E L P
EARLY LEARNING PARTNERSHIP

Measuring the quality of early learning programs

This brief is intended to provide basic information on measuring the quality of early learning programs using a range of tools and approaches.

WHY IS IT IMPORTANT TO MEASURE THE QUALITY OF EARLY LEARNING PROGRAMS?

Investment in early childhood programming is increasing in response to convincing evidence on the benefits of supporting the social, emotional, and cognitive development of young children. As countries around the world work to expand access to early learning opportunities, it is critical to ensure the quality of both the services and children's experiences. High-quality programs can improve outcomes for children and set them on a positive trajectory in life. Low-quality programs, in contrast, are unlikely to generate the desired outcomes and can even be detrimental to children's development.[1]

HOW DO YOU MEASURE THE QUALITY OF EARLY LEARNING?

High-quality early learning can be delivered in a range of settings, including formal preschools, community-based programs, and home-based child care. There is no single blueprint for achieving quality under these different models; however, there are common elements of quality to be considered across preschool settings.

It is important to measure both structural and process aspects of quality. Process quality refers to a child's day-to-day experiences in early learning settings and encompasses dynamic elements such as interactions with teachers, peers, and materials, the quality of daily routines, and the

implementation of the curriculum. Structural elements, in contrast, include a center's infrastructure and materials, health and safety aspects, characteristics of the relevant groups of children and teachers (such as group size and student/teacher ratio), and caregiver characteristics (such as teachers' level of education, experience, and salary). (See Table 1.)

TABLE 1 **Examples of structural and process variables[2]**

STRUCTURAL VARIABLES	Physical environment	Infrastructure, availability of equipment and materials, health and safety
	Group characteristics	Adult-child ratios, group size
	Caregiver variables	Initial education, training, mentoring/supervision, salary
PROCESS VARIABLES	Interactions	Caregiver-child and child-child interactions
	Program characteristics	Curriculum, quality of daily routines

Some of the common dimensions that are considered when measuring quality include the following (note that the dimensions can be defined and grouped in different ways):

Physical environment: The physical space of the classroom and school, including space for each child, the characteristics of play areas and furnishings, the availability of materials, and the safety of the facilities. This dimension often includes cleanliness and access to potable water and toilets.

Teaching and learning processes: This dimension, which can be called or include pedagogy, play, or teacher-child interactions, covers the approaches teachers take to engage and teach children (such as individualized learning vs. group learning), and especially the type and quality of interactions among teachers, children, and the environment. It can include

- *Play,* the emphasis of the program on creating opportunities for children to engage in free and group play; adequate toys and spaces to play; and
- *Teacher-child interactions,* the type and quality of interactions between teachers and children as well as among children.

Curriculum: This dimension measures the extent to which curriculum content addresses children's physical, socio-emotional, linguistic, and cognitive development needs and stimulates early literacy and numeracy skills.

Teacher and school characteristics: This dimension covers the amount and type of training, professional development, support, and compensation teachers receive; teachers' knowledge of child development principles; and whether the number of trained staff is sufficient to maintain appropriate teacher/child ratios for the age groups in the classroom. It also sometimes includes

- *Leadership:* The leadership skills (such as management, administration, communication, support and guidance) of principals, directors, and administrators; and

Inclusion: The extent to which the classroom is able to support participation by all children, including those with disabilities and special needs.

Family and community engagement: This dimension reflects the degree to which families and communities are able to engage in children's education and the life of the school.

Table 2 provides examples of constructs and items used to measure quality early learning.

TABLE 2 **Examples of quality constructs and items**

CONSTRUCT	SUB-CONSTRUCT	ITEM
Physical environment	Organization of physical space and materials	• Does each child have access to his/her own learning materials, like pencils and paper? • Are there places for children to play within classrooms? • Are there enough seats for all children?
	Cleanliness	• Does the environment promote good health practices (e.g., personal hygiene, including hand washing? • Do children have access to clean water and sanitation?
Play	Materials and opportunities for play	• Do all children have access to free time for play? • Do all children have access to materials for play inside the classroom?
Teacher-child interactions	Positive interactions	• Does the teacher smile, clap, or offer positive words of praise for children's efforts? • Are children allowed/encouraged to interact with one another?
Teacher characteristics	Knowledge, skills, qualifications of personnel	• Do educators/caregivers demonstrate knowledge of child growth, development, and learning, and if so, are they able to apply this knowledge to practice?
Family and community engagement	Engaging families and communities in program	• Are parent workshops on education matters held?

