# Monitoring learning in the early years

## A review of early childhood assessments to support global monitoring

### July 2024

Dan Cloney, David Jeffries, Stephanie Templeton, Ursula Schwantner, Juliette Mendelovits, Amy Berry, Syeda Kashfee Ahmed, Nicola Andrews & Bethany Davies

The Global Education Monitoring (GEM) Centre drives improvements in monitoring the quality of education worldwide.

# Contents

# List of tables

# List of figures

# List of abbreviations

| | |
|---|---|
| ACER | Australian Council for Educational Research |
| AEDC | Australian Early Development Census |
| ARNEC | Asia-Pacific Regional Network for Early Childhood |
| ASER | Annual Status of Education Report |
| ASQ-3 | Ages and Stages Questionnaires, Third Edition |
| CFI | Comparative Fit Index |
| CLA | Citizen-led Assessments |
| CREDI | Caregiver Reported Early Development Instruments |
| DAZ | Development-for-age-z-score |
| DFAT | Department of Foreign Affairs and Trade of the Government of Australia |
| DHS | Demographic and Health Surveys |
| DIF | Differential Item Functioning |
| EAP-ECDS | The East Asia-Pacific Early Childhood Development Scales |
| EAPs | Expected A Posteriori |
| ECDAS | Early Childhood Development Assessment Scale |
| ECDAS-CS | Early Childhood Development Assessment Scale – Caregiver Survey |
| ECDAS-DA | Early Childhood Development Assessment Scale – Direct Assessment |
| ECDI 2030 | UNICEF Early Childhood Development Index 2030 |
| ECE | Early Childhood Education |
| EDI | Early Development Instrument |
| EGMA | Early Grade Mathematics Assessment |
| EGRA | Early Grade Reading Assessment |
| eHCI | Early Human Capability Index |
| ELANA | Early Language, Literacy and Numeracy Assessment |
| GEM Centre | Global Education Monitoring Centre |
| GSED | Global Scales for Early Development |
| GSED LF | Global Scales for Early Development Long Form |
| GSED SF | Global Scales for Early Development Short Form |
| HCP | England's Healthy Child Programme |
| IAEG-ECD | UNICEF's Inter-Agency and Expert Group on ECD Measurement |
| ICAN | International Common Assessment of Numeracy |
| ICC | Item Characteristic Curves |
| IDELA | International Development and Early Learning Assessment |
| IELS | OECD's International Early Learning and Child Well-being Study |
| IRT | Item Response Theory |

| | |
|---|---|
| IYCD | Infant and Young Child Development |
| MAPs | Maximum A Posteriori |
| MELE | Measuring the Quality of Early Learning Environments |
| MELQO | Measuring Early Learning Quality and Outcomes |
| MICS | Multiple Indicator Cluster Surveys |
| MODEL | Measuring Child Development and Learning |
| OECD | Organisation for Economic Co-operation and Development |
| OSF | Open Society Foundations |
| PAL Network | People's Action for Learning Network |
| RMSEA | Root Mean Square Error of Approximation |
| SDG | Sustainable Development Goal |
| SDQ | Strengths and Difficulties Questionnaire |
| SEL | Wellbeing and Social and Emotional Learning |
| TLI | Tucker-Lewis Index |
| UIS | UNESCO's Institute for Statistics |
| UNESCO | The United Nations Educational, Scientific and Cultural Organization |
| UNICEF | United Nations Children's Fund |
| EAPRO | UNICEF's East Asian and Pacific Regional Office |
| WHO | World Health Organisation |

# Executive summary

## Introduction

The United Nation's Sustainable Development Goal (SDG) Indicator 4.2.1 aims to track the "proportion of children aged 24–59 months who are developmentally on track in health, learning and psychosocial well-being, by sex" (United Nations Department of Economic and Social Affairs, n.d.). The United Nations Children's Fund (UNICEF) is the custodian of the SDG 4.2 target, developing and maintaining the official assessment for reporting against SDG 4.2.1: the Early Childhood Development Index 2030 (ECDI2030). This review attempts to identify and critique a range of existing ECE assessments (including the ECDI2030) that could be utilised for SDG 4.2.1 reporting, highlighting their respective strengths and limitations.

The role that ECE plays in the growth and development of foundational cognitive and socio-emotional skills cannot be overstated, particularly with the resultant flow-on-effect ECE has on the later acquisition of important life skills (World Bank, 2018). Research has shown the positive effect that engagement in ECE can have on the children, particularly those from disadvantaged backgrounds (Heckman, n.d.; World Bank., 2018). Further, the quality is important in this regard as it reinforces the effects, giving rising concern to the lack of access to high quality ECE that many children still face today (UNESCO, 2021). A key indicator of the quality of ECE is children's developmental progress.

Children's developmental trajectories can vary widely in the early years and are influenced by a number of external factors (Yoshikawa & Kabay, 2015). Assessment of this development is typically approached from one of two perspectives. The first is the examination of individuals using screening methods for diagnoses of developmental delays for follow-up assessment or intervention (United Nations Children's Fund, 2020a). The second is at the system level, primarily concerned with capturing a snapshot of the development of a population. This second purpose is the focus of the review: population-wide assessments.

Gaining an adequate understanding of the level of development of a population is important for systems to develop and enact policies where they are most needed (UNESCO, n.d.). At a global level, it is important for monitoring progress towards a shared goal (for example, the SDGs). The critical role population assessments play in national and global education agendas highlights the need for them to be developed and implemented in a way that is valid, reliable and unbiased (Cloney et al., 2019). Methodological approaches that provide this type of data for monitoring ECE outcomes become important.

Going beyond the examination of a snapshot (that is, a single point in time) of the developmental level of a population, systems are interested in examining change in children's development over time. This also provides an indication of how the quality of ECE is progressing. To monitor system-level change, stakeholders need comparable

data. This can either be achieved by utilising a common assessment, or by enabling direct comparison of the results from different assessments. The latter can be achieved when the results from individual assessments are placed together on the same reporting scale, for example by using the pairwise comparison method (PCM: UNESCO Institute for Statistics & Australian Council for Educational Research, 2024). This approach also allows for internationally recognised benchmarks – for example, the Minimum Proficiency Levels (MPLs: Australian Council for Educational Research, 2022b) – to be applied from one assessment to another, once they are equated onto the same scale. The concept of linking assessments can be extended to monitoring growth over the lifespan: for example, young children (SDG 4.2.1), primary and secondary children (SDG 4.1.1) and adults (SDG 4.6.1).

To give substance to the quality of information that is available for systems to inform policy development, it is important to describe the skills that are expressed by the numerical values of an assessment scale: that is, to establish a described proficiency scale (Turner, 2014). Learning progressions are evidence-based descriptions of typical growth that draw on theory, subject matter expertise and other sources of evidence. Not only are these learning progressions useful for educators to inform practice, but they can be beneficial to the broader education community by providing a shared language articulating what meaningful and valuable learning looks like.

In their role as custodian of SDG 4.2 target on ECE outcomes, UNICEF has overseen the development of the Early Childhood Development Index 2030 (ECDI2030). The ECDI2030 is a population tool designed to capture the level of children aged 24 to 59 months against key defined developmental milestones. It consists of 20 items that span the core domains defined in SDG 4.2.1, namely health, learning, and psycho-social wellbeing. While ECDI2030 provides a simple approach to reporting against SDG 4.2.1, it has a number of constraints that limit its capacity providing systems with a nuanced picture of children's progress. The constraints include:

- A small item pool: Does not provide comprehensive construct coverage for a broader range of sub-skills within the domain as well as a wider range of abilities

- Global indicator only: Inability to report against health, learning and psycho-social wellbeing separately

- Reliance on raw score approach: Assumes complete data, requires adjusting for non-response to yield an equivalent raw score, and uses single and fixed form that is not dynamic and cannot be changed and compared over time (including a progression of learning over a continuum)

- Omission of direct assessment of children: Relies on parent/caregiver or educator reports about children's learning

These constraints make it difficult to translate results directly from the ECDI2030 to support initiatives for improving specific sub-domains of learning (for example, literacy or numeracy) or link to a continuum of learning that leads to later outcomes, in school and beyond the years of schooling.

This report reviews a range of assessments, including ECDI2030, with a view to identifying how they can be strengthened. The review of existing ECE assessments has been motivated by the Global Education Monitoring (GEM) Centre's aim to contribute to the objective measurement of skills and attributes in early childhood education. It is intended to benefit global stakeholders in the SDG 4 agenda, national governments, practitioners, and ultimately, children.

## Method

The review of early childhood assessments began with a search of academic and grey literature, to identify early childhood assessments that can be used to measure progress toward SDG 4.2.1. Domains that align with the SDG 4.2.1 target include:

- Health: gross motor development, fine motor development and self-care.

- Learning: expressive language, literacy, numeracy, pre-writing, and executive functioning.

- Psychosocial well-being: emotional skills, social skills, internalising behaviour, and externalising behaviour.

To be included in the review, the assessments were required to be explicitly related to learning and be targeted at children aged approximately 2-6 years.

Forty-eight assessments were initially identified in the literature search. Of these, 25 assessments were included in the critical review based on the inclusion criteria. Some of these assessments were grouped together for reporting purposes, resulting in 14 individual/groups of assessments, which were:

- Ages & Stages Questionnaires, Third Edition (ASQ-3), University of Oregon

- Caregiver Reported Early Development Instruments (CREDI), Harvard University

- Early Childhood Development Assessment Scale (ECDAS), The University of Hong Kong

- Early Childhood Development Index 2030 (ECDI2030), UNICEF

- Early Development Instrument (EDI), Offord Centre for Child Studies, McMaster University

- Early Grade Reading Assessment (EGRA) and Early Grade Mathematics Assessment (EGMA), RTI International

- Early Human Capability Index (eHCI), Telethon Kids Institute

- East Asia-Pacific Early Child Development Scale (EAP-ECDS), UNICEF, the Asia-Pacific Regional Network for Early Childhood (ARNEC), Open Society Foundations (OSF), The University of Hong Kong (Technical partners)

- Global Scales for Early Development (GSED), World Health Organisation (WHO)

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

- Infant and Young Child Development (IYCD), World Health Organisation (WHO)

- International Early Learning and Child Well-being Study (IELS), Organisation for Economic Cooperation and Development (OECD)

- Measuring Early Learning Quality and Outcomes (MELQO), UNESCO, UNICEF, Brookings Institution, World Bank

- Citizen-led assessments and common assessments of the People's Action for Learning (PAL) Network.

An evaluation framework was developed to ensure that a standardised approach was taken in reviewing and reporting on the assessments included. Existing frameworks for reviewing assessments were used as a basis. Each included assessment (or group of assessments) was reviewed against the key assessment components, with a detailed breakdown provided for each one. A classification for each assessment (or group of assessments) was also developed.

## Results and discussion

A number of strengths were identified for many of the assessments.

All of the assessments included in the review show general alignment with SDG 4.2.1 and could be used, or are already being used, for SDG reporting. Most of the assessments also have well-defined assessment frameworks, detailing the level of the assessments' coverage of the domain of interest for SDG 4.2.1 reporting. However, the level of construct coverage varies widely. Many assessments directly assess children's learning and development, whereas others rely on indirect approaches via parent/educator reports of children's learning. A few of the assessments are designed in a way that would potentially allow for them to be more fully developed over time to align with SDG 4.2.1 by examining trends over time, comparing across countries/systems, and being empirically aligned with ECDI2030 to provide a mapping of multiple assessments to an international benchmark. Three of the included assessments have published details of analyses which examined the cross-cultural/linguistic comparability of the assessment items using appropriate measurement invariance.

There are, however, some limitations common across the reviewed assessments.

Several assessments consist of small item pools, which limit the ability to provide adequate construct coverage, especially when there is a desire to cover a range of subdomains of a broader domain of interest. A large item pool allows for more targeted and fit-for-purpose testing to be done that suit the immediate context, whilst being able to be linked back to the common established metric. Several assessments employed raw

score approaches to measurement (for example, sum or average), which either assumes complete data, or requires adjusting for non-response to yield an equivalent raw score, whilst others provided guidance on transforming raw scores onto an ordinal/linear scale. Relying on a single raw score precludes the possibility of linking to a learning progression. Raw score reporting, in isolation, does not allow for interpreting results in a meaningful way; nor does it allow for monitoring trends over time via a continuum of learning (for example, linking SDG 4.2.1 to SDG 4.1.1). While some assessments used IRT-based approaches to illustrate the span of skills that higher ability children may possess as they progress further along the developmental continuum, none of the assessments included in the review have learning progressions associated with them. This results in an inability to interpret the results of the assessments by describing the underlying knowledge, skills and understandings children possess at particular locations on the scale and what they might be developing next. This has practical implications for the assessments' ability to inform policy and practice.

## Conclusion

This review has identified and critiqued a range of existing early years assessments that could be utilised for SDG 4.2.1. reporting. Whilst the ECDI2030 is the official measure for doing this, it should not preclude the use of other suitable assessments for reporting against these important international benchmarks. Some of these assessments go beyond the level of sophistication of the ECDI2030 in terms of scoring, scaling and measurement of children's skills. The approach used for SDG 4.1.1. reporting could also be adopted for SDG 4.2.1 where a a set of tools, methods and criteria have been developed to align and harmonise findings across different assessments (UNESCO Institute for Statistics, 2023; UNESCO Institute for Statistics & Australian Council for Educational Research, 2024).

The review highlights the strengths of some of the assessments in terms of large item pools and applying IRT-based approaches to scoring and scaling that allows for better comparability of scores across different versions of a test or across different populations (Adams, 2017); an advantage over classical test theory, which many of the existing assessments still apply. More sophisticated approaches could ultimately lead to the development of learning progressions, which provide useful information about descriptions of learning, what progress looks like and why progress is valued (Waters, 2019, para. 16). IRT-based approaches also allow for linking assessments in the early years with later assessments (school and post-school) to better describe learning as a continuum. Utilising this approach where assessments are linked together would allow systems to select the assessments that are most appropriate for their context, report against international benchmarks, and highlight the areas of need for their respective population.

# Introduction

Sustainable Development Goal 4 (SDG 4) aims to "ensure inclusive and equitable quality education and promote lifelong learning opportunities for all" (United Nations Department of Economic and Social Affairs, n.d.). In relation to Early Childhood Education (ECE), SDG Target 4.2 aims to "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" (United Nations Department of Economic and Social Affairs, n.d.). To operationalise this goal, the primary Indicator 4.2.1 sets out to measure the "proportion of children aged 24–59 months who are developmentally on track in health, learning and psychosocial well-being, by sex" (United Nations Department of Economic and Social Affairs, n.d.).

The United Nations Educational, Scientific and Cultural Organization (UNESCO) Institute for Statistics (UIS) has been mandated with monitoring the progress of countries towards achieving SDG 4. ACER is a technical partner of UIS and – through the GEM Centre, contributes to the development of public goods and activities that facilitate education systems' reporting against SDG 4 in a globally consistent way. Consistency in monitoring learning is essential so that education systems gain an understanding of their strengths and the challenges they face. The evidence provided can be used to inform the development of policies and practices that ultimately improve student learning.

The United Nations Children's Fund (UNICEF) is the custodian of SDG Target 4.2. As such, UNICEF has coordinated efforts to develop an assessment that "captures the achievement of key developmental milestones of children between the ages of 24 and 59 months" (UNESCO Institute of Statistics, 2022). This assessment is called the Early Childhood Development Index 2030, or ECDI2030. The ECDI2030 has been in use since March 2020. While the tool provides a quick snapshot for reporting, there is room for improvement, considering the breadth of skills that are developing as children grow.

This review of existing ECE assessments has been motivated by the GEM Centre's aim to develop appropriate definitions and to contribute to the objective measurement of skills and attributes in early childhood education. The report reviews a range of assessments, including ECDI2030, with a view to identifying how they can be strengthened. The report is intended to benefit global stakeholders in the SDG 4 agenda, national governments, practitioners, and ultimately, children.

# Background

## The importance of Early Childhood Education (ECE)

ECE plays a crucial role in advancing both individuals and society. The benefits of education for individuals and societies have long been understood and empirically demonstrated in research over the past 60 years. For individuals, benefits come in the

form of higher incomes, better health and greater life satisfaction (World Bank 2018). For society, benefits of education include higher economic growth, better functioning institutions and greater social cohesion (McMahon, 2010; World Bank., 2018). ECE is critical in this dynamic as during early childhood the foundational cognitive and socio-emotional skills are acquired that are needed for the further development of those skills later in life (World Bank 2018). Research shows that pre-primary education provides the highest return on investment of all education sub-sectors (UNICEF, n.d.). In particular, investing in ECE for children from disadvantaged backgrounds is most effective in reducing social costs (Heckman, n.d.). Early childhood programs can provide an opportunity for disadvantaged children to access nutrition, stimulation and lower-stress environments (World Bank., 2018). This helps with their biological development and supports their ability to learn, ultimately increasing self-sufficiency and productivity in the long term (Heckman, n.d.; World Bank., 2018). Children who have participated in ECE are also more likely to complete schooling and are less likely to repeat grades (UNICEF, n.d.; World Bank., 2018).

The quality of ECE matters: the higher the quality of the ECE programs, the greater the positive impact (UNICEF, 2020; Waldfogel, 2006). Even though there has been much progress in expanding access to ECE over the past decade, many children still lack access to high quality programs (UNESCO, 2021).

## Approaches to assessing early childhood development

There are 2 major approaches to assessing children's development. The use of each approach is based on its purpose. One approach is to capture a snapshot of development across a population. Assessments used for this purpose generally attempt to cover a broad definition of a child's development, including health, learning and psychosocial well-being (or socio-emotional well-being). The other approach is to screen or diagnose children for developmental delays. Assessments used for this purpose aim to identify individuals for further assessment or intervention and usually require highly trained professionals to administer the assessment, and extensive assessment time (United Nations Children's Fund, 2020a). As the purpose of this review is to identify assessments that have the potential to provide information about early childhood development across the global population, the review focuses on the first of these 2 approaches: population-wide assessments.

Population-wide assessment is valuable for a variety of actors in any education sector. Assessment can help professionals to determine the efficacy of their pedagogy, to identify gaps in children's development and learning, and to intervene and support children's learning in evidence-based ways (Cloney et al., 2019; UNICEF, 2021). Assessment can help governments to build policies that improve the curriculum, pedagogy and teaching resources, and direct budgets to where they are most needed (UNESCO, n.d.). The global community can use assessment to monitor progress towards a shared goal – such as SDG 4, or to prioritise and direct funding towards specific initiatives.

Yet, for an assessment to be truly useful and relevant for any stakeholder and within any sub-sector, it must be reliable, valid, and unbiased (Cloney et al., 2019). Assessment in the early childhood sector faces particular methodological challenges. One challenge arises from the fact that young children have relatively short attention spans, and their ability to self-regulate is still developing (National Research Council, 2000). To overcome this and improve the validity of the assessment, it is important to use direct assessment of a child's learning and development in a way that is engaging for the child (Cloney et al., 2019). That said, gaining additional insights from parents/caregivers and educators via indirect assessment is important in providing further context about children, whilst also triangulating information gained through direct assessment (Cloney et al., 2019). Using rigorous population-wide assessment to measure learning in early childhood helps advance outcomes for children and society more broadly.

## Approaches to measuring progress

Population-wide assessment of children's learning and development plays an important role in monitoring the quality of ECE, providing crucial evidence for educators and policy makers to enact change. In order to do this in a meaningful way, it is important to go beyond the representation of test-takers abilities in terms of a numerical value on a scale; evidence becomes meaningful when numerical scores are provided alongside a description of what these scores mean, to provide a substantive understanding of the learning and development assessed in a particular domain (Turner, 2014). This type of approach to describing proficiency scales is also aligned with recent research on learning progressions and can draw on this research to aim for greater impact. Learning progressions are evidence-based representations of growth that are typically developed by drawing on multiple sources of data, and ideally comprise both an underpinning numerical scale and qualitative descriptions of what different regions of the scale mean in terms of knowledge, understandings and skills. Research on learning progressions is a growing field, with scholars exploring how they can be used not only in assessment, but also in instructional planning, in the development and review of curricula, and as resources for the training and professional development of teachers and educators (Australian Council for Educational Research, 2022a; Australian Education Research Organisation, 2023; Confrey, 2019; Mosher & Heritage, 2017). The research on learning progressions points to a way forward for the international community in its efforts to promote the SDG 4 goal of quality education and lifelong learning opportunities for all (Adams et al., 2018; Jackson et al., 2018). In particular, the descriptions in learning progressions can be used to first negotiate and then represent a shared understanding of what meaningful and valuable learning looks like – an understanding that is not bound by national borders.

To monitor change and progress over time, it is important to adopt processes and methods that allow for results to be compared, including from different data collection efforts. There are multiple ways of doing this with the ultimate goal of bringing multiple assessments together onto the same scale (Adams et al., 2018; Turner et al., 2014). Recent efforts have been made to show how the pairwise comparison method (PCM: UNESCO

Institute for Statistics & Australian Council for Educational Research, 2024) can be used to link multiple assessments together. In addition to comparing outcomes across different assessments this method also allows for the transfer of global standards for reporting from one assessment to another. An example of this is   the Minimum Proficiency Level (MPL) thresholds that have been established on the ACER Learning Progressions for reading and mathematics during the International Standard Setting Exercise (ISSE) carried out by ACER in 2022 (Australian Council for Educational Research, 2022c). This approach opens up the possibility of obtaining internationally comparable data for one SDG 4 indicator (for example, 4.2.1) while countries measure and monitor learning in ways that work for them. It might also facilitate connections between the efforts to collect and interpret data for the SDG 4 indicators that are concerned with learners at different educational and life stages, ultimately linking assessments that are designed to assess young children (SDG 4.2.1), primary and secondary children (SDG 4.1.1) and adults (SDG 4.6.1), across the full range of abilities on a single scale. Such an approach makes possible the meaningful monitoring of learning progress across the lifespan.

## Global monitoring of ECE outcomes with ECDI2030

To ensure quality education (SDG 4) in the early childhood sector (Target 4.2), Indicator 4.2.1 aims to measure the "proportion of children aged 24 to 59 months who are developmentally on track in health, learning and psychosocial well-being, by sex" (United Nations Department of Economic and Social Affairs, n.d.). Since 2015, proxy indicators have been used for reporting on Indicator 4.2.1. UNICEF, as the custodian agency of SDG Indicator 4.2.1, has worked to develop and validate assessments that can monitor and track progress towards achieving SDG target 4.2 (Cappa et al., 2021). UNICEF manages up-to-date databases that include key indicators of early childhood development gathered through multiple sources, such as the Multiple Indicator Cluster Surveys (MICS), Demographic and Health Surveys (DHS) and other nationally representative household surveys (UNICEF, 2022).

To support the coordination and management of monitoring ECE outcomes against SDG Indicator 4.2.1, UNICEF formed the Inter-Agency and Expert Group on ECD Measurement (IAEG-ECD), involving representatives of national statistical offices and experts from various development agencies. A Technical Advisory Group has also been convened, including researchers from selected institutes, agencies and universities, who act in an advisory role to support the IAEG-ECD (Cappa, C., 2019). Through these efforts, the Early Childhood Development Index 2030 (ECDI2030), a nationally representative and internationally comparable and standardised ECD data collection tool, has been developed (UNICEF, 2021).

The ECDI2030 is a freely available standardised questionnaire designed to record the achievement of key developmental milestones by children between the ages of 24 and 59 months. It is a population-based assessment that is designed to be integrated into existing national data collection efforts such as MICS (UNICEF, 2019). Whilst the 20

items cover the core domains of health, learning and psychosocial well-being[1], a single summary score is calculated based on a series of age-appropriate cut scores. The questionnaire is typically administered to the mother or caregiver as part of broader, existing household surveys. The ECDI2030 was officially released in March 2020, and training and capacity building efforts have been undertaken by UNICEF to encourage uptake. As of 2022, at least 18 countries were preparing to implement tool in their national surveys (Petrowski et al., 2022).

Given the emphasis education systems and policy makers place on gaining a nuanced view of the progress children are making in their learning in terms of their underlying knowledge, skills and understandings, it is imperative to highlight some of the constraints that result from using the ECDI2030 as a standalone assessment. For example, the ECDI2030:

- has a relatively small item pool – just a few items per sub-domain– means that there is little information to describe the breadth of skills that are developing as children grow

- yields an indicator of a child's overall development: one number that encompasses health, learning and psychosocial wellbeing

- relies on raw score-based approaches to measurement

- does not directly assess a child's learning and development. Rather, the ECDI2030 relies on the child's mother or primary caregiver's perception and knowledge of the child. As previously mentioned, these perceptions are important given the key role that families and their community contexts play in a child's development. However, without direct assessment of the child, certain biases may be introduced, reducing the validity of the assessment results (Campbell et al., 2016).

These constraints, and the impact they can have on an assessment's validity and utility are not unique to the ECDI2030, as will be highlighted and discussed later in the report.

While the release of an official assessment for reporting against Indicator 4.2.1. marks progress towards measuring the quality of education globally, the data that ECDI2030 provides is unlikely to be optimally useful for governments and practitioners in implementing improvements in policy and practice. It is not apparent how policy makers and practitioners can translate results from this assessment to support early childhood education programs through improvements in specific sub-domains (for example, literacy or numeracy), or to link results to a continuum of learning and development that leads to school education outcomes, including SDG 4.1.1.

