

Results from Four MELQO Pilot Countries: Evidence of Within-Country Validity and Cross-Country Comparison of Child Development data

Overview

The Measuring Early Learning Quality and Outcomes (MELQO) project is designed to provide country-level evidence on child development and learning to meet the strong demand for national data among developing countries. While a range of tools are currently in use in these countries to measure child development, generally little attention has been paid to establishing technical standards for validity across measures. Any measure of child development should be backed up by analysis to establish validity, which is important for determining whether an assessment measures what we think it measures and thus to ensure that data are fit for application to policy.

This brief presents evidence on validity from four countries participating in MELQO in its early phases, all of which used the MODEL (Measure of Development and Early Learning) instrument. The MODEL tools have now been used in 13 countries; this brief is intended to provide an overview of the first participating countries as MELQO has progressed from a concept to a testable set of items.¹ Results from MELQO countries will be shared as they become available.

This brief focuses on three aspects of validity: concurrent validity, which as defined here refers to the extent to which MODEL scores correlate with expected predictors such as age and family socioeconomic status; internal validity or the degree to which items in the measures work together to measure underlying

constructs; and the extent to which MODEL scores demonstrated expected factor structures within and between countries using measurement invariance models. Changes made to the MODEL instrument based on the data are also presented.

Participating countries

Countries participating in the first round of MELQO pilots included Lao PDR, Madagascar, Mongolia, and Tanzania. Each country used a different mix of assessments depending on their needs; the sample sizes and assessment types used are noted in Table 1. Children were recruited from preschools in three countries (Tanzania, Madagascar, and Mongolia) and through home visits as part of a research study in one country (Lao PDR).

The primary reason for participating in the MELQO pilot varied by country. For Lao PDR and Madagascar, impact evaluations were underway, and MODEL scores were used as part of a larger evaluation. For Tanzania and Mongolia, the pilot samples reported here were the first phase of larger studies and were used to test the functioning of the scales before moving to nationally representative samples. In all cases, the use of MODEL tools was framed as a pilot, intended to generate evidence for their validity, with the expectation that they would be modified in response to the data.

Country	ASSESSMENT TYPE			AGE		GENDER (FEMALE)
	Direct assessment	Teacher report	Caregiver report	M (range)	Standard deviation	Number (%)
Lao PDR (n = 200)	X		X	4.34 (2 – 7)*	1.651	99/200 (49.5%)
Madagascar (n = 200)		X	X	4.79 (2.4 – 7.2)	0.938	100/200 (50%)
Mongolia (n = 533)	X	X		5.44 (4.5 – 6.4)	0.528	258/533 (48.4%)
Tanzania (n = 209; n = 85 for age)	X	X		5.28 (3.2 – 6.8)	0.853	100/209 (47.8)

This brief is the first in a series on early childhood measurement. The series builds on the experience of the Measuring Early Learning Quality and Outcomes (MELQO) modules, designed to provide country-level evidence on child development and learning in response to country demands for national data. This brief both summarizes findings and lessons learned to date and offers practical guidance for others wishing to conduct research using the MELQO modules. To access other MELQO briefs and for more information on the MELQO initiative, please visit ecdmeasure.org. This brief was written by Abbie Raikes and reflects the work of the MELQO global team, including but not limited to Paul Dizona, and Leslie Hawley, Magdalena Janus, Natalie Koziol, and Linda Platas.

Methods and findings

Methods

Evaluation steps and description of measures. As part of the MELQO project, experts outlined several sources of evidence considered central to establishing validity. These included (i) documentation of internal consistency, or the extent to which items index a common underlying construct; (ii) documentation of concurrent validity by identifying associations with family background, age, and child gender; and (iii) examination of cross-country validity through measurement invariance modeling, which establishes the degree to which items function similarly across countries.

The measures used included a direct assessment of children, using a tool administered by an observer, and a teacher/caregiver report, which is a survey on children's development that administered to teachers and/or caregivers through an interview. Both instruments contained items indexing children's cognitive development (in mathematics and literacy), executive function, and social/emotional development. A full list of items tested in each instrument is presented in the Annex.

