

The Measuring Early Learning Quality & Outcomes initiative: purpose, process and results

Abbie Raikes, Rebecca Sayre, Dawn Davis, Kate Anderson, Marilou Hyson, Evelyn Seminario & Anna Burton

To cite this article: Abbie Raikes, Rebecca Sayre, Dawn Davis, Kate Anderson, Marilou Hyson, Evelyn Seminario & Anna Burton (2019) The Measuring Early Learning Quality & Outcomes initiative: purpose, process and results, *Early Years*, 39:4, 360-375, DOI: [10.1080/09575146.2019.1669142](https://doi.org/10.1080/09575146.2019.1669142)

To link to this article: <https://doi.org/10.1080/09575146.2019.1669142>



View supplementary material [↗](#)



Published online: 29 Sep 2019.



Submit your article to this journal [↗](#)



Article views: 128




View related articles [↗](#)



View Crossmark data [↗](#)



The Measuring Early Learning Quality & Outcomes initiative: purpose, process and results

Abbie Raikes ^a, Rebecca Sayre^b, Dawn Davis^c, Kate Anderson^b, Marilou Hyson^d, Evelyn Seminario^d and Anna Burton^e

^aECD Measure, University of Nebraska Medical Center; ^bECD Measure/ Unbounded Associates, Washington, DC, USA; ^cECD Measure, University of Nebraska, Lincoln, NE; ^dIndependent Consultant, Philadelphia, PA, USA; ^eCollege of Arts and Human Sciences, University of Nebraska, Lincoln, NE, 68588-0312, USA

ABSTRACT

Measuring Early Learning Quality & Outcomes (MELQO) was initiated to address needs for child development and quality of early childhood education (ECE) data, specifically for low- and middle-income countries. Drawing from existing tools, MELQO convened a consortium to create open-source tools to be adapted to national contexts, simultaneously informing global and national population-level monitoring. Beginning with the rationale and context for creating MELQO, we outline the process of creating the tools; highlight results from psychometric evaluation; and describe how the data have been used to reveal patterns of inequity and levels of learning and classroom quality. While psychometric analyses of MELQO indicate that scores should not be used to compare quality of child development between countries, country experiences suggest that data from MELQO tools within countries identify priorities for improvement of preprimary classrooms. MELQO also highlights the importance of local leadership to generate high-leverage data on ECE.

ARTICLE HISTORY

Received 10 June 2018
Accepted 15 September 2019


KEYWORDS

International Early childhood development; measurement and monitoring; early childhood education

Introduction

The adoption of the United Nations' Sustainable Development Goals (SDGs), a set of 17 interconnected goals to promote sustainable development and equity within and across countries, ushered in a new era of global development. The SDGs were adopted by all 193 member states of the United Nations and include a target (Target 4.2) on early childhood development (ECD) as part of Goal 4 on lifelong learning. Target 4.2 reads, 'By 2030, ensure that all girls and boys have access to quality early childhood development, care and preprimary education so that they are ready for primary education.' The inclusion of Target 4.2 was spurred by a strong advocacy effort by early child development stakeholders, government ministers and researchers, and reflects the strength of the science demonstrating the significance of early childhood for lifelong health and well-being (e.g. Black et al. 2017).

CONTACT Abbie Raikes  abbie.raikes@unmc.edu  College of Public Health, University of Nebraska Medical Center, Omaha, NE 68198, USA

 Supplemental data for this article can be accessed [here](#).

© 2019 TACTYC

Yet one potential threat to the inclusion of ECD in the SDGs was the lack of global monitoring tools (UNESCO 2017¹). In the rush to include numerous important aspects of development across all sectors, targets without clear measurement plans risked being excluded from the agenda (SDSN 2015). Because there are no legal mechanisms to enforce adherence to United Nations' agendas, countries' voluntary reporting on progress towards global and national indicators across the 17 goals serves as a primary mechanism for encouraging country participation in the SDGs and in building a cohesive global effort to achieve the SDGs (Raikes et al. 2017). Global SDG monitoring requires data that are valid, relevant and comparable across countries, with data produced by tools that are relevant and administered in the same way across all settings. National monitoring, which is also part of the SDGs, is not intended to produce comparable data, and thus has a wider and more open-ended scope that is intended to be more responsive to country needs (de la Mothe, Espey, and Schmidt-Traub 2015).

In 2014, as discussions on the Sustainable Development agenda were beginning, the Measuring Early Learning Quality & Outcomes (MELQO) project was initiated to help generate actionable, feasible data on ECE and the quality of children's learning environments for national and global monitoring (UNESCO, UNICEF, Brookings Institution, and the World Bank 2017). While Target 4.2 is focused on early childhood development from birth through the start of school, MELQO focused only on early childhood education (ECE) for children roughly between four and 6 years, to begin with a smaller scope. Led by the Brookings Institution, UNICEF, UNESCO and World Bank, MELQO convened more than 60 experts, country stakeholders, and representatives of multilateral organizations. Now 5 years since its initiation, MELQO's contribution to global and national early childhood systems can be viewed from three perspectives: first, by examining the psychometric functioning of the tools; next, by assessing the impact of the data on ECE policies and programs, in particular the ability of data resulting from MELQO to identify inequities; and third, by its contribution to the wider field of ECE researchers, program evaluators and implementors, and other stakeholders who are interested in measurement in ECE. Below we outline the rationale for initiating MELQO, followed by a description of the development process. We offer an evaluation of the tools' psychometric properties; describe how stakeholders have used the resulting data to influence policy and practice; and reflect on the impact of MELQO within the broader early childhood community. We end with a discussion of lessons learned from MELQO and future directions for addressing measurement of child development and quality at the national level in low- and middle-income countries.

