

POLICY NOTE 1

# Measurement of young children's learning for program evaluation

## WHY MEASURE?

The global commitment to early learning has been expressed in the United Nations (UN) Sustainable Development Goals Agenda (SDG) (United Nations, 2016) and access to support for early learning is considered a human right for all children, whether provided by the family, community, or institutional programs (UNESCO, 2013). Inadequate cognitive stimulation has been identified as one of the key psychosocial risk factors associated with poor child development – a factor that is modifiable, with the right interventions (Walker et al., 2007). Thus, insights into how early learning supports may be delivered effectively in various contexts are essential.

To explore the effectiveness of different learning supports requires the measurement of outcomes. Otherwise, supports may be perceived as beneficial by various stakeholders, yet have no demonstrable impact on learning outcomes and may therefore be ineffective investments.

This policy note summarises findings regarding the measurement of learning outcomes in Early Childhood Education and Care (ECEC) based on a recent scoping review. The review aims to identify ECEC interventions in economically developing countries between 1998 and 2017 which have been effective in improving children's learning in the years before school (Jackson et al., 2019). For a summary evidence map visit - using the Firefox browser - <http://egmopenaccess.3ieimpact.org/evidence-maps/improving-young-childrens-learning-economically-developing-countries-scoping-review>.

## KEY MESSAGES

- ▶ Where children's learning before school is measured in low and middle income countries, approaches are highly variable and frequently draw on measures designed for and validated in economically developed countries.
- ▶ Many ECEC interventions do not provide evidence as to their impact on children's learning outcomes due to a lack of measurement.
- ▶ Early childhood education and care interventions and programs need to administer robust, cost-effective, fit-for-purpose measures of learning, validated in economically developing contexts, to obtain evidence as to their effectiveness.
- ▶ Measurement needs to occur at least at the beginning and the end of an intervention to assess impact.
- ▶ Examples of measures designed for low and middle income countries which are cost-effective and relatively easy to use include the International Development and Early Learning Assessment (IDELA), the Early Human Capability Index (EHCI), the Early Childhood Development Index (ECDI) (UNICEF, 2017a) and the Measuring Early Learning Quality and Outcomes (MELQO, also Measure of Development and Early Learning (MODEL) (UNESCO, UNICEF, Brookings Institution, & World Bank, 2017).

In line with the SDGs (Indicator 4.2.1), learning in the review is defined quite broadly across various domains including cognitive, socio-emotional, language and motor development.

### The need for measurement

While measurement is also needed for screening, diagnostic, teaching and learning as well as system-level monitoring purposes, it is an essential element of program evaluation. To build an evidence base of effective interventions requires information as to the extent to which the intended effects on outcomes have been achieved.

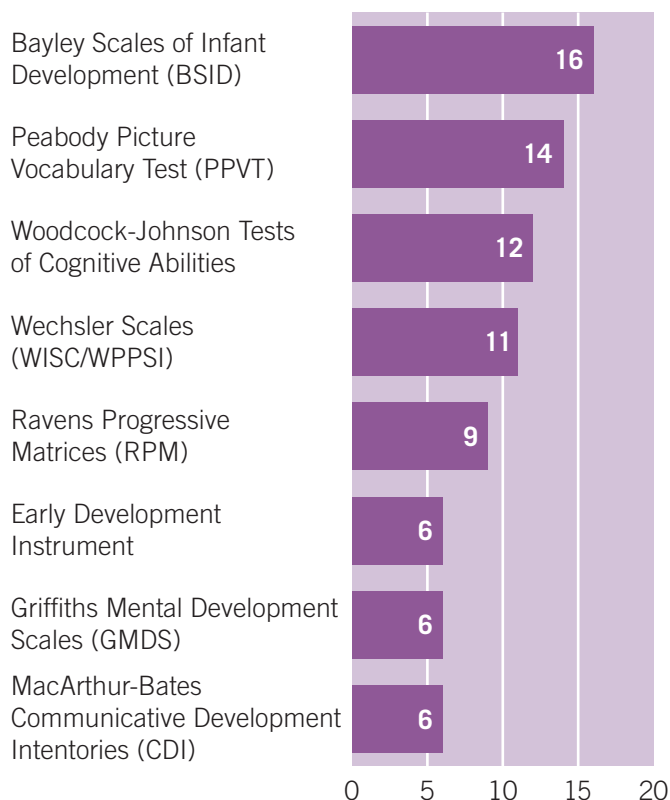
Yet, one of the most striking findings of the review is that many interventions fail to measure learning outcomes. Of the 772 studies initially identified, the most common reason for exclusion (n=145) is that they have not measured learning outcomes.