Creation of summary scores using scaled means. To create summary scores for further analyses, scaled means were calculated for the direct assessment and the teachers' and caregivers' reports. We chose to use scaled means because, in contrast to using raw data, doing so puts tasks on the same scale and allows

for easier comparisons among the diverse scales used across countries. For the direct assessment, the teacher report, and the caregiver report, children's scores were summarized by creating a score for each item. Scaled means were calculated for each task as follows:

- First, items were recoded to have a minimum response option of 0 if necessary;
- Then, the mean of each item was calculated; and
- Finally, the sample mean was divided by the maximum number of possible responses.

The scaled means have a possible range of 0 to 1, with values closer to 1 indicating easier tasks (meaning that on average, the mean for each task was higher).

Results: Functionality of MODEL items

Several steps were taken to document the functioning of the MELQO items. First, descriptive information was prepared to describe the ages and percent of boys and girls in each sample. Results indicated that on average, children were between ages four and five, which is considered the appropriate age for the MODEL assessment. Table 2 summarizes the analyses and findings on the functionality of MODEL items

Table 2. Psychometric analyses to test functionality of MODEL items

TYPE OF ANALYSIS	PURPOSE	FINDINGS	CONSEQUENT TOOL IMPROVEMENTS
Item response theory	To establish strength of relationship between items and overall scores on scale	Items showed different patterns in different countries; a few items were consistently hard in some countries and consistently easy in others.	Items that were consistently easy or difficult were evaluated for potential exclusion from the core set. Difficult items were kept to preserve workable items, and easy items were removed and recommended for younger samples.
Inter-correlations	To establish associations with key predictors of child learning and development	Reliable and moderate to large associations were found between scores on parent/teacher and direct assessments; with child age; and with family background, as available.	No changes made to tools based on this information.
Factor analysis	To identify underlying constructs within measures	One factor was identified, rather than distinct factors for multiple underlying constructs.	Items were added to literacy and mathematics assessments to increase odds of obtaining distinct factor structures.
Measurement invariance	Evaluate uniformity in factor structures across countries, as a step towards documenting relevance across countries.	There was partial evidence of measurement invariance, meaning the test should not be considered fully comparable across countries.	No changes were made to tools based on this information; will repeat analyses with a more uniform data set (expected 2018).

Results: Internal validity

Item Response Theory (IRT). To test the internal validity of the data, item-response theory was used to determine the item difficulty of the child assessment and the parent/teacher report across countries. IRT results can be used to remove items that do not differentiate well between children who are high scorers and those who are low scorers, that is, items that may not contribute valuable information to an overall score. They can also be used to examine the degree to which items are equally difficult across. For countries considering a larger administration of MODEL, IRT can also provide valuable information on which items may be redundant and therefore can be removed from the assessment.

Across the four countries, results indicated that the difficulty

rankings were generally consistent across countries for cognitive items, both in the direct assessment and in the teacher and caregiver reports. Direct assessment results indicated that backwards digit span, an executive function item, was very difficult for children in all countries, while measurement vocabulary was easy. Variation in direct assessment item difficulty occurred for some items; for example, number identification was easier for children in Tanzania than for children in Laos and Mongolia. Teachers and caregivers reported that name writing was consistently challenging across countries, while measurement vocabulary was again rated as easy. Social/emotional items were also analyzed, and items that were redundant with other items or showed limited added value to the overall scales were identified and targeted for removal or revision. Complete IRT results are presented in Tables 3 and 4.

Table 3. IRT Analysis for MODEL direct assessment

	TANZANIA		MONGOLIA		LAO PDR	
	Average score ^d	Difficulty rank ^e	Average score ^d	Difficulty rank ^e	Average score ^d	Difficulty rank ^e
Measurement vocabulary	0.92	20	0.72	19	0.84	20
Verbal counting	0.72	16	0.45	11	0.35	10
Producing a set	0.85	19	0.46	12	0.47	14
Number comparison ^a	0.70	14	0.85	20	0.75	19
Mental addition ^b	0.58	9	0.51	13	0.50	15
Spatial vocabulary	0.77	18	0.72	18	0.60	17
Mental transformations ^c	0.31	3	0.43	10	0.27	4
Number identification	0.76	17	0.36	7	0.30	7
Expressive vocabulary	0.53	8	0.41	9	0.39	12
Initial sound discrimination	0.33	4	0.25	5	0.14	2
Initial sound identification	0.15	2	0.15	4	0.15	3
Letter naming knowledge	0.45	6	0.12	3	0.33	8
Listening story comprehension	0.62	10	0.69	17	0.50	16
Name writing	0.69	13	0.39	8	0.40	13
Head, toes, knees, shoulders	0.47	7	0.25	6	0.28	5
Forward digit span	0.70	15	0.54	15	0.63	18
Backward digit span	0.11	1	0.05	1	0.08	1
General knowledge	0.44	5	0.11	2	0.34	9
Perspective-taking/empathy	0.64	11	0.52	14	0.30	6
Understanding feelings	0.68	12	0.65	16	0.35	11

