Journal of Applied Developmental Psychology 62 (2019) 122–134

Contents lists available at ScienceDirect

Journal of Applied Developmental Psychology

journal homepage: www.elsevier.com/locate/jappdp

Examination of school readiness constructs in Tanzania: Psychometric evaluation of the MELQO scales

Abbie Raikes[a,*], Natalie Koziol[b], Magdalena Janus[c], Linda Platas[d], Tara Weatherholt[e], Anna Smeby[f], Rebecca Sayre[g]

[a] University of Nebraska Medical Center, College of Public Health, United States
[b] MAP Academy, University of Nebraska, Lincoln, United States
[c] McMaster University, Canada
[d] San Francisco State University, United States
[e] RTI Inc., United States
[f] UNICEF, Tanzania
[g] World Bank, United States

ARTICLE INFO

Keywords:
Early childhood development and learning
Measurement
Low-income countries
Self-regulation

ABSTRACT

This study examines the development and learning of 684 Tanzanian children starting school, averaging 7 years of age. A primary goal was to evaluate the psychometric properties of a globally-informed measure of school readiness. Using multiple measures including newly-developed direct assessment, and teacher and parent reports of child development, we hypothesized that children's development and learning would demonstrate expected constructs of academic and social/emotional skills and associations with family and child characteristics. Children's direct assessment scores factored into five domains measuring pre-mathematics, pre-literacy, executive functioning, fine motor skills, and socioemotional knowledge. Teachers' reports of children's social/emotional abilities factored into three domains measuring children's social competence, attention/self-regulatory abilities, and problem behaviors. Structural analyses indicated that children's attentional/self-regulatory abilities were associated with their direct assessment scores. Future research should examine these constructs in other countries, with additional methodologies to examine cultural fit and relevance.

Introduction

Early childhood development and learning has a profound impact on learning throughout the school years and is one of the best leading indicators of future academic performance and economic success (Hanushek & Woessmann, 2008, 2012). Many countries are increasing investments in early childhood education to promote equity in learning and ensure that children have the basic skills and competencies necessary for school success. As evidence of this trend, as many as 48% of children now have access to pre-primary education globally, up from 30% in 2000 (World Bank, 2018). Pre-primary education may be especially important considering the notable discrepancies that emerge in cognitive development long before the start of school (Fernald, Weber, Galasso, & Ratsifandrihamanana, 2011). Early childhood development, and especially children's preparation for formal schooling, was also included in the United Nations' Sustainable Development Goals (SDGs). Inclusion of early childhood development in SDGs has sparked increased interest in defining and measuring competencies at the start of school, both as part of tracking progress towards global goals using globally-comparable instruments and to inform national-level policy-making (Raikes, Yoshikawa, Britto, & Iruka, 2017; UNESCO, 2017).

In response to the need for national-level, representative data on early childhood development before the start of formal schooling, the Measuring Early Learning Quality & Outcomes (MELQO) project was founded by three United Nations organizations and one non-governmental organization, UNESCO, UNICEF, the World Bank, and the Brookings Institution. The MELQO initiative supports low- and middle-income countries in generating feasible, actionable measurement of early childhood development and quality of pre-primary settings, to inform both global monitoring and provide nationally relevant data (Raikes, 2017). It resulted in a common item set, drawn from existing

* Corresponding author at: University of Nebraska Medical Center College of Public Health, 986075 Nebraska Medical Center, Omaha, NE 68198, United States.
E-mail addresses: abbie.raikes@unmc.edu (A. Raikes), nkoziol@unl.edu (N. Koziol), janusm@mcmaster.ca (M. Janus), tweatherholt@rti.org (T. Weatherholt), asmeby@unicef.org (A. Smeby), rsayre@worldbank.org (R. Sayre).

https://doi.org/10.1016/j.appdev.2019.02.003
Received 3 June 2018; Received in revised form 10 January 2019; Accepted 6 February 2019

measures used in low- and middle-income countries, that represented constructs deemed by a large consortium to be potentially relevant across countries. MELQO also developed procedures for adaptation used by low- and middle-income countries to adapt and measure child development (see UNESCO, 2017, for a complete description of MELQO).

Many existing measures of early childhood development and learning contain similar items and constructs (Raikes, 2017), lending weight to the possibility of globally-relevant measurement. Yet at present, we have little information on normative learning and development at the start of the school years in many countries, or how local patterns of children's learning and development fit with global expectations. Both universalist and socio-cultural theories have been applied to children's learning and development in the preschool years in sub-Saharan Africa (e.g., Kabay, Wolf, & Yoshikawa, 2017; Wolf & McCoy, 2017), and these theories predict that some elements of learning and development may be consistent across contexts while others, such as social/emotional development, may show greater sensitivity to culture and context.

To examine the applicability of global constructs across diverse contexts, the purposes of this study were two-fold: First, to psychometrically evaluate the properties of a measure of child development and learning drawn from existing global instruments in Tanzania, a low-income country in sub-Saharan Africa; and second, to describe associations child and family characteristics and child development and learning in Tanzania. These purposes are important for at least two reasons. First, data on child development and learning can help promote effective policies and practices (Raikes et al., 2017), but only if appropriately aligned with national priorities and cultural values (Serpell et al., 2017). Second, documenting associations between child and family characteristics and child development and learning can increase understanding of child development in diverse contexts, which in turn can help inform developmental science and help promote effective interventions to support young children's development.

Conceptual and empirical evidence of "school readiness" constructs in low- and middle-income countries

Broadly speaking, school readiness refers to the state of children's skills and competencies at the start of school that facilitate success later in schooling (Snow, 2006). Existing research on school readiness emphasizes the multidimensional nature of children's development, for example, the joint contributions of social-emotional and pre-academic skills (Blair, 2002; Duncan et al., 2007) and the vital contribution of children's social behavior to a successful transition to primary school (Raver, 2002). Most measures of child development and learning at the start of school have not been fully validated in many countries, especially low- and middle-income countries (Fernald, Prado, Kariger, & Raikes, 2017), which poses a challenge in identifying psychometrically-supported tools in the context of the strong political push for data on child development. Building psychometric evidence for a measure or set of measures ideally involves analyses of construct validity, or the extent to which scores represent the intended construct; and reliability, including the internal consistency of scores. Construct validity is a broad concept that requires consideration of many aspects, including internal (e.g., factor structure) and external (convergent/divergent) evidence (Messick, 1995).

Existing studies of tools purportedly measuring child development have established validity evidence primarily through two mechanisms: first, by documenting the factor structures of tools within countries (e.g., Wolf et al., 2017); and second, by demonstrating associations between children's scores and expected predictors of child development, including family background and early experiences such as attendance at preschool (e.g., Rao et al., 2013). Convergent evidence can also be established by demonstrating associations across measures of child development; for example, between direct assessments of children's development and learning, and parent or teacher reports. Evaluation of psychometric evidence using representative samples is especially critical, as small and/or selected populations can produce skewed distributions that make it difficult to assess whether the tool adequately measures the constructs of interest for most children. Within the context of the SDGs, a final standard for establishing validity evidence is policy relevance, or the extent to which measures produce data that are useful and applicable to policy decisions related to early childhood development (Raikes et al., 2017). Policy relevance can be evaluated by examining the process for adapting assessments to align with national standards, and by identifying how the resulting information can pragmatically be used in creating or modifying existing policies.

In relation to the first purpose of this study, psychometric evaluation the properties of MELQO tools for national measurement in Tanzania, some research reports on validity evidence of tools used for program evaluations and research studies in other African countries, but not at a national level and not within Tanzania. For example, using a locally-developed measure, McCoy, Zuilkowski, and Fink (2015) documented associations between Zambian children's in- and out-of-home learning experiences and learning domains, including executive function, fine motor, and receptive vocabulary. Using the International Development and Early Learning Assessment (IDELA), a global early childhood development measure designed for program evaluations, Wolf et al. (2017) found psychometric support for factors of school readiness including early literacy, mathematics, motor development and social/emotional development in Ethiopia and demonstrated invariance of the factor structure across different groups of children, such as those who attended preschool and those who did not.

In sum, emerging research suggests the feasibility and reliability of measuring children's development and learning in the preschool years in sub-Saharan Africa using tools that are both global and locally-developed. However, the evidence base is still relatively limited. Wolf et al. (2017) reported that few studies document factor structures of assessments used in low- and middle-income countries and concluded that the field needs more research to compare how measures function across countries, especially considering growing policy emphasis on measurement and the need to ensure adequate alignment to national policy goals. Few studies to date have examined the role of social and emotional development in school achievement in sub-Saharan Africa or documented the psychometric properties of measures used to index social and emotional development. For example, although Wolf and McCoy (2017) reported overall validity evidence for the factor structure of the IDELA in Ghana in a study examining the effects of parent socioeconomic status on school readiness, scores from the social/emotional subscale were not significantly associated with many predictors of pre-academic skills, such as caregiver stimulation. Research on social/emotional development in the United States, for example, has identified theoretical models to serve as the basis for measurement that include multiple aspects such as self-regulation, problem behaviors, and social competence (Darling-Churchill & Lippman, 2016). These factors emerge within samples from the United States (e.g., Rispoli, McGoey, Koziol, & Schreiber, 2013), but models of social and emotional development including multiple dimensions of early social/emotional development have not frequently been tested in sub-Saharan Africa. Finally, few studies in sub-Saharan Africa have examined constructs of child development and learning across national populations, which may show different patterns than samples of children participating in research studies or program evaluations.

