Abstract

The East Asia-Pacific Early Child Development Scales represent the first effort to create a developmental assessment tool on the basis of the diverse cultures and values of a range of countries within a world region. The Scales were administered to a representative sample of 7,757 children (3,869 girls), ranging in age from 36 to 71 months, from Cambodia, China, Mongolia, Timor-Leste, Papua New Guinea and Vanuatu. In all six countries, child development scores increased with age and urban children consistently performed better than rural children. The gap between urban and rural children widened with age in Cambodia. There were significant gender differences in total scores, favouring girls in four countries. Results illustrate commonalities and variations in trajectories of children's early development across contexts. Reasons for the findings and their implications are discussed.

Keywords: early child development; East Asia Pacific Early Child Development Scales (EAP-ECDS); measurement; low- and middle- income countries

1. Introduction

A burgeoning body of research indicates that experiences during the early childhood period influence health, development, and productivity throughout the life course. However, detailed knowledge regarding developmental patterns as well as risk and protective factors of early childhood development across different developmental domains in low- and middle-income countries (LMICs) is still lacking. The majority of our knowledge regarding child development is derived from Western, educated, industrialized, rich and democratic (WEIRD) societies (Henrich, Heine, & Norenzayan, 2010), but only 18% of children are born in high-income countries; the remaining 82% live in LMICs (UNESCO, 2015). The paucity of culturally appropriate assessment tools contributes to the dearth of valid data on early child development in LMICs. More robust research is needed to ensure that our knowledge of child development is applicable to children living in LMICs who face serious risks to their development (Richter et al., 2016).Moreover, children living in the world's poorest countries are twice as likely to be stunted and to die before their fifth birthday compared to children living in the world's richest countries (UNICEF 2015b). To reduce inequality, it is crucial that policy be informed by comparable child outcome data across different cultural contexts. In the area of early childhood, this has traditionally been limited to mortality and anthropometric indicators. Stunting data, for example, are available from countries across the world as assessment is universally agreed upon and straightforward to collect. It is also possible to compare educational achievements of school-age children across countries with tools such as the Programme for International Student Assessment (PISA) which has now been completed in more than 70 countries (OECD, 2014), the Trends in International Mathematics and Science Study (TIMMS) (Mullis, Martin, Foy & Arora. 2012) and the Progress in International Reading Literacy Study (PIRLS) (Mullis, Martin, Foy & Drucker,

2012). Emphases on literacy and numeracy has resulted in similar school curricula across the world and thus allows for the use of common measurement items and data comparison. No such cross-cultural comparisons have been possible for preschool aged children to date.

Against this background, this paper outlines the development of a culturally appropriate population-based measure of early child development in LMICs in the East Asia-Pacific - the East Asia-Pacific Early Child Development Scales (EAP-ECDS). Utilizing data from this tool, herein we aim to assess diversity and examine correlates of early childhood development within and across six LMICs in order to inform social policy with the ultimate aim of ensuring that the maximum number of children are able to reach their developmental potential.

1.1. Measuring holistic childhood development

The measurement of holistic child development, inclusive of cognitive and non-cognitive developmental domains, is complex as it must take into account culture, language and research traditions (Hambleton, Merenda & Speilberger, 2005). Such measures must not only be culturally appropriate in terms of administration method and materials, but also reflect the values and skills important within a particular cultural or national framework. This makes both regional and international comparison difficult, and indeed, challenging to implement at a global scale with developmental items standardized for children from different countries. This is most often due to concerns that tools from Western societies are not valid for use in other countries due to differences in culture and context and lack of consensus on assessment techniques and constructs to be measured (Rao, Sun & Becher, 2015).

A good example of an effort to measure and compare child development across countries is the Early Development Instrument (EDI): an 100-item population-based measure of holistic child development completed by teachers of children in their first year of full-time school (Janus

& Offord 2007; Janus et al., 2016). The EDI measures children's development across five key domains: physical health and wellbeing, language and cognition, social competence, emotional maturity, and communication skills and general knowledge. To date, the EDI has been used across more than 20 countries including Canada, Spain, Italy, Australia, Japan, and Indonesia, though to do so, adaptation and piloting was required to establish validity and reliability of the instrument within a country prior to cross-country comparison. The last few years have seen the development of three new tools designed to assess and compare child development across countries. First, the Inter-American Development Bank's Regional Project on Child Development Indicators (PRIDI) is a 21-item test of the development of children aged 24 to 59 months old. Completed in households, the PRIDI involves direct observation as well as parent report and measures children's cognitive, language, motor and socio-emotional development. The PRIDI has been validated in Costa Rica, Nicaragua, Paraguay and Peru (Verdisco, Cueto, Thompson, & Neuschmidt, 2015). Similarly, the International Development and Early Learning Assessment (IDELA), developed by Save the Children, is a population-level play-based direct assessment tool designed to measure the holistic early learning and development of children three to six years of age (Pisani, Borisova & Dowd, 2015). Specifically, the IDELA assesses children's motor skills, early language and literacy, early numeracy and problem solving, socio-emotional development as well as their approaches to learning. The IDELA has been used in more than 20 countries including Australia, England, Canada and the Unites States, and the tool's short length allows for easy adaptability to varied contexts, and thus international comparison of results. Third, a new population measure of child development has been developed as part of the Measurement of Early Learning and Quality Outcomes (MELQO) project, initiated in 2014 with a partnership between the Brookings Institution, World Bank,

UNICEF and UNESCO. The project's tool – the Measurement of Development and Early Learning (MODEL) – collects information on the development and learning of children aged three to seven years via a direct assessment as well as a parent and/or teacher report. More specifically, the MODEL assesses children's socio-emotional skills, a range of pre-academic skills such as language, early literacy and numeracy, as well as executive functioning. To date, the MODEL has been used in 10 LMICs including Colombia, Bangladesh, Sudan and Tanzania (World Bank, 2016b).

These efforts are encouraging considering the great need to develop culturally appropriate tools to validly measure early childhood development, particularly in LMICs. Great efforts have been exerted to ensure the PRIDI, IDELA and MELQO are sensitive to the contexts of LMICs. That stated, the conceptual roots of the EAP-ECDS are in contrast to that of the above projects; indeed it is arguable that the EAP-ECDS is a more comprehensive measure of development than the other tools.

1.2. The current study

Children's ability at school entry sets their academic trajectories and has important implications across the life course (see for example Barnett, 2011; Brinkman et al, 2013; Heckman, 2008; Duncan et al 2007). Millions of children in the East Asia-Pacific do not receive early education, and as a result, many children who enter school do not have the cognitive and non-cognitive skills required to learn effectively. Research demonstrates that these children have a higher likelihood of dropping out of school or repeating a grade due to a lack of readiness for formal schooling, and then experience the associated economic, health, and crime-related consequences (Louden, Chan, Elkins & Greaves, 2000; McCain & Mustard, 1999). Child

development data are needed in order to take stock of and address this issue, and importantly, a culturally-appropriate measure is required to achieve this task.

The EAP-ECDS were created to assess child development at a population level in the East Asia-Pacific region. Prior to the development of the Scales, there existed no measure of children's development that took into account cultural and contextual diversity within and across the region. First developed in 2010 and piloted throughout 2010-2012, the Scales were administered to representative samples of children from Cambodia, China, Mongolia, Vanuatu, Timor-Leste and Papua New Guinea in order to assess their validity and reliability (Rao et al., 2014; Rao, Sun & Becher, 2015). Together, these six countries were selected as they demonstrate the contextual and cultural diversity that exists across the East Asia-Pacific, and the importance of having a culturally-appropriate measure of child development that can highlight meaningful differences across the region. To illustrate: population size among these countries varies from 264,000 in Vanuatu to 1.4 billion in China; gross domestic product per capita income in 2014 ranged from US$1,094.60 in Cambodia to US$7,590 in China; prevalence of child stunting varies from 58% in Timor-Leste to 9% in China (UNESCO, 2014); and gross enrollment ratio in pre-primary education ranged from 15% in Cambodia to 85% in Mongolia in 2012 (UNESCO, 2015; World Bank, 2016). Differences across countries in levels of development, educational poverty, as well as cultural, contextual and linguistic differences all impact the measurement of child development.

