Early Learning and Development Standards Assessment Technical Report: Evidence for Validity and Reliability

Developed on behalf of Education Review Office by Kenji Kitamura and Shyam Prasad Acharya.

To cite this report: Kitamura, K. & Acharya, S. P. (2022). Early Learning and Development Standards Assessment Technical Report: Evidence for Validity and Reliability. Bhaktapur, Nepal: Education Review Office.

Draft

Table of Contents

1. Introduction *1*

1.1. Background of the ELDS **1**

1.2. Context of the ELDS assessment in Nepal **1**

1.3. Objectives of report **3**

1.4. Structure of report **3**

2. Validity and reliability of ELDS assessment score *3*

2.1. Method: The Process Model **3**

2.2. Phase I: Assessment context **4**

2.3. Phase II: Assessment framework (Domain specification) **5**

2.4. Phase III: Instrument design **12**

2.5. Phase IV: Validity and reliability of ELDS scores **12**

2.5.1. Unified validation Plan 12

2.5.2. Sample for the EDLS assessment data 14

3. Evidence of Content based validity *16*

4. Evidence of item quality (Item analysis) *22*

4.1. Analytic plan **22**

4.2. Results **23**

5. Evidence of validity of theorized internal structure *25*

5.1. Analytic plan **25**

5.2. Results **27**

5.2.1. Exploratory factor analysis within domains (RQ1) 27

5.2.2. Exploratory factor analysis across domains (RQ2, RQ3) 29

5.2.3. Confirmatory factor analysis 31

6. Evidence of score reliability (Internal consistency) *34*

7. Evidence of group difference *35*

8. Discussion and recommendation *36*

Appendix A. The ELDS Assessment Tool *39*

Appendix B. ECD Expert Panel Participants for the Content Validation *48*

Appendix C. Sample size calculation for the confirmatory sample *49*

References *50*

1. Introduction

1.1. Background of the ELDS

Under international momentum to support early childhood development (ECD), UNICEF initiated work on the Early Learning Development Standards (ELDS) in 2002, through a project entitled *Going Global with indicators of child well-being: using a standards approach* (UNICEF, 2017). The ELDS reflect what children should know and be able to do with respect to various domains of child development, including their physical, cognitive, social-emotional and language development (UNICEF, 2017).

Adoption of standards has been widespread since then, with over 30 countries supported by UNICEF to develop their national standards. While the adoption of standards was lauded, their ability to influence policy has been limited by a paucity of data on whether they are being achieved: while many countries know where they want children to be in terms of learning and development outcomes, they lack the data to show the status. In 2016 a key recommendation stemming from an evaluation of the ELDS program was to develop operational tools, which draw on ELDS and facilitate their better use in classrooms and policy formulation (UNICEF, 2017).

In Nepal, ELDS was adopted in 2008 and validation was conducted from 2009 to 2011 by the national working group (UNICEF, 2017). The standards focus on children aged 48- 60 months, corresponding to the target age of the preprimary education in Nepal. The standards were developed around the six domains which stem from the values emphasized by the country team: Physical, Social, Emotional, Cognitive, Language, and Cultural development. Nepal was able to develop comprehensive standards within the country since 2011, it was necessary to utilize effectively. (Shrestha et al., 2017).

1.2. Context of the ELDS assessment in Nepal

As Nepal has increased the investment in early childhood programs and policies, so has the interest in assessing children's learning and development outcomes. Also, the global goal for ECD (i.e., SDG target 4.2) urges countries to monitor the progress of the indicator 'proportion of children under 5 years of age who are developmentally on track in health, learning and psychosocial well-being, by sex' (United Nations, 2020). This requires an instrument to measure whether and to what extent children in Nepal are 'developmentally on track'.

The Education Review Office (ERO) under the Ministry of Education, Science and Technology (MoEST) of Nepal, took leadership to develop the tool to assess the status of pre-primary children based on the ELDS in 2017 with support from UNICEF Nepal. There were several rounds of consultations with key personnel involved in ECD planning and program implementation. Various national and international literature sources on assessment of early learning and development were reviewed to develop the ELDS assessment framework (Shrestha et al., 2017). The framework articulates domains of child's learning and development as well as more specific standards from subdomain through tasks relevant to the domain. Based on the framework, the ELDS assessment tool was prepared by experts and teachers in a workshop and had 44 items with some sub-tasks within items at the beginning. Later, it was piloted with over 500 children and analyzed to examine items' objectivity, feasibility, and relevancy, leading to discarding some items, with 28 items remaining after a consultation workshop (Bhattarai, 2017).

In the national ELDS workshop held by ERO, UNICEF, and Save the Children in April 2019, a standard setting was conducted to establish cut scores on the ELDS assessment scale to classify children into three levels of development status (development categories): On track, Progressing, and Struggling. The cut scores enabled the country to monitor to what extent children in Nepal are 'developmentally on track', corresponding to the global call based on SDG target 4.2. Also, the tool was scrutinized by measurement experts and other ECD stakeholders through the workshop, which resulted in minor revisions to items and administration and scoring guidelines. The revised tool has 26 items with 58 sub-tasks. Since the cut scores established in the 2019 workshop is for the ELDS assessment tool prior to the revision, the revised tool necessitated updated cut scores. Therefore, ERO and UNICEF held the workshop in 2020 June for the standard setting based on the current ELDS assessment tool.

Despite all the work for designing the tool and establishing cut scores for development categories, there was no reported document to evaluate the procedures employed and provide evidence of validity and reliability of ELDS assessment scores and cut scores for development categories to date. It is critical to evaluate the extent to which the assessment tool actually yields meaningful results in context by gathering evidence of validity and reliability. This report aims to fill the gap of the ELDS assessment tool by providing evidence regarding validity and reliability through qualitative and quantitative analysis.

1.3. Objectives of report

This report aims to conduct analysis and evaluation of the validity and reliability of scores produced by the ELDS assessment tool both qualitatively and quantitatively. The primary audience for this report includes:

1. The users of the ELDS assessment tool, which include but are not limited to technical or administrative officials from the government and non-governmental organizations involved in the assessment administration, data collection and analysis, report writing, and revision of the assessment.
2. Government officials involved in the monitoring of the national/regional progress towards national/international targets on child development and learning, including SDG target 4.2.
3. Audience reading reports that include the data from the ELDS assessment tool.

1.4. Structure of report

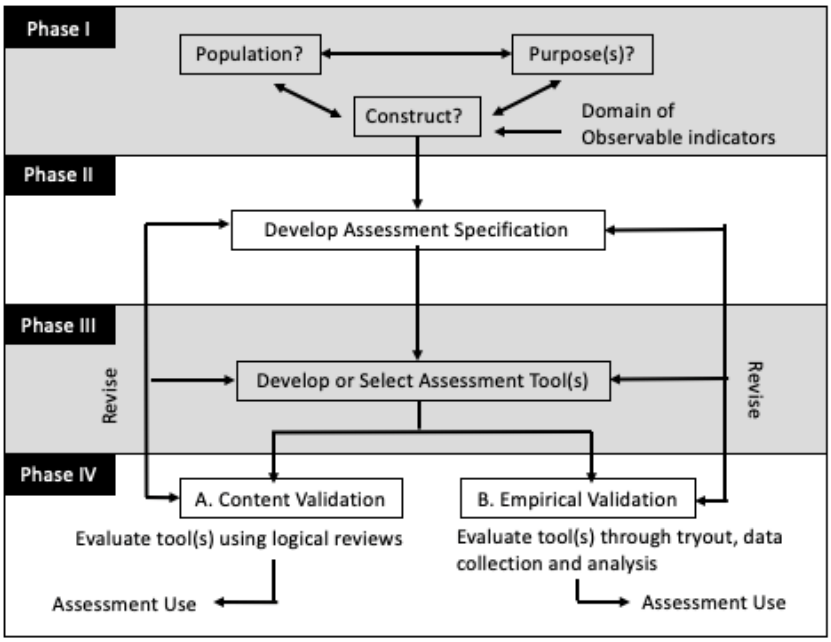
This report is organized in eight sections. The first section presents the background and sets the context of the ELDS assessment in Nepal, followed by the needs and objectives of this technical report. In the second section, the approaches to backwards evaluation of the validity and reliability of ELDS assessment scores are presented, following the Process Model (Chatterji, 2003). Then, the results of various validity and reliability analyses are described in Section 3 through Section 7. Lastly, the report concludes with discussions to inform further improvement of the ELDS assessment tool.

2. Validity and reliability of ELDS assessment score

2.1. Method: The Process Model

This section introduces the approach to evaluate validity and reliability of ELDS assessment scores through qualitative and quantitative analysis, following the Process Model (Chatterji, 2003). The Process Model is an iterative process for assessment design and validation. As illustrated in the figure 1, we followed the four phases, starting with the clear articulation of the assessment context, followed by the description of assessment specification and assessment tool. We also introduced various aspects of validity and reliability, some of which are addressed in depth in the following sections.

Figure 1. Flow Chart of the Process Model



Sources: Chatterji (2003).

2.2. Phase I: Assessment context

As a first phase of the Process Model, we clearly articulated the ELDS assessment contexts, in terms of constructs, population, and purposes, based on which the validity and reliability of the scores are evaluated. The context of the ELDS assessment tool in terms of Constructs, Population, and Purposes, was specified as follows:

Constructs: The construct that the ELDS assessment aims to tap into is children's learning and development outcomes. This construct is multidimensional, which consists of multiple domains that are theoretically distinct, such as physical, cognitive, social-emotional and language development (UNICEF, 2017). In the case of the ELDS in Nepal, six development domains were identified, building on the international ELDS convention as well as reflecting the values emphasized by the country: Physical, Social, Emotional, Cognitive, Language, and Cultural development (Shrestha et al., 2017). The general and specific indicator statements which describe what the intended constructs encompass in terms of expected skills and knowledge will be provided in the later section.

Population: As described in the introduction, the target population of the ELDS in Nepal is children of age 48-60 months. Therefore, the operationalized ELDS assessment tool also focused on the same population. However, given a practical limitation due to which data can

be collected from children when they completed ECED and entered Grade 1, the data is likely to have children above age of 4: the majority of children in Grade 1 is at age 5 or 6. Thus, we defined the population to which the ELDS assessment tool is used and validated as children aged 48-84 months.

Purposes: There are three purposes for the ELDS assessment tool, each of which targets different users as follows:

a. To help the government monitor their progress towards the national and international goals (e.g., SDG target 4.2) based on the ELDS, as well as inspire attention to and scale up of effective ECED initiatives (**Criterion referenced**)

b. To inform research that supports program quality and program improvement (**Norm referenced**)

c. To assess children's development against the ELDS to inform teaching and parenting practice (**Norm/Criterion referenced**)

For the Purpose 1 and 3, standards to understand whether children are developmentally on track are necessary. In contrast, the Purpose 2 requires scores with a certain variability so that they can sensitively reflect the impact of program improvement or any intervention. In short, both raw ELDS scores and development categories with cut scores are required for the above-mentioned purposes. This report mainly addresses the validity issues of raw scores for norm referenced evidence in the following sections while the same evidence is relevant for development categories.

2.3. Phase II: Assessment framework (Domain specification)

The second phase is to establish domain specification, against which the assessment tool is reviewed for validity in the later phase. For the ELDS assessment, there is an existing assessment framework (Shrestha et al., 2017). In the assessment development project led by ERO, there were several rounds of consultations with ECD experts who were also involved in the development of the Nepal ELDS. Furthermore, various national and international literature sources on assessment of early learning and development were reviewed (Shrestha et al., 2017). They included two Nepal ELDS related documents: Early Childhood Development Guidelines (Curriculum) (DOE, 2062 BS) and Early Learning Development Standards (DOE, 2069 BS). The following literature on international ELDS related assessments were also reviewed: International Development and Early Learning Assessment (IDELA) developed by Save the

Children (Pisani et al, 2015), East Asia-Pacific Early Child Development Scales (EAPECDS) developed by the UNICEF (Rebello, Britto & Hancioglu, 2016), and Measuring Early Childhood Outcomes developed by the World Bank (Laat, 2015).

Based on the review of these literature sources and consultation with ELDS stakeholders, above-mentioned six development and learning domains were identified as relevant constructs in the Nepal context. General indicator statements that come along with the assessment framework are provided here, followed by the established assessment framework, which include subdomains, aspects, standards (benchmarks), and particular skills and knowledge (tasks) (Shrestha et al., 2017). Six development and learning domains and their general definitions are as follows:

- **Physical development**: Children's physical health and ability to engage in daily activities.
- **Social development**: Children's ability to form positive relationships that give meaning to children's experiences at home, school, and larger community.
- **Emotional development**: Children's disposition and the emotional competence, rather than skill, for becoming involved in learning and acquiring knowledge.
- **Cognitive development**: Children's ability to understand and think about the physical and social world. In particular, this domain focuses on children's knowledge of objects in the world around them, logic and mathematical knowledge, knowledge of agreed-upon social conventions such as numbers and colors, and understanding and appreciation of the arts in their lives.
- **Language development**: Children's understanding and use of language, emerging reading and writing skills, and ability to communicate effectively.
- **Cultural development**: Children's value related to cultural diversity and respect for cultural heritage for the future of Nepal.

