40] It also encompasses the ability to follow socially prescribed norms and rules, to be goal-directed and to be able to delay gratification.[41] Therefore, it is considered as a cross-cutting subdomain (i.e., present in both the social and emotional subdomains of psychosocial well-being).

Measuring self-regulation in ECD is especially important because it impacts children's ability to focus attention, control impulses or emotions and navigate relationships with others[42] in order to cope effectively with environmental demands.[43] It has also been shown that self-regulation in childhood is associated with later mental health, life satisfaction and well-being, income and labour market outcomes, measures of physical health, obesity, smoking, crime and mortality.[44]

Health

The development of the health domain has followed a slightly different approach than the other two domains.

Following a review of existing conceptual frameworks on child health, a literature review and an assessment of data availability, there was an initial recommendation to create a composite indicator of health comprising gross motor development, immunization and stunting. The criteria used to select these components of health were that they should:

- directly relate to children rather than inputs to families/environment

- measure current inputs/outcomes and not past ones

- be applicable over a wide age range within the group of children under 5 years of age

- be equivalent and relevant across contexts for all children

- be objective if possible

Stunting was later removed from the domain following an analysis and discussion of the results from the field test in Mexico (more details on this below),

citing reasons that stunting is (a) already a separate SDG indicator and (b) considered a risk to ECD, not an outcome of it and that (c) psychometric analyses revealed it was not mapping onto the domain of health. Immunization was also subsequently dropped as it is also a separate SDG indicator and was felt to be an input (rather than an outcome) that contributes to healthy child development and also reflects coverage of services.

The TAG also decided to include self-care as a subdomain, which is linked to motor development but was seen as an important developmental outcome contributing to child health.

The initial conceptual framework was refined and reorganized in light of subsequent item testing and analysis (for example, the fine motor development subdomain was finally placed under the health domain and some subdomains were dropped, such as approaches to learning/play).

Selecting the initial pool of items

After identifying the most relevant subdomains and constructs for all three domains based on expert consultation, and taking into account the conceptual framework, the next step was to create an inventory of items from existing tools that aim to measure child development at population level (as opposed to those developed for other purposes such as individual child assessment). For tools with both caregiver-reported and direct assessment items, only those items based on the former were retained for consideration.

The resulting inventory included more than 500 items drawn from over 20 instruments and tools. This initial bank of items included a number of direct assessment tools such as the IDELA and EAP-ECDS. These were eventually removed given the parameter of identifying only items that rely on caregiver or teacher reports. The resulting set of items was grouped according to domain, subdomain and construct.

Members of the TAG who are academics and technical experts in the field of ECD measurement then participated in an exercise to rate these existing items against several predefined criteria that included whether the item has policy relevance, is intervenable/actionable, has cross-cultural applic-

ability, is pragmatic/easy to administer, can accurately be reported on by caregivers and is known to have strong empirical evidence/predictive validity.

Feedback received from the experts was consolidated into a single master database, and all items that were either not recommended by any experts or recommended by only one expert were removed. This resulted in a revised bank of items recommended by a minimum of two experts, which was then organized again by domain and within each domain by age (i.e., those items relevant only for 2-year-olds, 3-year-olds, 4-year-olds, 2–3-year-olds, 3–4-year-olds or 2–4-year-olds).

Observations from the experts indicated that the initial set of items was skewed towards the inclusion of items for younger children, i.e., those who are 2 years old. Therefore, a further process of selection was undertaken to identify additional items to include, ensuring that those selected received a recommendation from at least two experts. All of the MICS ECDI items were also included with the exception of two: the item about kicking, biting and hitting was excluded since a similar item was already included from another measure; and the item about being too sick to play was also removed given it has not been shown to be a good indication of physical development due to its vague nature.

The resulting bank of 61 items (34 on learning, 20 on psychosocial well-being and 7 on health) was retained as the final set to undergo cognitive testing.

Testing

Cognitive testing

The purpose of cognitive testing is to evaluate survey questions in order to assess respondent comprehension and interpretation, enhance cross-cultural appropriateness of items, reduce response bias and identify potential sources of measurement error.[45]

As a qualitative methodology, cognitive testing typically involves in-depth, semi-structured interviews with purposive (and hence non-representative) samples. The primary benefit of cognitive interviewing over non-qualitative evaluation methods is that it provides rich, contextual data on how respondents

comprehend the underlying construct in survey questions, recall needed information, judge their answer and, finally, map their answer onto one of the available response categories. The data generated from this process provide insights into whether a question and its response options are capturing the original intent and meaning of the questions. Cognitive testing is a crucial step in designing new measures to inform item selection and ensure data quality. It is especially important in the field of ECD measurement given the complexity and culturally driven nature of early childhood development.

The items on the MICS ECDI had not previously undergone cognitive testing. Therefore, cognitive testing was conducted on the original set of 10 MICS ECDI items in India and Jamaica in March and April 2016. This involved a total of 45 parents of children aged 2 to 4 years in Jamaica (n=20) and 2 to 5 years in India (n=25).

The main findings from this round of testing included:

- Respondents found many questions in the MICS ECDI to be confusing.

- Questions tended to be complex; respondents had difficulty mapping their experiences onto the simple, binary "yes/no" answer categories and, as a result, used varying patterns of interpretation both across and within countries.

- Many of these patterns appeared to be 'out-of-scope', indicating that some items were prone to measurement errors.

As a result of these findings, two items were dropped, and it was noted that the remaining question set would benefit from further revision and testing.

In 2017, UNICEF conducted cognitive testing on the larger set of 61 draft ECDI2030 items compiled from a variety of existing measures and tools during the desk review and expert rating exercise (see previous section for more details). The testing was conducted in four field sites (in Bulgaria, Mexico, Uganda and the United States) in collaboration with national implementing partners (Institute for Public Policies Studies in Bulgaria, National Institute of Public Health of Mexico, Uganda Bureau of Statistics and RTI International in the United States). Cognitive testing across countries followed a standard protocol

for the overall research process, including interview methodology and data analysis. The main findings from the cognitive testing are included here, but more detail can be found elsewhere.[46]

The research process was iterative, with the findings from the United States informing changes made to items and interview methodology in Bulgaria and Mexico and findings from these two countries then informing the testing in Uganda.

Participant recruitment
Within each country, the purposive sample of respondents was recruited based on diverse characteristics, including child age (at least one 2–4-year-old child), child sex, household location, level of caregiver education and child disability status. In Mexico, mothers of indigenous ethnicity were also specifically recruited. Local staff from the implementing partner in each country used a variety of recruitment procedures to identify potential participants. In Bulgaria and Mexico, staff worked through word-of-mouth invitations and snowball sampling to recruit mothers who were neighbours or acquaintances. In Uganda, staff from the local implementing agency worked with chairmen (local leaders) from sampled areas prior to the data collection to identify participants with the desired attributes.

Interviewer training
Training to prepare interviewers to conduct cognitive testing included two-day, in-country, interactive sessions that incorporated cognitive interviewing techniques, practice and review of the interview guide as well as review of project procedures and logistics. Interviewers in all countries received standard training from a core team of trainers. Selected interviews were observed by trainers during fieldwork, either in-person or using video conferencing, with suggestions and corrections provided to the interviewer after completion of the interview.

Sample
Interviews were conducted in 2017 with 146 respondents across the four countries: Bulgaria (30 interviews), Mexico (47 interviews), Uganda (39 interviews) and United States (30 interviews).

Interview script
Interviewers used a combination of the 'think-aloud' method and scripted or spontaneous verbal probes to collect data about the response process. Since cognitive testing took place in four rounds, the interview script was adjusted to reflect the changes to questions and to elicit information about specific issues raised during the preceding rounds. Mothers were asked to listen to each question as the interviewer read it aloud and then provide a response. Following this process, participants talked through what they thought about as they answered each question. Follow-up probes were both general (e.g., "In your own words, what is this question asking?") and question-specific, such as probes about terms or concepts that might have been confusing (e.g. "What does 'identify' mean as it is used in this question?"). These scripted probes allowed interviewers to explore respondents' understanding of the question and whether they comprehended constructs anticipated to be problematic. Spontaneous ad-hoc probes explored inconsistencies in mothers' responses. Response types included dichotomous (e.g., yes/no) and frequency scale (e.g., never, rarely, sometimes, often, always) options depending on the item.

Analysis
Information gathered from audio recordings of participants during interview probing was compared to item descriptions to determine matches and mismatches between the intention of the item and participants' answers and interpretation of the item. Information was also examined in terms of underlying patterns related to the child's age and other mother or child demographics.

Results
Key themes emerged from the four rounds of cognitive testing in Bulgaria, Mexico, Uganda and the United States that could potentially lead to systematic measurement error or response bias. Areas of concern arose within every phase (i.e., comprehension, retrieval, judgement and response) of the question-response process, which is the underlying theory that guides cognitive interviewing.[47]

Issues with comprehension: Participants experienced issues in understanding some questions. Four main issues related to item interpretation and comprehension are important to note:

- There was confusion over concepts and terms used in some questions. For example, respondents

interpreted the item "Can (name) identify all written numbers from 1 to 5?" in three ways, as illustrated in Figure 4.

- Respondents sometimes conflated their child's willingness to perform a task with their actual ability to perform it (e.g., ability to go to the bathroom alone vs. willingness to do so).

- Respondents showed some confusion in understanding the difference between a child performing tasks versus performing them correctly (e.g., counting 10 objects with mistakes vs. doing so correctly).

- There was a lack of clarity when words have more than one meaning (e.g., the word 'distraction' in the question "Does (name) get distracted easily?" has two meanings in Spanish).

Figure 4

Participants' interpretation during cognitive testing of the item on written numbers
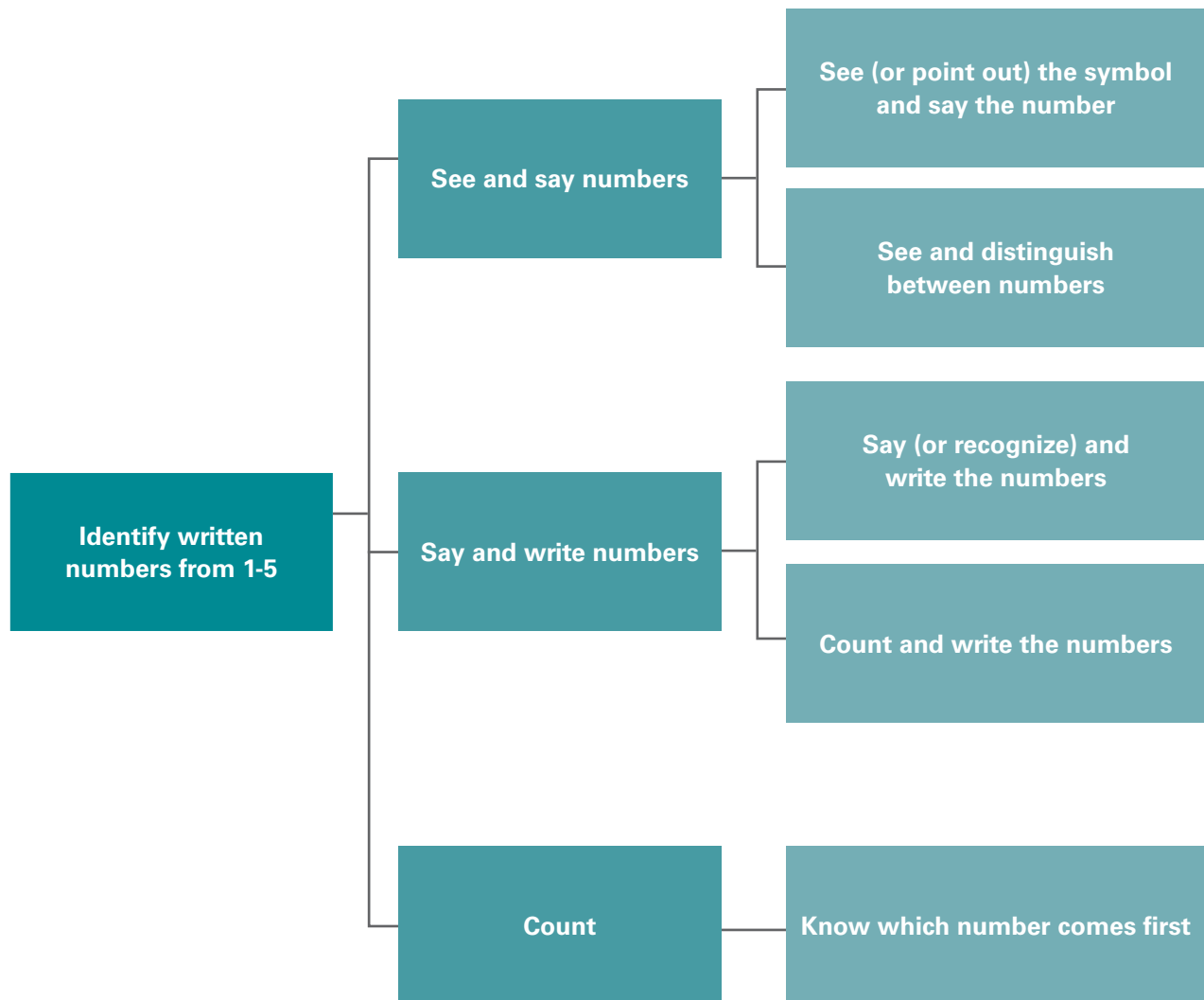
Issues with retrieval: In some cases, participants had never observed the behaviour being asked about (e.g., picking up a small object with two fingers) or were not familiar with objects referenced in the question as examples, such as animals that are less common in certain countries.

Issues with judgement: Some respondents, especially with younger children, based their answers on perceptions or suppositions rather than actual observation of the behaviour, such as when a child has not yet started to speak. Other concerns included participants varying in what they believe 'counts' as an affirmative response to the question, such as whether children needed to sing an entire song or draw an exactly straight line in order to affirm their child's ability.

Issues with response: Some questions were problematic because the questions provided only yes or no response options but participants wanted to give more variability in their answer (e.g., by responding 'sometimes'), such as whether a child becomes extremely withdrawn or shy in new situations. Respondents mentioned that scaled response options as opposed to a yes/no response option would more accurately allow representation of their children's behaviour overall.

Results from maternal and child subgroup analyses suggested potential interpretative differences based on maternal educational attainment and language/cultural background. For example, difficulty understanding some questions or certain words was more frequent among mothers with lower education levels. Subgroup analyses also revealed that the comprehension of some items was impacted by translation issues. In Mexico, for example, some questions were excessively wordy and therefore difficult to understand, such as, "Can (name) easily switch back and forth between activities such as going back to a game or playing with a toy after being interrupted?", which was not easily understood by almost all mothers and eventually dropped. Also, in Uganda, there was evidence that respondents based some of their answers on cultural expectations.

The different meanings that words can have in another language or their lack of appropriateness in the local context also interfered with the objective of some questions. For example, the word 'distracted'

in "Does (name) get distracted easily?" had another possible meaning in Spanish and, in Mexico, some mothers answered this item as if they were asked about whether the child is attracted to or likes a variety of things. Likewise, some examples were not appropriate across countries or contexts. For instance, the example provided in the item "Does (name) frequently act impulsively or without thinking (e.g., running into the street without looking)?" may have different implications in rural and in urban settings. In Uganda, a few mothers indicated that their child is always with them and others mentioned that they would not put the child in a position to be able to run into the street. In a similar way, when asked "Does (name) know that a goat weighs more than a mouse?", most mothers in an urban setting in Mexico mentioned that their children had never seen a goat and did not know what it was.

Informed by the results of the cognitive testing, recommendations for edits to item wording, response options, interviewer training instructions and other translation and implementation processes were discussed among the TAG and IAEG-ECD members at a technical consultation in January 2018. This resulted in a modified set of 58 items to undergo field testing (see Annex C).

Field testing

The primary goal of the field testing was to collect quantitative data to be able to assess the psychometric properties of the items in different cultural contexts. It was also an opportunity to test administration aspects of the ECDI2030 and its implementation tools, including training and instructions for interviewers, data analyses and tabulation plans.

