MEASURING QUALITY OF PRESCHOOL EDUCATION IN LOW- AND MIDDLE-INCOME COUNTRIES: ANALYSIS OF THE VALIDITY OF THE EARLY CHILDHOOD ENVIRONMENT RATING SCALE-REVISED IN COLOMBIA

By

Laura Betancur Cortes

B.S. in Psychology Universidad de Antioquia, Medellin, Colombia, 2011

M.A. in Psychology Universidad de Los Andes, Bogota, Colombia, 2013

Submitted to the Graduate Faculty of the
Kenneth P. Dietrich School of Arts and Sciences in partial fulfillment
of the requirements for the degree of
MS in Developmental Psychology

University of Pittsburgh

2016

UNIVERSITY OF PITTSBURGH
DIETRICH SCHOOL OF ARTS AND SCIENCES

This dissertation was presented

by

Laura Betancur Cortes

It was defended on

September 18, 2016

and approved by

Elizabeth Votruba-Drzal, Ph.D., Associate Professor Department of Psychology

Aidan G.C. Wright, Ph.D., Assistant Professor Department of Psychology

Shannon B. Wanless, PhD, Assistant Professor, School of Education

Dissertation Director: Elizabeth Votruba-Drzal, Ph.D., Associate Professor

MEASURING QUALITY OF PRESCHOOL EDUCATION IN LOW- AND MIDDLE-INCOME COUNTRIES: ANALYSIS OF THE VALIDITY OF THE EARLY CHILDHOOD ENVIRONMENT RATING SCALE-REVISED IN COLOMBIA

Laura Betancur Cortes, M.S.

University of Pittsburgh, 2016

In recent years, governments and international organizations have turned to preschool programs to address child poverty and related problems in low and middle income countries (LAMI). As preschool attendance has spread across the globe from 22% in 1999 to 45% in 2010, attention has shifted from preschool access to program quality, and researchers and policymakers have turned to the Early Childhood Environment Rating Scale-Revised (ECERS-R) to assess the quality of preschool experiences in their countries. However, important questions exist about the reliability and validity of this measure in LAMI countries. Colombia is among the countries that have expanded access to preschool programs in recent years, and have started to use the ECERS-R for assessing program quality. This study examines the reliability and validity of ECERS-R for preschool settings in Colombia, finding no evidence that the original scale structure is valid in the Colombian context. Additionally, it considered whether items on the ECERS-R could be used to generate a measure with stronger psychometric properties in Colombia. Evidence indicated that some items of the scale are not valid in the Colombian context but that others form three different factors related to routines, interactions and availability of materials for learning. Finally, analysis of predictive validity indicated that the *routines* factor significantly predicts children's gains in execute function and language skills over the kindergarten year. This evidence suggests that only some items of the ECERS-R consistently measure meaningful variability early education quality in Colombia, but that even low scores in routines are associated with gains in cognitive and social development.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1.0 INTRODUCTION

Over 550 million children under age five reside in low and middle-income (LAMI) countries, and 22% live in poverty (Grantham-McGregor et al., 2007; Unesco, 2015). These children are at risk for not reaching their full potential because they face environmental conditions that threaten their health, cognitive, and behavioral development (Grantham-McGregor et al., 2007; Engle et al., 2007). In recent years, governments and international organizations have turned to preschool programs to address child poverty and related problems (Engle et al., 2007; Myers, 1992, 2005; Ramey & Ramey, 1998; Unesco, 2015). In fact, due to the drastic expansion of government investment in preschool over the past decade, the proportion of poor children attending preschool in LAMI countries increased from 22% in 1999 to 45% in 2010 (Unesco, 2012). As preschool programs have spread across the globe, attention has shifted from preschool access to program quality, and researchers and policymakers have turned to measures of child care quality developed in the U.S. (such as the Early Childhood Environment Rating Scale-Revised, ECERS-R; Harms, Clifford, & Cryer, 1998) to assess the quality of preschool experiences in their countries (Wortham, 2013; Yoshikawa & Kabay, 2014). However, important questions exist about the reliability and validity of these measures in LAMI countries.

Colombia is among the countries that have expanded access to preschool programs in recent years. Nearly half of the 4.8 million children under age five in Colombia live in poverty (Rubio, Pinzón & Gutiérrez, 2010). Increasingly, these children are attending preschool before they enter primary school. Participation rates in early childhood education rose from 44% in 2010 (Rubio, Pinzon & Gutiérrez, 2010) to 60% in 2013 (Bernal & Quintero, 2014). Seeking to understand the effects of investment in childcare in Colombia, researchers and governmental institutions have started to use the ECERS-R for assessing program quality and its relation to child development (Bernal et al., 2009; Bernal et al., 2011; Maldonado-Carreño

& Votruba-Drzal, 2015). This study examines the reliability and validity of ECERS-R for preschool settings in Colombia.

Three specific aims are addressed in this study. First, we test whether the original structure of the ECERS-R is reproduced in Colombian preschool settings. Second, using the original set of items, we examine the number of factors that the scale measures in the Colombian sample. Third, we explore whether items on the ECERS-R can be used to generate a measure of child care quality that is related to child development. This study intends to provide information for measuring classroom quality in the context of a middle-income country.

1.1 USING THE ECERS-R TO ASSESS EARLY EDUCATION QUALITY

The most commonly used instrument for assessing preschool quality is the ECERS-R. The ECERS-R consists of 43 items that measure quality across seven dimensions: 1. space and furnishing, 2. personal care routines, 3. language and reasoning, 4. activities, 5. interaction, 6. program structure and 7. parents and staff. Scores for these items are based on odd numbers (from 1 to 7) as follows: 1 - inadequate, 3 - minimum, 5 – good, and 7 - excellent. The odd scores are assigned only if all the indicators for that score are meet, and the even and lower score is assigned if only half of them are met. Total score and individual sub-scale scores are generated by averaging across items (Ishimine & Wilson, 2009; Ishimine, Wilson & Evans, 2010).

Evidence has shown that the ECERS-R has good psychometric properties in the US and other high-income countries (Matherset al., 2007; Rentzou, 2010; Tietze, Cryer, Bairrão, Palacios, & Wetzel, 1996). However, recent evidence from the US indicates that the structure of the ECERS-R is not consistent. Although the scale is designed to measure seven different areas of quality, there are inconsistencies in the literature about how many distinct dimensions of quality the ECERS-R captures. Several studies find no evidence for distinct domains, but instead uncover a single quality factor (Beller, Stahnke, Butz, Stahl, & Wessels, 1996; Holloway, Kagan, Fuller, Tsou, & Carroll, 2001; Munton, Rowland, Mooney, & Lera, 1997; Phillipsen et al., 1997; Scarr et al., 1994). Other studies identify two factors:

Caregiving and Activities (Howes et al., 1992; Sakai et al., 2003), and still others find evidence of three factors: activities and language, interactions, and routines (Cassidy et al., 2005; Clifford et al., 2005; Early et al., 2006; Frede et al., 2007; Gordon, 2013; Perlman, Zellman, & Le, 2004; Sakai et al., 2003). The fact that so few studies validate the seven sub-scales, and some uncover only a single quality factor, suggests that the instrument may obscure specific features about quality that were collected during the observations (Gordon et al., 2013).

Beyond these concerns related to the reliability and validity of the measure, researchers have expressed major reservations about using the ECERS-R in LAMI countries. Serious concerns have been expressed that not all items on the ECERS-R are relevant for preschool settings in LAMI countries. Indeed, items on the ECERS-R are often dropped due to extreme low-scores. In fact, settings in LAMI countries tend to have very low aggregate scores, with average quality scores in the range of "inadequate" (scores of 1 or 2), and the highest scores only reaching the threshold for "acceptable" (scores of 3 or 4) (Aboud, 2006; Bernal et al., 2009; Campos et al. 2010). These low scores may be an indication of poor preschool quality in LAMI relative to early education settings in the U.S. However, these scores may also reflect cultural differences in perspectives about early education. The ECERS-R places great emphasis on child-selected activities and materials, which reflects U.S. values that are not central to collectivistic and underprivileged cultures. For example, items that are commonly dropped include child-centered activities, which are not as valued in collectivistic contexts, as well as provision for special needs, access to soft toys or cozy areas, acceptance of diversity, space for privacy, time for free play, use of video and computer, and provisions for taking naps (Mathiesen et al., 2011; Treviño et al., 2012; Villalon et al., 2002; Aboud, 2006; Li et al., 2014). Consequently, researchers from China (Li et al., 2014), India (Isley, 2000) and Cambodia (Rao & Pearson, 2007) have modified the scale so that it emphasizes whole-group teaching instead of the individualized approach of the ECERS-R.