This paper describes the development of the EAP-ECDS – the first regional early child development scale that was originated, developed and then validated specifically with and for children from the East Asia-Pacific. First we report on the Scales' reliability and validity across the six countries. Next, we consider the influence of age, gender and urban-rural residence on the

Scales using country-specific analyses. Research has long revealed gender differences across different domains of children's development (see for example Rose & Rudolph, 2006; Votruba-Drzal & Lindsay chase-Lansdale, 2004). Naturally, age also has a key influence on children's developmental outcomes (see for example (Pons, Lawson, Harris, & De Rosnay, 2003). Finally, urban-rural disparities in child development are also evident across a range of countries (Van de Poel, O'Donnell, & Van Doorslaer, 2007). Considering large differences in culture and context amongst the countries examined, we sought to better understand diversity in children's outcomes in the region and to explore if any of these correlates of early childhood development had a differential influence across countries. Based on previous literature, it was hypothesized that across all countries (i) older children would perform better on the Scales than younger children, (ii) that girls would on average, outperform boys, and (iii) that children living in urban areas would be developing better than those living in rural areas.

2. Methods

2.1. Participants

A total of 8,439 children from six countries (Cambodia, China, Mongolia, Papua New Guinea, Timor-Leste, and Vanuatu) participated in this study. With the exception of China, the sampling plan in each country was determined in conjunction with the National Census Department or National Statistics Institute. The sample in China was drawn from five provinces that were selected to represent a wide variety of levels of economic development (Guizhou, Heilongjiang, Jiangsu, Shanghai and Zhejiang). In China, an additional stage of sampling was used whereby preschools within communities were randomly sampled to identify children aged between 36 and 71 months who were representative of most children in the community. This procedure was used because a large majority (82% in 2014) of preschool-aged children were

enrolled in preschool in China (UN Development Programme, 2015). In all countries, the sample was stratified by age, gender, and urban/rural residence. Sampling of ethnic minority children (n=498) from all countries was insufficient to be included in analyses. Further, those with missing data in all domains of the Scales (n=26), those who were outside the age range of 36 to 71 months (n=144), those with missing data on urban or rural residence (n=13), and those with all domain scores outside 3 standard deviations of the mean for their age (n=1) were also excluded. As such, here we present data for 7,757 children (3,869 girls), ranging in age from 36 to 71 months (Table 1). With the exception of China wherein all participating children attended preschool, preschool participation rates varied across the remaining five countries. Mongolia, Cambodia, and Vanuatu had preschool participation rate of 50.2%, 41.8%, and 41.8%, respectively. Timor-Leste had a very low participation rate of 6.1% and that of Papua New Guinea was even lower (2.8%).

2.2. Measurement

The framework and items of the EAP-ECDS were developed based on the goals and values for children as defined by the countries themselves. The standards and indicators that were developed within each country were combined to develop a regional measure that represents the perspectives of each country. This approach avoided the application of inappropriate concepts that occurs when transferring a test developed in one culture to another. For example, the Scales include a domain focused on Cultural Knowledge and Participation which is seldom included in developmental scales of western societies. Notwithstanding the fact that religious holidays are celebrated in western countries, Asia has a variety of festivals and celebrations that may not be tied to religious events but to things like the harvest, the lunar calendar, and so on. In the EAP-ECDS children are asked to name festivals and culturally

relevant details of the celebration in terms of food, attire, music, history, customs and traditions. This method is also innovative in that it selected the most representative items across the nations in the region from a pool of items which were extracted from a well-recognized and homogeneous source from each country. This was done in order to guarantee the interpretative equivalence of the scale across countries. For example, measures of emergent literacy reflected differences in orthography. Given pan-cultural commonalities in goals for young children and our common biology and psychological needs, it is not surprising that similar competencies are valued across the region.

There were three phases in the development of the EAP-ECDS, the first of which being the initial development of the instrument. In Phase I, a comprehensive desktop review of the Early Learning and Development Standards (ELDS) of seven countries (Cambodia, Laos, Mongolia, Philippines, Thailand, Vanuatu, and Vietnam) in the region was conducted. ELDS reflect country-specific expectations for what children of different ages should know and be able to do (see for example Kagan, Castillo, Gomez & Gowani, 2015). The construction of the EAP-ECDS with a common set of items was underpinned and determined based on these standards. Seven developmental domains were identified based on an analysis of 1,738 indicators for 3-, 4-, and 5-year-olds from the ELDS. A 99-item scale was developed, with the number of items in each domain determined based on the number of indicators in each domain in the database. Thus, an attempt was made to reflect the importance given to each domain of development in a country's ELDS (Rao, Engle & Sun, 2011). Following these criteria, 99 most frequently mentioned indicators across the countries' ELDS were selected to form a 7-domain scale composed of 99 items: Cognitive Development (25 items, including items related to counting, addition/subtraction, short-term memory, concepts, behavioral inhibition, and knowledge of

shapes); Cultural Knowledge and Participation (10 items, including items related to nation, community and household knowledge, and acting according to moral codes); Language and Emergent Literacy (15 items, including items related to expressive language, grapheme knowledge, and pre-writing/drawing); Motor Development (10 items, including items related to walking, cup holding, ball throwing and catching, paper folding, and string beads); Health, Hygiene and Safety (10 items, including items related to personal hygiene, buttoning, knowledge of safety, knowledge of human body, and food hygiene behavior); Socio-Emotional Development (19 items, including items related to etiquette, social comprehension, emotional recognition, and perspective-taking); and Approaches to Learning (10 items, including items related to learning motivation, self-regulation, and persistence,). These selected ELDS indicators were converted into direct assessment items for use in the Scale with clear instructions and scoring schemes for testing children. The Scale was assumed to be applicable for children aged 36 to 71 months as it was constructed based on the ELDS for children of this age.

Phase II involved pilot testing of the Scales in three countries, and subsequent changes to the Scale based on the pilot experience. The 99-item scale was administered in the local language by trained assessors to 120 children aged 36 to 71 months, stratified by age, gender, and location, in China, Fiji and Mongolia. Test items were adapted for use in the three countries; as such there were three separate country Scales with slightly different items but equivalent test constructs. Based on results of systematic item analyses examining the discrimination and difficulty of each item as well as the appropriateness of item adaptation across each country, 85 items were selected and a revised testing protocol was developed as a result of the field administration experience.

This paper presents data from Phase III, in which the revised 85-item Scale was further validated with 8,439 children stratified by age, gender, and urban/rural residence in six countries in the region: Cambodia, China, Mongolia, Papua New Guinea, Timor-Leste and Vanuatu. A research team provided technical support to the six country teams at each stage of the validation process and ensured there was appropriate rigor in sampling, data collection and reporting. A test manual, model kit, and module for assessor training were developed, and specific guidelines were given on translation (including the use of back translation procedures), country adaptation, assessor selection and preparation, as well as sampling and data recording. Country teams attended training workshops and then adapted items to be country-specific (for example, different objects for counting, scripts, signs for danger, and pictures for sequencing, etc.), back-translated measures, conducted pilot assessments, took videotapes of children being assessed in the pilot study and sent them to the research team. This ensured standardized administration and scoring procedures were followed. Members of the research team also visited each country to carry out in-country training of the administration of the EAP-ECDS and evaluate the fidelity of the assessments. Particular attention was given to: training assessors to use standardized assessment processes; evaluating inter-assessor reliability; and minimizing bias and errors in the assessment process.

In Phase III, the EAP-ECDS were administered to children by individuals who had experience or training in early childhood education. Assessors were thoroughly familiar with the test materials and practiced administering and scoring the Scales under the supervision of an experienced assessor before using the test as a standardized measure. Prior to any test administration, the assessor administered the test and scored the child's performance parallel with the supervisor. Agreement (inter-observer reliability) between the assessor and the

supervisor was at least 85% prior to formal testing, and reliability between assessors or between the supervisor and the assessor was normally evaluated approximately every 20 test administrations.

The EAP-ECDS are untimed; total administration time was normally 45-60 minutes, though this was dependent upon a child's age, ability, temperament, mood, as well as rapport with the assessor. Items were administered in a fixed order in the local language.