Through the latter part of 2018 and early part of 2019, dedicated field testing was conducted on the draft set of ECDI2030 items in Mexico (see Figure 5) and State of Palestine. There was also a field test in Belize as part of the larger MICS6 pilot; as such, the sampling approach, instruments, eligible respondents and fieldwork procedures are detailed further elsewhere.[48]

The samples in each country were drawn using a stratified, probabilistic three-stage cluster design in order to ensure representativeness, and they included regional areas (West Bank and Gaza Strip in the State of Palestine), metropolitan, urban and rural areas (both indigenous and non-indigenous), as well

Table 1

Field testing in Mexico, State of Palestine and Belize

	Location	Sample size	Households surveyed	Response rate	Stratification	Instruments utilized	Translation
Mexico	• Puebla	Women: 1,455 Children: 1,641	1,513	92.8%	• Metropolitan • Urban • Rural-indigenous • Rural-non-indigenous	• MICS household, women's and children under 5 questionnaires • ECDI2030 draft set of questions • Anthropometry • ECD direct assessment	Spanish
State of Palestine	• West Bank • Gaza Strip	Women: 888 Children: 1,102	897	99.5%	• Region • Urban • Rural • Refugee camps	• MICS household, women's and children under 5 questionnaires • ECDI2030 draft set of questions	Arabic
Belize	• Toledo • Belize City South Side	Women: 583 Children: 214 under age 5, of which 120 aged 24-59 months	680	89.9% (for children under age 5)	None	• MICS household, women's and children under 5 questionnaires • ECDI2030 draft set of questions (only subset of items from the learning domain)	Spanish

as refugee camps, depending on the country. The final sample size across countries is shown in Table 1. The surveys were implemented using computer-assisted personal interviews (CAPI) loaded onto handheld tablets.

In general, non-response due to absence of eligible household members was between 7.1 and 10.2 per cent, depending on the country. Refusals to participate in the survey accounted for about 1 per cent for all countries. The response rate for the child questionnaire was close to 100 per cent, and almost 100 per cent of these questionnaires were answered by the child's mother.

Field test instruments
Given the ECDI2030 objectives, standard MICS questionnaires were adapted so as to include only the relevant sections. The field test instruments were thus comprised of three questionnaires: a household questionnaire, a women's questionnaire and a questionnaire for children under 5 years of age. These

instruments were available via paper questionnaire or the CAPI system.

An adult household member (over the age of 18) was asked to respond to the household questionnaire, which included modules to collect information on household characteristics, water and sanitation, handwashing and education of household members.

The women's questionnaire was administered only to mothers of children aged 2 to 4 years and collected data on mothers' background (education, etc.). In Mexico, modules on marriage/union and on depressive symptomology were also included in this questionnaire.

The questionnaire for children under 5 years of age was administered to mothers of children aged 2 to 4 years and included modules on child background characteristics (age, sex, etc.), birth registration, child functioning, child discipline and early childhood development.

The ECD module included the ECDI2030 draft set of 58 items as well as questions about the availability of books and playthings in the home, early stimulation and responsive care from adult household members, inadequate supervision and attendance in early childhood education. In Belize, only the learning domain items were tested instead of the full set of items.

In Mexico, data were also collected on height and weight for each child aged 2 to 4 years as well as on a direct assessment of ECD, as explained next.

Direct assessment of early childhood development
The items on the ECDI2030 do not rely on direct assessment. However, to compare mothers' reports on their child's abilities to the child's actual performance related to those abilities, the field test in Mexico also included direct assessment of children. This module comprised 13 tasks related to learning and gross motor development, as detailed in

Table 2. The direct assessment items were adapted from validated ECD measurement tools including the Bayley Scales of Infant and Toddler Development, Third Edition (BSID-III),[49] Regional Project on Child Development Indicators (PRIDI),[50] IDELA[51] and the draft Developmentally On Track (DOT) assessment tool[52] developed by The University of Hong Kong (HKU).

Fieldwork personnel training
Prior to fieldwork activities in each country, trainings took place that consisted of theoretical and practical sessions. The contents included a general introduction to the survey and application logistics (including management of the CAPI), visiting households and approaching eligible respondents (including how to address lack of response and related problems) and interview techniques, followed by simulated practice of the questionnaires on paper and on tablet. Time was also devoted to discussion and clarification of any concerns.

Table 2

Direct assessment items used in Mexico field test

Item	Domain	Adapted from
Child gives three blocks	Learning	IDELA
Child counts ten blocks	Learning	BSID-III
Child counts five blocks	Learning	BSID-III
Child identifies big and small objects	Learning	BSID-III
Child stacks three or more blocks	Learning	BSID-III
Child writes his/her name	Learning	DOT
Child describes position of a block (e.g., "in", "on", "above", "under" another object)	Learning	DOT
Child uses verbs in past tense	Learning	PRIDI
Child names ten objects correctly	Learning	BSID-III
Child recognizes ten objects correctly	Learning	BSID-III
Child jumps up with both feet leaving the ground	Gross motor development	BSID-III
Child throws a ball	Gross motor development	BSID-III
Child catches a ball	Gross motor development	PRIDI

In Mexico, training of nurses for administration of the ECD direct assessment was carried out by estimating measurement variance in relation to results yielded by gold-standard measurers. Reliability and accuracy of ECD direct assessment was measured by Cohen's Kappa and Bland-Altman graphs. Standardization was achieved by a Cohen´s Kappa value of at least 0.80 and agreement within 95 per cent limits in Bland-Altman[53] graphs. All nurses with out-of-range values for the direct assessment participated in an additional practical training and standardization session.

Pre-test
Prior to the final fieldwork, a pre-test was conducted in both Mexico and State of Palestine to test survey instruments, fieldwork logistics and procedures, including the standardization of fieldworker interview techniques. The CAPI process was also tested for system consistency and data transfer protocols. In Mexico, standardization of the assessment protocols for both the ECD direct measurement and anthropometric measurements was also pre-tested.

The pre-tests led to several modifications and improvements to the survey instruments and procedures, including adjustments to paper and CAPI questionnaires, additional information on the interviewer's and supervisor's manuals, refinement of the ECD direct assessment application protocol and scoring guidelines and identification of final training topics that required additional content and/or clarification.

Final fieldwork activities
To ensure the overall quality of the information collected, there were multiple levels of operational oversight and supervision, including continuous monitoring of data-quality indicators throughout the fieldwork.

Interviewers visited sampled households, asking for respondents' consent to participate in the survey. Within the household, mothers of children aged 2, 3 and 4 years were invited to participate in the interviews. Questionnaire application followed standard MICS protocol, with fieldworkers returning to the household at least three times, if needed, to complete all questionnaires. Failure to obtain household consent to participate was recorded as household non-response, and failure to complete questionnaires after three visits to the household was recorded as individual non-response.

Figure 5
Overview of field test activities in Mexico in 2018
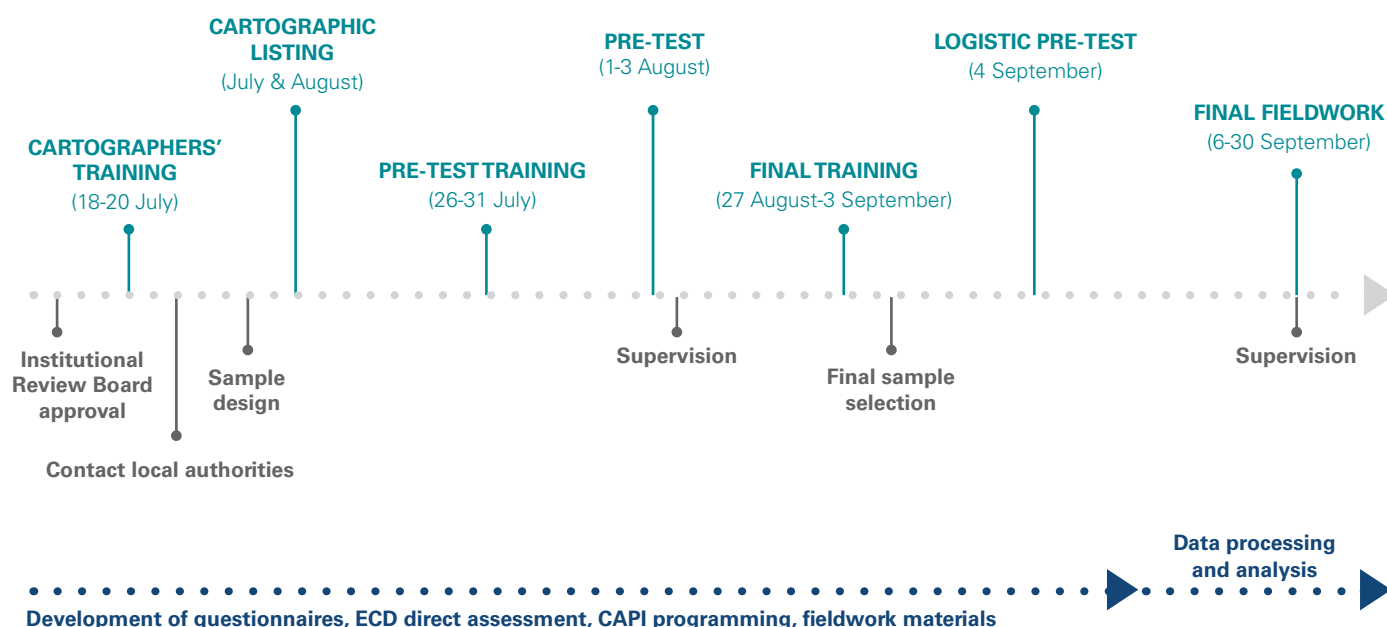
Field test data processing and initial analyses
For each field test, a data-processing team performed standard data consistency checks and produced data quality tables. Sampling coordinators generated household-, women- and child-specific sample weights considering the sample design, selection probabilities and non-response.

For each country, standard MICS indicators were calculated using the data collected by the different questionnaires. Using weighted data, a wealth index was calculated using principal component analysis of selected household characteristics (e.g., crowding, possession of durable goods in the house and having a bank account). The resulting index was used to classify households in wealth quintiles.

Initial data analyses focused on the Mexico and State of Palestine field tests, including estimating national-, regional- and state-level percentages and 95 per cent confidence intervals for all standard indicators, as well as generating results by key disaggregation variables (e.g., child's sex and age, maternal education level, area of residence and wealth quintile). Basic descriptive statistics were also produced for the draft ECDI2030 items to inspect overall endorsement of each question by age group and sex, as well as the proportion of 'don't know' responses for each item (Annex D).

Data processing and statistical analyses were performed in SPSS v.20, Stata v.15 and R version 3.6.0.

Harmonizing a global dataset

In addition to the data from the three ECDI2030 field tests, data from a further 30 countries were used to pre-screen candidate items and inform the analytic process to pare down and refine the draft set of items. Countries shared access to raw data or statistical summaries on candidate items (or their close analogues) to form a global dataset. Four countries had convenience sample data on 49 candidate items from HKU,[54] and 17 had convenience sample data on 24 candidate items from the Caregiver Reported Early Development Instrument (CREDI).[55] Data on 16 items from population-level data collected in Canada in 2008-2009 – the National Longitudinal Study of Children and Youth (NLSCY) and Survey of Young Canadians (SYC) – were also utilized. Ten countries

possessed data on one or more candidate items from the Measuring Early Learning Quality and Outcomes (MELQO), but country names were unknown as they had been annonymized in the dataset.

In total, data were harmonized for the 58 candidate items across 33 countries, from a collective sample of around 60,000 children aged 24 to 59 months (Table 3).

Psychometric analyses

Psychometric analyses were used to assess the performance of individual items in all countries with available data. Even though each step in the psychometric work involved different aims and methods, two overarching principles guided all the steps and item selection. First, content coverage was considered to ensure the selection of a sufficient number of items across all the conceptually defined subdomains. Each subdomain needed to be represented by at least one item on the final ECDI2030, and the total number of subdomains should be maintained. Age coverage was also established as a core criterion for the learning and health domains such that an item would not be dropped if there were no other items with similar probability of endorsement for the same age range. The specific aims and results of each analytic step are described in the following sections. The results of the psychometric analyses are detailed in another paper.[56]

Initial item screening
Aims: Pre-screening candidate items informed the analytic process to pare down and refine the draft set of items and remove any that had poor measurement properties.

Datasets used: Item selection of the 58 candidate items was based on information from all countries in the global dataset with available data (Table 3).

Methods: A decision-matrix was used to evaluate the difficulty and discrimination properties of the candidate items.

Item difficulty was assessed to ensure a selection of items with adequate variability in terms of difficulty across the ages of 24 to 59 months. To assess item difficulty, the percentage of children 'passing' each

item was estimated across each year of age. Two criteria were used in the decision-matrix related to difficulty: (1) At least a 10-percentage-point difference between children passing the item from ages 2 to 3 years old and (2) At least a 10-percentage-point difference between children passing the item from ages 3 to 4 years old. These criteria were only applied to items in the learning and health domains given that items within the psychosocial well-being domain were not expected to show a clear age progression.[57]

Item discrimination reflects the item's ability to discriminate between children with better and worse overall development. Point-biserial correlations were used as an initial approach to inspect item discrimination on the basis of two criteria: (1) Item-rest correlations of the item with the domain score of at least 0.3 and (2) Item-rest correlations of the item with the full set of items score of at least 0.25.

Each item was assigned a 'flag' on the basis of the extent to which the above criteria were met. Green-flagged items met all four criteria and red-flagged items did not meet any of the criteria. A yellow flag was also used for items meeting only some criteria or meeting all criteria but only in a few countries. Results are summarized in Annex E.

Table 3
Description of datasets used for initial item screening

Country and agency/survey	Age range	Number of children in the sample	Population-level probabilistic sampling
A. Mexico, National Institute of Public Health	2-4 years	1,641	Yes
B. State of Palestine, Palestinian Central Bureau of Statistics	2-4 years	1,099	Yes
C. CREDI (17 countries[1] aggregated)	2-3 years	16,317	No
D. The University of Hong Kong (Bangladesh, China, India and Myanmar)	3-4 years	633	No
E. Belize, UNICEF and the Statistical Institute of Belize	2-4 years	120	Yes
F. Canada, NLSCY	2-4 years	6,865	Yes
G. Canada, SYC2010	2-4 years	4,782	Yes
H. Country 1, MELQO	2-4 years	5,669	No
I. Country 2, MELQO[2]	2-4 years	112	No
J. Country 3, MELQO[2]	2-4 years	355	No
K. Country 4, MELQO[2]	2-4 years	205	No
L. Country 5, MELQO	2-4 years	1,388	No
M. Country 6, MELQO	2-4 years	5,269	No
N. Country 7, MELQO	2-4 years	10,152	No
O. Country 8, MELQO[2]	2-4 years	3,680	No
P. Country 9, MELQO[2]	2-4 years	344	No
Q. Country 10, MELQO	2-4 years	6,779	No

1. Information from Bangladesh, Brazil, Cambodia, Chile, Colombia, Ghana, Guatemala, India, Jordan, Lao People's Democratic Republic, Lebanon, Nepal, Pakistan, Philippines, United Republic of Tanzania, United States and Zambia.
2. Due to the low proportion of children aged 2 years old, the decision-matrix only considered information from children aged 3 and 4 years old.

In addition to the classical test theory-based screening using the decision-matrix, a subset of 9 items on the learning domain and 3 items on the health domain were compared to analogous direct-assessment items administered in Mexico. If the caregiver-reported responses and the direct assessment responses showed a large discrepancy (e.g., > 10% of responses in the off-diagonal cells), the item was flagged for follow-up and discussed with the TAG.

Main results: Items were dropped from the learning and health domains due to floor or ceiling effects in multiple countries. For example, "Pretends object is something else" was removed because it was highly endorsed in most countries regardless of child age and, since this was the only item in the approaches to learning subdomain, this subdomain was also effectively removed. Items were omitted in the psychosocial domain due to low correlations with other items in the domain and/or the overall ECDI2030. Items that addressed problematic internalizing or externalizing behaviours did not exhibit a strong correlation with other items in the pool or with child age, but they were retained to ensure content coverage of the psychosocial subdomains. Finally, it is important to note that the item "Recognizes three letters of the alphabet" was dropped despite being flagged as green (i.e., meeting all four criteria) because there was another item ("Recognizes five letters of the alphabet") that captures more advanced literacy skills and was needed for age coverage criteria.