^aQuantitative comparison of two sets (Mongolia and Lao PDR)

^bAddition with two sets (Mongolia and Lao PDR)

^cSpatial visualization (Mongolia and Lao PDR)

^dAverage score = Rescaled average sample score (0 = all children received the lowest score, 1 = all children received the highest score)

^eDifficulty rank = Rank of task difficulty within country (1 = hardest task, 20 = easiest task)

Table 4. IRT analysis for MODEL teacher assessment

	TANZANIA		MONGOLIA		MADAGASCAR	
	Average score	Difficulty rank	Average score	Difficulty rank	Average score	Difficulty rank
Verbal counting	N/A	N/A	N/A	N/A	N/A	N/A
Producing a set	0.89	10.5	0.65	4	0.73	6
Number identification	0.82	6.5	0.55	3	0.68	4
Shape naming	0.64	2	0.89	10	0.87	9
Mental addition	0.79	4	0.73	5	0.72	5
Measurement vocabulary, set 1	0.90	12	0.91	11	0.96	12.5
Measurement vocabulary, set 2	0.82	6.5	0.87	9	0.83	7
Spatial vocabulary	0.84	8	0.86	8	0.92	11
Time concepts	0.91	13	0.78	6	0.65	3
Motivation to read	0.8	5	0.96	12.5	0.85	8
Expressive vocabulary	0.89	10.5	0.85	7	0.96	12.5
Listening story comprehension	0.86	9	0.96	12.5	0.89	10
Letter naming knowledge	0.72	3	0.15	1	0.45	2
Name writing	0.62	1	0.16	2	0.29	1

Results: Concurrent validity

Intercorrelations

The next step was to document the correlations between the constructs tested and children’s demographic and family characteristics and, as data were available, to document the associations between the direct assessment results and those from the teacher and caregiver reports.

- **Direct assessment.** As predicted, children who scored higher on the direct assessment cognitive domains of math and literacy also scored higher on the direct assessment domains of social/emotional development in Tanzania, Lao PDR, and Mongolia. The domains of math and literacy were more strongly correlated with one another than with the social/emotional domain. In addition, the cognitive domains showed stronger relationships with children’s age than the social/emotional domain did.
- **Teacher and caregiver reports.** The teacher and caregiver reports of children’s development showed strong correlations between children’s mathematics and literacy skills. There were also reliable correlations, albeit smaller, between children’s social/emotional functioning and

their cognitive functioning; and between children’s ages and teacher and caregiver reports on children’s abilities in both the cognitive and the social/emotional domains.

- **Direct assessment and teacher and caregiver reports.** Correlations between the direct assessment and the teacher and caregiver reports indicated stronger concordance among these assessments concerning the cognitive domains than the social/emotional domains. Concerning children’s cognitive development, the average correlations between teacher and caregiver ratings and children’s direct assessment scores indicated a good deal of agreement in both Tanzania and Mongolia. In Lao PDR, the ratings were very high, which may have been related to the method of collecting the direct assessments; home visits were used to collect information on both the direct assessment and the teacher and caregiver reports, whereas the children were assessed separately from parent and teacher ratings in other countries.

Concerning social/emotional development, the ratings between the direct assessment and teacher and caregiver reports were much lower, suggesting that teachers, caregivers, and children’s responses to direct assessment items are potentially indexing different elements of children’s social/emotional functioning.