Measurement of school readiness in Tanzania

Tanzania, a low-income country in sub-Saharan Africa, represents an important opportunity to examine school readiness because children are not meeting basic academic milestones in primary school, and thus may have distinct learning profiles at the start of school. In 2016,

although the gross primary school enrollment rate is 85% of eligible children (UNESCO Institute for Statistics[1]), only 12% of Tanzanian second graders could read with comprehension and 8% met the benchmark for addition and subtraction (RTI International, 2016). To promote children's school success, Tanzania recently launched a National Education and Training Policy, which added one year of pre-primary as part of free and compulsory basic education. In 2015, 32% of all eligible children were enrolled in pre-primary education (World Bank, 2016), with plans to expand access dramatically in upcoming years. To enhance children's learning in the early school years, new curricula have recently been recently developed, outlining expectations for pre-primary and early primary education. The current study was initiated to generate empirical evidence on the characteristics of children, families and children's learning and development at the start of primary schooling, to serve as a baseline for national efforts to improve pre-primary and primary education, and to inform planning and financing of implementation of new curricula. Documenting social and emotional development is an important element of defining effective interventions to promote children's learning in Tanzania: with such low levels of learning, social/emotional development has been highlighted by policymakers as a potential window into why learning levels are so low.

In the present study, we examine school readiness within a nationally representative sample of Tanzanian children enrolled in the first year of primary school, as part of the Measuring Early Learning Quality & Outcomes initiative (MELQO). We posed two questions: First, drawing on global conceptualizations of school readiness, what are the psychometric properties of a "global" tool used in the Tanzanian context? Second, how do children's characteristics – specifically their social and emotional development and family backgrounds –.

contribute to learning and development at the start of school? Based on previous work and reflecting constructs of school readiness hypothesized to be universal in nature, we hypothesized that children's direct assessment and teacher-rated measures of child development and learning, including social/emotional development, would demonstrate factor structures similar to global conceptualizations of learning and development at the start of school, as well as associations between children's social/emotional development and their learning, and associations with family background. We also hypothesized that items from assessments used in other countries would show adequate alignment with Tanzania's national standards, reflecting consistency in the definition of school readiness across contexts.

Method

MELQO Tanzania (see UNESCO et al., 2017 pp. 19–26, for descriptions of the effort and its piloting) was initiated by the Tanzanian Ministry of Education, Science, and Technology (MOEST) with support from UNICEF Tanzania, the World Bank, and the Global MELQO team. Tanzanian officials were interested in data to 1) create a baseline of pre-primary quality and learning outcomes for the new pre-primary curriculum and financing of pre-primary being launched the same year; and 2) inform planning and prioritization of future investments in pre-primary education. Multiple meetings, workshops and field tests were conducted to review and adapt items drawn from the MELQO Measurement of Development Early Learning (MODEL) global item set (outlined in UNESCO, et al., 2017), and ensure alignment with national curricula, standards, culture and context, as guided by Ministry and curriculum officials and local experts.

A one-day workshop with representatives of MOEST, the Tanzania Institute of Education (TIE, responsible for curriculum and in-service teacher training), and local experts, was held to review alignment of the items with national curricula, resulting in a confirmation that the MELQO items adequately covered most domains of the new curriculum. The national curricula document outlines expectations for children's learning in the pre-primary years by areas of learning; for example, for language, literacy and communication, the document states that children should gain skills in speaking and listening; awareness about print, books, letters, sounds and words.

In addition, participants resolved that new items should be developed to cover the two domains (Caring for Health, Caring for the Environment) of the curriculum not covered by the MELQO items, resulting in the decision to develop items on nutrition, hygiene and safety (these items were not included in analyses because they were not hypothesized to be part of a global set of items on school readiness). Finally, participants agreed to include MELQO items indexing executive functioning, although not part of the curriculum, because it serves as a cognitive foundation for the acquisition of competencies across the curriculum. These new items were developed through consultation between the research team, the participants and a local child development expert, contracted to field test the new items.

Participants

Participants included children, their teachers and a family member from a nationally-representative sample of Tanzanian children entering grade 1 of primary school. Because policymakers were primarily interested in understanding the overall readiness of children for primary education and because a greater proportion of children are enrolled in primary than pre-primary education, children were sampled from primary school classrooms in the early months of the school year. Participants were identified by first randomly selecting twelve regions of the country, and then drawing the names of seventy schools from a national registry of schools (including both public and private) within each region, from which one entry-level primary school classroom and ten children within that classroom were randomly drawn for study participation. In partnership with the Tanzanian MOEST, a sampling frame was created that included all private and public schools serving children at the start of Standard 1, or formal education, that also provided pre-primary education (99% of primary schools offer pre-primary education). Probability proportional to size (PPS) sampling was used by selecting schools based on their total pre-primary pupil enrolment. By using PPS sampling of schools and selecting the same number of children from each school, children had equal probabilities of selection, which includes the benefits of low bias, improved precision, and a sample whose demographics are approximately the same proportions as the population of interest.

Study participants included 684 children (51.17% female) selected from 69 schools, and a family member (n = 568) who reported on family environments and teacher (n = 671) who reported on child development and learning. There were three reports of children's age: from parents; from teachers; and from children themselves. Because not all parents responded to surveys and because children's birthdates are not considered noteworthy dates in many Tanzanian families, teacher reports of children's ages were used (578 teacher-reported ages versus 365 parent-reported ages); children were not considered as reliable reporters of age as teachers given that some were very young. According to teachers, children ranged in age from 4 to 16 years, with an average age of 6.66 years (SD = 1.29). However, there were discrepancies between teacher reports of children's ages and reports by parents and children themselves, which could indicate low accuracy of age reporting. Approximately 70% of birth dates given by parent and teachers aligned within three months; and 40% aligned exactly. Overaged children were omitted from analyses involving age, as detailed in the data analysis plan section.

Materials and procedure

Procedures involved 1) adapting the MODEL measures, including a

[1] http://uis.unesco.org/country/TZ

direct assessment of child learning and a teacher report of child development, and surveys of family engagement, assets and demographics, to align with local context and policy questions, as outlined above; 2) piloting MODEL measures to ensure measure functioning and to inform additional modifications; and 3) conducting full data collection with a nationally representative sample. The adaptation of each measure began with a global set of items generated through the MELQO process (full process and justification for selection of items is outlined in UNESCO, 2017). All of the items used in both the direct assessment and the teacher report had been used in other measures previously used in low- and middle-income countries; the direct assessment consisted of items from commonly-used intelligence and school-readiness instruments, the IDELA (Wolf et al., 2017) as well as other measures such as the Head Shoulders Knees Toes task (McClelland et al., 2014), and the teacher report drew most frequently from the Canadian National Longitudinal Study of Children and Youth and the EDI (Janus & Offord, 2007).

Measures were adapted to ensure local relevance by convening stakeholders in early childhood education; reviewing all tools and ensuring alignment with national standards; piloting measures; and adjusting the tools again after obtaining preliminary results.

For piloting and full data collection, data collection was led by an international non-profit research firm, which trained a group of Tanzanian data collectors on the instruments in partnership with a Tanzanian data collection firm. All enumerators were trained to conduct parent and teacher interviews, and then were divided into two groups, one focused on the child direct assessment, and one group on the classroom observation (data not reported here). Enumerators were evaluated several times with accuracy checks and were judged to reach reliability standards based on participation in in-person training and two checks on reliability conducted at the end of the training period.

There were two phases of MELQO data collection. First, pilot data (N = 400) from direct assessments of children, teacher interviews, and parent interviews were collected in the fall of 2015 to analyze items for ill fit and redundancy, and a second pilot was conducted in 2016 to further inform final changes to the instruments. Results from the pilot are not reported in this study but were used to modify the instruments and identify ill-fitting items before using the tools again with a larger sample. Modifications to the tools included elimination from the direct assessment of an item that was too easy for all children; and adjustment of the teacher survey to emphasize social/emotional development rather than children's academic knowledge because the direct assessment would be used to capture this information. A second phase, with a representative sample and modified measures, was then initiated in 2017, and results are reported here.

Data on children's development and learning and family environments were collected through three methods: children were assessed using a direct assessment administered by a trained enumerator at the primary school; teachers were interviewed about each child who was selected to participate in the direct assessment; and enumerators interviewed parents about children's home environments. Schools were visited by teams of enumerators who assessed ten randomly-selected from each first-year primary classroom, Standard 1, to participate in the child assessment, and teachers of each randomly-selected child were asked to respond to a survey on that child's development. Parents were contacted and interviewed after the school-based data collection was complete. All respondents and enumerators were blind to responses from other respondents; for example, teachers and parents were not informed of children's results on the direct assessment. Details on administration of each measure are provided below.

Direct assessment of children

For the direct assessment (DA) of children, children were asked to respond to a series of questions and tasks measuring early literacy and mathematics knowledge and reasoning, executive function, fine motor skills, and knowledge of social environments and health information.

Table 1
Direct assessment items, abbreviations, and internal consistency evidence.

Abr	Item	Cronbach's alpha
SV	Spatial vocabulary	0.73
SET	Producing a set	0.71
NC	Number comparison	0.68
MA	Mental addition	0.66
MT	Mental transformation	0.32
NID	Number identification	0.91
EV	Expressive vocabulary	0.62
ISI	Initial sound identification	0.83
LNK	Letter naming	0.94
LC	Listening comprehension	0.67
HSKT	Head shoulders knees toes	0.94
FDS	Forward digit span	0.56
BDS	Backward digit span	0.76
CPY	Copy shapes	0.53
NW	Name writing	N/A (1 item)
UF	Understands feelings	0.71
PTC	Perspective-taking/Empathy	0.66

See Table 1 for a complete list of the DA tasks and abbreviations used in subsequent sections, as well as internal consistency evidence of the individual tasks based on Cronbach's alpha. Tanzanian officials and the research team created three new subtasks (11 items total) for health knowledge that were not included in the global core and thus not included in the present analyses. In all, 105 items across 22 subtasks were included in the direct assessment (mean percent correct score = 45.46% [SD = 18.89%]), with an average assessment time of 30 min (ranging from 13 min to over 2 h).