The results from application of the decision-matrix are summarized in Figure 6 and Table 4. A total of 22 items were dropped among candidate items due to poor measurement properties, reducing the 58-item pool to 36 items that performed well across settings.

Figure 6
Decision-matrix criteria
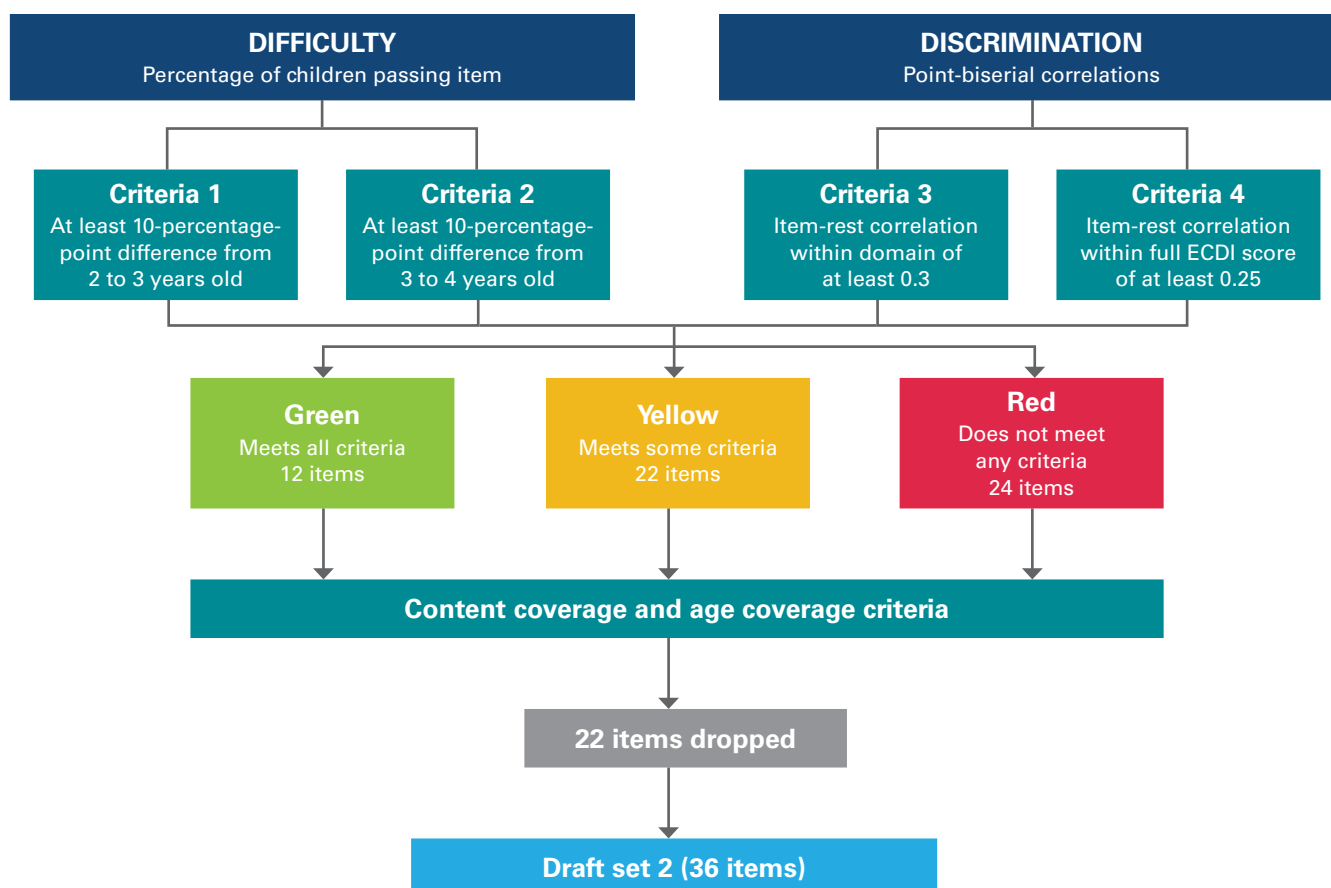
Only a single item was dropped based on the comparison with direct assessments ("Knows the difference between big and small"). Caregivers tended to underreport their children's ability to distinguish sizes compared to the direct assessment.

As expected, the psychosocial items related to problematic internalizing and externalizing behaviours were endorsed with similar frequency regardless of age. It was also the case that these items had low proportions of endorsement in all countries. For example, the response 'Always' was endorsed for the item "Seems to be sad or depressed" by about 2 per cent of caregivers in Mexico and by 3 per cent of caregivers in State of Palestine, and this was relatively constant over age. This led to a dilemma in deciding whether such items should be omitted because they were not strongly related to other items, or be retained because of their importance in defining the content coverage of the psychosocial domain. The consensus of the IAEG-ECD and TAG was to retain one item from each subdomain to ensure representation of all psychosocial subdomains, with the retained items having the better item-rest correlations overall.

A final review of the cognitive testing results also revealed that one item ("Says what others like or dislike") performed poorly as it did not appear to work as intended and the underlying concept did not seem to translate well into a survey question.

IRT test assembly
Aims: The objective of the automated test assembly (ATA) was to maximize the reliability of the ECDI2030 over the range (-1.65, 1.65) SD units on the item response theory (IRT) scale score.

Datasets used: Mexico and State of Palestine.

Methods: IRT-based test assembly was conducted separately in the two primary samples (Mexico and State of Palestine). All analyses utilized sampling weights reflecting selection probabilities and survey non-response. The IRT scale score of the ECDI2030 was standardized to have a mean of zero and a standard deviation of one in both samples.

The test information function (TIF) was used to operationalize reliability, and the objective function was computed as $TIF / (1 + TIF)$. The range of the IRT scale score was chosen to cover the middle 90

per cent of the normal distribution, because this range would be most relevant for determining on-track status, as defined by SDG 4.2.1. The TIF was computed using a unidimensional IRT model. In particular, the two-parameter logistic model (2PL) was used for items with binary response items, and the graded response model (GRM) was used for items with multiple categories.

ATA was used to maximize test reliability subject to the following constraints:

1) The final test form would contain 20 items based on logistical considerations.

2) Items from each of the content-based subdomains must appear on the final test form. This condition operationalized the content coverage requirements of SDG indicator 4.2.1.[58]

Applying these constraints, ATA was implemented in the R programming language using the mirt package to estimate IRT models and the eatATA package to run the test assembly. Resolving discrepancies between the ATA test forms in the two samples was informed by input and consultation with the IAEG-ECD and TAG, which focused on (a) the distribution of subdomain coverage by difficulty and (b) the face validity of the overall measure. This procedure resulted in the current version of the ECDI2030.

Main results: After screening, a total of 34 items remained in the item pool. These 34 items were then used for separate ATAs with the data from Mexico and State of Palestine. As shown in Annex F, the ATAs resulted in 17 out of 20 items in common across the two countries. Of the 17 items, 16 were included on the final ECDI2030. The remaining 4 items were selected based on input from the TAG. The final ECDI2030 form is shown in the last column in the table in Annex F.

Statistical properties of the ECDI2030
The methodological work resulted in a 20-item measure with marginal IRT-based reliability coefficients of .850 for the original ATA in Mexico and .839 for the final ECDI2030 and .880 for the original ATA in State of Palestine and .876 for the final ECDI2030. These results confirm that the ATA procedure yielded a single test form with reasonably good approximation to the original ATA forms in both

Table 4

Stages of item pool screening and items dropped at each stage

Analysis	Datasets used	Number of items analysed	Number of items dropped
Classical test theory-based screening (decision-matrix)	Mexico, State of Palestine, CREDI, HKU, Belize, Canada and MELQO	58	22
Comparison with direct assessment	Mexico	9 learning domain items, 3 health domain items	1
Psychosocial domain analyses and review of cognitive testing results	Canada, Mexico, State of Palestine, CREDI, HKU and MELQO; CT results from Bulgaria, Mexico, Uganda and United States	18 psychosocial domain items	1

countries. In addition, the final 20-item measure was approximately 90 per cent as efficient as the 34-item version (except at the higher end of the scale in State of Palestine), suggesting that the ATA procedure worked well at identifying a common minimum set of items. A summary of the statistical properties of the final ECDI2030 is presented in Annex G.

The IRT scale score was highly correlated with the unweighted total score in Mexico (0.96) and State of Palestine (0.95), suggesting that a unidimensional model to report a single overall score based on caregiver responses was appropriate for most of the items on the ECDI2030. Some items in the psychosocial domain (i.e., those measuring internalizing and externalizing behaviours) were problematic from a measurement perspective, given their low correlation with age. Nevertheless, these items were retained in the final measure due to content coverage requirements.

Correlations with concurrent validation measures revealed strong to moderate relationships in the anticipated direction in most cases (Annex G). For example, the correlation of the IRT scale score with child age was strong, positive and approximately linear across the 24-59 month age range. The main exceptions were the relationship of the IRT scale score with exposure to early stimulation in both countries (weak but positive correlations with the ECDI2030), and the relationship with wealth index quintiles in both countries (significant but small differences). However, comparing children from the wealthiest households with those from the poorest showed anticipated differences in ECDI2030 scores.

Standard-setting

Items on the ECDI2030 capture specific developmental constructs nested within the three domains of health, learning and psychosocial well-being. Because the intention was to generate a single summary score reflecting the interlinkages among these domains, and the holistic nature of ECD more broadly, it was necessary to define some criteria in order to transform the summative score obtained from the 20 items into a performance standard for classifying children as 'developmentally on track'. These specifically focused on defining expectations or 'standards' of what minimally on track children should be able to do by ages 24, 36 and 48 months, thereby reflecting accumulated development up to that age. The main methods and results of the standard-setting are outlined below and documented in greater detail elsewhere.[59]

To establish performance standards on the ECDI2030, a criterion-referenced standard-setting exercise using a modified Angoff approach was carried out. The Angoff method (and its variations) is a common and widely employed test-centred methodology for establishing criterion-referenced cut scores.[60] It was selected over other methodologies because having field test data from only two countries was not considered sufficient to generate norm-referenced standards. That said, all the psychometric information, including item parameters, was used to inform the standard-setting exercise, and experts calibrated their expectations considering the empirical information (i.e., impact data).

In the Angoff method, subject matter experts (SMEs) are tasked with reviewing each item, determining the knowledge or skills required to answer the item correctly and then making a judgement as to how a subject, at the minimum threshold for a performance level, will likely respond.

The first step of the standard-setting process was to conduct a pilot in Mexico in which six SMEs met for 1.5 days and practised this methodology using items previously discarded from the draft item set for the ECDI2030. The main recommendation generated by the pilot was to utilize a modified Angoff approach in which SMEs indicated the likelihood that the subject at each performance level would answer the item correctly.[61] For the ECDI2030 specifically, this translated into the following key questions:

- For items requiring a yes/no response: If you asked 100 mothers/caregivers of children aged 24 months (for example) who were minimally on track in their development, how many of them would endorse the item?

- For items that had more than two response options: If you asked 100 mothers/caregivers of children aged 24 months (for example) who were minimally on track in their development, how many out of 100 would mark each response option?

Following the pilot, a panel of 15 global SMEs convened (virtually and in-person) in August 2019. The members (Annex H) were identified based on nominations from the IAEG-ECD. They were all experts in ECD/developmental psychology and possessed significant experience in conducting standardized assessments of children under age 5 in primary or tertiary health care and/or educational settings, worked in clinical, educational or research settings and collectively represented 13 different countries spanning five continents.

To begin, training was provided in order for SMEs to develop a shared understanding of what it means for a child to be considered developmentally on track at each age (informed by their own experience and knowledge) and to understand how to translate their expectations into a standard-setting judgement. The panel also participated in some practice sessions prior to undertaking two rounds of standard-setting for the 36 items from the draft set of the ECDI2030.

During Round 1, SMEs individually reviewed the items and submitted their ratings via an online survey form. Following this, there was a facilitated discussion of selected items where greater disparity in ratings was observed as well as of items where there was a high degree of agreement. Additionally, the estimated impact (i.e., the percentage of children who would be identified as 'on track' by applying the group's recommended cut scores) using field test data from Mexico and State of Palestine was also presented to the group. During Round 2, experts completed a second round of ratings, informed by the feedback and discussion from the first round.

Values were summed across all items for each expert to determine their individually recommended cut score for each age group. The group's recommended cut score (mean, median, range) and variability was also calculated for each round and each age (Table 5). For example, in Round 1, the panel's recommended mean cut score for a 24-month-old who is on track was 7.0 items out of 36.

The standard-setting exercise was carried out based on the 36 items that were retained after the initial item screening. However, the psychometric analyses described above had reduced these to 20 items. Moreover, during subsequent analyses of the psychosocial domain, items were dropped and two items ("Gets along with other children" and "Seems to be sad or depressed") from the candidate item pool were reinstated to form the final measure.

As a result, only 18 of the 20 items on the final ECDI2030 were included in the standard-setting exercise. A calibration method was therefore applied to determine the difference in difficulty between the set of 18 items and the final 20-item ECDI2030, and then this relationship was used to obtain the standard-setting recommendations for the final ECDI2030. Table 6 shows the standard-setting results after calibrating the complete set of 20 items.

Identification of the final set of performance standards
The last step in identifying the performance standards for the ECDI2030 involved establishing the final cut scores to identify children developmentally on track. To inform this decision, field test data from Mexico and State of Palestine were once again used to generate performance profiles according to different expectations of children's performance based on the average calibrated cut scores.

Table 5
Standard-setting results by round

	Mean	Median	Standard error	Range
Round 1				
24 months	7.0	4.7	1.4	4.3 - 9.7
36 months	17.1	16.4	1.3	14.5 - 19.6
48 months	26.7	26.6	1.0	24.7 - 28.8
Round 2				
24 months	7.7	6.8	1.1	5.5 - 10
36 months	18.4	18.1	0.9	16.5 - 20.2
48 months	28.0	28.5	0.8	26.4 - 29.5

Table 6
Calibrated standard-setting results (20 items)

	Min	Max	Mean	Median	Standard error	Range
24 months	3.6	11.5	5.6	5.1	0.5	4.5 - 6.6
36 months	9.1	14.1	10.9	10.5	0.4	10 - 11.7
48 months	13.4	18.0	15.9	16.4	0.3	15.2 - 16.6

The proportion of developmentally on track children rendered by each performance profile was reviewed in each country, taking into account the underlying expectation about children's performance in general and a number of key disaggregation variables. The identification of the final cut scores also took two additional issues into consideration:

- A requirement on the minimum number of items within each domain: The idea of requiring a minimum number of items within each domain was to ensure that children could not be classified as 'on track' if they were lacking proficiency in one or more domains. It was decided, however, that this was not necessary because content coverage is already a core attribute of the ECDI2030 and was addressed in several ways throughout

the different stages of designing the measure. Moreover, further exploration of the field test data suggested that there were very few children who were lacking proficiency in one domain but were still classified as on track overall. Therefore, it was determined that the application of a single cut score for each age range best supported the holistic nature of ECD.

- The need to include intermediate cut scores for children between the ages of 24 and 35 months and 36 and 47 months: Given how quickly development occurs during these early ages, it was recognized that many children in these age groups might be identified as 'on track' if they were in the latter part of the year range (e.g., a 32-month-old judged against the expectations for a 24-month-old).

Therefore, the panel recommended the inclusion of different cut scores by 6-month age groups for the younger children.