Table 1. Assessment Framework for ELDS

Subdomains	Aspects	Standards	Tasks
Physical development			
Physical development	Gross motor skills	• Demonstrate coordination of body parts	• Walk along straight line balancing the whole body • Stretch, bend and touch own feet • Move some steps backward • Jump some steps with both feet • Hop turn by turn on one foot • Crawl on the ground • Climb up/down ladder, slopes, and steps • Lift and carry a small char of their size • Throw and catch small objects from short distance
	Fine motor skills	• Coordinate and use fine motor body-parts • Demonstrate hand-eye coordination • Use tool, instruments, objects with control	• Tear paper or leaves into pieces • Thread various small objects, such as beads and corncobs • Do up buttons • Click fingers • Color within a boundary • Do simple weaving • Draw lines, circles, and patterns • Fold paper • Cut and paste small objects • Mold and manipulate wet sand, dough or clay
	Sensory motor skills	• Show reactions and responses to something using senses	• Imitate, identify, and differentiate sounds • Follow given rhythm • Show a reaction to things with various feelings (smooth vs. rough, hard vs. soft, etc.) • See and distinguish objects far and near using each eye in turn • Respond to different tastes (sweet, sour, bitter, etc.)
Health and hygiene	Personal hygiene	• Demonstrate health and hygiene behaviors and practices	• Maintain their body, including nose, hand, nail, feet and teeth, neat and clean • Show awareness of dirt vs. clean objects and food • Follow proper toilet practice
	Safe place	• Have understanding of safe practices • Avoid harmful and dangerous objects	• Show awareness of danger of fire, electricity, sharp objects and poisons • Show familiarity with signs of danger and poisons and avoid them

	Food habit	• Have some understanding of health and hygienic food habits and practices	• Follow healthy and hygienic food habits • Avoid bad food habits
Language development			
Communication	Listening	• Listen to and respond to properly	• Listen and distinguish common sounds, such as human voices and animal voices • Listen and follow instructions in familiar language • Listen and respond in familiar language • Listen to short story
	Communicating with others	• Speak simple short sentences • Communicate with others, such as familiar peers and adults	• Speak fluently in familiar language using appropriate words with respect to geography, caste and ethnicity, cultures, and community • Participate in conversation in small groups • Ask and respond to questions • Express views and opinions • Talk with elders and younger children according to local norms and values • Wait for their turn and listen to others
Literacy	Pre-reading	• Demonstrate pre-reading skills	• Interpret the picture and pictorial story • Match letters and words with pictures • Recognize alphabets in Nepali and mother tongue
	Pre-writing	• Demonstrate pre-writing skills	• Draw pictures • Draw and copy patterns • Draw some basic shapes, such as line, half/full circle • Begin to form alphabets
Cognitive development			
Intellectual development	Cognitive skills and learning processes	• Exploration and investigate • Reflect on own learning for further learning	• Use puzzles for explanation and investigation • Use previous knowledge and skills for various day to day situation
	Classification and ordering	• Recognize and describe basic geometric and other shapes	• Describe familiar geometric shapes, such as squares, triangles, circles
Cognitive knowledge	Scientific exploration (weather related)	• Describe sun, moon, and sky	• Name and draw picture of sun, moon, star, and sky
	Living and non-living beings	• Know the parts of boy and their functions	• Name the parts of boy • Differentiate the living and non-living beings
	Scientific knowledge (materials)	• Identify and describe everyday materials	• Identify utensils of everyday use • Identify materials by their quality and property

	Scientific exploration (technology)	• Differentiate some technological materials and tools	• Name technological devices available in the context
	Scientific exploration (transportation)	• Describe type of transportation	• Name types of transportation available in the context
	Distance and direction	• Recognize and describe distance and direction	• Describe distance • Properly name direction
	Concept of time	• Recognize the concept of time	• Name the days of week and months of year
Mathematics	Numeracy	• Recognize and write basic numbers (i.e., 1 to 9)	• Recognize and write numbers 1 to 9
	Measurement	• Differentiate objects with different length and size	• Measure and compare objects with different lengths and size
Creativity	Creative arts	• Create and draw shapes • Sketch objects	• Draw and describe pictures
	Imagination (i.e., drama and role play)	• Use imagination in drama and role play • Create music using instrument	• Create scenario in drama and role play • Create rhythm
Social development			
Social development	Relationship with peers and adults	• Demonstrate relationship with peers and unfamiliar adults	• Interact easily with known family members • Show respect to family members • Interact positively with known community members e.g. shopkeeper, teacher • Greet and interacts with unfamiliar adults appropriately • Ask for help from unfamiliar adult when needed
	Life skills	• Introduce own selves • Take responsibilities in daily activities • Participate in learning and development activities • Demonstrate self • dependent in doing individual and in group work • Involve peers in performing given activities	• Tell owns name/address properly • Take out/put thins properly in right place • Shows curiosity to find out new things and keen to learn • Listen to other before doing something and do following the instruction give • Take and wait for turn of own • Initiate and lead activities • Perform activities independently and confidently • Involve in group work with peers
	Cooperation	• Involve in class/center's daily activities • Share things, feelings/thoughts in groups/with peers	• Help at home bringing things such as books, mug, clothes • Help in class/center to bring and put things back in their places • Take responsibility for owns belongings

			• Offers own things and thoughts to friends as they needed
	Social behavior	• Demonstrate appropriate behavior with peers/adults • Adjust in familiar and unfamiliar environments	• Show sensitivity to the feelings and needs of others • Acknowledge own mistakes • Show respect according to relationship • Mix up own selves in any group • Express comfort in any setting
	Self-concept	• Demonstrate sense of self • Show pride of own existence	• Introduce oneself mentioning name, gender home address etc. • Express owns likes and dislikes • Takes pride of owns abilities and skills • Takes pride of owns family, goods, home etc
Emotional development			
Emotional development	Emotional expressions	• Demonstrate different emotions • Respond to the feeling of others and empathies	• Express different emotions according to situations • Demonstrate satisfaction or pride of what s/he does, achieves, has got etc. • Demonstrate proper reaction when provoked, teased and appreciated etc • Appreciate or compliment others • Show empathy to others
	Emotional security and confidence	• Demonstrate security in different situations • Demonstrate confidence to new experience and learning	• Demonstrate confidence/security in familiar/unfamiliar settings • Knows/realizes that s/he is valued/accepted by others when s/he is called by name • Demonstrate proper coping strategies for different situations • Show independence in individual/group work • Express oneself confidently with eye contact
	Self-control and balancing	• Demonstrate balances over the emotions	• Show modified expression/behavior of emotions for different situations • Help/support other in trouble • Show situational expressions
Cultural Development			
Cultural Development	Family and community	• Know and follow the values and norms of their family and community • Follow and value their own family, community and ECD center routine	• Introduce himself or herself using name, gender, locality and nationality • Involve in their family events such as birthday, marriage etc. • Follow the daily routine of ECD center showing respect for local and national events and festivals.
	Natural and cultural heritage	• Show an appreciation to local rules and cultures	• Know about specific aspect of festival celebrated by her/his family community, e.g. special food

		• Demonstrate respect for natural and cultural heritage	• Talk about different ethnic, cultural dress worn in her/his family and community • Name three heritages of the locality. i.e. River, Pond, Mountain, Cave • Say basic do's and don'ts in the heritage
	Nation and nationality	• Understand and respect people's different needs and culture • Understand children born from Nepali parents are Nepali • Show respect to Nepali emblems	• Mixes up and play with children from different culture, ethnicity, community with respect • Identify and appreciate different ways of life in different region Mountain, Hill and Tarai of Nepal • Identify national flag, bird, animal etc • Sing national anthem and song

2.4. Phase III: Instrument design

Based on the assessment framework, the tool was prepared through expert- and teacher-workshops. Initially, the tool had 44 items with some sub-tasks within items covering the six domains, which were reduced to 28 items after a plot trial and a consultation workshop. In April 2019, the tool was revised through the national ELDS workshop, which resulted in 26 items with 58 sub-tasks. This revised tool is the one analyzed in this validation report. The translated items and subtasks of the tool can be seen in Appendix A.

Children's performance on tasks were scored by trained enumerators based on observations with standardized administration and scoring procedures. Enumerators scored 2 if children performed a task correctly, 1 if performed partially correctly, 0 if performed incorrectly, and 999 if they did not respond (later converted as 0). Domain scores were obtained by dividing the sum of task scores by maximum possible scores, which thus, correspond to the percentage of subtasks that the child performed correctly.

2.5. Phase IV: Validity and reliability of ELDS scores

2.5.1. Unified validation Plan

The phase four of the Process Model is based on the perspective of unitarian validity, which claims that all relevant evidence should be collectively evaluated in a unified manner rather than focusing only on a predetermined narrow set of evidence. Evaluation of the validity and reliability of scores produced through the ELDS assessment tool requires the following set of evidence (Table 2).

Table 2. Unified validation Plan

Validity Evidence	Validation Question	Rationale	Analytic Method
A. Content-Based Validity	To what extent does the evidence show *Content Relevance* and *Content Representativeness* of the items and the overall assessment vis-à-vis the ELDS and the theory about children's development?	To assure that the observable indicators of the constructs are defensible and consistent with existing knowledge, theory, or expert judgment.	Expert reviews via survey. Content Validity Index (CVI; adjusted kappa coefficient). Analysis of qualitative feedback.
B. Evidence of Item Quality (Item Analysis)	To what extent does the evidence show that the items function well to measure the construct?	To assure that items do not have flaws.	Field-test items. Item difficulty index and item homogeneity index.
C. Validity of Theorized Internal Structure (Factorial validity)	To what extent does the evidence show that the variance in item responses is explained	To assure that the domains are explained by mathematically derived dimensions that	EFA and CFA with randomly split data sets.

	by underlying "latent" constructs, per the theory of children's development?	are consistent with existing knowledge, theory or expert judgment.	
D. Discriminant Validity	To what extent does the evidence show that the tool measures the distinct domains of the construct as specified in domain specifications?	To assure that each domain score measures the intended distinct dimensions of the construct.	EFA to analyze the order and pattern of factors. CFA to analyze theoretically specified multi-domain factors in comparison to unidimensional factor models.
E. Evidence of group differences	To what extent does the evidence show that the ELDS scores discriminate by stage of development (age)?	To assure that domain scores and total scores can distinguish children whose developmental status is theoretically different.	Comparisons of mean scores by theoretically important grouping variables that are proxies for developmental stages
F. Score reliability (Internal Consistency)	To what degree does the evidence show that the total scores and domain scores are internally consistent?	To assure that scores can be replicated within tolerable margins of error.	Examine Cronbach's alpha and Guttman's lambda reliability of all domain scores.
G. Score reliability (Inter-Rater Reliability)	To what extent would the scores be replicable, regardless of the enumerators providing the ratings?	To assure that individual scores can be replicated within tolerable margins of error, even when conditions (enumerators) change.	Examine observer's agreement and kappa reliability of the score.
H. Predictive validity	To what extent does the evidence show that the total scores and domain scores predict a future criterion behavior of individuals assessed?	To assure that domain scores are correlated with theoretically related future criterion behavior, such as academic achievement in early grades.	Field-test scales with target population (panel data). Correlations between domain scores with academic achievement in early grades.

As shown in Table 2, there are various kinds of evidence that should be evaluated for the validity and reliability of ELDS assessment scores. However, some of them cannot be addressed in this report due to data constraints: i.e., inter-rater reliability tests were not conducted as a part of training and main data collections in the past studies; and all the ELDS assessment data collected in the past were cross-sectional (no follow up to create longitudinal data was done). Thus, a set of evidence presented in this report should not be considered exhaustive. More data should be collected in the future to investigate validity and reliability of ELDS assessment scores more comprehensively.

2.5.2. Sample for the EDLS assessment data

The validity evidence provided in this report addressed both content validation (qualitative) and empirical validation (quantitative). For the empirical validation analysis, we used the data of children collected in the ERO's annual ECD study in 2019. What follows is a description of the sampling procedure, followed by a note for limitations of the data.

First, ERO selected 15 districts in March 2019: three districts from Bagmati Province and two from each of other provinces that were not included in the previous two rounds of studies. Purposive sampling approach was employed, aiming to reflect the distribution of ECED centers in Nepal's geographical regions (Mountains, Hills and Terai). Based on the number of existing schools in the districts, ERO assigned the number of schools in each district (see third column of Table 3) although this assignment was not strictly proportional.

In consultation with ERO, the respective Education Development and Coordination Unit (EDCU), the district level education office under the MoEST, selected schools in early June. In the selection of schools, the EDCU first identified schools which had 12 or more students in Grade 1 who were in the connected ECED centers in the previous year. They also considered the balance between urban and rural areas when selecting schools. However, as there were no specific rules and criteria in selection, the EDCU's attempt to balance the sample between urban and rural setting might vary district to district. Most invited schools agreed to participate in the school (i.e., overall school level participation rate is 93.8%) (see Table 3).