Additionally, the heavy emphasis on access to materials and spaces in the ECERS-R place preschool settings in LAMI countries at a disadvantage because access to materials is often limited. Perhaps not surprisingly, some studies have showed that sub-scales related to interactions show more variability and higher scores than those related to materials (Aboud, 2006; Campos et al., 2010). The low scores obtained on the ECERS-R indicates that the scale

may fail to capture meaningful caregiving differences within collectivistic and underprivileged contexts.

Regarding the predictive validity of the scale, research in the U.S. has shown that higher scores on the ECERS-R are associated with better early cognitive development in children (Byrant, Burchinal, Lau, Sparling, 1994; Peisner-Feinberg & Burchinal, 1997; Peisner-Feinberg et al. 2001, Mashburn et al., 2008; Ishimine & Wilson, 2009) and long-term gains measured as late as second grade (Peisner-Feinberg et al., 1999). Indeed, the ECERS-R has greatly impacted preschool education in general as it has been used for research, improving teachers' practices, and even evaluating public policy (Fenech, 2011). However, the predictive qualities of the ECERS-R in LAMI countries have not been as thoroughly investigated. In fact, few studies explicitly link the measure to child's outcomes (Campos et al. 2010; Mathiesen et al., 2011). There are notable exceptions. In particular, studies in India (Rao, 2010), China (Li et al., 2014), and Cambodia (Rao & Pearson, 2007) find that children in high-quality preschools had significantly better emotional, linguistic and math outcomes than children in lower-quality settings. However, in these studies, the ECERS-R scale was transformed to meet some characteristics of each country. Evidence from Tanzania, Uganda, and Bangladesh show that even low scores on the ECERS-R were linked to improvements in children's cognitive skills (Malmberg, Mwaura, & Sylva, 2011; Moore, Akhter & Aboud, 2008). Thus, even low-quality (based on the ECERS-R) early education settings seem to benefit children in LAMI countries. Indeed, these settings may be a major improvement over other caregiving arrangements.

A review of the current literature about the use of the ECERS-R in LAMI countries indicates that: 1) its use has been increasing in recent years; 2) there is mixed evidence about the number of factors that the scale measures; 3) items are consistently dropped due to low variability across settings; 4) its high emphasis on child-selected activities and materials contrasts with the teacher-directed and limited resources of LAMI countries; and 5) evidence about its predictive validity is limited but some is promising. Thus, although psychometric proprieties of the ECERS-R are questioned, some evidence suggests that the ECERS-R may perform reasonably well when assessing quality in the context of LAMI countries. Examining the reliability and validity of the ECERS-R is important because the instrument has widespread use when it comes to making international comparisons. Moreover, use of the

ECERS-R allows researchers to participate in the international dialogue about preschool program quality (Pena, 2007).

1.2 THE PRESENT STUDY

The goal of this study is to strengthen knowledge about whether the ECERS-R has good psychometric qualities in LAMI countries by considering how it performs in two studies in Colombia. Specifically, this study addressed three aims. First, it tested whether the original structure of the ECERS-R was reproduced in Colombian preschool settings. Second, using the original set of items it examined the number of factors that the scale measures in the Colombian sample. Third, it considered whether items on the ECERS-R can be used to generate a measure of child care quality with stronger psychometric qualities in the Colombian context. In doing so, this study contributes to the growing literature aimed at strengthening understanding of how well the ECERS-R measures classroom quality in the context of a LAMI country.

2.0 METHODS

2.1 SAMPLE

The present study uses data from two studies conducted in Colombia that used the ECERS-R to assess preschool classroom quality. The first is the Early Education Quality and Child Development in Bogotá (EEQCDB) study which was conducted in Bogotá, the capital of Colombia. The second is the Impact Assessment of Child Development Centers (IACDC) study which was collected in several cities across the country. In tandem, these samples provide information for the use of the ECERS-R in different areas of the country and include three different types of child care: private, public and co-funded large centers. Descriptive statistics of the child outcomes and family characteristics by subsample are presented in Table 1. Interesting, although both samples are composed by low-income families, families in the EEQCDB earn a higher income while the families in the IACDC sample are more educated.

2.1.1 Early Education Quality and Child Development in Bogotá (Bogotá sample; EEQCDB)

The EEQCDB sample includes 233 children selected from 61 classrooms across early childhood care centers in Bogotá, Colombia (Maldonado-Carreño & Votruba-Drzal, 2015). The sample is comprised of 26 classrooms in private centers and 35 co-funded centers. Private settings are for-profit centers, paid for by parents and managed by independent members of the community and do not receive public funds. Co-funded centers are co-financed by the Bogotá government and by civil society organizations. Both types of centers provided services to low- and middle-class children in Bogotá.

Table 1: Descriptive statistics by sub-sample

	EEQCDB (Bogota) N=226		IACDC (National) N=458
	M(SD) or %		M(SD) or %
Male	54.87%		53.93%
Age (months)	49.70 (4.53)		50.78 (8.68)
Parent education			
Below HS	16.28%		39.52%
High school	41.40%		42.36%
College	42.33%		18.12%
Income	1.20 (0.92)		1.98 (0.75)
Cognitive development			
Eff. Control T1	-0.11 (2.44)	ASQ Cognitive T1	132.96 (86.04)
Eff. Control T2	-0.03 (2.55)	ASQ Cognitive T2	196.82 (46.91)
Ex. Function T1	0.05 (1.40)	Verbal	80.27 (12.36)
Ex. Function T2	-0.01 (1.55)	Associative memory	87.59 (16.00)
Math T1	6.29 (2.71)	Executive function	92.15 (11.69)
Math T2	7.57 (3.37)	Numeric reasoning	84.60 (12.40)
LW identif. T1	2.59 (2.12)	Receptive language	92.46 (12.53)
LW identif. T2	3.22 (2.98)		
Picture vocabulary T1	15.96 (3.57)		
Picture vocabulary T2	18.01 (3.10)		
Sound awareness T1	1.07 (1.39)		
Sound awareness T2	1.93 (1.90)		
Socioemotional development			
Emoc. Control T1	2.38 (0.59)	ASQ Socioemot T1	56.34 (24.76)
Emoc. Control T2	2.34 (0.53)	ASQ Socioemot T2	65.36 (33.30)
Positive emotion T1	2.07 (0.61)		
Positive emotion T2	2.23 (0.59)		
Social skills (parent) T1	103.28 (16.43)		
Social skills (parent) T2	105.00 (16.82)		
Behav probl (parent) T1	25.45 (12.13)		
Behav probl (parent) T2	23.21 (10.71)		

Children in the participating classrooms were about 45-months old on average. Children's development was assessed two times at a six-month interval. As the school year in Colombia starts in January and ends in December, data were collected at the beginning and middle of the pre-kindergarten year. The first measurement occurred between February and April of 2012 and second measurement occurred between July and August of 2013. A group of trained psychologists conducted the assessments in the centers. Researchers obtained parents' permission for children to participate in the study, and parents completed questionnaires that collected data on household demographic characteristics and children's socio-emotional outcomes. Among the 233 children in the sample, 58% had complete data

on all variables included in the analyses. Of cases without complete data, 30.47% had missing data on one variable, and the greatest number of missing variables for any one participant was 4 (5.15% of participants). The percentage of missing data for each variable in the analyses ranged from 0.4% to 16.30%.