---

[1] Domain: health. Sub-domain: gross motor development, fine motor development and self-care.
Domain: learning. Sub-domain: expressive language, literacy, numeracy, pre-writing, and executive functioning.
Domain: psychosocial well-being. Sub-domain: emotional skills, social skills, internalizing behaviour, and externalizing behaviour.

# Method

This review of globally available assessments in early childhood learning and development aims to identify assessments, including ECDI2030, that could be used to report on SDG Indicator 4.2.1., and to provide recommendations for the strengthening of these assessments for the benefit of all stakeholders. As such, this review aims to address the following research questions:

1) Which assessments could potentially support the monitoring and reporting of SDG Indicator 4.2.1?

2) What are the strengths of the assessments?

3) What are the limitations of the assessments?

This section describes the approach used to review the assessments.

The methodology for identifying assessments to include in this review and for evaluating the assessments in terms of the research questions comprised 3 stages:

- Stage one, outlining a search strategy to identify assessments

- Stage 2, defining the criteria for the inclusion or exclusion of an assessment from the review

- Stage 3, developing a review framework that defines the key components against which the assessments are evaluated.

## Stage one: Search strategy

The list below details the approaches taken to locate assessments for review:

- Search on key terms in databases such as EBSCO, Scopus, Web of science, Google Scholar etc. Key terms include: early childhood, early years, early develop*, early learn*, child*, assess*, measur*, tool, index, instrument. An asterisk represents a wildcard symbol that broadens the search by finding variations of the root word.

- Leverage existing assessment inventories and previous reviews in the area of early childhood. For example, (Ainley et al., 2021; Cloney et al., 2019; Fernald et al., 2017; Halle & Darling-Churchill, 2016; Munoz-Chereau, B et al., 2021).

- Search grey literature. For example, search repositories of large education-related organisations: UNICEF, OECD, the World Bank, Save the Children.

- Search reference lists of the found documents to identify other relevant reports and documents.

See Appendix A:  for a full list of search key terms and the search strategy.

In total, 47 assessments in addition to ECDI2030 were identified at this stage.

# Stage 2: Inclusion and exclusion criteria

To identify assessments that have the potential to be used in reporting progress against Indicator 4.2.1., researchers developed a set of inclusion and exclusion criteria that align with key requirements of Indicator 4.2.1. An assessment was included in the review if it fit all the inclusion criteria and did not fit any of the exclusion criteria. These criteria are summarised in Table 1.

**Table 1: Criteria for inclusion and exclusion of assessments in the review**

| Criteria | Include | Exclude |
|---|---|---|
| Age range | Assessments designed for children 2 to 6 years of age, targeting the range partially or fully. | Assessments for children <2 years of age only and those targeting children >6 years of age only. |
| Purpose | Assessments whose results are or can be used to measure a population and have some relation to education. | Assessments that are predominantly used for individual diagnostic/screening purposes[2]. |
| Originality | Newly developed assessments; or assessments that combine existing items from multiple sources. | Adaptations of the same assessment or those fully based on an existing assessment (for example, the Australian Early Development Census (AEDC) is excluded because it is based on the Canadian Early Development Index (EDI)). |
| Availability | Assessments that are open source, or free to use on request, or include one-off royalty fees. | Assessments that are used for commercial purposes, typically by specialist groups or for professional assessment of learning / developmental difficulties that require clinical, allied health, or other professional qualification to access (for example, assessments where one must pay individually for the scoring sheets). |
| Outcomes | Assessments that measure at least one domain explicitly related to learning (for example, pre-academic skills, language or literacy, arithmetic or numeracy, cognitive abilities, or approaches to learning), or include relevant learning items that belong to a broader domain (for example, holistic/global development). | Assessments that measure domains other than those explicitly related to learning, even if they are often associated with learning (for example, self-concept, identity, literacy interest, self-efficacy, psychological stress/difficulties). |
| Implementation | Assessments that have been trialled across multiple countries/languages; or have been designed for use across countries/languages and include an adaptation manual. | Assessments that have only been trialled in one country/language. |

---

[2] Assessments predominantly used as screening tools have been included if there are multiple sources of evidence that they are used as population measures (for example, ASQ - (Department of Health (UK), 2016; Kendall et al., 2014, 2019)).

| Criteria | Include | Exclude |
|---|---|---|
| Languages | The assessments must have English source versions available[3]. | Those only available in a language other than English. |
| Validity and reliability | Assessments that have been psychometrically validated, with details published in a research report/article. | Assessments that have not been psychometrically validated. |

Of the 48 assessments that were identified in total, 25 were included based on these criteria. Some of these 25 assessments are reviewed and presented in this report as a set, rather than as individual assessments. This is due to one or more of the features described in Table 2 below.

**Table 2: Features of assessments where multiple tools are presented together**

| Feature | Assessments |
|---|---|
| Developed by the same organisations and often implemented together | EGRA and EGMA<br><br>ECDAS-DA and ECDAS-CS |
| Earlier versions of the same assessment, or a similar assessment | Individual citizen led assessments (for example, ASER, UWEZO, Beekunko)<br><br>PAL Network's common assessments (for example, ICAN and ELANA) |
| Alternative versions of the same assessment | EDI (Canadian version predominantly used for review)<br><br>eHCI (Australia version predominantly used for review) |
| Alternative forms (short and long forms) | CREDI<br><br>GSED |
| Multiple forms using different administration methods for triangulation purposes | IELS |

In total, this report reviews 14 individual assessments or sets of assessments. For the full screening details see Appendix B: : List of assessments screened.

The following provides an example for an included and an excluded assessment to illustrate the criteria.

---

[3] As English is the language of this research study, this criterion was included for practical reasons.

The Early Grade Reading Assessment (EGRA) and Early Grade Mathematics Assessment (EGMA) are short assessments, orally administered to children in grades 1 to3, which aim to measure their foundational skills in reading and mathematics (RTI International, 2012). EGRA and EGMA are designed to be population-level assessments that are used for monitoring students' performance against curriculum, and for evaluating the effectiveness of curriculum changes, school-based interventions, and teacher professional development programs. The core EGRA consists of 4 subtasks: listening comprehension, letter identification (names and sounds), nonword reading, and oral reading fluency with comprehension. Phonemic awareness and familiar word reading are non-core subtasks that are also often administered. Other subtasks that are available and occasionally administered are dictation, phoneme segmentation, maze and cloze. The core EGMA includes items on number identification, number discrimination (relative size), number pattern identification, and addition and subtraction (including worded problems). Mean raw scores by module clearly identify the relative performance of sub-groups of students on a specific skill. Guides are available to assist countries in adapting, translating and implementing the assessments.

The Strengths and Difficulties Questionnaire (SDQ) is an assessment designed for children and young people between 3 and 16 years. It is primarily designed to assess the Wellbeing and Social and Emotional Learning (SEL) of children in this age group. The assessment comes in different versions and formats to cater to researchers, clinicians, and educationalists. Although the SDQ is widely used and has been translated into more than 80 languages, has been validated, and has strong psychometric properties (Yao, S et al., 2009; Muris, P et al., 2003; Goodman. R, 2001), it has been excluded for the purposes of this current review, as it is recommended to be used predominantly as a screening tool (Mieloo, C. et al., 2012) for at-risk groups to assess their psychological attributes.

## Stage 3: Definition of key assessment components

To ensure a standardised approach to evaluating each assessment, researchers defined key assessment components to serve as an evaluation framework. The evaluation framework was based on existing reviews of assessments or other frameworks for evaluating assessment (Ainley, J et al., 2021; Cloney et al., 2019; Fernald et al., 2017; GEM Centre, & UIS, 2017; Halle & Darling-Churchill, 2016; Jackson et al., 2019; Munoz-Chereau, B et al., 2021). While the focus of this paper is on early childhood assessments, the literature gathered at this step did not need to be specifically related to early

childhood development and learning, although early childhood was prioritised. See Appendix C: for a list of frameworks analysed in this step.

Table 3 presents the key assessment components used to evaluate the 14 assessments (or sets of assessments[4]) included in the review.

**Table 3: Key assessment components**

| Component | Details |
|---|---|
| Purpose(s) of the assessment (as stated by assessment publishers) | Population monitoring (yes, no)<br>Longitudinal cohort studies (yes, no)<br>Impact/program evaluation (yes, no)<br>Provides evidence for informing policy decisions (yes, no)<br>Individual diagnosis/screening (yes, no) |
| Age range | in years |
| Domain(s) – for example, literacy, language, numeracy, executive function, motor, social-emotional | High – Multiple learning domains OR One learning domain plus at least one health or psychosocial wellbeing domain[5]<br>Medium – One learning domain only<br>Low – Global/ holistic development domain only |
| Learning progression | yes/no |
| Assessment framework | High – Extensive details of construct coverage<br>Low – Limited details of construct coverage |
| Administration format | Direct child assessment (yes, no)<br>Indirect caregiver/educator report (yes, no) |
| Data collection | Paper/pen (yes, no)<br>Digital – for example, PC or tablet (yes, no) |
| Data scoring | High – Detailed information or advice available about data processing (including scoring) and reporting of converting raw scores onto a ordinal/linear scale<br>Medium – Basic scoring advice using raw scores (including benchmarks or cut scores)<br>Low – Advises the use of sum scores or no scoring guidance provided |
| Item type(s) - yes/no, correct/incorrect, partial credit, multiple choice, open text, constructed response, interactive | One type only<br>Multiple types |
| Background / contextual information collected | Child (yes, no)<br>Parent(s)/caregiver(s) and home learning environment (yes, no)<br>Engagement and experience in ECE or school (yes, no) |

---

[4] As described in Table 2.

[5] The definition used here is aligned with SDG 4.2 (and ECDI2030), where the following domains include, but are not limited to:
- Health: gross motor development, fine motor development and self-care.
- Learning: expressive language, literacy, numeracy, pre-writing, and executive functioning.
- Psychosocial well-being: emotional skills, social skills, internalizing behaviour, and externalizing behaviour.

| Component | Details |
|---|---|
| Length/time | (0-10 minutes; 10-20 minutes; >20 minutes) |
| Available languages | High – Translation/adaption guide available plus assessment forms in multiple languages available |
| | Medium – Translation/adaption guide available, but no available assessment forms in multiple languages |
| | Low – No translation/adaption guide or no available assessment forms in multiple languages |
| Availability/cost | High – Open source (freely available) |
| | Low – Incurs cost |
| Training materials/requirements/capacity development | High – Established training protocol with a requirement for enumerators to attend administration training |
| | Medium – Training materials provided |
| | Low – No/minimal training recommended with no/basic materials provided |
| Validation for use across different populations | High – Validated using appropriate measurement invariance analysis methods and shown to be either fully or partially invariant across countries/languages |
| | Low – Validated in different populations separately; or shown to be non-invariant across countries/languages; or no evidence of invariance across countries/languages found |

Appendix D: outlines each assessment (or group of assessments) based on these key assessment components.

# Results and discussion

## What assessments are available to support the monitoring and reporting of SDG Indicator 4.2.1?

The literature search identified 48 assessments (see Appendix B: : List of assessments screened), of which 25 were included in the review as individual assessments or a set of assessments. In total, 14 reviews were undertaken.

The 14 assessments / sets of assessments that were reviewed are:

- Ages & Stages Questionnaires, Third Edition (ASQ-3), University of Oregon

- Caregiver Reported Early Development Instruments (CREDI), Harvard University

- Early Childhood Development Assessment Scale (ECDAS), The University of Hong Kong

- Early Childhood Development Index 2030 (ECDI2030), UNICEF

- Early Development Instrument (EDI), Offord Centre for Child Studies, McMaster University

- Early Grade Reading Assessment (EGRA) and Early Grade Mathematics Assessment (EGMA), RTI International

- Early Human Capability Index (eHCI), Telethon Kids Institute

- East Asia-Pacific Early Child Development Scale (EAP-ECDS), UNICEF, the Asia-Pacific Regional Network for Early Childhood (ARNEC), Open Society Foundations (OSF), The University of Hong Kong (Technical partners)

- Global Scales for Early Development (GSED), World Health Organisation (WHO)

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

- Infant and Young Child Development (IYCD), World Health Organisation (WHO)

- International Early Learning and Child Well-being Study (IELS), Organisation for Economic Cooperation and Development (OECD)

- Measuring Early Learning Quality and Outcomes (MELQO), UNESCO, UNICEF, Brookings Institution, World Bank

- Citizen-led assessments and common assessments of the People's Action for Learning (PAL) Network.

Given the criteria for inclusion in this review, all reviewed assessments contain the necessary elements that allow for the potential to report against SDG Indicator 4.2.1. That said, the assessments vary in the extent to which they adhere to the key assessment components:

- Purpose: While all assessments are able to measure at the population level, and most were specifically developed for this purpose, some assessments were designed to have multiple purposes, including population monitoring. One assessment, the ASQ-3, is specifically used for developmental screening but has also been used in England as a population measure (Department of Health (UK), 2016).

- Age range: All assessments target the age range 24 to 59 months, at least in part.

- Domain(s): All assessments have strong overlap with the focus areas of learning and development. Almost all framework documents reviewed put a strong emphasis on the holistic nature of learning and explicitly mention aspects of health, learning and psychosocial wellbeing (typically described in terms of social and emotional skills).

- Learning progression: None of the assessments reviewed explicitly referred to a learning progression.

- Assessment framework: Most assessments were found to have well-established assessment frameworks.

- Administration format: Some assess a child's development indirectly, administering expert informant questionnaires to parents/caregivers or educators, while others use one-on-one direct assessments of children.

- Data collection: Some assessments are administered using paper and pen, others are administered digitally, while many offer the option to choose which method best suits the context.

- Data scoring: Detailed scoring guides are provided for most assessments.

- Item type(s): The reviewed assessments predominantly include dichotomous items with some using a mix of dichotomous and other item types: for example, multiple choice, count, likert or open-ended items.

- Background/contextual information collected: All assessments use items which gather information relating to the background of the child under examination (age, gender, language spoken). Many of them also include items relating to the background of the parent(s)/caregiver(s) and the home learning environment. Some of the assessments include items which capture information relating to the child's engagement and experience with ECE, kindergarten or school.

- Length/time: Some assessments are relatively short, taking less than 5 minutes, and others are longer, taking 20 minutes or more to administer.

- Available languages: Every assessment included in this review has been trialled in multiple different contexts (for example, across districts, countries, languages). Most of the assessments were originally developed in English; however, all have been trialled in multiple languages, have published the assessment in multiple languages or provide detailed adaptation (including translation) guides for use in different contexts.

- Availability/cost: Most of the assessments are freely available for download and use.

- Training materials/requirements/capacity development: In all cases, the importance of enumerator training is emphasised; however, the amount of training suggested or required differs between assessments, from simply reviewing administration guides to attending 5 days of extensive training.

- Validation for use across different populations: Many of the assessments have used either classical test theory or item response theory methods to assess the validity and reliability of the measurement domains under examination for each context.

An overview of each assessment based on the key assessment components is provided in Appendix D: : Summaries of included assessments. Specific details on how each assessment compared to the key assessment components is described in Appendix E: Summary of assessments against key reporting components.

# What are the strengths of these assessments?

## General alignment with SDG Indicator 4.2.1.

This review has found 14 assessment options that can currently be used, or have the potential to be used, for reporting against SDG Indicator 4.2.1. As mentioned above, they all cover (at least partially) the correct age range; they can be used to measure a population; they focus on at least one of the targeted learning and development domains; they have been designed for use across multiple countries/languages; they have been psychometrically validated (to varying degrees of robustness); and they are available to use at no cost or for a once-off fee.

## Well-established assessment frameworks

Most assessments were found to have well-established assessment frameworks. Many of these provide at least partial coverage of SDG Indicator 4.2.1, particularly relating to learning. The degree of coverage is highly variable, however. This is mostly due to the definitions of the constructs. The ECDI2030, for example, reports holistic learning and development as a single, unidimensional (or overall) score with no disaggregation by domain. The EDI, by comparison, reports against more general but disaggregated (multidimensional) constructs including *Language and Cognitive Development* and *Communication Skills and General Knowledge*. While these domains are related to more traditional domains of learning observed in learning frameworks (such as literacy and numeracy frameworks), they are more holistically defined in a way that is typical of epidemiological or population health measures. Other assessments are more strongly aligned with domains of learning and use academic and cognitive constructs – for example, (Early) Literacy, Mathematics, and Executive Function (IDELA, IELS, MELQO, PAL, EGRA and EGMA).

## Direct assessment of children

Many of the reviewed assessments directly assess the child's learning and development (ECDAS-DA, EGRA and EGMA, EAP-ECDS, GSED-LF, IDELA, IELS, MELQO-MODEL, and PAL). Others rely solely on indirect methods – parent/caregiver or educator reports – to gain information about children's learning and development (ASQ, CREDI, ECDI2030, EDI, eHCI and IYCD). Most of the assessments that do employ direct assessment of children also conduct parent/caregiver or teacher surveys to triangulate the data and gain further information about the children's contexts. Gathering data directly from the child as well as from other close sources helps to strengthen the validity of the assessment's results and their interpretation.

## Possibility for measurement expansion

A small number of assessments use approaches that potentially allow for sophisticated population measurement. Some of these assessments, including the GSED, CREDI, PAL and IELS, are at least partially aligned with SDG Indicator 4.2.1, and because of their more sophisticated analytic underpinnings could be developed over time to align more fully, such that trends could be quantified (and the quality of trends assessed), and

where international comparisons could be made. Further, these measures could be empirically aligned with the ECDI2030 to provide a mapping to SDG Indicator 4.2.1 from a variety of assessments.

## Ability to administer the assessment across countries/languages

Based on the psychometric properties of an assessment (albeit to varying degrees of robustness), some researchers and assessment publishers have gone on to suggest that the assessment is able to be used for cross-cultural comparisons, as the results of the individual analyses show that the psychometric properties are adequate for each of the individual contexts studied. This applies to MELQO (Pushparatnam et al., 2021) and GSED (World Health Organization, 2023h). However, only 3 of the included assessments – EDI: (Duku et al., 2015), IDELA (P. F. Halpin et al., 2019) and IELS: (OECD, 2021) – have published details of the complex analyses examining the cross-cultural/linguistic comparability of the assessment items using appropriate measurement invariance analyses (for example, multigroup CFA, differential item functioning (DIF)). PAL Network's ICAN and ELANA have also undergone measurement invariance testing but details are not yet publicly available.

With respect to administration support, many assessments have been translated into several languages and/or make administration materials available to support consistent implementation across contexts. These materials include training guides, adaptation and translation guides and directions regarding enumeration, data entry and reporting. Assessments that provide this level of administrative support include ECDI2030, MELQO, IDELA, GSED, PAL, CREDI, and EGRA and EGMA.

### Published validation studies

Several assessments have published validation studies, including those using factor analytic models (for example, MELQO) and item response theory (IRT) models (for example, CREDI, GSED) to assess the functioning of items, the correlation between subdomains, and the distribution of abilities observed in samples.

## What are the limitations of these assessments?

### Use of small item pools

Most of the assessments have relatively small item pools, the ECDI2030 for example being a single static form of 20 items. Exceptions are the GSED (long form) with 155 items, CREDI (long form) with 100 items, EDI with 103 items, core EGRA includes 96 items, EAP-ECDS LF includes 85 items, ELANA with over 700 items, IELS with 157 items, and IYCD with 100 items. The GSED and ELANA are the only tools to explicitly make adaptive testing[6] available, though many tools use basal and ceiling rules to

---

[6] "In conventional assessments, students answer the same questions. But in adaptive assessments, students' responses determine the content, resulting in personalised test pathways. The result is a highly detailed picture of achievement, enabling more precise diagnosis of student needs and targeting teaching and policy where it is needed most." (Masters, 2021)

attempt to limit assessment length and appropriately target the items (for example, ASQ-3, CREDI). This has implications for the coverage of the construct as more items provide an opportunity to cover a broader range of sub-skills within the domain as well as a wider range of abilities.  For example, the ability to count or recognise shapes may act as an indicator of mathematical ability, but the specific information is lost in the single number reported. This makes it difficult to use the results to inform policy focussing on how early childhood programs and professionals can improve children's learning outcomes. Some large-scale assessment programs in education have developed sophisticated methods and large item pools to ensure that the broad and multidimensional aspects of learning and development can be described, and rich profiles of strengths and limitations can be identified. This is particularly relevant for population measurement where there is a substantial opportunity to rotate many items across a sample or census. There is also an opportunity to break down the assumption that assessment can only be used for one-off population reporting or evaluation, or formative purposes. Instead, a mature item pool can develop fit-for purpose assessments on demand that meet the requirement at hand and can link the result back to a common, valid, and reliable metric.

## Use of raw scores

Many of the existing assessments related to SDG 4.2.1 rely on raw score-based approaches to measurement, typically using summed or averaged scores to represent the ability of children on the scale. This assumes complete data, or requires adjusting for non-response to yield an equivalent raw score (as implemented in ASQ-3). Some assessments rely on methodological practices that use single and fixed form assessments that are not dynamic and cannot be changed and compared over time – leading to the proliferation of assessments using the same item pool with items added or subtracted, creating new data sets each time. This can also result in limited trend capacity, including comparisons across time and settings, as such assessments are either restricted to the same age group for re-administration, or require the administration of very long tests to cover the full scope of skills. However, some assessments provide varying levels of guidance and resources to transform raw scores onto an ordinal/linear scale, including CREDI, EAP-ECDS, GSED, IYCD, IELS, MELQO MODEL, ICAN and ELANA. For example, CREDI and GSED assessments: both come with pre-written syntax and instructions to apply an IRT scoring model using fixed item parameters to raw scores to yield scale scores (maximum a-posteriori (MAPs) for CREDI and expected a-posteriori (EAPs) for GSED).

Providing one overall raw score that represents multiple aspects of learning and development does not allow for results to be linked to a progression of learning and development that leads to school education outcomes. SDG Target 4.2 refers to the progress children make in their learning and development as they move through ECE by explicitly linking access to quality early childhood education in preparation for primary school. SDG Indicator 4.1.1 focuses on school education outcomes, specifically minimum proficiency levels in reading and mathematics. Restricting the reporting of

SDG Indicator 4.2.1 to one overall raw score precludes what is implied in SDG 4.2: the linking across indicators 4.2 and 4.1to describe a continuum of learning.

## Absence of learning progressions

Some of the assessments (EGRA/EGMA, IELS, IDELA, MELQO, PAL) imply progression in their reporting even though they do not have formal described scales or learning progressions. They include many items spanning a large ability range that can show where children are at in their learning and where they can go next. In the measures that use IRT approaches (CREDI, EAP-ECDS, GSED, IYCD, IELS, MELQO MODEL, ICAN and ELANA) the reporting shows that the items span a range of difficulties and therefore imply the kind of skills that children with more advanced skills may demonstrate as their proficiency grows. Other assessments report outcomes as a function of age expectation (for example, the DAZ score in the GSED, ASQ-3) or as developmental cut-offs. For example, ECDI2030 has established raw score cut-offs that represent being developmentally on track as a function of chronological age (that is, we would expect higher raw scores and therefore more advanced learning and development as children get older). Indeed, ECDI2030 is used within the context of the SDG 4.2.1 indicator to set multiple benchmarks. This approach suggests that there is not one single level of being on-track for learning and development – as a single "cut-off" might imply – but rather that there is continuity of learning and development over the early years and beyond.

Ultimately, none of the assessments included in this review currently have learning progressions associated with them, leading to a lack of clarity about what it means for an individual student or a population group to be placed at a certain location on a scale. As a result, it is hard to know what knowledge, skills and understandings children possess and what they might do next to continue developing. This has limitations for policy use: it is not clear what a certain score on these assessments means in practical terms. It is difficult to align learning frameworks with assessment results, resulting in an inability to effectively use assessment results to inform the development or revision of principles, practices and outcomes that educational leaders use to assist in curriculum design and education delivery.

# Conclusion

This report is a review of assessments of early learning that could be used to inform SDG Indicator 4.2.1. ECDI2030 has been identified as the official measure of SDG Indicator 4.2.1, but many other assessments of early learning and development also exist, are in use, or are in development. Many of these assessments have similar properties to the ECDI2030, including the purpose of measuring a population rather than individuals, targeting young children, and assessing a wide range of learning and development. Some of these assessments go beyond the ECDI2030 by using more sophisticated scoring and scaling techniques, capturing a wider breadth of skills, and directly involving children in their assessment.

The identification of a single assessment as the official measure of SDG Indicator 4.2.1 is unlikely to prevent the use of other assessments for a similar purpose. For example, the US is reporting against "Early math skills among US kindergarteners (typically age 5)"[7]. As we have seen, a number of assessments capable of yielding reporting against 4.2.1 are available for use with relatively few restrictions. Use of different assessments is recognised and accommodated to report the parallel school level SDG 4 indicator, SDG Indicator 4.1.1, for which a set of tools, methods and criteria have been developed to align and harmonise findings (UNESCO Institute for Statistics, 2023; UNESCO Institute for Statistics & Australian Council for Educational Research, 2024). This approach could be adopted for SDG Indicator 4.2.1 to provide a consistent yet flexible approach that maximises countries' capacity to participate in SDG 4.2.1 reporting.

All of the identified assessments were included in this review because they were able to satisfy the inclusion/exclusion criteria. This in itself suggests that each of them are useful for reporting against SDG 4.2.1. However, there is significant scope to continue to grow and improve direct assessment of early childhood learning and development within the context of reporting against SDG Indicator 4.2.1. This includes expanding item pools and using sophisticated scoring and analytic methods, specifically the use of IRT. IRT enables the capability of individuals to be placed within a broad standardised scale that can include those with very high or very low abilities. Furthermore, it allows for better comparability of scores across different versions of a test or across different populations (Adams, 2017). These are advantages over classical test theory, which many of the existing assessments still apply.