Measurement of early childhood development

Even before the start of the SDGs, ECD measurement took place in many countries through research studies and routine monitoring of child outcomes, classroom quality, and/or teacher characteristics (such as levels of education and completion of training; OECD 2015), and with a variety of different measures (see Fernald et al. 2017 for a review). These tools were used for several purposes, including population-level monitoring, or tracking of trends across a population of children to monitor equity within and across populations, often to inform policies; program evaluation, or to document the impacts of interventions; and research.

When MELQO began in 2014, at least 10 measures of child development and learning at the start of primary school and the quality of ECE environments had been developed and used in low- and middle-income countries, mostly for program evaluation or research, but some for national-level monitoring as well (a list of measures appears in UNESCO, UNICEF, Brookings Institution, and the World Bank 2017). For use at the population level, UNICEF had developed an initial version of the MICS Early Childhood Development Index (ECDI), a household survey that relies on caregiver-report of child development. Several population-level tools designed for regional or programmatic use were in development or in use in 2014, including the Early Development Instrument (EDI; Janus and Offord 2007), the International Development and Early Learning Assessment (IDELA; Pisani, Borisova, and Dowd 2015), the East Asia Pacific Child Development Scales (EAP-CDS; Rao et al. 2019), the Early Learning Outcomes Measure, developed for use in South Africa (ELOM; Snelling et al. 2019), the PRIDI in Latin America (Verdisco et al. 2014), and other efforts (Raikes, Anderson, and Atinc 2014).

But there were at least two issues with the established measures and ways of using them. First, although many high-income countries regularly monitor program quality and/or child development for the purposes of quality improvement (Atchison et al. 2017), routine national-level measurement for the purposes of tracking trends in child development or quality of learning provisions was still relatively rare (for exceptions, see Canada's use of the EDI; Janus and Offord 2007 and Australia's use of the AEDI for an example of census-level monitoring of child development; Brinkman et al. 2007, 2014). Yet generating global and national data on ECE plays an important role in ensuring equity in education, by providing information on the extent to which education systems are achieving stated goals of reducing gaps in learning due to region, gender, and family income, among other sources of inequity. Second, few attempts had been made to look across these measures to identify common elements that could inform both global and national monitoring. The array of measures led to a range of choices for governments, but this array also precluded the possibility of using the data to inform regional or global monitoring in a cohesive manner.

MELQO's rationale and process

MELQO was initiated as the SDGs were emerging, in response to the recognition that measurement is an essential element of a new global agenda. Several multi-stakeholder initiatives such as the Learning Metrics Task Force (LMTF) and the Sustainable Development Solutions Network (SDSN) called for the inclusion of indicators of early development and school readiness in the SDGs, based on extensive expert and practitioner consultations (LMTF 2013; SDSN 2015). The idea behind MELQO was that a small 'common core' of items could be identified to inform global monitoring, supplemented by a larger group of items with potential relevance to cultural and contextual influences within each country. By building on existing measurement tools and developing a common core, the MELQO Consortium hoped to harmonize rather than replace the national, regional and global efforts that were already gaining momentum. The Consortium was interested in promoting measurement to: 1) ensure that early childhood remained in the SDGs by generating options for ECD measurement; 2) harmonize the many tools of child development and learning, and quality of learning environments;

and 3) promote data-driven systems that would lead to more equitable and higher-quality ECE. The expectation was that the resulting tools would be open-source and freely available, thus accelerating measurement of Target 4.2 at the global and national levels. While the final language of SDG 4.2 indicators focused on measuring access to preprimary education and child development and learning ('the proportion of children who are developmentally on-track'), MELQO was designed to integrate both 1) quality of learning environments and 2) child development, and learning to ensure that, beyond access, there was sufficient emphasis on whether preprimary environments were high enough quality to promote children's development and learning.

Who was involved in MELQO?

A large group of people representing non-profit organizations, governments, universities, and multi-lateral organizations was involved in creating the tools, identified through their engagement with early childhood projects around the world (see UNESCO, UNICEF, Brookings Institution, and the World Bank 2017, for a complete list of participants). A smaller group of experts led development of the work to develop tools and items, which were reviewed by the MELQO Consortium. Representatives from governments and NGOs in the pilot countries helped adapt, test, and refine the tools and items. MELQO's first meeting, in November 2014, drew together representatives of the host institutions and a small group of researchers from various parts of the world to provide input on the goals and process for MELQO. An independent evaluation (Salzano and Kjaersgaard 2016) found that MELQO was able to mobilize this broad group of experts, researchers and practitioners in the international community, including some of the major players in the field of ECE, due to their shared interest and perceived need for ECE measurement. Building the consortium generated enthusiasm for the work and leveraged the technical, implementation, and convening capacity of the lead agencies, partner organizations, and participating governments.

What did MELQO do?

MELQO first developed modules, or groups of measurement items, for child development and learning, the area in which most of the progress in developing global and regional tools had been made. Members of the MELQO Consortium initiated a detailed review of items and measures that had been used in more than one country, ideally in low- and middle-income settings. It was apparent from the initial review that many measures of child development and learning relied on similar items, suggesting that MELQO could identify a common core of items from existing measures. This was the starting point for developing a common core, which was then translated into a suite of tools to measure child development and learning.

While the child development and learning tools were being finalized, a separate working group on quality of learning environments in programs, schools, and centers convened to review the items and measures that had been used in more than one country. There was interest in including both the structural elements of quality, such as teacher qualifications and ratios, and the elements of process quality, or children's experiences in classrooms with teachers, peers, and engagement with materials (e.g. Burchinal 2018). It was clear that there were some ECE observational quality measures with applicability across countries (e.g. Aboud 2008; Yoshikawa et al. 2015) and some

cross-country research on quality and child development (Montie, Xiang, and Schweinhart 2006), but the items and scales had a smaller evidence base across low- and middle-income countries and the measures were wider and more disparate in content and structure than child development measures, and there was more evidence from some parts of the world than others. As a result, the development of the quality scale was informed by theories of child development and learning, in addition to drawing upon existing measures of ECE quality, and evidence from cross-country work.