In addition to the direct assessment of children's development through the EAP-ECDS, data were also collected from the caregiver of each child via a parent questionnaire. Caregivers in Cambodia, Mongolia, Papua New Guinea, Timor-Leste and Vanuatu were interviewed individually in order to complete the questionnaire. Due to higher levels of adult literacy in China, caregivers completed the questionnaire in small groups under the supervision of members of the research team who were there to assist if required. Information was collected regarding family background (parental education, occupation, and family wealth), children's participation in early childhood education programs, children's health and hygiene practices, as well as information regarding the home learning environment. Caregivers were also asked to rate their child's performance on the seven domains assessed by the EAP-ECDS. Direct assessment of children has an advantage over teacher and/or parent reports as observable behavior is evaluated, and often, direct assessment tests are used as a gold standard to assess the validity of adult reports (Bedford, 2014). The collection of both direct assessment and parent report data in this case however, has been utilized to further demonstrate the validity of the EAP-ECDS. Finally, children's height, weight, and body fat were measured in order to determine nutritional status; however these data are not presented in the current paper.

2.3. Measurement equivalence in cross-cultural research

Measurement equivalence – "whether or not, under different conditions of observing and studying phenomena, measurement operations yield measures of the same attribute" (Horn & McArdle, 1992) – a critical feature in the development and validation of a cross-cultural instrument, can only be achieved if the tool taps into the same construct and is administered with the same procedures across different cultures (Kankaras & Moors, 2010). Particularly important are interpretative equivalence – ensuring constructs of an instrument serve similar functions within cultural groups involved (Singh, 1995) – and procedural equivalence, or the degree of similarity of administration procedures across cultures.

Vandenberg and Lance (2000) further proposed three hierarchical levels of measurement equivalence: (i) configural equivalence which suggests the similarity of data structure across cultures but does not address the similarities required for direct comparisons of the results across groups; (ii) metric equivalence which assumes the equivalence of measurement units or intervals of the scale across cultural groups (Steenkamp & Baumgartner, 1998), as such valid comparisons of relationships of the latent variables and other variables of interest can only be made when satisfactory metric equivalence has been achieved; and (iii) scalar equivalence which demands that the scales of the same latent construct also have the same origin, so that full comparisons can be made with the measurement results (Meredith, 1993; Steenkamp & Baumgartner, 1998).

Few measures of early childhood development have undergone rigorous validation processes to ensure measurement equivalence has been achieved before they are administered in different cultural groups. This study, however, followed a stringent approach to develop and validate the EAP-ECDS to ensure satisfactory measurement equivalence (interpretative, procedure, configural, metric and scalar equivalence) was achieved, indicating that the tool is appropriate for use in program evaluation and national monitoring.

Item analysis has previously been conducted to ensure that assessment items differentiate between the ability of respondents, and that there was not any systematic bias affecting scores across countries (Rao et al., 2014). Analysis of item information curve (IIC) and test information function (TIF) graphs indicated the latent ability level at which assessments offered the most information. In the Cognitive Development, Socio-emotional Development, Language and Emergent Literacy, and Cultural Knowledge and Participation domains, information concentrated on mid to high levels of ability in all six countries. In the Health, Hygiene and Safety domain, information was concentrated on middle levels of ability, and in the Motor Development domain, information was concentrated on lower levels of ability in all six countries. However, there was more variation across countries in information concentration in the Approaches to Learning domain, although in most countries it was concentrated on middle levels of ability (Rao et al., 2014).

3. Results

3.1. Reliability

Data on the reliability of the EAP-ECDS have been published previously in the grey literature (Rao et al., 2014; Rao, Sun & Becher, 2015). Table 2 presents means for the domain scores in all countries for our sample. Alphas ranged from 0.78 to 0.94 and indicated a satisfactory level of reliability when the Scales were adapted for use in all six countries.

3.2. Validity

The Scales' content validity was assessed through soliciting the advice of experts in the region on earlier versions of the Scale. Suggestions were discussed by the team prior to the Phase III validation and items were modified or retained accordingly. A correlational analysis was performed to assess the consistency between children's performance and parents' assessment of

children's development. The results were significant, however the strength of the relationship between parents' ratings and EAP-ECDS total scores varied across countries (r = 0.46 in Cambodia, r = 0.26 in China, r = 0.46 in Mongolia, r = 0.23 in Papua New Guinea, r = 0.43 in Timor-Leste, and r = 0.30 in Vanuatu).

Exploratory factor analyses were conducted to examine the factor structure of EAP-ECDS scores across countries. Results show one main factor in each country, with eigenvalues of the first factor ranging from 2.27 in China to 3.53 in Vanuatu (Table 3). Confirmatory factor analysis was then conducted to test goodness of fit for a single factor model. Results showed either adequate fit (RMSEA test) or good fit (CFI, TLI, and SRMR tests) (Table 4).

As a developmental scale, we expected the EAP-ECDS to demonstrate differences in scores by age, gender, and from different areas of residence. Therefore, we conducted a series of country-specific 6 (age at 6-month intervals) × 2 (gender) × 2 (urban/rural residence) MANOVAs with children's scores in each of the seven domains as dependent variables. The MANOVA analyses indicated significant main effects of age and urban residence in all countries, with a significant interaction effect for age by urban residence in three countries: Cambodia, China and Mongolia. A gender main effect was found to be significant in four countries: Cambodia, China, Mongolia and Vanuatu, but not in Papua New Guinea or Timor-Leste. Gender was not found to interact with either age or residence location in any country, therefore gender interaction terms were not included in the results.

To explore the magnitude and direction of the age, residence, and gender effects on EAP-ECDS scores, we ran a series of country-specific linear regressions on the individual domain scores, as well as the composite score. As expected, older children performed consistently better

than younger children in all countries (see Figure 1 for the composite score and Figure 2 for domain scores by country).

Urban children generally performed better than rural children in Cambodia, China, Mongolia and Timor-Leste except in the Motor Development domain, although the differences on individual domains were mostly not significant at the 5% level (Figure 3). The better performance of rural children as compared to urban children in Papua New Guinea was only statistically significant for Motor Development. In Cambodia, where urban children performed better than rural children in five of the seven domains, we found that the gap between urban and rural children widened significantly with age in the domain of Language and Emergent Literacy, with a widening trend in the other domains (except for Motor Development) (see Figure 4a). A weak widening trend was also noted in Mongolia for Socio-Emotional Development (not shown), whereas in China, any urban-rural gap at age 36 months appears to diminish by age 71 months (Figure 4b).

In the four countries with a significant gender main effect, girls generally did better than boys in all domains except in Motor Development (see Figure 5). In contrast, boys did a little better on average than girls in Papua New Guinea, although the differences were only statistically significant for Cognitive and Motor Development. There were no statistically significant differences between genders in any of the domains in Timor-Leste.

4. Discussion

Great disparities in child development exist both within and across regions and countries within the East Asia-Pacific. Measurement is crucial to assess and monitor the status of children's development, and to guide effective intervention in order to address impairments identified. However, there is a paucity of culturally appropriate and sensitive tools to measure

early childhood development in LMICs. This paper describes one effort to address this shortcoming. The EAP-ECDS represents the first effort to create a child development measurement tool on the basis of the culture and values of a range of countries from within a world region, avoiding the transfer of contextually inappropriate concepts that often occurs when a measurement tool from one culture is adapted for use in another.

Results demonstrate that all seven Scale domains achieved satisfactory internal consistency across all six countries. Interestingly, the strength of the relationship between parents' ratings of child development and EAP-ECDS scores varied across countries. This could be due to variations in parental education and understanding of children's developmental milestones, or differing amounts of time spent between parents and children, and parents' subsequent knowledge of their child's capability and development. It could also be explained by varied levels of social desirability bias amongst parent responses due to cultural differences. Importantly, this highlights how differences within a region can impact the performance of an assessment tool.

Both the exploratory and confirmatory factor analyses demonstrated a one-factor structure of the EAP-ECDS. This is supported by the fact that different domains of early child development are interrelated with each other and a holistic approach is usually adopted in early childhood education.