2.1.2 Impact Assessment of Child Development Centers (national sample; IACDC)

A total of 445 children were drawn from 113 classrooms across Colombia (Bernal & Quintero, 2014). Data were collected as part of the assessment of a national strategy that moved children from small community nurseries to large child care centers in Colombia. The transition was implemented as part of a government plan for providing children with better childcare services by transferring children from home-based childcare provided by a mother from the community to an institutional program with superior infrastructure and better-trained personnel. The present sample includes children from 14 cities in Colombia who were cared for in15 new centers. On average, children were about 44 months of age. Child development was assessed first when children were still in home-care and several months after the transition to center-based care. Baseline data collection occurred in all locations between the end of November 2010 and beginning of May 2011 (the end and beginning of the academic year). Follow-up data collection took place in two different stages. The first stage took place in November and December 2011 and the second stage took place from September to November 2012. The two-stage process was due to financial constraints of the funding agency; therefore, the number of months between the baseline and second assessment varies across the centers. In six of the centers (142 children) the follow up was 9-10 months later, and in the remaining nine centers (303 children) the follow up was made from 18 to 19 months later. Child outcome data were collected by psychologists or social workers who were trained and assessed for reliability by the evaluation team. At both times, children were assessed in a small area of the center. Parental permission to participate in the study and parent questionnaires on household demographic characteristics and children's socio-emotional outcomes were collected in the center or in the home when it was not possible in the center. Data on the child's attendance was collected directly from school records available

to the evaluation team that was cross-referenced with parental questionnaires. Among the 458 children in the sample, 70% had complete data on all variables included in the analyses. Of cases without complete data, 17.5% had missing data on one variable, and the greatest number of missing variables for any one participant was 3 (4.80% of participants). The percentage of missing data for each variable in the analyses ranged from 1.09% to 19.43%.

2.2 MEASURES

2.2.1 Child care quality

In both studies, child care quality was measured using the Early Childhood Environmental Rating Scale - Revised (ECERS-R; Harms, Clifford & Cryer, 1998), which has been described previously. The ECERS-R provides a global measure of preschool classroom quality by scoring seven sub-scales, based on 43 items. The ECERS-R has shown predictive validity across cognitive domains (Peisner-Feinberg et al., 2001; Burchinal, Roberts, Riggin, Zeisel, Neebeand Bryant, 2000) and social-emotional domains (Sylvia et al., 2006).

2.2.2 Child outcomes

2.2.2.1 Bogotá sample; EEQCDB For this sample, three sub-tests of the Woodcock-Muñoz III Tests of Achievement (WM-III) were used for assessing language skills (Muñoz-Sandoval, Woodcock, McGrew, Mather & Schrank, 2005). The scale has been translated into Spanish, adapted for Latin American contexts and used to evaluate effects of early childhood interventions on cognitive development in infants and older children (Fernald et al. 2009). The sub-tests used have showed good psychometric qualities in Latin-American contexts (Schrank, et al., 2011): 1. Letter-word identification ($\alpha = .91$), which assesses the child's ability to identify letters and isolated words; 2. Picture-vocabulary sub-scale ($\alpha = .77$), which

assesses the ability to recognize and name drawn objects; and 3. Sound awareness (α = .81), which measures phonological knowledge.

Children's mathematical abilities were assessed through the 15-item Math Assessment (Klibanoff et al., 2006). This non-standardized assessment is similar to the Test of Early Mathematics Ability (TEMA-2; Ginsburg & Baroody, 1990). However, the multiple-choice format facilitates shorter administration time (approximately 10 minutes) than the TEMA-2. This measure consists of 15 multiple-choice items and evaluates skills such as understanding the concept of ordinal and cardinal numbers, performing calculations, nomination of figures, understanding the concept of "half" and recognition of numerical symbols. Each item is scored as right or wrong, and a total sum score of all correct scores is calculated (α =.85).

To measure executive function and effortful control skills, the Preschool Self-Assessment Regulation (PSRA; Smith-Donald, Raver, Hayes, Richardson, 2007) was used. The PSRA assesses self-regulation in the emotional (managing excitement, frustration, distress), attentional (focusing and shifting attention, executive control), and behavioral (impulsivity, ability to wait) domains using a brief, structured battery of seven tasks. The PSRA tests executive function using three tasks that require children to filter competing stimuli: Pencil Tap, Balance Beam, and Tower Task Turn Taking (α =.87). To assess effortful control, four delay tasks were used: Toy Wrap, Toy Wait, Snack Delay, and Tongue Task (α =.85).

Finally, socio-emotional development was measured using the Social Skills Improvement System, SSIS (Gresham & Elliott, 2008). For the SSIS assessment, parents and teachers rate children's socio-emotional and behavioral development. The SSIS assesses children's social skills, behavior problems, and academic competence (Gresham & Elliott, 2008). The instrument evaluates whether students possess age-appropriate social skills and identifies any behavioral problems that may interfere with the acquisition of social skills. It consists of a questionnaire answered by the child's parent and teacher, who assigns a score of 0-3 based on the frequency with which a situation occurs: never (0), rarely (1) often (2) almost always (3). Although there are different forms for parents and teachers, both consist of about 140 items and take approximately 25 minutes to complete. The items reflect two main constructs: social skills (α =.93) and behavior problems (α = .74).

2.2.2.2 National sample; IACDC To test cognitive development, five sub-tests of the Woodcock-Muñoz III Tests of Achievement (WM-III; Muñoz-Sandoval et al., 2005) were administered to children at the follow-up assessment. General verbal ability (i.e., higher-order, language-based acquired knowledge and the ability to communicate that knowledge) was measured by the verbal comprehension subscale (α =.90). Receptive language, which reflects the ability to attend to the sound structure of language by analyzing and synthesizing speech sounds, was measured by the concept formation subscale (α =.94). Associative memory (i.e., the ability to store and retrieve associations) was assessed by the delayed recall subscale (α =.92). The auditory attention subscale (α =.88) was used to assess executive function skills, including response inhibition, cognitive flexibility, and planning. Finally, the number reversal subscale (α =.87) measured mathematical reasoning skills. Specifically, this sub-test measures the ability to reason about mathematical relationships and number properties.

Children's general development was assessed at baseline and follow-up by the Ages & Stages Questionnaire (ASQ)-Third edition. The ASQ-3 evaluates children's on-time achievement of key developmental milestones, is completed by parents, and assesses children from birth through age 6. The ASQ-3 focuses on cognitive development and the identification of children at risk for developmental delays. It assesses development across several domains, including fine motor, gross motor, communication and problem solving skills (Squires and Bricker, 2009). The ASQ-3 has been used for early developmental screenings in low- and middle-income countries (α = .80) (Bernal, 2015; Rubio-Codina, Araujo, Attanasio, Muñoz and Grantham- McGregor, 2015). Parents completed the ASQ-3 questionnaires during interviews with researchers.

To assess children's socio-emotional development in the IACDC sample, researchers used the Ages and Stages Questionnaires for the Socio-Emotional domain (ASQ: SE; Squires, Bricker & Twombly, 2009). The ASQ: SE is a parent-report measure for children ages 6 to 60 months. The questionnaires focus on socio-emotional development and the identification of children at risk for social-emotional difficulties. It evaluates functioning across multiple domains: self-regulation, compliance, communication, adaptive functioning, autonomy, affect, and interactions with others. Like the ASQ-3, the ASQ: SE has been used

for early childhood developmental assessments in low- and middle-low income countries (α =.84) (Handal, Lozoff, Breilhand & Harlow, 2007; Heo, Squires & Yovanoff, 2007). To reduce the impact of disparities in literacy level, the ASQ: SE data was collected through parent interviews.

2.2.3 Household demographic information

For both samples, demographic questionnaires were administered to parents at baseline and follow-up. The questionnaires were sent home to parents to complete. Information about income and parental education at follow-up were included in the analyses. Parental education was coded as a three level categorical variable indicating if the highest level of education of any of the parents was below high school, high school degree, or college degree. Income was coded as the number of monthly minimum wage units per household during the year of data collection. In other words, the total income reported by each household was assessed by diving parent's monthly income by the minimum monthly wage for the year of data collection as defined by the Colombian law. This transformation helps adjust for the large fluctuations in inflation that the Colombian economy faces every year and ensures that the income data collected in different years are equivalent measures of economic resources.

2.3 STATISTICAL ANALYSIS

The first aim of the study was to test the structural validity of the measure in the Colombian context. The structural validity of the ECERS-R was tested using confirmatory factor analysis (CFA) on a data set that merged the two samples. Initially, a seven-factor model was estimated to determine if there is evidence of the original sub-scales. The factors were allowed to correlate freely with an oblique rotation.

The second goal of the study was to examine the number of factors that the scale measures in the sample. For this, an exploratory factor analysis was estimated to determine

the best number of factors (between one and seven) that fits the data. Several indices were used for assessing model fit: Root mean square error of approximation (RMSEA) of 0.06 or lower, and a comparative fit index (CFI) and the Tucker-Lewis index (TLI) of at least .90 (Hu & Bentler, 1999). After finding a number of factors that fits the data well, the predictive validity of the measure was tested using structural equation modeling to fit path models from the ECERS-R quality factors to developmental outcomes in the participating children.