Teacher interviews

Teachers were asked to report on 21 items (see Table 2 for complete list of items and abbreviations) assessing children's social/emotional and self-regulatory skills using a 3-point scale of 0 = never, 1 = sometimes and 2 = often/always. The mean response across items (after reverse-scoring negatively-worded items) was 1.38 (SD = 0.31). Cronbach's alpha was 0.82. For thirteen children who participated in

Table 2
Teacher-reported child social/emotional and self-regulatory skills items.

Abr	Item
sr1	Is (*name*) often easily distracted (i.e., how often does his/her concentration wander)?
sr2	When asked to do several things, how often does (*name*) remember all the instructions?
sr3	How often does (*name*) plan ahead?
sr4	How often does (*name*) stop an activity when told to do so?
sr5	How often does (*name*) rudely intrude on others?
sr6	How often does (*name*) keep working at something until s/he is finished?
sr7	How often does (*name*) have difficulties doing things that s/he does not like?
sr8	How often does (*name*) explore the function of new objects?
sr9	How often does (*name*) accept responsibility for his/her actions?
sr10	How often does (*name*) show consideration of other people's feelings?
sr11	Does (*name*) often get along with other children s/he plays with?
sr12	How often does (*name*) offer to help someone who seems to need help?
sr13	Does (*name*) often have difficulty taking turns when playing together with others?
sr14	How often does (*name*) share with his/her peers?
sr15	Does (*name*) often adjust easily to transitions? (for example a new teacher or classroom)
sr16	How often does (*name*) settle down after periods of exciting activity?
sr17	When interacting with others, for example, sharing toys, does (*name*) show self-control?
sr18	Would you say (*name*) kicks, bites, or hits other children or adults?
sr19	How often is (*name*) upset when left by parents/guardians?
sr20	Would you say that (*name*) is often sad or unhappy?
sr21	How often does (*name*) describe his or her feelings? For example, “I'm happy…” or “I'm sad…”

Table 3
Parent-reported family engagement in activities with child items.

Abr	Item
fe1	Read books to or looked at picture books with (*name*)?
fe2	Told stories to (*name*)?
fe3	Sang songs to (*name*) or with (*name*), including lullabies?
fe4	Took (*name*) outside the home, compound, yard or enclosure?
fe5	Played with (*name*)?
fe6	Named, counted, or drew things to or with (*name*)?

the direct assessment, teacher reports were not completed, resulting in a response rate of 98%.

Parent interviews

Because parents' literacy level affects their ability to appropriately respond to a questionnaire, enumerators interviewed all parents in children's home environments rather than asking parents to fill out a survey. To obtain an indicator of socioeconomic status, parents were asked to report on the presence or absence of 16 assets in the family home (M = 4.30 assets, SD = 2.58). These assets included electricity, radios, television, non-mobile phones, refrigerators, watches, mobile phones, heating/cooling, running water, gas stoves, bicycles, motorcycles or scooters, three-wheelers, vehicles, animal-drawn carts, and boats, which was deemed by Tanzanian stakeholders to be a more appropriate and useful indicator of socio-economic status than household income (see Filmer & Pritchett, 1998). Parents were also asked to report on whether the child attended pre-primary in the previous year before attending primary school (93.84%), and on the types of activities that the child had experienced at home in the past three days (see Table 3), including reading books, singing and other activities (M = 3.81 activities [SD = 1.87] out of a possible 6 activities, which were considered to represent the measures of family engagement). Cronbach's alpha was 0.73 for the six items. Enumerators conducted a total of 568 parent interviews, resulting in a response rate of 83%.

Data analysis plan

Quantitative methods were used to evaluate validity evidence relating to the three research questions. Analyses proceeded in two stages. First, measurement models were estimated to evaluate the factor structure and reliability evidence of (a) the direct assessment, (b) teachers' report of children's social/emotional and self-regulatory skills, and (c) parents' report of their engagement with the child. Whereas the focus of this study was on child development and learning constructs, psychometric analyses involving the family engagement items were needed to support the use of this variable in the structural analyses. Second, bivariate and multivariate analyses relating child and family characteristics to child development and learning outcomes were performed to evaluate generalizability and convergent evidence of the scores.

M*plus* Version 7.11 was used to analyze the data (Muthén & Muthén, 1998-2015). To account for nesting of children within schools, cluster-robust standard errors and test statistics were computed. For all tests, the statistical significance level was set at $\alpha = 0.05$. Due to item- and person-level missing data, the analytic sample size varied across analyses, as described below.

Measurement analyses

Categorical confirmatory factor analysis (CCFA) was performed to evaluate the hypothesized factor structures. Items (for the direct assessment, see Table 1 for item codes and description) were analyzed in their raw form with the following exceptions: 1) Verbal Counting (vc) was recoded from a 0–30 range to 0 = 0, 1 = 1–9, 2 = 10–19, 3 = 20–29, 4 = 30; 2) Number Identification items (ni6-ni10) were summed to create a single polytomous item due to extreme collinearity among the individual items; and 3) Expressive Vocabulary items (ev1-ev2) were recoded from a 0–10 range to 0 = 0, 1 = 1–2, 2 = 3–4, 3 = 5–6, 4 = 7–8, 5 = 9–10. Two items were excluded from analyses. Head, Toes, Knees, Shoulder item htks2 was removed due to computational problems and instability of parameter estimates upon its inclusion, and Backward Digit Span item bds7 was removed because no child correctly answered it.

Weighted least squares mean- and variance-adjusted (WLSMV) estimation was used in conjunction with a probit link function and M*plus*'s Theta parameterization. In the presence of item-level missingness, WLSMV implements a pairwise present analysis such that cases were dropped only if they had missing data on all items. The resulting analytic sample sizes were $N = 684$ (DA), $N = 669$ (teacher-report), and $N = 561$ (parent-report).

Global model fit was assessed via the chi-square test of exact fit, root mean square error of approximation (RMSEA) and corresponding 90% confidence interval, and comparative fit index (CFI). Non-significant chi-square tests, RMSEA values close to 0 (e.g., < 0.05 or < 0.08 [Browne & Cudeck, 1993]) and CFI values close to 1 (e.g., > 0.90 [Bentler, 1990] or > 0.95 [Hu & Bentler, 1999]) provide tentative support for the model, though strict adherence to specific cutoffs is not recommended (Kline, 2016). Local model fit was assessed via univariate Lagrange multiplier tests and by examining the parameter estimates and general pattern of results. Chi-square difference tests (via M*plus*'s DIFFTEST option) were performed to compare the fit of nested models. Due to the ample theoretical and past empirical support for the hypothesized factor structures, a cross-validation approach—using part of the sample for model building purposes and the remainder for validation purposes—was not applied. Using the full sample throughout was necessary to maintain as large as possible sample size to number of estimated parameters ratio (a particular concern for the DA analyses due to the large number of items and complex relationships among items). A limitation of this strategy, however, is that data-driven modifications to the hypothesized structures may not generalize beyond this sample. To help mitigate this limitation, modifications were made only when they could be supported by theory or otherwise justified (e.g., a residual correlation due to common item wording).

Total information functions (TIFs) were estimated to evaluate the precision of scores. TIFs provide an indication of how well the set of items discriminates among respondents at each point along the latent continuum. Information can be translated to a reliability metric through the formula: $\rho = I(\theta)/[I(\theta) + Var(\theta)]$.

Structural analyses

In the structural analyses, *maximum* a posteriori (MAP) factor scores, based on the best-fitting measurement models, were used as estimates of child development and learning, teacher-reported child social/emotional and self-regulatory skills, and parent-reported family engagement. Although theoretically possible, simultaneous estimation of the measurement and structural parameters would have been ill-advised when considering the number of parameters to be estimated in relation to the sample size.

Pearson product-moment correlations and mean differences, depending on the distribution of the covariate, were estimated to evaluate the bivariate zero-order relationships between child/family characteristics (including age, gender, preprimary attendance, social/emotional and self-regulatory skills, family assets, and family engagement) and child development and learning outcomes. A path analysis was subsequently performed to evaluate the unique associations between child/family characteristics and child outcomes. The child outcomes, assumed to be normally distributed, were considered simultaneously, and residual variances were allowed to correlate.

Full information maximum likelihood (FIML) estimation was used for the structural analyses such that cases were retained if they had at least partial data. Predictors with missing data (all but gender) were treated as endogenous variables in order to avoid listwise deletion that

would otherwise occur for cases with missing data on the predictors. Overaged children (children older than 8 years, as reported by teachers, who were enrolling in primary school for the first time; $n = 37$) were omitted from all analyses involving age to prevent distorted estimates of age effects. For this same reason, children with missing age information ($n = 93$) were omitted from these analyses.[2] These criteria resulted in analytic sample sizes of $N = 554$ (bivariate associations with age), $N = 684$ (all other bivariate associations), and $N = 554$ (path analysis).

Results

Measurement analyses

Results are organized by research question. See supplementary material for the full set of parameter estimates from the final models, appearing in Tables 1s - 8s.

Question 1: What do results of psychometric analyses reveal on the validity of the tools for use in Tanzania?

Child development and learning (DA)

Procedures for factor analyzing the DA generally followed those of Wolf et al. (2017). First, separate analyses were performed to evaluate the factor structure of the five hypothesized domains. Because items were nested within subtasks, a bifactor parameterization, in which all items load on the primary domain factor in addition to a subtask "method" factor (where all method factors are assumed to be uncorrelated with each other and with the domain factor), was expected to provide better fit to the data than a single domain-level factor. Second, the best-fitting domain-specific models were combined into a single analysis and hypotheses about the relationship among domains were tested.