The final set of MELQO tools developed by the Consortium consists of two suites of tools: Measure of Early Development and Learning (MODEL) and Measure of Early Learning Environments (MELE). The MODEL tools include a child direct assessment, parent report of child development and teacher report of child development. The MELE tools include a classroom observation, teacher interview, parent interview, and director interview (see UNESCO, UNICEF, Brookings Institution, and the World Bank 2017).

Was MELQO in pursuit of globally comparable measures?

A key question posed to the Consortium in the first phase of work was the desired and expected degree of cross-country comparability in measurement. Consortium members were asked whether they thought measures should prioritize global comparability, which would require using the same set of items in all countries; have a core set of items that would then allow country adaptation; compare only at the 'construct' level, meaning that constructs could be the same across countries, but items would vary; or attempt no degree of comparability, which would lead to an item bank only. Measures of child development and learning, and quality of learning environments, were envisioned as having two distinct paths towards comparability: whereas the domain of child development and learning was thought to potentially have a core set of comparable items that would be relevant across countries, it was not assumed that quality of learning environments could or should be measured in similar ways across countries. Rather, the consortium agreed that some constructs would apply across countries, but with items generated to reflect local context.

While the consortium had relatively little difficulty agreeing on this approach, it was in many respects an experiment. While many child assessments relied on similar items, at that point, little work had been done to look at item level functioning across countries, specifically whether items demonstrated measurement invariance, or the degree to which measures performed similarly across countries. Measurement of quality in ECE classrooms was even more nascent, with less research and fewer measures tested across countries, and less agreement in the field on how to define and measure quality. MELQO launched the tool development and testing phase with two shared goals across consortium members: 1) to use existing measures to generate a common core of items that could jointly serve to support national and global monitoring; and 2) to generate processes for country-level adaptation to ensure that measures would be aligned to country goals, contexts and priorities, and useful for informing policy.

Evaluating psychometric properties of MELQO tools

Ensuring measures meet adequate psychometric standards is essential for generating reliable data, especially when used to inform policy and practices. There is little work

outlining acceptable technical standards for measures intended for use at the population level, especially considering the SDGs and related policy interest in globally comparable data (Raikes et al. 2017). While a full evaluation of psychometric properties for population-level monitoring requires a range of dimensions (Yoshikawa et al. 2018), here we provide a brief summary of initial psychometric results.

Pilot and national level testing

Tools were tested in four pilot countries, beginning in 2015. Countries selected which measures they wanted, leading to a mix of measures used. MODEL results, which were not based on representative samples, demonstrated the technical challenges of moving towards global measurement of child development and learning. Items showed different levels of difficulty and concurrent validity, with some countries showing much stronger associations between child characteristics and family environments than others (World Bank 2017a). Since the data came from non-representative samples, however, caution should be used when interpreting results. MELE piloting took place through program evaluations in three countries. Results were used to refine the constructs through factor analyses (unpublished data).

The pilot-tested MELQO tools were used in several countries as part of nationally or regionally representative studies beginning in 2016. Implementing entities varied: studies were led by international research firms (i.e. RTI took the lead in Tanzania; see Weatherholt 2018, for a report on study findings) or small teams of consultants were hired to train data collectors and implement studies. The level of training and expertise among research teams varied considerably. Guidance was provided on the adaptation process, which included convening local stakeholders, reviewing national standards and goals for ECE, revising the tools for contextual or cultural relevance, and clarifying among stakeholders how the data would be used to inform policy and practice and data analyses and reporting. All materials, including the tools, tool manuals, and an implementation toolkit, were made publicly available in 2017, on the ecdmeasure.org website.

Psychometric evaluation

Results from these studies can be used to evaluate the two elements of psychometric properties outlined above, internal structure and relations with other variables. Analyses indicated that the factor structure of MODEL's direct assessment varied across countries. Upon failing to find a consistent factor structure of child development and learning items, further analyses to establish global comparability were not warranted. However, analyses indicated acceptable psychometric properties at the national level for the direct child assessment and the teacher-rated social/emotional scale (for example, in Tanzania, see Raikes et al. 2019). Analyses of MELE data indicated adequate internal consistency, evidenced by adherence to a hypothesized factor structure, but relations with child development and teacher characteristics are inconsistent (Raikes, et al., in preparation). In some countries, expected associations between MODEL scores on child development and learning and MELE scores on quality of classrooms emerge (e.g. Maldonado et al. 2019; Ogando et al. 2019).

What conclusions can be drawn from psychometric results?

MELQO results suggest that ECE measurement tools may function differently across countries. This conclusion is largely consistent with findings from other tools intended

to produce globally comparable data on child development and learning: for example, IDELA has also reported evidence of test functioning within but not across countries (Halpin et al. 2019). But the lack of global comparability may not be bad news: Drawbacks to globally comparable data on ECD have been identified (Urban 2018), and several national-level adaptations of MELQO have produced relevant and impactful data, as outlined below. The MELE findings are in large part consistent with those from high-income countries. Commonly used early childhood classroom observation scales, such as ECERS (Early Childhood Environmental Rating Scale), do not demonstrate consistent technical properties when used in high-income countries, such as the United States (Gordon et al. 2013; Early et al. 2018), yet still may contribute to the process of improving ECE, for example, by informing professional development (e.g. Burchinal 2018). It is also possible that the emphasis on beginning with global constructs obscured the local and more culturally relevant definitions of quality (e.g. Dahlberg, Moss, and Pence 2013) that would have led to measurement with greater relevance and sensitivity to local context. Finally, it is important to note that the level of training of data collectors, which differed between countries, may have contributed to the lack of consistency in findings.