These additional considerations informed calibration of the final set of standards from within the ranges recommended by the global panel. For the 24-month-old standard, the upper end of the recommended range (6.6 rounded to 7) was selected on the basis of the estimated impact from the field test data. Similarly, for the 48-month-old standard, the lower end of the recommended range (15.2 rounded to 15) was chosen based on the estimated impact. Finally, the intermediate performance standards were identified as the median score within the range (i.e., the difference between the performance standard at the start of an age level and the performance standard for the next age level). Thus defined, the final age-specific cut scores adopted to identify children developmentally on track were:

• Children aged 24-29 months: 7 of the 20 items

• Children aged 30-35 months: 9 of the 20 items

• Children aged 36-41 months: 11 of the 20 items

• Children aged 42-47 months: 13 of the 20 items

• Children aged 48-59 months: 15 of the 20 items

Figure 7 illustrates the impact of the application of these cut scores to the field test data from State of Palestine. The dots in the chart represent the probabilities of obtaining different ECDI2030 scores for children between 24 and 59 months. Each dot represents a month of age, and different colours are used to depict the probability for certain age groups (e.g., light blue dots for children aged 24 months and yellow dots for children aged 59 months). As expected, the chart shows that the ECDI2030 score is a function of age, with older children being more likely to score higher. Likewise, as ECDI2030 scores increase, the probabilities of obtaining such scores decrease, also as a function of age. In addition to this, the chart also identifies how the final age-specific cut scores are reflected in these probabilities. It indicates that 84 per cent of children aged 24 to 59 months in the country are developmentally on track in health, learning and psychosocial well-being.

The reasonableness of the final cut scores will be monitored and assessed in light of the availability of global data once more countries have implemented the ECDI2030. With the collection of data from a sufficiently large enough number of countries, age-specific norms could be established, thus producing a set of norm-referenced standards to compare to the criterion-referenced standards determined by the standard-setting exercise.

Figure 7

Predicted probabilities for ECDI2030 score by month of age and proportion of children developmentally on track according to final cut scores, State of Palestine
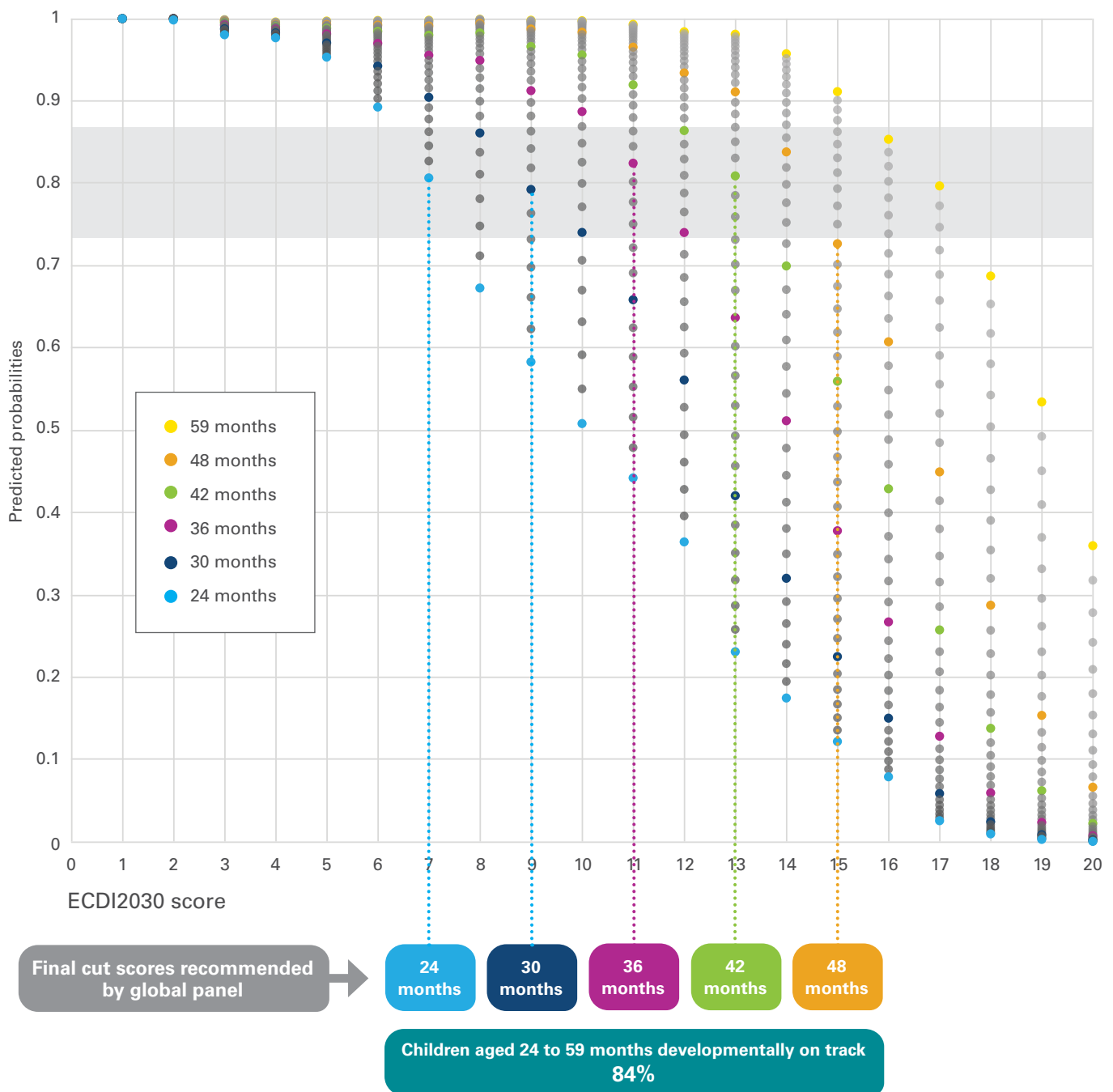
PART FOUR

The final ECDI2030

The final ECDI2030 is comprised of 20 items covering 3 domains and 12 subdomains (Figure 8). The learning domain is made up of 5 subdomains as measured by 11 items, the psychosocial well-being domain is made up of 4 subdomains as measured by 5 items and the health domain has 3 subdomains measured by 4 items.

Figure 8
Content coverage of the ECDI2030

3 DOMAINS

12 SUBDOMAINS

Learning

EXPRESSIVE LANGUAGE

NUMERACY

LITERACY

PRE-WRITING

EXECUTIVE FUNCTION

Psychosocial well-being

EMOTIONAL DEVELOPMENT

SOCIAL DEVELOPMENT

EXTERNALIZING

INTERNALIZING

Health

SELF-CARE

GROSS MOTOR DEVELOPMENT

FINE MOTOR DEVELOPMENT

20 items
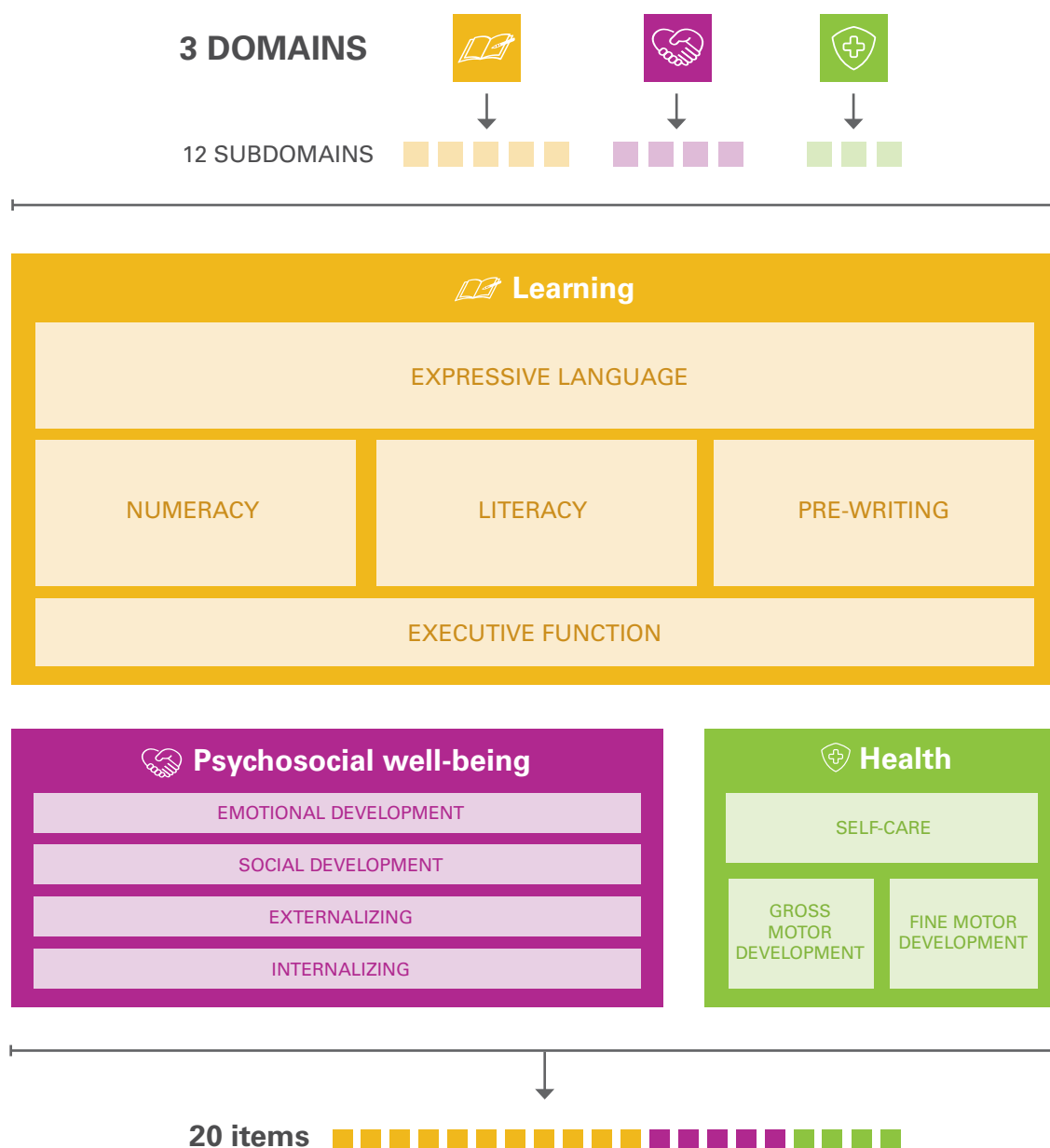
Figure 9 shows the age coverage of the ECDI2030 and the variation of item difficulty across age groups. The main objective of the chart is to exemplify how the final set of items allows for capturing the acquisition of milestones across domains and ages.

For each item, the chart shows the percentage of children for whom the item was endorsed or, in other words, the percentage of children who 'pass' the item. The different colours quantify the proportion of children passing the item by month of age: purple shows the age band by when 25 per cent to

50 per cent of children have passed the item and thus corresponds to the most difficult items in the ECDI2030; teal shows the age band by when 50 per cent to 75 per cent of children have passed the item; and dark blue shows the age band by when 75 per cent to 90 per cent of children have passed the item and thus corresponds to the less difficult items in the ECDI2030.

Even though a pronounced variation by age for the items in the psychosocial well-being domain is not expected, the probabilities of endorsement are also

Probabilities of endorsement of the final ECDI2030 items by month of age, Mexico field test data
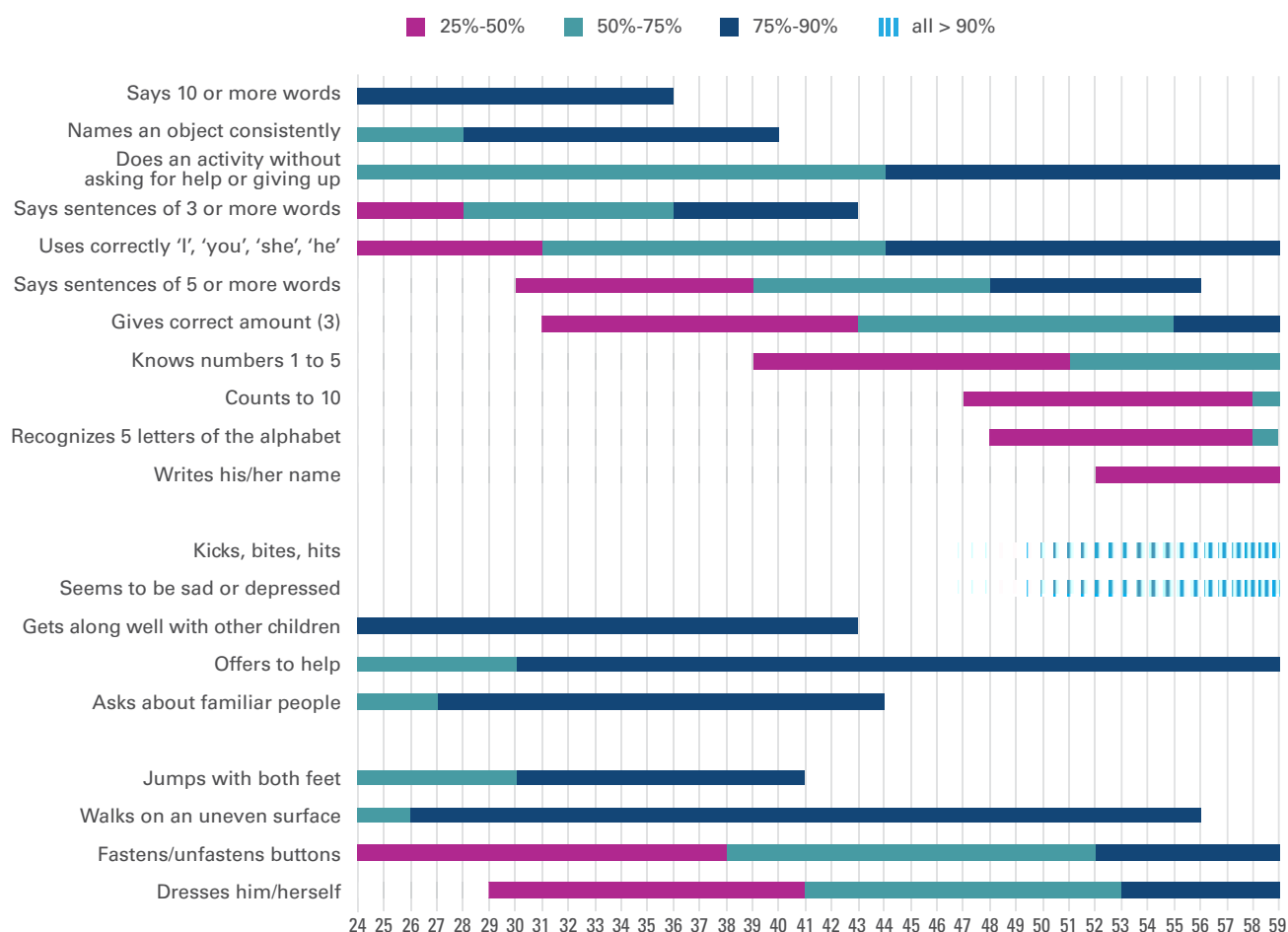
Figure 10
Calculation of the ECDI2030 indicator

The 20 items in the ECDI2030 are used to generate **one single indicator**

$$\text{ECDI2030} = \frac{\text{Number of children aged 24 to 59 months who have achieved the minimum number of milestones expected for their age group}}{\text{Total number of children aged 24 to 59 months}}$$

SDG 4.2.1

Proportion of children aged 24 to 59 months who are developmentally on track in health, learning and psychosocial well-being

SDG indicator 4.2.1

Total indicator value and disaggregated values

81%

71%

Total on track 75%

Poorest children

Richest children

24 to 29 months · 30 to 35 months · 36 to 41 months · 42 to 47 months · 48 to 59 months

7 · 9 · 11 · 13 · 15

MINIMUM NUMBER OF MILESTONES EXPECTED BY AGE GROUP

---

shown for these items. In this domain, a blue stripped bar is used for two items ("Kicks, bites, hits" and "Seems to be sad or depressed"), which measure manifest behaviours rather than developmental milestones.