Domain-specific models. Evaluations of local fit of the hypothesized bifactor models resulted in only a few minor modifications. For Pre-Mathematics, a residual correlation was allowed between Spatial Vocabulary items sv3 and sv4, and Producing a Set item pas1 was removed from its subtask factor. For Executive Functioning, a residual correlation was allowed between Forward Digit Span items fds1 and fds2. For Fine Motor Skills, the Copying subtask factor was deemed unnecessary. Lastly, for Socioemotional Knowledge, Perspective-Taking/Empathy item pte1 was removed from its subtask factor. Table 4 provides global fit information for the final domain-specific models. Although the chi-square test of exact fit was significant for the Pre-Mathematics, Pre-Literacy, and Executive Functioning models, RMSEA and CFI suggested adequate close fit and incremental fit (relative to a null model), respectively. The pattern of factor loadings also supported the bifactor parameterization, with items generally loading significantly on both the primary and subtask factors.

Multi-domain models. The best-fitting domain-specific models were combined into a single unconstrained model in which the five hypothesized domains were allowed to freely correlate (DL_M1). A residual correlation (hypothesized a priori) was allowed between the Socioemotional Knowledge domain and Expressive Vocabulary subtask to account for common method variance (i.e., the necessity of expressive vocabulary in completing the Perspective-Taking/Empathy and Understanding Feelings subtasks). Local fit statistics also led to inclusion of a residual correlation between the Number Identification and Letter Name Knowledge subtasks.

Table 4 provides global fit information for DL_M1, in addition to four other plausible, and more parsimonious, factor structures: (a) a single-factor model in which all items loaded on a single domain (DL_M2), (b) a three-factor model in which the Pre-Mathematics, Pre-Literacy, and Executive Functioning domains were collapsed into a single domain (DL_M3), (c) a four-factor model in which the Pre-Mathematics and Pre-Literacy domains were collapsed into a single domain (DL_M4), and (d) a hierarchical factor model in which the relationships among domains was accounted for through the inclusion of a single hierarchical factor (DL_M5). All five models demonstrated adequate close and incremental fit to the data. However, the unconstrained model fit significantly better than the more parsimonious models, despite large correlations among the domains (see Table 5).

Figs. 1-5 provide the standardized path coefficients for the final multi-domain model (DL_M1). All items loaded significantly on their respective primary domains with the exception of Letter Name Knowledge item lnk8 (the letter "V"), Forward Digit Span item fds1 (repeat the numbers 1 and 6), and Name Writing with respect to the Fine Motor domain. Fig. 1s in the Supplement illustrates the total information functions. The functions peaked at −0.35 SD (maximum reliability = 0.96) units away from the mean for Pre-Mathematics, −0.38 SD (maximum reliability = 0.98) for Pre-Literacy, 0.93 SD (maximum reliability = 0.97) for Executive Functioning, −0.54 SD (maximum reliability = 0.62) for Fine Motor Skills, and − 0.08 SD (maximum reliability = 0.90) for Socioemotional Knowledge. This indicates that the items do best at discriminating among children with average to slightly below average Pre-Mathematics, Pre-Literacy, Fine Motor Skills, and Socioemotional Knowledge, and above average Executive Functioning. Reliability was ≥0.70 for children between −2.96 SD and 1.37 SD units from the mean for Pre-Mathematics, −1.73 SD and 3.00 SD for Pre-Literacy, the entire continuum except between −2.83 SD and − 2.45 SD for Executive Functioning, and between −0.89 SD and 1.34 SD for Socioemotional Cognition. Overall, with the exception of the four-item Fine Motor Skills domain, results suggest that the domain scores can be estimated with adequate precision.

Child social/emotional and self-regulatory skills (teacher-report)

Table 4 provides the global fit statistics and indices for each model. Results provided moderate support for the hypothesized 3-factor solution (SR_M1) representing Attention/Self-Regulation (sr1-sr4, sr6-sr8, sr15-sr16), Problem Behaviors (sr5, sr18-sr20), and Social Competence (sr9-sr14, sr17, sr21). However, inspection of the factor loadings and modification indices suggested the presence of local misfit. Taking into account both empirical and theoretical considerations, the model was modified (SR_M2) as follows: a residual correlation was allowed between items related to the child intruding on others (sr5) and kicking/biting/hitting others (sr18); items related to the child having difficulties doing things s/he does not like (sr7) and taking turns when playing with others (sr13) were specified to load on the Problem Behaviors factor instead of the Attention/Self-Regulation and Social Competence factors, respectively; and an item related to the child's distractibility (sr1) was allowed to cross-load on the Problem Behaviors factor in addition to loading on the Attention/Self-Regulation factor. SR_M2 fit the data well as evidenced by strong global and local fit. Although the correlation between the Social Competence and Attention/Self-Regulation factors was large ($r = 0.86$), collapsing the factors (SR_M3) significantly worsened model fit.

Reliability and fit indices ultimately favored SR_M2. Fig. 6 provides the standardized path coefficients, where all loadings were significantly different from zero. See Fig. 2s in the Supplement for the corresponding total information functions. The functions peaked at 0.65 SD (maximum reliability = 0.68), 0.11 SD (maximum reliability = 0.85), and − 0.22 SD (maximum reliability = 0.87) units away from the mean for the Problem Behaviors, Attention/Self-Regulation, and Social Competence factors, respectively. This indicates that the items do best at

[2] Sensitivity analyses were performed to determine how including children with missing age information impacted the results. Differences in inferential conclusions are noted in the Results section. Complete results for the sensitivity analyses are available upon request.

Table 4
Global fit statistics and indices for categorical confirmatory factor analysis models.

Model	χ^2 (df)	*p*	RMSEA (90% CI)	CFI	$\Delta\chi^2$ (Δdf)	*p*
Final domain-specific child development and learning models						
Pre-Numeracy	290.71 (210)	< 0.001	0.024 (0.017, 0.030)	0.989	–	–
Pre-Literacy	697.71 (464)	< 0.001	0.027 (0.023, 0.031)	0.991	–	–
Executive Functioning	510.41 (228)	< 0.001	0.043 (0.038, 0.047)	0.987	–	–
Fine Motor Skills	5.45 (2)	0.066	0.050 (0.000, 0.103)	0.984	–	–
Socioemotional Cognition	5.91 (3)	0.116	0.038 (0.000, 0.083)	0.998	–	–
Multi-domain child development and learning models						
DL_M1. Unconstrained correlated five-factor[a]	4353.64 (3650)	< 0.001	0.017 (0.015, 0.019)	0.978	–	–
DL_M2. Single domain factor	4460.17 (3662)	< 0.001	0.018 (0.016, 0.020)	0.975	107.20 (12)	< .001[b]
DL_M3. Correlated three-factor (pre-numeracy, pre-literacy, and executive functioning collapsed)	4401.59 (3657)	< 0.001	0.017 (0.015, 0.019)	0.976	45.90 (7)	< .001[b]
DL_M4. Correlated four-factor (pre-numeracy and pre-literacy collapsed)	4371.58 (3654)	< 0.001	0.017 (0.015, 0.019)	0.977	19.39 (4)	.001[b]
DL_M5. Hierarchical factor	4363.37 (3655)	< 0.001	0.017 (0.015, 0.019)	0.977	15.44 (5)	.009[b]
Teacher-reported child self-regulation models						
SR_M1. Correlated three-factor	389.55 (186)	< 0.001	0.040 (0.035, 0.046)	0.924	–	–
SR_M2. Empirically-adjusted correlated three-factor[a]	294.78 (184)	< 0.001	0.030 (0.023, 0.036)	0.958	–	–
SR_M3. Correlated two-factor (social competence and self-regulation collapsed)	322.95 (186)	< 0.001	0.033 (0.027, 0.039)	0.949	27.81 (2)	< .001[c]
Family engagement models						
FE_M1. One-factor	28.95 (9)	< 0.001	0.063 (0.038, 0.089)	0.977	–	–
FE_M2. One-factor w/residual correlation (fe2 w/ fe3)[a]	18.66 (8)	0.017	0.049 (0.020, 0.078)	0.988	12.43 (1)	< .001[d]

[a] Final model.
[b] Comparison model was DL_M1.
[c] Comparison model was SR_M2.
[d] Comparison model was FE_M1.

Table 5
Correlations among child development and learning domains based on unconstrained five-factor model (DL_M1).

	Pre-mathematics	Pre-literacy	Executive functioning	Fine motor skills	Socioemotional cognition
Pre-mathematics	1.00				
Pre-literacy	0.90	1.00			
Executive functioning	0.88	0.84	1.00		
Fine motor skills	0.69	0.77	0.58	1.00	
Socioemotional Cognition	0.68	0.75	0.74	0.46	1.00

Note. $N = 684$. All correlations significantly different from 0 ($p < .001$).