Source: These items and variations appear in a number of the assessments included in Annex 1.

What are the different methods available to collect data?

Because of the complex nature of measuring quality, various forms of data collection are used. There are three main mechanisms: direct observation, self-reporting by qualified informants, and reviews of existing documentation. Utilizing more than one mechanism can provide more information and improve the precision of quality measurement, and some measurement tools combine different mechanisms.

Direct observation: Direct observation is the best way to effectively capture the interactions and experiences at a school or center. This method requires more training, resources, and time than other methods, since observers must be well-trained to collect information on specific elements of quality defined in the instrument and usually spend more than an hour in a given school. Videotaping or photography can be used to capture information that can be coded and interpreted at a later time and—in some cases—by another person (often with a higher degree of training and expertise).

Self-reporting by qualified informant: Structured interviews or surveys are used in this method to gather information on the dimensions of interest. Informants can be teachers, school administrators, parents, or children. Examples of information that can be acquired from teachers or administrators in this way include the percentage of students with disabilities and the level of education required to be a teacher in the school.

Review of documentation: This kind of review can be useful to gather information on structural aspects of the early learning environment. Examples of documents to review include curricula, classroom guidelines, and safety protocols.

How do we score results?

A variety of different scoring methods can be used to evaluate the data you have collected using the methods described above. These include checklists, Likert scales, and time sampling.

Checklists are fast and easy methods for collecting data. They mainly use yes-or-no options, and can be used to record observations for individuals, groups, or a whole class.

Likert scales are rating scales that usually offer four or more response options, allowing enumerators to represent a range of performance levels. Effective Likert scales use descriptors with clear options, such as frequencies (e.g., "frequently, sometimes, rarely, never"), and avoid subjective or relative descriptors, such as "fair" or "good."

Time sampling is a method by which enumerators repeatedly evaluate the setting in specific time increments (e.g., one minute, 10 minutes) in order to measure aspects of quality throughout the day.

How do we select an instrument?

Defining why you are interested in measuring quality will help you determine the type of measurement tool you need. For example, your purpose could be to assess the quality of an intervention that is being evaluated to better understand its impact. The purpose of measuring quality could also be to identify critical areas of improvement and resources needed, or to inform key policy decisions (including budget, staffing, and licensing), or to contribute to quality enhancement processes (such as for accreditation). The purpose of your measurement will have implications for the resources, time, and stakeholders you will need to conduct the assessment.

It is important to keep in mind that the best tools to measure quality will have reliable and valid measures (see Box 1). For example, when adapting items to the cultural context, you cannot assume that the reliability and validity of the original items will be automatically passed on to the adapted items. Instead, you will likely need expert guidance to ensure the reliability and validity of the modified tool. It is recommended to document reliability before and during fieldwork to ensure that scores are consistent over time.

Annex 1 provides examples of other commonly used tools, the dimensions they measure, and the countries where they have been used, among other details. All the tools featured there have been validated, used in at least one developing country, and adapted on at least one occasion to be culturally appropriate. They are all used to help identify what improvements are needed in preschool settings and can track improvements in quality when changes are made.

Annex 2 shares information on the Measure of Early Learning

Environments (MELE) module, which was produced by the Measuring Early Learning Quality and Outcomes (MELQO) initiative. This tool includes items and technical guidance to support countries in measuring the quality of early learning.

What are the key considerations when choosing an instrument?[3]

Following are the key considerations that matter when one is choosing an instrument to measure the quality of early learning programs.