The EAP-ECDS were able to distinguish between the development of children of different ages and gender. Age was the best predictor of child development in both country-combined and country-specific analyses, with older children consistently performing better than younger children in all countries as hypothesized. Children in China had the best development scores, with three-year-olds in China having similar levels of development to five-year-olds in

both Papua New Guinea and Timor-Leste (Figure 1). Further, increases in child development with age appear steeper in Cambodia, China and Mongolia relative to that in other countries. The results reflect the dramatic development in early years and slight differences across countries may be related to the contextual differences in each country.

Girls are generally found to have better developmental outcomes than boys especially in domains related to cognitive and language development (Matthews, Ponitz, & Morrison, 2009). Similarly, in our results, girls were found to be developing better than boys in Cambodia, China, Mongolia, and Vanuatu on all domains except for Motor Development, with the largest gender disparities in Vanuatu and Mongolia. In contrast, in Papua New Guinea boys had better Cognitive and Motor skills than girls, whereas in Timor-Leste no significant differences between boys and girls were found.

The Scales also highlighted developmental differences between children living in urban versus rural areas as hypothesized, though these disparities were not as pronounced as those found for gender. Interestingly, rural children had significantly higher Motor Development scores than urban children in Papua New Guinea; and in Cambodia, urban children were found to have significantly higher scores on domains of Cognitive Development, Cultural Knowledge and Participation, Language and Emergent Literacy, Health, Hygiene and Safety and Socio-Emotional Development, relative to children living in rural areas. Further in Cambodia, the disparity between rural and urban children widened as age increased due to the development of children living in urban areas increasing more with age than that of rural children. These findings could be the result of a wide range of phenomena, however, we speculate this is because parents of children in urban areas were better educated and the children had access to higher quality health and early education services.

There are several limitations to this study. First, data from the ethnic minority populations in different countries was insufficiently representative to be included the analyses. Second, although a bottom-up approach was used to determine items and domains of the EAP-CDS, the factor analyses identified only one composite factor that included items from several domains. Given the results of the factor analysis, caution should be exercised in making comparisons of children's performance across the different domains of the EAP-ECDS. Third, the EAP-ECDS are a child development tool developed for a region and competencies particularly important for an individual country may not be sufficiently represented.

Importantly, this research exemplifies great diversity in child development across the region, likely a reflection of a range of differences across the six countries including culture, parenting practices, social and economic policy, availability and quality of early education services, and so on. The EAP-ECDS are a psychometrically robust and contextually appropriate child assessment measure for East Asian countries. They can be used at a population level to help determine where there is inequality in children's development within and across countries in the region. The Scales and their data will be highly useful for resource planning and mobilization, as well as for national monitoring of child development. Together, these actions lay the foundation for improvement of educational and social policy with the goal of promoting equitable child development outcomes. Finally, the EAP-ECDS and their data make an important contribution to our gap in knowledge around child development in LMICs where the vast majority of the world's children live.

References

Bedford, H. (2014). *Measures of child development: A review*. London: UCL Institute of Child Health, Centre for Paediatric Epidemiology and Biostatistics.

Brinkman, S., Gregory, T., Harris, J., Hart, B., Blackmore, S., & Janus, M. (2013). Associations between the Early Development Instrument at Age 5, and Reading and Numeracy Skills at Ages 8, 10 and 12: a Prospective Linked Data Study. *Child Indicators Research*, 6(4), 695-708.

Barnett, W.S. (2011). Effectiveness of early educational intervention. *Science,* 333, 957-978.

Duncan, G.J., Dowsett, C.J., Claessens, A., Magnuson, K., Huston, A.C., Klebanov, P. et al (2007). School readiness and later achievement. *Developmental Psychology*, 43(6), 1428-1446.

Engle, P. L., Rao, N., & Petrovic, O. (2013). Situational analysis of young children in a changing world. In P. R. Britto, P. L. Engle, & C. M. Super (Eds). *Handbook of early childhood development research and its impact on global policy* (pp.35-64). New York: Oxford University Press.

Hambleton, R.K., Merenda, P.F., & Speilberger, C.D. (2005). Adapting educational and psychological tests for cross cultural assessment. Mahwah: Lawrence Erlbaum Associates Inc.

Heckman, J.J. (2008). Schools, Skills, and Synapses. *Economic Enquiry*, 46(3), 289-324.Henrich, J., Heine, S.J., & Norenzayan, A. (2010). Most people are not WERID. *Nature*, 466, pp 29.

Henrich, J., Heine, S.J., & Norenzayan, A. (2010). Most people are not WERID. *Nature*, 466, pp 29.

Horn, J. L., & McArdle, J. J. (1992). A Practical and Theoretical Guide to Measurement Invariance in Aging Research. *Experimental Aging Research*, 18, 117–144.

Janus, M., Harrison, L. J., Goldfeld, S., Guhn, M., & Brinkman, S. (2016). International research utilizing the Early Development Instrument (EDI) as a measure of early child development: Introduction to the Special Issue. *Early Childhood Research Quarterly*, *35*, 1-5.

Janus, M. & Offord, D.R. (2007). Development and psychometric properties of the Early Development Instrument (EDI): A measure of children's school readiness. Canadian Journal of Behavioural Science, 3, 1-22.

Kagan, S.L., Castillo, E., Gomez, R.E & Gowani, S. (2015). Understanding and Using Early Learning Standards for Young Children Globally. International Journal of Child Care and Education Policy, 7(2), 53-66.

Kankaras, M., & Moors, G. (2010). Researching measurement equivalence in cross-cultural studies. *Psihologija, 43*, 121-136. doi: 10.2298/PSI1002121K.

Lawn, J.E., Blencowe, H., Oza, S., You, D., Lee, A.C.C, Waiswa, P., Lalli, M., Bhutta, Z., Barros, A.J.D., Christian, P., Mathers, C., Cousens, S.M. (2014). Every Newborn: progress, priorities, and potential beyond survival. *The Lancet*, 384(9938), 189-205.

Louden, W., Chan, L., Elkins, J., & Greaves, D. (2000). Mapping the Territory - primary students with learning difficulties: literacy and numeracy. Canberra: Department of Education, Employment and Workplace Relations.

Matthews, J. S., Ponitz, C. C., & Morrison, F. J. (2009). Early gender differences in self-regulation and academic achievement. *Journal of Educational Psychology, 101*(3), 689-704. doi: 10.1037/a0014240

McCain, M., & Mustard, J. F. (1999). The Early Years Study-Reversing the Real Brain Drain. In The Canadian Institute for Advanced Research to the Ontario Government (Ed.), (Vol. 2013). Toronto, Ontario: The Canadian Institute for Advanced Research to the Ontario Government.Meredith, W. (1993). Measurement Invariance, Factor Analysis and Factorial Invariance, *Psychometrika, 58*, 525–543.

Mullis, I.V.S., Martin, M.O., Foy, P., & Arora, A. (2012). TIMSS International Results in Mathematics. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Mullis, I.V.S., Martin, M.O., Foy, P., & Drucker, K.T. (2012). The PIRLS 2011 International Results in Reading. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

OECD (2014). PISA 2012 Results in Focus: What 15-year-olds know and what they can do with what they know. Paris, OECD.

Pisani, L., Borisova, I., & Dowd, A.J. (2015). International Development and Early Learning Assessment (IDELA). Save the Children Technical Working Paper. Retrieved from http://www.savethechildren.org/atf/cf/%7B9def2ebe-10ae-432c-9bd0-df91d2eba74a%7D/IDELA%20TECHNICAL%20WORKING%20PAPER_V4.PDF

Inter-American Development Bank (2015) *Regional Project on Child Development Indicators (PRIDI)*

Rao, N. & Pearson, E. (2015). Assessment of child development across cultures. *Assessment & Development Matters*, 7 (3), 8-9.

Rao, N., Sun, J., & Becher, Y. (2015). Assessing early development and learning across cultures: The East Asia Pacific – Early Child Development Scales. *Assessment & Development Matters, 7,* 21-25.