Table 3. Province and district-wise number of sampled schools and children

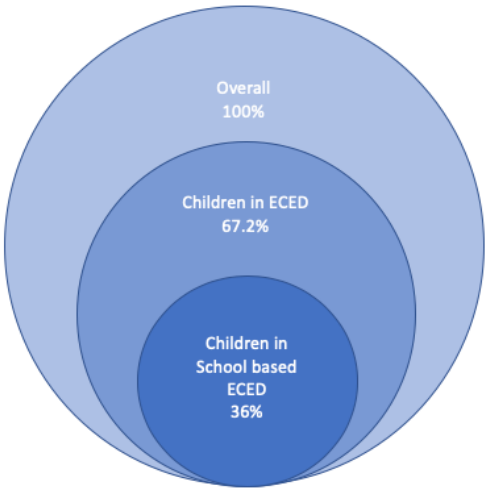
Province	District	Sampled school	Participated schools	Participated children
Province 1	Morang	50	48	566
	Solukhubu	20	20	167
Province 2	Saptari	50	50	539
	Rauthat	50	50	600
Bagmati	Sindhuli	35	35	371
	Bhaktapur	35	35	417
	Dhading	30	16	181
Gandaki	Mustang	10	9	69
	Nawalparasi East	40	40	474
Lumbini	Bardia	50	48	559
	Rolpa	20	20	240
Karnali	Mugu	20	10	79
	Jajarkot	20	19	223
Sudur-paschim	Kanchanpur	50	50	526
	Darchula	20	19	218
Total	15	500	469	5229

Note: The number of registered school based ECED/community ECED/PPE is based on 2019/2020 Flash report I. The current study selected the sample from school based ECED.

In mid-June, the EDCU selected 12 children studying in Grade 1 with ECED experience from each school. When there were more than 12 eligible children, 12 children were randomly selected. In cases where less than 12 students enrolled or presented on the day of assessment, all presented students were included. This selection of children from 469 schools resulted in 5229 children who participated in the assessment. The data from these children and ECED centers (those connected to the sampled schools) were collected in mid-June 2019, which was around two months after children's promotion from ECED to Grade 1. As the sampled children were already in Grade 1, the sample includes many children whose age were above the original ELDS target age (i.e., 48 to 60 months). Also, there was a small proportion of children who were younger than 48 months. In this report, we focused on the data of 4344 children who were at the age of 48 to 84 months at the time of the data collection as described in the section 2.2. In the empirical analysis, we removed 29 observations who did not respond to all assessment tasks, resulting in a sample size of 4315.

The limitations of this study include the fact that the sample focused only on community school-based ECED centers and that the selection of districts and ECED centers was not random. First of all, there was a substantial number of children who did not have access to ECED. The national net enrolment ratio (NER) in ECED was 67.2%, and this ratio was further low in Province 2 (34.1%) in the school year of 2019-20. Also, among three types of ECED programs: the data focused only on public school-based ECED centers: other types are community-based ECED centers and institutional (private) ECED programs. Public school-based ECED centers are the major modality, which accounted for 78 % of ECED programs nationally and 79 % in the studied 15 districts. The proportion of children in public school-based ECED (54%) was even smaller because institutional ECED programs tended to have larger numbers of children than does public school-based ECED. This focus of the data exclusively on children in public school-based ECED limits the representativeness, affecting the generalizability of findings. The population to which this study can be generalized is children who went to public school-based ECED centers in the selected districts, not all children in Nepal (see Figure 2). In addition, the extent to which this study can be generalized to such a population without bias is limited due to the non-random sampling approach. It is recommended to interpret the results of analysis in this report as preliminary findings rather than decisive evidence for validity and reliability of the ELDS assessment scores. Further studies are needed with representative data to rigorously judge the quality of the ELDS assessment scores.

Figure 2. The proportion of study population to overall children.



Note: The proportion of children in school based ECED is calculated based on 2019/20 Flash report: the proportion of children in any type of ECED (national NER: 67.2%) multiplied by the proportion of children in school based ECED to those in any type of ECED in participating districts (54.2%).

3. Evidence of Content based validity

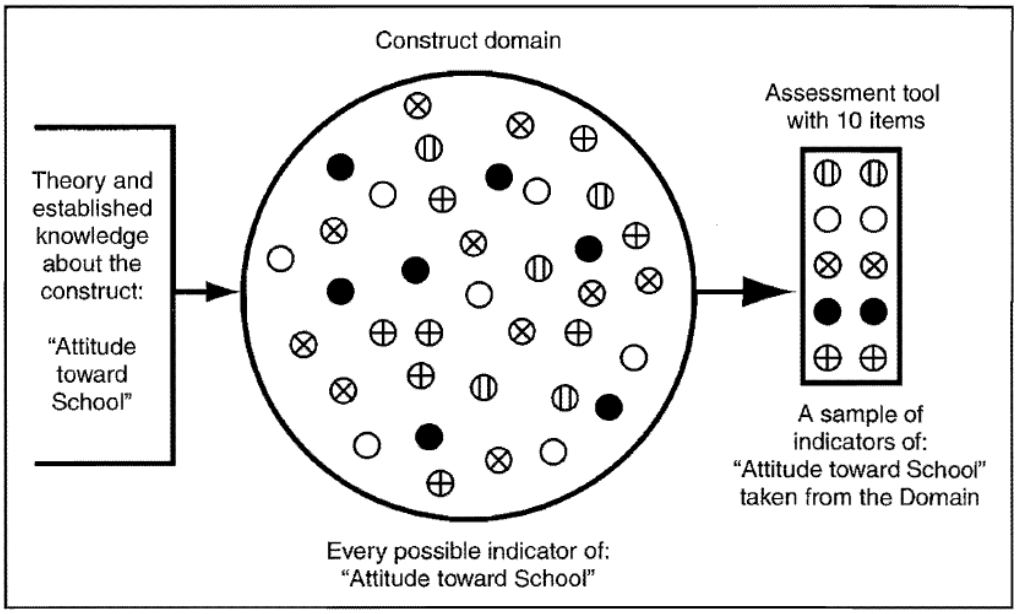
The literature suggests that one should establish content-based validity of assessments as a part of the development process, that is, before one uses them (Chatterji, 2003). However, we conducted content-based validity analysis backwardly on the ELDS assessment tool in use. Content-based validity consists of two aspects: content relevance and content representativeness (Messick,1989; Messick, 1994; Chatterji, 2003). Content relevance analyzes if items are truly indicators of the construct of interest based on logical reviews of content experts and theory (Chatterji, 2003). Content representativeness is about the proportional sampling of assessment items from all possible items of the construct (Chatterji, 2003).

In the earlier step (i.e., Phase 2), the observable indicators were specified that represent the constructs we wish to measure in the assessment framework (i.e., the domain specification). “The term "domain" refers to the entire gamut of all possible items or tasks by which a theoretical construct can be represented” (Chatterji, 2003, p. 58). An assessment tool measuring that construct would be made up of only a sample of the items/tasks from the larger domain (Chatterji, 2003).

See Figure 3 as an illustration of what a sample of indicator from the specified domain means to understand content relevance and content representativeness. For an instrument to be content relevant, we have to ensure that the assessment tasks truly reflect the construct and that all the different aspects of skills, knowledge, or behaviors in the domain are included in the

sample of items that make up the assessment (Chatterji, 2003). For the instrument to be content representative, we need to proportionately sample each different aspect of skills, knowledge, or behaviors in the domain (Chatterji, 2003).

Figure 3. Sampling Indicators from a Construct Domain to Ensure Content Relevance and Content Representativeness.



Source: Chatterji (2003)

Take the physical domain of the ELDS assessment tool as an example. According to the assessment framework (see Table 1), the physical domain consists of two sub-domains and six aspects. Furthermore, each aspect has a few standards which are relatively specific. The assessment framework also contains a list of tasks that are considered to reflect the standards. For the physical domain of the ELDS assessment tool to be content relevant, we need to ensure that all the different aspects of skills, knowledge, or behaviors in the domain as per specified framework are included in the sample of items that make up the assessment and that the included assessment tasks reflect all the constructs. Also, for the physical domain of the ELDS assessment tool to have content representativeness, the proportion of items in different sub-domains and aspects is consistent with the distribution of indicators in the theoretically defined assessment framework.

Assuming that the existing domain specification (the assessment framework) is both content relevant and content representative, we focused on the analysis of content relevance and representativeness of the assessment tool that consists of the sampled items. The review of the tool for the content validation was conducted by the ECD expert panel in April 2022, using the content validation survey. In the survey, experts in the panel were asked to independently

evaluate each of 58 subtasks (grouped into 26 items) with the following four categories: 1 = Not relevant (major modifications needed); 2 = Somewhat relevant (some modifications needed); 3 = Quite relevant (no modifications needed but could be improved with minor changes); and 4 = Highly relevant (no modifications needed). Then, for each task, the content validity index (CVI) was computed as the number of experts giving a rating of either 3 or 4, divided by the number of experts, which is the proportion in agreement about relevance Polit et al. (2007).

A total of eight experts were invited, and all of them agreed to participate in the content validation project (see Appendix B for the summary characteristics of the participating experts). The selection of experts was conducted based on the predetermined criteria as follows:

- One has to have received specialized training in the field of early childhood development (e.g., development psychology) in graduate level: at least some relevant graduate level coursework or desirably completed graduate level degree (i.e., Master's or Doctoral degree) in related discipline.
- If one has only some relevant graduate level coursework but has not completed graduate level degree in the related discipline, she/he has to have at least a few years of professional experience related to young children.

Based on their responses to the content validation survey, we calculated CVI and adjusted kappa, following the procedure recommended in Polit et al. (2007). Both indices focus on expert's agreement on the relevance of items to established domain specification. The latter index adjusted for chance agreements on relevance, which is slightly different from the usual kappa index that adjusted for agreements of any type. Based on the suggestions provided by the literature (Cicchetti & Sparrow, 1981; Fleiss, 1971; Polit et al., 2007), we evaluated content relevant of each task as follows: Fair = k* of .40 to .59; Good = k* of .60–.74; and Excellent = k > .74. Following Polit et al. (2007)'s recommendation, tasks with CVI of less than .78 were considered candidates for revision. In this process, 24 out of 58 tasks were considered irreverent and candidates for revision (see Table 4).

Qualitative feedback from experts were also collected to understand reasons for low relevancies and to inform the improvement of irrelevant tasks. Common reasons for low relevancies provided include the followings: tasks tap into other domains of development than specified; and scoring rules are not specific or standardized sufficiently, leading to an inability for different scoring categories to reflect different levels of specific skills.

Table 4. Ratings on Tasks by Experts

	CVI	k*	Evaluation
Physical Domain			
12: Hopping	0.875	0.87	Excellent
13: Folding paper	1.000	1.00	Excellent
14: Forming shape	1.000	1.00	Excellent
15A: Throwing ball	0.875	0.87	Excellent
15B: Catching ball	0.875	0.87	Excellent
16: Daily hygiene	0.500	0.31	Unacceptable
Language Domain			
5A: Copying letters (straight line)	1.000	1.00	Excellent
5B: Copying letters (circle)	1.000	1.00	Excellent
5C: Copying letters (क)	0.750	0.72	Good
5D: Copying letters (ख)	0.875	0.87	Excellent
5E: Copying letters (A)	1.000	1.00	Excellent
5F: Copying letters (B)	1.000	1.00	Excellent
7A: Following instruction	0.750	0.72	Good
8A: Naming objects (spoons)	0.750	0.72	Good
8B: Naming objects (Banana)	0.750	0.72	Good
8C: Naming objects (Umbrella)	0.750	0.72	Good
8D: Naming objects (Scissors)	0.750	0.72	Good
9A: Identifying initial letter/sound of the words (pot)	0.750	0.72	Good
9B: Identifying initial letter of the words (scale)	0.750	0.72	Good
9C: Identifying initial letter of the words (pigeon)	0.750	0.72	Good
10A: Identifying alphabet (ख)	1.000	1.00	Excellent
10B: Identifying alphabet (घ)	1.000	1.00	Excellent
10C: Identifying alphabet (D)	0.875	0.87	Excellent
10D: Identifying alphabet (B)	0.750	0.72	Good
11A: Comprehending story 1	0.875	0.87	Excellent
11B: Comprehending story 2	0.750	0.72	Good
11C: Comprehending story 3	0.875	0.87	Excellent
Cognitive Domain			
3A: Arranging puzzle	1.000	1.00	Excellent
4A: Knowledge of numbers (1)	1.000	1.00	Excellent
4B: Knowledge of numbers (3)	1.000	1.00	Excellent
4C: Knowledge of numbers (4)	1.000	1.00	Excellent
4D: Knowledge of numbers (6)	1.000	1.00	Excellent
6A: Creative Art (legs)	0.625	0.52	Fair
6B: Creative Art (eyes)	0.625	0.52	Fair
6C: Creative Art (tails)	0.625	0.52	Fair
17A: Identifying color (red)	0.500	0.31	Unacceptable
17B: Identifying color (yellow)	0.500	0.31	Unacceptable
17C: Identifying color (blue)	0.500	0.31	Unacceptable
18A: Concept of time (yesterday)	1.000	1.00	Excellent
18B: Concept of time (tomorrow)	1.000	1.00	Excellent
19A: Identifying of means of transportation	0.875	0.87	Excellent
19B: Identifying of means of communication	0.875	0.87	Excellent
20A: Identifying direction (right)	0.875	0.87	Excellent
20B: Identifying direction (left)	0.750	0.72	Good
21A: Knowledge of size (big)	1.000	1.00	Excellent
21B: Knowledge of size (small)	1.000	1.00	Excellent
21C: Knowledge of length (long)	1.000	1.00	Excellent
21D: Knowledge of length (short)	1.000	1.00	Excellent
Socioemotional/Cultural Domain			
1A: Greeting	0.750	0.72	Good
2A: Self-expression (name)	1.000	1.00	Excellent

2B: Self-expression (age)	1.000	1.00	Excellent
2C: Self-expression (address)	0.875	0.87	Excellent
22A: Friendship	0.750	0.72	Good
23A: Identification of emotion (happy)	0.625	0.52	Fair
23B: Identification of emotion (sad)	0.750	0.72	Good
24A: Empathy	1.000	1.00	Excellent
25A: National anthem	0.625	0.52	Fair
26A: Festivals	0.875	0.87	Excellent

Additionally, we analyzed the coverage of items against the established framework, which is a part of content relevance as well as content representativeness. As shown in Table 5, there are a number of aspects that do not have corresponding items in the assessment tool, suggesting that the assessment tool can measure only a part of the construct of our interest. In other words, the tool does not fully reflect all aspects of the constructs and, therefore, has a limited content representativeness.