Using MELQO data to generate nationally relevant evidence

Beyond psychometric properties, policy relevance and data use were identified as top goals for MELQO tools. We selected country examples from diverse parts of the world to demonstrate the process of adapting and implementing MELQO tools and using the data to inform policy and practice. As outlined below, countries chose to participate in MELQO based on interest within ministries and availability of funding².

Country demand for MELQO

At the heart of any data collection exercise is demand for data – at the national level, specifically the expectation that having more information will improve governments' abilities to address key policy and programmatic concerns. Data on quality of learning environments are often unavailable in low- and middle-income countries, and many rely on proxy variables such as teacher-child ratios, compliance with operating hours, and infrastructure standards to monitor quality, if quality is monitored at all (Anderson et al. 2017). Many countries interested in MELQO wanted data on the overall levels of learning and development among young children in preprimary or early primary grades, information on the routines, practices, and pedagogical approaches in preprimary classrooms and data on the extent to which national efforts to reform curricula or teacher training have resulted in improvements in classrooms or ideally, children's learning. As outlined below, some countries had specific questions on equity (i.e. whether classrooms in some parts of the country were better-resourced than others), and on the degree to which children's early learning and development aligned with national standards or curricula expectations for the preprimary and primary grades.

Ministries of Education in Tanzania and Zanzibar were the first to undertake a nationally representative study beginning in 2014. MELQO was initiated as the Government of Tanzania was undergoing preprimary policy reforms, including a new policy to universalize fee-free preprimary education as well as a launch of a new national

competency-based preprimary curriculum. The MELQO study provided baseline information to the government and early childhood stakeholders on school readiness and the quality of preprimary education in Tanzania.

In Indonesia, MELQO data were generated as part of a large-scale effort to evaluate a pilot teacher-training initiative supported by the World Bank and the Government of Indonesia. The MELE tool was adapted and piloted in poor rural districts to help the government understand the baseline quality of classrooms and the extent to which an improved, district-based teacher training system may influence the quality of teaching in village preschool classrooms.

In Peru, where MELQO was initiated to generate data on the quality of preprimary classrooms, the Ministry of Education began with measurement of classrooms in one major metropolitan area, which was used to adapt and test the national tools. The Ministry of Education was invested in expanding the monitoring and evaluation system and used the adapted MELQO tools in preprimary classrooms as part of a larger effort to document the quality and outcomes of the education system.

In Lesotho, MELQO was initiated as the first-ever effort to produce national-level data on child development and quality of learning environments. The Ministry of Education partnered with the World Bank to collect national MELQO data that could serve as a baseline for tracking progress towards goals for ECD; and would inform the selection of critical elements to include in the education management information system (EMIS). Lesotho was interested in building capacity within the Ministry and its partners to engage in more in-depth monitoring of child development and quality of learning environments. In addition, the Ministry of Education had recently developed and age-validated their Early Learning and Development Standards and planned to use the MELQO data to inform an upcoming policy and curriculum review.

Country adaptation of global ECE concepts

As outlined above, the MELQO Consortium began with the generation of global concepts of quality in preprimary learning environments and child development that the group believed would have relevance across settings. Yet the research was limited, as many countries were in the early stages of building monitoring systems that incorporated observations of early education settings (OECD 2015). As a result, MELQO provided an opportunity to evaluate the extent to which anticipated constructs would be deemed culturally and contextually relevant across countries. National quality service standards and curricular standards, when available, were considered central to assessing this alignment. The MELQO process recommended a careful review of the standards prior to meeting with country stakeholders, to provide an overview of points of alignment and mismatch for discussion. Countries vary in the format and development of standards, leading to different experiences aligning standards to the MELQO tools.

In Tanzania, for example, national standards were used as the basis for MELQO adaptation. The MODEL and MELE tools were aligned to specific competencies within the national preprimary curriculum. The tools aligned well with the domains measured by MELQO and slight adaptations were made to fit to context and culture (for example, stimuli depicting Tanzanian children in the child direct assessment). Several competencies in the curriculum, including hygiene, nutrition, and safety, were not directly covered in the MELQO instruments and were subsequently added to the Tanzania MELQO direct

assessment, as the government was interested in understanding how well children were learning about these curricular areas. In addition, the teacher and head teacher interviews were adapted to answer key policy questions related to teacher background and professional development.

In Indonesia, the concepts emphasized in the MELE instrument appeared to match well with the government's expectations for ECE quality. Because the Indonesian team intended to use the instrument within a large-scale pilot study of preschool quality with local observers, the MELE was pared down to focus on items that were more feasible to use across a large population, but still maintained a good deal of alignment with the core constructs. The MELE was mapped to the Indonesian national teacher competencies and appeared to align well with many of them. A national team of early childhood leaders helped make adaptations to some MELE items to align better with government expectations for what children should learn and how teachers should promote their progress. An item was also added to reflect the government's new expectations about the use of thematic curriculum.

Similarly, in Peru, the MELQO modules aligned well with the Ministry of Education's expectations for children's early learning and development. Each of the domains measured by MODEL was aligned with core competencies in the national preprimary curriculum. The Ministry found that even when certain competencies were not described in precisely the same language as MODEL items, the elements in the curriculum related well to what was measured within the MODEL. While the MELQO tools demonstrated alignment with national early learning priorities, the Ministry framed the MELQO study as an assessment of global child development competencies, not the curriculum, emphasizing that MELQO was a measure of child development based on international evidence on children's early learning rather than a curriculum-based test.