Importantly, these improvements would allow for the development of learning progressions that link measures of early childhood education with school education, thus articulating that learning is a continuum (GEM Centre, 2021). By identifying the state of a student's learning progress, educators can be informed of the appropriate teaching and support for each learner. Big picture statements typically included in learning progression provide policymakers, curriculum developers and groups involved in setting performance expectations with useful information about the descriptions of the domain(s), describe what progress looks like, and highlight why progress is valued (Waters, 2019, para. 16). Thus, assessments applying learning progressions can be used for large-scale measurement, as well as for formative purposes.

Room for improvement was identified for most of the assessments reviewed. By having a variety of validated measurement options at hand, governments can choose which assessments best suit their monitoring needs and their context. Measuring progress towards ensuring inclusive and equitable quality education and lifelong learning for all by 2030 (SDG 4), is most worthwhile when the resulting data provides policymakers and practitioners with an understanding of student learning levels and gaps – the evidence needed for developing strategies to improve learning for all.

---

[7] Sourced from U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 2010–11 (ECLS-K:2011): https://sdg.data.gov/4-2-1/.

# References

ACER. (2020a). *The Early Grade Mathematics Assessment: Assessing children's acquisition of basic numeracy skills in developing countries*. https://research.acer.edu.au/assessgems/12

ACER. (2020b). *The Early Grade Reading Assessment: Assessing children's acquisition of basic literacy skills in developing countries*. https://research.acer.edu.au/assessgems/17

Adams, R. (2017). Working with UIS towards a global measurement scale for learning. *International Developments*, 7, 6–7. ACEResearch.

Adams, R., Jackson, J., & Turner, R. (2018). *Learning progressions as an inclusive solution to global education monitoring*. Australian Council for Educational Research. https://research.acer.edu.au/cgi/viewcontent.cgi?article=1032&context=monitoring_learning

Ages & Stages Questionnaires®. (2018). *Commonly used parent-report developmental screening tools*. Paul H. Brookes Publishing Co., Inc. https://agesandstages.com/wp-content/uploads/2019/01/ASQ-Comparison-Chart-2018.pdf

Ages & Stages Questionnaires®. (n.d.a). *For parents*. Ages & Stages Questionnaires (ASQ). https://agesandstages.com/wp-content/uploads/2015/02/asq-3-48-month-sample.pdf

Ages & Stages Questionnaires®. (n.d.b). *Training*. Ages & Stages Questionnaires (ASQ). https://agesandstages.com/wp-content/uploads/2015/02/asq-3-48-month-sample.pdf

Ages & Stages Questionnaires®,  author. ), Diane. (2009). *48 Month Questionnaire*. Paul H. Brookes Publishing Co., Inc. https://agesandstages.com/wp-content/uploads/2015/02/asq-3-48-month-sample.pdf

Ainley, J, Lietz, P, Jeffries, D, Andrews, N, & Ahmed, S. K. (2021). *Review and advice on assessment across the South Australian Public education system: Measuring what matters.* Australian Council for Educational research.

Ainley, J., Lietz, P., Jeffries, D., Andrews, N., & Ahmed, S. K. (2021). *Review and advice on assessment across the South Australian Public education system: Measuring what matters.* Australian Council for Educational Research.

ASER Centre. (2022). *Annual Status of Education Report (Rural) 2022*. https://img.asercentre.org/docs/ASER%202022%20report%20pdfs/All%20India%20documents/aserreport2022.pdf

Australian Council for Educational Research. (2020). *International Early Learning and Child Well-Being Study*. Australian Council for Educational Research. https://research.acer.edu.au/assessgems/20

Australian Council for Educational Research. (2022a). *How to use ACER's learning*

*progressions*. https://mfb.acer.edu.au/au/pat/pat-insights/how-to-use-acers-learning-progressions

Australian Council for Educational Research. (2022b). *Pairwise comparison method: Concept note* (3; GEM). https://research.acer.edu.au/gem/3

Australian Council for Educational Research. (2022c). *Pairwise comparison method: Concept note*. Australian Council for Educational Research.
https://research.acer.edu.au/gem/3
1st pass

Australian Early Development Census. (2017). *AEDC guide to completing the Australian version of the Early Development Instrument, adapted for the Australian Early Childhood Educational Development Experience (AECEDE) Research Project*. Department of Education and Training.
https://www.aedc.gov.au/Websilk/Handlers/ResourceDocument.ashx?id=a7582764-db9a-6d2b-9fad-ff0000a141dd

Australian Education Research Organisation. (2023). *Early childhood learning trajectories: The evidence base*. https://www.edresearch.edu.au/resources/early-childhood-learning-trajectories-evidence-base

Beggs, C. (2016, May 11). *Overview of EGRA and EGMA*. Global Alliance to Monitor Learning. https://gaml.uis.unesco.org/wp-content/uploads/sites/2/2018/08/4_USAID_20160511.pdf

Brinkman, S. A., & Kinnell, A. (2013). *The Early Human Capability Index*. Telethon Kids Institute. https://ehci.telethonkids.org.au/siteassets/media-docs---ehci/ehci-by-country/ehci-australiaenglish.pdf

Brinkman, S. A., & Kinnell, A. (2022). *About the eHCI*. Early Human Capability Index (eHCI). https://ehci.telethonkids.org.au/about/

Campbell, S. B., Denham, S. A., Howarth, G. Z., Jones, S. M., Whittaker, J. V., Williford, A. P., Willoughby, M. T., Yudron, M., & Darling-Churchill, K. (2016). Commentary on the review of measures of early childhood social and emotional development: Conceptualization, critique, and recommendations. *Journal of Applied Developmental Psychology*, *45*, 19–41.

Cappa, C. (2019). *Monitoring SDG 4.2. 1: A briefing from the INTER-agency and expert group on ECD measurement*.
In APHA's 2019. Annual Meeting and Expo (Nov. 2-Nov. 6)

Cappa, C., Petrowski, N., De Castro, E. F., Geisen, E., LeBaron, P., Allen-Leigh, B., Place, J. M., & Scanlon, P. J. (2021). 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*, *18*(22), 12181. https://doi.org/10.3390/ijerph182212181

Cloney, D., Jackson, J., & Mitchell, P. (2019). *Assessment of children as confident and involved learners in early childhood education and care: Literature review*. Victorian Curriculum and Assessment Authority. https://www.vcaa.vic.edu.au/Documents/earlyyears/EYLitReviewLearning.pdf SG done

Confrey, J. (2019). *A synthesis of research on learning trajectories/progressions in mathematics* (Future of Education and Skills 2030: Currriculum Analysis, p. 58). Organisation for Economic Co-operation and Development. https://www.oecd.org/education/2030-project/about/documents/A_Synthesis_of_Research_on_Learning_Trajectories_Progressions_in_Mathematics.pdf

Consortium for Research on Early Childhood Development and Education (CORE). (2017). *Assessing Early Development and Learning: The East Asia-Pacific Early Childhood Development Scales (EAP-ECDS).EAP-ECDS BRIEF October 2017.* The University of Hong Kong. https://hkece.edu.hku.hk/content/uploads/2020/11/EAPECDS-brief_Oct-25-2017.pdf

CORE. (2022). *Child Assessment—ECDAS*. COnsortium for REsearch on Early Childhood Development and Education. https://hkece.edu.hku.hk/ecdas/

CREDI team. (2023). *CREDI Materials*. https://credi.gse.harvard.edu/credi-materials

Department of Health (UK). (2016). *Factsheet: Developing a public health outcome measure for children aged 2 – 2½ using ASQ-3™*. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/550487/ASQ-3_September_2016.pdf

Drummond, T., & Nakamura, P. R. (2021). An examination of early grade reading assessments in Central Asia: Using factor analysis to determine the latent data structure in Kyrgyz, Russian and Tajik. *Reading and Writing*, *34*, 1343–1366. https://doi.org/10.1007/s11145-020-10110-9

Duku, E., Janus, M., & Brinkman, S. (2015). Investigation of the cross-national equivalence of a measurement of early child development. *Child Indicators Research*, *8*, 471–489. https://doi.org/10.1007/s12187-014-9249-3

Fernald, L. C. H., Prado, E., Kariger, P., & Raikes, A. (2017). *A toolkit for measuring early childhood development in low and middle-income countries.* World Bank. https://openknowledge.worldbank.org/handle/10986/29000

GEM Centre. (2021). *Monitoring, evaluation and learning framework: Global education monitoring centre.* Australian Council for Educational Research.

GEM Centre, & UIS. (2017). *Principles of good practice in learning assessment*. Australian Council for Educational Research. https://research.acer.edu.au/monitoring_learning/35/

GEM Centre & UIS. (2017). *Principles of good practice in learning assessment. Australian Council for Educational Research.* Australian Council for Educational Research. https://research.acer.edu.au/cgi/viewcontent.cgi?article=1035&context=monitoring_learning

Gladstone, M., Lancaster, G. A., McRay, G., Cavallera, V., Alves, C. R. L., Maliwichi, L., Rasheed, M. A., Dua, T., Janus, M., & Kariger, P. (2021). Validation of the Infant and Young Child Development (IYCD) indicators in three countries: Brazil, Malawi and Pakistan. *International Journal of Environmental Research and Public Health*, *18*(11), 1–19. https://doi.org/10.3390/ijerph18116117

GSED Team. (2019). The global scale for early development (GSED. *Early Childhood Matters*, *14*, 80–84.

GSED Team. (2021). *The global scales for early development (GSED) project*. WHO. https://cdn.who.int/media/docs/default-source/mental-health/gsed-project_overview_october2021.pdf?sfvrsn=28c5c484_11

Halle, T. G., & Darling-Churchill, K. E. (2016). Review of measures of social and emotional development. *Journal of Applied Developmental Psychology*, *45*, 8–18. https://doi.org/10.1016/j.appdev.2016.02.003

Halpin, P., De Castro, E. F., Petrowski, N., & Cappa, C. (2023). *Monitoring early childhood development at the population level: The ECDI2030* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/6qcjb

Halpin, P. F., Wolf, S., Yoshikawa, H., Rojas, N., Kabay, S., Pisani, L., & Dowd, A. J. (2019). Measuring early learning and development across cultures: Invariance of the IDELA across five countries. *Developmental Psychology*, *55*(1), 23–37. https://doi.org/10.1037/dev0000626

Heckman, J. J. (n.d.). *Invest in early childhood development: Reduce deficits, strengthen the economy.* The Heckman Equation. Retrieved January 23, 2024, from https://heckmanequation.org/wp-content/uploads/2013/07/F_HeckmanDeficitPieceCUSTOM-Generic_052714-3-1.pdf

Jackson, J., Adams, R., Schwantner, U., & Waters, C. (2018). The role of learning progressions in global scales. *Research Developments, ACER*. https://rd.acer.org/article/the-role-of-learning-progressions-in-global-scales

Jackson, J., Ahmed, S. K., Carslake, T., & Lietz, P. (2019). *Improving young children's learning in economically developing countries: What works, why, and where? Scoping review.* Australian Council for Educational Research. https://research.acer.edu.au/cgi/viewcontent.cgi?article=1038&amp;context=monitoring_learning

Jacobusse, G., Buuren, S., & Verkerk, P. H. (2006). An interval scale for development of children aged 0–2 years. *Statistics in Medicine*, *25*(13), 2272–2283. https://doi.org/10.1002/sim.2351

Janus, M., Brinkman, S., & Duku, E. (2011). Validity and psychometric properties of the Early Development Instrument in Canada, Australia, United States, and Jamaica. *Social Indicators Research*, *103*, 283–297. https://doi.org/10.1007/s11205-011-9846-1

Janus, M., Brinkman, S., Duku, E., Hertzman, C., Santos, R., Sayers, M., Schroeder, J., & Walsh, C. (2007). *The Early Development Instrument: A population-based measure for communities—A handbook on development, properties, and use*. Offord Centre for Child Studies.

Janus, M., & Offord, D. R. (2007). Development and psychometric properties of the Early Development Instrument (EDI): A measure of children's school readiness. *Canadian Journal of Behavioural Science*, *39*(1), 1–22. https://doi.org/10.1037/cjbs2007001

Janus, M., & Reid-Westoby, C. (2016). Monitoring the development of all children: The Early Development Instrument. *Early Childhood Matters*, *7*, 40–45.

Kan, S., Fahez, M., & Valenza, M. (2022). *Foundational literacy and numeracy in rural Afghanistan: Findings from a baseline learning assessment of accelerated learning centres*. UNICEF. https://www.unicef-irc.org/publications/pdf/Foundational-literacy-and-numeracy-in-rural-Afghanistan_Findings-from-a-baseline-learning-assessment-of-accelerated-learning-centres.pdf

Kendall, S., Nash, A., Braun, A., Bastug, G., Rougeaux, E., & Bedford, H. (2014). *Evaluating the use of a population measure of child development in the Healthy Child Programme Two Year Review*. UCL Institute of Child Health, University of Hertfordshire. https://www.ucl.ac.uk/children-policy-research/sites/children_policy_research/files/evaluating-the-use-of-a-population-measure-of-child-development-in-the-healthy-child-two-year-review.pdf

Kendall, S., Nash, A., Braun, A., Bastug, G., Rougeaux, E., & Bedford, H. (2019). Acceptability and understanding of the Ages & Stages Questionnaires®, Third Edition, as part of the Healthy Child Programme 2-year health and development review in England: Parent and professional perspectives. *Child: Care, Health and Development*, *45*(2), 251–256. https://doi.org/10.1111/cch.12639

Lancaster, G. A., McRay, G., Kariger, P., Dua, T., Titman, A., Chandna, J., McCoy, D. C., Abubakar, A., Hamadani, J. D., Fink, G., Tofail, F., Gladstone, M., & Janus, M. (2018). Creation of the WHO Indicators of Infant and Young Child Development (IYCD): Metadata synthesis across 10 countries. *BMJ Global Health*, *3*, 1–15. https://doi.org/10.1136/bmjgh-2018-000747

Masters, G. (2021). Adaptive assessment: A fairer way to test. *ACER Discover*. https://mfb.acer.edu.au/au/discover/article/adaptive-assessment-a-fairer-way-to-test

McCoy, D. C., Sudfeld, C. R., Bellinger, D. C., Muhihi, A., Ashery, G., Weary, T. E., Fawzi, W., & Fink, G. (2017a). Development and validation of an early childhood development scale for use in low-resourced settings. *Population Health Metrics*, *15*(1), 3. https://doi.org/10.1186/s12963-017-0122-8

McCoy, D. C., Sudfeld, C. R., Bellinger, D. C., Muhihi, A., Ashery, G., Weary, T. E., Fawzi, W., & Fink, G. (2017b). Development and validation of an early childhood development scale for use in low-resourced settings. *Population Health Metrics*, *15*. https://doi.org/10.1186/s12963-017-0122-8

McCoy, D. C., Waldman, M., Team., C. R. E. D. I. F., & Fink, G. (2018). Measuring early childhood development at a global scale: Evidence from the Caregiver-Reported Early Development Instruments. *Early Childhood Research Quarterly*, *45*, 58–68. https://doi.org/10.1016/j.ecresq.2018.05.002

McCoy, D. C., Zuilkowski, S. S., Yoshikawa, H., & Fink, G. (2017). Early Childhood Care and Education and School Readiness in Zambia. *Journal of Research on Educational Effectiveness*, *10*, 482–506. https://doi.org/10.1080/19345747.2016.1250850

McMahon, W. W. (2010). *The external benefits of education.* Elsevier.

Mosher, F., & Heritage, M. (2017). *A hitchhiker's guide to thinking about literacy, learning progressions, and instruction* (RR 2017-2; CPRE Research Reports). Consortium for Policy Research in Education. http://www.cpre.org/sites/default/files/ccii_final_reportvaug11.pdf

Munoz-Chereau, B, Ang, L, Dockrell, J, Outhwaite, L, & Heffernan, C. (2021). Measuring early child development across low and middle-income countries: A systematic review. *Journal of Early Childhood Research*, *19*(4), 443–470.

National Research Council. (2000). *How people learn: Brain, mind, experience, and school: Expanded edition*. The National Academies Press. Doi:10.17226/9853

OECD. (2021). *International Early Learning and Child Well-being Study: Technical Report*. OECD. https://www.oecd.org/education/school/early-learning-and-child-well-being-study/IELS_2018_Technical_Report%20with%20coverpdf.pdf

Offord Centre for Child Studies. (2018). *Early Development Instrument: A population based measure for communities—Ontario 2017/2018*. https://edi-offordcentre.s3.amazonaws.com/uploads/2019/01/EDI-ON-ENG-2018.pdf

Offord Centre for Child Studies. (2019a). *Early Development Instrument: Interpretation toolkit*. https://edi-offordcentre.s3.amazonaws.com/uploads/2019/03/EDI-interpretation-toolkit.pdf

Offord Centre for Child Studies. (2019b). *EDI Domains and Subdomains*. Early Development Instrument. https://edi.offordcentre.com/researchers/domains-and-subdomains/

Offord Centre for Child Studies. (2019c). *EDI FAQ*. Early Development Instrument. https://edi.offordcentre.com/researchers/faq/

Offord Centre for Child Studies. (2019d). *EDI International implementations*. Early Development Instrument. https://edi.offordcentre.com/partners/international/

Offord Centre for Child Studies. (2019e). *EDI Newfoundland*. Early Development Instrument. https://edi.offordcentre.com/teachers/edi-implementation-

resources/newfoundland/

Offord Centre for Child Studies. (2019f). *EDI Resources*. Early Development Instrument. https://edi.offordcentre.com/resources/

Offord Centre for Child Studies. (2019g). *History of the EDI*. Early Development Instrument. https://edi.offordcentre.com/about/history-of-the-edi/

Offord Centre for Child Studies. (2019h). *How to use the EDI*. Early Development Instrument. https://edi.offordcentre.com/researchers/how-to-use-the-edi/

Offord Centre for Child Studies. (2019i). *Interested in using the EDI?* Early Development Instrument. https://edi.offordcentre.com/contact/interested-in-using-the-edi/

Offord Centre for Child Studies. (2019j). *Requirements / Responsibilities of EDI users*. Early Development Instrument. https://edi.offordcentre.com/researchers/requirementsresponsibilities-of-edi-users/

Offord Centre for Child Studies. (2019k). *What is the EDI?* Early Development Instrument. https://edi.offordcentre.com/about/what-is-the-edi/

Organisation for Economic Co-operation and Development. (2018). *Early learning matters*. OECD. https://www.oecd.org/education/school/Early-Learning-Matters-Project-Brochure.pdf

PAL Network. (2020a). *ICAN: International Common Assessment of Numeracy. Background, features and large-scale implementation*. People's Action for Learning Network. https://palnetwork.org/download/94/ican-full-report/9749/2020_pal-network_ican-full-report.pdf

PAL Network. (2020b). *ICAN policy linking for measuring global learning outcomes: Workshop report*. People's Action for Learning Network. https://palnetwork.org/download/94/ican-full-report/10279/ican-policy-linking-report.pdf

PAL Network. (2022a). *Citizen led assessments*. https://palnetwork.org/citizen-led-assessments/

PAL Network. (2022b). *Common assessments*. https://palnetwork.org/common-assessments/

PAL Network. (2022c). *Tests and tools*. https://palnetwork.org/tests-and-tools/

Petrowski, N., Cappa, C., & Dia, E. (2022). Two years on: Progress and country uptake of the ECDI2030. *UNICEF Data Blog*. https://data.unicef.org/data-for-action/two-years-on-progress-and-country-uptake-of-the-ecdi2030/

Phair, R. (2021). *International Early Learning and Child Well-being Study Assessment Framework* (246; OECD Education Working Papers). OECD Publishing. https://doi.org/10.1787/af403e1e-en

Pisani, L., Borisova, I., & Dowd, A. J. (2017). *International Development and Early Learning Assessment Technical Working Paper*. Save the Children US.

Pisani, L., Borisova, I., & Dowd, A. J. (2018). Developing and validating the International Development and Early Learning Assessment (IDELA). *International Journal of Educational Research, 91*, 1–15. https://doi.org/10.1016/j.ijer.2018.06.007

Pushparatnam, A., Luna Bazaldua, D. A., Holla, A., Azevedo, J. P., Clarke, M., & Devercelli, A. (2021). Measuring Early Childhood Development among 4–6 year olds: The identification of psychometrically robust items across diverse contexts. *Frontiers in Public Health, 9*, 569448. https://doi.org/10.3389/fpubh.2021.569448

Raghaven, C., & Santiago, E. (2016). *The East Asia Pacific Early Childhood Development Scales (EAP-ECDS): Improving the quality of early childhood development through systems-level monitoring*. UNICEF EAPRO and ARNEC. https://arnec.net/static/uploads/1544519683421-EAPR%20ECD%20Scales%20One%20Pager%20April%202016%20image_ModifiedV1.0Final-2.pdf

Raikes, A., Koziol, N., Janus, M., Platas, L., Weatherholt, T., Smeby, A., & Sayre, R. (2019). Examination of school readiness constructs in Tanzania: Psychometric evaluation of the MELQO scales. *Journal of Applied Developmental Psychology, 62*, 122–134. https://doi.org/10.1016/j.appdev.2019.02.003

Rao, N., Chan, S. W. Y., Lee, D., & Becher, Y. (2020a). *Early childhood development assessment scale – Direct assessment (ECDAS-DA*. Faculty of Education, The University of Hong Kong.

Rao, N., Chan, S. W. Y., Lee, D., & Becher, Y. (2020b). *Early childhood development assessment scale—Caregiver survey (ECDAS-CS*. Faculty of Education, The University of Hong Kong.

Rao, N., Chan, S. W. Y., Su, Y., Richards, B., Cappa, C., Castro, E. F., & Petrowski, N. (2021). *Measuring being "Developmentally on track": Comparing direct assessment and caregiver report of early childhood development in*. Early Education and Development. https://doi.org/10.1080/10409289.2021.1928446

Rao, N., Su, Y., & Chan, S. W. Y. (2023). Reliability and validity of the East Asia-Pacific early child development scales: A longitudinal validation study in China. *Child Psychiatry & Human Development*. https://doi.org/10.1007/s10578-023-01526-9

Rao, N., Sun, J., Ng, M., Becher, Y., Lee, D., Ip, P., & Bacon-Shone, J. (2014). *Validation, finalization and adoption of the East Asia-Pacific Early Child Development Scales (EAP-ECDS). UNICEF*. https://arnec.net/static/uploads/EAP-ECDS

Richards, B., Rao, N., & Chan, S. W. Y. (in press). Measuring indicators of Sustainable Development Goal Target 4.2.1: Factor structure of a direct assessment tool in four Asian countries. *Oxford Review of Education*.

Richter, J., & Janson, H. (2007). A validation study of the Norwegian version of the Ages and Stages Questionnaires. *Acta Paediatrica, 96*(5), 748–752.

Richter, L., Black, M., Britto, P., Desmond, C., Devercelli, A., Dua, T., Fink, G., Heymann, J., Lombardi, J., Lu, C., Naicker, S., & Vargas-Baron, E. (2019). Early childhood

development: An imperative for action and measurement at scale. *BMJ Global Health*, 4, 154–160. https://doi.org/10.1136/bmjgh-2018-001302

RTI International. (2010). *Snapshot of School Management Effectiveness (SSME) FAQs*. USAID. https://www.globalreadingnetwork.net/sites/default/files/media/file/SSME_FAQs_revised_6July2012.pdf

RTI International. (2012). *EGRA EGMA: Tools to measure and improve student learning*. https://www.globalreadingnetwork.net/sites/default/files/media/file/EGRA_EGMA_-_World_Bank_presentation_-_Wed_5th_Dec_2012_FINAL.pdf

RTI International. (2014a). *Early grade mathematics assessment (EGMA) Toolkit*. USAID. https://ierc-publicfiles.s3.amazonaws.com/public/resources/EGMA%20Toolkit_March2014.pdf

RTI International. (2014b). *The Primary Math and Reading (PRIMR) initiative: Endline impact evaluation*. USAID. https://pdf.usaid.gov/pdf_docs/PA00K27S.pdf

RTI International. (2016). *Early grade reading assessment (EGRA) Toolkit* (2nd Edition). USAID). https://ierc-publicfiles.s3.amazonaws.com/public/resources/EGRA%20Toolkit%20V2%202016.pdf

RTI International, & International Rescue Committee. (2011). *Guidance notes for planning and implementing EGRA*. USAID.