Table 5. The Assessment Framework and Corresponding Items

Aspects	Standards	Sampled Tasks
Physical Development		
Gross motor skills	• Demonstrate coordination of body parts	• ***Item 12***. Hopping on one-foot • ***Item15***. Throwing and catching ball
Fine motor skills	• Coordinate and use fine motor body-parts • Demonstrate hand-eye coordination • Use tool, instruments, objects with control	• ***Item 13***. Folding paper • ***Item 14***. Forming shape
Sensory motor skills	• Show reactions and responses to something using senses	None
Personal hygiene	• Demonstrate health and hygiene behaviors and practices	• ***Item 16***. Following daily hygiene activities
Safe place	• Have understanding of safe practices • Avoid harmful and dangerous objects	None
Food habit	• Have some understanding of health and hygienic food habits and practices	None
Language Development		
Listening	• Listen to and respond to properly	• ***Item 7***. Following instruction • ***Item 11***. Comprehending story
Communicating with others	• Speak simple short sentences • Communicate with others, such as familiar peers and adults	None
Pre-reading	• Demonstrate pre-reading skills	• ***Item 8***. Identifying and naming objects • ***Item 9***. Identifying initial letter of the words • ***Item 10***. Identifying alphabet.
Pre-writing	• Demonstrate pre-writing skills	• ***Item 5***. Copying letters
Cognitive Development		

Cognitive skills and learning processes	• Explore and investigate • Reflect on their learning	• ***Item 3***. Arranging puzzle
Classification and ordering	• Recognize and describe basic geometric and other shapes	• ***Item 17***. Identifying color
Scientific exploration (weather related)	• Describe sun, moon, and sky	None
Living and non-living beings	• Know the parts of boy and their functions	None
Scientific knowledge (materials)	• Identify and describe everyday materials	None
Scientific exploration (technology)	• Differentiate some technological materials and tools	• ***Item 19***. Identification of means of transportation and communication
Scientific exploration (transportation)	• Describe type of transportation	• ***Item 19***. Identification of means of transportation and communication
Distance and direction	• Recognize and describe distance and direction	• ***Item 20***. Identifying direction
Concept of time	• Recognize the concept of time	• ***Item 18***. Concept of Time
Numeracy	• Recognize and write basic numbers (i.e., 1 to 9)	• ***Item 4***. Knowledge of numbers
Measurement	• Differentiate objects with different length and size	• ***Item 21***. Knowledge of size, quantity, and length
Creative arts	• Create and draw shapes • Sketch objects	• ***Item 6***. Creative Art
Imagination (i.e., drama and role play)	• Use imagination in drama and role play • Create music instrument	None
Social Development		
Relationship with peers and adults	• Demonstrate relationship with peers and unfamiliar adults	• ***Item 22***. Identifying relationship with friends
Life skills	• Introduce own selves • Take responsibilities in daily activities • Participate in learning and development activities • Demonstrate self • dependent in doing individual and in group work • Involve peers in performing given activities	• ***Item 1***. Greeting
Cooperation	• Involve in class/center's daily activities • Share things, feelings/thoughts in groups/with peers	None
Social behavior	• Demonstrate appropriate behavior with peers/adults • Adjust in familiar and unfamiliar environments	None
Self-concept	• Demonstrate sense of self • Show pride of own existence	• ***Item 2***. Expressing oneself

Emotional Development		
Emotional expressions	• Demonstrate different emotions • Respond to the feeling of others and empathies	• ***Item 23***. Identification of emotion • ***Item 24***. Empathy
Emotional security and confidence	• Demonstrate security in different situations • Demonstrate confidence to new experience and learning	None
Self-control and balancing	• Demonstrate balances over the emotions	None
Cultural Development		
Family and community	• Know and follow the values and norms of their family and community • Follow and value their own family, community and ECD center routine	None
Natural and cultural heritage	• Show an appreciation to local rules and cultures • Demonstrate respect for natural and cultural heritage	• ***Item 26.*** Naming festivals
Nation and nationality	• Understand and respect people's different needs and culture • Understand children born from Nepali parents are Nepali • Show respect to Nepali emblems	• ***Item 25***. National values and norms (national anthem)

4. Evidence of item quality (Item analysis)

4.1. Analytic plan

As a first step of empirical validation analysis, we conducted item analysis to obtain the evidence of item quality. The question that we aim to address through this analysis is: *To what extent does the evidence show that the items function well to measure the construct?*. We employ two indices: item difficulty index (*p*) and item homogeneity index (adjusted item-test: *r*). When we calculated item difficulty index, *p* value, we recorded task scores from 2, 1, 0 to 1, 0.5, 0, and took the average of task scores of the 4315 children so that the *p* value corresponds to the percentage of children who performed the task correctly, while taking into account the partially correct performance (i.e., the score of 0.5). Adjusted item-test correlation correlates the task distribution with the domain score distribution, with the task deleted from the domain score calculation, which therefore, is also referred to as task-rest correlation.

In the item analysis and the following empirical validation analysis, we combined the Social development, Emotional development, and Cultural development into one domain for the following two reasons. First, theoretically, they can be conceptualized as the same construct. More practical reason is that we need to address the reliability issue derived from the insufficient items and subtasks for these three domains: that is, since each domain has two or

three items, it is likely that the resulting domain score does not reach the required level of reliability.

4.2. Results

Table 6 reports the results of item analysis. In the second column, item difficulty index (*p*) for each subtask is reported. Across domains, there are a number of subtasks with the value of *p* value higher than .7 or approaching .9. Such high values of *p* value indicate that these subtasks are easy for most of the studied children. As children in the sample were at target age or overaged and attended ECED, it is expected that most of them achieved these tasks. However, such easy subtasks, and especially those with *p* values of more than .9, have very small variability (see the third column of Table 6) because most children perform them correctly. Thus, they render the variability of the domain score small, which may make the use of raw domain scores for research purposes difficult with low reliability. The number of relatively difficult subtasks, those with *p* values of less than .5, is small: less than a half of the children performed the following five tasks correctly: *11B and 11C* (language domain), *18A and 18B* (cognitive domain), *and 26A* (SE/cultural domain).

In the last column of Table 6, item homogeneity (*r*) for each subtask is reported. According to Chatterji (2003), when tasks show adjusted item-test correlations of .40 or better, they are considered to be well contributed to the domain score. More than a half of the tasks (i.e., 25 out of 58) showed the item-test correlation of less than .40. Furthermore, there were six subtasks which had .30 or lower item-test correlations: *8A* (language domain), *4A, 19A, 21A, and 21D* (cognitive domain), *and 26A* (SE/cultural domain). Among the four domains, the cognitive and SE/Cultural domains had a high proportion of tasks with insufficiently low item-test correlations. Especially, more than half of the tasks in the and SE/Cultural domain showed correlations of less than .40. This finding raised a concern that the tasks were not homogeneous enough to compose a construct. The potential reasons for this issue are that there were multiple constructs within the domain and/or that some tasks were not relevant indicators of the constructs and/or had some malfunctions.

Table 6. Item Difficulty and Homogeneity Statistics

	p (Item Mean)	Item Variance	Adjusted Item-Total Correlation
Physical Domain			
12A: Hopping	0.85	0.30	0.39
13A: Folding paper	0.76	0.31	0.48
14A: Forming shape	0.80	0.30	0.40

15A: Throwing ball	0.86	0.26	0.50
15B: Catching ball	0.76	0.32	0.42
16A: Daily hygiene	0.62	0.34	0.35
Language Domain			
5A: Copying letters (straight line)	0.76	0.43	0.33
5B: Copying letters (circle)	0.81	0.39	0.44
5C: Copying letters (क)	0.90	0.30	0.59
5D: Copying letters (ख)	0.88	0.32	0.60
5E: Copying letters (A)	0.88	0.32	0.60
5F: Copying letters (B)	0.85	0.36	0.59
7A: Following instruction	0.56	0.37	0.32
8A: Naming objects (spoons)	0.96	0.19	0.21
8B: Naming objects (Banana)	0.97	0.17	0.31
8C: Naming objects (Umbrella)	0.97	0.17	0.34
8D: Naming objects (Scissors)	0.95	0.22	0.34
9A: Identifying initial letter/sound of the words (pot)	0.64	0.48	0.55
9B: Identifying initial letter of the words (scale)	0.57	0.49	0.57
9C: Identifying initial letter of the words (pigeon)	0.59	0.49	0.58
10A: Identifying alphabet (ख)	0.82	0.39	0.53
10B: Identifying alphabet (घ)	0.75	0.43	0.55
10C: Identifying alphabet (D)	0.74	0.44	0.59
10D: Identifying alphabet (B)	0.79	0.41	0.60
11A: Comprehending story 1	0.70	0.46	0.42
11B: Comprehending story 2	0.41	0.49	0.37
11C: Comprehending story 3	0.50	0.50	0.38
Cognitive Domain			
3A: Arranging puzzle	0.68	0.37	0.40
4A: Knowledge of numbers (1)	0.95	0.22	0.28
4B: Knowledge of numbers (3)	0.93	0.26	0.40
4C: Knowledge of numbers (4)	0.92	0.27	0.41
4D: Knowledge of numbers (6)	0.87	0.34	0.41
6A: Creative Art (legs)	0.57	0.49	0.41
6B: Creative Art (eyes)	0.72	0.45	0.42
6C: Creative Art (tails)	0.57	0.49	0.42
17A: Identifying color (red)	0.76	0.42	0.38
17B: Identifying color (yellow)	0.71	0.45	0.46
17C: Identifying color (blue)	0.62	0.48	0.47
18A: Concept of time (yesterday)	0.34	0.47	0.42
18B: Concept of time (tomorrow)	0.39	0.49	0.40
19A: Identifying of means of transportation	0.82	0.38	0.28
19B: Identifying of means of communication	0.80	0.40	0.31
20A: Identifying direction (right)	0.63	0.48	0.43
20B: Identifying direction (left)	0.61	0.48	0.44
21A: Knowledge of size (big)	0.98	0.13	0.27
21B: Knowledge of size (small)	0.97	0.16	0.35
21C: Knowledge of length (long)	0.96	0.20	0.31
21D: Knowledge of length (short)	0.95	0.21	0.29
Socioemotional/Cultural Domain			
1A: Greeting	0.91	0.23	0.31
2A: Self-expression (name)	0.96	0.16	0.36

2B: Self-expression (age)	0.81	0.34	0.41
2C: Self-expression (address)	0.84	0.31	0.43
22A: Friendship	0.79	0.32	0.45
23A: Identification of emotion (happy)	0.94	0.22	0.35
23B: Identification of emotion (sad)	0.95	0.22	0.34
24A: Empathy	0.68	0.46	0.36
25A: National anthem	0.69	0.37	0.37
26A: Festivals	0.43	0.38	0.27

5. Evidence of validity of theorized internal structure

5.1. Analytic plan

Through factor analysis, we seek the evidence for validity of theorized internal structure in addressing the following question: *To what extent does the evidence show that the variance in item responses is explained by underlying "latent" constructs as per the established ELDS framework?* Also, we aim to collect the evidence for discriminant validity to show the distinctness of the specified four domains in addressing the following question: *To what extent does the evidence show that the tool measures the four distinct domains of the construct as specified in domain specifications?*

In the following analysis, we broke the two research questions above into the following three sub-questions. 1) *Are the subtasks of the ELDS assessment consistent with the hypothesis of a single domain-level factor?* This question addresses whether the individual domains measured by the ELDS assessment tool are in fact each unidimensional. *2) Do any subtasks measure a domain other than their intended domain?* This addresses the specificity of the subtasks as indicators of their target domain. 3) *How are the ELDS domains related to one another?* In particular, *is there evidence for four distinct domains (i.e., discriminant validity)?*

To address the research questions, we employed the step-by-step approach introduced in the existing studies on a similar assessment tool (Halpin et al., 2019; Wolf et al., 2017). We randomly divided the 4315 observations into two subsamples: an exploratory sample and a confirmatory sample. The purpose of the exploratory sample was to allow for multiple variations on initial models to be fitted to find a "proposed model" for each domain, as well as for the overall ELDS assessment. The purpose of the confirmatory sample was to ensure that the proposed models demonstrated generalizability. The half of the sample, 2157, were used for the exploratory sample and the other half for the confirmatory sample (see Appendix C for the statistical power calculation).