In Lesotho, the country had recently undertaken a process to develop Early Learning and Development Standards, which served as the basis for identifying elements of MELQO to use for child development monitoring. While the overall degree of alignment was adequate, it is important to note that some key areas of the national standards were difficult to include in MELQO – for example, children's knowledge of national symbols, local plants and cultural traditions were not included in the MELQO assessments of child development, and it was challenging to define how best to include them going forward. In addition, Lesotho's Education Policy required that they report children's progress towards meeting standards and the specific impact of malnutrition on child development. Because of those priorities, data were collected on children's nutritional and health status.

Beyond alignment with standards, both the MODEL and MELE tools have required adaptation to cultural contexts. The items included in the assessments include words, pictures, and letters or numbers that require adaptation (see MELQO Implementation Toolkit for further detail). Perhaps more significantly, some of the concepts included in MELQO have varying levels of relevance across settings – for example, play-based learning has proven difficult to define and measure in contexts, such as Ethiopia, where there is little reliance on play-based approaches in many classrooms (Rossiter et al. 2018). This suggests that more work is needed to examine the definitions and manifestations of various pedagogical approaches before moving to large-scale measurement. Some constructs, such as safety standards, may be globally relevant

(meaning that the idea of safety is universal) but locally defined, as what constitutes 'unsafe' in one location may not be the same in other locations. Moreover, the constructs defined in measurement may be aspirational, creating a new standard to work towards, but producing data with high degree of skew and few if any classrooms or children achieving mid- or high-end scores. Overall, national standards can serve as an important starting point for ensuring that the global tools are adapted to reflect national context. When no standards are available, MELQO can help inform the creation of the standards – but that presents a delicate balance, as measurement should ideally follow key policy constructs as opposed to driving them (e.g. Early et al. 2018).

The process of adapting measures and aligning to national standards can be significantly more effective when more stakeholders are involved. Establishing national taskforces or committees comprised of early childhood stakeholders facilitates opportunities for inputs, as well as buy-in, from local experts. In addition, involving stakeholders at all levels of the early childhood system can ensure the tools reflect the reality of local classrooms. In early phases of MELQO, teachers were not consistently included in the adaptation process. Now we make a recommendation to include teachers and use qualitative methods to benefit from their insights on the definitions and manifestations of quality.

How can MELQO data help improve ECE?

Two themes emerge when looking across countries to see how MELQO data have been used for policy application. First, MELQO data highlight inequities, both in terms of influences on child development and learning in the preprimary years and the quality of children's learning environments. Second, despite the lack of cross-country comparability of scores resulting from the measures, the descriptive data from MELQO studies can be used to point towards specific and actionable areas for improvement within ECE systems.

Uncovering patterns of inequity

MELQO data from several countries have documented the profound influence of family background and income on child development. In Tanzania, Brazil, and Cambodia, for example, inequities due to wealth, region, and family background emerged from MELQO data. In Tanzania, using the MELE showed that children living in rural areas were substantially more likely to receive lower-quality ECE (RTI 2018); in Cambodia, using a battery of MELQO items supplemented with other tests, wealth gradients in child development by age 3 years were identified and shown to grow over time (Berkes et al. 2019); wealth gradients were also evident using an early version of MODEL in Bangladesh (Pisani, Borisova, and Dowd 2015). In Brazil, preprimary-aged children who were from indigenous backgrounds scored significantly lower on MODEL than non-indigenous children; maternal education was also strongly associated with MODEL scores (Ogando et al. 2019). While it is important to note that MODEL items may not have been culturally responsive, the alignment of MODEL items with national standards suggests that MODEL is sensitive to gaps in child development due to inequity.

Overall level of quality in ECE classrooms

MELQO data have provided insight into the status of preprimary classrooms. While the mix of strengths and challenges differs by country, MELE data identified the challenges in ensuring quality ECE. In Ethiopia, teachers relied on a high degree of repetition, with limited materials for some types of play and learning (Rossiter et al. 2018). Baseline [or pre-training] data from Indonesia's pilot teacher-training project showed very low levels of teacher verbal interaction with children, almost no book-reading, and high emphasis on whole-class teaching. Classrooms in Brazil similarly faced challenges in reaching high levels across all indicators of quality, with strengths in engaging children in learning and additional work to institute individualized instruction (Ogando et al. 2019). Colombian classrooms were observed to be strong in structural quality, including health and safety and highly trained teachers, but observations of teachers' instructional practices revealed room for growth in specific learning areas, such as early math and literacy (Maldonado 2019). In Lesotho, results revealed low levels of literacy activities, with storybook readings occurring in less than 20% of the observations, and no or only rote-based literacy instruction occurring in over half the classrooms.

What makes data useful?

MELQO results can be extremely useful in simply describing learning environments, teachers' characteristics and children's development. The use of descriptive data, for example, the percentage of teachers who read a storybook to children during the observation or the percentage of classrooms where children have access to materials, can be a powerful tool to help policymakers better understand the realities of preprimary classrooms. In Indonesia, the World Bank team, in consultation with national and local governments, used MELE data on teachers' characteristics in classrooms to identify areas where more support for teaching quality was needed within the new teacher-training program being piloted (such as more use of open questions and more intentional support for literacy and math development). In Peru, the Ministry was able to provide information on the day-to-day routines and activities that was essential for enhancing professional development. In Tanzania, results have helped identify where professional development may be useful, and MELQO data have also helped inform improvements to the national curriculum. In Lesotho, MELQO data have been used by the Ministry of Education and Training to support collaborative efforts among national, multi-sectoral partners to address identified areas of need, specifically around teacher training, materials, and integration of services such as health/nutrition and child protections into ECE.