Save the Children. (2017). *IDELA: Classroom Environment Tool*. https://resource-centre-uploads.s3.amazonaws.com/uploads/IDELA_Classroom-Environment-Tool_2017_Final.pdf

Save the Children. (2019). *International Development and Early Learning Assessment (IDELA): Adaptation & Administration Guide*. https://idela-network.org/wp-content/uploads/2019/04/Adaptation-Administration-Guide-and-IDELA-Tool_2019.pdf

Save the Children. (2022). *International Development and Early Learning Assessment*. IDELA. https://idela-network.org/

Save the Children. (n.d.). *Steps for IDELA Data Cleaning and Basic Analysis*. https://idela-network.org/wp-content/uploads/Members/Data_Collection/Steps-for-IDELA-Data-Cleaning-and-Analysis-in-Excel.docx

Seiden, J., Waldman, M., McCoy, D.C., & Fink, G. (2021). *CREDI data management and scoring manual*. https://inee.org/sites/default/files/resources/credi_scoring_manual.pdf

Sincovich, A., Gregory, T., Zanon, C., Santos, D. D., Lynch, J., & Brinkman, S. A. (2019). Measuring early childhood development in multiple contexts: The internal factor structure and reliability of the early Human Capability Index in seven low and middle income countries. *BMC Pediatrics*, *19*(471), 1–14.

https://doi.org/10.1186/s12887-019-1852-5

Squires, J., Bricker, author. ), Diane, Twombly, author. ), Elizabet, Potter, contributor. ), LaWand, & Paul H. Brookes Publishing Co., P., (publishe. (2009). *Ages & stages questionnaires: A parent-completed child monitoring system*. Baltimore, Maryland : Paul H. Brookes Publishing Company. https://trove.nla.gov.au/version/165719364

Squires, J., Twombly, E., Bricker, D., & Potter, L. (2009). ASQ-3 Technical Report. In *ASQ-3™ User's Guide*. Brookes Publishing.

Telethon Kids Institute. (2022a). *eHCI in action*. Early Human Capability Index (eHCI). https://ehci.telethonkids.org.au/about/action/

Telethon Kids Institute. (2022b). *FAQ*. Early Human Capability Index (eHCI). https://ehci.telethonkids.org.au/faq/

Telethon Kids Institute. (2022c). *Resources*. Early Human Capability Index (eHCI). https://ehci.telethonkids.org.au/resources/

Telethon Kids Institute. (2022d). *Scoring guide*. Early Human Capability Index (eHCI). https://ehci.telethonkids.org.au/scoring/

Telethon Kids Institute. (2022e). *The tool*. Early Human Capability Index (eHCI). https://ehci.telethonkids.org.au/tool/

Turner, R. (2014). *Described proficiency scales and learning metrics* (4; Assessment GEMS). Australian Council for Educational Research. https://www.acer.org/files/04_AssessGEMs_Described-Proficiency-Scales-and-Learning-Metrics.pdf

Turner, R., Adams, R., Schwantner, U., Cloney, D., Scoular, C., Anderson, P., Daraganov, A., Jackson, J., Knowles, S., O'Connor, G., Munro-Smith, P., Zoumboulis, S., & Rogers, P. (2014). *Development of reporting scales for reading and mathematics: A report describing the process for building the UIS reporting scales*. The Australian Council for Educational Research. https://research.acer.edu.au/monitoring_learning/33/

UNESCO. (n.d.). *Assessment for improved learning outcomes*. Retrieved January 23, 2024, from https://www.unesco.org/en/learning-assessments

UNESCO. (2021). *Inclusion in early childhood care and education: Brief on inclusion in education* (0000379502). https://unesdoc.unesco.org/ark:/48223/pf0000379502

UNESCO Institute for Statistics. (2023). *Policy linking for measuring global learning outcomes: Toolkit – Linking assessments to the global proficiency framework*. UNESCO. https://www.edu-links.org/sites/default/files/media/file/policy_linking_for_measuring_global_learning_outcomes_toolkit.pdf

UNESCO Institute for Statistics & Australian Council for Educational Research. (2024). *Pairwise comparison method toolkit: A toolkit for countries to measure global learning outcomes*. Australian Council for Educational Research.

https://research.acer.edu.au/gem/20/

UNESCO Institute of Statistics. (2022). *SDG indicator metadata*.
https://unstats.un.org/sdgs/metadata/files/Metadata-04-02-01.pdf

UNESCO, UNICEF, Brookings Institution, & World Bank. (2017). *Measuring Early
Learning Quality and Outcomes (MELQO): Overview*.
https://www.brookings.edu/wp-content/uploads/2017/06/melqo-measuring-
early-learning-quality-outcomes.pdf

UNICEF. (2019). *The formative years: UNICEF's work on measuring ECD*.
https://data.unicef.org/resources/the-formative-years-unicefs-work-on-
measuring-ecd/

UNICEF. (2020). *Guidance on the Importance of Quality in Early Childhood Learning and
Education in Latin America and the Caribbean*.

UNICEF. (2021). *Early Childhood Development Index 2030: A new tool to measure SDG
Indicator 4.2.1*. UNICEF Data. https://data.unicef.org/wp-
content/uploads/2023/09/ECDI2030_Technical_Manual_Sept_2023.pdf

UNICEF. (2022). *SDG indicator metadata*.
https://unstats.un.org/sdgs/metadata/files/Metadata-04-02-01.pdf

UNICEF. (n.d.). *Early childhood education*. https://www.unicef.org/education/early-
childhood-education

United Nations Children's Fund. (2020a). *ECDI2030: Frequently asked questions*. UNICEF.
https://data.unicef.org/wp-content/uploads/2020/07/ECDI2030-Frequently-
Asked-Questions.pdf

United Nations Children's Fund. (2020b). *ECDI2030: Instructions for interviewers*.
UNICEF. https://data.unicef.org/wp-content/uploads/2021/03/ECD-Interview-
Guidelines-EN-3_12.pdf

United Nations Children's Fund. (2021). *A brief introduction of the ECDI2030 — A new tool
to measure SDG 4.2.1*. UNICEF.
https://www.unicef.cn/media/21141/file/ECDI2030%20%E2%80%93%20Early%20
Childhood%20Development%20Index%202030.pdf

United Nations Children's Fund. (n.d.). *ECDI2030: Overview of the ECDI2030*.
https://www.unescwa.org/sites/default/files/event/materials/PPT%20SDG%204.2.
1_E.pdf

United Nations Department of Economic and Social Affairs. (n.d.). *Ensure inclusive and
equitable quality education and promote lifelong learning opportunities for all.* Goals.
Retrieved January 23, 2024, from
https://sdgs.un.org/goals/goal4#targets_and_indicators

Vameghi, R., Sajedi, F., Kraskian Mojembari, A., Habiollahi, A., Lorenzhad, H. R., &
Delavar, B. (2013). Cross-cultural adaptation, validation and standardization of
Ages and Stages Questionnaire (ASQ) in Iranian children. *Iranian Journal of Public*

*Health*, 42(5), 522–528.

Waldfogel, J. (2006). *Early Childhood Policy: A Comparative Perspective.*

Waldman, M., McCoy, D. C., Seiden, J., Cuartas, J., Fink, G., & CREDI Field Team. (2021). Validation of motor, cognitive, language, and socio-emotional subscales using the Caregiver Reported Early Development Instruments: An application of multidimensional item factor analysis. *International Journal of Behavioral Development*, 23(1), 1–10.

Waters, C. (2019). *Learn more about learning progressions*. ACER Discover. https://www.acer.org/au/discover/article/learn-more-about-learning-progressions

World Bank. (2018). *World Development Report 2018: Learning to Realize Education's Promise*. World Bank.
publisher: World Bank Group

World Health Organization. (2022). *Global Scales for Early Development*. https://www.who.int/teams/mental-health-and-substance-use/data-research/global-scale-for-early-development

World Health Organization. (2023a). *Global Scales for Early Development v1.0: Long Form (directly administered)*. https://apps.who.int/iris/bitstream/handle/10665/366270/WHO-MSD-GSEDpackage-v1.0-2023.5-eng.pdf

World Health Organization. (2023b). *Global Scales for Early Development v1.0: Long Form (directly administered): Item guide*. https://apps.who.int/iris/bitstream/handle/10665/366269/WHO-MSD-GSEDpackage-v1.0-2023.6-eng.pdf

World Health Organization. (2023c). *Global Scales for Early Development v1.0: Long Form (directly administered): User manual*. https://apps.who.int/iris/bitstream/handle/10665/366271/WHO-MSD-GSEDpackage-v1.0-2023.7-eng.pdf

World Health Organization. (2023d). *Global Scales for Early Development v1.0: Scoring guide*. https://apps.who.int/iris/bitstream/handle/10665/366272/WHO-MSD-GSEDpackage-v1.0-2023.8-eng.pdf

World Health Organization. (2023e). *Global Scales for Early Development v1.0: Short form (caregiver-reported)*. https://apps.who.int/iris/bitstream/handle/10665/366265/WHO-MSD-GSEDpackage-v1.0-2023.2-eng.pdf

World Health Organization. (2023f). *Global Scales for Early Development v1.0: Short form (caregiver-reported): Item guide*. https://apps.who.int/iris/bitstream/handle/10665/366266/WHO-MSD-GSEDpackage-v1.0-2023.3-eng.pdf

World Health Organization. (2023g). *Global Scales for Early Development v1.0: Short form*

*(caregiver-reported): User manual*.
https://apps.who.int/iris/bitstream/handle/10665/366267/WHO-MSD-
GSEDpackage-v1.0-2023.4-eng.pdf

World Health Organization. (2023h). *Global Scales for Early Development v1.0: Technical
report*. https://apps.who.int/iris/bitstream/handle/10665/366268/WHO-MSD-
GSEDpackage-v1.0-2023.1-eng.pdf

Yoshikawa, H., & Kabay, S. (2015). *The evidence base on early childhood care and education in
global contexts: Paper commissioned for the EFA Global Monitoring Report 2015,
Education for All 2000-2015: Achievements and challenges.* UNESCO.
https://unesdoc.unesco.org/ark:/48223/pf0000232456

Zirakashvili, M., Gabunia, M., Tatishvili, N., Ediberidze, T., Lomidze, G., Chachava, T.,
& Hix-Small, H. (2018). Cultural Adaptation and Psychometric Validation of the
Ages and Stages Questionnaires for use in Georgia. *Journal of Child and Family
Studies*, *27*, 739–749.

# Appendix A: Search strategy and key terms

**Table 4: Search strategy and key terms**

| POPULATION | Intervention (assessment tool) | | Comparison | OUTCOMES OF INTEREST |
|---|---|---|---|---|
| | **Domains** | **Sub-components** | | |
| **Early childhood 2-6 years**<br>Toddler<br>Pre-schooler<br>Child<br>_____<br>**Settings**<br>Early learning centre<br>Preschool<br>Kindergarten<br>Reception<br>Elementary<br><br>Home<br><br>Community<br>_____<br>**Keywords that might exclude a tool**<br>Professional<br>Clinical<br>Psychologist<br>Psychiatric | Global development<br>Health<br>Learning<br>Cognitive<br>Psycho-social wellbeing | Initiative and Curiosity, Attentiveness, Engagement and Persistence, Creativity, Creating and materials, Being imaginative and expressive, Processing and Using Information, Exploration of the Arts, Using the Arts and Self Expression, Listening, Attention, and Understanding, Communicating and Speaking, Comprehension, Emergent Reading, Writing, Number Knowledge, Measurement, Patterns, Geometry and Spatial Thinking, Gross Motor, Fine Motor, Self and Emotional Awareness, Self-Management, Self-Regulation, Social Understanding and Relationships, Community, People and Relationships, Change Over Time, Past and Present, Environment, Natural world, Economics, Technology, Discover, Act, Integrate<br><br>_____<br>**Keywords that might exclude a tool**<br>Metacognition<br>Diagnostic<br>Screening[8] | Validated<br>Trialled<br>Adapted<br>Multiple languages<br>Translated<br>Population assessment | **Primary outcome**<br>Child outcomes<br>Cognitive outcomes<br>Learning outcomes<br>• Pre-literacy<br>• Pre-numeracy<br>• Cognitive<br>• School readiness<br>Social outcomes<br>Social and emotional learning (SEL)<br>_____<br>**Secondary outcomes** (if measured alongside primary outcomes)<br>• Emotional regulation<br>• Physical health and development<br>_____<br>**Keywords that might exclude a tool**<br>Commercial products<br>Pay per assessment |

---

[8] Diagnostic tools and screeners were excluded, unless they are used widely as population measures

# Appendix B: List of assessments screened

**Table 5: Criteria for including assessments in the review**

| Assessment tool | Age range | Domains | Purpose | Originality | Availability | Outcomes | Implementation | | Languages | Validity and reliability | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Is this a population assessment? | Is this a newly developed assessment or does it combine existing items from multiple sources? | Is this available for use? | Are learning domains assessed? | Has this been trialled across multiple countries/languages? | Are adaptation manuals available? | Is this available in English? | Has this been psychometrically validated? | Have details been published? | Decision |
| **The People's Actions for Learning (PAL) Network Assessments** | | | | | | | | | | | | |
| · *Annual Status of Education Report (ASER) – India, Nepal, and Pakistan* | 5-16 years | Literacy and numeracy | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Include |
| · *BEEKUNKO –Mali* | 6-14 years | Literacy and numeracy | Yes | No | Yes | Yes | No | Yes | Yes | Yes | Yes | Include under ASER |
| · *Uwezo –Kenya, Uganda and Tanzania* | 6-16 years | Literacy and numeracy | Yes | No | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Include under ASER |
| · *Jangadoo –Senegal* | 5-16 years | Literacy and numeracy | Yes | No | Yes | Yes | No | Yes | Yes | Yes | Yes | Include under ASER |
| · *Medición Independiente de Aprendizajes- Mexico* | 5-16 years | Literacy and numeracy | Yes | No | Yes | Yes | No | Yes | Yes | Yes | Yes | Include under ASER |
| · *LEARNigeria- Nigeria* | 5-16 years | Literacy and numeracy | Yes | No | Yes | Yes | No | Yes | Yes | Yes | Yes | Include under ASER |

| Assessment tool | Age range | Domains | Purpose | Originality | Availability | Outcomes | Implementation | | Languages | Validity and reliability | | Decision |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Is this a population assessment? | Is this a newly developed assessment or does it combine existing items from multiple sources? | Is this available for use? | Are learning domains assessed? | Has this been trialled across multiple countries/languages? | Are adaptation manuals available? | Is this available in English? | Has this been psychometrically validated? | Have details been published? | |
| · **IID/BRAC Survey-Bangladesh** | 5-16 years | Literacy and numeracy | Yes | No | Yes | Yes | No | Yes | Yes | Yes | Yes | Include under ASER |
| · **TPC Mozambique-Mozambique** | 5-16 years | Literacy and numeracy | Yes | No | Yes | Yes | No | Yes | Yes | Yes | Yes | Include under ASER |
| · **VIdA Nicaragua-Nicaragua** | 5-16 years | Literacy and numeracy | Yes | No | Yes | Yes | No | Yes | Yes | Yes | Yes | Include under ASER |
| · **Early Language & Literacy and Numeracy Assessment (ELANA)** | 4-10 years | Literacy, numeracy | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Include |
| · **International Common Assessment of Numeracy (ICAN)** | 5-16 years | Numeracy | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Include |
| **Caregiver Reported Early Development Instruments (CREDI)** | 0-3 years | Motor, Language, Cognition, Social-Emotional, and Mental Health | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Include |

| Assessment tool | Age range | Domains | Purpose | Originality | Availability | Outcomes | Implementation | | Languages | Validity and reliability | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Is this a population assessment? | Is this a newly developed assessment or does it combine existing items from multiple sources? | Is this available for use? | Are learning domains assessed? | Has this been trialled across multiple countries/languages? | Are adaptation manuals available? | Is this available in English? | Has this been psychometrically validated? | Have details been published? | Decision |
| East Asia-Pacific Early Child Development Scale (EAP-ECDS) | 3-5 years | Cognitive, motor, language, SEL, Pre-academic skills, approaches to learning | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Include |
| Early Child Development Index (ECDI) 2030 | 3-5 years | Language/cognitive), physical, social-emotional, and approaches to learning | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Include |
| Early Human Capability Index (eHCI) | 3-5 years | Cognitive, language, SEL, pre-academic, approaches to learning | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Include |
| Early Development Instrument (EDI) | 4– 6 years | Cognitive, Language, Motor, SEL | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Include |

| Assessment tool | Age range | Domains | Purpose<br>Is this a population assessment? | Originality<br>Is this a newly developed assessment or does it combine existing items from multiple sources? | Availability<br>Is this available for use? | Outcomes<br>Are learning domains assessed? | Implementation<br>Has this been trialled across multiple countries/languages? | Are adaptation manuals available? | Languages<br>Is this available in English? | Validity and reliability<br>Has this been psychometrically validated? | Have details been published? | Decision |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| · *Australian Early Development Census AEDC* | 4- 6 years | Cognitive, motor, language, SEL | Yes | No | Yes | Yes | Based on EDI | No | Yes | Yes | Yes | Include under EDI |
| · *Chinese Early Development Instrument (CEDI)* | 5 years | School readiness | Yes | No | Yes | Yes | No; but based on EDI (Canada) | Yes | Yes | Yes | Yes | Include under EDI |
| **Early Grade Reading Assessment (EGRA) / Early Grade Math Assessment (EMGA)** | 6-10 years | Pre-academic (language) and Pre- academic (math) | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Include |
| **Global Scale for Early Development (GSED)** | 0-4 years | Child development; Communication, fine and gross motor | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Include |
| **International Development Learning Assessment (IDELA)** | 3.5–6 years | Cognitive, Language, Motor, SEL, Executive functioning, pre-academic, approaches to learning | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Include |

| Assessment tool | Age range | Domains | Purpose<br>Is this a population assessment? | Originality<br>Is this a newly developed assessment or does it combine existing items from multiple sources? | Availability<br>Is this available for use? | Outcomes<br>Are learning domains assessed? | Implementation<br>Has this been trialled across multiple countries/languages? | Are adaptation manuals available? | Languages<br>Is this available in English? | Validity and reliability<br>Has this been psychometrically validated? | Have details been published? | Decision |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **International Early Learning and Child Well-being Study (IELS)** | 5 years | Emergent literacy, emergent numeracy, some self-regulation skills and some social-emotional skills | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Include |
| **Measuring Early Learning Quality Outcomes (MELQO) Measuring Child Development and Learning (MODEL)** | 4-6 years | SEL, Executive function, Pre-academic, and approaches to learning | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Include |
| **Early Childhood Development Assessment Scales (ECDAS) - direct assessment (DA) and caregiver survey (CS)** | 3- 5 years | Learning, psychosocial well-being and health | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Include |
| **Infant and Young Child Development (IYCD) Indicators** | 0-3 years | Motor, language and cognitive, and social-emotional | Yes | Yes | Yes | Yes | Yes | Unclear | Yes | Yes | Yes | Include |

| Assessment tool | Age range | Domains | Purpose | Originality | Availability | Outcomes | Implementation | | Languages | Validity and reliability | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Is this a population assessment? | Is this a newly developed assessment or does it combine existing items from multiple sources? | Is this available for use? | Are learning domains assessed? | Has this been trialled across multiple countries/languages? | Are adaptation manuals available? | Is this available in English? | Has this been psychometrically validated? | Have details been published? | Decision |
| **Ages & Stages Questionnaire (ASQ)** | 0-5 years | Communication, gross motor, fine motor, problem solving, and personal-social | Yes, also used as a diagnostic screening tool. | | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Include |
| **School Readiness Indicators Initiative-Getting Ready** | 0-8 years | Physical Well-Being and Motor Development. Social and Emotional Development. Approaches to Learning. Language Development. Cognition and General Knowledge. | Yes | | Yes | Yes | No | No | Yes | Yes | Yes | Exclude |
| **Regional Project on Child Development Indicators (Programa Regional de Indicadores de Desarrollo Infantil; PRIDI)** | 2-5 years | Cognitive, motor, language, SEL | Yes | | Yes | Yes | Yes | Yes | No | Yes | Yes | Exclude |

| Assessment tool | Age range | Domains | Purpose | Originality | Availability | Outcomes | Implementation | | Languages | Validity and reliability | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Is this a population assessment? | Is this a newly developed assessment or does it combine existing items from multiple sources? | Is this available for use? | Are learning domains assessed? | Has this been trialled across multiple countries/languages? | Are adaptation manuals available? | Is this available in English? | Has this been psychometrically validated? | Have details been published? | Decision |
| Berkeley Puppet Interview Academic (BPI- A) | 4–8 years | Wellbeing | No | | No | Yes | Yes | Yes | Yes | Yes | Yes | Exclude |
| British Abilities Scales (BAS) | 4-17.9 years | Cognitive, language, pre-academic/ academic | No | | No | Yes | Yes | Yes | Yes | Yes | Yes | Exclude |
| Children's Behavior Questionnaire (CBQ) | 3-7 years | Temperament | No | | No | No | No | Yes | Yes | Yes | Yes | Exclude |
| Child and Adolescent Functional Assessment Scale (CAFAS) | 5-19 years | Problem behaviors, strengths, and goals | No | | No | No | No | No | Yes | Yes | Yes | Exclude |
| Child Behavior Checklist and related tools from the Achenbach System of Empirically Based Assessment (CBCL/ASEBA) | 6-18 years | Behaviors and mood | No | | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Exclude |

| Assessment tool | Age range | Domains | Purpose | Originality | Availability | Outcomes | Implementation | | Languages | Validity and reliability | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Is this a population assessment? | Is this a newly developed assessment or does it combine existing items from multiple sources? | Is this available for use? | Are learning domains assessed? | Has this been trialled across multiple countries/languages? | Are adaptation manuals available? | Is this available in English? | Has this been psychometrically validated? | Have details been published? | Decision |
| Child and Youth Wellbeing Index | | Wellbeing | | | | | | | | | | Exclude, no learning outcomes so no further analysis. |
| Comprehensive Developmental Inventory for Infants and Toddlers (CDIIT) | 0.3-6 years | Cognitive, motor, language, social | No | | No | Yes | No | | | Yes | Yes | Exclude |
| Dynamic Indicators of Basic Early Literacy Skills (DIBELS) | 5-9 years | Initial Sound Fluency (ISF), Phoneme Segmentation Fluency (PSF), Letter Naming Fluency (LNF), and Nonsense Word Fluency (NWF). | No | | No | Yes | No | | Yes | Yes | Yes | Exclude |
| Early childhood longitudinal study, birth Cohort (ECLS-B) | 0-6 years | Early reading and math, fine motor | No | | Yes | Yes | No | No | Yes | Yes | Yes | Exclude |
| Early Childhood Environment Rating Scale (ECERS) | 2.5–5 years | Child interactions. language ability, | Yes | | Yes | No | Yes | Yes | Yes | Yes | Yes | Exclude |

| Assessment tool | Age range | Domains | Purpose | Originality | Availability | Outcomes | Implementation | | Languages | Validity and reliability | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Is this a population assessment? | Is this a newly developed assessment or does it combine existing items from multiple sources? | Is this available for use? | Are learning domains assessed? | Has this been trialled across multiple countries/languages? | Are adaptation manuals available? | Is this available in English? | Has this been psychometrically validated? | Have details been published? | Decision |
| | | math readiness, and social skills. | | | | | | | | | | |
| **Early Learning and Developmental Standards (ELDS)-UNICEF** | 4-5 years | Socio-emotional, logic and reasoning, approaches to learning and language and literacy development | No | | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Exclude |
| **Health Behaviour in School-aged Children (HBSC)** | 11, 13 and 15 years | | | | | | | | | | | Exclude, out of age range so no further analysis. |

| Assessment tool | Age range | Domains | Purpose | Originality | Availability | Outcomes | Implementation | | Languages | Validity and reliability | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Is this a population assessment? | Is this a newly developed assessment or does it combine existing items from multiple sources? | Is this available for use? | Are learning domains assessed? | Has this been trialled across multiple countries/languages? | Are adaptation manuals available? | Is this available in English? | Has this been psychometrically validated? | Have details been published? | Decision |
| **Hong Kong Early Child Development Scale (HKECDS)** | 3-6 years | Personal, Social and Self-Care, Language Development, Pre-academic Learning, Cognitive Development, Gross Motor, Fine Motor, Physical Fitness, Health and Safety, and Self and Society | Yes | | Yes | Yes | No | Yes | Yes | Yes | Yes | Exclude |

| Assessment tool | Age range | Domains | Purpose: Is this a population assessment? | Originality: Is this a newly developed assessment or does it combine existing items from multiple sources? | Availability: Is this available for use? | Outcomes: Are learning domains assessed? | Implementation: Has this been trialled across multiple countries/languages? | Implementation: Are adaptation manuals available? | Languages: Is this available in English? | Validity and reliability: Has this been psychometrically validated? | Validity and reliability: Have details been published? | Decision |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Individualized Classroom Assessment Scoring System (inCLASS)** | 3-5 years | Child behavior and the social context: teacher and peer interactions, task orientation, and conflict interactions | No | | No | Yes | Yes | Yes | Yes | Yes | Yes | Exclude |
| **Mayr and Ulich's Positive development and resilience in kindergarten; (PERIK)** | 4-8 years | Wellbeing, resilience | | | | No | | | | | | Exclude, out of age range and no learning domain so no further analysis. |
| **Malawi Developmental Assessment Tool (MDAT)** | 0-6 years | Child development, health and neurodisabilities | No | | | No | | | | | | Exclude |
| **Parents' Evaluation of Developmental Status (PEDS)** | 0–8 years | Development, behavior, social-emotional /mental-health, and autism | Yes | | Yes | No | Yes | Yes | Yes | Yes | Yes | Exclude |

| Assessment tool | Age range | Domains | Purpose | Originality | Availability | Outcomes | Implementation | | Languages | Validity and reliability | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Is this a population assessment? | Is this a newly developed assessment or does it combine existing items from multiple sources? | Is this available for use? | Are learning domains assessed? | Has this been trialled across multiple countries/languages? | Are adaptation manuals available? | Is this available in English? | Has this been psychometrically validated? | Have details been published? | Decision |
| **Preschool Developmental Assessment Scale (PDAS)** | 3-6 years | | | | | | No | | | | | Exclude, used in Hong Kong only so no further analysis |
| **Strengths and Difficulties Questionnaire (SDQ)** | 3-16 years | Wellbeing, SEL | Yes | | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Exclude |
| **Wechsler Preschool and Primary Scales of Intelligence (WPPSI)** | 2.5-7.6 years | Cognitive, Language, Executive functioning | No | | No | No | Yes | Yes | | Yes | Yes | Exclude |

# Appendix C: Frameworks used for evaluating assessments

- Assessment of children as confident and involved learners in early childhood education and care: Literature review (Cloney et al., 2019)

- Principles of good practice in learning assessment (GEM Centre & UIS, 2017)

- Review and advice on assessment across the South Australian Public education system: Measuring what matters  (Ainley et al., 2021)

- Improving young children's learning in economically developing countries: What works, why, and where?: Scoping Review (Jackson et al., 2019)

- A toolkit for measuring early childhood development in low and middle-income countries (Fernald et al., 2017, p. 9).