Finally, the third aim of the study was to explore whether some items on the ECERS-R can be used to construct a measure with stronger psychometric properties than the original measure. Based on existing research, it was anticipated that the seven original factors would not be supported and that some items may not load onto any factor due to limited variability. In effort to create a stronger measure, ECERS items were dropped based on low variability and lack of cultural importance according to a questionnaire completed by Colombian professionals (see Appendix A). After deciding which items should be dropped, an exploratory factor analysis (EFA) was estimated on the item covariance matrix. The EFA examines how many distinct aspects of quality are being measured by the ECERS-R in the Colombian context. Several fit indices were used to assess fit to the data: a non-significant chi-square, RMSEA, CFI, and TLI.

Unfortunately, in these data none of the models examined in the EFA was a good fit for the data. A close examination of the data showed that this sample frequently obtained low scores in the ECERS-R, which resulted in limited item-level variability and non-normal distributions in the data. Additional. Low correlations among indicators were also found in the data. This may give rise to poor model fit since SEM factor analysis requires normal distributions and homogeneity of variance. Additionally, it is not appropriate running a factor analysis when there are few correlations above 0.3 between the indicators (Tabachnick & Fidell, 2001). The lack of correlations indicates that the indicators do not share enough variance or that they are not closely related to suppose there is a latent factor. Hence, it becomes impractical to try to determine if they are arranged in a factor. In this study, out of the 325 correlations among the items, 30 correlations meet this requirement. Thus, it was decided to use a principal components analysis (PCA) to estimate the number of components present in the items.

The PCA method is useful for reducing items into fewer components in this particular case because its requirements for finding fewer components are less strict than the factor analysis (FA). As the purpose of FA is to identify or to model latent constructs that predict the observed indicators, it assumes that indicators include both the shared variance that constitutes the factor and residual variance. Differently, PCA is a reduction method that transforms the original indicators into fewer components that include both shared and residual variance (Fabrigar et al., 1999). Thus, whereas in FA the shared variances are analyzed, in PCA all of the observed variance is analyzed with no assumption of any particular statistical hypothesis. In other words, the PCA was used in this case because it is more a geometrical technique that calculates constructs using the variance in the measured variables and, unlike factor analysis, does not assumes normality in the covariation among indicators variables.

After obtaining evidence of the best factor structure for the samples, maximum likelihood with missing values (MLMV) analysis were used to fit path models from the new ECERS-R quality factors to developmental outcomes in children. The MLMV method uses all available information in the data and avoids dropping subjects because of missing data. This model was estimated in the two samples separately because each sample has different child outcomes measures. The analyses were clustered at the teacher level to adjust standard errors for the nesting of children in classrooms (Murnane & Willett, 2010). Across measurement and path models, outcomes were conditioned on baseline outcomes. The only exception was the set of five Woodcock-Muñoz III sub-tests from the given that were only measured at the follow-up in the IACDC sample. Several control variables were included in the model: child age, gender, household income and parental education. In the case of the EEQCDB sample, an indicator for private (in comparison with co-funded) centers was added. A diagram of the model is presented in Figure 2.1. This model was estimated separately for each outcome.

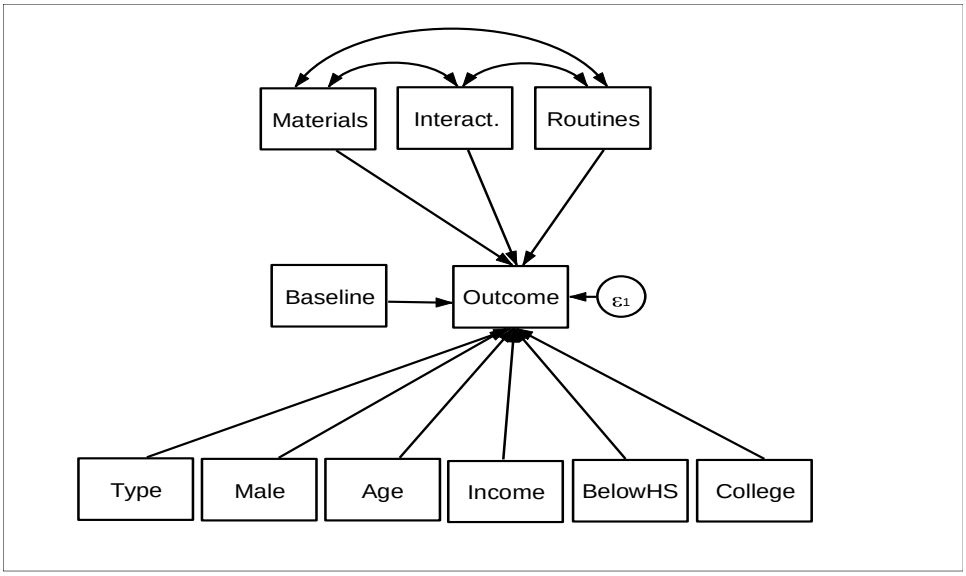

Figure 1: Diagram of the MLMV model for testing predictive validity

3.0 RESULTS

3.1 CONFIRMATORY AND EXPLORATORY FACTOR ANALYSIS

Descriptive statistics for the ECERS-R items are shown in Table 2. In this sample, the mean scores on each item are generally low, ranging from 1 to 2 which indicates that quality is "inadequate" according to ECERS-R standards. Although generally all items had low scores, it is important to highlight that items related with Interactions and Parent & Staff seem to have higher scores than the other items on the measure.

Table 2: Descriptive statistics of the ECERS-R items

Items	M	SD	Min	Max
Space and furnishing				
Indoor Space	2.01	1.31	1	7
Furniture Care	2.11	1.52	1	7
Furniture Relaxation	1.24	0.77	1	5
Room Arrangement	1.28	0.70	1	7
Space Privacy	1.54	0.84	1	4
Child Display	2.33	1.13	1	7
Space Gross Motor	2.22	1.04	1	6
Gross Motor Equipment	1.83	1.11	1	7
Care Routines				
Greeting/Departing	1.79	1.46	1	7
Meals/Snacks	1.53	1.03	1	5
Nap/Rest	1.86	2.15	1	7
Toileting	1.14	0.64	1	7
Health Practices	1.89	1.09	1	7
Safety Practices	1.31	0.88	1	7
Language and Reasoning				
Books/Pictures	1.58	0.90	1	4
Encourage to communicate	1.63	0.88	1	6
Language for Reasoning	1.63	0.90	1	4
Informal Language	1.99	1.18	1	7
Activities				
Fine Motor	1.75	0.86	1	5
Art	1.79	0.98	1	7
Music/Movement	2.28	0.79	1	5
Blocks	1.37	0.67	1	6
Sand/Water	1.21	0.53	1	4
Drama Play	1.46	0.64	1	4

Nature/Science	1.43	0.59	1	4
Math/Number	1.21	0.65	1	4
Tv/video/computer	1.39	0.59	1	4
Accepting Diversity	1.29	0.75	1	5
Interactions				
Gross Motor Supervision	2.12	1.22	1	6
General Supervision	2.15	1.27	1	7
Discipline	2.23	1.16	1	6
Staff-Child Interactions	2.49	1.29	1	7
Children Interactions	1.78	1.12	1	7
Program Structure				
Schedule	1.83	0.63	1	4
Free Play	1.66	0.84	1	5
Group Time	1.27	0.60	1	4
Provision for disabilities	1.29	0.75	1	4
Parents and Staff				
Provision for parents	1.91	1.30	1	7
Staff personal needs	1.18	0.51	1	4
Staff professional needs	2.52	1.74	1	7
Staff interactions	3.89	1.82	1	7
Staff supervision	3.70	1.89	1	7
Professional growth	2.35	1.22	1	7

The first goal of this study was to test whether the original structure of the ECERS-R was reproduced in Colombian preschool settings. For this, a confirmatory factor analysis was performed testing the original 7-factor structure. However, the model failed to converge. Second, using the original set of items, we examined the number of factors that the scale measures in Colombia. For this, a 7-factors EFA was estimated. Fit indices of the seven estimated models are presented in Table 3. No model showed to be a good fit for the data. As was predicted from previous uses of the ECERS-R in LAMI countries, it seems the original set of items in the ECERS-R does reliably measure classroom quality in Colombia.