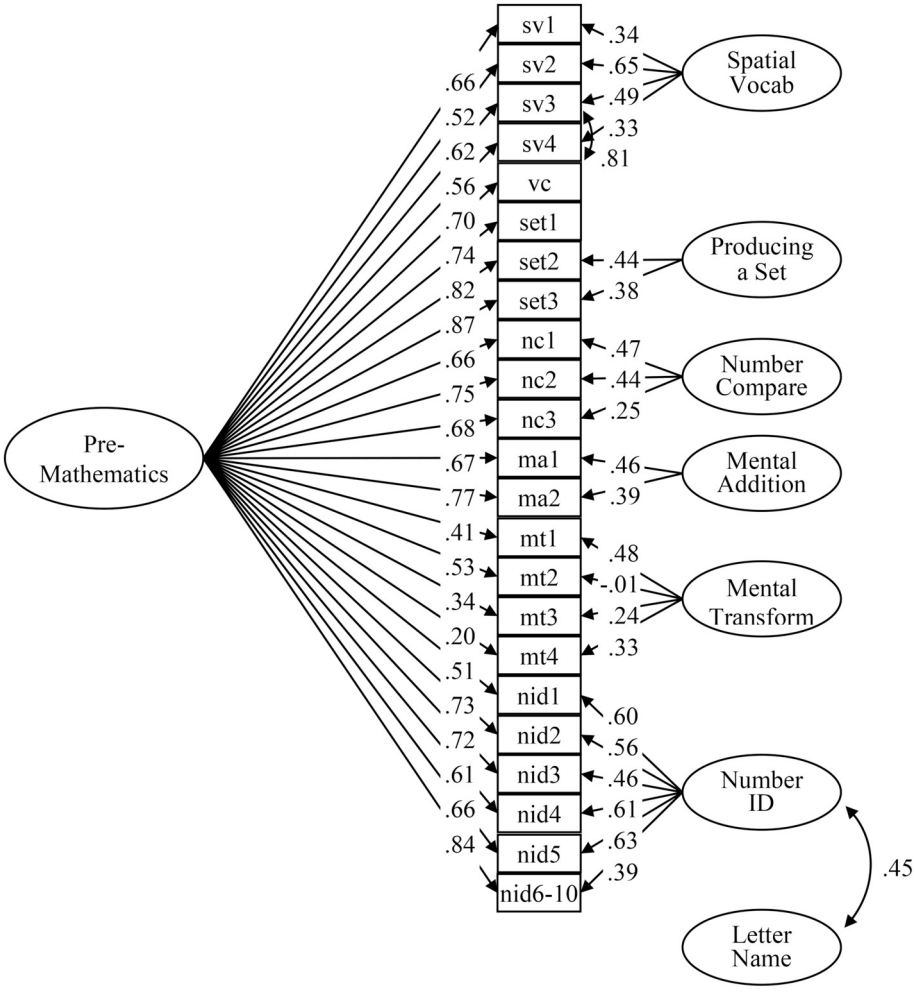

Fig. 1. Standardized path coefficients for the Pre-Mathematics domain of the final multi-domain child development and learning model (DL_M1). vc = verbal counting. Correlations between Pre-Mathematics and Pre-Literacy ($r = 0.90$), Executive Functioning ($r = 0.88$), Fine Motor Skills ($r = 0.69$), and Socioemotional Cognition ($r = 0.68$) omitted from diagram for simplicity. $N = 684$.

discriminating among children with approximately average Attention/Self-Regulation and Social Competence, and slightly more Problem Behaviors. For Attention/Self-Regulation, reliability was ≥0.70 for children between −2.60 SD and 1.27 SD units from the mean, and for Social Competence, reliability was ≥0.70 for children between −2.53 SD and 0.74 SD units from the mean. Overall, evidence is limited for using the current set of self-regulation items to measure problem behaviors, and thus this factor was not considered as a covariate in the structural analyses.

Family engagement with the child (parent-report)

Global fit statistics and indices are provided in Table 4. The hypothesized one-factor model (FE_M1) demonstrated adequate fit, but modification indices indicated local misfit in the form of a conditional dependency between two items related to telling stories (fe2) and singing songs/lullabies (fe3). Adding a residual correlation (FE_M2) significantly improved global fit and eliminated local misfit.

See Fig. 7 for the standardized path coefficients corresponding to the final model. All loadings were significantly different from zero. Fig. 3s in the Supplement illustrates the total information function summed across the six family engagement items. The function peaks at −0.47 SD (maximum reliability = 0.85) indicating that the items do best at discriminating among families who are slightly below average on the latent trait. Reliability was ≥0.70 for families between −1.56 SD and 0.61 SD away from the mean on the latent family engagement continuum.

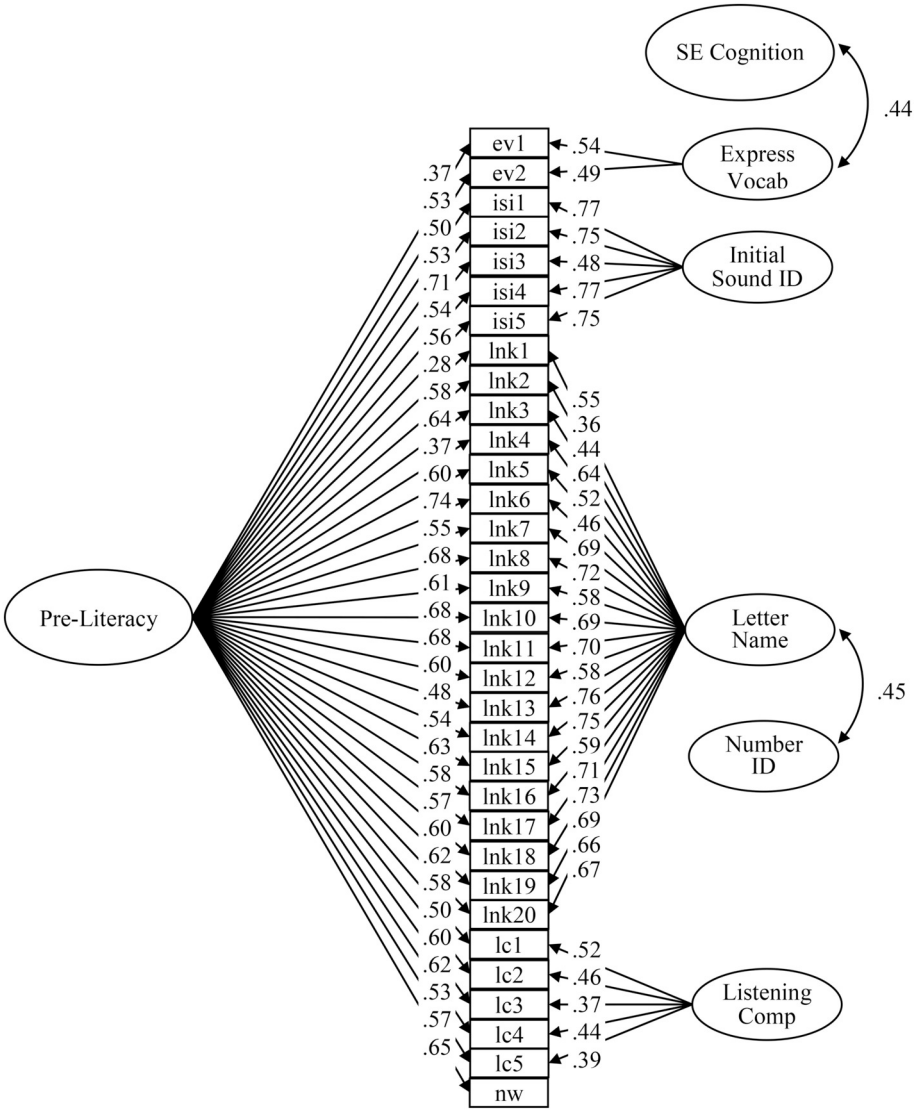

Fig. 2. Standardized path coefficients for the Pre-Literacy domain of the final multi-domain child development and learning model (DL_M1). nw = name writing. SE = socioemotional. Correlations between Pre-Literacy and Pre-Mathematics (r = 0.90), Executive Functioning (r = 0.84), Fine Motor Skills (r = 0.77), and Socioemotional Cognition (r = 0.75) omitted from diagram for simplicity. N = 684.

Structural analyses

Question 2: How do child characteristics including social/emotional development and family assets and home environments contribute to children's learning and development?

Bivariate zero-order relationships

See Table 6 for the bivariate associations between child/family characteristics and child outcomes. Child attention/self-regulation and social competence and family assets were significantly and positively related to direct assessment outcomes. Correlations ranged from r = 0.30 to 0.36 for attention/self-regulation, r = 0.26 to 0.32 for social competence, and r = 0.38 to 0.43 for family assets. Family engagement was significantly and positively related to three of the five outcomes (pre-literacy, executive functioning, and socioemotional cognition), with the significant correlations ranging from r = 0.09 to 0.12. Gender differences emerged in only one domain, with girls demonstrating significantly less socioemotional cognition than boys (standardized mean difference [SMD] = −0.15). Child age and preprimary attendance were not significantly related to child outcomes.

Path analysis

See Table 7 for the estimated path coefficients. The model was saturated, so fit could not be established. Taken together, the predictors accounted for the most variability in children's pre-literacy skills (R^2 = 0.26) and the least variability in children's socioemotional cognition (R^2 = 0.19). After controlling for the other variables in the model, family assets were significantly associated with each of the child outcomes, with standardized estimates ranging from $\widehat{\beta}$ = 0.33 to 0.37. Child attention/self-regulation was significantly associated with all but fine motor skills, with standardized estimates for the significant effects ranging from $\widehat{\beta}$ =. 33 to 0.39.[3] Age was significantly associated with Pre-Mathematics skills ($\widehat{\beta}$ = 0.12) and Executive Functioning ($\widehat{\beta}$ = 0.10), but not the other child outcomes.[4] After controlling for the other variables

[3] When children with missing age information were included in the analyses, the association between child attention/self-regulation and fine motor skills was significant, and the association between attention/self-regulation and socioemotional cognition was non-significant.

[4] There were no significant associations with age when children with missing age information were included in the analyses.