Purpose. What is the purpose of your measurement exercise? Has the instrument been used before for the purpose you've defined?

Adaptation. Was the tool designed for use in the country or region where you are working, or will you need to adapt the tool to country context, culture, or language—or all three?

Cost. How much does the tool cost? Plan for training, data collection, in-country workshops or meetings, and, sometimes, adapting and contextualizing tools to fit local settings. Even when the cost of an assessment is low or free, implementation costs can add up, especially in the case of direct observation.

Training. What level of training is required to administer the tool? Some tools require that interviewers complete a specific (and sometimes costly) training as well as pass reliability tests (such as the Classroom Assessment Scoring System or CLASS; see Annex 1) whereas others do not require formal training.

Repeatability. Will your project have funds to use the same tool more than once to track progress? Conducting the same assessment repeatedly can yield valuable information on trends in quality.

How do we interpret and use the findings?

Once you have identified what needs to be improved using the right tools, it is important to discuss your findings with all the relevant stakeholders to explore options for making improvements in early learning. Your findings can be used by policymakers, teachers, and other stakeholders to select priority areas and determine how to improve the quality of settings, for example. In that case, the elements to be improved might be expanding and improving the physical space, providing additional classroom materials, integrating nutrition and health programs, improving the management of early childhood care and education programs, increasing teacher qualifications and training, or for strengthening connections with parents and community members—or some combination of these elements.

How can we integrate quality measurements into a national monitoring system?

Ideally, early learning quality measurement will become integrated into a national monitoring system to track changes in quality over time; inform standard setting, curricula, and teacher training; and influence decisions about resource allocation for quality improvement. Key considerations include the following.

Having a designated government agency. There should be a unit in government that is intimately involved in the design and implementation of the first quality measurement effort, with capacity, willingness and mandate, to maintain ongoing quality measurement efforts.

BOX 2

RELIABILITY & VALIDITY

Reliability refers to the extent to which a test will consistently provide similar scores when and if administered to a child or group of children over time.

Validity refers to how well a test or assessment measures what it intends to measure.

Concurrent validity refers to the extent to which the results of measurement correspond to those of a previously established measurement for the same construct.

Predictive validity refers to the extent to which results of a test are related to later performance that the test was designed to predict.

Having personnel on hand to collect data. Who is available in country to collect data regularly over time? This could be local education officials who already inspect education facilities (e.g. district education officers) or, in some cases, local university graduate students with expertise in child development.

Ensuring that new data can be integrated. What are school inspections covering at present and how can additional quality measures or data collection processes be integrated most effectively? Similarly, it is important to know what information system is currently available and in use in the country. Can indicators of preschool classroom quality be integrated into school census efforts that are ongoing and channeled into a national education management information systems (EMIS)? Are other related quality measurement efforts routinely carried out with which you could align yours?

How much will measuring quality cost and how long will it take? [4]

Budgets and timelines can vary significantly depending on the country context, the tool being used, sample size, data collection methodology, and, especially, the purpose of measurement. For example, measurement for a project evaluation could be less expensive than measurement for a nationally representative study with a larger sample size. In thinking about cost, it is important to consider expenses related to licenses for tools, adaptation workshops and pilot testing, translation, fees for approval of modified instruments (where relevant), and training and salaries of enumerators/assessors (which varies depending on their level of training and qualifications), as well as data collection, analysis, and dissemination.

Measuring the quality of settings can take time. Table 3 shows an example of a daily data collection schedule from Ecuador.

TABLE 3 **Example of daily data collection[5]**

HOUR	OBSERVER	SUPERVISOR
7:30	Arrival and preparation, selection of group to be studied	
8:00-12:00	Filming/coding instrument(s)	Structural quality questionnaire
12:00-14:00	Caregiver interview	
14:00-15:00	Coordinator interview	

Table 4 shows the types of budget items needed to prepare for and measure quality, with examples from both a project evaluation and a nationally representative study.