The 20 items are used to calculate a single score by applying the final set of performance standards that were adopted. By way of example, the mother of a child between the ages of 24 and 29 months would need to endorse at least 7 of the 20 items on the

ECDI2030 in order for that child to be classified as 'developmentally on track'. The resulting data reflect the proportion of children aged 24 to 59 months who are developmentally on track in health, learning and psychosocial well-being (i.e., SDG indicator 4.2.1). And because the data can be disaggregated by key demographics (such as wealth quintiles, as illustrated in Figure 10) and subnational areas, the use of the ECDI2030 to measure indicator 4.2.1 can also help advance the SDG commitment to leave no one (child) behind.

Annexes

Annex A. Membership of the Inter-agency Expert Group on ECD Measurement (IAEG-ECD) and Technical Advisory Group (TAG)

IAEG-ECD members

Claudia Cappa, Senior Adviser, Statistics and Monitoring, Division of Data, Analytics, Planning and Monitoring, UNICEF (Chair)

Marta Rubio Codina, Senior Economist and Child Development Specialist, Inter-American Development Bank

Rami Al Dibs, Director of Health Statistical Department, Palestinian Central Bureau of Statistics

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

Alaka Holla, Senior Economist and Program Manager, Strategic Impact Evaluation Fund, World Bank Group

Dafna Kohen, Assistant Director, Health Analysis Division, Statistics Canada

Silvia Montoya, Director, UNESCO Institute for Statistics

Rowena Phair, Project Leader, Education and Social Progress, OECD

Lauren Pisani, (former) Advisor in Learning Research, Save the Children

Adelle Pushparatnam, Education Specialist, World Bank Group

Ruth Argelia Vázquez Salas, Researcher, National Institute of Public Health, Mexico

TAG members

Santiago Cueto, Principal Investigator, Grupo de Analisis para el Desarrollo (GRADE), Peru

Edward Frongillo, Professor and Director, Global Health Initiatives, University of South Carolina, United States

Melissa Gladstone, Senior Lecturer in Neurodevelopmental Paediatrics and International Child Health, University of Liverpool, United Kingdom

Peter Halpin, Associate Professor, Quantitative Methods, University of North Carolina, United States

Magdalena Janus, Professor, Department of Psychiatry and Behavioural Neurosciences, Offord Centre, McMaster University, Canada

Gillian Lancaster, Professor of Medical Statistics, School of Medicine, Keele University, United Kingdom

Dana Charles McCoy, Associate Professor, Graduate School of Education, Harvard University, United States

Abigail Raikes, Assistant Professor and Director, Global Early Childhood Development Research, University of Nebraska Medical Center, United States

Nirmala Rao, Professor, Faculty of Education, The University of Hong Kong, Hong Kong, China

Ghassan Shakhshir, Chair of the Advisory Expert Group, Palestine National Child Institute, An Najah University, State of Palestine

Hollie Hix-Small, Associate Professor of Special Education and Coordinator of the Early Intervention/Early Childhood Special Education (EI/ECSE) Program, Portland State University, United States

Annex B. Subdomains and constructs in the learning and psychosocial well-being domains

Domain	Subdomain	Construct	Definition and importance to ECD
Learning	Early numeracy	Counting (verbal and set production)	Rote counting refers to children having knowledge of the names and order of counting words. However, this does not indicate an understanding of quantity (i.e., just because a child can count to five does not mean he/she can count five objects). Being able to accurately count requires a number of skills that include knowledge of the counting word sequence; an understanding of one-to-one correspondence; knowledge that the last number word in the count is the cardinal value of the set; knowledge that objects can be counted in any order; and an understanding that any collection of objects can be counted.[62] Along with other skills related to number sense, both verbal counting and set production are predictive of later achievement in mathematics.[63]
		Number identification	Number identification, which represents the ability to identify numerals and understand their corresponding numericities, is important to measure because this skill is at the core of many other numerical competencies.[64] At its most basic level, it is the ability to form an accurate association between the written word of a number and its symbol equivalent (e.g., knowing that 'two' and '2' have the same meaning).[65] Numerical identification has been found to be predictive of later math abilities.[66]
		Number/size discrimination	Number/size discrimination refers to a child's ability to compare things of different numbers or size, e.g., "that spoon is bigger" or "she is taller than he is". This ability to discriminate is important because measurement-related procedures have the potential to later serve as cognitive tools that children can use to organize the way they reason in math problem-solving.
	Language, literacy and communication	Letter/alphabet knowledge	This construct involves skills such as the ability to recognize and name letters and knowledge of the sounds associated with them. Development of these skills during early childhood is critical since this knowledge forms the foundation of later literacy skills such as reading and spelling. In fact, among other reading-readiness skills, letter identification has been recognized as the strongest predictor of later literacy achievement.[67]
		Phonological awareness	The concept of phonological awareness has been defined as "a skill that allows kids to recognize and work with the sounds of spoken language".[68] In the preschool years, this can be demonstrated by children's abilities to identify words that rhyme or to clap out the number of syllables in a name. Development of phonological awareness is crucial given its strong and consistent links with early reading and spelling success in every language in which it has been researched.[69]
		Expressive language	This construct is defined as "the ability to communicate verbally with others".[70] There is great variety in the development of expressive language capacities, resulting in a range of what is defined as 'normal' development.[71] Measuring expressive language is especially important because children who demonstrate difficulty in expressing themselves have been found to be at greater risk for language, social and academic problems in later years.[72]
		Receptive language	This construct refers to the ability to hear, listen and comprehend oral information. Children who have difficulties understanding others may find it challenging to follow instructions and/or appropriately respond to questions or requests. Having foundational oral language skills are also critical for general reading achievement at older ages.[73]
	Executive function	Mental flexibility	This set of skills helps children to both sustain attention when it is required as well as shift attention in response to different demands and settings. Sustained attention has been found to modestly, but consistently, predict academic achievement outcomes, both in preschool and in the early grades, even after controlling for the effects of other cognitive and language abilities.[74] Children's inability to pay attention or 'do what is expected of them' has been associated with anti-social behaviours, peer rejection and lower academic achievement.[75]

Domain	Subdomain	Construct	Definition and importance to ECD
Learning	Executive function	Working memory	Working memory refers to the ability to hold information in the mind and manipulate that information in order to perform tasks.[76] It is an important component of ECD since, along with attention, it has been found to be associated with academic outcomes and years of schooling, with stronger relationships documented in the early childhood period.[77] Working memory was also identified as the best predictor of reading, spelling and math skills six years after assessment.[78]
		Inhibition	Inhibition, also known as inhibitory control, has been identified as one of the first executive function skills to emerge, developing quickly during the preschool years.[79] Measuring inhibition during ECD is important because control skills help children suppress inappropriate behaviours and thoughts that may distract them from instruction.[80] Further, when measured prior to kindergarten age, inhibitory control skills predict math skills in kindergarten. When measured in kindergarten, inhibitory control predicts math and early literacy skills in primary school.[81]
Psycho-social well-being	Social development	Social competence	Social competence is a construct broadly defined as "children's ability to be effective in their social interactions with respect to achieving their goals".[82] It is distinct from emotional competence or regulatory competence in that it is the enactment, or behavioural manifestation, of other competencies.[83] Social competence is a composite term, representing a multitude of skills, which is advantageous in that it leads to many broad, describable and long-term outcomes but is disadvantageous in that many of the associated skills are interlinked and therefore difficult to differentiate from one another. A substantial body of evidence links social competence to a range of later life outcomes and demonstrates how it affects children across other developmental domains.[84] It has also been found that children with early social competence skills are more likely to have higher IQs, positive self-worth and better mental health.[85] Lastly, social competence (along with emotional competence) provides the foundation for mastery of a wide range of skills crucial for successful academic achievement.[86]
		Relationship skills	Relationship skills have been defined as the "ability to establish and maintain healthy and rewarding relationships with diverse individuals and groups. The ability to communicate clearly, listen well, cooperate with others, resist inappropriate social pressure, negotiate conflict constructively, as well as seek and offer help when needed".[87] Relevant group relations in the preschool years include interaction beyond the dyad, friendships, gender segregation, dominance hierarchies and peer acceptance and rejection.[88] Measuring relationship skills is especially important since a bi-directional influence is said to exist between peer relationships and psychosocial and emotional disorders in children. That is, peer problems may contribute to the onset of disorders, while children with disorders may find themselves "at odds with their peers from the very first years of life".[89]
		Interpersonal conflict resolution	Interpersonal conflict resolution refers to the strategies and methods used by individuals and groups to peacefully negotiate interpersonal disputes. It includes the ability to express emotions relating to an interpersonal conflict, e.g., sharing a toy. Interpersonal conflict resolution skills in the early years develop when a child learns to inhibit aggressive behaviours. While aggression manifests early and declines over the first five years, it is considered abnormal when it is pervasive, frequent and severe.[90] Capturing interpersonal conflict resolution skills is necessary because of its salience on later aggression and its impact on a child's later life outcomes. Further, persistent and highly aggressive behaviour is associated with co-occurring language problems, impulsivity, hyperactivity, poorly regulated negative emotions and defiance. The directionality of these associations, however, is not clear.[91]

Domain	Subdomain	Construct	Definition and importance to ECD
Psycho-social well-being	Social development	Prosocial behaviour	Prosocial behaviour, also termed as helpful or kind behaviour, has been defined as "voluntary actions specifically intended to benefit or improve the well-being of another individual or group of individuals".[92] Examples of prosocial behaviour include being respectful, helping, sharing, consoling, comforting, cooperating and protecting others from real or potential harm.[93] Prosocial behaviour is considered to be a construct that lies within the moral and cultural dimensions of development and exists across different subnational, national, cultural and faith contexts. Measuring prosocial behaviour in early childhood is especially important because its development and manifestation during these years is associated with peer acceptance, empathy, self-confidence and emotion regulation skills.[94] Prosocial behaviour in early childhood has also been shown to be an antecedent to later positive relationship trajectories.[95]
		Social cognition	Social cognition refers specifically to "the way in which people process, remember, and use information in social contexts to explain and predict their own behaviour and that of others".[96] Skills such as the ability to understand, describe and predict people's mental states all constitute a pathway for children to develop strong social cognition.[97] Social cognition in the early years is critical because its related skills may impact the quality of relationships and school success. Children with better social cognition tend to be better communicators, more socially competent, more popular with peers, happier at school and academically more advanced.[98] Further, the literature shows that children who are not able to discern the thoughts and feelings of others have a higher probability of acting aggressively and experiencing peer rejection.[99]
		Theory of mind	Theory of mind refers to "the understanding of epistemic mental states such as knowledge and belief, as well as motivational mental states such as desire and emotion, and their consequences on people's behaviours".[100] In other words, it is the ability to attribute mental states to both oneself and others and to recognize others' perspectives even if they differ from one's own. It is the cognitive aspect of interpreting other people's intentions, desires and beliefs, while empathy is the emotional aspect. Theory of mind has important links with children's success in social interactions and in academics. Children with strong theory of mind are better communicators, better at resolving conflicts with friends and engage in more complex pretend play. They are rated more socially competent by teachers, are happier in school, more popular with peers and produce school work that is more advanced in some ways.[101]
	Emotional development	Empathy	Empathy is defined as "the ability to perceive, understand, and react to other people's emotions appropriately".[102] It refers specifically to being able to feel what another person is feeling or might be expected to feel. Measuring empathy is especially important because it is a precursor to prosocial behaviours. Empathy is distinguished from theory of mind, which describes the mind of the child comprehending the mental states of others, while empathy refers to the 'heart' of how that child comprehends the mental states of others.[103]
		Emotion knowledge	This construct is closely linked to the concept of emotion recognition, which refers to the "awareness of feelings".[104] Emotion knowledge is the ability to recognize one is experiencing an emotion, comprehending one's emotional experience within the constraints of emotion scripts and the social context, and realizing that one's own inner and outer emotional states may differ from that of others. Emotion knowledge has been recognized as being key for children when moving into the world of peers and getting along with peers and adults.[105] Therefore, a child's emotional knowledge is said to be both a precursor to social competence and a skill that expands with social competence.[106] Further, it has been posited that emotion knowledge is a construct that has predicted academic success.[107] Children who are weak on this construct are unable to discern the thoughts and feelings of others and therefore have a higher probability of behaving aggressively and experiencing peer rejection.[108]

Annex C. Draft set of ECDI2030 items for field testing

Summary	Domain	Subdomain
Knows numbers 1 to 5	Learning	Numeracy
Gives correct amount (3)	Learning	Numeracy
Counts to 5	Learning	Numeracy
Counts to 10	Learning	Numeracy
Knows difference between big/small	Learning	Numeracy
Recognizes 3 letters of the alphabet	Learning	Literacy
Recognizes 5 letters of the alphabet	Learning	Literacy
Recognizes 10 letters of the alphabet	Learning	Literacy
Recognizes 4 simple words	Learning	Literacy
Stacks small objects (3)	Learning	Pre-writing
Fastens/unfastens buttons	Health	Fine motor skills
Writes his/her name	Learning	Pre-writing
Names an object consistently	Learning	Expressive language
Says 10 or more words	Learning	Expressive language
Says 15 or more words	Learning	Expressive language
Says 20 or more words	Learning	Expressive language
Asks using 'what,' 'which,' 'where,' 'who'	Learning	Expressive language
Says sentences of 3 or more words	Learning	Expressive language
Says sentences of 4 or more words	Learning	Expressive language
Says sentences of 5 or more words	Learning	Expressive language
Uses correctly 'I,' 'you,' 'she,' 'he'	Learning	Expressive language
Uses correctly 'on,' 'in,' 'under'	Learning	Expressive language
Asks using 'why'	Learning	Expressive language
Uses past tense	Learning	Expressive language
Identifies 7 objects	Learning	Expressive language
Sings short song or repeats from memory	Learning	Expressive language
Talks about what he/she sees, hears or does	Learning	Expressive language
Does an activity without asking for help or giving up	Learning	Executive function
Stops when told "no" or "stop doing that"	Learning	Executive function
Acts impulsively (reversed)	Learning	Executive function
Follows instructions of more than one step	Learning	Executive function
Switches from one activity to another	Learning	Executive function
Concentrates on activity	Learning	Executive function

Summary	Domain	Subdomain
Pretends object is something else	Learning	Approaches to learning
Kicks, bites, hits	Psychosocial	Externalizing behaviour
Becomes withdrawn, shy	Psychosocial	Internalizing behaviour
Shows respect	Psychosocial	Social skills
Likes to meet new adults	Psychosocial	Social skills
Takes turns	Psychosocial	Social skills
Seeks help	Psychosocial	Social skills
Gets along with familiar children	Psychosocial	Social skills
Shares things	Psychosocial	Social skills
Asks about familiar people	Psychosocial	Social skills
Offers to help	Psychosocial	Emotional skills
Gets interested in a person who is sick	Psychosocial	Emotional skills
Helps with household chores	Psychosocial	Social skills
Says what others like or dislike	Psychosocial	Emotional skills
Calms down after periods of exciting activity	Psychosocial	Emotional skills
Shows when he/she needs to use the bathroom	Psychosocial	Emotional skills
Seems to be sad or depressed	Psychosocial	Internalizing behaviour
Says when he/she is unhappy, sad, angry	Psychosocial	Emotional skills
Destroys things (reversed)	Psychosocial	Externalizing behaviour
Dresses him/herself	Health	Self-care
Jumps with both feet	Health	Gross motor skills
Throws ball or stone	Health	Gross motor skills
Catches ball	Health	Gross motor skills
Walks on an uneven surface	Health	Gross motor skills
Runs without falling or bumping into objects	Health	Gross motor skills

Annex D. Results from field testing of draft set of ECDI2030 items in Mexico and State of Palestine