- Measuring early child development across low and middle-income countries: A systematic review (Munoz-Chereau, B et al., 2021, p. 443).

- Review of measures of social and emotional development by Halle and Darling-Churchill (2016).

# Appendix D:  Summaries of included assessments

## Ages & Stages Questionnaires, Third Edition (ASQ-3)

The Ages and Stages Questionnaires, Third Edition (ASQ-3) (Squires, Bricker, et al., 2009) is a developmental screening instrument designed to assess babies and children and determine when further evaluation is required. It was developed by researchers at the University of Oregon and this third edition of the tool has been in use since 2009.

### Purpose(s)

The ASQ-3 is a screening tool with the purpose of identifying babies and children who would benefit from in-depth evaluation for developmental delays. However, the tool has also been used to monitor populations in England's Healthy Child Programme (HCP) since 2009 (Kendall et al., 2019).

### Age range

The ASQ-3 screens babies and children between the ages of 2 months and 60 months of age.

The 21 questionnaires are designed for the following ages:

2, 4, 6, 8, 9, 10, 12, 14, 16, 18, 20, 22, 24, 27, 30, 33, 36, 42, 48, 54, and 60 months of age.

### Domain(s)

The ASQ-3 reports against multiple domains, including learning.

**Table 6: Summary of the ASQ-3 domains**

| Domain | Description | Sample question (from 48-month questionnaire) |
|---|---|---|
| Communication | The child's language skills, both what the child understands and what he or she can say | Does your child tell you at least 2 things about common objects? For example, if you say to your child, "Tell me about your ball," does she say something like, "It's round. I throw it. It's big"? |
| Gross motor | How the child uses their arms and legs and other large muscles for sitting, crawling, walking, running, and other activities | Does your child climb the rungs of a ladder of a playground slide and slide down without help? |
| Fine Motor | The child's hand and finger movement and coordination | Does your child unbutton one or more buttons? (Your child may use his own clothing or a doll's clothing.) |
| Problem solving | How the child plays with toys and solves problems | If you place five objects in front of your child, can he count them by saying, "one, two, three, four, five," in order? (Ask this question without providing help by pointing, gesturing, or naming.) |
| Personal-Social | The child's self-help skills and interactions with others. | Does your child serve herself, taking food from one container to another using utensils? For example, does your child use a large spoon to scoop applesauce from a jar into a bowl? |

Source: (Ages & Stages Questionnaires®, n.d.a)

## Learning progression

There is no learning progression associated with the ASQ-3.

## Assessment framework

The ASQ-3 provides extensive details of assessment development and how the items provide construct coverage. The ASQ-3 technical report (Squires, Twombly, et al., 2009) provides detail about the development of the ASQ-3.

## Administration

The ASQ-3 uses indirect assessment of the child where the parents/caregivers of the child complete the questionnaire.

## Data collection

The ASQ-3 provides the option to use either paper/pen or digital (online) formats to collect data.

## Data scoring

The ASQ-3 provides basic scoring advice using raw scores. Professionals, paraprofessionals, or clerical staff complete the scoring of the parent questionnaire, calculating the total score for each domain. The scorer then compares the calculated scores with the cut-offs on the scoring sheet. The cut-off scores give an indication of the child's development status. The higher the score, the more positive the outcome. Figure 1 illustrates the summary score sheet.

**1. SCORE AND TRANSFER TOTALS TO CHART BELOW:** See *ASQ-3 User's Guide* for details, including how to adjust scores if item responses are missing. Score each item (YES = 10, SOMETIMES = 5, NOT YET = 0). Add item scores, and record each area total. In the chart below, transfer the total scores, and fill in the circles corresponding with the total scores.

| Area | Cutoff | Total Score | 0 | 5 | 10 | 15 | 20 | 25 | 30 | 35 | 40 | 45 | 50 | 55 | 60 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Communication | 34.60 | | ● | ● | ● | ● | ● | ● | ● | ○ | ○ | ○ | ○ | ○ | ○ |
| Gross Motor | 38.41 | | ● | ● | ● | ● | ● | ● | ● | ● | ○ | ○ | ○ | ○ | ○ |
| Fine Motor | 29.62 | | ● | ● | ● | ● | ● | ● | ○ | ○ | ○ | ○ | ○ | ○ | ○ |
| Problem Solving | 34.98 | | ● | ● | ● | ● | ● | ● | ● | ○ | ○ | ○ | ○ | ○ | ○ |
| Personal-Social | 33.16 | | ● | ● | ● | ● | ● | ● | ● | ○ | ○ | ○ | ○ | ○ | ○ |

**3. ASQ SCORE INTERPRETATION AND RECOMMENDATION FOR FOLLOW-UP:** You must consider total area scores, overall responses, and other considerations, such as opportunities to practice skills, to determine appropriate follow-up.

If the baby's total score is in the ▢ area, it is above the cutoff, and the baby's development appears to be on schedule.
If the baby's total score is in the ▨ area, it is close to the cutoff. Provide learning activities and monitor.
If the baby's total score is in the ■ area, it is below the cutoff. Further assessment with a professional may be needed.

**Figure 1 - Summary sheet generated by the ASQ-3 (Squires et al., 2009, p. 157)**

## Item type(s)

The ASQ-3 assessment contains only one type of item. The assessment items are closed questions, requiring the parent to answer 'Yes', 'Sometimes' or 'Not Yet'. For example,

> Does your child stack a small block or toy on top of another one? (18-month questionnaire, fine motor area) (Ages & Stages Questionnaires®, 2018)

Once the assessment is completed, some 'Overall' questions are included which allow parents to explain their answers. For example,

> Do you think your child talks like other toddlers her age? If no, explain. (Ages & Stages Questionnaires®, 2009)

## Background/contextual information

The ASQ-3 collects background/contextual information about the child. Questions consist of the child's name, date of birth, gender and, if the child was born more than 3 weeks prematurely, the number of weeks premature (Ages & Stages Questionnaires®, 2009). Some basic information is also collected about the parent/caregiver completing the questionnaire, like name, address and their relationship to the child (Ages & Stages Questionnaires®, 2009).

## Length/time

The ASQ-3 contains approximately 30 items per questionnaire and is typically completed by parents in 10 to 15 minutes. It then takes 2 to 3 minutes for professionals to score.

## Available languages

The ASQ-3 is available in multiple languages: Arabic, Chinese, English, French, Spanish and Vietnamese. Online only versions are also available in Hmong and Somali. Guidelines are provided for the translation/adaptation of the assessment.

## Availability/cost

The ASQ-3 incurs a cost to access the assessment materials.

The ASQ-3 starter kit costs $US295.00. It contains:

1) 21 paper masters of the questionnaires and scoring sheets
2) CD-ROM with printable PDF questionnaires
3) the ASQ-3 User's Guide
4) a laminated ASQ-3 Quick Start Guide that keeps scoring and administration basics close at hand.

The complete package costs $US1269.75. It is in English only and includes additional resources such as a scoring and referral DVD, home visit DVD, learning activities, materials kit, and family and professional annual subscription to online platforms.

## Training materials/requirements/capacity development

The ASQ-3 provides training materials to users of the tool. The ASQ-3 is designed to be easy to use by providers with varying levels of education and expertise. A DVD delivers training that is short, inexpensive and can be viewed as often as is required. Additionally, virtual training sessions are available, or programs may request to engage with on-site training (Ages & Stages Questionnaires®, n.d.b).

## Validation for use across different populations

The ASQ-3 has only been validated in different populations separately.

The ASQ-3 has been validated using a research sample that included 15,138 children reflecting the U.S. population in terms of race, ethnicity, and socio-economic groups (Squires, Twombly, et al., 2009). The technical report provides extensive detail about the psychometric studies conducted to establish the validity and reliability of the screening tool.

Several studies have been undertaken to evaluate the ASQ in cultures other than the US. For example, Vameghi et al (2013) evaluated a Persian version of the ASQ to determine that the validity and reliability were adequate in screening for developmental disorders in Iran. Similarly, Richter and Janson (2007) concluded their study of Norwegian children by confirming the literature supporting the construct validity of the ASQ. However, other studies like Zirakashvili (2018) found that there were 'significant differences in cut off points between the Georgian ASQ-3 and US reference population in most domains across age groups'.

# Caregiver Reported Early Development Instruments (CREDI)

The CREDI assessment is designed to collect population-level data on early childhood development for children from birth to age 3. The CREDI was developed by researchers and is based in Harvard University. It was developed in 2013. CREDI was designed to be "culturally neutral" for use by low-and-middle-income countries and has already been used across many countries in Asia, Africa, and the Americas (McCoy, Sudfeld, et al., 2017a).

## Purpose(s)

The main purpose pf the CREDI assessment is to monitor populations. The Short Form version of the tool is designed to make cross-national comparisons of developmental status and progress for the youngest children (aged 0 to 3 years), particularly those within vulnerable groups. The Long Form version of the tool is designed for program evaluation or exploratory research.

## Age range

This assessment is suitable for assessing children aged 0 to 3 years.

## Domain(s)

The CREDI is able to report against multiple domains including multiple learning domains.

**Table 7: Summary of the CREDI domains**

| Domain | Description | Sample question |
|---|---|---|
| Motor | Children's ability to use fine and gross movements to explore and engage with their environments | "Can the child pick up a small object (for example, a small toy or stone) with just his/her thumb and a finger?" |
| Language | Children's ability to communicate their needs and desires, and understand what others are saying to them | "Can the child say five or more separate words (for example, names like "Mama" or objects like "ball")?" |
| Cognition | Children's ability to pay attention, remember information, perceive, and discriminate between objects and people in their environment, solve problems, and acquire basic knowledge | "Does the child usually put objects or toys back where they belong after using them?" |
| Social-Emotional | Children's ability to control their behaviours and emotions, understand their feelings, and get along well with others | "Does the child greet neighbors or other people he/she knows without being told (for example, by saying hello or gesturing hello)?" |
| Mental Health | The absence of behaviours related to aggression, anxiety, and distress | "Does the child cling excessively to his/her caregiver, even in a safe setting?" |

Both CREDI forms (Short Form and Long Form) cover items from all 5 of these domains, however, the short form reports development on an aggregate/holistic scale, whereas the long form gives a measure of each domain. Therefore, the short form assesses children using items from each of the 5 domains but reports only on overall/global development.

## Learning progression

There is no learning progression associated with the CREDI.

## Assessment framework

The CREDI provides extensive details of assessment development, including theoretical underpinnings and how the items provide construct coverage. Associated articles (McCoy, Sudfeld, et al., 2017a; McCoy et al., 2018) describe the theoretical foundations of the domains being covered and how the items developed for the assessment cover the constructs.

## Administration

The CREDI uses indirect assessment of the child where enumerators complete the questionnaires based on an interview with the parents/caregivers of the child.

## Data collection

The CREDI provides the option to use either paper/pen or digital formats to collect data. However, the paper-based format is the most commonly used. The digital format is a

computer-based online survey that allows the parent/caregiver to respond directly to the questionnaire without the need for an enumerator.

## Data scoring

The CREDI provides detailed information or advice about data processing (including scoring) and converting raw scores onto an ordinal/linear scale. A data management and scoring manual (Seiden, J. et al., 2021) is available for downloading from the CREDI website. The manual supports users to calculate composite CREDI scores to use in their data analysis, as well as to accurately interpret these scores. Users can generate scores either by uploading data to the CREDI Scoring App or by downloading the CREDI package and scoring data locally using the statistical software, R (Waldman, M. et al., 2021). Both these options use the responses on each CREDI item to generate overall, domain, and norm-referenced scores for use in later analysis. The manual (Seiden, J. et al., 2021) only covers information for using the CREDI Scoring App.

## Item type(s)

The CREDI assessment contains only one type of item. All the items require a yes/no (dichotomous) response.

## Background/contextual information

The CREDI collects background/contextual information about the child and the caregiver. The questions about the child consist of their age. The questions about the caregiver include whether they are literate in the given language.

## Length/time

The Short Form of the CREDI contains 20 questions and is typically completed by the caregiver in 5 minutes. The Long Form of the CREDI contains up to 100 questions and is typically completed by the caregiver in 15 minutes.

## Available languages

The CREDI is available in multiple languages: Armenian, Cebuano, Chinese, Filipino, French, Hindi, Ilonggo, Japanese, Khmer, Korean, Mandarin, Nepali, Portuguese, Spanish, Swahili. Guidelines are provided for the translation/adaptation of the assessment through consultations with the CREDI team.

## Availability/cost

The CREDI is an open-source assessment. No access or use fees or royalties are involved. Open resource free download: https://sites.sph.harvard.edu/credi/

## Training materials/requirements/capacity development

The CREDI provides training materials to users of the tool. Minimal training is required for using this assessment, such as a half-a-day training to go over all the key materials. The materials that are needed to understand, administer, and score the forms can be found on the CREDI website (CREDI team, 2023).

Both the CREDI Short Form and the CREDI Long Form have been validated as scales that authors suggest should be used exactly as they are written as they are designed to be culturally and linguistically neutral (McCoy et al., 2018; Waldman, M. et al., 2021). However, no studies were found that provide evidence for measurement invariance across countries/languages.

Pilot Round 1 targeting children ages 18–35 months took place in Tanzania (Ifakara) and included a total of 70 items selected from an initial set of 92 items (McCoy, Sudfeld, et al., 2017b; McCoy, Zuilkowski, et al., 2017). Based on these results, and feedback from the local team and advisory panel, 2 main revisions were made to the CREDI prior to Pilot Round 2: (i) the age range was expanded to cover all children under age 3; and (ii) additional items that are more difficult for 18- to 35-month-olds were included to avoid ceiling effects (McCoy et al., 2018). A revised set of 117 items was tested in a sample of 4,472 0- to 35-month-old children living in Bangladesh, Brazil, Laos, Tanzania (Dar es Salaam), United States, and Zambia as part of Pilot Round 2 (McCoy et al., 2018). After completion of Pilot Round 2, additional item revisions were made, and further items were added to ensure appropriate coverage in all domains and age ranges (McCoy et al., 2018). As part of Pilot Round 3 a total of 147 items were then administered to a sample of 993 children in Lebanon, Jordan, and Pakistan (McCoy et al., 2018). After a final round of revisions/additions, a set of 149 items was administered to a sample of 7,807 children in Brazil, Cambodia, Chile, Colombia, Ghana, Guatemala, India, Nepal, Philippines, and the United States as part of Pilot Round 4 (McCoy et al., 2018). Some qualitative pilot testing was further conducted with in Ghana, Guatemala, Hong Kong, Laos, and Lebanon (McCoy et al., 2018).

# Early Childhood Development Assessment Scale (ECDAS)

The Early Childhood Development Assessment Scale – Direct Assessment (ECDAS-DA: Rao et al., 2020a) and the Early Childhood Development Assessment Scale – Caregiver Survey (ECDAS-CS: Rao et al., 2020b) are assessments designed to measure early childhood development in line with SDG 4.2.1 reporting (Rao et al., 2021), drawing from the same pool of items that were considered when the ECDI2030 was developed. Both assessments were designed for the purpose of assessing how different modes (that is, direct assessment against caregiver report) of measuring early childhood outcomes compare in terms of their psychometric properties. The assessments were developed by researchers at The University of Hong Kong and was piloted in Bangladesh, China, India and Myanmar (Rao et al., 2021).

## Purpose(s)

The purpose of ECDAS- DA and ECDAS-CS is to monitor populations. The tools have been piloted in 4 countries and results suggest that the ECDAS-DA can be used for both longitudinal studies and impact evaluations in low- and middle- income countries.

## Age range

The age range suggested for these assessments is 24 months to 59 months (that is, in line with SDG 4.2.1).

## Domain(s)

The ECDAS assessments contains items related to multiple learning domains, however domains are aggregated into one overall score for reporting.

**Table 8: Summary of the ECDAS domains**

| Domain | Sub-domains |
|---|---|
| Learning | literacy, expressive language, pre-writing, and numeracy |
| Psychosocial wellbeing | emotional skills, social skills, and externalizing |
| Executive functioning and health | gross motor and care |

## Learning progression

There is no learning progression associated with ECDAS-DA or ECDAS-CS.

## Assessment framework

The ECDAS-DA and ECDAS-CS provide limited details of assessment development, including theoretical underpinnings and how the items provide construct coverage.

## Administration

The ECDAS-DA uses direct assessment of the child where trained assessors individually administer the assessment. The ECDAS-CS uses indirect assessment where the questionnaire is either completed by the enumerator in an interview with the parents/caregivers, or the parents/caregivers complete the questionnaire themselves.

## Data collection

The ECDAS-DA and ECDAS-CS instruments are currently administered using paper/pen. A tablet based version of the surveys is being developed (CORE, 2022).

## Data scoring

The ECDAS-DA and ECDAS-CS provide basic scoring advice using raw scores.

The ECDAS-DA and ECDAS-CS both use a single sum-score to represent children's development (Rao et al., 2021). For the ECDAS-CS, all of the items that are used for SDG 4.2.1 reporting (in line with the ECDI2030) are scored dichotomously (0 – no, 1 – yes). Most of the ECDAS-DA items are also scored dichotomously (0 – incorrect, 1 – correct). Some offer partial credit but are still scaled to range between 0 and 1.

To be consistent with reporting in line with SDG 4.2.1, a method of determining the proportion of children who were "developmentally on track" was employed (Rao et al.,

2021). However, given that not all the 20 ECDI2030 items were included in the assessments, the process that UNICEF (2021) uses could not be followed. Instead, mean age-adjusted scores calculated from "residuals from a regression model predicting the ECDAS-DA and ECDAS-CS scores from age and age squared" (Rao et al., 2021, p. 9) were used to compare results across countries. Three different "developmentally on track" indicators using these scores were used, including:

> (i) not more than 2 SDs below the mean age-adjusted ECDAS-DA [and ECDAS-CS] score, (ii) not more than 1.5 SDs below the mean age-adjusted ECDAS-DA [and ECDAS-CS] score, and (iii) not more than 1 SD below the mean age-adjusted ECDAS-DA [and ECDAS-CS] score (Rao et al., 2021, p. 9)

## Item type(s)

The ECDAS assessments contain multiple item types. Most of the items included in both assessments are dichotomous (yes/no or correct/incorrect). Some items offer opportunities for partial credit, such as writes own name (ECD11: 1 - letter/symbol, 2 – half a name, 3 - first or family name in full). For the ECDAS-DA most items contain sub-tasks where the scores are summed together to create an aggregated score for each overall item. For example, item 10 asks children to look at a picture and point to specific objects, with a score of 1 received for each object up to a maximum of 7.

## Background/contextual information

The ECDAS-CS collects background/contextual information about the child (for example, age, gender, language, and birth-related details), the parents/caregivers and the home learning environment (for example, parental education and occupation, household possessions) and about engagement and experience in ECE or school (for example, length of time in the centre).

## Length/time

The ECDAS-DA contains 36 items and the ECDAS-CS includes 50 items, relating to childhood development.

## Available languages

The ECDAS forms are available in multiple languages. They were originally developed in English, and have been translated into Chinese, Bengali, Hindi and Myanmar. No translation/adaptation guidelines are provided, however a description of the quality assurance process undertaken during translation is available.

## Availability/cost

The ECDAS-DA and ECDAS-CS are not currently publicly available, but can be provided upon request from the developers.

### Training materials/requirements/capacity development

The ECDAS has an established training protocol and requires enumerators to attend administration training. The initial testing of the assessment involved assessors who were qualified and experienced in ECE completing a 2-day training workshop (Rao et al., 2021).

### Validation for use across different populations

The ECDAS-DA and ECDAS-CS were shown to be non-invariant across countries/languages. The ECDAS-DA and ECDAS-CS were tested across 4 different countries, namely Bangladesh, China, India and Myanmar (Rao et al., 2021). Cronbach's Alpha reliability coefficients of 0.75 and 0.62 were found for ECDAS-DA and ECDAS-CS, respectively (Rao et al., 2021). Another article which reports the factor structure of the ECDAS-DA in 4 different countries is currently being developed (Richards et al., in press).

# Early Childhood Development Index 2030 (ECDI2030)

The ECDI2030 was designed to collect population-level data to measure the impact of government action relating to access to quality early childhood development, care and pre-primary education. It measures the proportion of children (between the ages of 24 and 59 months) who are on course to achieve the internationally agreed upon developmental milestones in learning, health and psychosocial wellbeing (United Nations Children's Fund, 2020b). UNICEF convened experts to design the instrument, which has been in use since 2020. At least 18 countries have incorporated it into their national surveys.

### Purpose(s)

The purpose of ECDI2030 is to monitor populations. The data is designed to provide evidence that can inform policy decisions, in particular, to report progress against SDG Indicator 4.2.1.

### Age range

The ECDI2030 is designed to target caregivers of children between the ages of 24 and 59 months.

### Domain(s)

The ECDI2030 assessment contains items related to multiple learning domains, however domains are aggregated into one overall score for reporting.

**Table 9: Summary of the ECDI2030 domains**

| Domain | Number of questions | Description | Sub-domains |
|---|---|---|---|
| Learning | 11 | The early pre-academic skills and competencies critical to the later acquisition of more complex skills and academic success | Expressive language<br>Literacy<br>Numeracy<br>Prewriting<br>Executive functioning |
| Psychosocial wellbeing | 5 | Competencies and behaviours related to forming and maintaining healthy interpersonal relationships with adults and peers and regulating and expressing emotions in socially and culturally appropriate ways | Emotional skills<br>Social skills<br>Internalizing behaviour<br>Externalizing behaviour |
| Health | 4 | Skills and milestones related to fine and gross motor development and self-care. | Gross motor<br>Fine motor<br>Self-care |

Source: (United Nations Children's Fund, 2020b, p. 6)

## Learning progression

There is no learning progression associated with the ECDI2030.

## Assessment framework

The ECDI2030 provides extensive details of assessment development, including theoretical underpinnings and how the items provide construct coverage. Associated documents (P. Halpin et al., 2023; United Nations Children's Fund, 2020b, n.d.) describe the theoretical foundations of the domains being covered and how the items developed for the assessment cover the constructs.

## Administration

The ECDI2030 uses indirect assessment of the child where enumerators complete the questionnaires based on an interview with the mother of the child. In cases where the mother is deceased or is not a household member, an alternative caregiver may be interviewed.

## Data collection

The ECDI2030 is administered using paper/pen.

## Data scoring

The ECDI2030 provides basic scoring advice using raw scores.

The ECDI2030 module is not designed to report on individual domains separately. Rather, it is meant to produce a single summary score that captures the interlinked developmental concepts embedded in the 3 domains specified in SDG 4.2.1.

This single summary score is then compared to the 'cut-score' which defines the minimum number of milestones achieved for the age-group. That is the mother/caregiver needs to respond positively to a pre-determined minimum number of items according the child's age as follows:

- 7 items for children aged 24 to 29 months;

- 9 items for children aged 30 to 35 months;

- 11 items for children aged 36 to 41 months;

- 13 items for children aged 42 to 47 months; and

- 15 items for children aged 48 to 59 months.

(United Nations Children's Fund, 2021)

The sum score can be comprised of any combination of items across the domains being measured. This means, for example, that a child who gets all 11 learning items correct but none of the others is considered at the same developmental level as a child who gets 2 learning items and the remaining 9 psychosocial wellbeing and health items correct.

From the sum scores, a resulting indicator is established by using the number of children who have achieved the minimum number of milestones for their age group as the numerator, and the total number of children as the denominator.

### Item type(s)

The ECDI2030 assessment contains only one type of item. The ECDI2030 items are composed of closed-answer questions. Most items (18 out of 20) use a dichotomous yes/no response scale, and the 2 last items (ECD19 and ECD20) use an ordinal response scale, for example, with 5 options.

### Background/contextual information

The ECDI2030 collects a limited amount of background/contextual information about the child. Prior to the interview an 'age-check' is carried out to ensure that the child is between 24 and 59 months. It is a population-based assessment that is designed to be integrated into existing national data collection efforts such as MICS (UNICEF, 2019).

### Length/time

The ECDI2030 contains 20 items and is typically completed in approximately 3 minutes.

### Available languages

The ECDI2030 assessment is available in multiple languages: Arabic, Chinese, English, French, Portuguese, Portuguese (Brazilian), Russian and Spanish. Implementation tools are also provided in all of these languages except Portuguese (Brazilian). Guidelines are provided for the translation/adaptation of the assessment.

### Availability/cost

The ECDI2030 is an open-source assessment. No access or use fees or royalties are involved.

### Training materials/requirements/capacity development

The ECDI2030 has an established training protocol and requires enumerators to attend administration training.