Table 3: Fit indices of exploratory 7-factor analysis with original items

Model	Free param	df	χ^2	χ^2 p	RMSEA			CFI	TLI
						[90% CI]	p		
1-factor	129	860	2359	0.000	0.10	[0.095, 0.105]	0.000	0.33	0.24
2-factor	171	818	2066	0.000	0.09	[0.089, 0.099]	0.000	0.44	0.39
3-factor	212	777	1831	0.000	0.09	[0.083, 0.094]	0.000	0.53	0.45
4-factor	252	737	1646	0.000	0.08	[0.079, 0.090]	0.000	0.60	0.50
5-factor	291	698	1497	0.000	0.08	[0.075, 0.087]	0.000	0.64	0.54
6-factor	329	660	1310	0.000	0.07	[0.069, 0.081]	0.000	0.71	0.60
7-factor	366	623	1183	0.000	0.07	[0.066, 0.078]	0.000	0.75	0.64

3.2 USE OF ECERS-R ITEMS FOR GENERATING A VALID MEASURE

Structural validity. Although the set of original items did not show evidence of structural validity, this study explored whether some items on the ECERS-R could be used to generate a measure of child care quality with stronger psychometric qualities for the Colombian context. For this, ECERS-R items were dropped if they had low variability and lack of cultural relevance according to a questionnaire completed by Colombian professionals (see Appendix A for seeing answers to the questionnaire). Additionally, the entire *Parents and Staff* subscale was dropped because these items are not directly observed but self-reported by teachers or staff. Out of the original 43 items, 27 were maintained for this analysis: Indoor space, furniture for care, space for gross motor, gross motor equipment, greeting/departing, meals/Snacks, nap/rest, toileting, health practices, safety practices, books/pictures, encourage to communicate, language for reasoning, informal language, fine motor, art, music/movement, blocks, nature/science, math/number, gross motor supervision, general supervision, discipline, staff-child interactions, children interactions, free play, and drama play (see Appendix A for a complete list of the items dropped and kept).

After selecting the set of items to be analyzed, an exploratory factor analysis was estimated. Fit indices of the six models tested are presented in Table 4. Here it can be seen that none of the models fits the data appropriately. In this case, it may be that the little variability in the data prevents a good model fit given that this type of estimation requires variability and significant correlations between the indicators.

Table 4: Fit indices of exploratory 6-factor analysis with selected items

Model	Free param	df	χ^2	χ^2 p	RMSEA			CFI	TLI
						[90% CI]	p		
1-factor	81	324	888	0.000	0.10	[0.092, 0.108]	0.000	0.45	0.41
2-factor	107	298	724	0.000	0.09	[0.080, 0.099]	0.000	0.59	0.51
3-factor	132	273	588	0.000	0.08	[0.072, 0.091]	0.000	0.69	0.61
4-factor	156	249	487	0.000	0.07	[0.064, 0.084]	0.000	0.77	0.67
5-factor	179	226	414	0.000	0.07	[0.059, 0.080]	0.000	0.82	0.72
6-factor	201	204	356	0.000	0.07	[0.054, 0.077]	0.014	0.85	0.75

Given the limitations of doing an SEM analysis with low-variability, as was previously explained, we turned to a PCA to examine the number of components. Results shown in Table 5, indicate that items could be reduced into three different components characterized by 1) the availability of spaces and materials for learning, 2) routines, and 3) interactions and language. In addition, Cronbach's alpha coefficients are presented in Table 5. Generally, all items loaded into one of these three factors, with the exception of the *Nap/Rest* item which loaded negatively into the routines factor and decreased the reliability of the factor so this item was dropped.

Table 5: Results of principal component analysis

Items	**Components**		
	Materials and activities	**Interactions**	**Routines**
Drama play	.705*		
Furniture for care	.636*		
Fine motor	.621*		
Blocks	.555*		
Math/number	.555*		
Gross motor equipment	.540*		
Books/pictures	.538*	.301	
Encourage to communicate	.527*		
Nature/science	.438*	.306	-.376
Music and movement	.401*		
Free play	.388*	.383	
Indoor space	.381*		
Art	.355*		
General supervision		.723*	
Staff-child interactions		.719*	
Children interactions		.617*	
Informal language		.607*	
Discipline		.604*	.473
Gross motor supervision		.568*	
Language to reasoning		.527*	
Space for gross motor	.337*		.586
Safety practices			.555*
Meals/Snacks			.543*
Health practices			.508*
Toileting		.373	.453*
Nap/Rest			-.385
Greeting/Departing			.326*
Cronbach's alpha	.765	.787	.500
Correlations			
Interactions	0.352**		
Routines	0.343**	0.343**	

* Item selected into factor **Correlation is significant at the 0.01 level (2-tailed)

The reliability of the factors was good in the case of spaces and materials for learning, and interactions and language factors, but was low for the routines factor. Nevertheless, the factor was maintained because consistent loadings were obtained in the principal component analysis and because obtaining a reasonable Cronbach's alpha requires higher correlations among items than those obtained in this particular case. Factors scores were created by averaging the item's scores into each factor.

Descriptive statistics of the final ECERS scores are presented in Table 6. Statistics of the new factors show that even after dropping items with low-variability or no cultural relevance, the quality is still very low and factors have low variability. Across both samples, the interactions factor tended to have the highest scores and the most variability.

Table 6: Descriptive statistics by sub-sample

	EEQCDB (Bogota) N=226	IACDC (National) N=458
	M (SD) or %	M (SD) or %
ECERS-R factors		
Materials	1.88 (0.67)	1.70 (0.31)
Interactions	2.24 (0.92)	1.96 (0.69)
Routines	1.82 (0.81)	1.48 (0.48)

Predictive validity. After obtaining the best factor structure for the Colombian sample, the predictive validity of the three factors was tested by predicting gains in children's cognitive and socio-emotional development during the kindergarten year. Analysis included controls for children baseline scores, age at the follow-up assessment, and family characteristics. Maximum likelihood with missing values (MLMV) analysis were used to fit path models from the new ECERS-R quality factors to developmental outcomes in children. Across measurement and path models, outcomes were conditioned on baseline outcomes if available. Several control variables were included in the model: age, gender, household income and parental education. In the case of the EEQCDB sample, an indicator for co-funded (in comparison with private) centers was added.

Results for the predictive validity of the new ECERS-R factors are presented in Table 7. The three factors had distinct associations with children outcomes. The materials and activities factor show no significant associations with children's gains in any developmental

domains. In contrast, the interactions and language factor unit increase in this factor predicts gains at the end of kindergarten of 0.24 SD in sound-awareness.

The routines factor is more consistently linked to child development in both samples. In the EEQCDB sample, higher scores in routines relate to gains in both cognitive and socioemotional skills. More specifically, a one-unit increase in the routines factor is associated with 0.23 SD higher executive function, 0.13 SD more gains in picture-vocabulary, 0.24 SD gains in emotional control, and a reduction of 0.13 SD in behavioral problems. In the case of the IACDC sample, results indicate that children attending classrooms with higher routines gain more cognitive skills over the kindergarten year, such that one-unit increase in the routines factor is associated with 0.42 SD gain in associative memory, 0.23 SD increase in executive function, and 0.49 SD increase in receptive language.

Given the differentiated results that each of the factors had, in order to see if there were unique associations with each of the factors, another set of models was tested including the factors separately with all of the covariables. Most of the results were similar to those of the models including the three factors simultaneously. However, some effects changed. When included separately, the interactions factors showed significant but negative associations with Letter-Word identification, and significate and positive associations with Picture-Vocabulary. Similarly, the routines factor was, when included separately, positively associated with Picture-Vocabulary.