Rao, N., Sun, J., Ng, M., Becher, Y., Lee, D., Ip, P., & Bacon-Shone, J. (2014). *Validation, Finalization and Adoption of the East Asia-Pacific Early Child Development Scales (EAP-ECDS).* UNICEF.http://www.arnec.net/wp-content/uploads/2015/07/EAP-ECDS-Final-Report1.pdf

Rao, N., Engle, P.L. & Sun, J. (2011). *Revised constructs and indicators for the East Asia-Pacific Early Child Development Scales*. Singapore: Asia Pacific Regional Network for Early Childhood.

Singh, J. (1995). Measurement Issues in Cross-Cultural Research. *Journal of International Business Studies, 26,* 597–619.

Steenkamp, J.E.M., & Baumgartner, H. (1998). Assessing Measurement Invariance in Cross-National Consumer Research. *Journal of Consumer Research, 25,* 78–90.

UN Development Programme (2015). Human development data (1990–2015). http://hdr.undp.org/en/data (accessed Sept 6, 2016).

UNICEF (2015a). *Progress for Children. Beyond Averages: Learning from the MDGs*. Report Number 11. ISBN: 978-92-806-4806-5.

UNICEF (2015b). The state of the world's children 2015: Reimagine the future: Innovation for every child. New York; UNICEF.

UNESCO (2014). EFA Global Monitoring Report. Teaching and Learning: Achieving quality for all 2013/2014. Paris: UNESCO.

UNESCO (2015). *EFA Global Monitoring Report. Education for All 2000-2015: Achievements and Challenges*. United Nations Educational, Scientific and Cultural Organization: Paris, UNESCO.

Vandenberg, R.J., & Lance, C.E. (2000). A Review and Synthesis of the Measurement Invariance Literature: Suggestions, Practices, and Recommendations for Organizational Research. *Organizational Research Methods, 2*, 4–69.

Verdisco, A., Cueto, S., Thompson, J., &, Neuschmidt, O. (2015). *PRIDI Regional Project of Child Development Indicators. Urgency and Possibility Results: First Initiative to Comparative Data on Child Development Latin America*. Inter-American Development Bank: Washington, DC. Retrieved from https://publications.iadb.org/bitstream/handle/11319/6849/PRIDI.%20Urgency%20and%20Possibility..pdf?sequence=1

Wachs, T. D., & Rahman, A. (2013). The nature and impact of risk and protective influences on children's development in low-income countries. In P. R. Britto, P. L. Engle, & C. M. Super (Eds). *Handbook of early childhood development research and its impact on global policy* (pp. 85-122). New York: Oxford University.

World Bank. (2016). *World Development Indicators.* Retrieved from *http://data.worldbank.org/data-catalog/world-development-indicators*.

World Bank. (2016b). Measuring child development and early learning. Concept note. Retrieved from *http://documents.worldbank.org/curated/en/659701473955877219/pdf/108286-BRI-PUBLIC-ELPIBMeasuringCDvprint.pdf*.

Table 1

Participant count and age range from each country by location and gender

Country	Location	Gender	N	Mean	SD	Min	Max
Cambodia	Rural	Female	273	53.4	10.4	36.0	71.8
		Male	277	53.5	10.1	36.1	71.9
	Urban	Female	325	54.1	10.4	36.6	71.9
		Male	323	54.1	10.5	36.0	71.9
China	Rural	Female	401	54.7	10.2	36.0	72.0
		Male	422	55.2	10.2	37.1	71.9
	Urban	Female	404	54.1	10.3	36.1	71.7
		Male	394	54.4	10.2	36.3	71.8
Mongolia	Rural	Female	301	53.2	10.1	36.1	72.0
		Male	306	53.3	9.7	36.0	71.5
	Urban	Female	313	54.3	9.9	36.0	71.9
		Male	312	54.0	10.0	36.3	71.7
Papua New Guinea	Rural	Female	584	53.3	10.1	36.0	70.0
		Male	582	53.2	9.9	36.0	70.2
	Urban	Female	308	52.1	10.1	36.0	71.3
		Male	288	52.7	10.2	36.0	70.2
Timor-Leste	Rural	Female	296	53.8	10.1	36.1	71.7
		Male	293	53.2	10.0	36.0	71.7
	Urban	Female	296	53.1	10.3	36.1	71.8
		Male	300	52.7	10.4	36.0	71.6
Vanuatu	Rural	female	322	55.2	9.4	36.0	71.8
		male	356	54.1	9.9	36.0	71.0
	Urban	female	46	56.6	8.3	38.6	70.6
		male	35	54.8	9.9	36.8	68.8
Total			7757	53.8	10.1	36.0	72.0

Table 2

Means and Standards Deviations of Total EAP-ECDS scores across urbanicity and gender in different countries

Country	Location	Gender	N	Mean	SD
Cambodia	Rural	Female	273	50.53	16.64
		Male	277	50.10	16.61
	Urban	Female	325	58.33	18.31
		Male	323	56.38	18.21
China	Rural	Female	401	66.44	16.34
		Male	422	65.31	16.80
	Urban	Female	404	69.02	15.28
		Male	394	67.44	15.38
Mongolia	Rural	Female	301	54.75	16.21
		Male	306	52.65	16.76
	Urban	Female	313	57.18	16.78
		Male	312	53.71	17.62
Papua New Guinea	Rural	Female	584	41.76	19.00
		Male	582	42.38	19.34
	Urban	Female	308	35.14	17.58
		Male	288	37.83	15.95
Timor -Leste	Rural	Female	296	35.60	13.97
		Male	293	35.48	14.50
	Urban	Female	296	37.33	14.26
		Male	300	36.38	14.00
Vanuatu	Rural	Female	322	51.58	19.94
		Male	356	47.75	19.32
	Urban	Female	46	51.95	20.17
		Male	35	43.64	22.20
Total			7757	50.53	20.34

Table 3

Results of exploratory factor analysis for each country individually

	Eigenvalue	Proportion	Cumulative		Eigenvalue	Proportion	Cumulative
Cambodia				Papua New Guinea			
Factor 1	3.05	1.07	1.07	Factor 1	3.42	1.07	1.07
Factor 2	0.26	0.09	1.16	Factor 2	0.15	0.05	1.11
Factor 3	0.03	0.01	1.17	Factor 3	0.05	0.02	1.13
Factor 4	-0.04	-0.01	1.16	Factor 4	0.00	0.00	1.13
Factor 5	-0.12	-0.04	1.12	Factor 5	-0.09	-0.03	1.10
Factor 6	-0.14	-0.05	1.07	Factor 6	-0.15	-0.05	1.05
Factor 7	-0.19	-0.07	1.00	Factor 7	-0.17	-0.05	1.00
China				Timor-Leste			
Factor 1	2.27	1.18	1.18	Factor 1	3.18	1.09	1.09
Factor 2	0.20	0.11	1.28	Factor 2	0.15	0.05	1.14
Factor 3	-0.01	-0.01	1.28	Factor 3	0.03	0.01	1.15
Factor 4	-0.10	-0.05	1.23	Factor 4	-0.02	-0.01	1.14
Factor 5	-0.12	-0.06	1.16	Factor 5	-0.06	-0.02	1.12
Factor 6	-0.14	-0.07	1.09	Factor 6	-0.15	-0.05	1.07
Factor 7	-0.17	-0.09	1.00	Factor 7	-0.20	-0.07	1.00
Mongolia				Vanuatu			
Factor 1	3.35	1.07	1.07	Factor 1	3.53	1.07	1.07
Factor 2	0.16	0.05	1.12	Factor 2	0.15	0.04	1.11
Factor 3	0.06	0.02	1.14	Factor 3	0.04	0.01	1.13
Factor 4	-0.04	-0.01	1.12	Factor 4	-0.03	-0.01	1.12
Factor 5	-0.05	-0.02	1.11	Factor 5	-0.07	-0.02	1.09
Factor 6	-0.15	-0.05	1.06	Factor 6	-0.13	-0.04	1.05
Factor 7	-0.18	-0.06	1.00	Factor 7	-0.18	-0.05	1.00