All interviewers should be trained to administer the questionnaire on paper and additional training is provided for surveys using Computer-Assisted Personal Interviews (United Nations Children's Fund, 2020b). The manual should be reviewed by the fieldwork personnel before the training commences and the training session should take around 3 hours.

### Validation for use across different populations

No studies were found which show that the ECDI2030 is invariant across countries/languages.

All the items in the ECDI2030 were derived from instruments that had been previously validated across different cultural backgrounds. Additional to this, cultural adequacy was one of the criteria used in the initial item selection. To further ensure the cultural validity of the items, the selected items were subjected to cognitive testing. These qualitative interviews were designed to highlight and address any interpretation or adequacy issues and to ensure that the item measures what it intends to measure.

Psychometric analyses have been reported following some trialling in Mexico and Palestine (P. Halpin et al., 2023), resulting in the current version of the ECDI2030. An initial item pool of 58 items was established from pre-existing assessments. Following a series of analytic steps (for example, Classical Test Theory and Item Response Theory for examining item functioning), the item pool was reduced to 20 items.

The psychometric properties (for example, reliability, relative efficiency, concurrent validity, gender differential item functioning (DIF), item characteristic curves (ICCs) of the final set of 20 ECDI2030 items were tested using Mexico and Palestine data. In addition, the authors (P. Halpin et al., 2023) examined the suitability of using raw scores as a proxy for scale scores using correlational and proportion of shared variance analyses. They suggested that the results of this analyses shows that raw scores are an adequate proxy for scale scores.

## Early Development Instrument (EDI)

The Early Development Instrument (EDI) assesses the holistic development (skills and behaviour) of children in the transition from an early age to formal schooling (Offord Centre for Child Studies, 2019k). The EDI was first developed in Canada in 1998 by researchers at the Offord Centre for Child Studies at McMaster University (Offord

Centre for Child Studies, 2019g), and is now in use across many parts of the world (Offord Centre for Child Studies, 2019k).

Australia was the first country outside of Canada to implement the EDI (Janus & Reid-Westoby, 2016), and has gone on to become one of the most prevalent users of the EDI. However, given that the Australian version (and other versions administered in other countries) of the EDI has been based predominantly on the original Canadian version of the EDI, the components reported on in this review will explicitly refer to the original version, unless there is a particular component that would benefit from referring to other (for example, adaptation or cross-cultural validation).

## Purpose(s)

The purpose of the EDI is to monitor populations. The assessment is designed to be able to be used by multiple stakeholders: it can provide governments with evidence for informing policy decisions, and it can provide educators with a diagnosis/screening tool for individual children to create better target their teaching.

## Age range

The EDI has been validated for use with children aged 3.5 to 6.5 years (Offord Centre for Child Studies, 2019k).

## Domain(s)

The EDI is able to report against multiple domains including multiple learning domains.

**Table 10: Summary of the EDI domains**

| Domain | Number of items | Description |
|---|---|---|
| Physical health and wellbeing | 13 | Fine and gross motor, physical independence and physical readiness for school day |
| Social competence | 26 | Overall social competence, responsibility and respect, approaches to learning and readiness to explore new things |
| Emotional Maturity | 30 | Prosocial and helping behaviour, anxious and fearful behaviour, aggressive behaviour and hyperactivity and inattentive behaviour |
| Language and cognitive development | 26 | Basic literacy, advanced literacy, basic numeracy and interest in literacy/numeracy and memory |
| Communication skills and general knowledge | 8 | Communication and general knowledge |

Source: (Offord Centre for Child Studies, 2019b)

## Learning progression

There is no learning progression associated with the EDI.

### Assessment framework

The EDI provides extensive details of assessment development, including theoretical underpinnings and how the items provide construct coverage. Associated documents (Janus & Offord, 2007; Offord Centre for Child Studies, 2019b) describe the theoretical foundations of the domains being covered and how the items developed for the assessment cover the constructs.

### Administration

The EDI uses indirect assessment of the child where "individuals who know the children well in an early learning setting" (Offord Centre for Child Studies, 2019h, para. 1) (for example, educators) complete the questionnaire.

### Data collection

In Canada and Australia, the EDI instrument is administered using a digital format. It is not clear from the documentation whether a particular method of data collection is preferred; however, it is a requirement of using the EDI that users "share an electronic copy of raw data collected using the EDI with the [Offord Centre for Child Studies], following a signed data-sharing agreement" (Offord Centre for Child Studies, 2019j, para. 3).

### Data scoring

The EDI provides basic scoring advice using raw scores.

Publishers of the EDI suggest that it is appropriate to "collect the data categorically, and then convert it into a mean as if the data were continuous" (Offord Centre for Child Studies, 2019c, para. 13). They do note, though, that categorical data is rescaled to ensure equal weighting of items in the calculation of mean scores (Offord Centre for Child Studies, 2019c). In addition to reporting mean scores, the assessment publishers suggest presenting the results of domain-specific scores from the EDI in terms of dividing the population into 'vulnerable' (below the bottom 10% of baseline scores), 'at risk' (between the bottom 10%- 25% of baseline scores), and 'on track' (above the bottom 25% of baseline scores), where the baseline refers to the very first full implementation of the EDI in a particular context (for example, country, region) (Offord Centre for Child Studies, 2019a). Using the baseline scores as a comparison for all future rounds of data collection allows users "to determine whether children's developmental outcomes are getting better or worse" (Offord Centre for Child Studies, 2019a, p. 8).

### Item type(s)

The EDI assessment contains multiple item types. Items are dichotomous (for example, yes/no) and polytomous (for example, very good or good/average/poor or very poor; often or very true/sometimes or somewhat true/never or not true) (Offord Centre for Child Studies, 2019f). Each item also allows for a 'don't know' response, but enumerators are warned that these are treated as missing data and may "render the

relevant section of the questionnaire unscoreable" (Australian Early Development Census, 2017, p. 22).

## Background/contextual information

The EDI collects background/contextual information about the child and their engagement and experience in ECE or school (for example, enrolment status, grade repetition). Questions about the child include date of birth, sex, postcode, grade, class type, special needs, other educational needs (for example, exceptional, gifted, IEP, special education), language status, first language, enrolment status, grade repetition, the child's special skills, problems and kindergarten history (Offord Centre for Child Studies, 2018).

## Length/time

The EDI contains 103 items and is typically completed in approximately 7 to 20 minutes (Janus & Reid-Westoby, 2016).

## Available languages

The EDI assessment is available in multiple languages: English and French. Requirements for the translation/adaptation of the assessment into other languages are stipulated by the assessment publishers. Thes include:

- consultation with local ECD experts to establish the relevance of the items

- modifications to the EDI to suit the contextual needs of the user

- conducting a pilot

- evaluating the validity and reliability of the adapted instrument (for example, test-retest reliability and construct validity) (Offord Centre for Child Studies, 2019h).

## Availability/cost

The EDI incurs a cost to use the assessment materials.

The EDI has a copyright, and therefore, requires a license for administration (Offord Centre for Child Studies, 2019i). In addition to the costs associated with licensing, there are also costs for scoring the EDI and assistance with the adaptation of the instrument based on the intended setting and sample size (Offord Centre for Child Studies, 2019i).

## Training materials/requirements/capacity development

The EDI has an established training protocol and requires enumerators to attend administration training. The training session aims to "ensure accurate, consistent interpretation of items, as well as inform respondents about the purpose of data collection, how results will be used, and the logistics of the data collection process" (Janus et al., 2007, p. 11). A variety of additional resources is also available to assist enumerators to undertake the EDI. Resources include videos, guides and manuals, and

presentations (the resources provided to Newfoundland (Canada) teachers are available at: (Offord Centre for Child Studies, 2019e)).

### Validation for use across different populations

The EDI has been validated using appropriate measurement invariance analysis methods and shown to be either fully or partially invariant across countries/languages.

The EDI has been administered in many different contexts in different ways, including regional collections (for example, province, state, county, country), district-wide collections (for example, municipalities and neighbourhoods) and smaller samples for research or pilot projects (Offord Centre for Child Studies, 2019d).

The first fully developed version of the EDI has been psychometrically examined using multilevel confirmatory factor analysis (Janus & Offord, 2007). The results showed: low intra class correlations for each domain (low between class/teacher variance), high teacher reliability estimates (using hierarchical liner modelling methodology: 0.76-0.84), and good internal consistency estimates (Cronbach's alpha: 0.84-0.96) (Janus & Offord, 2007).

Alternate versions of the EDI have been psychometrically tested and validated (inc. internal consistency, test-retest reliability, construct validity, predictive validity and differential item functioning) in a number of different countries (Janus & Reid-Westoby, 2016). One study presented the results of psychometric analyses (confirmatory factor analyses) across 4 different countries, namely Canada, United States, Australia and Jamaica (Janus et al., 2011). However, the authors simply presented the results of the analyses (inc. internal consistency estimates (Cronbach's alpha), factor loadings, goodness of fit statistics, and convergent validity) for each country separately (Janus et al., 2011), which doesn't explicitly assess measurement invariance across countries. A later publication by the same authors showed that the fit of the models to the data was below acceptable thresholds for several of the domains, resulting in cross loadings being introduced to establish measurement models that would meet the desired criteria (Duku et al., 2015). Only 3 of the 5 domains were able to be used for subsequent analyses (Duku et al., 2015). Using these newly established models, measurement invariance across countries was examined, with partial invariance established for 2 domains (social competence and emotional maturity), and full invariance established for the language and cognitive development domain (Duku et al., 2015).

# Early Grade Reading Assessment (EGRA) and Early Grade Mathematics Assessment (EGMA)

The Early Grade Reading Assessment (EGRA) and Early Grade Mathematics Assessment (EGMA) are short orally administered assessments which aim to measure young children's (Grades 1-3) foundational skills in reading and mathematics (RTI International, 2012). The assessments were developed by RTI International and have

been used in many countries globally. EGRA has been in use since 2007 and the current iteration of EGMA (Core EGMA) has been in use since 2011.

## Purpose(s)

The purpose of EGRA and EGMA is to monitor populations. Data is primarily used to monitor and evaluate the effectiveness of curriculum changes, school-based interventions and teacher professional development programs (RTI International, 2014a). Components from both EGRA and EGMA can also be used by teachers and schools as a formative diagnostic assessment (RTI International, 2014a, 2016).

## Age range

EGRA and EGMA target students in Grades 1-3 (ACER, 2020a, 2020b).

## Domain(s)

The EGRA and EGMA are able to report against multiple learning domains.

**Table 11: Summary of the EGRA and EGMA domains**

| Domain | Sub-domains |
| --- | --- |
| Literacy (core) | Listening comprehension |
| | Letter identification (names and sounds) |
| | Nonword reading |
| | Oral reading fluency with comprehension |
| Literacy (non-core) | Phonemic awareness |
| | Familiar word reading |
| | Dictation |
| | Phoneme segmentation |
| | Maze |
| | Cloze |
| Numeracy | Number identification |
| | Number discrimination (relative size) |
| | Number pattern identification |
| | Addition |
| | Subtraction |

## Learning progression

There is no learning progression associated with the EGRA or EGMA.

## Assessment framework

The EGRA and EGMA toolkits (RTI International, 2014a, 2016) both provide extensive details of assessment development, including theoretical underpinnings and how the items provide construct coverage. The toolkits highlight the theoretical foundations of the domains being covered using key literature in the field. They describe the changing

nature of the assessments over time as they were tested and revised (due to ongoing validation analyses in differing contexts). The toolkits also explain how the EGRA and EGMA are designed to provide construct coverage.

## Administration

EGRA and EGMA are orally administered and use direct assessment of the child where trained assessors individually administer the assessment.

## Data collection

EGRA and EGMA provide the the option to use either paper/pen or digital formats to collect data. The assessment is orally administered with the student referring to paper-based materials (for example, a text to read aloud) while the enumerator records the responses on paper or on tablet.

## Data scoring

EGRA and EGMA use sum scores in their reporting.

Scoring guides are provided for all items in EGRA and EGMA materials. The toolkits provide guidance on how to use multiple items to produce a score on subtasks. For example, for letter identification (sounds) in EGRA "three data points are used to calculate the total correct letter sounds and diphthongs/digraphs per minute (clspm):" (RTI International, 2016, p. 45)

$$clspm = (Total\ letter\ sounds\ identified - Total\ incorrect) / [\frac{(60 - Time\ remaining\ on\ device)}{60}]$$

Given that reporting is not aggregated across modules it does provide substantial information about student achievement. The mean raw scores by module clearly identify the relative performance of sub-groups of students on a specific skill, such as letter naming or word reading, with the maximum score shown. This form of reporting allows for formative interpretation: for example, a mean score of 1.5 for reading comprehension (out of a possible score of 5) suggests weakness in that skill.

## Item type(s)

EGRA and EGMA contain multiple item types. The items are typical of orally delivered assessments in the early years, with the use of observation checklist-style scoring (for example, 1 – correct; 0 – incorrect; missing - no response, count/tally). Given the emphasis placed on fluency in the EGRA, some items are timed. The speed and accuracy with which students answer questions feed into the scoring of subtasks in the assessments.

## Background/contextual information

EGRA and EGMA collect background/contextual information about the child. Questions consist of the child's grade, age and gender.

The Snapshot of School Management Effectiveness (SSME) tool (see RTI International, 2010) is also commonly used in conjunction with the EGRA and EGMA, to collect background/contextual information about engagement and experience in school. Topics include school management and teaching practice (ACER, 2020a).

### Length/time

The number of items included in the EGRA varies widely depending upon the content chosen by the administering organisations (RTI International, 2016). The core EGRA assessment includes 96 items (ACER, 2020a). The appropriate administration length for the EGRA and EGMA is deemed to be 15-20 minutes (ACER, 2020a; RTI International, 2016). Stop rules are used and are based on errors made and different time-based instructions (for example, 'STOP if the child doesn't respond after 5 SECONDS'). Therefore, the actual number of administered items depends on the start and stop rules so the administration of the assessment could take less than 15-20 minutes.

### Available languages

EGRA and EGMA are available in English only. However, guidelines are provided for the translation/adaptation of the assessment to suit the contextual needs of the population under examination.

As of 2016, the EGRA had been used by over 30 organisations in more than 120 different languages across more than 70 countries (Beggs, 2016). It was also stated that the EGMA had been administered in 22 countries, being adapted into 24 different languages (Beggs, 2016).

### Availability/cost

EGRA and EGMA are an open-source assessments. No access or use fees or royalties are involved.

### Training materials/requirements/capacity development

EGRA and EGMA have an established training protocol and require enumerators to attend administration training. Extensive details and suggestions are provided in the EGRA and EGMA toolkits on the training needs of assessors (RTI International, 2014a, 2016). Considerations are given for recruitment of training participants, planning the training event, components of assessor training, training methods and activities, school visits, assessor evaluation process, and measuring assessors' accuracy (RTI International, 2016). It is suggested that assessor training can take up to 5 days, depending on the number and length of the assessments, number of available trainers, assessors prior experience, budget, and time (RTI International, 2016). More detailed information is also available in Guidance Notes for Planning and Implementing EGRA (RTI International & International Rescue Committee, 2011).

**Validation for use across different populations**

No studies were found which show that the EGRA or EGMA is invariant across countries/languages. In fact, assessment authors recommend against cross-language comparison due to differences in phonology across languages (RTI International, 2014a, 2016; RTI International & International Rescue Committee, 2011). However, both assessments have been trialled and validated in a number of countries separately during field testing and main studies using psychometric methods (for example, Afghanistan (Kan et al., 2022), Kenya (RTI International, 2014b)).

Despite recommending against comparing across languages, a cross-linguistic comparability study has recently been done, evaluating the underlying latent structure of EGRA domains for Kyrgyz, Russian and Tajik languages (Drummond & Nakamura, 2021). Using factor analysis techniques with representative samples of students in Grade 2 and Grade 4, they showed that there were 2 underlying reading constructs that were common across the 3 language groups, namely decoding and language comprehension (Drummond & Nakamura, 2021).

# Early Human Capability Index (eHCI)

The early Human Capability Index (eHCI) is a holistic assessment of early childhood development aimed at measuring the development of 3-5 year old children from diverse cultures and contexts (Brinkman & Kinnell, 2022). The majority of the use of the eHCI so far has been to provide a snapshot of children's development in low- and middle-income countries for the purpose of informing program and policy level decision making. The assessment was developed by researchers at the Telethon Kids Institute and has been in use since 2015 (Telethon Kids Institute, 2022a).

The Australian version of the eHCI (Brinkman & Kinnell, 2013) will be referred to (where appropriate) in this review.

**Purpose(s)**

The purpose of eHCI is to monitor populations. Data is primarily used to inform program and policy-level decision-making, to evaluate the impact of interventions, and for longitudinal studies of cohorts (Brinkman & Kinnell, 2022).

**Age range**

The eHCI is designed to assess children between 3 and 5 years old (Brinkman & Kinnell, 2022).

**Domain(s)**

The eHCI assessment contains items related to multiple domains including multiple learning domains, however domains are aggregated into one overall score for reporting.

**Table 12: Summary of the eHCI domains**

| Domain | Example question |
|---|---|
| General verbal communication | Can this child communicate their needs by crying or pointing? |
| Approaches to learning | Does this child show more curiosity about something new in comparison to something familiar? |
| Numeracy and concepts | Can this child recognise geometric shapes (for example, triangle, circle, square)? |
| Formal literacy - reading | Can this child follow reading directions? (i.e. left to right, top to bottom) |
| Formal literacy – writing | Can this child scribble on a page using a pen/pencil/crayon? |
| Cultural knowledge | Can this child identify 2 local animals? |
| Social and emotional skills | Is this child happy to share their toys and belongings? |
| Perseverance | Does this child perform tasks independently? |
| Physical health | Is this child frequently sickly? |

Source: (Brinkman & Kinnell, 2022)

## Learning progression

There is no learning progression associated with the eHCI.

## Assessment framework

The eHCI provides limited details of assessment development, including theoretical underpinnings and how the items provide construct coverage. The eHCI manual (Telethon Kids Institute, 2022d) includes some background information on early childhood development and human capability, which provides some context to the development of the assessment.

## Administration

The eHCI uses indirect assessment of the child where "parents/caregivers, childcare workers, teachers, allied health and other health or early childhood practitioners" (Telethon Kids Institute, 2022e, para. 2) of the child complete the questionnaire.

## Data collection

The eHCI is administered using paper/pen.

## Data scoring

The eHCI uses sum scores in its reporting.

A number of resources are available for each version of the eHCI to assist users with scoring the eHCI, including a user manual, data entry sheet, scoring guide, and syntax for data analysis using either SPSS, Stata, SAS, or R (Telethon Kids Institute, 2022d). The scoring guide indicates that each of the items included in the assessment (except the item about height, weight and disabilities/special needs) is scored dichotomously (0 or 1) and the average (mean) is taken for each of the individual domains to create scale scores (Telethon Kids Institute, 2022d). Items that should be reverse scored are also highlighted

in the scoring guide. A scale score can only be calculated if no more than 20% of the items for each domain are missing (Telethon Kids Institute, 2022d). Two aggregated summary indicators are also suggested which provide assessments of literacy/numeracy and overall development (Telethon Kids Institute, 2022d). Assuming the provided scoring sheet and syntax are used as recommended, scale scores for each of the listed domains will be produced.

### Item type(s)

The eHCI assessment contains only one type of item for the purpose of calculating scale scores. The items are dichotomous, either yes/no, can already/can't yet, or yes/not yet.

### Background/contextual information

Depending on the version used, the eHCI collects background/contextual information about the child, the parents/caregivers and the home learning environment, and about engagement and experience in ECE or school.

The Australian version of the eHCI collects background/contextual information about the child (for example, date of birth, gender), and the parents/caregivers and the home learning environment (for example, mother's education). Some other versions of the eHCI collect additional information required for the context in which they were being implemented, including fathers' education level, the location/village/community where the child lives, and the child's engagement in pre-school/kindergarten (Telethon Kids Institute, 2022e).

### Length/time

The Australian version of the eHCI contains 60 items and is typically completed in approximately 10 minutes (Telethon Kids Institute, 2022b).

### Available languages

The eHCI assessment is available in multiple languages: English (8 versions for implementation in different contexts), Portuguese, Chinese, Kiribati, Samoan, Tongan, and Tuvaluan (Telethon Kids Institute, 2022e). There are no guidelines provided for the translation/adaptation of the assessment.

### Availability/cost

The eHCI is an open-source assessment. No access or use fees or royalties are involved and it is available for download from the Telethon Kids Institute website (Telethon Kids Institute, 2022e).

### Training materials/requirements/capacity development

The eHCI provides training materials to users of the tool. Minimal training is required to administer the eHCI (Telethon Kids Institute, 2022b). Multiple resources are freely available to assist with assessment administration, including a scoring guide, a user

manual (with detailed descriptions and examples for each item), scoring sheets and scoring syntax (Telethon Kids Institute, 2022d).

### Validation for use across different populations

The eHCI has been shown to be non-invariant across countries/languages. The eHCI has been administered in 7 low- and middle-income countries: Tonga, Samoa, Lao PDR, Tuvalu, Brazil, Kiribati and China (Telethon Kids Institute, 2022a). Various results from the studies that used the eHCI are available in different formats, including case studies, study reports, research snapshots and conference presentations (Telethon Kids Institute, 2022c). In 2019, the psychometric properties of the eHCI were examined across the 7 countries (Sincovich et al., 2019). Confirmatory factor analytic methods in MPlus software were used to examine the underlying factor structure in each of the countries separately (Sincovich et al., 2019). Analyses included fit statistics (Root Mean Square Error of Approximation (RMSEA), Comparative Fit Index (CFI) and Tucker-Lewis Index (TLI)) which were used to show that each of the models (specified using the 9 theoretically defined domains) fitted the data adequately (Sincovich et al., 2019). Factor loadings for literacy and numeracy domains tended to be consistently high across all countries (Sincovich et al., 2019). However, the factor loadings for the other developmental domains varied across countries (Sincovich et al., 2019). Internal consistency estimates (Cronbach's alpha and ordinal alpha) also showed substantial variability across countries (Sincovich et al., 2019).

# East Asia-Pacific Early Child Development Scale (EAP-ECDS)

The East Asia-Pacific Early Child Development Scale (EAP-ECDS) is a regional instrument, based on the UNICEF's Early Learning and Developmental Standards (ELDS) for measuring early childhood development in the East Asia and Pacific region (Rao et al., 2014). The University of Hong Kong (HKU), The Asia-Pacific Regional Network for Early Childhood (ARNEC), and the Open Society Foundation (OSF) collaborated to develop and validate the EAP-ECDS assessment. The instrument was developed between 2010 and 2014 and has been validated across 6 countries in the region: Cambodia, China, Mongolia, Timor-Leste, Papua New Guinea, and Vanuatu.

### Purpose(s)

The purpose of the EAP-ECDS is to monitor populations. Its primary use is to "inform evidence-based decision-making about ECD policies, services, and programmes" (Raghaven & Santiago, 2016, p. 1).

### Age range

The EAP-ECDS is designed to assess children aged between 3 and 5 years.

## Domain(s)

The EAP-ECDS is able to report against multiple domains including multiple learning domains. The 7 domains covered are:

- Cognitive development
- Social-emotional development
- Motor development
- Language and emergent literacy
- Health, hygiene, and safety
- Cultural knowledge and participation
- Approaches to learning.

## Learning progression

There is no learning progression associated with the EAP-ECDS.

## Assessment framework

The EAP-ECDS provides extensive details of assessment development, including theoretical underpinnings and how the items provide construct coverage in the technical report (Rao et al., 2014).

## Administration

The EAP-ECDS uses direct assessment of the child where trained assessors individually administer the assessment.

## Data collection

The EAP-ECDS is administered orally where results are recorded using paper/pen.

## Data scoring

The EAP-ECDS provides detailed information about data processing (including scoring) and converting raw scores onto an ordinal/linear scale. Scoring forms are available and the assessors score these forms manually.

## Item type(s)

The EAP-ECDS assessment contains only one type of item. All items are open-ended. Many of them "require oral responses from the child or necessitate the manipulation of test materials by the child" (Consortium for Research on Early Childhood Development and Education (CORE), 2017, p. 4). For example, for cognitive development, the child "Named at least 4-6 simple geometric shapes," or for language and emergent literacy, the child "Wrote about half of their name without model."

## Background/contextual information

The EAP-ECDS collects background/contextual information about the child, the parents/caregivers and the home learning environment and about engagement and experience in ECE or school. The parent questionnaire has 70 items collecting information about parents' age, education and occupation, as well as home resource availability, enrolment in early childhood settings, and parental engagement in their child's learning.

## Length/time

The short form includes 33 items and the long form includes 85 items. The items are administered in a fixed order. However, the assessment allows for some country level adaptation and flexibility through options for adding extra items (in the actual Assessment form or Parent Questionnaire). The assessment is untimed, with the long form usually taking around 45 - 60 minutes.

## Available languages

The EAP-ECDS assessment has been conducted in multiple languages: English, Khmer, Chinese, Mongolian, Tok Pisin, Tetum and Bislama. No guidelines were found for translation/adaptation of the assessment.

## Availability/cost

The EAP-ECDS is available to countries in the East Asia-Pacific region through the UNICEF EAPRO funding.

## Training materials/requirements/capacity development

The EAP-ECDS has an established training protocol and requires assessors to attend administration training. Assessors are also required to have some experience or training in early childhood education. An Instruction Manual, Scoring Forms, and training videos are available. The assessor needs to be familiar with all the assessment materials and is expected to practise administering and scoring under the supervision of an experienced assessor, before going into the field.