### Measures used

The studies ultimately included in the review (n=109) employ a total of 46 different measures of learning outcomes. Figure 1 illustrates the variety of measures used to assess children’s learning outcomes in at least three studies in the review. Most of the tools involve measuring cognitive abilities as do the three most commonly used ones (see Table 1). Also, many of the instruments are used only partially – by employing selected tasks or sub-scales – or adapted or translated for the language and context in which they are applied.

Where multiple measures are used, analysis is typically presented for each of the instruments. Therefore, the variation in measures used is even greater than Figure 1 suggests.

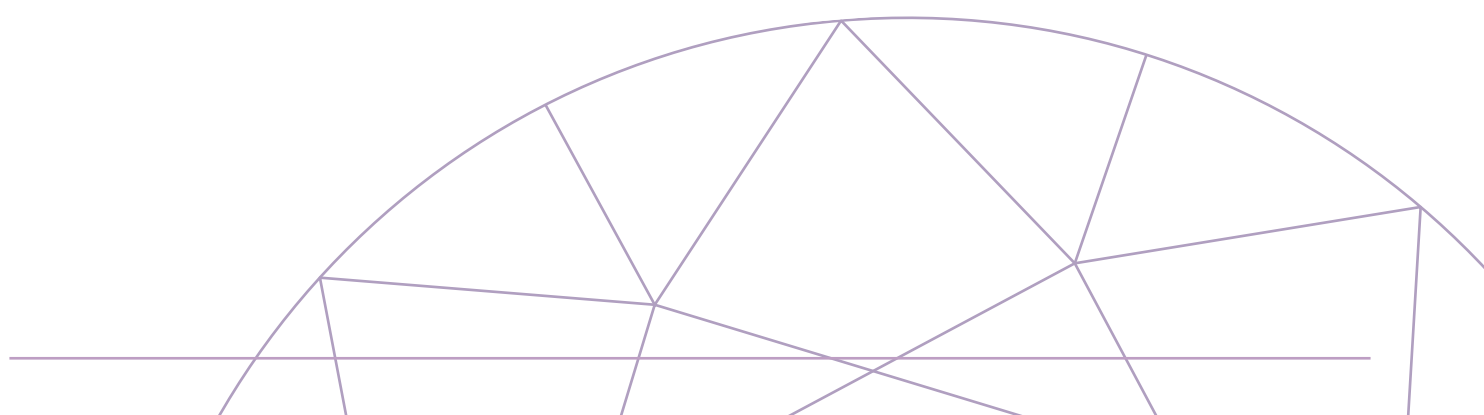
**Figure 1** Measures of learning outcomes used in three or more studies



Source: Jackson et al., 2019

**Table 1** Characteristics of commonly used instruments

	Bayley Scales of Infant Development (BSID)	Peabody Picture Vocabulary Test (PPVT)	Woodcock-Johnson Tests of Cognitive Abilities
<b>Age (years)</b>	0.1–3.5	2.5–90	2–90
<b>Domains</b>	Cognitive and motor skills	Language	Cognitive
<b>Method</b>	Direct child assessment, caregiver/teacher report	Direct child assessment	Direct child assessment
<b>Admin training</b>	Specialist	Moderate	Specialist
<b>Regions used</b>	Africa, Asia, Central and South America	Africa, Asia, Central and South America	Africa, Central and South America



Considering the importance of the measurement tool to the likely outcomes of a study, it is surprising that relatively few studies provide a clear rationale for their choice of instrument. Where studies do give reasons, these include:

- ▶ The instrument has either been validated in their context (Fernald, Weber, Galasso, & Ratsifandrihamanana, 2011; Nair et al., 2009; Powell, Baker-Henningham, Walker, Gernay, & Grantham-McGregor, 2004; Tessier et al., 2009) or in a similar context (Walker, Grantham-McGregor, Powell, & Chang, 2000).
- ▶ The instrument predicts later learning (Rolla San Francisco, Arias, & Villers, 2005).
- ▶ Children enjoy the assessment (American Institutes for Research, 2013).
- ▶ Age of the children involved. For example, the Bayley Scales of Infant Development (BSID) is designed for children aged 2–42 months requiring administration by parents or caregivers while some of the longitudinal studies use school assessments or measures designed for any age (e.g. Peabody Picture Vocabulary Test, PPVT, and Raven’s Progressive Matrices).

The studies in the review illustrate that the reasons for the variability in measures used include pragmatic, language or cultural considerations.

## REASONS FOR VARIABILITY IN INSTRUMENTS USED

### Pragmatic

In **Colombia**, Bernal and Fernández (2013) reported that the high costs of standardised tests led them to use parent-reported child outcomes for most of their large sample. A study in KwaZulu-Natal, **South Africa**, reported that developmental assessments were confusing to implement for ECEC staff with limited training (Hodgson, Papatheodorou, & James, 2014).

### Language

**Latin America** had by far the least custom-designed instruments with only one study taking this approach. In this region, the Woodcock-Muñoz (Spanish adaptation of the Woodcock–Johnson Tests of Cognitive Abilities) was administered far more than any other instrument. While this may reflect the concentration of studies in this region around one long-standing project, it may also illustrate the benefits (e.g. comparability) of having a reliable, valid measure of learning available in the local language when evaluating interventions.

### Cultural

In **Uganda**, expressive language items were removed from the Bayley Scales of Infant Development (BSID) because children were too shy to speak to researchers.

**Table 2** Ideal characteristics of an Early Childhood Development (ECD) assessment

<b>Ideal 1</b>	The test score represents the child’s true ability.
<b>Ideal 2</b>	The test is appropriate, interpretable, and has high reliability and validity in all contexts and cultures.
<b>Ideal 3</b>	The test shows variance in scores at all ages and ability levels.
<b>Ideal 4</b>	The test is easy to administer.
<b>Ideal 5</b>	The test can be administered quickly and at low cost.
<b>Ideal 6</b>	The test provides information on all developmental domains.
<b>Ideal 7</b>	The test score is relevant to a child’s practical function in daily life and therefore relevant to policy and program design.
<b>Ideal 8</b>	The test is a good indicator of future success.
<b>Ideal 9</b>	The brain systems and neural mechanisms underlying test performance are well-understood.
<b>Ideal 10</b>	The impact of health, nutrition, and environmental factors on the test score is well-understood.

Source: Fernald et al., 2017, p. 63

Ultimately, the measures used in the interventions cover a wide range of child development constructs, using instruments originating from clinical (e.g. health centres/hospitals) as well as educational settings. Each of them could be argued to measure an important aspect of children’s learning. However, given that the measured constructs, abilities or skills are likely to differ in complexity, it is also likely that the measures vary considerably in their propensity to show improvement as a result of the interventions.

In summary, the instruments used to measure children’s learning in economically developing contexts vary widely and draw frequently on measures which have been validated in the Global North.

## Implications

The lack of and diversity in outcome measures by Early Childhood Education and Care (ECEC) interventions hinders efforts to build an evidence base regarding their effectiveness.

To address this issue, policy- and decision-makers need to increase the uptake of robust, cost-effective, fit-for-purpose measures of children’s learning that have been validated in economically developing contexts from the outset of any intervention.

To this end, ‘A toolkit for measuring early childhood development in low-and middle-income countries’ has been developed by the World Bank (Fernald et al., 2017) as a helpful resource to guide the selection of assessments in ECEC. The toolkit not only proposes ten ideal characteristics of an early childhood development assessment (see Table 2) but also describes existing

measures, for the evaluation of programs or interventions as well as for system-level monitoring and the screening of individuals. Furthermore, it provides a step-by-step approach for the adaptation of existing instruments and the development of new instruments.