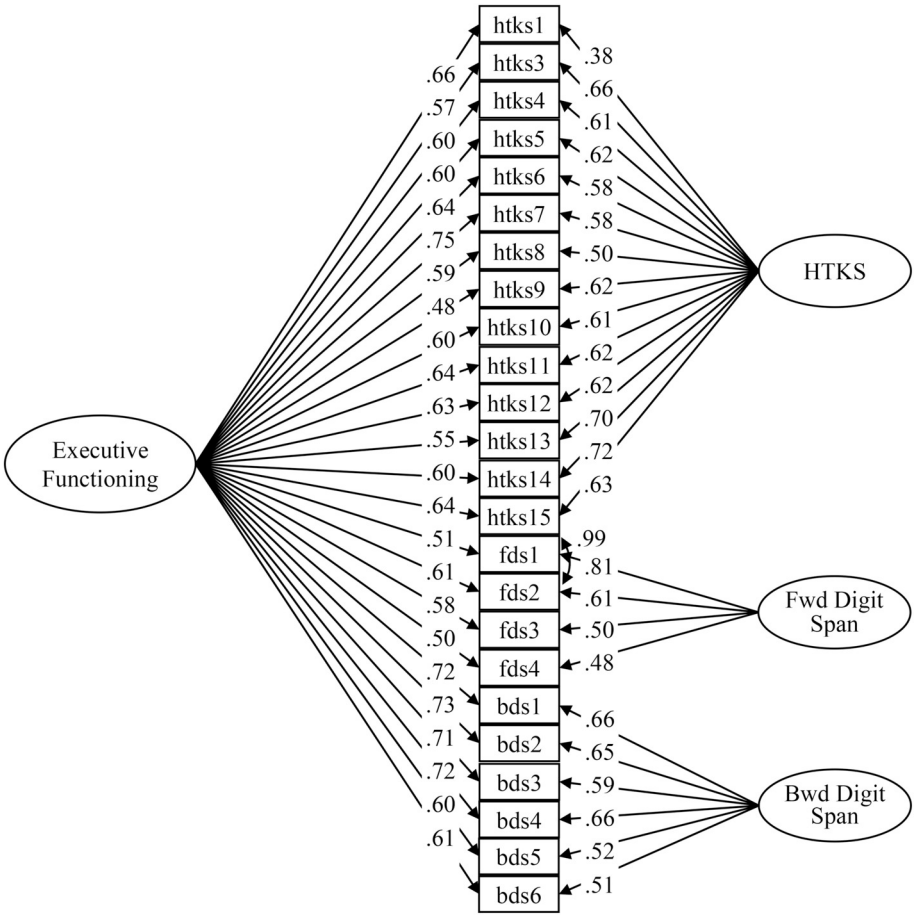

Fig. 3. Standardized path coefficients for the Executive Functioning domain of the final multi-domain child development and learning model (DL_M1). HTKS = head, toes, knees, shoulders. Correlations between Executive Functioning and Pre-Mathematics (r = 0.88), Pre-Literacy (r = 0.84), Fine Motor Skills (r = 0.58), and Socioemotional Cognition (r = 0.74) omitted from diagram for simplicity. N = 684.

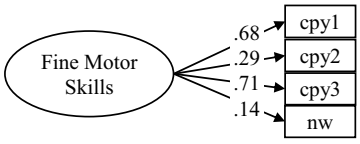

Fig. 4. Standardized path coefficients for the Fine Motor Skills domain of the final multi-domain child development and learning model (DL_M1). cpy = copying. nw = name writing. Correlations between Fine Motor Skills and Pre-Mathematics (r = 0.69), Pre-Literacy (r = 0.77), Executive Functioning (r = 0.58), and Socioemotional Cognition (r = 0.46) omitted from diagram for simplicity. N = 684.

in the model, child sex, pre-primary attendance, social competence, and family engagement were not associated with child outcomes.

Discussion

The purpose of this study was to examine children's development and learning at the start of primary school, beginning with a global frame of school readiness and examining its applicability to the Tanzanian context. Establishing validity evidence of child development and learning measures is essential for accurate assessment of progress towards country and global goals for early childhood development. This study adds to a small but growing body of work evaluating the validity evidence of tools to measure child development in low- and middle-income countries. Results also demonstrate that child development in this low-income country is influenced by predicted factors based on research in mostly high-income countries—an important confirmation of cross-country relevance and the global call for investment in early childhood development (Shonkoff, Radner, & Foote, 2017).

This study brings several new and important contributions. First, analyses of two measures—a teacher report of children's social/emotional and self-regulatory skills and a direct measure of children's development and learning at the start of formal schooling administered by trained enumerators—supported factors of children's learning and development at the start of school that were largely consistent with expectations based on existing conceptualizations of school readiness. Second, these measures showed predicted associations with family

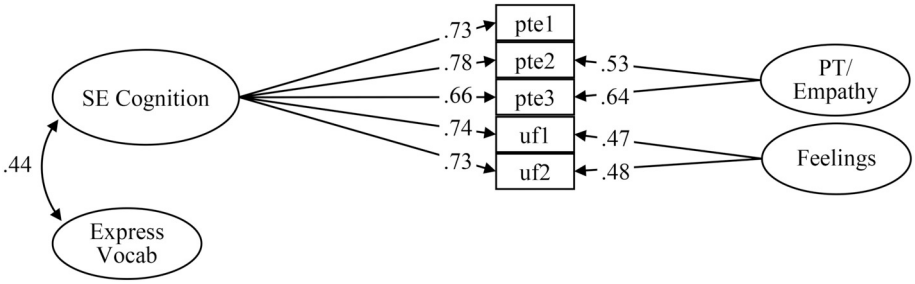

Fig. 5. Standardized path coefficients for the Socioemotional (SE) Cognition domain of the final multi-domain child development and learning model (DL_M1). pte = perspective-taking/empathy. uf = understanding feelings. Correlations between Socioemotional Cognition and Pre-Mathematics (r = 0.68), Pre-Literacy (r = 0.75), Executive Functioning (r = 0.74), and Fine Motor Skills (r = 0.46) omitted from diagram for simplicity. N = 684.

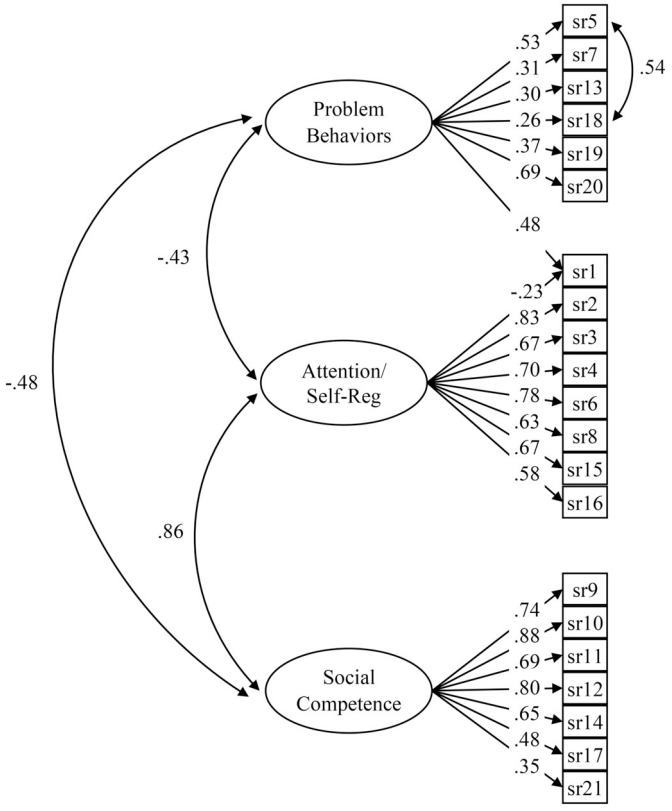

Fig. 6. Standardized path coefficients for the final teacher-reported child social/emotional and self-regulatory skills model (SR_M2). $N = 669$.

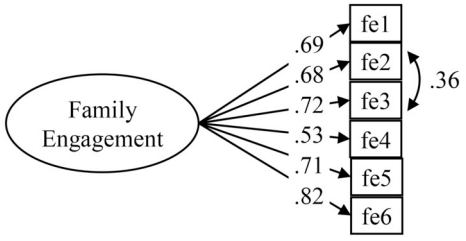

Fig. 7. Standardized path coefficients for the final family engagement model (FE_M2). $N = 561$.

assets, consistent with existing work on the impact of family assets on child development. Third, social competence and self-regulatory skills as rated by teachers were associated with children's learning as measured by the direct assessment. Overall, findings from this study demonstrate validity evidence of the MELQO MODEL instruments, including both an enumerator-administered test of child development and

Table 7
Estimated unique associations between child/family characteristics and child development and learning outcomes.