With regards to the first question, separate models were specified for each domain using the exploratory sample. We examined whether subtasks on the same item related to one another, after controlling for the domain-level factor (i.e., in the unidimensional models). In the ELDS assessment tool, subtasks are nested within items, which are nested within domains. Since subtasks on the same items use the same directions and materials and require similar responses, correlation among subtasks on the same item could occur. We fit bi-factor models to inspect whether the factor pattern of the residual factors was consistent with the item structure: subtasks on the same item loaded on the same residual factor due to influences of common directions, materials, and response styles.

Once domain-level structures were tested and proposed models were found, we conducted the analysis with all the four domains to address the second and third questions. We analyzed all subtasks simultaneously by combining the proposed domain-level models. Two models were fit to the full assessment, using the exploratory sample. These models differed in terms of the correlations among the domain factors.

The first model combined the domain-level models without placing any restrictions on the covariances of the domain factors. We inspected whether any subtasks loaded on more than one domain by examining overall model fit and model misspecification indices, which addressed the second research question. This first model provided a reference model for chi-square difference testing of the two following models. We therefore refer to it as the "Unconstrained Model."

The second model replaced the four domain-level factors with a single factor to test whether the four domains were really providing unique information, or whether the ELDS assessment tool only measures a single overarching construct. We refer to this as the "Unidimensional Model." The fit of the Unidimensional Model was assessed using chi-square difference testing against the Unconstrained Model. Such tests concerned the third research question regarding discriminant validity: the inability to reject the Unidimensional Model suggests that the ELDS assessment subtasks measured only one construct rather than the four distinct constructs as per the ELDS assessment framework (i.e., a test of redundancy). The final analytic step in our factor analyses involved assessing the out-of-sample generalizability. We fit the models developed for the within- and across-domain analyses, using the confirmatory sample.

In all the analysis, the subtask level data were modeled as ordinal scale. We conducted the analysis based on the robust weighted least squares estimator, using the lavaan package of R (Rosseel, 2012). In particular, we employed a weighted least square mean-and variance-

adjusted (WLSMV) test statistics, which tends to perform well with ordinal scale (Kline, 2015). We scaled latent response variables (those underlying responses on the corresponding indicators) by fixing their total variance to one (i.e., delta scaling), and scaled the latent factors by fixing their variance to one (i.e., unit variance identification). Thus, the loading of response variables on the latent factors, when they depend on a single factor (e.g., all pattern coefficients in the unidimensional models), is interpreted as structural coefficients, and their squared values are proportions of variance of latent response variables explained by the factor.

We employed the following indices for models' global-fit: comparative fit index (CFI), root mean square error of approximation (RMSEA), and its 90 percent confidence intervals. Following the convection, we used a cutoff value close to .95 or higher on the CFI and close to or below .05 on the RMSEA as criteria of good fit (Kline, 2015). We also examined correlations between residuals based on modification indices following the approach introduced in Halpin et al. (2019).

5.2. Results

Some subtasks showed high correlations with others, which could potentially cause an inadmissible solution in the factor analysis. In the Cognitive domain, one subtask: *Cog20A*; Identifying direction (right), was collinear with the other subtask on that item (i.e., *Cog20B*). The high correlation between these two subtasks (r = .86) suggests that most children who were able to identify the direction of left were also able to identify the direction of right. Additionally, *Cog21B* and *Cog21C* were collinear with the other subtasks on that item (r = .59~.79). In the Language domain, the following four subtasks showed high correlations with others: *Lang5C* (correlated with *Lang5D*: r = .82), *Lang5E* (correlated with *Lang5F*: r =.77, and with *Lang5D*: r = .73), and *Lang9B* (correlated with *Lang9A*: r = .76, and with *Lang9C*: r = .82). To avoid an inadmissible solution due to the high correlation, we omitted these six subtasks: *Cog20A*, *Cog21B*, *Cog21C*, *Lang5C*, *Lang5E*, and *Lang9B*, from all analyses.

5.2.1. Exploratory factor analysis within domains (RQ1)

First, we specified the unidimensional models for each of the four domains. The first panel of Table 7 reports fit indices of the unidimensional models. The results suggest that models did not fit any of the ELDS domains well. An inspection of the modification indices from these models revealed a clear pattern: there were a number of large residual correlations between subtasks on the same item. This finding suggested that subtasks on the same item were related

to one another even after controlling for the domain-level factor possibly due to common directions, materials, and response styles.

Table 7. Summary of Goodness of Fit for Proposed Models at The Domain Level

Domain	(df)	RMSEA (90% CI)	CFI	χ^2-diff (df)	p-value
Unidimensional Models					
Physical	195.19 (9)	.098 (.086, .110)	.938	NA	NA
Language	3947.10 (135)	.114 (.111, .118)	.886	NA	NA
Cognitive	4676.91 (135)	.125 (.122, .128)	.730	NA	NA
Socioemotional	599.99 (35)	.087 (.081, .093)	.851	NA	NA
Bi-factor Models					
Physical	68.77 (8)	.059 (.047, .073)	.980	93.12 (1)	<.001
Physical*	43.62 (7)	.049 (.036, .064)	.988	NA	NA
Language	498.45 (122)	.038 (.034, .041)	.989	1610.60 (13)	<.001
Cognitive	460.38 (125)	.035 (.032, .039)	.980	2408.40 (13)	<.001
Socioemotional	192.36 (31)	.049 (.043, .056)	.957	323.22 (4)	<.001

To address such residual correlations based on the assessment structure, we employed bi-factor models. Besides the domain factor, we included factors corresponding to items when there are multiple subtasks within the item. That is, each subtask was modeled as loading on two factors, the domain factor and a specific factor corresponding to its item. The item factors were assumed to be uncorrelated with the domain factor and with each other.

For items that had a single subtask, the item factor was omitted. For items with only two subtasks, both subtasks' factor loadings on the item factor were fixed to one so that the model identified the variance of the item factor. Furthermore, we used equality constraints with some additional subtasks on the same item to address the issue of inadmissible solutions. Inadmissible solutions or estimate's failure to converge are evident when an estimate has a negative variance or when the standard error of a parameter estimate is large (Kline, 2015). Potential causes of these issues include few subtasks within items (Kline, 2015). In the current analyses, this issue occurred on *Item 8 Naming objects* and *Item 4 Knowledge of numbers*. For subtasks in these items, we fixed the loading of subtasks on the item factor.

The bottom panel in Table 7 shows the results of fit indices of bi-factor models. For the Language, Cognitive and Socioemotional domains, the RMSEA were lower than the threshold of .05 while the Physical domain showed marginal levels of fits (RMSEA > .05). Also, for all domains, CFI was larger than .95. Furthermore, chi-square tests for nested models that compared the bi-factor models against corresponding unidimensional models revealed that bi-factor models fit the data significantly better.

To understand the potential causes of the marginal fit of the Physical domain, we inspected modification indices based on the bi-factor model. The indices revealed that residuals of subtasks between *Phy13A* and *Phy14A* had a substantial correlation. As *Phy13A* and *Phy14A* intended to tap into the same aspect of the Physical domain: Fine motor skills, and response style was similar, this correlation was expected. Therefore, we allowed the residuals of subtasks *Phy13A* and *Phy14A* to be correlated in subsequent analyses, which resulted in adequate level of fit based on all indices (see Physical* in the bottom panel in Table 7). In the subsequent analysis, we used the respecified models for the Physical domain.

5.2.2. Exploratory factor analysis across domains (RQ2, RQ3)

Using the domain level models described above with modifications, we fit all the domains in the same model. First, we specified an unconstrained model in which the four domains are intercorrelated as theoretically suggested in the domain specification. To address the second research question in the factor analysis, we inspected modification indices to see if there are any subtasks that measure a domain other than their intended domain. Then, in addressing the third research question, this theoretically specified unconstrained model was compared to atheoretical single factor unidimensional model. This comparison seeks evidence for four distinct domains as intended in the domain specification (i.e., discriminant validity).

Table 8 summarizes the goodness of fit indices of the two models. In the first row, the unconstrained model showed marginal model fit (CFI < 0.95). Examination of modification indices revealed that the residual of one subtask in the Physical domain (*Phy16A. Daily hygiene*) had sizable associations with other three domain factors. The residual of this subtask was also correlated with residuals of some subtasks in the Language domain (e.g., *Lang11B* and *Lang11C*). Such cross-domain correlations were somewhat expected as this subtask can be considered a language-related construct rather than health/motor-related construct: that is, *Phy16A* measured children's ability to listen to questions and properly respond verbally.

Similarly, the residual of *SE26A* was found to be associated with numerous subtasks in the Language and Cognitive domains. This evidence indicates that *SE26A* might measure whether children listened to questions and properly respond verbally based on their memory about festivals. A relatively small loading of *SE26A* on the socioemotional domain factor (0.44) also suggested that this item did not capture children's social-emotional skills well. Another subtask in the Socioemotional domain, *SE23B*, was also found that its residual was correlated with many subtasks in the Language and Cognitive domains. Unlike *SE26A*, *SE23B* did not

require children to verbally respond, and the other subtask of the same item, *SE23A*, did not show such residual correlations with subtasks in other domains. Thus, it is not clear what caused such cross-domain correlations, indicating a deeper inspection and discussion of the item and administration protocol.

The inspection of modification indices also showed that the residual of *Lang5A* was associated with other three domain factors and with a few subtasks in the Physical domains (e.g., *Phy13* and *Phy14*). Such residual associations might imply that the ability to draw a straight line was more related to fine motor skills rather than writing skills under the Language domain, which was also supported by a relatively small loading of *Lang5A* on the latent language domain factor (0.42). Furthermore, the residual of *Lang7A* had noticeable associations with other three domain factors, particularly strongly with the Cognitive and Socioemotional domains. This finding may indicate that *Lang7A* tapped into cognitive-related skills (e.g., paying attention) and socioemotional-related construct (e.g., singing the song in front of unfamiliar enumerators without being shy).

Lastly, in the Cognitive domain, residuals of *Cog18A* and *Cog18B* were found to be associated with a number of subtasks in the Language domain (e.g., *Lang8B, Lang8C, Lang8D, Lang11B,* and *Lang11C*). Similar to the issues with *Phy16* and *SE26*, these two subtasks in the Cognitive domain seemed to tap into children's ability to listen to questions and properly respond verbally.

The inclusion of these subtasks rendered the distinction of domains less clear (i.e., problematic discriminant validity). The modified Unconstrained model (Unconstrained model* in Table 8), in which the seven subtasks were omitted, showed adequate fit. Thus, we decided to omit these seven subtasks in subsequent analyses (i.e., the Unidimensional Model based on the exploratory sample and all models based on the confirmatory sample).

Table 8. Summary of Goodness of Fit for Proposed Models for The Overall Model

Model	(df)	RMSEA (90% CI)	CFI	χ^2-diff (df)	p-value
Unconstrained Model	4247.57 (1239)	.034 (.032, .035)	.941	NA	NA
Unconstrained Model*	3279.60 (915)	.030 (.029, .031)	.961	NA	NA
Unidimensional Model	3279.61 (921)	.034 (.033, .036)	.948	301.55 (6)	<.001

The goodness of fit indices was reported for the Unidimensional Model in the second row of Table 8. Note that this model maintains bi-factor model, taking item factors into account to address the subtask-item structure of the assessment. The overall fit of the model was

marginal in terms of CFI, and the chi-square difference test for nested models revealed that the unidimensional model fit the data significantly worse than the unconstrained model ($p < .001$). This statistical test that compared the theoretically specified four factor model to atheoretical unidimensional model provided empirical evidence for discriminant validity: that is, with this sample, the ELDS assessment tool measured the four distinct domains of children's development and learning outcomes as indicated in the domain specification. Replacing the four constructs with a single construct did not show acceptable fit to the data, indicating that keeping the four unique domains as distinct is important.

5.2.3. Confirmatory factor analysis

As a final step, we considered the out-of-sample generalizability of the results reported in the exploratory analyses using the confirmatory sample. Note that the seven subtasks found to be problematic (residual correlations with subtasks of other domains) in the exploratory cross-domain models were omitted in this confirmatory domain-level analysis. As shown in Table 9, all domain-level bi-factor models were replicated with acceptable goodness of fit in the confirmatory sample. Parameter estimates are summarized in Figures. 4–7. These figures also include pattern coefficients.