## Validation for use across different populations

The EAP-ECDS has been validated in different populations separately, but evidence for invariance across countries/languages has not been found. The original version of the assessment was validated in Cambodia, China, Mongolia, Papua New Guinea, Timor-Leste, and Vanuatu with a total sample of 8,439 children (4,215 girls) (Rao et al., 2014). The results reported suggest high internal consistency for all domains in all 6 countries. A recent longitudinal study in China has reported on additional psychometric properties of the EAP-ECDS short form (Rao et al., 2023), including test-retest reliability ($r$ coefficient and ICC for raw scores for the EAP-ECDS scale were 0.58 and 0.73, respectively), predictive validity analyses (correlations of 0.59, 0.43 and 0.50 for school readiness, language and literacy and mathematics, respectively) and regression analyses

(EAP-ECDS scale score was a significant predictor of school readiness ($\beta$=0.26, *SE*=0.05), language and literacy ($\beta$=0.18, *SE*=0.07) and mathematics ($\beta$=0.22, *SE*=0.07)).

# Global Scales for Early Development (GSED)

The Global Scales for Early Development (GSED) are a set of assessments designed to measure the developmental level of children aged 0-3 years for the purpose of population monitoring and program evaluation. The GSED were developed by a team of researchers assembled by the World Health Organization. The assessments have been validated in 3 countries as of 2023 with further validation currently in progress in 4 additional countries (World Health Organization, 2023h).

### Purpose(s)

The purpose of the GSED Short Form (GSED SF) is to monitor populations. The purpose of the GSED Long Form (GSED LF) is to evaluate programs (GSED Team, 2021). Caregiver reported assessments of the household (household form (GSED HF)) and psychosocial behaviour (psychosocial form (GSED PF)) are currently being developed and tested, but this is not the focus of this review.

### Age range

The GSED are designed to assess children aged between 0 and 3 years (GSED Team, 2019).

### Domain(s)

The GSED assessment contains items related to multiple domains including multiple learning domains, however domains are aggregated into one overall score for reporting. The domains are:

- Cognitive development
- Motor development
- Language development
- Socio-emotional development.

(World Health Organization, 2022)

### Learning progression

There is no learning progression associated with the GSED.

### Assessment framework

The GSED provides limited details of assessment development, including theoretical underpinnings and how the items provide construct coverage. A technical report for the GSED is available which provides some basic information about the items included in the assessments and their ability to provide construct coverage (World Health

Organization, 2023h). However, a dedicated assessment framework with extensive details of intended construct coverage is not available.

### Administration

The GSED-LF uses direct assessment of the child where trained assessors individually administer the assessment. The GSED-SF uses indirect assessment of the child where enumerators complete the questionnaires based on an interview with the parents/caregivers of the child.

### Data collection

The GSED provides the option to use either paper/pen or digital formats to collect data. The preferred option is to use the digital format via tablet as it makes use of interactive media (for example, audio, visuals, videoclips) (GSED Team, 2021).

### Data scoring

The GSED provides detailed information about data processing (including scoring) and converting raw scores onto an ordinal/linear scale.

The GSED was developed with the aim of producing a single scale for early child development (GSED Team, 2019), termed a developmental score (D-score), similar to that proposed by Jacobusse, van Buuren and Verkerk (2006).

A scoring guide is available for the GSED which includes explanations of the different scoring options, links to a website (https://d-score.org) and R package (https://CRAN.R-project.org/package=dscore) for automated scoring and associated syntax (with annotations) (World Health Organization, 2023d). The scoring software allows for 2 different types of scores to be produced:

- D-score –single developmental score representing holistic development on an interval scale

- DAZ – development-for-age-*z*-score which is "age-independent and is scaled such that at each age, the distribution of scores is normally distributed with a mean of 0 and a variance of 1" (World Health Organization, 2023d, p. 1)

The D-score is calculated using the expected a-posteriori (EAP), which is the mean of the posterior distribution for each case, constructed from the IRT and latent regression model fit to the data. Sub-domain scores can also be produced but these are not recommended to be used for practical applications (World Health Organization, 2023d).

The publishers note that the DAZ is not currently appropriate for benchmarking or determining whether children are developmentally on-track or not (World Health Organization, 2023d). However, the DAZ has been used for some validation analyses.

## Item type(s)

The GSED SF and GSED LF assessments contain only one type of item. All items are dichotomous (yes/no) items, with an option of pass/don't know responses (World Health Organization, 2023e, 2023a).

## Background/contextual information

The GSED SF and GSED LF collect background/contextual information about the child. Questions consist of the child's age (in months), the child's date of birth and the administration date.

## Length/time

The GSED SF contains 139 items in total, however the actual number of administered items depends on the start and stop rules. The GSED SF is typically completed in approximately 5 to 10 minutes. The GSED LF contains 155 items in total, however again, the actual number of administered items depends on the start and stop rules. The GSED LF is typically completed in fewer than 30 minutes. (L. Richter et al., 2019).

The GSED Team (2021) is currently trialling an adaptive testing approach in 3 countries to reduce the number of items administered to each child.

## Available languages

The GSED is available in multiple languages upon request (World Health Organization, 2023h). The English version is the only publicly available version. Guidelines are provided for the translation/adaptation of the assessment.

## Availability/cost

The GSED version 1.0 is an open-source assessment. No access or use fees or royalties are involved and the assessment forms and supporting materials are available from the WHO website.

## Training materials/requirements/capacity development

The GSED has an established training protocol and requires assessors to attend administration training. Assessors are required to have the following experience/qualifications (World Health Organization, 2023g, 2023c):

- Completed secondary school
- Experience building rapport with children and their families
- Familiarity with local customs
- Fluency in the administration language.

Detailed user manuals and item guides are available for both the GSED SF and the GSED LF (World Health Organization, 2023g, 2023c, 2023f, 2023b). However, it is still suggested that assessors undergo the relevant "training and certification to ensure all

administration rules are being followed and that they are familiar with tablet use prior to conducting the assessments" (World Health Organization, 2023g, p. 2). In-person or online training (approximately 2-3 days in length) is available in English upon request, with self-paced training currently being developed (World Health Organization, 2023g).

**Validation for use across different populations**

The GSED has been shown to be non-invariant across countries/languages.

A number of validation analyses have been conducted in a diverse range of contexts, which includes internal reliability, external reliability, concurrent validity, convergent validity, known groups validity and short-term predictive validity (World Health Organization, 2023h). The main validation phase (including adaptive testing) has been undertaken in Bangladesh, Pakistan and Tanzania with a total sample of 4,349[9] children being used for analyses (World Health Organization, 2023h). In this study, the Rasch model was applied to the GSED SF, GSED LF, and the combined forms (CB). Internal reliability was measured using test information, showing that the reliability of the GSED SF >0.8, GSED LF >0.8 and the GSED CB >0.9 for most of the scale. Inter-rater reliability and test-retest reliability was >0.97 for all countries across all forms (GSED SF, GSED LF and GSED CB). A 2-parameter logistic (2PL) model was estimated and used to evaluate concurrent and convergent validity. Concurrent validity was examined using Pearson correlations between the D-score for each country and the sub-domains and overall domain of the Bayley-III, showing correlations >0.87 for all combinations. Results from the convergent validity analyses showed that there were statistically significant (at 5% level) correlations between the DAZ scores and compared variables of interest (for example, wealth, birth weight, gestational age, maternal education, various childhood development scales) in the expected directions for total scores (aggregated across countries). However, there were some non-significant findings at the country-level. Known groups validity analyses showed that there were statistically significant (5% level) differences (none of the Cis for odds ratios contained zero) for all comparison groups of interest (for example, premature birth, low birth weight, maternal drug use during pregnancy, stunted growth). Short-term predictive validity was examined for the DAZ scores 6 months apart, showing a correlation of 0.59 across the countries in aggregate.

Differential item functioning and assessing the cultural comparability of items are currently being planned (GSED Team, 2021).

Future validation work is planned for Brazil, China, Cote D'Ivoire and the Netherlands, which may result in revisions to the assessment forms (GSED Team, 2021).

# International Development and Early Learning Assessment (IDELA), Save the Children

---

[9] Smaller sub-samples were used for different types of validation analyses depending upon the amount of data available and resource constraints.

The International Development and Early Learning Assessment (IDELA) aims to help governments and global actors identify and scale effective early childhood care and development programs. The assessment was developed by Save the Children between 2011 and 2015 and has been used in the evaluation of early childhood programs in 76 countries.

## Purpose(s)

The purpose of IDELA is to monitor populations. Data is used to monitor and evaluate programs and to provide evidence for informing policy decisions.

## Age range

The IDELA is designed to assess children aged between 3.5 and 6 years (Save the Children, 2022).

## Domain(s)

The IDELA is able to report against multiple domains including multiple learning domains.

**Table 13: Summary of the IDELA domains**

| Domain | Examples |
|---|---|
| Motor development (fine and gross motor skills)) | Hopping<br>Copying shapes<br>Folding paper<br>Drawing |
| Emergent language and literacy | Print awareness<br>Expressive vocabulary<br>Letters<br>Phonological awareness<br>Listening comprehension |
| Emergent mathematics and numeracy | Number sense<br>Shapes<br>Sorting<br>Problem solving<br>Comparison<br>Simple operations |
| Social-emotional development | Empathy<br>Emotional awareness<br>Self awareness<br>Solving conflict<br>Peer relationships |

Source: (Save the Children, 2022)

The IDELA long form also includes items measuring additional. These items measure: Inhibitory control, memory, and children's learning approaches.

### Learning progression

There is no learning progression associated with the IDELA.

### Assessment framework

The IDELA provides extensive details of assessment development, including theoretical underpinnings and how the items provide construct coverage. Pisani (2018) described the theoretical foundations of the domains being covered and how the items developed for the assessment cover the constructs.

### Administration

The IDELA uses direct assessment of the child where trained enumerators individually administer the assessment.

### Data collection

The IDELA provides the option to use either paper/pen or digital formats to collect data.

### Data scoring

The IDELA uses sum scores in its reporting.

IDELA is scored as each item is administered. For administrators using the tablet-based version, this scoring occurs automatically, while those using the paper-based version need to score items manually, immediately after the child provides a response (Pisani et al., 2017).

Administrators use provided scoring guidance and scoring sheets to record responses. Detailed guidance is also provided to clean and analyse data, using Microsoft Excel as an example (Save the Children, n.d.). Sub-domain scores and total scale scores are calculated based on raw scores, expressed as percentage correct.

### Item type(s)

The IDELA assessment contains multiple item types. The majority of the items are dichotomous: correct/incorrect or I don't know. The remaining items are continuous. For these items, the enumerator counts a child's response and records the number in the scoring sheet. The purpose of these items is to gain an understanding of the range of responses a child can provide or the depth of skills they have. The maximum number of responses that can be recorded is 10, even if the child provides responses beyond this.

### Background/contextual information

The IDELA collects background/contextual information about the child, the parents/caregivers and the home learning environment (Caregiver Survey) and about engagement and experience in ECE or school (Classroom Environment Tool). Questions about the child include their full name, sex, and age (from birth certificate if available). It is left to the discretion of the project team whether further background information about the child is required (Save the Children, 2019).

Save the Children recommend also administering a Caregiver Survey and a Classroom Environment Tool. The caregiver survey heavily mirrors the fourth round of the Multiple Indicator Cluster Survey (MICS4) developed by UNICEF. It is administered through an interview with the child's primary caregiver and takes approximately 20-30 minutes. Similar to the child assessment, additional items and sections are available to meet specific contextual needs, such as disability (Pisani et al., 2017). The classroom environment tool is designed to measure classroom quality within early childhood centres. The focus is on learning environments for children aged 3.5 – 6 and is not considered appropriate for classrooms or centres with very young children (0-3 year olds) (Save the Children, 2017).

### Length/time

The IDELA contains 22 core items and is typically completed in approximately 35 minutes (Save the Children, 2022).

### Available languages

The IDELA assessment is available in more than 50 languages. New or updated translations are received by the assessment authors monthly (Save the Children, 2022). Guidelines are provided for the translation/adaptation of the assessment.

### Availability/cost

The IDELA is an open-source assessment. No access or use fees or royalties are involved.

### Training materials/requirements/capacity development

The GSED has an established training protocol and requires enumerators to attend administration training. IDELA provides a comprehensive training program that typically lasts 4 to 5 days. Training includes in-person exercises and hands-on field training. Enumerator training begins with practice using IDELA with peers in a controlled setting, and then pilot testing in locations with young children in communities similar to those in the study sample. When selecting enumerators, priority is given to those with previous experience working with young children. However, no formal training is required (Pisani et al., 2017).

IDELA also offers a series of training videos that provide guidance on preparation, tasks and scoring. The videos also provide examples of interactions between children and assessors, showing the diverse range of behaviours that children might demonstrate during an assessment. The videos, however, are not designed to replace the training and are meant to used alongside the full IDELA toolkit of guides, translations and data analysis tools (Save the Children, 2022).

Enumerators are also supported by a dedicated team of IDELA experts and a community of Master Trainers located around the world.

### Validation for use across different populations

The IDELA has been shown to be non-invariant across countries/languages. In terms of validation across countries, Halpin et al. (2019) conducted a measurement invariance analysis (using a multigroup CFA approach in Mplus) of the IDELA assessment across 5 diverse low- and middle-income countries (Afghanistan, Bolivia, Ethiopia, Uganda and Vietnam), to determine its useability across different populations. The results of the study provided initial evidence to support the use of IDELA for the purpose of program evaluation and within-country monitoring. It concluded, however, that it was not suitable for making international comparisons. Scalar invariance models were rejected under a number of conditions, showing that full invariance across countries could not be established. Partial invariance was also tested, but results showed that there was DIF for most of the items across the IDELA domains for 2 or 3 of the 5 countries examined. The findings suggested that while the conceptual model underpinning IDELA generalised across the 5 countries, the domains measured by IDELA tend to be highly correlated with one another. The researchers emphasised that this was not an issue specific to the IDELA, but was, "reflective of cultural and contextual variation in expectations about child development at the level of specific skills and competencies" (P. F. Halpin et al., 2019, p. 36). The researchers suggested that it would be useful to "reconcile the richness and nuance of multidomain conceptualizations of ELD with the complexity of making generalizations in international settings" (P. F. Halpin et al., 2019, p. 35), and that additional research might support the identification of a core subset of items that could be used to support cross-country comparisons.

## Infant and Young Child Development (IYCD)

The Infant and Young Childhood Development (IYCD) is an assessment that aims to measure early development across contexts in low- and middle-income settings. It was developed by a team of researchers convened by the World Health Organisation in 2018 and has been validated in 3 countries.

### Purpose(s)

The purpose of IYCD is to monitor populations (Gladstone et al., 2021; Lancaster et al., 2018).

### Age range

The ICYD targets caregivers of children aged 0 and 3 years (Gladstone et al., 2021; Lancaster et al., 2018).

### Domain(s)

The IYCD reports against multiple domains including a learning domain.

**Table 14: Summary of the IYCD domains**

| Domain | Sub-domains | Number of items | Example |
|--------|-------------|-----------------|---------|
| Motor | Fine and gross | 40 | Does your child dress him/herself (except for shoelaces, buttons and zippers)? |
| Language and cognitive | Expressive and receptive | 30 | Does your child tell a story? |
| Socio-emotional | | 20 | Does your child have difficulty taking turns? |
| Behaviour | | 10 | |

Source: (Gladstone et al., 2021)

## Learning progression

There is no learning progression associated with the IYCD.

## Assessment framework

The IYCD provides limited details of assessment development, including theoretical underpinnings and how the items provide construct coverage. Lancaster et al. (2018) provides some description of how the team ensured items covered each construct and the targeted age range.

## Administration

The IYCD uses indirect assessment of the child. The parents/caregivers of the child complete the questionnaire (Gladstone et al., 2021; Lancaster et al., 2018).

## Data collection

The IYCD is administered in digital format using open data kit (ODK) software on a tablet based system (Gladstone et al., 2021).

## Data scoring

The IYCD provides detailed information about data processing (including scoring) and converting raw scores onto an ordinal/linear scale.

The main validation study uses IRT methodology to create scores for the IYCD (Gladstone et al., 2021). More specifically, a "generalized partial credit model (GPCM) using an empirical histogram prior to account for the non-normality in the ability (development) distribution was fitted to the data using the R package MIRT" (Gladstone et al., 2021, p. 7). Using these results, the effect of age on the latent scores was removed using the LMS (lambda, mu, sigma) method, to create 'development-for-age z-scores' (DAZ scores), which were then used in subsequent analyses (Gladstone et al., 2021).

## Item type(s)

The IYCD assessment contains dichotomous items only (pass/fail).

## Background/contextual information

The IYCD collects background/contextual information about the child (for example, height, weight, head circumference and mid-upper arm circumference), and the parents/caregivers and the home learning environment (wealth index, maternal education, home environment) (Gladstone et al., 2021).

## Length/time

The latest version of the IYCD contains 100 developmental items: 40 motor, 30 language and cognitive, 20 social-emotional and 10 behavioural[10] (Gladstone et al., 2021). Main study data collection allowed for different starting points dependent upon the child's age (Gladstone et al., 2021). It is unclear from the materials located for the review how long it would take to administer the IYCD.

## Available languages

The IYCD assessment was translated into multiple languages for the main validation study, which was conducted in Brazil, Malawi and Pakistan. It is not apparent without access to the IYCD site what languages are available to users. Guidelines for translation/adaptation of the assessment have not been found.

## Availability/cost

The IYCD is an open-source assessment. No access or use fees or royalties are involved.

## Training materials/requirements/capacity development

The IYCD has an established training protocol and requires enumerators to attend administration training. For the main validation study undertaken by the assessment authors (the only publicly available validation reporting that was found), training was provided to enumerators over a 2-3 day period prior to administering the assessments (Gladstone et al., 2021). The training materials used during these training workshops are also available via the IYCD website.

## Validation for use across different populations

Measurement invariance testing across countries/languages has not been reported for the IYCD.

The main validation study for the IYCD used data from Brazil, Malawi and Pakistan to examine the psychometric properties of the IYCD (Lancaster et al., 2018). Item functioning was examined by plotting proportion correct (for each item) for different age groups and examining the developmental trajectory, whilst comparing across the 3 countries (Gladstone et al., 2021). Items that were not able to reflect expected developmental trajectory, or substantially varied between countries were subjected to expert review (Gladstone et al., 2021). The expert review was able to determine whether the differences were due to measurement bias, or simply differences in ability between

---

[10] Behavioural items not used to compute scale scores due to lack of developmental progression on the social-emotional scale, but retained for importance reasons (Gladstone et al., 2021).

countries (Gladstone et al., 2021). Inter-rater and test-retest reliability analyses were also undertaken for the whole sample (due to sample sizes being too small to conduct separately), along with cognitive testing to ensure that items were well understood, resulting in 100 (out of 121) items being retained in the final validated assessment (Gladstone et al., 2021).

# International Early Learning and Child Well-being Study (IELS)

The International Early Learning and Child Well-being Study (IELS) measures 5-year-old children's key developmental and learning outcomes for the purpose of population level monitoring and cross-country comparisons (Australian Council for Educational Research, 2020). The assessment is administered by a consortium of organisations on behalf of the OECD. The first cycle of the assessment was undertaken in 2018 in 3 countries. A second cycle is currently under development, scheduled for administration in 8 to 10 countries in 2024. The following review is based on the first cycle.

### Purpose(s)

The purpose of IELS is to monitor populations. Data is designed to provide evidence for informing policy decisions.

### Age range

IELS is designed to assess 5 year olds enrolled in an early learning centre or school (Australian Council for Educational Research, 2020).

### Domain(s)

The IELS reports against multiple domains including multiple learning domains. Ten distinct scale scores are produced, as indicated by the numbers in brackets in Table 15.

**Table 15: Summary of the IELS domains**

| Domain | Sub-domains |
|---|---|
| Emergent literacy (1) | Listening comprehension<br>Phonological awareness<br>Vocabulary |
| Emergent numeracy (2) | Numbers and counting<br>Working with numbers<br>Measurement<br>Shape and space<br>Pattern |
| Self-regulation | Working memory (3)<br>Mental flexibility (4)<br>Inhibition (5) |
| Social and emotional skills | Emotion identification (6)<br>Emotion attribution (7)<br>Prosocial behaviour (8)<br>Disruptive behaviour (9)<br>Trust (10) |

**Learning progression**

There is no learning progression associated with the IELS.

**Assessment framework**

The IELS provides extensive details of assessment development, including theoretical underpinnings and how the items provide construct coverage. The Assessment Framework document (Phair, 2021) highlights the theoretical foundations of the domains being covered using key literature in the field and explains how the IELS is designed to provide construct coverage.

**Administration**

The IELS uses direct assessment of the child where trained assessors administer the assessment to groups of children. The IELS also collects background information about children through parent/caregiver and educator report questionnaires (Phair, 2021).

**Data collection**

The IELS direct assessment is administered in digital format using a tablet-based application. The study administrators show children how to navigate the tablet, help them attempt practice questions, and aim to ensure their well-being (Phair, 2021).

The parent/caregiver and educator questionnaires are administered using a combination of digital (online) and paper/pen formats (OECD, 2021).

### Data scoring

The IELS provides detailed information about data processing (including scoring) and converting raw scores onto an ordinal/linear scale.

IRT methodology is used for the analysis of data and production of scale scores and plausible values for population level reporting (OECD, 2021). Scale scores are reported for each sub-strand separately, meaning countries could use the data to identify specific areas of concern (for example, comparing relative results across different sub-domains).

### Item type(s)

The IELS assessment contains multiple item types. The item types are "stimulus, simple multiple choice, free form, short response items, re-usable templates, and complex dynamic items interactive on-screen elements" (OECD, 2021, p. 29).

### Background/contextual information

The IELS collects background/contextual information about the child, the parents/caregivers and the home learning environment, and about engagement and experience in ECE or school. The questionnaires consist of questions about the child's age, gender, immigrant background, their parent's SES, family composition, parental activities with children, parent-child relationship, home learning resources, age of entry, duration, frequency, continuity, and ECE type (Organisation for Economic Co-operation and Development, 2018).

### Length/time

The IELS contains a total of 157 items across the 4 overarching domains (OECD, 2021). The administration of the items for each domain is undertaken in 4 separate sessions of approximately 25-35 minutes, in 2 sessions per day, over a 2 day period (OECD, 2021). Parent/caregiver questionnaires (24 questions, 177 items) take approximately 30 minutes and educator questionnaires (12 questions, 86 items) take around 10 minutes per child (Australian Council for Educational Research, 2020).

### Available languages

The IELS has been conducted in several languages: English (England, USA), Estonian (Estonia) and Russian (Estonia) (OECD, 2021). Guidelines for the translation/adaptation of the assessment are not publicly available. Any country that requires the IELS materials to be adapted (including translation) receives a user manual, video tutorials and interactive online training about how to use the translation software (OECD, 2021). All translated materials are independently verified by the international study consortium's linguistic quality assurance partner (OECD, 2021).

### Availability/cost

The IELS assessment is not publicly available. The IELS is an OECD assessment program that requires a region to sign on and commit to substantial costs for participation.

### Training materials/requirements/capacity development

The IELS has an established training protocol and requires administrators to attend administration training on the purpose of the study and administration procedures (Phair, 2021).

Additionally, a school/centre coordinator, who oversees all activities and processes within their site, and international quality assurance managers, who conduct site visits to a sub-sample of centres/schools, receive training in the purpose of the study, the purpose of their role and their responsibilities and deliverables (OECD, 2021) (Phair, 2021).

### Validation for use across different populations

The IELS was validated using appropriate measurement invariance analysis methods and shown to be fully or partially invariant across countries/languages.

The IELS was developed under the guiding principle that it would be a valid and reliable assessment which is "comparable across countries, languages, cultural contexts and over time" (OECD, 2021, p. 17). National reports for each of the participating countries (England, Estonia and USA) were produced for the first iteration of the IELS (Australian Council for Educational Research, 2020). An international report provides aggregated results for all participating countries, including comparisons made between countries (Australian Council for Educational Research, 2020). The IELS technical report provides a more-in-depth account of the design, administration and analyses for the IELS (OECD, 2021). More specifically, details of differential item functioning (DIF) analyses are highlighted. If DIF was present, one of 3 actions was taken: (OECD, 2021, p. 129)

1) Note DIF and undertake no further treatment.

2) Remove the item from the assessment pool.

3) Keep the item in the assessment pool but free the parameters around the grouping variable indicating DIF. This essentially removes the item from the international pool and replicates the item over groups of interest.

The DIF analyses resulted in one item being removed (from the prosocial behaviour domain), and 9 items' parameters were freed across languages (8 from the literacy domain and one from the trust domain) (OECD, 2021). This meant that the numeracy, inhibition, mental flexibility, working memory, emotion identification, emotion attribution, prosocial behaviour[11] and disruptive behaviour were shown to be fully invariant across country and language, while the literacy and trust domains were therefore shown to be partially invariant across country and language (OECD, 2021).

## Measuring Early Learning Quality and Outcomes (MELQO)

---

[11] Note: one item was removed from this domain.