Percentage of mothers endorsing the items, by child age, Mexico

Domain-subdomain	Construct	Item	Don't know or missing	Age			
				2 years	3 years	4 years	Total
Learning - Expressive language	Single-word verbal communication	**Says 10 or more words**	0.16	84.52	93.70	96.31	91.56
		Says 15 or more words	0.22	56.72	82.20	85.12	74.74
		Says 20 or more words	0.22	38.17	66.57	70.60	58.53
	Early sentence construction	**Says sentences of 3 or more words**	0.02	55.10	89.98	95.75	80.40
		Says sentences of 4 or more words	0.00	38.54	83.07	91.02	71.04
		Says sentences of 5 or more words	0.15	21.05	65.79	82.64	56.81
	Object recognition	**Names an object consistently**	0.00	79.84	92.03	96.99	89.71
		Identifies 7 objects	0.00	92.44	95.73	98.16	95.49
	Person recognition	**Uses correctly 'I,' 'you,' 'she,' 'he'**	0.05	48.27	72.46	84.78	68.73
	Pragmatics	Uses correctly 'on,' 'in,' 'under'	0.01	41.96	72.19	77.77	64.09
	Temporal understanding	Uses past tense	0.00	20.46	56.83	76.62	51.67
	Asking questions	Asks using 'what,' 'which,' 'where,' 'who'	0.20	53.40	77.04	87.58	72.87
		Asks using 'why'	0.00	45.29	82.36	95.02	74.46
	Other	Sings short song or repeats from memory	0.01	65.47	84.27	95.27	81.87
		Talks about what he/she sees, hears or does	0.00	59.02	87.21	96.99	81.26
Learning - Literacy	Letter identification	Recognizes 3 letters of the alphabet	0.13	15.35	29.72	57.34	34.63
		Recognizes 5 letters of the alphabet	0.12	4.79	16.62	35.26	19.22
		Recognizes 10 letters of the alphabet	0.00	1.01	2.37	7.75	3.81
	Reading words	Recognizes 4 simple words	0.11	7.13	8.93	16.39	10.95
Learning - Numeracy	Counting	**Gives correct amount (3)**	0.27	24.06	46.70	72.27	48.14
		Counts to 5	0.19	15.48	42.64	74.65	44.83
		Counts to 10	0.11	3.46	19.97	38.20	20.87
	Number identification	**Knows numbers 1 to 5**	0.21	13.93	30.35	57.57	34.44
	Number/size discrimination	Knows difference between big/small	0.26	60.17	77.96	93.84	77.61
Learning - Pre-writing	Writing	**Writes his/her name**	0.02	2.43	6.91	29.61	13.38
	Object manipulation	Stacks small objects (3)	0.27	85.76	93.35	94.91	91.37
Learning - Executive function	Following directions	Follows instructions of more than one step	0.24	73.14	82.72	83.37	79.76
	Mental flexibility	**Does an activity without asking for help or giving up**	0.00	58.62	73.20	82.71	71.69
		Switches from one activity to another	0.01	72.88	78.57	84.80	78.86
		Concentrates on activity	0.00	84.19	90.82	91.28	88.77

Domain-subdomain	Construct	Item	Don't know or missing	Age			
				2 years	3 years	4 years	Total
Learning - Executive function	Inhibition/control	Stops when told "no" or "stop doing that"	0.00	69.85	71.38	71.53	70.93
		Acts impulsively (reversed)	0.11	44.77	53.69	56.70	51.78
Learning - Approaches to learning play		Pretends object is something else	0.04	69.82	84.07	74.36	75.92
Psychcosocial - Social development	Sharing	Shares things:					
		Always	0.00	39.59	36.99	45.54	40.85
		Sometimes		46.08	54.70	44.32	48.19
		Never		14.33	8.31	10.14	10.96
	Respect	Shows respect:					
		Always	0.23	44.62	47.05	52.29	48.08
		Sometimes		52.28	49.19	45.38	48.88
		Never		2.89	3.34	2.27	2.81
	Getting along with others	**Gets along with other children**	0.34	83.04	90.17	93.34	88.91
	Helpfulness	**Offers to help**	0.05	69.90	81.68	88.17	80.04
		Helps with household chores:					
		Always	0.04	43.37	46.81	52.10	47.52
		Sometimes		40.68	47.27	43.56	43.77
		Never		15.89	5.92	4.28	8.66
	Attachment/trust	Becomes withdrawn, shy:					
		Never	0.03	25.64	31.71	28.93	28.72
		A few times a year		19.49	15.70	22.48	19.34
		Monthly		23.86	23.60	19.25	22.16
		Weekly		19.89	15.21	18.76	18.01
		Daily		11.12	13.75	10.52	11.74
		Seeks help:					
		Always	0.19	57.80	53.04	53.47	54.77
		Sometimes		37.96	41.56	41.38	40.30
		Never		4.17	5.40	4.68	4.74
	Other	Likes to meet new adults	1.15	64.66	65.42	69.76	66.69
Psychosocial - Emotional development	Empathy	Gets interested in a person who is sick:					
		Always	0.60	35.00	48.65	51.42	45.08
		Sometimes		42.99	43.09	44.11	43.41
		Never		20.85	8.11	3.99	10.90

Domain-subdomain	Construct	Item	Don't know or missing	Age 2 years	3 years	4 years	Total
Psychosocial - Emotional development	Aggressiveness	**Kicks, bites, hits:**					
		Not at all	0.02	41.83	51.50	62.63	52.19
		The same or less		38.73	39.19	29.45	35.62
		More		13.07	5.69	5.63	8.13
		A lot more		6.33	3.62	2.29	4.05
		Destroys things (reversed)	0.00	68.51	68.43	71.10	69.39
	Self-control/ regulation	Shows when he/she needs to use the bathroom	0.00	74.85	92.81	93.51	87.08
		Says when he/she is unhappy, sad, angry	0.05	41.42	66.04	80.13	62.79
	Emotion regulation	Calms down after periods of exciting activity	0.03	75.99	76.87	83.15	78.78
	Patience	Takes turns	0.61	42.78	54.35	64.33	54.00
	Affect	**Seems to be sad or depressed:**					
		Never	0.00	57.93	56.27	47.78	53.84
		Sometimes		41.65	41.12	49.90	44.38
		Always		0.42	2.61	2.33	1.78
	Emotion recognition	**Asks about familiar people**	0.00	78.61	92.30	94.89	88.65
		Says what others like or dislike	0.24	29.68	55.37	81.91	56.13
Health - Self-care		**Dresses him/herself**	0.05	24.94	54.79	74.50	51.77
		Fastens/unfastens buttons	0.04	36.68	54.59	77.88	56.80
		Jumps with both feet	0.40	73.79	90.81	97.25	87.41
Health - Gross motor		Throws ball or stone	0.00	96.74	97.57	99.58	98.00
		Catches ball	0.39	56.98	67.94	72.78	65.99
		Walks on an uneven surface	0.32	76.88	85.77	88.87	83.90
		Runs without falling or bumping into objects	0.36	81.65	84.21	94.46	86.95
Health - Stunting		No versus yes	2.32*	79.08	83.43	84.18	81.92

*The child was not measured as he/she was not present or it was not allowed.

Note: Items in bold are in the final ECDI2030.

Percentage of mothers endorsing the items, by child age, State of Palestine

Domain-subdomain	Construct	Item	Don't know or missing	Age			Total
				2 years	3 years	4 years	
Learning - Expressive language	Single-word verbal communication	**Says 10 or more words**	0.00	86.71	92.55	97.13	91.85
		Says 15 or more words	0.32	79.41	86.61	93.59	86.18
		Says 20 or more words	0.12	71.37	85.55	91.84	82.27
	Early sentence construction	**Says sentences of 3 or more words**	0.12	75.30	94.80	98.50	88.68
		Says sentences of 4 or more words	0.00	59.53	86.59	96.95	79.79
		Says sentences of 5 or more words	0.26	39.34	77.15	92.97	68.09
	Object recognition	**Names an object consistently**	0.31	80.11	95.42	96.71	90.09
		Identifies 7 objects	0.00	87.15	94.99	98.56	93.21
	Person recognition	**Uses correctly 'I,' 'you,' 'she,' 'he'**	0.13	63.00	85.73	96.14	80.58
	Pragmatics	Uses correctly 'on,' 'in,' 'under'	0.25	76.42	93.04	95.58	87.62
	Temporal understanding	Uses past tense	0.16	35.01	62.14	73.69	55.71
	Asking questions	Asks using 'what,' 'which,' 'where,' 'who'	0.14	69.79	91.25	94.62	84.28
		Asks using 'why'	0.14	39.42	75.13	87.34	65.69
	Other	Sings short song or repeats from memory	0.14	75.55	87.56	94.61	85.34
		Talks about what he/she sees, hears or does	0.00	67.71	89.78	94.62	83.06
Learning - Literacy	Letter identification	Recognizes 3 letters of the alphabet	1.07	4.67	15.52	55.95	24.63
		Recognizes 5 letters of the alphabet	0.64	1.45	10.47	49.90	19.95
		letters of the alphabet	0.02	0.42	2.84	36.14	12.80
	Reading words	Recognizes 4 simple words	0.83	3.30	11.13	28.28	13.78
Learning - Numeracy	Counting	**Gives correct amount (3)**	3.74	30.28	64.85	87.85	59.36
		Counts to 5	0.95	31.54	61.50	86.10	58.26
		Counts to 10	0.65	15.34	37.01	68.34	39.09
	Number identification	**Knows numbers 1 to 5**	0.36	21.44	41.19	67.47	42.34
	Number/size discrimination	Knows difference between big/small	2.23	56.62	81.55	94.56	76.42
Learning - Pre-writing	Writing	**Writes his/her name**	0.00	3.54	8.46	33.02	14.62
	Object manipulation	Stacks small objects (3)	0.20	83.91	92.42	96.67	90.61
Learning - Executive function	Following directions	Follows instructions of more than one step	0.60	79.94	90.47	89.94	86.34
	Mental flexibility	**Does an activity without asking for help or giving up**	1.35	53.61	72.08	85.81	69.61
		Switches from one activity to another	0.62	76.13	86.64	93.17	84.82
		Concentrates on activity	0.95	82.05	91.23	94.61	88.88

Domain-subdomain	Construct	Item	Don't know or missing	Age			
				2 years	3 years	4 years	Total
Learning - Executive function	Inhibition/control	Stops when told "no" or "stop doing that"	0.01	76.70	76.97	77.50	77.04
		Acts impulsively (reversed)	2.12	58.05	65.06	64.71	62.31
Learning - Approaches to learning play		Pretends object is something else	0.71	76.89	77.68	83.13	79.16
Psychcosocial - Social development	Sharing	Shares things:					
		Always	1.29	37.85	48.27	50.90	45.21
		Sometimes		47.59	43.59	42.20	44.64
		Never		12.84	7.57	5.43	8.85
	Respect	Shows respect:					
		Always	1.48	42.40	64.88	61.07	55.18
		Sometimes		44.74	28.83	33.84	36.45
		Never		9.24	5.90	5.09	6.89
	Getting along with others	**Gets along with other children**	1.13	85.84	89.50	90.74	88.53
	Helpfulness	**Offers to help**	3.30	65.74	76.29	89.59	76.66
		Helps with household chores:					
		Always	0.37	45.93	47.57	45.72	46.35
		Sometimes		39.70	47.12	50.14	45.32
		Never		13.38	5.32	4.14	7.97
	Attachment/trust	Becomes withdrawn, shy:					
		Never	0.99	50.19	50.91	43.81	48.32
		A few times a year		25.94	24.08	32.70	27.59
		Monthly		10.44	11.30	11.59	11.07
		Weekly		7.77	9.89	7.70	8.38
		Daily		4.05	3.31	3.48	3.65
		Seeks help:					
		Always	0.34	52.18	44.87	43.75	47.25
		Sometimes		42.84	47.18	52.14	47.17
		Never		4.08	7.95	4.11	5.24
	Other	Likes to meet new adults	1.49	67.29	79.29	81.05	75.35
Psychosocial - Emotional development	Empathy	Gets interested in a person who is sick:					
		Always	1.43	46.92	61.79	60.25	55.70
		Sometimes		37.48	30.65	33.87	34.27
		Never		12.72	7.44	4.93	8.60

Domain-subdomain	Construct	Item	Don't know or missing	Age 2 years	3 years	4 years	Total
Psychosocial - Emotional development	Aggressiveness	**Kicks, bites, hits:**					
		Not at all	0.27	24.01	32.46	30.39	28.60
		The same or less		58.94	52.06	57.03	56.27
		More		12.51	10.65	9.84	11.08
		A lot more		3.85	4.80	2.75	3.77
		Destroys things (reversed)	0.14	71.77	73.98	79.27	74.88
	Self-control/regulation	Shows when he/she needs to use the bathroom	0.09	60.81	92.70	96.54	81.96
		Says when he/she is unhappy, sad, angry	0.83	66.48	79.94	83.51	76.05
	Emotion regulation	Calms down after periods of exciting activity	0.27	72.27	81.22	81.05	77.80
	Patience	Takes turns	1.24	49.76	53.82	65.07	55.97
	Affect	**Seems to be sad or depressed:**					
		Never	1.21	71.01	74.13	67.41	7076
		Sometimes		27.02	21.52	31.01	26.68
		Always		1.42	1.77	0.87	1.34
	Emotion recognition	**Asks about familiar people**	0.27	88.14	95.44	94.55	92.40
		Says what others like or dislike	3.09	26.91	56.00	67.38	48.77
Health - Self-care		**Dresses him/herself**	0.28	42.17	77.12	90.09	68.21
		Fastens/unfastens buttons	1.11	31.27	66.32	78.70	57.18
		Jumps with both feet	0.67	76.42	95.09	97.82	88.96
Health - Gross motor		Throws ball or stone	0.02	94.47	96.93	98.41	96.49
		Catches ball	1.01	80.67	94.21	93.22	88.79
		Walks on an uneven surface	0.86	73.55	91.88	94.54	85.86
		Runs without falling or bumping into objects	0.06	72.05	84.64	90.54	81.83

Note: Items in bold are in the final ECDI2030.

Annex E. Results of the decision-matrix

	Flw	Domain	Result according to decision-matrix
Recognizes 3 letters of alphabet	Green	Learning	Item passed all criteria but was dropped as it was considered that another item made it redundant
Counts to 5	Yellow	Learning	Item was too easy in CREDI and HKU and not discriminant in NLSCY 2008
Likes to meet new adults	Yellow	Psychosocial	Item was not discriminant in Mexico and HKU
Gets along with other children	Yellow	Psychosocial	Item was not discriminant in Mexico
Shows when he/she needs to use the bathroom	Yellow	Psychosocial	Item was not discriminant in Mexico and HKU
Catches ball	Yellow	Health	Item was too easy for children aged 3 to 4 years
Recognizes 10 letters of alphabet	Red	Learning	Item was too difficult and with little variation across ages in Mexico, State of Palestine and some MELQO countries
Recognizes 4 simple words	Red	Learning	Item was too difficult in all countries except HKU
Stacks small objects (3)	Red	Learning	Item was too easy in all countries with almost no variation across ages
Says 15 or more words	Red	Learning	Item was too easy and less discriminant than another similar item
Says sentences of 4 or more words	Red	Learning	Item was too easy in Belize and Mexico
Identifies 7 objects	Red	Learning	Item was too easy and with little variation with age in Mexico, State of Palestine and HKU
Stops when told "no" or "stop doing that"	Red	Learning	Item was not discriminant in most countries
Acts impulsively (reversed)	Red	Learning	Item was not discriminant in most countries
Switches from one activity to another	Red	Learning	Item was not discriminant in Mexico
Concentrates on activity	Red	Learning	Item was not discriminant in Mexico and CREDI
Pretends object is something else	Red	Learning	Item was not discriminant in most countries
Becomes withdrawn, shy	Red	Learning	Item was not discriminant in any country
Seeks help	Red	Psychosocial	Item was not discriminant in any country
Seems to be sad or depressed	Red	Psychosocial	Item was not discriminant in most countries
Destroys things (reversed)	Red	Psychosocial	Item was not discriminant in most countries
Throws ball or stone	Red	Psychosocial	Item was too easy and with low discrimination in many countries

Annex F. Results of ATA in Mexico and State of Palestine and the final version of the ECDI2030

Item	ATA Mexico	ATA State of Palestine	Final ECDI2030
Knows numbers 1 to 5	+	+	+
Gives correct amount (3)	+	+	+
Counts to 10	+	+	+
Recognizes 5 letters of the alphabet	+	+	+
Fastens/unfastens buttons	+	+	+
Writes his/her name	+	+	+
Names an object consistently	–	+	+
Says 10 or more words	+	–	+
Says 20 or more words	–	–	–
Asks using 'what,' 'which,' 'where,' 'who'	–	–	–
Says sentences of 3 or more words	+	+	+
Says sentences of 5 or more words	+	+	+
Uses correctly 'I,' 'you,' 'she,' 'he'	–	+	+
Uses correctly 'on,' 'in,' 'under'	–	+	–
Asks using 'why'	+	–	–
Uses past tense	–	–	–
Sings short song or repeats from memory	–	–	–
Talks about what he/she sees, hears or does	+	–	–
Does an activity without asking for help or giving up	+	+	+
Follows instructions of more than one step	–	–	–
Kicks, bites, hits	+	+	+
Shows respect	–	–	–
Takes turns	–	–	–
Gets along with other children	–	–	+
Asks about familiar people	+	+	+
Offers to help	+	+	+
Gets interested in a person who is sick	+	+	–
Helps with household chores	–	–	–
Calms down after periods of exciting activity	–	–	–
Seems to be sad or depressed	+	+	+
Dresses him/herself	+	+	+
Jumps with both feet	+	+	+
Walks on an uneven surface	+	+	+
Runs without falling or bumping into objects	–	–	–

Note: '+' indicates the item was included on the test form, '-' indicates that the item was not included on the test form.