TABLE 4 Sample budget

	TASK	ESTIMATE OF TIME NEEDED	COUNTRY A (PROJECT EVALUATION)	COUNTRY B (NATIONALLY REPRESENTATIVE STUDY)
PREPARATION	Preliminary meeting	1–4 days, including planning and execution	US$500	US$1,000
ADAPTATION	Expert time for adaptation[a]	10 days (often ~US$400–600/day)	US$4,000 – US$6,000	US$5,000 – US$7,000
	Local meeting(s) (including per diems, space, pre-piloting)	5 days	US$1,000 – US$3,000	US$2,000 – US$5,000
	Translation of tools	Depends on language, length of tool	US$2,000 – US$4,000	US$2,000 – US$4,000
TRAINING	International expert time for training (master trainers)	10 days (often ~US$400–600/day)	US$3,000 – US$5,000	US$4,000 – US$7,000
	Travel (including experts, TTL)	1 week mission for each participant	US$4,000 – US$20,000	US$4,000 – US$20,000
	Space rental, materials, etc.	Depends on country	US$500 – US$2,000	US$500 – US$2,000
DATA COLLECTION, ANALYSIS, AND DISSEMINATION	Data collection (transport, enumerators' salary/per diem,[b] etc.), materials (tablets, questionnaires, etc.)	Depends on country	US$30,000 – US$45,000 (sample size[c] = 80 classrooms)	US$120,000 – US$200,000[d] (sample size = 600 classrooms)
	Data analysis[e] and report writing	~4–6 weeks	US$4,000 – US$10,000	US$5,000 – US$20,000
	Dissemination	1 event (US$1,000–3,000)	US$1,000	US$2,000
OTHER	General TTL travel	1–4 missions in one year	US$5,000 – US$20,000	US$5,000 – US$20,000
TOTAL			**US$55,000 – US$126,500**	**US$150,500 – US$288,000**

[a] Includes time to update tools and prepare manual, if needed.

[b] Depends on enumerators' level of training required.

[c] Note this is NOT a representative sample.

[d] Note this amount can vary *greatly* depending on the country context, transportation costs, etc.

[e] Can include psychometric analysis, statistical analysis, etc.

Table 5 is an approximate timeline, meant to illustrate the steps needed to undertake the measurement process and how this would be sequenced over time. In reality, depending on the specifics of the quality measurement objective, the steps could be combined or expanded and could happen more or less quickly than noted in the table.

TABLE 5 Approximate timeline		MONTH 1	MONTH 2	MONTH 3
LAUNCH AND PLANNING	Hold meeting(s) to identify main research questions of government (local/regional/global experts)	X		
	Appoint task force or focal point from government	X		
	Review existing curriculum and service delivery standards and align with assessment domains		X	
	Propose study design based on government priorities		X	X
	Map out project timeline and budget		X	X
	Write and release ToR for data collection			
	Get ethical approval as needed			
	Gather necessary background data for sampling purposes			
	Select contractors and finalize terms of contract			
ADAPTATION AND PRE-FIELD TESTING	Translate and back-translate tools			
	Hold in-person meeting(s) with national experts (including curriculum developers, academics conducting ECD research, school inspectorate (or equivalent)) to adapt items/measures and align with curriculum			
	Test tools on small sample and further revise and adapt tools as needed			
PILOTING/ DATA COLLECTION	Train enumerators for data collection (training should include access to children and classrooms so enumerators can practice using the instruments)			
	Collect data			
	Clean data			
ANALYSIS AND SYNTHESIS OF RESULTS AND DISSEMINATION	Analyze findings			
	Synthesize findings in easily digestible reports for different audiences			
	Distill findings for policy makers based on current policy plans			
	Disseminate findings through meetings and other events with national policy makers, academics, civil society, donor partners, regional/district education officials, teachers/principals/school managers, and parents/general public			
	Ensure findings are accessible/understandable for teachers and parents, etc.			