The Measuring Early Learning Quality and Outcomes (MELQO) assessments were designed for population modelling at scale, with a focus on measuring the development and learning of children and the quality of early learning environments from low-and-middle income countries. The assessment contains 2 modules: one for measuring early childhood development and learning (MODEL), and one for measuring the quality of early learning environments (MELE). The information provided in the review of MELQO will focus on the MODEL. The development of the assessment was undertaken by a consortium of researchers led by UNESCO, the World Bank, the Center for Universal Education at the Brookings Institution and UNICEF following a meeting in 2013. The assessments have been in use since 2015 across numerous countries.

## Purpose(s)

The purpose of MELQO is to monitor populations. Data is designed to be used to provide evidence for informing policy decisions.

## Age range

The MELQO is designed to assess children between the ages of 4 and 6 years.

## Domain(s)

The MELQO-MODEL is able to report against multiple domains including multiple learning domains.

**Table 16: Summary of the MELQO-MODEL domains**

| Domain | Description |
|---|---|
| Early mathematics | Number and operations, measurement and spatial relations |
| Early literacy | Motivation, expressive language, alphabet knowledge and receptive language |
| Executive function | Working memory, inhibition and fine motor |
| Social-emotional skills | Self-regulation, social understanding/pro-social behaviour, social competence and emotional wellbeing |

## Learning progression

There is no learning progression associated with the MELQO.

## Assessment framework

The MELQO provides extensive details of assessment development, including theoretical underpinnings and how the items provide construct coverage. The technical manual for the MODEL (UNESCO et al., 2017) describes the theoretical foundations of the domains being covered and how the items developed for the assessment cover the constructs.

## Administration

The MELQO-MODEL uses direct assessment of the child where trained assessors observe each child individually. The MELQO-MODEL also uses indirect assessment of the child where both the parents/caregivers of the child and the child's are interviewed.

## Data collection

The MELQO assessments provide the option to use either paper/pen or digital formats to collect data.

## Data scoring

The MELQO-MODEL provides detailed information about data processing (including scoring) and converting raw scores onto an ordinal/linear scale. Scoring guidelines are provided on enumerator booklets.

## Item type(s)

The MELQO-MODEL assessment contains multiple item types. For children's learning and development domains, a combination of dichotomous (for example, 0 – incorrect, 1 – correct), partial credit (for example, 0 – incorrect, 1 – self corrects (initially incorrect response, followed changing response to correct answer), 2 – correct) and Likert type (0 – never, 1 – sometime, 2 – often/always) items are used.

## Background/contextual information

The MELQO-MODEL collects background/contextual information about the child (for example, child health), the parents/caregivers and the home learning environment (for example, socio-economic conditions, parent education, household composition, home learning environment/parent involvement, neglect) and about engagement and experience in ECE or school (for example, participation in early learning).

## Length/time

The following times are estimated for administration:

- Direct child assessment – 25 minutes
- Teacher child report – 10-15 minutes
- Parent/caregiver report – 15-25 minutes.

## Available languages

The MELQO assessments are available in multiple languages: English, Spanish, French and Kiswahili. Guidelines are provided for the translation/adaptation of the assessment.

## Availability/cost

The MELQO assessments are open-source assessments. No access or use fees or royalties are involved.

### Training materials/requirements/capacity development

The MELQO assessments have established training protocols and require enumerators to attend administration training. Technical manuals (see UNESCO et al., 2017) outline how implementing agencies can appropriately implement the MELQO modules. The manual describes the necessary processes and procedures for planning, adapting, field testing, collecting data, analysing, and using results to inform policy decisions.

### Validation for use across different populations

Evidence of the MELQO-MODEL being invariant across countries/languages has not been found.

The MELQO assessments have been administered in many countries with several reports and publications being released which detail the psychometric properties of the assessments (for example, Tanzania - Raikes et al., 2019).

Using empirical analyses, a recent study showed that 20 (out of 82) caregiver report items and 87 (out of 146) child direct assessment items from the MELQO-MODEL were cross-culturally relevant for 4-6 year olds (Pushparatnam et al., 2021). Data was taken from 16,015 caregivers and 24,533 children from 12 countries, including Ethiopia, Kenya Laos, Lesotho, Madagascar, Mongolia, Nigeria, Pakistan, Sudan, Tanzania, a Central American country (anon), and a South American country (anon). Once data from all countries were harmonised, items were analysed for face validity, content validity, and a range of other psychometric properties using classical test theory (CTT) and item response theory (IRT) methods. A number of criteria were used to determine which items would satisfy the requirements for 'cross-cultural validation' for 4-6 year olds. Some examples of these were:

1) CTT difficulty between 0.1 and 0.9

2) Item-rest correlation >0.1 and item-total correlation >0.3

3) Increase Cronbach's Alpha coefficient with item inclusion

4) CFA standardised factor loadings >0.4

5) Positive regression coefficient for age

6) 1PL and 2PL IRT item difficulty between -3.0 and 3.0 (dichotomous items) and 2PL/GRM item discrimination >0.5.

Following these analyses, items were characterised as either Tier 1 (met all criteria), Tier 2 (did not meet one criterion), Tier 3 (did not meet 2 or 3 criteria) or Tier 4 (did not meet more than 3 criteria). Items in Tiers 1-3 were classified as adequate items for cross-cultural comparability. It is important to note though that in this study that only items that were deemed to be medium difficulty (between 40% and 60% correct responses) were retained. In addition, the authors noted that invariance testing has not been done, but this was a priority for future analyses using MELQO data.

# Individual citizen-led assessments and common assessments of the People's Action for Learning (PAL) Network

The People's Action for Learning (PAL) Network is a south-south partnership of 17 member organisations working to promote children's foundational learning across Africa, Asia, and America. Assessment tools (PAL Network, 2022c) have been developed by the PAL Network members to generate evidence on the learning levels of children. The PAL Network supports 2 types of assessment programs: citizen-led assessments (CLAs: PAL Network, 2022a) and common international assessments (PAL Network, 2022b).

The CLAs are listed in Table 17.

**Table 17: Individual Citizen-led Assessments**

| Name of the assessment | Age group | Country | Purpose | Most recent cycle |
|---|---|---|---|---|
| Annual Status of Education Report (ASER) | 5 to 16 years | India, Nepal, and Pakistan | Assess children's foundation literacy and numeracy skills for lifelong learning. | 2021 |
| ASER Young Children | 3 to 8 years | India | Assess enrolment and learning status of younger children. | 2018 |
| BEEKUNKO | 6 to 14 years | Mali | Assessing children in basic literacy and numeracy skills. | 2014 |
| Uwezo | 6 to 16 years | Kenya, Uganda and Tanzania | Same as above | 2021 |
| Jangadoo | 6 to 16 years | Senegal | Same as above | 2021 |
| Medición Independiente de Aprendizajes | 5 to 16 years | Mexico | Same as above | 2016 |
| Let's Evaluate, Assess, and Report Nigeria (LEARNigeria) | 5-15 years | Nigeria | Same as above | 2017/18 |
| IID/BRAC Survey | Primary school-aged | Bangladesh | Assess basic learning competency of children from primary classes: Reading competency of Bangla and English, and simple arithmetic competency. | 2015 |
| TPC Mozambique | Grade 2 | Mozambique | Assessing children in basic literacy and numeracy skills. | 2016 |
| VIdA Nicaragua | 5 to 13 years | Nicaragua | Basic reading, arithmetic and social interaction skills | 2017 |

The citizen-led assessments of children's learning are carried out annually to obtain basic data on literacy and numeracy levels in 3 continents and 13 countries around the

world: Kenya, Uganda, Tanzania, India, Bangladesh, Nepal, Pakistan, Mali, Senegal, Mozambique, Nigeria, Nicaragua and Mexico.

The goal is to collect data that is generalisable at national/regional/district levels, with the aim of stimulating citizens and governments to take action to improve learning levels of children in their communities.

Given that the basic CLA assessments are quite similar in scope this review only looks in-depth at the Annual Status of Education Report (ASER), as it is the most widely used assessment of this type.

The PAL Network's common international assessments are shown in Table 18.

**Table 18: Common PAL Network assessments**

| Name of the assessment | Age group | Country | Purpose | Most recent cycle |
|---|---|---|---|---|
| ICAN (International Common Assessment of Numeracy) | 5 to 16 years | 13 low and middle-income countries across Africa, America and Asia. | Generate internationally comparable data on numeracy skills in early primary grades. | 2019/ 2020 |
| ELANA (Early Language, Literacy and Numeracy Assessment) | 4-10 years | 12 countries (India, Pakistan, Bangladesh, Nepal, Kenya, Uganda, Tanzania, Mozambique, Mali, Senegal, Nicaragua and Mexico) | A common literacy and numeracy assessment for early years. | 2022-24 (Field trials) |

The common assessments (ICAN and ELANA) are designed to be much more detailed and in-depth assessments of literacy and numeracy than the CLAs. Therefore, both assessments are reviewed.

## Annual Status of Education Report (ASER)

The ASER assessment aims to assess children's basic foundation literacy and numeracy skills for lifelong learning, whether they attend school or not. ASER was developed by Pratham in India with the first survey conducted across the country in 2005.

*Purpose*

The purpose of ASER is to monitor populations. Data is designed to be used to provide evidence for informing policy decisions.

*Age range*

ASER is designed to assess children from the age of 5 to 16 years.

*Domains*

The ASER is able to report against multiple domains including multiple learning domains. The assessment was initially designed to assess children's foundational literacy and numeracy skills. More recently ASER has included assessment of cognitive, and social and emotional development.

*Learning progression*

There is no learning progression associated with the ASER.

*Assessment framework*

The ASER provides extensive details of assessment development, including theoretical underpinnings and how the items provide construct coverage. The ASER India report (ASER Centre, 2022) describes the theoretical foundations of the domains being covered and how the items developed for the assessment cover the constructs.

*Administration*

The ASER uses direct assessment of the child in the household setting where trained assessors individually administer the assessment.

*Data collection*

The ASER is administered using paper/pen.

*Data scoring*

The ASER provides basic scoring advice using raw scores. This information is contained within scoring manuals.

*Item types*

The ASER assessment contains multiple item types. The reading assessment form predominantly consists of items where the child is asked to identify letters or read passages aloud. For numeracy, items are either open response (for example, solve a division problem) or require identifying numbers.

*Background/ contextual information*

The ASER collects background/contextual information about the child, the parents/caregivers and the home learning environment (for example, type of house, number of members, family resources, parental education).

*Length/time*

The latest (2023) ASER child assessment contained 31 items and is typically completed in approximately 10 to 15 minutes.

*Available languages*

The ASER assessment is available in multiple languages. In India alone, the 2022 assessment is available in 19 languages. Guidelines are provided for the translation/adaptation of the assessment.

As many of the PAL Network countries have adjusted the ASER assessment form and customised their own CLAs, these are available in many different languages.

*Availability/cost*

The ASER assessment is free to use for all PAL Network countries.

*Training materials/ requirements/ capacity development*

The ASER has an established training protocol and requires enumerators to attend administration training. A 2-stage approach to training is used where national training is conducted first, then those trained at that stage administer training at the district level. Training materials and manuals are provided to enumerators to support their administration of the assessment.

*Validation for use across different populations*

The ASER has been validated in individual populations. Table 17 describes the countries in which ASER and assessments very similar to ASER have been used. Various psychometric analyses have been done for these assessments to evaluate their validity and reliability. However, measurement invariance analyses were not able to be located for ASER.

## The International Common Assessment of Numeracy (ICAN) and the Early Language, Literacy and Numeracy Assessment (ELANA)

The International Common Assessment of Numeracy (ICAN) and the Early Language, Literacy and Numeracy Assessment (ELANA) were designed to be used for population monitoring across countries and jurisdictions, generating internationally comparable data on academic skills in school. The ICAN measures numeracy only, whilst the ELANA measures both language and literacy, and numeracy.

ICAN was developed through a collaborative effort among PAL Network member organisations in 13 low- and middle-income countries across Africa, America and Asia. It was administered in 2019/2020.

ELANA was developed through a collaborative effort among PAL Network member organisations in 12 low- and middle-income countries across Africa, America and Asia. Multiple field trials (with different purposes) have been administered since 2022..

*Purpose*

The purpose of the ICAN and the ELANA is to monitor populations. Data is used to inform policy decisions. The ICAN measures numeracy only, whilst the ELANA measures both numeracy and language and literacy.

*Age range*

The ICAN is designed to assess 5- to 16 year olds, while the ELANA is designed to assess 4 to 10 year olds.

*Domains*

The ICAN and ELANA are able to report against multiple learning domains.

The ICAN includes items from the numeracy sub-domains of number knowledge, geometry, measurement and data display. The ELANA extends the work done in the ICAN to include more numeracy items and a significant number of language and literacy items which cover the sub-domains of listening comprehension, decoding and reading comprehension.

*Learning progression*

There is no learning progression associated with the ICAN or ELANA.

*Assessment framework*

The ICAN and ELANA provide limited details of assessment development, including theoretical underpinnings and how the items provide construct coverage. An assessment framework which includes details of construct coverage is available for the ICAN (PAL Network, 2020a). A policy linking report is also available (PAL Network, 2020b) which describes how the ICAN has been aligned to the Global Proficiency Framework (GPF) and Sustainable Development Goal (SDG) 4.1.1, which define minimum proficiency levels that learners are expected to demonstrate more generally. The ELANA is still in development and similar outputs are expected to be produced following the main study.

*Administration*

The ICAN and the ELANA use direct assessment of the child where trained assessors individually administer the assessment.

*Data collection*

The ICAN is administered in digital format using a tablet application. The ELANA will also be administered using a tablet application.

*Data scoring*

The ICAN and the ELANA provide detailed information about data processing (including scoring) and converting raw scores onto an ordinal/linear scale.

For ICAN, manuals are provided to trained enumerators which includes scoring forms. For ELANA, automated scoring is done through the adaptive test via the tablet. Post-hoc data analysis is conducted to generate ability levels for all children included in the studies.

*Item types*

The ICAN and ELANA assessments contain multiple item types. For ICAN, there are correct/incorrect, multiple choice and counting item types.  For ELANA, additional

items such as multiple choice and open response utilising drag and drop, verbal, gesturing and select functions via the tablet are used.

### Background/contextual information

The ICAN and ELANA collect background/contextual information about the child (for example, gender, age, school enrolment status), and the parents/caregivers and the home learning environment (for example, parents' education, some basic household information, and information about sampled communities).

### Length/time

The ICAN contains 26 items and the child assessments are typically completed in approximately 15 minutes. Progressive assessment administration procedures are used in which only children who can do easier items are given more advanced ones. On average when the household survey is included, it takes 20-30 minutes per household to complete.

The ELANA will employ an adaptive test design, in which children will undertake items that are progressively more or less challenging, based on performance as they progress through the assessment. The literacy and numeracy components include 45 items (on average, depending on form assignment) and 30 items, respectively.

### Available languages

The ICAN is available in multiple languages: Kiswahili, Portuguese, English, Kamba, French, Wolof, Spanish, Hindi, Bangla, Nepali and Urdu. Guidelines are provided for the translation/adaptation of the assessment.

The ELANA is available in multiple languages: French, Hindi, Nepali, Urdu, English, Spanish, Kiswahili, Bangla and Portuguese.

### Availability/cost

The ICAN and ELANA assessments are free to use for all PAL Network countries.

### Training materials/ requirements/ capacity development

The ICAN and ELANA have established training protocols and require enumerators to attend administration training. Manuals and training are available for PAL Network members, supported by the PAL Network secretariat.

### Validation for use across different populations

The ICAN and ELANA have been validated using appropriate measurement invariance analysis methods and shown to be either fully or partially invariant across countries/languages.

Table 18 describes the countries in which ICAN and ELANA have been used. Validity and reliability analyses for both ICAN and ELANA, including differential item

functioning (DIF) for gender and language/country, has been undertaken. However, results of these analyses have not been made publicly available.

# Appendix E: Summary of assessments against key reporting components

| Key reporting components | | ASQ | CREDI | ECDAS | ECDI2030 | EDI | EGRA and EGMA | eHCI | EAP-ECDS | GSED | IDELA | IYCD | IELS | MELQO | PAL In-country citizen-led assessments (e.g., ASER) | PAL Common assessments (e.g., ICAN and ELANA) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Purpose(s) (as stated by assessment publishers) | Population monitoring | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| | Longitudinal cohort studies | No | No | Yes | No | No | No | Yes | No | No | No | No | No | No | No | No |
| | Impact/ program evaluation | No | Yes | Yes | No | No | No | Yes | No | Yes | Yes | No | No | No | No | No |
| | Informing policy decisions | No | No | No | Yes | Yes | Yes | Yes | Yes | Yes | Yes | No | Yes | Yes | Yes | Yes |
| | Individual diagnosis/ screening | Yes | No | No | No | Yes | Yes | No | No | No | No | No | No | No | No | No |
| Age range (in years) | | 0 - 5.5 | 0 - 3 | 2 - 6 | 2 - 6 | 3.5 - 6.5 | 6 - 8* | 3 - 5 | 3 - 5 | 0 - 3 | 3.5 - 6 | 0 - 3 | 5 | 4 - 6 | 5 - 16 | 4 - 16 |
| Domain(s) | | High | High | Low | Low | High | High | Medium | High | High | High | High | High | High | High | High |
| Learning progression? | | No | No | No | No | No | No | No | No | No | No | No | No | No | No | No |
| Assessment Framework | | High | High | Low | High | High | High | Low | High | Low | High | Low | High | High | High | Low |
| Administration format | Direct child assessment | No | No | Yes | No | No | Yes | No | Yes | Yes | Yes | No | Yes | Yes | Yes | Yes |
| | Indirect caregiver/ educator report | Yes | Yes | Yes | Yes | Yes | No | Yes | No | Yes | No | Yes | No | Yes | No | No |
| Data collection format | Paper/pen | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | No | Yes | Yes | Yes | No |
| | Digital | Yes | Yes | Yes | No | Yes | Yes | No | No | Yes | Yes | Yes | Yes | Yes | No | Yes |
| Data scoring | | Medium | High | Medium | Medium | Medium | Low | Low | High | High | Low | High | High | High | Medium | High |
| Item type(s) | | Singular | Singular | Multiple | Singular | Multiple | Multiple | Singular | Singular | Singular | Multiple | Singular | Multiple | Multiple | Multiple | Multiple |

| | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Background/ contextual information collected? | Child | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| | Parent(s)/ caregiver(s) and/ or home learning environment | No | No | Yes | No | No | No | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| | Engagement and/ or experience in ECE/ school | No | No | Yes | No | Yes | Yes | Yes | Yes | ? | Yes | ? | Yes | Yes | Yes | Yes |
| Administration time | | Medium | Short (short form), Medium (long form) | ? | Short | Medium | Medium | Short | Long | Short (short form), Medium (long form) | Long | ? | Long | Medium (direct assessment) | Medium | Medium |
| Available languages/ translation and adaption guide | | High | High | Low$ | High | High | Medium | Medium | Medium | High | High | Low$ | Low$ | High | High& | High& |
| Availability/ cost | | Low | High | Low^ | High | Low | High | High | High | High | High | Low | Low | High | Low | High |
| Training | | Medium | Medium | High | High | High | High | Medium | High | High | High | High | High | High | High | High |
| Validation for use across different populations | | Low | Low | Low | Low | High | Low | Low | Low | Low# | Low | Low | High | Low# | Low | High |

# Glossary

| Term | Definition |
|------|------------|
| Administration materials | Manuals relating to the administration of the tests and contextual instruments (otherwise known as field guidelines or field operations manuals) as well as important supporting documents such as student attendance forms (sometimes referred to as student tracking forms). |
| Assessment materials | Test forms, questionnaires, interviews, observation forms |
| Benchmark | A standard set as part of the assessment program (for example, performance levels) or from outside the assessment program (for example, SDG 4 learning outcome targets) against which to assess performance on the test. |
| Bias | A systematic distortion of results that is based on factors unrelated to ability. |
| Citizen-led assessments | Assessments typically run by non-government organisations and conducted in households rather than educational institutes. |
| Cluster (test and questionnaire design) | A small group of test/questionnaire items that are grouped together and treated as a block during test construction. |
| Cluster (sampling) | A sampling technique used when 'natural' but relatively homogeneous (similar) groupings are evident in a population of interest. |
| Cognitive testing | Assessment that collects information about what the participant knows, understands and can do in a particular learning domain, or domains. |
| Cognitive skills/abilities | Skills, sometimes called 'processes', 'cognitive domains' or 'aspects', are the ways of thinking, or intellectual approaches, that develop as individuals become increasingly proficient in a learning domain. |
| Contextual information | Data collected through questionnaires/interviews/observations on a range of topics that are useful to policy and in understanding the test results in context. |
| Contextual instruments | A set of items used to collect information about the personal characteristics, background, attitudes and values of participants in their contexts (for example, home, classroom, school). |
| Correlation | Indication of a relationship between 2 phenomena/variables. |
| Cross-sectional | An assessment where data are collected from individuals at a single point in time. While some assessment designs may collect data from, for example, a student cohort as they progress through school, that data is not tied to specific individuals. |
| Cycle (assessment) | All activities related to a single main survey assessment administration within a program with repeated administrations designed to assess learning over time. |
| Data cleaning | The process of identifying discrepancies and errors in the database and correcting or removing them. This process includes verification and validation of the data. |
| Data collection | The process of gathering data—in the case of large-scale assessments, the process of administering tests and contextual instruments to participants. |

| Differential item functioning | When the probability of answering an item correctly depends on the sub-population the respondent belongs to rather than her/his ability level. |
|---|---|
| Field trial | Administration of items under test conditions, used to test the items' validity and the administration procedures. Occurs before the main survey and uses a sample as similar as possible to the target population. |
| Fit statistics | Indicators of model fit for both person data (that is, a participant's response pattern) and item data (that is, the pattern of responses to an item). |
| Form (test and questionnaire) | The group of test and/or questionnaire items that is presented to each participant. There may be just one group of items for all participants (i.e one form), or participants may receive one of several different groups of items (that is, one of several forms). |
| Free-form item | An item with unrestricted response format. |
| Index (pl. indices) | A scaled indicator of a measure that is composed of several values or other measures. For example, a socioeconomic index might be composed of income, health factors, education level and other components. |
| Internal consistency | Internal consistency as a type of reliability estimate assumes that the test is unidimensional, or measuring a single construct. |
| Items | The questions or tasks used in an assessment. |
| Item difficulty | The difficulty of an item as hypothesised by test developers and confirmed by statistics. |
| Item discrimination | The ability for an item to group participants of different abilities. For example, participants who perform well overall on a test should also have a high chance of answering a particular item correctly. |
| Item pool | The total set of cognitive or contextual items written for an assessment. |
| Learning domain | The area of learning that is the focus of an assessment. This may be a curriculum area (for example, mathematics or science), or more generic areas of learning (for example, reading, writing or problem-solving). |
| Mean | The arithmetic average. |
| Metadata | A record of all the information related to an item, including the item code, the learning domain and skills the item is assessing, the estimated difficulty level and the item descriptor. |
| Model fit | How well the overall distribution of the observed data (the data collected from participants) reflects the expected distribution according to the measurement model being used to analyse the data. |
| Multiple-choice item | An item that presents several options as answers, from which the participant selects one. |
| Parameter | A characteristic that defines a population, such as its variability or its average. A characteristic that defines a sample is called a statistic. |
| Plausible values | A set of values drawn randomly from the marginal posterior distribution of scores that is used to represent performance in large-scale, sample-based assessments. |
| Policy Linking | A methodology that links different assessments to a common scale by aligning items with the GPF. This allows for cross national reporting against the SDGs. |
| Population | See 'target population'. |
| Questionnaire | See 'contextual instruments'. |

| Raw data | Data that comes directly from the source of the data and have not been processed in any way. These could be student responses on a test such as actual choices in a multiple-choice item, or actual words written in a short- response type of test. |
|---|---|
| Reliability | The consistency and accuracy of test and contextual measures and results over replications of the testing procedure (American Educational Research Association et al., 2014). |
| Reporting variable | Contextual factors that have been identified as important in accounting for the variance in performance across the target population, with the aim of discussing the outcomes in results reports. An example of a reporting variable could be gender, geographic location, or socioeconomic status. |
| Scale | A numeric or substantive description of progress in learning. |
| Scoring | The process of classifying responses and allocating (usually numerical) codes to represent the various categories of response. |
| Scoring guide | The description of the scoring categories that are used to categorise and score a participant's answer. |
| Skills | The ways of thinking, or intellectual approaches, that develop as individuals become increasingly proficient in a learning domain (sometimes called 'processes', 'cognitive domains' or 'aspects'. |
| Stimulus material | The prompt or context on which one or more items is based. For example, in a reading test, the stimulus is often a prose text made up of one or more paragraphs. In a mathematics test, the stimulus may be a diagram or a graph. |
| Sub-population | Groups of people within the larger population who are separated into mutually exclusive categories according to a particular characteristic. |
| Sub-scale | A numeric or substantive description of progress in learning within a particular sub-domain or strand. |
| Target population | A particular group of people that the assessment is attempting to describe or measure outcomes for. For example, an assessment may aim to measure reading ability of Grade 6 students in government schools in a particular region. This group of people is referred to as the target population. |
| Test targeting | In the context of test design, test targeting refers to the process in which item difficulties are matched with the ability levels of the target population. |
| Trends | The change in assessment results over time. |
| Validity | The extent to which the assessment instruments measure what they claim to be measuring for a specified population, and the extent to which interpretations made from the data analysis are correct and appropriate for the proposed use of the data (American Educational Research Association et al., 2014). |
| Variance | A numerical measure of how the data values are dispersed around the mean. |