Annex G. Statistical properties of the final ECDI2030 and associations with child background characteristics, Mexico and State of Palestine

Statistical properties of the final ECDI2030, Mexico

Domain	Item	Percentage of children passing the item			Point-biserial correlations		Item-level analyses	
		2 years	3 years	4 years	Domain	Age	Discrimi-nation	Difficulty
Learning	Can (name) say at least ten or more words like "Mama" or "ball"?	84%	94%	96%	0.79	0.33	1.90	-1.88
	Can (name) speak using sentences of three or more words that go together, for example, "I want water" or "The house is big"?	54%	90%	96%	0.87	0.63	2.95	-0.98
	Can (name) speak using sentences of five or more words that go together?	20%	66%	82%	0.80	0.61	2.22	-0.21
	If you show (name) an object he/she knows well, such as a cup or animal, can he/she consistently name it?	79%	92%	97%	0.72	0.45	1.85	-1.74
	Can (name) correctly use any of the words 'I,' 'you,' 'she,' or 'he', for example, "I go to the store," or "He eats rice"?	47%	72%	85%	0.76	0.44	1.72	-0.68
	Can (name) recognize at least five letters of the alphabet?	4%	17%	35%	0.83	0.53	2.01	1.14
	If you ask (name) to give you three objects, such as three stones or three beans, does (he/she) give you the correct amount?	23%	46%	72%	0.77	0.49	1.61	0.07
	Can (name) count 10 objects, for example, 10 fingers or blocks, without mistakes?	3%	19%	39%	0.83	0.51	1.95	1.08
	Does (name) know all numbers from 1 to 5?	14%	29%	58%	0.76	0.51	1.59	0.58
	Can (name) write his/her own name?	2%	7%	29%	0.80	0.60	1.84	1.52
	Can (name) do an activity such as colouring without repeatedly asking for help or giving up too quickly?	59%	73%	82%	0.60	0.30	0.90	-1.20
Psychosocial well-being	Does (name) get along well with other children?	83%	90%	93%	0.85	0.22	0.77	-3.17
	Does (name) ask about familiar people other than parents when they are not there, for example, "Where is Grandma?"?	78%	92%	95%	0.81	0.37	1.48	-1.86
	Does (name) offer to help someone who seems to need help?	69%	82%	88%	0.90	0.27	0.98	-1.68
	How often does (name) seem to be very sad or depressed?	98%	97%	98%	0.49	-0.11	0.03	-146.01
	Compared with children of the same age, how much does (name) kick, bite or hit other children or adults?	81%	90%	92%	0.83	0.23	0.62	-3.41
Health	Can (name) dress him/herself, that is, put on pants and shirt without help?	25%	55%	74%	0.87	0.51	1.26	-0.07
	Can the child fasten and unfasten buttons without help?	36%	54%	78%	0.94	0.44	1.19	-0.29
	Can (name) jump up with both feet leaving the ground?	74%	90%	97%	0.82	0.50	1.10	-2.12
	Can (name) walk on an uneven surface, for example, a bumpy or steep road, without falling?	77%	86%	88%	0.72	0.22	0.75	-2.44

Associations between the final ECDI2030 sum score and child background characteristics, Mexico

	Mean	Lower bound	Upper bound
Mexico	13.2	12.9	13.5
24 to 35 months	10.1	9.6	10.6
36 to 41 months	13.4	13.1	13.7
42 to 59 months	15.8	15.4	16.1
Girls	13.6	13.0	14.1
Boys	12.5	12.4	13.2
Mother's education level primary or less	13.1	12.7	13.7
Mother's education level upper secondary or higher	13.4	12.4	14.2
Children not attending early childhood education[1]	12.9	12.4	13.3
Children attending early childhood education[1]	15.2	14.9	15.5
Children without early stimulation	11.7	11.0	12.4
Children with early stimulation	13.6	13.3	14
Children without books	12.4	12.0	12.8
Children with books	14.7	14.3	15.2
Children who are stunted	12.6	11.8	13.4
Children who are not stunted	13.3	12.9	12.7
Children living in poorest 20% of households	13.1	12.6	13.3
Children living in richest 20% of households	13.9	13.2	14.4

[1] These results refer to children aged 36 to 49 months only.

Statistical properties of the final ECDI2030, State of Palestine

Domain	Item	Percentage of children passing the item			Point-biserial correlations		Item-level analyses	
		2 years	3 years	4 years	Domain	Age	Discrimi-nation	Difficulty
Learning	Can (name) say at least ten or more words like "Mama" or "ball"?	86%	93%	97%	0.79	0.44	1.48	-2.17
	Can (name) speak using sentences of three or more words that go together, for example, "I want water" or "The house is big"?	75%	95%	99%	0.87	0.65	2.98	-1.40
	Can (name) speak using sentences of five or more words that go together?	39%	77%	93%	0.80	0.69	3.45	-0.51
	If you show (name) an object he/she knows well, such as a cup or animal, can he/she consistently name it?	80%	95%	97%	0.72	0.57	2.46	-1.58
	Can (name) correctly use any of the words 'I,' 'you,' 'she,' or 'he', for example, "I go to the store," or "He eats rice"?	63%	86%	96%	0.76	0.65	3.86	-0.92
	Can (name) recognize at least five letters of the alphabet?	2%	11%	50%	0.83	0.49	4.97	0.87
	If you ask (name) to give you three objects, such as three stones or three beans, does (he/she) give you the correct amount?	30%	65%	88%	0.77	0.61	2.26	-0.29
	Can (name) count 10 objects, for example, 10 fingers or blocks, without mistakes?	15%	37%	68%	0.83	0.57	2.96	0.31
	Does (name) know all numbers from 1 to 5?	21%	41%	67%	0.76	0.54	2.29	0.24
	Can (name) write his/her own name?	3%	9%	33%	0.80	0.38	2.07	1.36
	Can (name) do an activity such as colouring without repeatedly asking for help or giving up too quickly?	54%	72%	86%	0.60	0.56	1.53	-0.76
Psychosocial well-being	Does (name) get along well with other children?	86%	89%	91%	0.85	0.35	0.70	-3.16
	Does (name) ask about familiar people other than parents when they are not there, for example, "Where is Grandma?"?	88%	95%	95%	0.81	0.44	1.36	-2.34
	Does (name) offer to help someone who seems to need help?	66%	76%	90%	0.90	0.44	1.13	-1.30
	How often does (name) seem to be very sad or depressed?	99%	98%	99%	0.52	0.07	0.21	-20.67
	Compared with children of the same age, how much does (name) kick, bite or hit other children or adults?	84%	84%	88%	0.78	0.05	0.32	-5.52
Health	Can (name) dress him/herself, that is, put on pants and shirt without help?	42%	77%	91%	0.86	0.50	1.17	-0.82
	Can the child fasten and unfasten buttons without help?	31%	66%	79%	0.84	0.52	1.23	-0.30
	Can (name) jump up with both feet leaving the ground?	76%	95%	98%	0.78	0.43	1.50	-1.87
	Can (name) walk on an uneven surface, for example, a bumpy or steep road, without falling?	73%	92%	95%	0.87	0.43	1.57	-1.60

Associations between the final ECDI2030 sum score and child background characteristics, State of Palestine

	Mean	Lower bound	Upper bound
State of Palestine	14.1	13.7	14.4
24 to 35 months	11.1	10.6	11.6
36 to 41 months	14.3	13.7	14.8
42 to 59 months	16.4	16.0	16.8
Girls	14.5	14.0	14.9
Boys	13.6	13.2	14.1
Children not attending early childhood education[1]	14.8	14.4	15.2
Children attending early childhood education[1]	17.7	17.3	18.1
Children without early stimulation	13.0	12.2	13.9
Children with early stimulation	14.2	13.8	14.5
Children without books	13.5	13.1	13.8
Children with books	16.2	15.5	16.8
Children living in poorest 20% of households	13.4	12.8	14.0
Children living in richest 20% of households	15.0	14.2	15.7

[1] These results refer to children aged 36 to 49 months only.

Annex H. Membership of the global panel on ECD standard-setting

Claudia Regina Lindgren Alves, Professor, Department of Pediatrics, School of Medicine, Universidade Federal de Minas Gerais, Minas Gerais, Brazil

Yvonne Becher, Chief Executive/Director, Programme Development and Learning, The Child Development Centre, Hong Kong

Maureen Black, Professor, University of Maryland School of Medicine & RTI International, Baltimore, United States

Gauri Divan, Director, Child Development Group, Sangath, Delhi, India

Kirsten Donald, Senior Specialist, Division of Developmental Paediatrics, Red Cross War Memorial Children's Hospital and Deputy Director, Neuroscience Institute, University of Cape Town, Cape Town, South Africa

Melissa Gladstone, Senior Lecturer in Neurodevelopmental Paediatrics and International Child Health, University of Liverpool, Liverpool, United Kingdom

Frances Page Glascoe, Professor of Pediatrics, Vanderbilt University, Nashville, United States

Jennifer Grisham-Brown, Professor and Faculty Director of Early Childhood Laboratory, University of Kentucky, Lexington, United States

Gwendoline Kandawasvika, Senior Consultant, Primary Health Sciences Department, Faculty of Medicine and Health Sciences, University of Zimbabwe, Harare, Zimbabwe

Shazia Maqbool, Professor & Chair Developmental-Behavioural Paediatrics Department, The Children's Hospital and Institute of Child Health, Lahore, Pakistan

Fahmida Tofail, Scientist & Senior Consultant Physician, International Centre for Diarrhoeal Disease Research, Dhakka, Bangladesh

Erika Marcela Osorio Valencia, Head of the Department of Developmental Neurobiology, National Institute of Perinatology, Mexico City, Mexico

Meta Van den Heuvel, MD, PhD, Hospital for Sick Children, Toronto, Canada

Tao Xin, Deputy Director of the National Assessment Center for Education Quality, Ministry of Education, Beijing, China

Pia Zeinoun, (former) Assistant Professor, American University of Beirut, Beirut, Lebanon

Endnotes

1. United Nations Children's Fund, 'Early Childhood Development Index 2030: A new tool to measure SDG indicator 4.2.1', undated, <https://data.unicef.org/resources/early-childhood-development-index-2030-ecdi2030>, accessed 3 January 2022.

2. World Health Organization, United Nations Children's Fund and World Bank Group, *Nurturing Care for Early Childhood Development: A framework for helping children survive and thrive to transform health and human potential*, WHO, Geneva, 2018.

3. *The Lancet*, 'Advancing Early Childhood Development: From science to scale', Series Papers, 4 October 2016, <www.thelancet.com/series/ECD2016>, accessed 3 January 2022.

4. For an overview of these tools, see: Fernald, Lia C. H., et al., *A Toolkit for Measuring Early Childhood Development in Low- and Middle-Income Countries*, World Bank, Washington, D.C., 2017, and the accompanying 'ECD Measurement Inventory', <https://documents.worldbank.org/en/publication/documents-reports/documentdetail/384681513101293811/a-toolkit-for-measuring-early-childhood-development-in-low-and-middle-income-countries>, accessed 3 January 2022.

5. For a detailed overview of the development of the MICS ECDI, see: Loizillon, Anaïs, et al., 'Development of the Early Childhood Development Index in MICS Surveys', MICS Methodological Papers, No. 6, Data and Analytics Section, Division of Data, Research and Policy, United Nations Children's Fund, New York, 2017.

6. Ibid.

7. *The Lancet*, 'Early Child Development in Developing Countries 2011', Series Papers, 24 September 2011, <www.thelancet.com/series/child-development-in-developing-countries-2>, accessed 3 January 2022.

8. United Nations Department of Economic and Social Affairs, Statistics Division, 'Ninth Meeting of the Inter-Agency and Expert Group on the Sustainable Development Goal Indicators, 26-28 March 2019: Report', STA/441/2/162A/3, 22 May 2019, <https://unstats.un.org/sdgs/meetings/iaeg-sdgs-meeting-09/>, accessed 3 January 2022.

9. United Nations Children's Fund, 'Meeting Report: Technical Consultation on SDGs indicator 4.2.1 – Revising the Early Childhood Development Index (ECDI), 14-15 September 2016', UNICEF, New York, 2017, unpublished.

10. Borisova, Ivelina, and Nell O'Donnell, *Early Literacy and Math Toolkit: Teacher training component – Facilitator's guide,* Save the Children USA, New York, 2013.

11. Starr, Ariel, Melissa E. Libertus and Elizabeth M. Brannon, 'Number Sense in Infancy Predicts Mathematical Abilities in Childhood', *Proceedings of the National Academy of Scienc*es, vol. 110, no. 45, 2013, pp. 18116–18120.

12. Child Trends and Center for Child Health Research, *Early Child Development in Social Context: A chartbook*, The Commonwealth Fund, Fairfield, CT, 2004.

13. Duncan, Greg J., et al., 'School Readiness and Later Achievement', *Developmental Psychology*, vol. 43, no. 6, 2007, pp. 1428–1446.

14. Ibid.

15. Devercelli, Amanda Epstein et al., 'Measuring Child Development and Early Learning', Early Learning Partnership Guidance Note, World Bank, Washington, D.C., July 2016, <https://documents.worldbank.org/en/publication/documents-reports/documentdetail/659701473955877219/measuring-child-development-and-early-learning>, accessed 3 January 2022.

16. 'Meeting Report: Technical Consultation on SDGs indicator 4.2.1'.

17. Moore, Kristin Anderson, et al., 'Flourishing from the Start: What is it and how can it be measured?', Child Trends Research Brief, March 2017, <https://www.childtrends.org/publications/flourishing-start-can-measured>, accessed 21 June 2023.

18. *Early Child Development in Social Context.*

19. Duncan et al., 'School Readiness and Later Achievement'.

20. *Early Child Development in Social Context.*

21. Fernald et al., *A Toolkit for Measuring Early Childhood Development in Low- and Middle-Income Countries.*

22. Duncan et al., 'School Readiness and Later Achievement'.

23. *Early Child Development in Social Context.*

24. Halle, Tamara G., et al., *In the Running for Successful Outcomes: Exploring the evidence for thresholds of school readiness*, Office of the Assistant Secretary for Planning and Evaluation, US Department of Health and Human Services, Washington, D.C., 2012.

25. Anderson, Kate, and Abbie Raikes, 'Key Measurement Questions for SDG 4.2.1', Discussion paper for Global Alliance to Monitor Learning (GAML) Task Force 4.2, unpublished, 2017.

26. 'Meeting Report: Technical Consultation on SDGs indicator 4.2.1'.

27. Jurado, Maria-Beatriz, and Monica Rosselli, 'The Elusive Nature of Executive Functions: A review of our current understanding', *Neuropsychology Review*, vol. 17, no. 3, 2007, pp. 213–233.