The highlight of the toolkit is a separate measurement inventory which covers 147 instruments for children from birth to eight years. The easily searchable inventory contains information regarding domains assessed, age range for which the tool is appropriate, method of administration, assessment purpose, origin and locations of use, logistics and cost. <https://openknowledge.worldbank.org/handle/10986/29000>

The toolkit also covers new measures of early childhood learning outcomes which have been designed specifically for use in low and middle income country contexts. Examples (see Table 3) include the International Development and Early Learning Assessment (IDELA), the Early Human Capability Index (EHCI), the Early Childhood Development Index (ECDI) (UNICEF, 2017a) and the Measuring Early Learning Quality and Outcomes (MELQO, also Measure of Development and Early Learning, MODEL) (UNESCO, UNICEF, Brookings Institution, & World Bank, 2017). These measures have been validated in diverse contexts and tend to be cheaper and simpler to use than more complex standardised instruments.

Findings from the scoping review suggest that the use of these measures is still emerging, with only two studies using the ECDI, and none using the MELQO-MODEL. However, this is likely to be a result of the time required from the implementation of interventions, the publication of studies to their inclusion in the review.

**Table 3** International measures for evaluating program and interventions, specifically in economically developing countries

	The International Development and Early Learning Assessment (IDELA)	The Early Human Capability Index (EHCI)	Early Childhood Development Index (ECDI)	Measuring Early Learning Quality and Outcomes (MELQO)
Age (years)	3.5–6	3–5	3–5	4–6
Domains	Cognitive, language, motor skills, social/emotional, executive functioning, (pre-) academic	Cognitive, language, social/emotional, (pre-) academic, approaches to learning	Language, motor skills, social/emotional, (pre-) academic	Social/emotional, executive functioning, (pre-) academic, approaches to learning
Method	Direct child assessment, caregiver/teacher report	Caregiver/teacher report	Caregiver/teacher report	Direct child assessment, caregiver/teacher report
Admin training	Moderate	Minimal	Minimal	Moderate
Regions used	Africa, Asia	Asia, Central and South America	Africa, Asia, Central and South America	Africa, Asia

Source: Fernald, et al., 2017

In summary, this policy note illustrates:

- ▶ The need to measure outcomes when evaluating the effectiveness of interventions in Early Childhood Education and Care (ECEC) – at least at the beginning and the end of an intervention.
- ▶ The measures used in the evaluation of interventions in ECEC over the last 20 years.
- ▶ ECEC outcome measures specifically designed for use in low and middle income countries.

---

## FURTHER READING

Jackson, J., Ahmed, S. K., Carslake, T., & Lietz, P. (2019). *Improving young children's learning in economically developing countries: A scoping review*. Camberwell, Vic.: Australian Council for Educational Research (ACER), Centre for Global Education Monitoring. [https://research.acer.edu.au/monitoring\\_learning/38/](https://research.acer.edu.au/monitoring_learning/38/)

Fernald, L. C., Prado, E., Kariger, P., & Raikes, A. (2017). *A toolkit for measuring early childhood development in low and middle-income countries*. Washington, D.C.: International Bank for Reconstruction and Development/The World Bank.



Research team: Jennifer Jackson, Syeda Kashfee Ahmed, Toby Carslake, Petra Lietz

The review team is grateful for the support from ACER colleagues, the ACER Cunningham Library and the many individuals who provided valuable advice on this review through external stakeholder consultations.

Photo credit: Hector Conesa

This document is an output from research jointly funded by the Australian Council for Educational Research (ACER) and the Australian Department of Foreign Affairs and Trade (DFAT) as part of ACER's Global Education Monitoring Centre (GEM). The views and opinions expressed in this document are those of the authors and do not necessarily reflect the views of DFAT or the Australian Government.

[dfat.gov.au](http://dfat.gov.au) • [acer.org/au/gem](http://acer.org/au/gem)

---