Table 7: Results of model with new factors predicting child's kindergarten gains

	EEQCDB (Bogota)										IACDC (National)						
	Effortful control	Executive function	Math	Letter-word id	Picture vocabu	Sound Awaren	Emotion control	Positive emotion	Social skills	Behavioral problems	ASQ Cognitive	ASQ Socio	verbal	Associati memory	Executive function	Numeric reason	Recept. Langu
Baseline	0.33***	0.26***	0.37***	0.57**	0.59***	0.19**	0.45***	0.37***	0.56***	0.44***	-0.27***	0.14**					
	(0.07)	(0.07)	(0.06)	(0.19)	(0.06)	(0.07)	(0.07)	(0.07)	(0.07)	(0.08)	(0.08)	(0.05)					
Type	0.02	0.10	-0.54***	-0.51**	0.02	0.06	-0.10	-0.03	-0.12	0.44**							
	(0.20)	(0.15)	(0.16)	(0.16)	(0.14)	(0.17)	(0.15)	(0.17)	(0.16)	(0.14)							
Male	-0.24+	-0.23+	-0.29*	0.07	0.09	-0.32	-0.32**	-0.23	0.003	0.01	0.01	-0.04	0.22**	0.10	0.07	0.01	0.41***
	(0.12)	(0.13)	(0.12)	(0.08)	(0.09)	(0.14)	(0.11)	(0.14)	(0.13)	(0.14)	(0.09)	(0.09)	(0.08)	(0.08)	(0.08)	(0.10)	(0.08)
Age	0.03*	0.03*	0.02	0.02*	0.02+	0.02	-0.013	0.002	0.01	0.01	0.04***	0.01	-0.05***	-0.03**	-0.06	0.003	0.01+
	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)	0.0172	(0.01)	(0.01)	(0.01)	(0.01)	0.00573	(0.01)	(0.01)	(0.01)	(0.01)
Min. wage	-0.04	-0.01	0.04	0.03	0.07	0.16+	-0.02	0.03	-0.04	0.10	-0.06	-0.04	0.22*	0.29*	-0.09	0.09	0.26
	(0.07)	(0.08)	(0.07)	(0.06)	(0.06)	(0.09)	(0.08)	(0.07)	(0.07)	(0.08)	(0.10)	(0.12)	(0.09)	(0.12)	(0.11)	(0.11)	(0.09)
BelowHS	0.11	0.27	0.32	0.05	-0.001	0.27	-0.28	-0.01	0.08	-0.02	-0.15	0.14	0.07	0.13	-0.01	0.06	0.14
	(0.20)	(0.20)	(0.20)	(0.11)	(0.14)	(0.20)	(0.19)	(0.20)	(0.22)	(0.19)	(0.15)	(0.16)	0.13818	(0.12)	(0.14)	(0.14)	(0.15)
College	0.19	0.39+	0.34+	0.25+	0.16	0.21	-0.05	0.10	0.01	0.01	0.07	-0.11+	0.11	0.10	-0.10+	0.11	0.01
	(0.22)	(0.23)	(0.19)	(0.13)	(0.17)	(0.22)	(0.20)	(0.22)	(0.22)	(0.23)	(0.06)	(0.06)	(0.08)	(0.07)	(0.06)	(0.06)	(0.06)
Materials	-0.01	-0.16	0.03	0.11	-0.01	-0.02	-0.07	0.032	-0.13	-0.03	0.01	-0.17	-0.14	-0.08	-0.24+	-0.08	-0.29
	(0.09)	(0.12)	(0.08)	(0.14)	(0.06)	(0.10)	(0.10)	(0.08)	(0.10)	(0.08)	(0.15)	(0.21)	(0.16)	(0.18)	(0.15)	(0.14)	(0.20)
Interact.	-0.15	-0.04	-0.03	-0.14	0.10	0.24**	-0.13	-0.11	0.15+	0.02	0.03	-0.10	0.01	-0.07	-0.15*	0.07	-0.06
	(0.11)	(0.08)	(0.08)	(0.09)	(0.08)	(0.08)	(0.09)	(0.10)	(0.08)	(0.06)	(0.08)	(0.08)	(0.05)	(0.12)	(0.07)	(0.07)	(0.08)
Routines	0.15	0.23*	0.06	-0.06	0.13+	-0.11	0.24**	0.07	-0.02	-0.13*	0.19	0.09	0.09	0.42***	0.23**	0.11	0.49***
	(0.11)	(0.10)	(0.09)	(0.09)	(0.07)	(0.07)	(0.08)	(0.09)	(0.08)	(0.05)	(0.14)	(0.12)	(0.10)	(0.11)	(0.09)	(0.10)	(0.14)
_cons	-1.63*	-1.84*	-0.75	-0.82	-1.73*	-1.72*	0.65	-0.02	-0.54	-0.38	-2.52**	0.26	1.96***	0.60	3.76**	-0.58	-1.04+
	(0.79)	(0.82)	(0.68)	(0.53)	(0.66)	(0.78)	(0.72)	(0.82)	0.71)	(0.65)	(0.82)	(0.48)	(0.48)	(0.49)	(0.50)	(0.51)	(0.58)

***p <0.001 **p <0.01 *p<0.05 +p<0.10

Table 8: Results of separate new factors predicting kindergarten gains

	EEQCDB (Bogota)										IACDC (National)						
	Effortful control	Executive function	Math	Letter-word id	Picture vocabu	Sound Awaren	Emotion control	Positive emotion	Social skills	Behavioral problems	ASQ Cognitive	ASQ Socio	verbal	Associati memory	Executive function	Numeric reason	Recept. Langu
Materials	0.009	-0.096	0.048	0.044	0.084	0.028	-0.024	0.019	-0.095	-0.066	-0.151	0.051	-0.090	0.151	-0.049	0.006	-0.049
	(0.09)	(0.12)	(0.88)	(0.12)	(0.07)	(0.09)	(0.09)	(0.08)	(0.09)	(0.08)	(0.19)	(0.18)	(0.14)	(0.22)	(0.26)	(0.15)	(0.26)
Interact.	-0.067	0.040	0.013	-0.149*	0.16*	0.18*	-0.026	-0.069	0.108	-0.052	-0.118	0.064	0.006	-0.046	-0.034	0.70	-0.034
	(0.09)	(0.070)	(0.07)	(0.07)	(0.06)	(0.07)	(0.09)	(0.08)	(0.07)	(0.06)	(0.08)	(0.09)	(0.05)	(0.12)	(0.09)	(0.06)	(0.09)
Routines	0.058	0.166*	0.052	-0.11+	0.17***	0.017	0.148*	0.013	0.020	-0.119*	0.052	0.15	0.059	0.385***	0.415**	0.102	0.415**
	(0.06)	(0.08)	(0.07)	(0.06)	(0.05)	(0.07)	(0.07)	(0.06)	(0.07)	(0.05)	(0.09)	(0.14)	0.08	(0.10)	(0.14)	(0.10)	(0.14)

***p <0.001 **p <0.01 *p<0.05 +p<0.10

4.0 DISCUSSION

The goal of this study was to examine the reliability and validity of ECERS-R for preschool settings in Colombia. Three aims were addressed in this study. First, it examines whether the original structure of the ECERS-R was reproduced in Colombian preschool settings, finding no evidence that the original structure of the scale is valid in the Colombian context. Second, using the original set of items, it examined the number of factors that the scale measures in the Colombian sample, finding no evidence that all the items in the scale are valid for measuring quality in Colombia. Third, this study explores whether items on the ECERS-R can be used to generate a measure of child care quality that is related to child development. Evidence indicates that some items have low variability and seem to not be valid in the Colombian context. After dropping these items, the remaining items consistently measure three factors: routines, interactions and availability of materials for learning. These results suggest that only some items of the ECERS-R consistently measure early education quality in Colombia. Analysis of predictive validity indicated that better routines are consistently associated with gains in executive function and language skills over the kindergarten year. This indicates that even low scores in routines can potentially improve children's cognitive and social development.

The low scores obtained with the ECERS-R suggest that the instrument should be used with caution in LAMI countries and that some items can be culturally inappropriate. The scores obtained in this study are similar to those obtained in studies in other LAMI countries, in which average levels of quality measured by the ECERS-R are generally "inadequate" or "acceptable" (Aboud, 2006; Bernal et al., 2009; Campos et al. 2010). Three important aspects may contribute to the low scores in this context: the stop-scoring system that the ECERS-R uses, the emphasis on structural aspects of early education quality, and the cultural differences in perspectives about early education.

First, the rating system of the ECERS-R is based on the *stop-scoring* system (Gordon, Fujimoto, Kaestner, Korenman, & Abner, 2012), which requires observers to stop scoring if

indicators are not endorsed, which may obscure important information about classrooms. The stop-scoring system assumes that each sub-scale contains indicators that are qualified in ascending order. Thus, the upper-level indicators of a sub-scale are scored only if the lower elements are already endorsed. If lower-level indicators are not met, the scoring stops. Unfortunately, this scoring system exhibits disorder, such that higher-level indicators are often achieved even though the lower-level indicators are not satisfied (Gordon et al., 2012). This means that when one low quality indicator is endorsed a setting is assigned a low score on this dimension, even if they satisfy indicators of higher quality on this same dimension (Gordon et al., 2012). Thus, the disorganization in the ECERS-R's items along with the stop-scoring used during data collection can contribute to the low scores found in this research. Ideally, in future research, all indicators would be scored until the end and these should all be incorporated into the final score.