Table 9. Summary of Goodness of Fit for Proposed Models at The Domain Level

Domain	(df)	RMSEA (90% CI)	CFI
Bi-factor Models			
Physical	12.78 (3)	.039 (.019, .062)	.996
Language	439.59 (94)	.041 (.037, .045)	.987
Cognitive	307.85 (95)	.032 (.028, .036)	.985
Socioemotional	117.041 (17)	.052 (.044, .061)	.966

Figure 4. Diagram of the final bi-factor model of the Physical domain

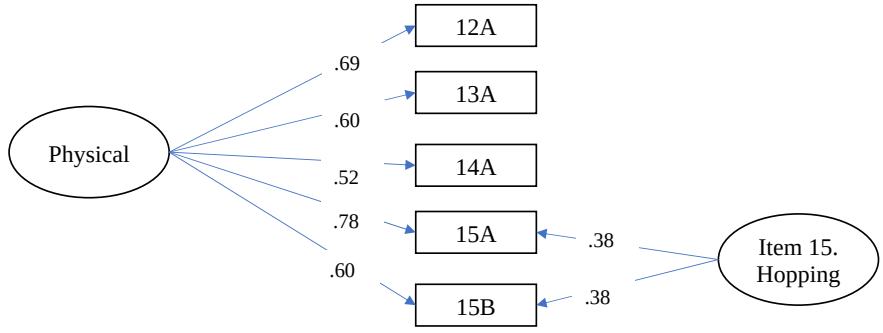


Figure 5. Diagram of the final bi-factor model of the Language domain

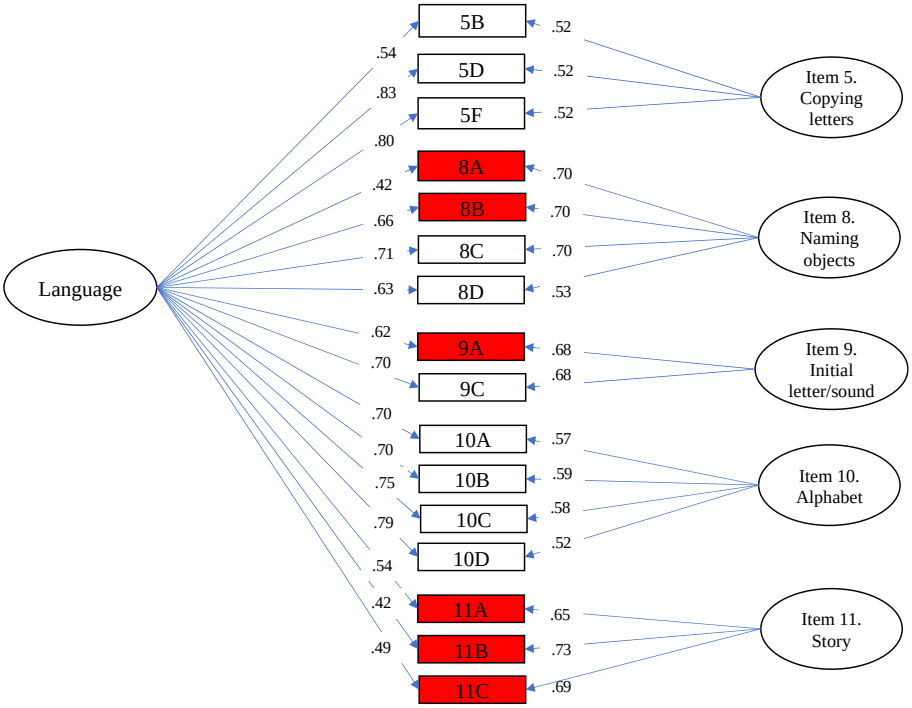


Figure 6. Diagram of the final bi-factor model of the Cognitive domain

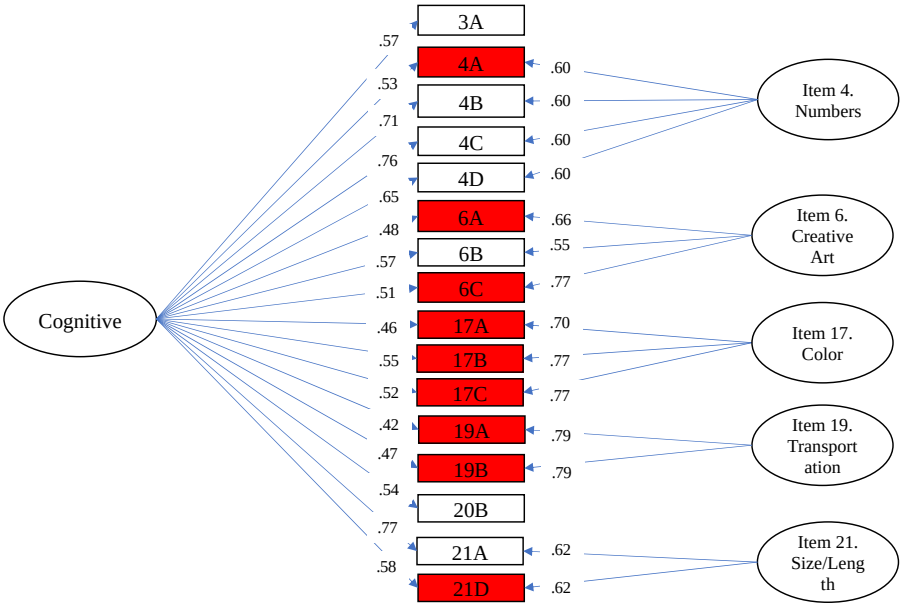
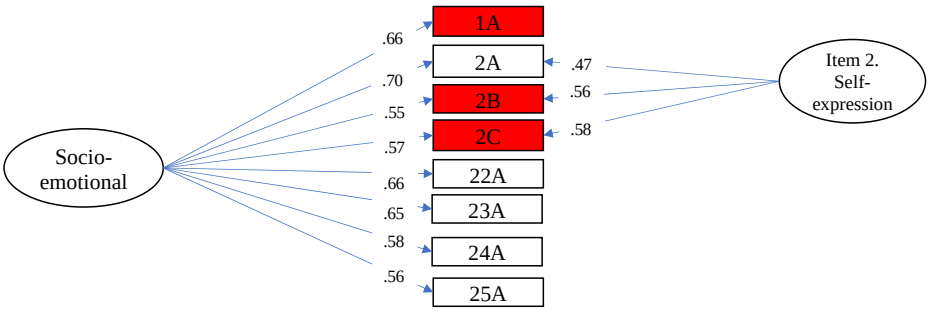


Figure 7. Diagram of the final bi-factor model of the Socioemotional domain



We looked at relative loadings of subtasks on domain factors and on specific item factors. A key assumption is that performance on a subtask is more strongly related to the domain it is intended to measure than its specific items. This is important because if a subtask loads more strongly onto its item factor than the general domain factor, this suggests that the item factor is the principal source of variation. As shown in Figures. 4–7, numerous subtasks showed stronger loadings on their item factors in the Language, Cognitive, and Socioemotional domains (highlighted in red). Such evidence indicated that these subtasks captured children's abilities to perform specific tasks according to the given directions, materials, and response styles rather than intended aspects of developmental constructs. One possible cause is that administration and scoring procedures were not properly designed and standardized to measure the defined development constructs, which should be carefully inspected by ECD and measurement experts.

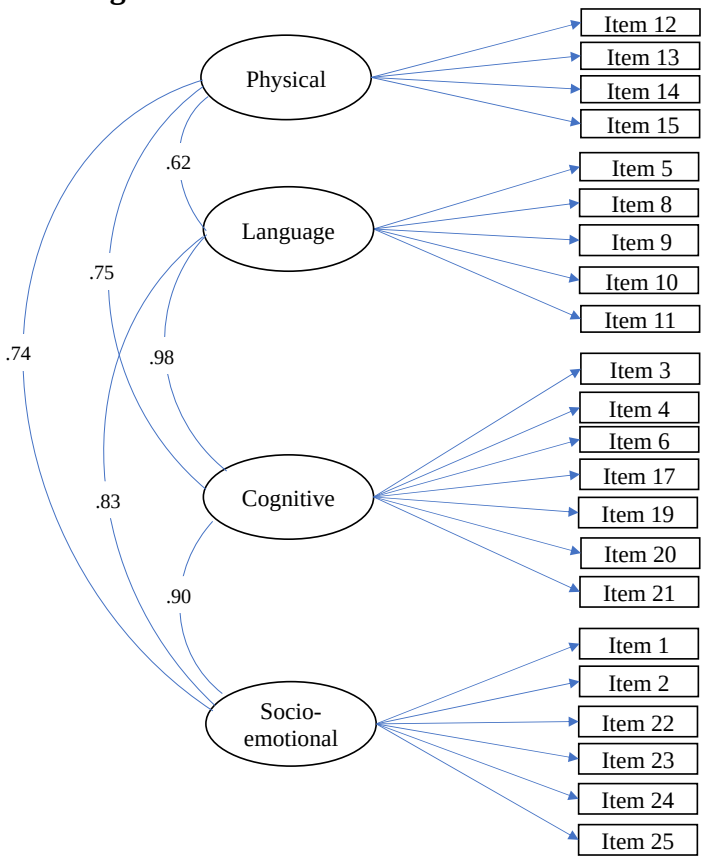
As shown in Table 10, the same conclusions about the overall model were drawn in the confirmatory sample. The Unconstrained Model fit the data well overall. The chi-square test against the Unconstrained Model again rejected the Unidimensional Model. Parameter estimates for the Unconstrained Model are reported in Figure 8. As with the exploratory sample, the correlations among the domains were very large, (r = [.624, .981]). Interpreting the squared correlation coefficients in terms of proportions of shared variance, the highest value of 96% shared variance was between the Language and Cognitive domains. The Physical domain had the lowest correlations with all other domains, with proportions of shared variance ranging between 39% with the Language domain, and 56% with the Cognitive domain.

These strong intercorrelations raised a concern about distinctions among domains as it suggested that a simpler, unidimensional model might be relevant despite the theory. However, the rejection of the Unidimensional Model against the Unconstrained Model provides empirical evidence for discriminant validity: that is, with this sample, the ELDS assessment tool measured the four distinct domains of children's development and learning outcomes as indicated in the domain specification.

Table 10. Summary of Goodness of Fit for Proposed Models for The Overall Model

Model	(df)	RMSEA (90% CI)	CFI	χ^2-diff (df)	p-value
Unconstrained Model	2772.75 (915)	.031 (.030, .032)	.950	NA	NA
Unidimensional Model	3234.93 (921)	.034 (.033, .035)	.938	282.64 (6)	<.001

Figure 8. Diagram of the final bi-factor model of the Socioemotional domain



6. Evidence of score reliability (Internal consistency)

We conducted analysis on internal consistency of the ELDS assessment domain scores to address the following question: *To what degree does the evidence show that the domain score is internally consistent?*

We employed two indices for internal consistency: Cronbach's alpha. Table 11 shows that the domain scores for all the domains exceed or are close to the acceptable level of reliability (i.e., ≥.70). The internal consistency of the Physical and Socioemotional/cultural domains are relatively low. As internal consistency is influenced by the number of tasks that constitute the domain score, the relatively small number of tasks within these domains are likely to be the main cause of low internal consistency.

Table 11. Reliability of the ELDS domain scores

	Original	After omission

Domain	N of tasks	Cronbach's Alpha	N of tasks	Cronbach's Alpha
Physical	6	.69	5	.67
Language	21	.87	16	.84
Cognitive	21	.81	16	.77
SE/cultural	10	.69	8	.66

After omitting a total of 13 tasks (ranging from 1 to 5 across domains) that were found to be problematic through previous empirical analysis, the Physical domain and Socioemotional/cultural domain showed the internal consistencies that are notably below the acceptable level. This finding indicates that a few tasks relevant to these two domains should be added to ensure that each domain score is reliable.

7. Evidence of group difference

It is critical that the ELDS assessment be sensitive to potential differences in children's development by their background characteristics as logically and theoretically expected. Thus, we aimed to collect the evidence of group difference by addressing the following question: *To what extent does the evidence show that the ELDS assessment scores capture differences in children's development by stage of development (age)?*

The literature on the similar assessment tool (i.e., IDELA) analyzes group differences in terms of age, socio-economic status (SES), home learning environment and ECED program learning environment as they are hypothesized or known dimensions of importance (Pisani, Borisova & Jo Dowd, 2015). We are in agreement that the ELDS assessment tool should be sensitive to all these children's characteristics and backgrounds as well as relevant interventions. However, due to the data limitation, we focused only on age as an important aspect of children's development.

The top panel in Table 12 provides the correlations between the ELDS assessment domain scores and children's age. The bottom panel shows the correlations between the IDELA domain scores and children's age across various countries for the comparison. Results showed that all of ELDS development domain scores had small sizes of positive correlations with children's age. These correlations were all statistically significant at the 99 percent confidence level. Compared to the IDELA scores, the degree of correlations of the ELDS assessment scores were small, which thus does not provide the strong evidence of group difference in terms of age. This finding may suggest that the ELDS assessment tool cannot reliably and validly capture the difference in children's chronological development, which can be because the tool

was developed based on the standards for the narrower range of age (i.e., the original target of the standards is children aged 48-60 months). To better establish the evidence of group difference from a comprehensive perspective, more data on household SES, home learning environment, and ECED program environment need to be collected, and correlations with the ELDS assessment domain scores with these important variables should be examined.

Table 12. Correlations between ELDS/IDELA Domain Scores and Child's Age

ELDS	Language	Cognitive	Physical	SE/Cultural
Nepal ELDS	.12	.17	.14	.14
IDELA	Literacy	Numeracy	Motor	SE
Bangladesh	.09	.19	.15	.06
Bhutan	.38	.32	.39	.40
Egypt	.23	.25	.30	.18
Ethiopia	.12	.12		.05
Mozambique	.30	.26		.15
Pakistan	.27	.38		.30
Rwanda	.12	.18	.24	.21
Zambia	.32	.32	.40	.31

Source: Pisani, L., Borisova, I., & Dowd, A. J. (2015). International development and early learning assessment technical working paper. Save the Children.