MONTH 4	MONTH 5	MONTH 6	MONTH 7	MONTH 8	MONTH 9	MONTH 10	MONTH 11	MONTH 12
X								
X								
X								
X	X	X						
X								
	X							
		X	X					
			X					
				X	X			
					X			
						X		
							X	
							X	
								X
								X

HELPFUL TIPS

Distinguish between minimum and ideal standards.
Consider making a distinction between minimum standards and ideal standards. It is critical to identify the minimum health and safety standards to which any center must adhere to ensure children's basic safety and security (such as that each classroom has light, ventilation, and a clear exit). In some countries these requirements are called minimum standards. Then there are the kinds of things that are helpful to have in each classroom, but not absolutely essential (such as a chair for each child). Ensuring the balance between minimum and ideal standards can help ensure that providers who could provide more affordable programs of sufficient quality are not forced out of the market.

Understand the early childhood education context.
Before beginning the quality measurement, it is critical to understand key contextual variables. These include school location (urban/rural), the school system's organizational structure and human resources, dates and schedules of operation, daily routines, and the monitoring and evaluation system, among other variables.

Ensure that selected dimensions of quality reflect country priorities.
A country's priorities, upcoming policy decisions, and political economy issues should guide the selection of quality dimensions to measure to ensure that results will be useful for decision making. Even if this process takes additional time at the outset, it will pay off in the usefulness and relevance of the results.

Time quality measurement appropriately.
Start the measurement of quality about one month after the beginning of the school year, to ensure that teachers have time to adjust to new students and early kinks can get worked out.

Select tools that are feasible to implement.
When the main purpose of the measurement is monitoring, simpler quality measures are recommended. If financial or human resources are limited, simplified tools and adapted approaches can be used. For example, short rating scales can be used along with videotaping, which can later be used to code the more complex constructs.

Adapt tools.
The tools presented here can be applied in a number of settings but should be adapted to be fully responsive to local values and context. Adaptation may involve translation, modification of content, or adaptations to the process of administration.

Ensure that enumerators are appropriately trained.
Enumerators should have significant knowledge of child development and early childhood education settings and receive the necessary training in the specific quality measurement tool.

TOOL	DESCRIPTION	DOMAINS
Arnett Caregiver Interaction Scale (CIS)	For children ages 36 – 60 months	4 domains: sensitivity; harshness; detachment; and permissiveness (26 items)
CLASS (Classroom Assessment Scoring System) Pre-K	CLASS Pre-K is an observation-based tool that assesses classroom quality in settings for children ages 36 months to kindergarten. All observers are required to receive prior training and are encouraged to use videotape footage.	10 dimensions of classroom quality across these 3 domains: emotional support; classroom organization; and instruction support
Early Childhood Environment Rating Scale (ECERS-R)	Developed at the University of North Carolina at Chapel Hill, ECERS is designed to assess group programs for children ages 2 to 5. The tool is mostly used for policy development, program evaluation, improvement advocacy, and training. ECERS-R includes classroom observation and a teacher's interview from the enumerator. The enumerator or observer should receive prior training.	7 domains: space and furnishings; personal care routines (health and safety); language and reasoning; activities; interactions; program structure; and parents and staff.
Acei Global Guidelines Assessment (GGA)	The GGA is an instrument designed to help ECCE professionals systematically assess the quality of their programs. GGA is mostly used for self-assessment by centers, to design new early childhood programs, or to improve existing programs.	The assessment includes environment and physical space, curriculum content and pedagogy, educators and caregivers, partnerships with families and communities, and children with special needs.
International Step By Step (ISSA) Principles of Quality Pedagogy	Principles of Quality Pedagogy is a tool developed by ISSA to define quality in ECCE teaching practices and classroom environment and is primarily used for planning and improvement.	7 focus areas, which include: interactions; family and community: inclusion, diversity and values of democracy; assessment and planning; teaching strategies; learning environment; professional development.
Measure Early Learning Environments (MELE)	The MELE is used to measure the quality of early learning environments for children ages 3 to 6. It includes a classroom observation tool, teacher/director survey, and parent survey.	The MELE addresses environment and materials; teacher-child interactions; pedagogy and approaches to learning; family and community engagement; inclusion; and play.
Stallings Classroom Snapshot instrument (or Stanford Research Institute Classroom Observation System)	A tool to gather information on the interaction between teachers and students in the classroom.	The focus areas are teachers' use of instructional time; teachers' use of materials; core pedagogical practices; and teachers' ability to keep students engaged.