Second, a large number of ECERS-R's items focus on structural characteristics of the centers, which contrast with the limited availability of resources in low- and middle-income countries generally and in the sample analyzed for this study specifically, which primarily includes preschool settings for low-income children. The scarcity of resources could be leading to low scores on items related to structural aspects of the settings, such as materials for activities and availability of spaces. In future research, in order to more carefully test the ECERS-R behaves in the Colombian context, it may be appropriate to include classrooms with greater material resources. The inclusion of classrooms with higher income would allow researchers know if the lack of variability found in this study may be caused by material homogeneity of the sample and also if the instrument is sensitive to classrooms with more marked differences. However, even if the sample used in this study was homogeneous as it was generally low-income, this finding could also be indicating that the ECERS-R has limited capability for capturing variability in high-poverty contexts. This situation raises the question about whether the measure is not adjusted for the lower availability of resources more generally found in LAMI countries. Addressing this issue is important, given that in LAMI countries the investment for improvement and assessing education is dedicated to high-risk and low-income populations.

Third, the emphasis that the ECERS-R places on child-selected activities contrasts with the values in collectivistic cultures. For example, the ECERS-R rates high classrooms in

which children spend time in small groups, are allowed to select the activities, decide on the spaces they want to be, and choose which peers they want to work with. In contrast, in the Colombian context, children are primarily in a whole group and the teacher decides when children can access materials and dictates which activities children engage in. In fact, in this study, items assessing whether children have space for privacy and if children they work in small groups were dropped because these items do not reflect the whole-group arrangements and teacher-directed activities that characterize classrooms in collectivistic cultures (Isley, 2000; Leyva et al., 2015; Li et al., 2014; Rao & Pearson, 2007).

This study evaluated which items are not culturally relevant or had too little variability, finding that 27 items out of the 43 original items could be selected for evaluating the validity of the instrument. At this point, it is very important to highlight that dropping out items based the low-variability alone was necessary in order to perform statistical analysis but it is not an adequate justification for truly suggesting that an item is not important in a given context. As it was argued, some items related individualistic values were dropped due to lack of cultural relevance: furniture for relaxation, space for privacy, room arrangement, sand/water, provision for disabilities, group time, tv/video/computer. However, the lack of cultural relevance this is not necessarily the case of items such as child display, accepting diversity, and nap time. These set of items intent to measure dimensions of education that are are important for child development in the context of Colombia, but around 85% of the classrooms obtained scores of 1 or 2 in these items. Thus, the validity of these items remain to be tested with a more heterogeneous sample.

This study did not find evidence that the ECERS–R measures a single global aspect of quality or seven subscales of quality. Instead, the factor analyses revealed that the scale measures three factors: materials for learning, interactions and language, and routines. Materials for learning is composed mostly of items belonging to the “spaces for care” and “activities”, which may reflect the general availability of resources in the classroom. The interactions factor is mainly composed by items from the original “language” and “interactions” subscales, which may reflect the classroom general quality of social and instructional interactions between teacher and students. Finally, the routines factor is composed of items mostly from the “routines” sub-scale in the ECERS-R, with the exception of the nap/rest item, which was dropped. Evidence of these three factors has been found in

previous research in the US (Cassidy et al., 2005; Early et al., 2006; Frede et al., 2007; Gordon: 2013; Sakai et al., 2003), suggesting that the instrument may be working similarly in the US context.

Each of these dimensions showed differential associations with children's outcomes, although they tended to be modest. Materials for learning factor reached low scores on average and analysis of predictive characteristics showed no significant associations with children gains. Interactions positively related to gains in sound-awareness and social skills. The routines factor showed more consistent associations with children's development in both samples. In the EEQCDB sample, higher scores on routines were associated with gains in both cognitive and socioemotional skills, whereas in the IACDC sample, children attending to classrooms with higher routines score showed larger gains in associative memory, executive function, and receptive language. Thus improvements in the consistency of schedules and practices, and better availability of activities have the potential to improving child outcomes.

The effect sizes of the ECERS-R factors on the developmental gains were modest in general. In fact, similar small effect sizes have been found for the ECERS-R (Gordon et al., 2013), which may reflect the weak psychometric properties of the scale. However, it is important to highlight that even if associations between factors and children's gains were modest, this study found evidence that the routines factor was consistently associated with children's outcomes. These results suggest that increases in the quality of education, even within the low range, have potential to impact children's development.

The modest but consistent predictive qualities of some factors suggest that focusing in improving classroom quality is important, and clearly supports the need to strengthen the measurement of child care quality in LAMI countries. In the meantime, researchers should use the ECERS-R in LAMI countries with caution as it did not show good structural or predictive validity in the Colombian context. Results obtained by the ECERS-R in this study indicate that it is urgent to develop measures of child care quality that adequately reflect the experiences of children in early care and education settings in Colombia. Besides the little variability and lack of cultural relevance of some items, this study also found lack of effects of the materials factor in both of the samples. This may suggest that the focus of the ECER'S on structural factors maybe is not adequate for measurement in low income contexts. In

contrast, this study suggests that quality improvement efforts are perhaps best targeted at interactions and routines instead of structural factors.

International organizations have been making efforts on developing an appropriate measure for LAMI countries. With the urgency of having a suitable measure of early childhood education quality in low- and middle-income countries, the Measuring Early Learning and Quality Outcomes (MELQO, 2015) project has develop a tool with culturally adaptable items. The process of development started in Tanzania, but it is intended to be used at a global level. In fact, since 2015, the project has been working along with a group of Colombian researchers (Maldonado, 2015) in order to obtain feedback about the adaptability of the MELQO in the context of Colombia. Results of a pilot study suggest that some changes are required in order to adapt the instrument to specificities of Colombian education. This could be an indication that quality measurement at the global level maybe possible, but that a meaningful measurement requires respect for cultural differences and adjustment in lower resource countries.

In conclusion, the ECERS-R results in this research indicated to have little evidence of structural and predictive validity in the Colombian context, although some items seem to be associated with child's gains over the kindergarten year. The ECERS-R, then, still should be used with caution, given that it includes items that may be culturally inappropriate and it may not capture differences in quality in high poverty contexts.

APPENDIX A

Table 9: Items selected into final analysis according to answers to questionnaire and variability