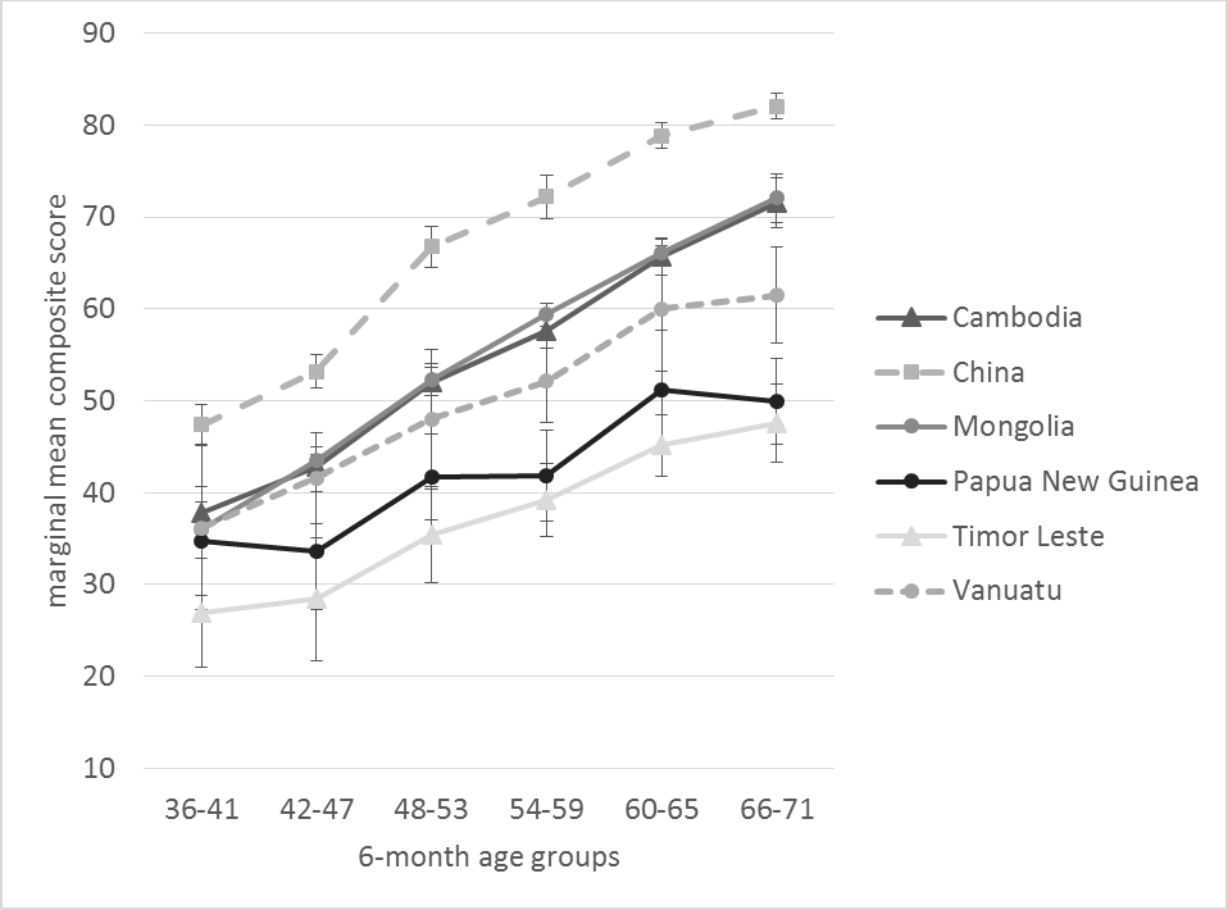
Table 4

Results of confirmatory factor analysis for a single factor model

	Result	Fit
Population error		
Root mean squared error of approximation	0.068	Adequate
Baseline comparison		
Comparative fit index	0.973	Good
Tucker-Lewis index	0.959	Good
Size of residuals		
Standardized root mean squared residual	0.024	Good

Figure 1

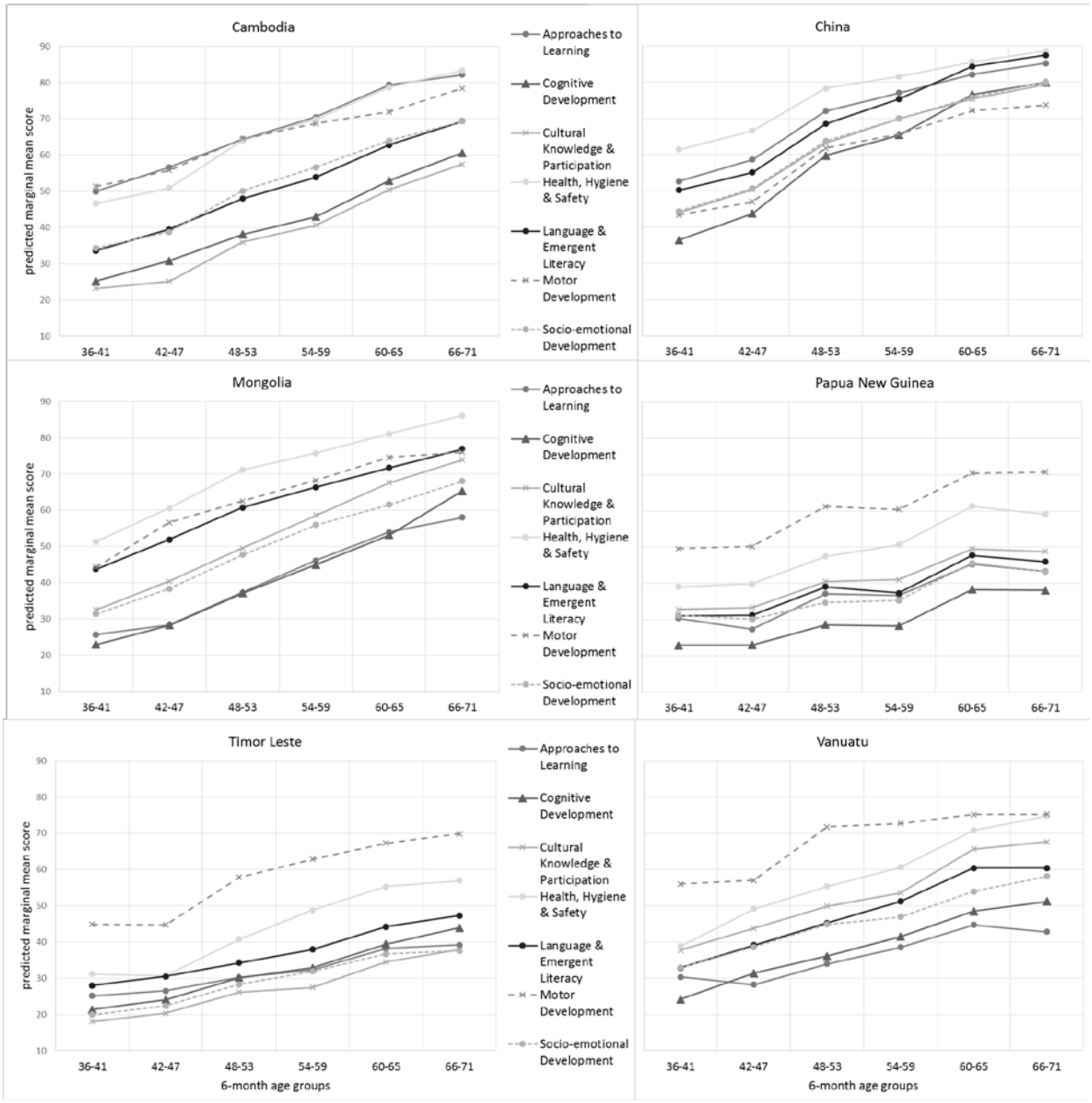
Country-specific age trajectories of marginal means[a] of composite score



[a] Marginal means by 6-month age category were obtained from a series of country-specific linear regressions of the composite EAP-ECD score controlling for urban-rural location, gender, and an interaction term for age category by urban-rural location. Standard errors of estimates are adjusted for clustering at the province level.