LENGTH	COST	COUNTRIES AND LANGUAGES	CONTACT
90 minutes: 2 observation cycles of 45 minutes, on separate occasions	Manual and score sheets are readily accessible and at no cost	Used in Bermuda and the United States	Instructions available at http://www.eec.state.ma.us/docs1/qris/20110121_arnett_scale.pdf
80 minutes (suggested): 4 classroom observations of 20 minutes each	Training costs vary from US$670 to US$1,500 per person for 2 to 5 days of training. Manuals cost US$50 each, and a pack of 10 scoring forms costs US$25.	Mainly used in the United States. Also used in Chile, Finland, and Portugal	Contact: Paul H. Brookes Publishing Co, www.brookespublishing.com
2 to 3 hours; contains 43 items and 7 subscales	US$19.95 for the manuals and US$59 for the training videos.	Used in Austria, Bangladesh, Canada, Chile, England, Germany, Greece, Hungary, Iceland, India, Italy, Korea, Portugal, Russia, Singapore, Spain, Sweden, and 7 Caribbean countries. A Spanish version of the tool is available.	http://ers.fpg.unc.edu/early-childhood-environment-rating-scale-ecers-r Contact: Richard Clifford, dick.clifford@unc.edu
N/A	Freely available	Used in more than 35 countries, including Canada, Guatemala, India, Kenya, Mexico, Peru, Sierra Leone, and Thailand. Available in Arabic, Chinese, English, French, Greek, Nepali, Russian, and Slovak.	For more information, visit http://acei.org/images/stories/GGAenglish.pdf and http://acei.org/acei-news/acei-global-guidelines-assessment.html Contact: Belinda Hardin, bjhardin@uncg.edu and Doris Bergen, bergend@muohio.edu
N/A	Freely available	Used in more than 27 countries, mostly in Eastern Europe, including Moldova, Slovenia, and Tajikistan.	For more information, visit: http://www.issa.nl/content/issa-quality-principles Contact: Liana GHENT, lghent@issa.hu
Depends on how module is adapted by each country	Free	Used in Nicaragua, Tanzania	Contact: Lucy Bassett lbassett@worldbank.org; Abbie Raikes araikes@unicef.org; Kate Anderson klanderson@brookings.edu
10 "snapshot" observations of 15 seconds each, over the course of one class period.	N/A	Used in Brazil, Colombia, Honduras, Jamaica, Mexico, Peru, and the United States, among others	

The Measure of Early Learning Environments (MELE) was developed by the Measuring Early Learning Quality and Outcomes (MELQO)[7] initiative led by the World Bank, UNICEF, UNESCO, and the Center for Universal Education at the Brookings Institution; it engaged experts and partners from around the world. The MELE was developed through a consultative process, drawing on the best experiences and tools for measuring the quality of early learning environments to date, and its items will share similarities with items from many of the tools profiled in Annex 2.

Given the state of evidence on quality in early learning environments and the strong cultural influences on what defines "good quality"—the number of items considered universally cross-culturally relevant was small—the MELE focuses on constructs rather than on specific items. A range of items are available for countries to select from, based on their own notion of quality.

The MELE includes a classroom observation tool, a teacher and director survey, and a parent survey. Quality is measured through six constructs:

- **Environments and materials:** Safety and cleanliness of the physical space; access to clean drinking water, nutritious meals, and adequate sanitation facilities; and whether a variety of culturally relevant and meaningful learning materials are available (such as visual displays, books, art supplies, and musical instruments).

- **Teacher–child interactions:** Children experience daily interaction with teachers and school staff who are nurturing, emotionally supportive, trained in pedagogy and early childhood development (ECD), and attuned to children's individual needs.