	$\hat{b}$	$\widehat{SE}$	p	$\hat{\beta}$	R^2
Pre-Mathematics[a,b,c,d] regressed on					0.24
Child age	0.10	0.05	0.028	0.11	
Child gender	−0.08	0.07	0.233	−0.10	
Child preprimary attendance	−0.05	0.14	0.740	−0.06	
Child attention/self-regulation	0.39	0.12	0.001	0.41	
Child social competence	−0.10	0.13	0.441	−0.11	
Family assets	0.11	0.02	< 0.001	0.33	
Family engagement	0.03	0.05	0.480	0.03	
Pre-Literacy[a,e,f,g] regressed on					0.27
Child age	0.07	0.05	0.114	0.08	
Child gender	−0.05	0.06	0.419	−0.06	
Child preprimary attendance	−0.04	0.14	0.784	−0.05	
Child attention/self-regulation	0.43	0.11	< 0.001	0.45	
Child social competence	−0.14	0.13	0.270	−0.15	
Family assets	0.12	0.02	< 0.001	0.36	
Family engagement	0.04	0.05	0.417	0.03	
Executive Functioning[b,e,h,i] regressed on					0.24
Child age	0.08	0.04	0.068	0.09	
Child gender	−0.08	0.06	0.228	−0.10	
Child preprimary attendance	−0.10	0.14	0.492	−0.12	
Child attention/self-regulation	0.39	0.12	0.001	0.42	
Child social competence	−0.12	0.13	0.340	−0.13	
Family assets	0.11	0.02	< 0.001	0.33	
Family engagement	0.06	0.05	0.202	0.06	
Fine Motor Skills[c,f,h,j] regressed on					0.23
Child age	0.09	0.05	0.057	0.10	
Child gender	0.03	0.06	0.651	0.04	
Child preprimary attendance	−0.01	0.15	0.950	−0.01	
Child attention/self-regulation	0.31	0.12	0.010	0.35	
Child social competence	−0.06	0.14	0.665	−0.07	
Family assets	0.11	0.02	< 0.001	0.34	
Family engagement	−0.01	0.05	0.847	−0.01	
Socioemotional Cognition[d,g,i,j] regressed on					0.19
Child age	0.04	0.04	0.318	0.05	
Child gender	−0.11	0.06	0.095	−0.14	
Child preprimary attendance	−0.18	0.16	0.240	−0.23	
Child attention/self-regulation	0.31	0.11	0.004	0.34	
Child social competence	−0.10	0.12	0.388	−0.11	
Family assets	0.11	0.02	< 0.001	0.33	
Family engagement	0.00	0.05	0.970	0.00	

Note. $N = 459$. Only children between the ages of 4 and 8 were included in the analyses. Cluster-robust SEs estimated to adjust for nesting of children within schools. Reference groups are males and non-attenders of preprimary. StdYX standardization used for continuous predictors (age, social competence, attention/self-regulation, assets, engagement) and StdY standardization used for dichotomous predictors (gender, preprimary attendance).[a]Residual $r = 0.95$. [b]Residual $r = 0.95$. [c]Residual $r = 0.78$. [d]Residual $r = 0.73$. [e]Residual $r = 0.92$. [f]Residual $r = 0.85$. [g]Residual $r = 0.81$. [h]Residual $r = 0.69$. [i]Residual $r = 0.79$. [j]Residual $r = 0.56$.

Table 6
Bivariate associations with child development and learning outcomes.

	Pre-mathematics	Pre-literacy	Executive functioning	Fine motor skills	Socioemotional cognition
Child age (years)[a]	$r = 0.07$	$r = 0.03$	$r = 0.05$	$r = 0.03$	$r = 0.00$
Child gender[b]	$SMD = -0.13$	$SMD = -0.11$	$SMD = -0.12$	$SMD = -0.09$	$SMD = -0.15$*
Child preprimary attendance[c]	$SMD = 0.12$	$SMD = 0.16$	$SMD = 0.08$	$SMD = 0.19$	$SMD = -0.03$
Child attention/self-regulation	$r = 0.35$***	$r = 0.36$***	$r = 0.36$***	$r = 0.31$***	$r = 0.30$***
Child social competence	$r = 0.32$***	$r = 0.32$***	$r = 0.32$***	$r = 0.28$***	$r = 0.26$***
Family assets	$r = 0.38$***	$r = 0.43$***	$r = 0.38$***	$r = 0.40$***	$r = 0.39$***
Family engagement	$r = 0.09$	$r = 0.10$*	$r = 0.12$*	$r = 0.06$	$r = 0.09$*

Note. $N = 684$ except for age ($N = 554$) and preprimary attendance ($N = 568$). r = Pearson product-moment correlation. SMD = standardized mean difference using StdY standardization. Cluster-robust SEs estimated to adjust for nesting of children within schools. [a]Only children between the ages of 4 and 8 were included in the bivariate analysis. [b]Reference group = males; [c]Reference group = non-attenders of preprimary.
***$p < .001$. **$p < .01$. *$p < .05$.

learning, and a teacher report of social/emotional and self-regulatory skills for use in low-income countries.

This study is one of the first to show strong, consistent associations between teacher ratings of social/emotional and self-regulatory skills and children's learning in a low-income country, as assessed through a direct assessment, at the start of primary school. These findings are consistent with work from the United States, documenting the importance of self-regulation in children's early academic achievement (e.g., Garon, Bryson, & Smith, 2008; Mashburn et al., 2008), and builds on previous research from Tanzania, demonstrating the importance of teacher/child relationships in children's school adjustment (Shavega, Brugman, & van Tuijl, 2014; Shavega, van Tuijl, & Brugman, 2015). The study design, which integrates three independent sources of information on children's development and family environments from parents, teachers and the child, adds to the strength of the findings. Results from this study provide compelling evidence that social/emotional and self-regulatory skills are essential elements of children's learning and development at the start of school, and teachers can provide dependable ratings of these skills. However, further modifications of this scale should include revised approaches to measuring behavior problems, which were not as successfully indexed by the items as other domains of social/emotional development. In sum, though, as demonstrated in high- and middle-income countries, teacher ratings could be a promising avenue for collecting information on children's development at the start of formal schooling. Not surprisingly, given the strong influence of the Early Development Instrument (EDI) on the design of the teacher report measure, the results align with the reliability and validity evidence of the teacher-reported EDI (Janus & Offord, 2007). A focus on social/emotional and self-regulatory skills may provide valuable information about children's readiness for formal education, independent of assessments of pre-academic skills.

Results also demonstrate that powerful inequities exist within populations of young children in Tanzania, consistent with results from other samples (e.g., Fernald et al., 2011). Family assets were consistently one of the strongest predictors of children's development and learning, regardless of pre-primary education, and even after accounting for social/emotional skills that promote children's learning. Although family home environments have been shown to have a profound influence on children's learning in other sub-Saharan African countries (Wolf & McCoy, 2017), we did not find that family engagement had a significant effect on learning after controlling for other family and child characteristics, nor did inclusion of the family engagement scale mitigate the impact of family assets on children's learning. This could be due to limitations in our measurement of family engagement, which was derived from a global scale and may not correspond to parents' views of early learning (e.g., Kabay et al., 2017), and the scale also may have focused on activities that were more appropriate for younger children, given the age of children entering primary school in Tanzania. But failure to confirm an association between family engagement and child development and learning also may reflect the complexity of the role of family engagement in learning across diverse contexts. In another sample of Tanzanian families, Ngorosho (2010), for example, found a significant effect of family assets on primary school learning but did not find effects of family engagement. Together, results suggest that family assets exert a powerful influence on child learning, and that definitions and measurement of family engagement may require more contextually-bound definitions to capture adequately. We also did not see a significant impact of pre-primary education on children's learning, which is not consistent with numerous studies documenting the positive impacts of pre-primary education on children's learning in developing countries (Rao, Sun, Chen, & Ip, 2017). Two reasons for the lack of findings in this study include the high percentage of children who had access to pre-primary education, which created low variability in exposure within our sample; and highly variable quality in children's learning environments by region in Tanzania (Mtahabwa & Rao, 2010). Future studies should examine the possible roles of family learning environments and quality in pre-primary education in mitigating the impact of family assets on children's learning.

As well, of note is the high rate of undernutrition among children in Tanzania, estimated at 34% in 2016 (Development Initiatives, 2018) and its negative impacts on children's development (Sudfeld et al., 2015), which was not measured in this study. Thus it is possible that as many as a third of the children in our sample have also suffered from undernutrition, which in turn has profound implications for their development and learning at the start of school. Future studies should integrate health and nutrition information into studies of early childhood development at the start of school especially using longitudinal designs, to raise awareness within the education sector of the importance of addressing health and nutrition status in the first years of life.

Overall, the SDGs have the potential to instigate notable progress in promoting early childhood development. While measurement is one mechanism of many, it is critical to evaluate carefully the standards and evidence for validity of measures intended to inform policy and practice. Both psychometric properties and the alignment of the content of measures with local policies and practices is important. This study provides a unique contribution to the early childhood literature, by establishing that the MODEL scale demonstrates psychometric strength and is appropriate for application to national policy by the Tanzanian government, as evidenced by the review and approval of the items as workable indicators of the national curricula. The combination of these forms of validity evidence lends weight to the idea of a global conceptualization of school readiness that may be broadly useful to policymakers in many countries. Although the findings point in the direction of validation of a global approach to school readiness, it is important to note that the evidence documented here should be cross-validated with other methodologies and theoretical perspectives. For example, previous theoretical work has highlighted the significance of cultural settings, and especially degree of industrialization, on cognitive development (Gauvain & Munroe, 2009). This study, by presenting a global frame rather than working from the ground up to develop a locally-defined frame, did not offer an opportunity to test alternative definitions of school readiness.

The growing emphasis on school readiness has the potential to unlock new resources and attention on the importance of young children's development. As countries develop early childhood systems and the emphasis on reaching global and national goals in education becomes stronger, measurement may become a higher priority. Our findings document validity evidence of the MELQO MODEL instruments in Tanzania and will be complemented by work underway to evaluate the scales in other countries. Results also demonstrate the important contribution of children's social and emotional development, and the value of teacher ratings for providing an indication of the overall functioning of children at the start of the primary school years. Perhaps most importantly, in the context of global goals to reduce inequities in education, it is critical to note that family assets play a strong and consistent role in influencing children's learning and development at the start of primary school, even when pre-primary education is provided. Future work should explore the roles of health and nutrition status, as well as the overall quality of pre-primary education, as possible routes to increasing equity in learning outcomes for all children.

Funding

This work was supported by Dubai Cares, the Children's Investment Fund Foundation, UNICEF and the World Bank.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.appdev.2019.02.003.