Three points emerge. First, it is important to note that the descriptive data are especially useful for policy and programmatic development when closely aligned to national standards and expectations, to ensure that all agree that the constructs measured and reported upon are important. It may be more challenging to effectively use descriptive data when national standards have not yet been developed, or when the standards indicate broad goals with little specificity. Second, a primary question for many countries is, 'Do quality early childhood services lead to better learning in the primary grades?' The multifaceted and complex set of influences on child development means that this question is challenging to answer. Monitoring studies generate associative data that do not allow causal associations to be established – and the inability to draw causal links can be

frustrating if policymakers would like to know about policy impacts, for example, the extent to which a curricular reform led to changes in teacher practices, classroom quality or child learning and development. Representative studies are also constrained by the extent to which sampling data are available; sampling data may not be reliable for key questions (for example, comparisons between children who have attended preprimary and those who have not, as the children who haven't attended are not listed in databases and therefore cannot be sampled). In some MELQO countries, complex statistical models were needed to accurately address policymakers' questions, which produced more reliable results, but which were more difficult for policymakers to understand and utilize. The balance between easily interpretable and applicable results and complex models required to achieve them is challenging to resolve.

Conclusion

The strong demand for measurement signals growing appreciation of, and investment in, ECD. Countries continue to express interest in building data-driven early childhood systems, creating an opportunity to build cutting-edge systems to promote high-quality early childhood settings and child learning. MELQO's experience helps clarify five key lessons in the realities of using data on ECD in low- and middle-income countries.

First, although the science on early childhood is strong and continues to grow, effects described in small-scale studies do not consistently translate across populations (Shonkoff, Radner, and Foote 2017), meaning that more national-level studies on ECE are needed, especially with longitudinal designs outlining the contributions of family backgrounds, child characteristics, health and nutrition, and quality learning environments to child development. While such studies require substantial investments, MELQO experiences also highlight how valuable those data can be – and that each country is unique. The complex associations between child development and quality in ECE generate multi-layered and mixed results, and ample time and resources should be set aside to fully mine the data and identify application to policy, ideally with a consistent group of country stakeholders at the table from start to finish.

Second, ongoing investments in tool development and validation are essential for effective use of global measurement tools. MELQO achieved its goal of creating an open-source set of tools, but the first phases of the project were not sufficient for fully validating the tools. The strong push for creating policy-relevant data may also have created too much pressure to move forward quickly and without adequate room for trial and error. That said, there are people registered to access the MELQO tools in nearly every country of the world, and the MELQO tools seem to have been a useful starting point for researchers, governments and NGOs, who then adapt the tools to meet their needs (for example, see Tenorio 2019). At the same time, we cannot overstate the value of piloting and validation. As the design of the MELQO tools allows for adaptation to country context and our findings have indicated different patterns across countries, adequate piloting and validation efforts are necessary to ensure the tool is accurately measuring the desired constructs and that findings are able to address and inform specific country needs.

Third, for greatest impact, measurement should not be a one-time exercise. Instead, measurement may be most valuable to countries when there are multiple groups involved in evaluating and testing tools over time, in close collaboration with government ministries who can identify research questions with the most policy relevance and help translate research tools into sustainable monitoring systems. Local capacity-building at the country level should be a high priority for future investments in measurement. Such capacity-building investments could include support for local universities to build expertise in early childhood measurement; partnerships between ministries in common regions to promote knowledge-sharing; and efforts to ensure that global tools are adequately adapted and integrated with local tools and priorities. These investments will support the creation of effective monitoring and improvement systems that ministries can use to track trends and impacts of policy or programming initiatives.

Fourth, it is important that any measurement effort is aligned with the policies and political context of the countries where it is implemented. The independent evaluation of MELQO conducted in 2015–2016 found that the MELQO tools were well-received in countries where there were already national policies on ECE and measurement, where there was technical support from multi-lateral agencies, and the content of the tools were aligned with the country's education sector plans (Salzano and Kjaersgaard 2016). However, in the time since the evaluation was conducted, we have found that multi-lateral support, including funding, must be complemented by a strong project lead or group of advocates within the government who champion and sustain the initiative. It is also important to note the risks of global measurement, as global initiatives like MELQO (and other global tools) run the risk of overtaking local tool-development efforts, which are critical for long-term adoption and sustainability of measurement.

As we go forward, questions about how to make ECE data relevant and actionable are paramount – including but not limited to MELQO. Recent work on encouraging use of data in education systems points to the importance of integrating data into education management information systems (EMIS), with a faster turn-around time and easier access to results (Custer et al. 2018).

Data on ECE clearly have an important role in the Sustainable Development agenda. One of MELQO's original purposes was to help inform global monitoring. Because this set of responsibilities has now been taken on by UNICEF through the Multiple Indicator Cluster Survey (MICS) Early Childhood Development Index, MELQO's contribution has shifted from global monitoring to national monitoring, with emphasis on helping countries create tools that will fit their needs. In some countries, the use of the MELQO tools is the first opportunity to systematically examine patterns in child development and learning and in the quality of children's learning environments. In many ways, the work of the MELQO Consortium represented an important point in the early childhood community, by recognizing the value of data for early childhood. MELQO's results and experiences contribute to paving the way for future efforts to promote data-driven early childhood systems, ideally through increased emphasis on capacity-building and investment in the design and testing of local as well as global tools.

Notes

1. <https://gemreportunesco.wordpress.com/2016/10/24/target-4-2-what-is-at-stake-for-monitoring-progress-on-early-childhood-education/>.
2. Full country reports, including descriptions of demographic information and policy context, are not available for all countries. Where available, the citations have been included.

Acknowledgments

The MELQO Initiative has relied upon the efforts of several individuals. A special thanks to Drs. Frances Aboud, Magdalena Janus, Linda Platas, Nirmala Rao, and Hiro Yoshikawa (in alphabetical order) for their contributions as lead developers of the tools. Funding for this project was received from the Children's Investment Fund Foundation, Dubai Cares, Porticus Foundation, World Bank and UNICEF.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the Children's Investment Fund Foundation; Porticus Foundation; Dubai Cares; UNICEF; World Bank Group.