8. Discussion and recommendation

Based on this analysis, we recommend several revisions to the ELDS assessment tool. First, we suggest the evaluation and removal of subtasks that are highly collinear with other subtasks. This reduction could allow for the addition of different items to the Physical and Socioemotional domain as these domains have a relatively smaller number of subtasks, likely leading to low reliability of the domain scores. As discussed in section 5, the following subtasks showed high correlations: *Cog20A*, *Cog21B*, *Cog21C*, *Lang5C*, *Lang5E*, and *Lang9B*. They were highly associated with other subtasks in the same item, suggesting that if the child can perform well on one subtask, it is highly likely that he/she can do well on other subtasks in the same item: having such highly correlated subtasks do not add any additional information to measure children's development, and thus, they should be removed.

Second, we recommend that subtasks that were found to be substantially correlated with subtasks in other domains than the domain they intend to measure should be removed or revised. The inclusion of such subtasks rendered the distinction of domains less clear: e.g., what is defined as the Socioemotional domain scores may reflect children's cognitive and language development. Such blurry distinctions across domain scores makes it impossible to provide

fine grained evidence and policy implication regarding specific challenges children may face. The factor analysis found the following seven subtasks correlated with subtasks in other domains: *Phy16A, SE26A, SE23B, Lang5A, Lang7A, Cog18A* and *Cog18B*. This empirical evidence for domain contaminations was corroborated by experts from a theoretical point of view, particularly for *Phy16A, SE26A, Lang5A, and Lang7A*. For instance, one expert stated as follows: "Measuring socio-emotional development - does the child feel comfortable enough to sing in front of an audience?"

Among the seven subtasks, *SE23B* and *Lang5A* have other subtasks in the same item that were found to function well (tapping into the intended constructs but no other constructs). Thus, these two subtasks should be removed while keeping other well-functioning subtasks in the same item. Other five subtasks do not have peer subtasks in the same items, and thus, entire items need to be revised or removed. In the revision, one should refer to the evidence from the factor analysis and experts' feedback to understand with which domains the given subtasks were contaminated: for instance, *SE26A* was found to be associated with the Language and Cognitive domains. This subtask intends to measure children's cultural development, which was defined as children's value related to cultural diversity and respect for cultural heritage for the future of Nepal (Shrestha et al., 2017). However, the empirical evidence revealed that it might rather tap into children's ability to listen to questions and properly respond verbally (language) based on their memory about festivals (cognitive). Thus, one should revise this task by making sure that it would capture more precisely children's value and knowledge related to cultural diversity and respect for cultural heritage in a way that does not substantially require their language and cognitive skills.

Third, we recommend that administration and scoring procedures of many items should be reviewed and revised so that specific directions, materials, and scoring procedures do not affect children's performance in a way that is not related to intended development constructs. The empirical evidence from the factor analysis found that numerous subtasks showed stronger loadings on their item factors in the Language, Cognitive, and Socioemotional domains, indicating that specific directions, materials, and response styles were the main sources of information these subtasks measured rather than the intended developmental domains. It is important to review all administration and scoring procedures to avoid any factors other than defined developmental domains affecting children's performance that are scored. In particular, one should ask the following questions when reviewing the administration and scoring procedures: *Does this task with the given direction and materials require children's skills that are not directly related to the defined development domains?*; *Is this scoring rule relevant and*

accurate to reflect different levels of children's skills in the defined domains? Once administration procedures and scoring rules are designed, it is also critical to standardize enumerators' work by emphasizing the procedures and rules in training and evaluating their understanding through fidelity checks and inter-rater-reliability (IRR) tests.

Lastly, we suggest that a few tasks should be added in the Physical and Socioemotional domains according to the assessment framework. The current tool has a smaller number of subtasks in the two domains than other domains (Cognitive and Language), making the domain scores of the Physical and Socioemotional domains insufficiently reliable. However, one should avoid simply adding any tasks that may not precisely measure these domains as doing so may increase score internal consistency but lead to ambiguous distinctions across the four domains; the problem pointed out in the second recommendation. Particular attention should be paid such that new subtasks in the Physical and Socioemotional domains are not contaminated with the Cognitive and Language domains: minimal cognitive and language skills are required for children to perform well on the subtasks in the Physical and Socioemotional domains.

While the first recommendation may rely mostly on the empirical evidence provided in this report, the other recommendation should be jointly guided by review and discussion among ECD experts. The evidence from the empirical evidence in this report is limited and cannot suggest particular revisions to be made while they provide insights into the problems and potential causes.

Appendix A. The ELDS Assessment Tool

Education Review Office
Early Learning and Development Standard Assessment Tool

Section A: Background Information

1) Name of ECD center/ school: ……. School's code no.:

2) Name of child: …… Girl ☐ Boy ☐
3) Child's age: …… month (up to the month of Chaitra 2075)
4) Child's Mother Tongue: ……..

Section B: Learning and Development Standard Assessment

Simple Guide to Enumerator

All the activities of Section B should be conducted personally by the enumerators for the quality and reliability. Before conducting any activities give them instruction and procedure clearly and give example as well. For the management and preparation of the working environment, take help from facilitator. For the operation and assessment of activities, score the point in the given method below.

Please Remember, all the **procedure and instruction** written in bold letters are to be read/ perform by the enumerators as it is. To build the rapport with the children, enumerators are requested to conduct following two activities.

Activity 1: Ask all the children to stand in the semi-circle position (U shaped) and demonstrate a clap in any rhythm and ask them to follow. Then, with the help of facilitator, practice a rhyme that every child knows.

Activity 2: Request children to stand up collectively and conduct activity by asking to touch nose, raise both hands, raise right foot, raise left foot in a child friendly tone, and after praising and thank them with clap, **provide a pencil, eraser, and sharpener to each**.

Then, **my name is …… I live in……. I am also involved in the teaching young children like you. I have come to talk with you and see what kind of games you know that you can do. Everyone, please perform properly what you know and tell me.**

Now call child separately and ask to perform the activities according to questionnaire and assess and score according to learning and development.

Assessment Questionnaire

S. no. 1. Social Development: Life skill **Activity**: Greeting
Materials: None
Procedure and Instruction

- Join hands and greet Namaskar to the children and observe if they reply the greeting or not; then introduce yourself (Name and Address).
- (Join Hand) **Namaskar nanu/babu (name of child).** (wait for a bit). **My name is …… I live in……**
- Score according to child's performance.

Reply Namaskar with expression (2)	Reply Namaskar without expression or provide other answer (1)	Do not reply (-)

S. no. 2. Social Development: Self-concept **Activity**: Expressing oneself **Materials**: None
Procedure and Instruction

- Ask the name of child, age and address.
- **Now I will ask you some questions. Are you ready? …….. Thank you.**
  - **What is your name?**
  - **How old are you?**
  - **Where do you live?**
- Score according to child's performance.
- In first question, if the child can tell his/her first name or full name, give full score in both conditions. In second and third question, mark according to the table.

Question	Answer clearly (2)	Answer with hesitation (1)	Do not response (-)
A. What is your name?			
B. How old are you?			
C. Where do you live?			

S. no. 3. Cognitive Development. **Activity**: Arranging puzzle
Materials: Picture of Dog and its 4 equal dimensional puzzles
Procedure and Instruction

- Attract the attention and show the picture of the Dog. Then hand over the pieces of puzzle to the children. Now, ask them to arrange the puzzle according to the picture. Give them 1 minute to complete.
- **Today we will play several kinds of games. At first, let's play puzzle. Look at me.** (show the picture of Dog) **This is dog. Now, I will give you these pieces of puzzle. You have to make a dog from these pieces.**

- Score according to child's performance.

Materials	Arrange all pieces correctly (2)	Arrange two or three pieces correctly (1)	Could not arrange (0)	Do not response (-)
Puzzle				

S. no. 4. Cognitive Development **Activity**: Knowledge of numbers
Materials: Separate cards of the different pictures of objects with numbers 1, 4, 3, 6
Procedure and Instruction

- Show the pictures to the children and chronologically ask them how many …… are there.

- **Children, please look at these papers. Here are some pictures. (**Showing them pictures) **How many cars are there?** (Showing second picture) **How many fish are there?** Ask continuously.

- Score according to child's performance.

Materials	Correct answer (2)	Wrong answer (0)	Do not response (-)
First (Car)			
Second (Fish)			
Third (Cow)			
Forth (Cat)			

S. no. 5. Language Development **Activity**: Copying letters **Materials**: Pencil and A4 sheet

Procedure and Instruction

- Provide a pencil and A4 sheet to each child to write. Enumerators should show the sheets of straight line (horizontal), semicircle, Nepali and English alphabets A, B, क, ख, chronologically and ask them to the copy the sign and character in the given paper.
  Look, here is something written in the paper. Can you see it? ……. Thank you. Now, please copy them properly as shown in the paper.
- Score according to child's performance

S.no.	Character/ Alphabet	Written correctly (2)	Not written correctly (0)	Do not response (-)
A	———			
B	)			
C	A			
D	B			
E	क			
F	ख			

S. no. 6. Cognitive Development **Activity**: Creative Art

Materials: A card with the picture of a cat without legs, eyes, and tail.

Procedure and Instruction

- Attract the attention and provide the card with the picture of the cat (cat without legs, eyes and tail). Ask children to draw the missing part of the cat with the pencil. Provide 1 minute for the activity.

Children, look at this paper. Here is a picture of a cat. But some parts of the cat is missing. Look at the picture properly and complete the picture by drawing the missing part of the cat.

- Score according to child's performance. Rather than quality, focus if the parts are drawn in the correct place.

Material	Draw in correct place (2)	Do not draw in correct place (0)	Do not response (-)
A. Legs			
B. Eyes			
C. Tail			

S. no. 7. Language Development: Listening **Activity**: Following instruction

Materials: None

Procedure and Instruction

- Sing the rhyme "फूल फुल्यो रेलीमाई आँगनैभरि". Ask the children to sing it.

Now I will sing you a song. First, listen to the song and then sing by yourself. (Sing the rhyme)

- Score according to child's performance.

Sing all the words correctly (2)	Sing with partial words correctly (1)	Sing incorrectly or say cannot sing (0)	Do not response (-)

S. no. 8. Language Development: Pre-reading **Activity**: Identifying and naming objects
Materials: A card with the picture of spoon, banana, umbrella and scissors.
Procedure and Instruction

- Show the picture of Spoons, Banana, Umbrella and Scissors one by one as listed in the assessment tool below and ask children to identify the pictures and name them.

Now, we will play another game. I will show you pictures one by one. Please tell me name of the pictures.

- Score according to child's performance

Pictures	Correct name (2)	Incorrect name (0)	Do not response (0)
Spoon			
Banana			
Umbrella			
Scissors			

S. no. 9. Language Development: Pre-reading **Activity**: Identify initial letter of the words
Materials: None
Procedure and Instruction

- Demonstrate the sample of word "rabbit" first. Then, according to the assessment tool show them the words pot, scale (weight measuring), and pigeon in order and ask the first letter of the words.

  Now, we will play the initial alphabet identification game. First, I will demonstrate you. You observe. R is the initial letter of the word Rabbit. Now your turn. I will tell the word, you have to tell me the beginning alphabet of the word. The word is Pot ……. Thank you. Next is scale ……. Thank you. Word is Pigeon …… Thank you.

- Score according to child's performance

Words	Correct alphabet/ sound (2)	Incorrect alphabet/ sound (0)	Do not Response (-)
Pot			
Scale			
Pigeon			

S. no. 10. Language Development: Pre-reading **Activity**: Identifying alphabet
Materials: Alphabet card

Procedure and Instruction

-Show the Nepali and English alphabet cards (ख, घ) and (D, B) in order and let the child identify them.

Now, let's identify alphabets. I will show alphabets. You answer what those alphabets are. (show ख) **……. Thank you.** (show घ) **……. Thank you.** (show D) **……. Thank you.** (show B) **……. Thank you.**

- Score according child's performance.

Shown alphabets	Correct answer (2)	Incorrect answer (0)	Do not response (-)
First Nepali alphabet ख			
Nepali alphabet घ			
First English alphabet D			
English alphabet B			

S. no. 11. Language Development: Listening Comprehension **Activity**: Comprehending story
Materials: Short story
Procedure and Instruction

- Read aloud the short story in fair rhyme, sequentially ask questions, and score as mentioned in the assessment tool.
- **Now I am going to tell you a story. Listen to the story carefully. Then I will ask you some questions. You have to answer the questions. Ok listen.**

  In one house there were older sister and younger brother. Sister's name was Gita and brother's name was Mahesh. They used to read together.
  - **How many are there in the story?..... Thank you.**
  - **What is their relationship?..... Thank you.**

    What is the name of Geeta's brother?..... Thank you.
- Score according to child's performance

Questions	Correct answer (2)	Incorrect answer (0)	Do not response (-)
A. How many persons are there in the story?			
B. What is their relationship?			
C. What is the name of Geeta's Brother?			