References

Bentler, P. M. (1990). Comparative fit indices in structural models. *Psychological Bulletin, 107*, 238–246. https://doi.org/10.1037/0033-2909.107.2.238.

Blair, C. (2002). School readiness: Integrating cognition and emotion in a neurobiological conceptualization of children's functioning at school entry. *American Psychologist, 57*(2), 111.

Browne, M. W., & Cudeck, R. (1993). Alternate ways of assessing model fit. In K. A. Bollen, & J. S. Long (Eds.). *Testing Structural Equation Models* (pp. 136–162). Newbury Park, CA: Sage Publications.

Darling-Churchill, K. E., & Lippman, L. (2016). Early childhood social and emotional development: Advancing the field of measurement. *Journal of Applied Developmental Psychology, 45*, 1–7. https://doi.org/10.1016/j.appdev.2016.02.002.

Development Initiatives (2018). *2018 Global Nutrition Report: Shining a Light to Spur Action on Nutrition.* Bristol, UK: Development Initiatives.

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., ... Sexton, H. (2007). School readiness and later achievement. *Developmental Psychology, 43*(6), 1428.

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

Fernald, L. C. H., Weber, A., Galasso, E., & Ratsifandrihamanana, L. (2011). Socioeconomic gradients and child development in a very low income population: Evidence from Madagascar. *Developmental Science, 14*(4), 832–847. https://doi.org/10.1111/j.1467-7687.2010.01032.x.

Filmer, D., & Pritchett, L. (1998). *Estimating wealth effects without expenditure data–or tears: An application to educational enrollments in states of India.* Washington, DC: World Bank Policy Research Working Papers No1994.

Garon, N., Bryson, S. E., & Smith, I. M. (2008). Executive function in preschoolers: A review using an integrative framework. *Psychological Bulletin, 134*(1), 31–60. https://doi.org/10.1037/0033-2909.134.1.31.

Gauvain, M., & Munroe, R. L. (2009). Contributions of societal modernity to cognitive development: A comparison of four cultures. *Child Development, 80*, 1628–1642. https://doi.org/10.1111/j.1467-8624.2009.01358.x.

Hanushek, E. A., & Woessmann, L. (2008). The role of cognitive skills in economic development. *Journal of Economic Literature, 46*(3), 607–668. https://doi.org/10.1257/jel.46.3.607.

Hanushek, E. A., & Woessmann, L. (2012). Schooling, educational achievement, and the Latin American growth puzzle. *Journal of Development Economics, 99*, 497–512. https://doi.org/10.1016/j.jdeveco.2012.06.004.

Hu, L.-T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*, 1–55. https://doi.org/10.1080/10705519909540118.

Janus, M., & Offord, D. (2007). Development and psychometric properties of the early development instrument (EDI): A measure of children's school readiness. *Canadian Journal of Behavioural Science, 39*(1), 1–22. https://doi.org/10.1037/cjbs2007001.

Kabay, S., Wolf, S., & Yoshikawa, H. (2017). “So that his mind will open”: Parental perceptions of early childhood education in urbanizing Ghana. *International Journal of Educational Development, 57*, 44–53. https://doi.org/10.1016/j.ijedudev.2017.08.013.

Kline, R. B. (2016). *Principles and Practice of Structural Equation Modeling* (4th ed.). New York, NY: Guilford Press.

Mashburn, A. J., Pianta, R. C., Hamre, B. K., Downer, J. T., Barbarin, O. A., Bryant, D., ... Howes, C. (2008). Measures of classroom quality in prekindergarten and children's development of academic, language, and social skills. *Child Development, 79*(3), 732–749. https://doi.org/10.1111/j.1467-8624.2008.01154.x.

McClelland, M. M., Cameron, C. E., Duncan, R., Bowles, R. P., Acock, A. C., Miao, A., & Pratt, M. E. (2014). Predictors of early growth in academic achievement: The head-toes-knees-shoulders task. *Frontiers in Psychology, 5*, 599.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50*, 741–749. https://doi.org/10.1037/0003-066X.50.9.741.

Mtahabwa, L., & Rao, N. (2010). Pre-primary education in Tanzania: Observations from urban and rural classrooms. *International Journal of Educational Development, 30*(3), 227–235. https://doi.org/10.1016/j.ijedudev.2009.10.002.

Muthén, L. K., & Muthén, B. O. (1998-2015). *Mplus User's Guide* (Version 7)Los Angeles, CA: Author.

Ngorosho, D. (2010). Reading and writing ability in relation to home environment: A study in primary education in rural Tanzania. *Child Indicators Research, 4*(3), 369–388. https://doi.org/10.1007/s12187-010-9089-8.

Raikes, A. (2017). Measuring child development and learning. *European Journal of Education, 52*(4), 511–522. https://doi.org/10.1111/ejed.12249.

Raikes, A., Yoshikawa, H., Britto, P. R., & Iruka, I. (2017). Children, youth and developmental science in the 2015–2030 global sustainable development goals. *Social Policy Report, 30*(3), https://doi.org/10.1002/j.2379-3988.2017.tb00088.x.

Rao, N., Sun, J., Chen, E. E., & Ip, P. (2017). Effectiveness of early childhood interventions in promoting cognitive development in developing countries: A systematic review and meta-analysis. *Hong Kong Journal of Paediatrics (New Series), 22*(1), 14–25.

Rao, N., Sun, J., Ng, S. S. N., Ma, K., Becher, Y., Lee, D., ... Ip, P. (2013). The Hong Kong early child development scale: A validation study. *Child Indicators Research, 6*(1), 115–135.

Raver, C. (2002). Emotions matter: Making the case for the role of young children's emotional development for early school readiness. *Society for Research in Child Development Social Policy Report, 16*(3), 3–19. https://doi.org/10.1002/j.2379-3988.2002.tb00041.x.

Rispoli, K. M., McGoey, K. E., Koziol, N. A., & Schreiber, J. B. (2013). The relation of parenting, child temperament, and attachment security in early childhood to social competence at school entry. *Journal of School Psychology, 51*(5), 643–658. https://doi.org/10.1016/j.jsp.2013.05.007.

RTI International (2016). *Assistance to Basic Education: All Children Reading (ABE-ACR) Findings Report, Tanzania National Early Grade Reading Assessment (EGRA).* (Retrieved from https://ierc-publicfiles.s3.amazonaws.com/public/resources/FdngsRpt_TZNtnlEGRA_08-Nov2016_FNL.pdf).

Serpell, R., Jere-Folotiya, J., Chansa-Kabali, T., Munachaka, J., Maumbi, M. N., Yalukanda, C., ... Lyytinen, H. (2017). A culturally sensitive approach to promoting initial literacy development in Africa: Ongoing and planned research and development at the University of Zambia's Centre for Promotion of literacy in Sub-Saharan Africa (CAPOLSA). In A. Abukar, & F. J. R. van de Vijver (Eds.). *Handbook of Applied Developmental Science in Sub-Saharan Africa* (pp. 313–334). New York, NY: Springer.

Shavega, T. J., Brugman, D., & van Tuijl, C. (2014). Children's behavioral adjustment in pre-primary schools in Tanzania: A multilevel approach. *Early Education and Development, 25*(3), 356–380. https://doi.org/10.1080/10409289.2013.807722.

Shavega, T. J., van Tuijl, C., & Brugman, D. (2015). Teachers’ perception of children’s behavioral adjustment in Tanzanian preprimary schools and their relationship to teachers’ cultural beliefs regarding obedience, cooperation, and play. *Infants & Young Children, 28*(1), 88–108. https://doi.org/10.1097/IYC.0000000000000026.

Shonkoff, J. P., Radner, J. M., & Foote, N. (2017). Expanding the evidence base to drive more productive early childhood investment. *The Lancet, 389*(10064), 14–16. https://doi.org/10.1016/S0140-6736(16)31702-0.

Snow, K. L. (2006). Measuring school readiness: Conceptual and practical considerations. *Early Education and Development, 17*(1), 7–41. https://doi.org/10.1207/s15566935eed1701_2.

Sudfeld, C. R., McCoy, D. C., Fink, G., Muhihi, A., Bellinger, D. C., Masanja, H., ... Fawzi, W. W. (2015). Malnutrition and its determinants are associated with suboptimal cognitive, communication, and motor development in Tanzanian children–3. *The Journal of Nutrition, 145*(12), 2705–2714. https://doi.org/10.3945/jn.115.215996.

UNESCO (2017). *Overview of MELQO: Measuring Early Learning Quality Outcomes.* Jointly published by UNESCO, UNICEF, World Bank & Brookings InstitutionParis: UNESCO.

Wolf, S., Halpin, P., Yoshikawa, H., Dowd, A. J., Pisani, L., & Borisova, I. (2017). Measuring school readiness globally: Assessing the construct validity and measurement invariance of the international development and early learning assessment (IDELA) in Ethiopia. *Early Childhood Research Quarterly, 41*, 21–36. https://doi.org/10.1016/j.ecresq.2017.05.001.

Wolf, S., & McCoy, D. C. (2017). Household socioeconomic status and parental investments: Direct and indirect relations with school readiness in Ghana. *Child Development*. https://doi.org/10.1111/cdev.12899 Advance online publication.

World Bank (2016). *SABER: Early Childhood Development. World Bank Global Education Practice, Washington, DC.* (Retrieved from http://wbgfiles.worldbank.org/documents/hdn/ed/saber/supporting_doc/brief/SABER_ECD_Brief.pdf).

World Bank (2018). *Gross enrolment data on pre-primary education.* (Retrieved from https://data.worldbank.org/indicator/SE.PRE.ENRR).