Figure 2

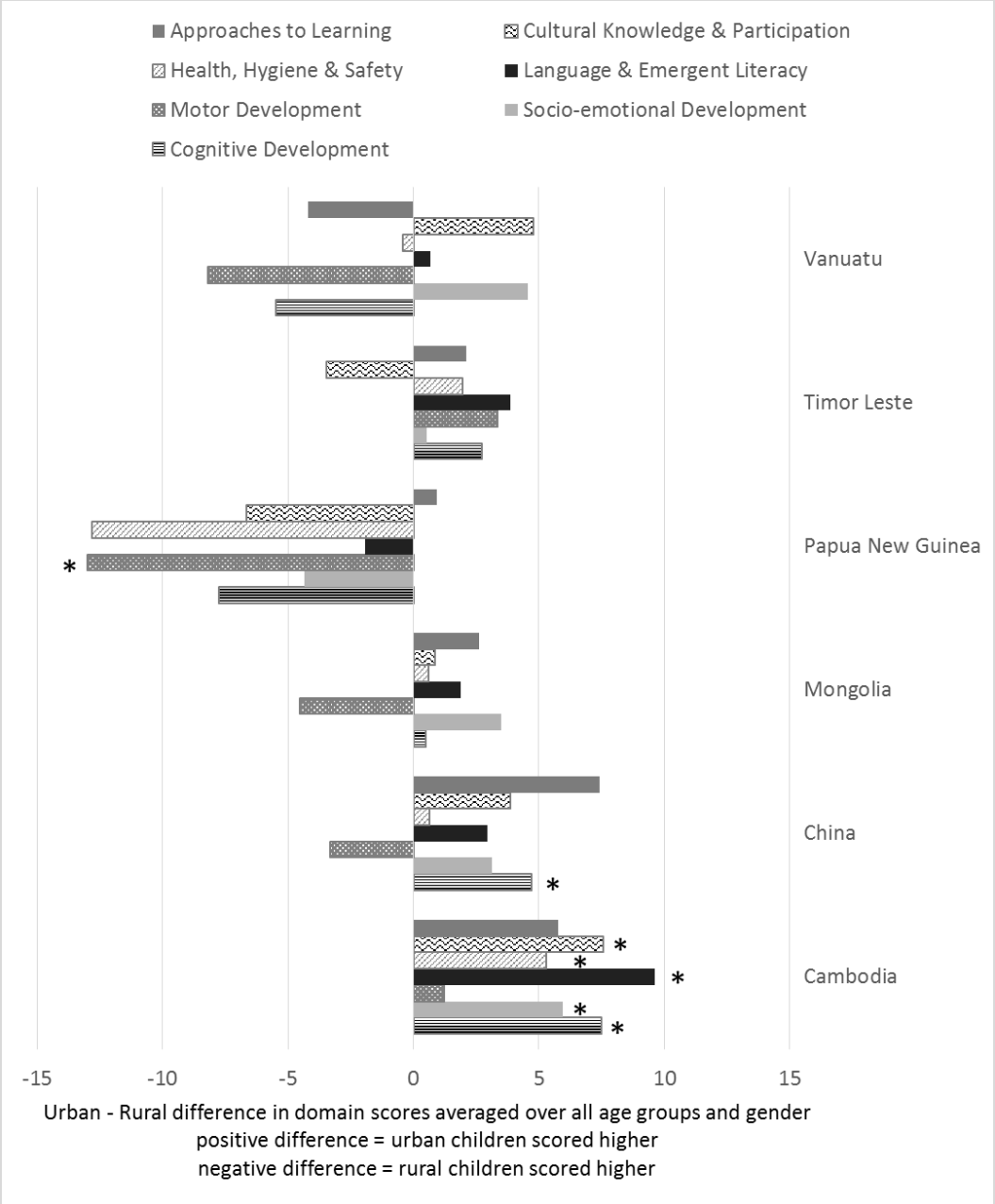
Country-specific age trajectories of marginal means[a] of domain scores



[a] Marginal means by 6-month age category were obtained from a series of country-specific linear regressions for each domain score controlling for urban-rural location, gender and an interaction term for age category by urban-rural location.

Figure 3

Estimated mean difference[a] in domain scores between urban and rural by country



[a] Score differences by urban-rural location are the estimated marginal effects obtained from linear regressions for each of the domain scores controlling for 6-month age category and gender. Standard errors of estimates are adjusted for clustering at the province level. An asterisk indicates that the estimated difference in scores by residence is statistically significant to the 5% level.

Figure 4a:

Age trajectories of the marginal means[a] of the domain-specific scores in Cambodia by urban-rural residence

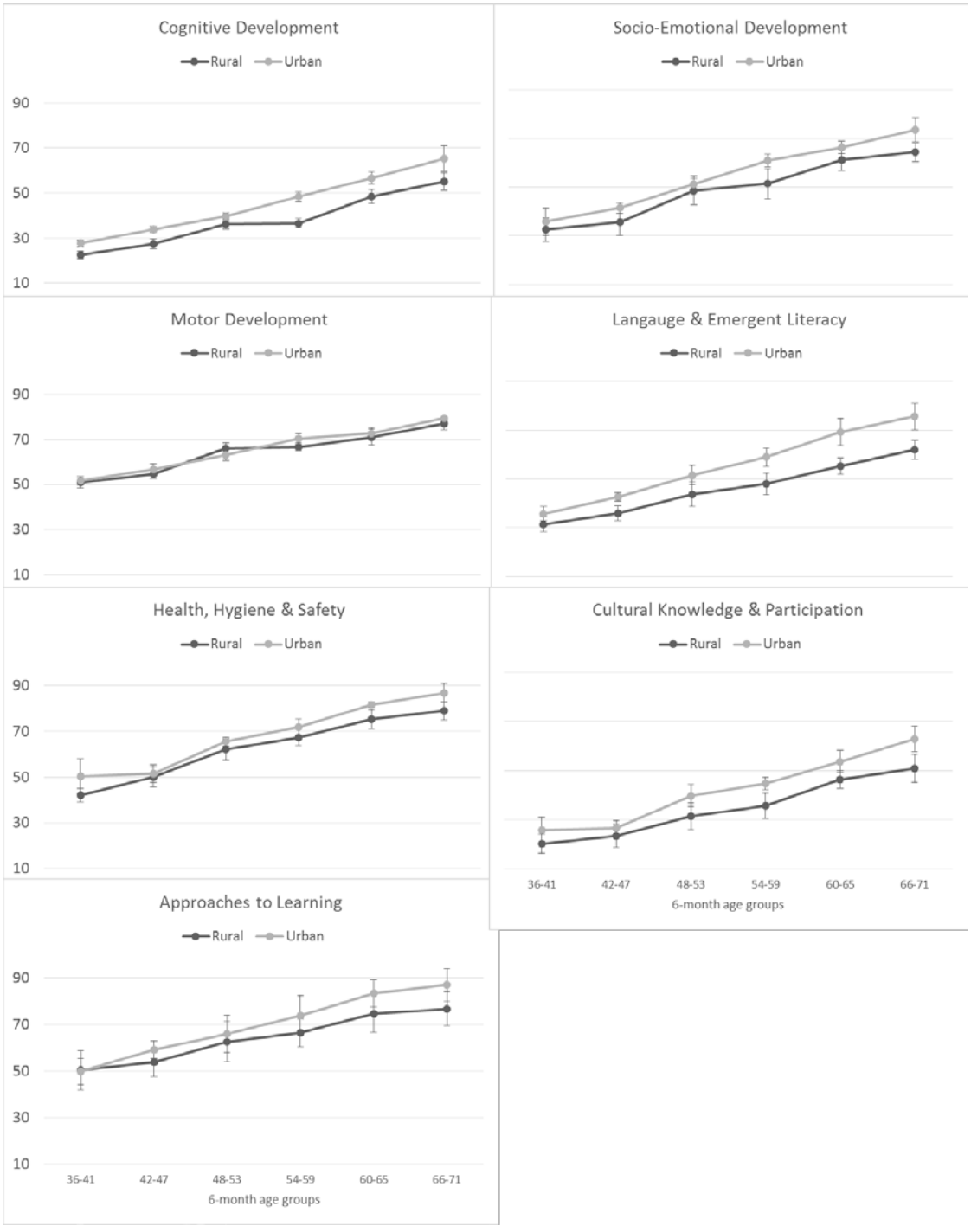
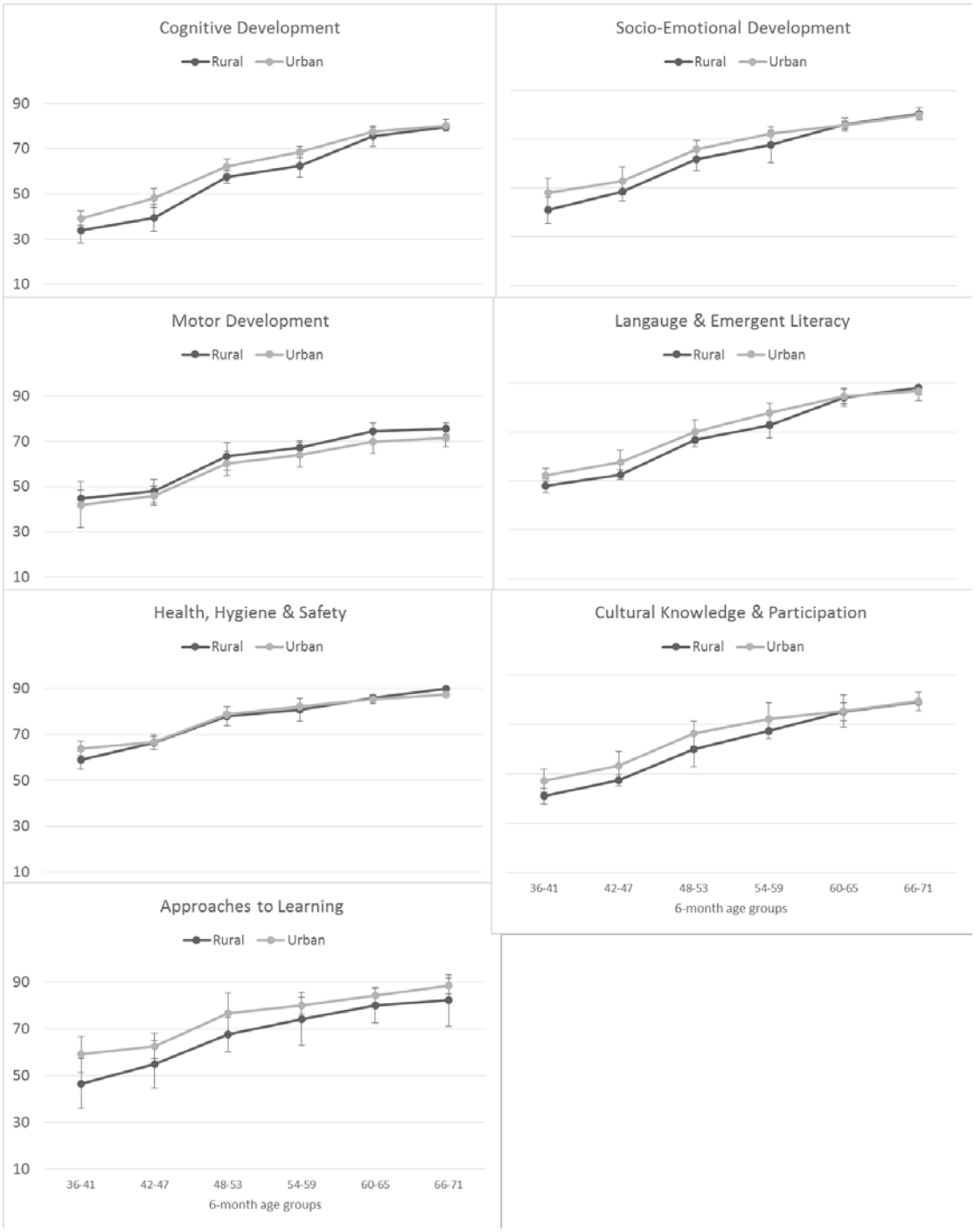