28. Administration for Children and Families, *Head Start Early Learning Outcomes Framework: Ages birth to five*, US Department of Health and Human Services, Washington, D.C., 2015; Fernald et al., *A Toolkit for Measuring Early Childhood Development in Low- and Middle-Income Countries.*

29. Ibid.

30. Jones, Stephanie M., et al., 'Assessing Early Childhood Social and Emotional Development: Key conceptual and measurement issues', *Journal of Applied Developmental*

Psychology, vol. 45, 2016, pp. 42–48; Fernald et al., *A Toolkit for Measuring Early Childhood Development in Low- and Middle-Income Countries.*

31. Rimm-Kaufman, Sara, and Lia Sandilos, 'School Transition and School Readiness: An outcome of early childhood development', in *Encyclopedia on Early Child Development,* edited by Richard E. Tremblay, Michel Boivin and Ray DeV. Peters, Centre of Excellence for Early Childhood Development, Montreal, 2017, <www.child-encyclopedia.com/school-readiness/according-experts/school-transition-and-school-readiness-outcome-early-childhood>, accessed 3 January 2022.

32. Best, John R., Patricia H. Miller and Jack A. Naglieri, 'Relations between Executive Function and Academic Achievement from Ages 5 to 17 in a Large, Representative National Sample', *Learning and Individual Differences*, vol. 21, no. 4, 2011, pp. 327–336.

33. Glewwe, Paul, Sofya Krutikova and Caine Rolleston, 'Do Schools Reinforce or Reduce Learning Gaps Between Advantaged and Disadvantaged Students? Evidence from Vietnam and Peru', *Economic Development and Cultural Change*, vol. 65, no. 4, 2017, pp. 699-739.

34. *Head Start Early Learning Outcomes Framework.*

35. Rimm-Kaufman and Sandilos, 'School Transition and School Readiness'.

36. Halle, Tamara G., and Kristen E. Darling-Churchill, 'Review of Measures of Social and Emotional Development', *Journal of Applied Developmental Psychology*, vol. 45, 2016, pp. 8–18.

37. Rubin, Kenneth H., and Linda Rose-Krasnor, 'Interpersonal Problem Solving and Social Competence in Children,' in *Handbook of Social Development: A lifespan perspective*, edited by Vincent B. Van Hasselt and Michel Herson, Plenum Press, New York, 1992, pp. 283–323.

38. Save the Children, *Social Emotional Learning Brief*, Save the Children, undated, <https://resourcecentre.savethechildren.net/pdf/226._sc_social_emotional_learning_-_brief.pdf/>, accessed 3 January 2022.

39. *Early Child Development in Social Context.*

40. Anderson Moore et al., *Flourishing from the Start.*

41. Goodman, Alissa, et al., 'Social and Emotional Skills in Childhood and Their Long-term Effects on Adult Life: A review for the Early Intervention Foundation', Institute for Education Cabinet Office Child Poverty and Social Mobility Commission, London, 2015.

42. Halle et al., *In the Running for Successful Outcomes.*

43. Campbell, Susan B., et al., 'Commentary on the Review of Measures of Early Childhood Social and Emotional Development: Conceptualization, critique, and recommendations', *Journal of Applied Developmental Psychology*, vol. 45, 2016, pp. 19–41.

44. Goodman et al., 'Social and Emotional Skills in Childhood and Their Long-term Effects on Adult Life'.

45. Miller, Kristen, et al., eds., *Cognitive Interviewing Methodology*, John Wiley and Sons, Hoboken, NJ, 2014.

46. Cappa, Claudia, et al., 'Identifying and Minimizing Errors in the Measurement of Early Childhood Development: Lessons learned from the cognitive testing of the ECDI2030', *International Journal of Environmental Research and Public Health*, vol. 18, no. 22, 2021, pp. 1-13.

47. Tourangeau, Roger, 'Cognitive Science and Survey Methods: A cognitive perspective', in *Cognitive Aspects of Survey Design: Building a bridge between disciplines*, edited by Thomas B. Jabine et al., National Academy Press, Washington, D.C., 1984.

48. Khan, Shane, M., et al., 'Results of a Multi-Topic Field Test in Belize 2019', *MICS Methodological Papers*, No. 10, Data and Analytics Section, Division of Data, Analytics, Planning and Monitoring, United Nations Children's Fund, New York, 2019.

49. Bayley, Nancy, *Bayley Scales of Infant and Toddler Development*, 3rd ed., Harcourt Assessment, San Antonio, TX, 2006.

50. Inter-American Development Bank, 'Regional Project on Child Development Indicators (PRIDI)', <https://www.iadb.org/en/sector/education/pridi/home#:~:text=WHAT%20IS%20PRIDI%3F,aged%2024%20to%2059%20months>, accessed 3 January 2022.

51. Save the Children, 'International Development and Early Learning Assessment (IDELA)', <https://idela-network.org>, accessed 3 January 2020.

52. Rao, Nirmala, et al., *Developmentally on Track Assessment Scale and Caregiver Survey*, Faculty of Education, The University of Hong Kong, 2020.

53. Bland, J. Martin, and Douglas G. Altman, 'Measuring Agreement in Method Comparison Studies', *Statistical Methods in Medical Research*, vol. 8, no. 2, 1999, pp. 135–160.

54. Information from Bangladesh, China, India and Myanmar (children aged 3 to 4 years old).

55. Information from Bangladesh, Brazil, Cambodia, Chile, Colombia, Ghana, Guatemala, India, Jordan, Lao People's Democratic Republic, Lebanon, Nepal, Pakistan, Philippines, United Republic of Tanzania, United States and Zambia (children aged 2 to 3 years old).

56. Halpin, Peter, et al., 'Monitoring Early Childhood Development at the Population Level: The ECDI2030', preprint, 2023, <https://psyarxiv.com/6qcjb/>, accessed 24 February 2023.

57. Factor analyses were initially performed on all items, but following technical consultation with the TAG and IAEG-ECD, it was agreed that the results from factor loading would lead to similar, but less informative, results as those derived from the discrimination functions.

58. With regards to this constraint, a simple application would require that at least one item from each subdomain be included on the final test form. However, this would result in the final measure being comprised almost entirely of items from the expressive language subdomain. This was deemed an inadequate distribution of items by content domains. Therefore, it was required instead that a minimum of 4 items from the health domain and 5 items from the psychosocial domain be included on the final test form. This ensured that the representation of these two domains on the final version

of the ECDI2030 was approximately proportional to their representation in the screened item pool. Additionally, the maximum number of items from the expressive language subdomain was restricted to 5.

59. Petrowski, Nicole, et al., 'Establishing Performance Standards for Child Development: Learnings from the ECDI2030', submitted for publication, 2023.

60. Angoff, William H., 'Scales, Norms, and Equivalent Scores', in *Educational Measurement*, 2nd ed., edited by Robert L. Thorndike, American Council on Education, Washington, D.C., 1971, pp. 508–600.

61. Plake, Barbara S., and Gregory J. Cizek, 'Variations on a Theme: The modified Angoff, extended Angoff, and yes/no standard setting methods', in *Setting Performance Standards: Foundations, methods and innovations*, 2nd ed., edited by Gregory J. Cizek, Routledge, New York, 2012.

62. Gelman, Rochel, and Charles R. Gallistel, *The Child's Understanding of Number*, Harvard University Press, Cambridge, MA, 1978. As cited in United Nations Educational, Scientific and Cultural Organization, et al., *Overview: MELQO Measuring Early Learning Quality and Outcomes*, UNESCO, Paris, United Nations Children's Fund, New York, and Brookings Institution and the World Bank, Washington, D.C., 2017.

63. Jordan, Nancy C., et al., 'Early Math Matters: Kindergarten number competence and later mathematics outcomes', *Developmental Psychology*, vol. 45, no. 3, 2009, pp. 850–867; Clements, Douglas H. and Julie Sarama, *Learning and Teaching Early Math: The Learning Trajectories Approach*, 2nd ed., Routledge, New York, 2014.

64. Geary, David C., 'From Infancy to Adulthood: The development of numerical abilities', *European Child and Adolescent Psychiatry*, vol. 9, no. 2, 2000, pp. S11–S16.

65. Mix, Kelly S., Catherine M. Sandhofer and Arthur J. Baroody, 'Number Words and Number Concepts: The interplay of verbal and nonverbal quantification in early childhood', in *Advances in Child Development and Behavior*, vol. 33, edited by Robert V. Kail, Elsevier Academic Press, New York, 2005, pp. 305–346.

66. Chard, David J., et al., 'Using Measures of Number Sense to Screen for Difficulties in Mathematics: Preliminary findings', *Assessment for Effective Intervention*, vol. 30, no. 2, 2005, pp. 3–14.

67. McBride-Chang, Catherine, 'The ABCs of the ABCs: The development of letter-name and letter-sound knowledge', *Merrill-Palmer Quarterly*, vol. 45, no. 2, 1999, pp. 285–308; Schatschneider, Christopher, et al., 'Kindergarten Prediction of Reading Skills: A longitudinal comparative analysis', *Journal of Educational Psychology*, vol. 96, no. 2, 2004, pp. 265–282; Wagner, Richard K., Joseph K. Torgesen and Carol A. Rashotte, 'Development of Reading-Related Phonological Processing Abilities: New evidence of bidirectional causality from a latent variable longitudinal study', *Developmental Psychology*, vol. 30, no. 1, 1994, pp. 73–87.

68. Johnson, Kelli, 'What Is Phonological Awareness?', undated, <www.understood.org/en/articles/phonological-awareness-what-it-is-and-how-it-works>, accessed 3 January 2022.

69. Bradley, Lynette, and Peter Bryant. 'Categorizing Sounds and Learning to Read: A causal connection', *Nature*, vol. 301, no. 5899, 1983, pp. 419–421; Roth, Froma P., Deborah L. Speece and David H. Cooper, 'A Longitudinal Analysis of the Connection Between Oral Language and Early Reading', *Journal of Educational Research*, vol. 95, no. 5, 2002, pp. 259–272; Wagner, Richard K., and Joseph K. Torgesen, 'The Nature of Phonological Processing and Its Causal Role in the Acquisition of Reading Skills', *Psychological Bulletin*, vol. 101, no. 2, 1987, pp. 192–212.

70. *Early Child Development in Social Context*.

71. Ibid.

72. Ibid.

73. Duncan et al., 'School Readiness and Later Achievement'; Stanovich, Keith E., Anne E. Cunningham and Dorothy J. Feeman, 'Intelligence, Cognitive Skills, and Early Reading Progress', *Reading Research Quarterly*, vol. 19, no. 3, 1984, pp. 278–303.

74. Duncan et al., 'School Readiness and Later Achievement'.

75. *Early Child Development in Social Context*.

76. *Head Start Early Learning Outcomes Framework*.

77. Stipek, Deborah, and Rachel A. Valentino, 'Early Childhood Memory and Attention as Predictors of Academic Growth Trajectories', *Journal of Educational Psychology*, vol. 107, no. 3, 2015, pp. 771–788.

78. Alloway, Tracy P., 'Working Memory Is a Better Predictor of Academic Success than IQ', *Psychology Today*, 21 December 2010.

79. Diamond, Adele, and Bertha Doar, 'The Performance of Human Infants on a Measure of Frontal Cortex Function, the Delayed Response Task', *Developmental Psychobiology*, vol. 22, no. 3, 1989, pp. 271–294; Diamond, Adele, Natasha Kirkham and Dima Amso, 'Conditions under which Young Children Can Hold Two Rules in Mind and Inhibit a Prepotent Response', *Developmental Psychology*, vol. 38, no. 3, 2002, pp. 352–362.

80. Alexander, Karl L., Doris R. Entwisle and Susan L. Dauber, 'First Grade Classroom Behavior: Its short- and long-term consequences for school performance', *Child Development*, vol. 64, no. 3, 1993, pp. 801–814; Blair, Clancy, 'School Readiness: Integrating cognition and emotion in a neurobiological conceptualization of children's functioning at school entry', *American Psychologist*, vol. 57, no. 2, 2002, pp. 111–127; Sáez, Leilani, et al., 'Relations among Student Attention Behaviors, Teacher Practices, and Beginning Word Reading Skill', *Journal of Learning Disabilities*, vol. 45, no. 5, 2012, pp. 418–432.

81. Blair, Clancy, and Rachel Peters Razza, 'Relating Effortful Control, Executive Function, and False Belief Understanding to Emerging Math and Literacy Ability in Kindergarten', *Child Development*, vol. 78, no. 2, 2007, pp. 647–663.

82. Blandon, Alysia Y., Susan D. Calkins and Susan P. Keane, 'Predicting Emotional and Social Competence during Early Childhood from Toddler Risk and Maternal Behavior', *Development and Psychopathology*, vol. 22, no. 1, 2010, pp. 119–132.

83. Campbell et al., 'Commentary on the Review of Measures of Early Childhood Social and Emotional Development'.

84. Ibid.

85. *Early Child Development in Social Context*.

86. Campbell et al., 'Commentary on the Review of Measures of Early Childhood Social and Emotional Development'; Fernald et al., *A Toolkit for Measuring Early Childhood Development in Low- and Middle-Income Countries*.

87. Collaborative for Academic, Social, and Emotional Learning, 'CASEL's SEL Framework: What are the core competence areas and where are they promoted', 1 October 2020, <https://sel4ma.org/resource/casels-sel-framework-what-are-the-core-competence-areas-and-where-are-they-promoted-2/>, accessed 21 June 2023.

88. Hay, Dale F., Alexandra Payne and Andrea Chadwick, 'Peer Relations in Childhood', *Journal of Child Psychology and Psychiatry*, vol. 45, no. 1, 2004, pp. 84–108.

89. Ibid.

90. Keenan, Kate, 'Development of Physical Aggression from Early Childhood to Adulthood', in *Encyclopedia on Early Childhood Development* [online], edited by Richard Tremblay, 2012.

91. Ibid.

92. Knafo-Noam, Ariel, ed., 'Prosocial Behaviour: Synthesis', in *Encyclopedia on Early Childhood Development* [online], edited by Richard Tremblay, Michel Boivin and Ray DeV Peters, 2016.

93. Dunfield, Kristen A., 'A Construct Divided: Prosocial behavior as helping, sharing, and comforting subtypes', *Frontiers in Psychology*, vol. 5, 2014, pp. 1-13.

94. Knafo-Noam, 'Prosocial Behaviour: Synthesis'.

95. Ladd, Gary W., 'School Transitions/School Readiness: An outcome of early childhood development ~ Perspective: Children's social and scholastic development – Findings from the Pathways Project,' in *Encyclopedia on Early Childhood Development* [online], edited by Richard Tremblay, Michel Boivin and Ray DeV Peters, 2009.

96. Bulgarelli, Daniela, and Paola Molina, 'Social Cognition in Preschoolers: Effects of early experience and individual differences', *Frontiers in Psychology*, vol. 7, 2016, pp. 1-11.

97. Astington, Janet W., and Margaret J. Edward, 'The Development of Theory of Mind in Early Childhood', in *Encyclopedia on Early Childhood Development* [online], edited by Richard Tremblay, Michel Boivin and Ray DeV Peters, 2010.

98. Ibid.

99. Denham, Susanne A., et al., 'Preschool Emotional Competence: Pathway to social competence?', *Child Development*, vol. 74, no. 1, 2003, pp. 238–256.

100. Wang, Zhenlin, and Lamei Wang, 'The Mind and Heart of the Social Child: Developing the empathy and theory of mind scale', *Child Development Research*, vol. 1, 2015, pp. 1–8.

101. Astington and Edwards, 'The Development of Theory of Mind in Early Childhood'.

102. Wang and Wang, 'The Mind and Heart of the Social Child'.

103. Ibid.

104. 'Measuring Child Development and Early Learning'.

105. Denham, Susanne A., 'Dealing with Feelings: How children negotiate the worlds of emotions and social relationships', *Cognition, Brain, Behavior*, vol. 11, no. 1, 2007, pp. 1–48.

106. Campbell et al., 'Commentary on the Review of Measures of Early Childhood Social and Emotional Development'.

107. Denham, 'Dealing with Feelings'.

108. Denham et al., 'Preschool Emotional Competence: Pathway to social competence?'; Fernald et al., *A Toolkit for Measuring Early Childhood Development in Low- and Middle-Income Countries*.

for every child

unicef

for every child