ORCID

Abbie Raikes  <http://orcid.org/0000-0003-2403-6270>

References

- Aboud, F. E. 2008. "A Social-cognitive Developmental Theory of Prejudice." In *Handbook of Race, Racism, and the Developing Child*, edited by S. M. Quintana and C. McKown, 55–71. Hoboken, New Jersey: John Wiley & Sons, .
- Anderson, K., A. Raikes, S. Kosaraj, and A. Solano. 2017. *National Early Care and Education Monitoring Systems*. Washington, DC: Brookings Institution. https://www.brookings.edu/wp-content/uploads/2017/04/global_20170427_early-childhood-care.pdf
- Atchison, B., L. Diffey, A. Rafa, and M. Sarubbi. 2017. "Equity in Education: Key Questions to Consider." http://www.ecs.org/wp-content/uploads/Equity_in_Education_Key_questions_to_consider.pdf
- Berkes, J., A. Raikes, A. Bouguen, and D. Filmer. 2019. "Joint Roles of Parenting and Nutritional Status for Child Development: Evidence from Rural Cambodia." *Developmental Science* e12874. doi:10.1111/desc.12874.
- Black, M. M., S. P. Walker, L. C. Fernald, C. T. Andersen, A. M. DiGirolamo, C. Lu, D. C. McCoy, et al. 2017. "Early Childhood Development Coming of Age: Science through the Life Course." *The Lancet* 389 (10064): 77–90. doi:10.1016/S0140-6736(16)31389-7.
- Brinkman, S. A., S. Silburn, D. Lawrence, S. Goldfeld, M. Sayers, and F. Oberklaid. 2007. "Investigating the Validity of the Australian Early Development Index." *Early Education and Development* 18 (3): 427–451. doi:10.1080/10409280701610812.
- Brinkman, S. A., T. A. Gregory, S. Goldfeld, J. W. Lynch, and M. Hardy. 2014. "Data Resource Profile: The Australian Early Development Index (AEDI)." *International Journal of Epidemiology* 43 (4): 1089–1096. doi:10.1093/ije/dyu085.

- Burchinal, M. 2018. "Measuring Early Care and Education Quality." *Child Development Perspectives* 12 (1): 3–9. doi:10.1111/cdep.2018.12.issue-1.
- Custer, S., E. King, T. Atinc, L. Read, and T. Sethi. 2018. *Toward Data-driven Education Systems: Insights into Using Information to Measure Results and Manage Change*. Washington, DC: Brookings Institution. <https://www.brookings.edu/research/toward-data-driven-education-systems-insights-into-using-information-to-measure-results-and-manage-change/>
- Dahlberg, G., P. Moss, and A. Pence. (2013). *Beyond quality in early childhood education and care: Languages of evaluation*. Routledge
- de la Mothe, E., J. Espey, and G. Schmidt-Traub. 2015. "Measuring Progress on the SDGs: Multi-level Reporting." *Global sustainable development report brief*. New York, NY: United Nations Division for Sustainable Development. <https://sustainabledevelopment.un.org/content/documents/6464102-Measuring%20Progress%20on%20the%20SDGs%20%20%20Multi-level%20Reporting.pdf>
- Early, D. M., J. Sideris, J. Neitzel, D. R. LaForett, and C. G. Nehler. 2018. "Factor Structure and Validity of the Early Childhood Environment Rating Scale—Third Edition (ECERS-3)." *Early Childhood Research Quarterly* 44: 242–256. doi:10.1016/j.ecresq.2018.04.009.
- Fernald, L. C., E. Prado, P. Kariger, and A. Raikes. 2017. *A toolkit for measuring early childhood development in low and middle-income countries*.
- Gordon, R. A., K. Fujimoto, R. Kaestner, S. Korenman, and K. Abner. 2013. "An Assessment of the Validity of the ECERS-R with Implications for Measures of Child Care Quality and Relations to Child Development." *Developmental Psychology* 49 (1): 146. doi:10.1037/a0027899.
- Halpin, P. F., S. Wolf, H. Yoshikawa, N. Rojas, S. Kabay, L. Pisani, and A. J. Dowd. 2019. "Measuring Early Learning and Development across Cultures: Invariance of the IDELA across Five Countries." *Developmental Psychology* 55 (1): 23. doi:10.1037/dev0000626.
- Janus, M., and D. R. Offord. 2007. "Development and Psychometric Properties of the Early Development Instrument (EDI): A Measure of Children's School Readiness." *Canadian Journal of Behavioural Science/Revue Canadienne Des Sciences Du Comportement* 39 (1): 1. doi:10.1037/cjbs2007001.
- Learning Metrics Task Force. 2013. *Toward Universal Learning: Recommendations from the Learning Metrics Task Force*. Montreal and Washington, DC: UNESCO Institute for Statistics and the Brookings Institution.
- Maldonado, C., M. C. Rey, A. M. Nieto, A. Mateus, H. Yoshikawa, L. Ponguta, and S. L. Kagan. 2019. "Quality of Early Childhood Education in Colombia: Results from a National Study." Panel presentation at the annual Comparative International Education Society conference, San Francisco, CA.
- Montie, J. E., Z. Xiang, and L. J. Schweinhart. 2006. "Preschool Experience in 10 Countries: Cognitive and Language Performance at Age 7." *Early Childhood Research Quarterly* 21 (3): 313–331. doi:10.1016/j.ecresq.2006.07.007.
- Ogando, L., B. Abuchaim, D. Santos, and A. Meraz. 2019. "Quality of Early Childhood Education in Brazil: Adaptation of MELQO." Panel presentation at the annual Comparative International Education Society conference, San Francisco, CA.
- Organisation for Economic Co-operation and Development. 2015. *Starting Strong IV: Monitoring Quality in Early Childhood Education and Care*. Paris, France: OECD Publishing.
- Pisani, L., I. Borisova, and A. J. Dowd. 2015. "International Development and Early Learning Assessment Technical Working Paper." <http://repositorio.minedu.gob.pe/bitstream/handle/123456789/4344/International%20Development%20and%20Early%20Learning%20Assessment%20Technical%20Working%20Paper.pdf?sequence=1&isAllowed=y>
- Raikes, A., H. Yoshikawa, P. R. Britto, and I. Iruka. 2017. "Children, Youth and Developmental Science in the 2015–2030 Global Sustainable Development Goals." *Social Policy Report* 30 (3). doi:10.1002/j.2379-3988.2017.tb00088.x.
- Raikes, A., K. Anderson, and T. M. Atinc. 2014. "Global Momentum for Measuring Early Learning Outcomes and Program Quality." <https://www.brookings.edu/blog/education-plus-development/2014/11/17/global-momentum-for-measuring-early-learning-outcomes-and-program-quality/>