Factoring

To determine whether the items in the direct assessment and the teacher and caregiver reports confirmed expected factor structures, confirmatory and exploratory factor analyses were conducted. Results indicated a one-factor solution, or a structure in which all items loaded onto one factor of “school readiness.” There was no differentiation between factors indexing mathematics, literacy, and executive function, for example. For social/emotional development, additional analyses were undertaken, resulting in a set of teacher-rated social/emotional items that was subsequently used in revisions of the tools.

- **Measurement invariance.** A test of measurement invariance, or the extent to which items function similarly across countries, was then performed. While items showed some degree of comparability in terms of item difficulty and the overall strength of correlations, the measurement invariance models suggested that the items’ internal structure was not the same in all countries. Some direct assessment items performed better across countries than others; these included name writing, listening comprehension, and general knowledge, which were all challenging items. This may account for the increased consistency across countries. (Statistics describing model results can be accessed on the MELQO portal at ecdmeasure.org.)

Tool improvements

After conducting analyses, changes to the MODEL instruments were made (see Table 2). Several social/emotional items were removed and revised; and items on the direct assessment that were consistently easy (such as measurement vocabulary) were targeted for inclusion only when the sample of children participating in MELQO was anticipated to be below four years of age. The most challenging items, such as backwards digit span and name writing, were retained to ensure adequate range for administration in the first year of primary; these will be evaluated in further revisions of the MODEL tools.

Conclusions and Next Steps

In sum, the results offer preliminary evidence of the reliability of the MODEL items. Items from both the direct assessment and the teacher and caregiver reports do show expected relationships with other domains of school readiness and with expected variations in the strength of the associations

between cognitive development and social/emotional development. Cognitive development items were more reliably related to one another within and across the direct assessment and the teacher and caregiver reports than social/emotional development. But social/emotional development also demonstrated an important association with cognitive development. Associations between social/emotional and cognitive development confirm scientific work emphasizing the importance of looking at all domains of children’s development.

Future work on the MODEL tool will focus on confirming the factor structures within and across countries. Because the first tests of MODEL were based on pilot data, there were limitations in the data that will be addressed in future analyses. Work to develop a more comprehensive battery of social/emotional items is also underway to increase the number of items used and develop a holistic picture of children’s social/emotional development.

Notes

¹ As of November 2017, the Measure of Development and Early Learning (MODEL) module has been used in Bangladesh, China (Yunnan Province), Peru, Ethiopia, Kenya, Lao PDR, Lesotho, Liberia, Madagascar, Mongolia, Nicaragua, Sudan, and Tanzania/Zanzibar. This brief focuses on data from an initial set of countries for which data were available (Lao PDR, Mongolia, Madagascar, and Tanzania).