Figure 4b:

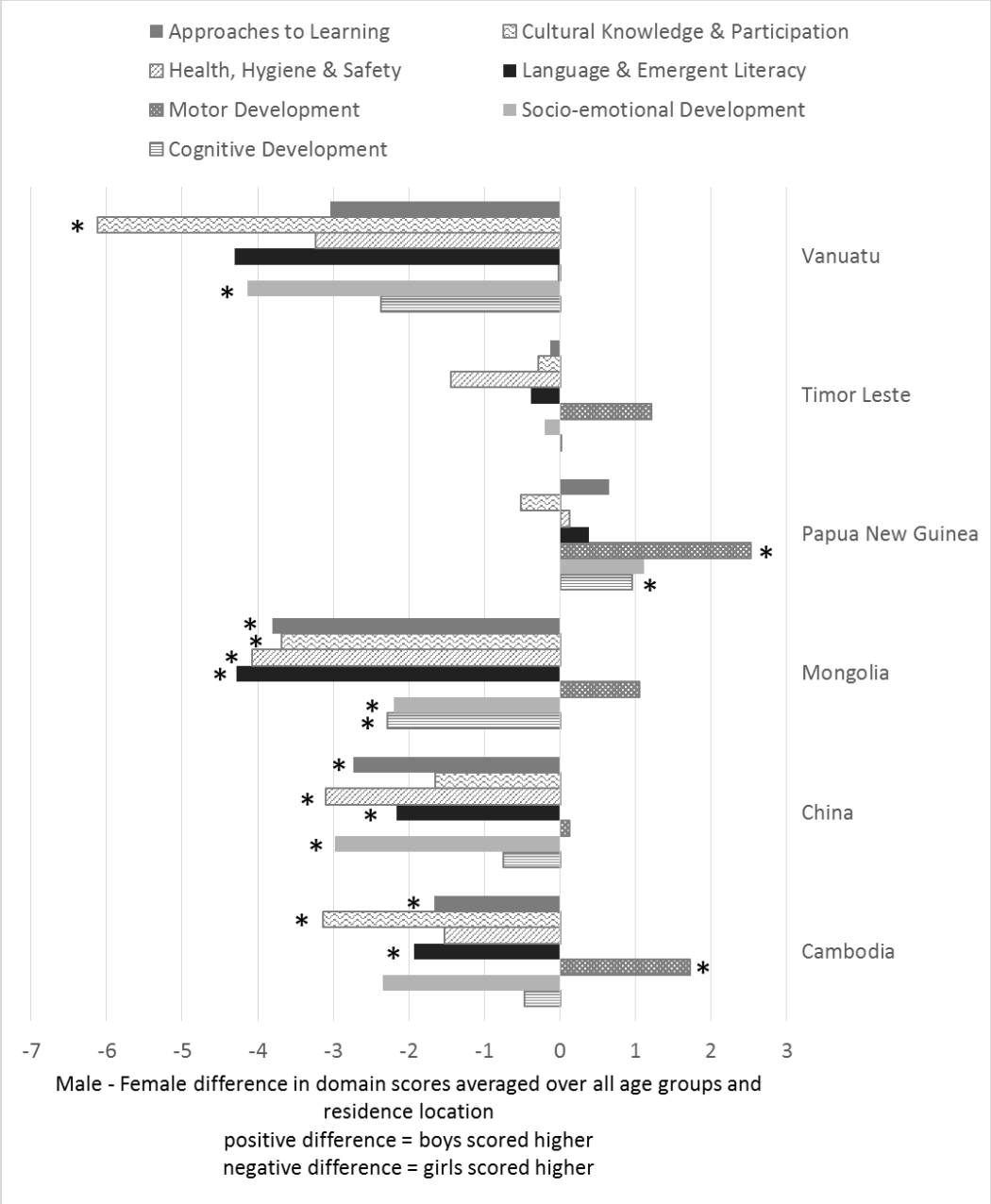
Age trajectories of the marginal means [a] of the domain-specific scores in China by urbanicity



[a] Marginal means by 6-month age category and residence location were obtained from a series of linear regressions for each domain score controlling for urban-rural location, gender and an interaction term for age category by urban-rural location.

Figure 5

Estimated mean difference[a] in domain scores between boys and girls by country



[a] Score differences by gender are the estimated marginal effects obtained from linear regressions for each of the domain scores controlling for 6-month age category and urban-rural residence. Standard errors of estimates are adjusted for clustering at the province level. An asterisk indicates that the estimated difference in scores by residence is statistically significant to the 5% level.