S. no. 12. Physical Development: Gross motor **Activity**: Hopping on one-foot
Materials: None
Procedure and Instruction

- First, show the activity by hopping for 5 times with any one leg. Ask children to hop 5 times with lifting one leg.

  Now, we will play lifting one leg up game. First, I will perform. Observe carefully. First, I will hop five times by lifting legs (hop five times). **Now, you also hop for 5 times by lifting one leg at one place.**
- Score according to child's performance

Performed activity	Hop 5 times correctly (2)	Hop less than 5 times correctly (1)	Hop incorrectly (0)	Do not response (-)
Jump				

S. no. 13. Physical Development: Fine motor **Activity**: Fold paper
Materials: A4 sized paper
Procedure and Instruction

- Give A4 sized papers and conduct the activity by letting children fold the paper along with them (first horizontal fold and then vertical fold).

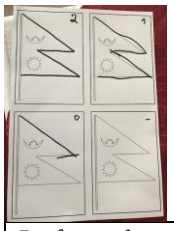
  Children now, fold the paper with me. Look at me. Now, along with me you have to fold the paper as I do.

- Score according to child's performance. While folding paper difference up to 1 centimeter could be considered as right and provide full score.

Performed Activity	Fold straightly in 4 equal parts (2)	Fold properly in 2 equal parts (1)	Could not fold properly in equal parts (0)	Do not perform (-)
Folding Activity				

S. no. 14. Physical Development: Fine motor **Activity**: Forming shape
Materials: Pencil and the dotted picture of the flag
Procedure and Instruction

- Provide children with a dotted flag picture in quarter sized A4 sheets and a pencil. Then ask them to shape by joining dots. Give 1-minute time.
  I will give you a paper. The paper has a dotted figure of flag. the flag by joining the dots properly.



- Score according to child's performance

Performed Activity	Prepare flag by joining dots properly (2)	Prepare flag by joining dots improperly (1)	Do not join dots completely (o)	Do not perform (-)
Form flag by joining dots				

S. no. 15. Physical Development: Gross motor **Activity**: Throwing and catching ball
Materials: Soft ball with 6 to 8 centimeter dimension
Procedure and Instruction

- Show the ball thrown by one person, caught by another. You and child should be in distance facing each other and the child will pass the ball three times towards you and you catch it. You will throw the ball and the child catches it.
  Now, our turn to play ball. Child, I will throw this ball towards you. Catch properly. Then you too throw the ball toward me similarly, I will catch.
- Score according to child's performance.

Performed Activity	Throw 3 times properly (2)	Throw less than 3 times properly (1)	Throw 3 times improperly (0)	Do not perform (-)
A. Throwing task				
Performed Activity	Catch 3 times properly (2)	Catch less than 3 times properly (1)	Could not catch all 3 times (0)	Do not perform (-)
B. Catching Task				

S. no. 16. Physical Development: Personal Hygiene **Activity**: Following daily hygiene activities
Materials: None
Procedure and Instruction

- Ask children in which condition the personal cleanliness activity is carried out.
  Children in which situation/ condition, do you wash your hand with soap and water?
- Score according to child's performance. Appropriate situations are: before the meal; after the meal; and after going to toilet; or related answer

Activity	States 3 situation (like A. before taking meal, B. after taking meal and C. after going to toilet) (2)	States less than 3 situations (1)	States incorrect situation (0)	Do not respond (-)

To follow daily cleanliness				

S. no. 17. Cognitive Development **Activity**: Identifying color
Materials: Red, yellow, blue and green colored cards
Procedure and Instruction

- Put red, yellow, blue and green colored paper's pieces in front of children. Conduct the activity by sequentially asking the cards; give me red colored cards, put that red paper aside and give me that yellow paper, put that paper in the same place, then give me blue paper.
- **Look here. Here are some papers.**
  - **Give me red colored cards? Thank you.**
  - **Give me yellow colored cards? Thank you.**
    Give me blue colored cards? Thank you.
- Score according to child's performance.

Materials	Correct answer (2)	Incorrect answer (0)	Do not respond (-)
Red Paper			
Yellow Paper			
Blue Paper			

S. no. 18. Cognitive Development **Activity**: Concept of Time **Materials**: None
Procedure and Instruction:

- Tell the day of the assessment. Ask them in order; Which day was yesterday? Which day is tomorrow?
  Children today is ……. day. Now tell me which day was yesterday? Thank you. Which day is tomorrow?
- Score according to child's performance

Materials	Correct answer (2)	Incorrect answer (0)	Do not respond (-)
Yesterday			
Tomorrow			

S. no. 19. Cognitive Development **Activity**: Identification of means of transportation and communication
Materials: Picture of television and mobile phone in and Picture of bus and airplane.
Procedure and Instruction:

- Show the paper with the picture of television and mobile phone and another paper with the picture of bus and airplane, then ask about their usage.

Children look here. There are two pictures in the paper.
(Showing picture of bus and airplane) **For what purpose these are the use? What can we do with them?**
(Showing picture of television and mobile phone) **For what purpose these are the use? What can we do with them?**

- Score according to the clarity of the usage.

Materials	Correct answer (2)	Incorrect answer (0)	Do not respond (-)
A. Mode of Transportation			
B. Mode of communication			

S. no. 20. Cognitive Development **Activity**: Identifying direction
Materials: Picture of television and bus.

Procedure and Instruction

- Put the picture of bus on one side and television on another side of the child and ask what is on the right side? And again, what is on the left side? Observe the response.
  Children, what is on your right side? ……. Thank you. Now what is on your left side? ..…Thank you.
- Score according to child's performance

Direction	Correct answer (2)	Incorrect answer (0)	Do not respond (-)
A. Right			
B. Left			

S. no. 21. Cognitive Development **Activity**: Knowledge of size/quantity and length
Materials: Picture of different sized (big, medium, and small) house and pencil.
Procedure and Instruction

- First, show the picture of big, medium and small sized houses and ask which house is the biggest and which is the smallest? Likewise, show the picture of long, medium and short pencil. Ask which pencil is the longest and which is the shortest?
  Children, I will show you picture. (show picture of house) W**hich house is the biggest? …. Thank you. Which house is the smallest? ….. Thank you.** (show picture of pencil) **Likewise, which pencil is the longest?..... Thank you. Which pencil is the shortest?..... Thank you.**
- Score according to child's performance

Shape	Correct answer (1)	Incorrect answer (0)	Do not respond (-)
A. Big			
B. Small			
C. Long			
D. Short			

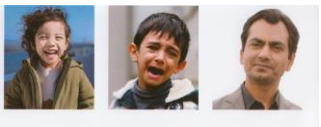
S. no. 22. Social Development **Activity**: Identifying relationship with friends
Materials: None
Procedure and Instruction

- Ask the name of child's friends.
  Children, tell me the names of your three friends?
- Score according to child's performance

Tell name of 3 friends clearly (2)	Tell name of 2 or 1 friends clearly (1)	Tell, has no friend (0)	Do not respond (-)

S. no. 23. Emotional Development **Activity**: Identification of emotion
Materials: Pictures with happy or laughing and crying person's face
Procedure and Instruction

- Show the picture of happy, normal, and sad faces once and ask children to identify happy face. Similarly, ask them to identify crying face.
  Now I will show you some pictures. (show the picture) **Look at these pictures.**



Now show me which is happy or laughing picture?..... Thank you. Similarly, show my which picture is sad or crying?

- Score according to child's performance

Emotional Situation	Correct answer (2)	Incorrect answer (0)	Do not respond (-)
A. Happy			

B. Crying			

S. no. 24. Emotional Development **Activity**: Empathy
Materials: A picture of crying child lying on the ground
Procedure and Instruction

- Show the picture of crying child lying on the ground
  What will you do if your friend cried as shown in the picture?
- Score according to child's performance.

Find out the problem and try to help or express similar intention (2)	Only try to say or say don't do anything (0)	Do not response (-)

S. no. 25. Cultural Development **Activity**: National values and norms (national anthem)
Materials: None
Procedure and Instruction

- Sing the Nepali national anthem "सयौ थुँगा फूलका हामी एउटै माला नेपाली" Then let children sing from start or other one stanza and evaluate accordingly.
  Children, listen I will sing the national anthem and then you sing it. सयौ थुँगा फूलका हामी एउटै माला नेपाली Now, you sing.
- Score according to child's performance.

Response	Can sing or say words of second line (2)	Can sing half words with confusion (1)	Can sing less than half words (0)	Do not response (-)
National Anthem				

S. no. 26. Cultural Development **Activity**: Naming Festival **Materials**: None
Procedure and Instruction

- Ask children what kinds of festivals are celebrated in their community or by them?
  Tell me, what are the festivals that are celebrated in your home or neighbor?
- Score according to child's performance.

Can name three festivals (2)	Can name 2 or 1 festivals (1)	Can name other than festivals (like marriage, birthday) (0)	Do not response (-)

Thank you for Today.

Appendix B. ECD Expert Panel Participants for the Content Validation

Expert	National vs. International	Organization	ECD related degree	Years of professional experience related to young children
1	National	International Organization	PhD in Child Development (on going)	20
2	National	University	Master's in child development and gender socialization (on going)	3
3	National	University	None	5
4	National	University	PhD in Child Development	20
5	National	Government	M.Phil. in Early Childhood Development	20
6	International	International Organization	M.A. in Education	10
7	National	School	Diplomas from private training institutions	24
8	National	International Organization	PhD in Regional Cooperation Studies/ Sociology of Education	10

Appendix C. Sample size calculation for the confirmatory sample

Following the approach in the Wolf et al. (2017), we conducted power analyses at the domain level to calculate the minimum sample size of the confirmatory sample. Results of the power analyses are summarized in Table B1. Degrees of freedom for each domain-level model were computed using a bi-factor model. Minimum sample size was computed for α =.05 and power =.80, and using RMSEA =.05 for the alternative distribution, values above were considered to represent poor model fit. We selected RMSEA =.00 as the value of the null distribution, which were considered to represent acceptable model fit.

Based on the minimum sample size of the Motor domain, which had the fewest degrees of freedom (i.e., *df* = 3), the minimum sample size of the confirmatory sample was found to be 1455. As the sample for the current study has 4315, whose half satisfies the minimum sample size, we randomly selected 2158 children (half of the entire sample) to be in the confirmatory sample. The remaining half, N = 2157, were assigned to the exploratory sample.

Table B1. Power analysis for factor analysis determining sample size of confirmatory sample.

Domain	Degree of freedom	Minimum sample size	Power at 2158
Physical	3	1455	0.94
Language	96	168	~1
Cognitive	95	169	~1
SE/cultural	22	397	~1

References

Bhattarai, G. P. (2017). A Report on ECD Assessment in Nepal 2017. Sanothimi, Bhaktapur, Nepal: ERO.

Bronfenbrenner, U. (1986). Ecology of the family as a context for human development: Research perspectives. *Developmental psychology*, 22(6), 723.

Chatterji, M. (2003). Designing and using tools for educational assessment. Allyn & Bacon.

Cicchetti, D. V., & Sparrow, S. A. (1981). Developing criteria for establishing interrater reliability of specific items: applications to assessment of adaptive behavior. *American journal of mental deficiency*.

Department of Education (DOE). (2062 BS). Early Childhood Development Guidelines (Curriculum). Sanothimi Bhaktapur, Nepal: Author.

Department of Education (DOE). (2069 BS). Early Learning and Development Standards for the Children from 48 to 60 Months Old Sanothimi Bhaktapur, Nepal: Author.

Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological bulletin*, 76(5), 378.

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.

Kline, R. B. (2015). Principles and practice of structural equation modeling. Guilford publications.

Laat, J. D. (2015). Measuring Early Childhood Outcomes. Washington D.C: The World Bank.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13-103). New York: Macmillan.

Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational researcher*, 23(2), 13-23.

Polit, D. F., Beck, C. T., & Owen, S. V. (2007). Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. *Research in nursing & health*, 30(4), 459-467.

Pisani, L., Borisova, I. & Jo Dowd, A. (2015). International Development and Early Learning Assessment. Technical Working Paper. NA: Save the Children.

Rebello Britto, P. & Hancioglu, A. (2016). Measuring Early Childhood Development in the SDG. Webinar Presentations. New York: UNICEF.

Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling and more. Version 0.5–12 (BETA). *Journal of statistical software*, 48(2), 1-36.

Shrestha, K., Pradhananga, D., & Bhattarai, G. P. (2017). Devising a Framework for Assessing Early Childhood Learning and Development in Nepal. *Educational assessment*, 57.

UNICEF. (2017). Early Learning and Development Standards and School Readiness: Evaluation Report. NY: Author.

United Nations. (2020). Global indicator framework for the Sustainable Development Goals and targets of the 2030 agenda for sustainable development. Author. https://unstats.un.org/sdgs/indicators/Global%20Indicator%20Framework%20after%202021%20refinement_Eng.pdf

Wolf, S., Halpin, P., Yoshikawa, H., Dowd, A. J., Pisani, L., & Borisova, I. (2017). Measuring school readiness globally: Assessing the construct validity and measurement invariance of the International Development and Early Learning Assessment (IDELA) in Ethiopia. *Early Childhood Research Quarterly*, 41, 21-36.