- Raikes, A., N. Koziol, M. Janus, L. Platas, T. Weatherholt, A. Smeby, and R. Sayre. 2019. "Examination of School Readiness Constructs in Tanzania: Psychometric Evaluation of the MELQO Scales." *Journal of Applied Developmental Psychology* 62: 122–134. doi:10.1016/j.appdev.2019.02.003.
- Rao, N., J. Sun, B. Richards, A. M. Weber, A. Sincovich, G. L. Darmstadt, and P. Ip. 2019. "Assessing Diversity in Early Childhood Development in the East Asia-Pacific." *Child Indicators Research* 12 (1): 235–254. doi:10.1007/s12187-018-9528-5.
- Rossiter, J., B. Hagos, P. Rose, T. Teferra, and T. Woldehanna. 2018. "Early Learning in Ethiopia: Equitable Access and Learning." *System Diagnostic Report for World Bank Early Learning Program*.
- RTI. 2018. *Measuring Early Learning and Quality Outcomes (Final Report)*. Mainland Tanzania: World Bank.
- Salzano, C., and J. Kjaersgaard. 2016. *Formative Evaluation: The Effectiveness, Efficiency and Sustainability of the Measuring Early Learning and Outcomes (MELQO) Project – Phase 1*. London: Children's Investment Fund Foundation. file:///Users/katie/Desktop/MELQO_Formative_Evaluation_Report.pdf
- Shonkoff, J. P., J. M. Radner, and N. Foote. 2017. "Expanding the Evidence Base to Drive More Productive Early Childhood Investment." *The Lancet* 389 (10064): 14–16. doi:10.1016/S0140-6736(16)31702-0.
- Snelling, M., A. Dawes, L. Biersteker, E. Girdwood, and C. Tredoux. 2019. "The Development of A South African Early Learning Outcomes Measure: A South African Instrument for Measuring Early Learning Program Outcomes." *Child: Care, Health and Development* 45 (2): 257–270. doi:10.1111/cch.12641.
- Sustainable Development Solutions Network (SDSN). 2015. *Indicators and a Monitoring Framework for the Sustainable Development Goals*. New York: SDSN.
- Tenorio, A. 2019. "Aiding the Quest for Quality: Digitizing and Adapting the MELE Tool for Instant Feedback and Continuous Quality Improvement." Panel presentation at the annual Comparative International Education Society conference, San Francisco, CA: UNESCO Institute for Statistics (2016). http://uis.unesco.org/sites/default/files/documents/the-cost-of-not-assessing-learning-outcomes-2016-en_0.pdf
- UNESCO, UNICEF, Brookings Institution, and the World Bank. 2017. *Overview: MELQO: Measuring Early Learning Quality and Outcomes*. Paris: UNESCO Publishing.
- Urban, M. 2018. "The Shape of Things to Come and What to Do about Tom and Mia. Interrogating the OECD's International Early Learning and Child Well-Being Study from an Anti-colonialist Perspective." *Policy Futures in Education* 17 (1): 87–101. doi:10.1177/1478210318819177.
- Verdisco, A., S. Cueto, J. Thompson, and O. Neuschmidt. 2014. "PRIDI: Urgency and Possibility." <https://publications.iadb.org/en/publication/16927/pridi-urgency-and-possibility>
- Weatherholt, T. 2018. "Understanding Preprimary Quality in Tanzania: Data from the MELQO Study." Panel presentation at the annual CIES (Comparative and International Education Society) conference, Mexico City, Mexico, March 25–29.
- World Bank. 2017a. *Results from Four MELQO Pilot Countries: Evidence of Within-country Validity and Cross-country Comparison of Child Development*. Washington, DC: World Bank. http://ecdmeasures.org/wp-content/uploads/2018/04/MELQO-BRIEF-1_MODEL-Results.pdf
- World Bank. 2017b. *Overview: MELQO Measuring Early Learning Quality and Outcomes*. Washington, DC: World Bank Global Education Practice. Retrieved from <http://unesdoc.unesco.org/images/0024/002480/248053e.pdf>
- Yoshikawa, H., A. J. Wuermli, A. Raikes, S. Kim, and S. B. Kabay. 2018. "Toward High-Quality Early Childhood Development Programs and Policies at National Scale: Directions for Research in Global Contexts." *Social Policy Report* 31 (1): 1–36. doi:10.1002/sop2.2018.31.issue-1.
- Yoshikawa, H., D. Leyva, C. E. Snow, E. Treviño, M. Barata, C. Weiland, M. C. Arbour, et al. 2015. "Experimental Impacts of a Teacher Professional Development Program in Chile on Preschool Classroom Quality and Child Outcomes." *Developmental Psychology* 51 (3): 309. doi:10.1037/a0038785.