- **Pedagogy and Approaches to learning:** Child-centered teaching encourages curiosity, persistence, attentiveness, cooperation, participation and active engagement; children engage in age-appropriate play, activities and routines; curriculum addresses children's physical, socio-emotional, linguistic, and cognitive development needs and stimulates early literacy and numeracy skills.

- **Family and community engagement:** Programs share information, promote positive relationships, and create opportunities for parent and community engagement; families and the local community are actively involved in planning, decision making, and action to improve early childhood care and education (ECCE).

- **Inclusion:** All children and families have access to high-quality ECCE services; teachers speak the home language of the majority of students; teachers are trained in providing ECCE to children with disabilities and special needs and in fostering age-appropriate development for all children.

- **Play:** Children are given ample opportunity to explore and enjoy learning through play, with time for pretend play and interactions with peers.

Ideally, the MELE can be used to measure a nationally representative distribution of the quality of early learning environments (usually for children aged 3-6), which could then inform policy planning and budgeting, including: resource allocation, curriculum design and teacher training programs, early learning standards, and ongoing monitoring of quality.

The MELE module, along with manuals, guides, and other resources, is available upon request.

ANNEX 3 Resources

For more information on instruments or tools to measure the quality of early learning settings, please consult:

Bouguen, A., D. Filmer, K. Macours, and S. Naudeau. 2013. *Impact Evaluation of Three Types of Early Childhood Development Interventions in Cambodia*. Policy Research working paper no. WPS 6540, Impact Evaluation series no. IE 97. Washington, DC: World Bank.

Britto, P. R., H. Yoshikawa, and K. Boller. 2011. "Quality of Early Childhood Development Programs in Global Contexts: Rationale for Investment, Conceptual Framework and Implications for Equity," *Social Policy Report* (Society for Research in Child Development), vol. 25, no. 2

López Boo, F., M. Caridad Araujo, and R. Tomé. 2016. *How Is Child Care Quality Measured? A Toolkit*. Washington, DC: Inter-American Development Bank.

Naudeau, S., N. Kataoka, A. Valerio, M. J. Neuman, and L. K. Elder. 2011. *Investing in Young Children: An Early Childhood Development Guide for Policy Dialogue and Project Preparation.* Washington, DC: World Bank.

Neuman, M. J., and A. Devercelli. 2013. *What Matters Most for Early Childhood Development: A Framework Paper.* SABER (Systems Approach for Better Education Results) working paper no. 5. Washington, DC: World Bank.

Raikes, A. 2014. *Early Childhood Care and Education: Addressing Quality in Formal Pre-Primary Learning Environments*. Paris: UNESCO.

UNICEF. 2014. *A Framework and Tool Box for Monitoring and Improving Quality* (draft). ECD Framework PART II (2012). (Accessed online August 2014.)

World Bank. 2015. *User Guide – Conducting Classroom Observations Analyzing Classroom Dynamics and Instructional Time Using the Stallings "Classroom Snapshot" Observation System."* Washington, DC.

Young, M., and L. Richardson, eds. 2007. *Early Child Development: From Measurement to Action–A Priority for Growth and Equity.* Washington, DC: World Bank.

NOTES

[1] Bouguen et al. 2013.

[2] Adapted from Neuman and Devercelli 2013 and Naudeau et al. 2011.

[3] These considerations were compiled based on anecdotal findings from the experience pre-piloting MELE, the MELQO tool, specifically through conversations with MELQO's Technical Advisory Group for Child Development and Early Learning Outcomes.

[4] This section on budget and timeline was also compiled based on experience pre-piloting the MELQO.

[5] The example in Table 3 is from López Boo, Araujo, and Tomé 2016.

[6] Information on MELE and MELQO in this annex appears thanks to the MELQO Core Team and Technical Advisory Group.

[7] Note that the MELQO initiative also developed a module, called the Measure of Development of Early Learning (MODEL), to assess child development and learning. The two modules (MODEL and MELE) are complementary and are both meant to be adapted to align with national systems and standards and to be used to inform policy decisions to improve early childhood development.