Annex: MODEL Tasks

Direct Assessment Tasks

DOMAIN	ABBREVIATION	TASKS	CONSTRUCT	SAMPLE <i>(exact wording can be accessed in MELQO portal)</i>	MAINTAINED IN CORE VERSION?
Early literacy skills	EV	Expressive vocabulary	Expressive language	Can you name some things you can eat that you buy at the market? Can you tell me the names of some animals you know?	Yes
	FWPC	Familiarity with print concepts	Print awareness	Can you show me how you would open book to read? Can you show me where I should start reading?	Yes
	ISD	Initial sound discrimination	Phonemic awareness	"Hats" starts with /h/. What else starts with /h/: ham, bread, or bean?	Yes
	ISI	Initial sound identification	Phonemic awareness	What is the first sound in "mouse"?	Yes
	LNK	Letter name knowledge	Alphabet knowledge	What letter is this? (while pointing to sheet of paper with letters)	Yes
	LSC	Listening comprehension - story	Language comprehension	I am going to tell you a story. After the story, I will ask you some questions about what happened.	Yes
	NW	Name writing	Print knowledge	Can you write your name here?	Yes
Early mathematics skills	MA	Mental addition	Number/operations	If you have three stones and I give you two more stones, how many stones will you have altogether?	Yes
	MT	Mental transformation	Spatial awareness	Look at these pieces. Now look at these shapes. If you put the pieces together, they will make one of these shapes. Point to the shape that the pieces make.	Yes
	MV	Measurement vocabulary	Measurement	(Show picture of three elephants). Put your finger on the smallest elephant.	Yes
	NC	Number comparison	Number/operations	Which number is greater, 3 or 5? 8 or 6? 4 or 7?	Yes
	NI	Number identification	Number/operations	I will point to a number and I'd like you to tell me which one it is.	Yes
	PAS	Producing a Set	Number/operations	Please give me 3 stones. Now please give me 6 stones. Now please give me 14 stones.	Yes
	QCTS	Quantitative comparison of two sets	Number/Operations	Put your finger on the picture with the most tomatoes.	No
	SN	Shape naming	Spatial awareness	What is the name of this shape?	No
	SV	Spatial vocabulary	Spatial awareness	Point to the picture with the ball on/under/in front of/next to the chair. (4 items)	Yes
	SVI	Spatial visualization	Spatial awareness	This is a picture of a dog and we are going to try to make this picture with these pieces. Can you try to join the pieces together to make this picture?	No
	VC	Verbal counting	Number/operations	How high can you count?	Yes
	Executive function	BDS	Backward digit span	Working memory	I am going to say a list of numbers, one after another. After you hear the numbers, I want you to repeat them after me backwards.
FDS		Forward digit span	Working memory	I am going to say a list of numbers, one after another. After you hear the numbers, I want you to repeat them after me in the same order.	Yes
HTKS		Head, toes, knees, shoulders	Inhibitory control	Please do the opposite of what I say: When I say touch your head, you touch your toes instead.	Yes
Social-emotional development	PTE	Perspective-taking/empathy	Social/emotional	Imagine this girl in the picture is a friend and she fell down and she is hurt. How do you think she is feeling right now? What would you do to help her feel better?	Yes
	UF	Understanding feelings	Social/emotional	Please tell me what makes you feel sad or want to cry? Please tell me what kinds of things make you feel happy?	Yes
	GEN	General knowledge	General knowledge	What is the name of the neighborhood/ community/village that you live in?	No

Teacher/Caregiver Assessment Tasks

DOMAIN	ABBREVIATION	TASKS	CONSTRUCT	SAMPLE <i>(exact wording can be accessed in MELQO portal)</i>	MAINTAINED IN CORE VERSION?	
Early literacy skills	EV	Expressive vocabulary	Expressive language	Can (name) communicate his/her own needs/what s/he wants in a way understandable to adults and peers?	Yes	
	LNK	Letter naming knowledge	Alphabet knowledge	Can (name) identify at least 10 letters of the alphabet?	Yes	
	LSC	Listening story comprehension	Language comprehension	Can (name) understand on first try what is being said to	Yes	
	him/her?	No	Phonemic awareness	What is the first sound in "mouse"?	Yes	
	MOT	Motivation to Read	Motivation	Would you say (name) is interested in reading (inquisitive/curious about the meaning of printed material)?	No	
	NW	Name writing	Print knowledge	Can (name) write his/her own name?	Yes	
	Early mathematics skills	MA	Mental addition	Number/operations	Can (name) add three and two together?	No
MV1		Measurement vocabulary, set 1	Measurement	Does (name) know that a giraffe is taller than a cat?	Yes	
MV2		Measurement vocabulary, set 2	Measurement	Can (name) know that an elephant weighs more than a mouse?	Yes	
NI		Number identification	Number/operations	Can (name) identify written numerals up to 5? (If yes ask	Yes	
10. If yes again, then ask 20.)		No	Number/operations	I will point to a number and I'd like you to tell me which one it is.	Yes	
PAS		Producing a Set	Number/operations	Can (name) count 10 objects?	Yes	
QCTS		Quantitative Comparison of Two Sets	Number/operations	Does (name) know that 8 is more than 2?	Yes	
SN		Shape naming	Spatial awareness	Can (name) name shapes like circles, triangles and squares	Yes	
SV		Spatial vocabulary	Spatial awareness	Does (name) understand phrases like next to and in front of?	No	
SVI		Spatial visualization	Spatial awareness	Can (name) complete a five-piece puzzle?	No	
TC		Time concepts	Measurement	Does (name) understand the concepts of: Today, Yesterday, Tomorrow?	Yes	
VC		Verbal counting	Number/operations	How high can (name) count?	No	
Social-emotional development		SR-SE	Self-regulation/socio-emotional	Socio-emotional	How often does child follow through on instructions? For example, does (name) finish his/her chores?	Yes