	Importance						Variability		Selected
	1. Not important	2	3	4	5. Extrem important	Total	M	SD	
Indoor space	2.94%	0.00%	8.82%	41.18%	47.06%	4.29	2.01	1.31	Y
Furniture for care	2.94%	2.94%	47.06%	29.41%	17.65%	3.56	2.11	1.52	Y
Furniture for relaxation	2.86%	2.86%	25.71%	37.14%	31.43%	3.91	1.24	0.77	N
Room arrangement	2.86%	14.29%	11.43%	28.57%	42.86%	3.94	1.28	0.7	N
Space for privacy	17.14%	22.86%	28.57%	22.86%	8.57%	2.83	1.54	0.84	N
Child display	0.00%	5.71%	40.00%	42.86%	11.43%	3.17	2.33	1.13	N
Space for gross motor	0.00%	0.00%	0.00%	25.71%	74.29%	4.74	2.22	1.04	Y
Gross motor equipment	0.00%	0.00%	0.00%	35.29%	64.71%	4.65	1.83	1.11	Y
Greeting/departing	0.00%	5.88%	5.88%	32.35%	55.88%	4.38	1.79	1.46	Y
Meals/snacks	0.00%	2.94%	0.00%	26.47%	70.59%	4.65	1.53	1.03	Y
Nap/rest	0.00%	8.82%	11.76%	17.65%	61.76%	4.32	1.86	2.15	Y
Toileting	2.94%	2.94%	5.88%	20.59%	67.65%	4.47	1.14	0.64	Y
Health practices	3.03%	3.03%	6.06%	33.33%	54.55%	4.33	1.89	1.09	Y
Safety practices	0.00%	0.00%	14.71%	17.65%	67.65%	4.53	1.31	0.88	Y
Books/pictures	2.94%	0.00%	26.47%	26.47%	44.12%	4.09	1.58	0.9	Y
Encourage communic.	0.00%	3.03%	18.18%	36.36%	42.42%	4.18	1.63	0.88	Y
Language reasoning	2.94%	0.00%	8.82%	26.47%	61.76%	4.44	1.63	0.9	Y
Informal language	3.03%	3.03%	9.09%	45.45%	39.39%	4.15	1.99	1.18	Y
Fine motor	2.94%	5.88%	8.82%	26.47%	55.88%	4.26	1.75	0.86	Y
Art	2.94%	2.94%	17.65%	14.71%	61.76%	4.29	1.79	0.98	Y
Music/Movement	2.94%	2.94%	8.82%	29.41%	55.88%	4.32	2.28	0.79	Y
Blocks	3.03%	0.00%	21.21%	33.33%	42.42%	4.12	1.37	0.67	Y
Sand/Water	0.00%	14.71%	14.71%	35.29%	35.29%	3.91	1.21	0.53	N
Drama play	5.88%	2.94%	20.59%	29.41%	41.18%	3.97	1.46	0.64	Y
Nature/science	2.94%	2.94%	11.76%	38.24%	44.12%	4.18	1.43	0.59	Y
Math/number	0.00%	0.00%	11.76%	38.24%	50.00%	4.38	1.21	0.65	Y
Tv/video/computer	8.82%	14.71%	17.65%	35.29%	23.53%	3.5	1.39	0.59	N
Accepting diversity	18.18%	6.06%	27.27%	18.18%	30.30%	3.36	1.29	0.75	N
Gross motor supervis.	2.94%	2.94%	11.76%	23.53%	58.82%	4.32	2.12	1.22	Y
General supervision	0.00%	2.94%	2.94%	23.53%	70.59%	4.62	2.15	1.27	Y
Discipline	5.88%	2.94%	2.94%	41.18%	47.06%	4.21	2.23	1.16	Y
Staff-Child interactions	3.03%	0.00%	0.00%	24.24%	72.73%	4.64	2.49	1.29	Y
Children interactions	0.00%	0.00%	2.94%	32.35%	64.71%	4.62	1.78	1.12	Y
Schedule	0.00%	0.00%	11.76%	41.18%	47.06%	4.35	1.83	0.63	Y
Free play	0.00%	0.00%	9.09%	30.30%	60.61%	4.52	1.66	0.84	Y
Group time	0.00%	8.82%	44.12%	32.35%	14.71%	3.12	1.27	0.6	N
Provision for disabili.	2.94%	2.94%	76.47%	11.76%	5.88%	4.56	1.29	0.75	N

BIBLIOGRAPHY

Aboud, F. E. (2006). Evaluation of an early childhood preschool program in rural Bangladesh. *Early Childhood Research Quarterly, 21*(1), 46–60

Beller, E. K., Stahnke, M., Butz, P., Stahl, W., & Wessels, H. (1996). Two measures of the quality of group care for infants and toddlers. *European Journal of the Psychology of Education, 11*, 151–167.

Bernal, R. & C. Quintero. (2014). Characterization of Children Younger than Five in Colombia based on the Longitudinal Colombia Survey ELCA. Unpublished manuscript Universidad de Los Andes.

Bernal, R., C. Fernández, Flórez, C.E., Gaviria, A., Ocampo, P.R., Samper, B., and Sánchez, F. (2009). Evaluación de impacto del Programa Hogares Comunitarios de Bienestar del ICBF. Documento CEDE, Ediciones Uniandes, 16, 1-29.

Bernal, R. and A. Camacho (2011). La Importancia de los Programas de Primera Infancia en Colombia. [The Importance of Early Childhood Programs in Colombia]. Imprenta Nacional de Colombia.

Bruininks, R., & Bruininks, B. (2005). Bruininks-Oseretsky Test of Motor Proficiency-2nd edition manual. Minneapolis, MN: NCS Pearson.

Bryant, D., Burchinal, M., Lau, L., & Sparling, J. (1994). Family and classroom correlates of head start children's developmental outcomes. *Early Childhood Research Quarterly, 9*, 289-309.

Burchinal, M.; Roberts, J.E.; Riggin, R.; Zeisel, S. A.; Neebe, E., and Bryant, D. (2000). Relating quality of center-based child care to early cognitive and language development longitudinally. *Child Development, 71*(2), 339-57.

Campos, M. M.; Esposito, Y.; Bhering, E.; Gimenes, N., & Abuchaim, B. (2011). Early childhood education quality: a Brazilian study in six capitals. *Cadernos de Pesquisa, 41*(142), 20-54

Cassidy, D. J., Hestenes, L. L., Hegde, A., Hestenes, S., & Mims, S. (2005). Measurement of quality in preschool child care classrooms: An exploratory and confirmatory factor analysis of the Early Childhood Environment Rating Scale-Revised. *Early Childhood Research Quarterly, 20,* 345-360.

Clifford, R. M., Barbarin, O., Chang, F., Early, D., Bryant, D., Howes, C., et al. (2005). What is pre-kindergarten? Characteristics of public pre-kindergarten programs. *Applied Developmental Science, 9*(3), 126-143.

Early, D. M., Bryant, D. M., Pianta, R. C., Clifford, R. M., Burchinal, M. R., Ritchie, S., & Barbarin, O. (2006). Are teachers' education, major, and credentials related to classroom quality and children's academic gains in pre-kindergarten? *Early Childhood Research Quarterly*, *21*(2), 174–195.

Engle, P., M. Black, J. Behrman, M. Cabral de Mello, P. Gertler, L. Kapiriri, R. Martorell, M. Eming Young &The International Child Development Steering Group. (2007). Child development in developing countries: strategies to avoid the loss of development potential in more than 200 million children in the developing world. *Lancet. Child Development Series, 369.*

Marianne Fenech, M. (2011). An Analysis of the Conceptualisation of 'Quality' in Early Childhood Education and Care Empirical Research: Promoting 'Blind Spots' as Foci for Future Research. *Contemporary Issues in Early Childhood, 12*(2), 102.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, *4*(3), 272-299.

Fernald, L. C. H., Kariger, P., Engle, P., and Raikes, A. (2009). *Examining early child development in low-income countries*. Washington DC: The World Bank.

Frede E, Jung K, Barnett WS, Lamy CE, Figueras A. (2007). The Abbott Preschool Program Longitudinal Effects Study (APPLES): Interim report. National Institute for Early Education Research; New Brunswick, NJ. Available at http://www.nj.gov/education/ece/research/apples2.pdf

Ginsburg, H. P., & Baroody, A. J. (1990). *Test of Early Mathematics Ability, Second Edition: Examiner's manual*. Austin, TX: PRO-ED, Inc.

Gordon, R. A., Fujimoto, K., Kaestner, R., Korenman, S., & Abner, K. (2013). An assessment of the validity of the ECERS-R with implications for measures of childcare quality and relations to child development. Developmental Psychology, 49,146–160.

Gresham, F.M., & Elliott, S.N. (2008). Social Skills Improvement System (SSIS) Rating Scales Manual. Minneapolis, MN: Pearson.

Grantham-McGregor, S.; Cheung, Y. B.; Cueto, S.; Glewwe, P-; Richter, L.; Strupp, B., & the International Child Development Steering Group. (2007). Developmental potential in the first 5 years for children in developing countries. Lancet, 369, 60-69.

Handal, A. J., Lozoff, B., Breilh, J., and Harlow, S. D. (2007). Sociodemographic and nutritional correlates of neurobehavioral development: A study of young children in a rural region of Ecuador. *Revista Panamericana de Salud Pública, 21,* 292-300.

Harms, T., Clifford, R. M., & Cryer, D. (1998). *Early Childhood Environment Rating Scale-Revised Edition*. New York: Teachers College Press.

Heo, K. H., Squires, J., and Yovanoff, P. (2007). Cross-cultural adaptation of a pre-school screening instrument: Comparison of Korean and US populations. *Journal of Intellectual Disability Research, 52*(3), 195-206.

Holloway, S. D., Kagan, S., Fuller, B., Tsou, L., & Carroll, J. (2001). Assessing child-carequality with a telephone interview. *Early Childhood Research Quarterly, 16*(2), 165–189.

Howes, C., Burchinal, M., Pianta, R., Bryant, D., Early, D., Clifford, R., et al. (2008). Ready to learn? Children's pre-academic achievement in pre-kindergarten programs. *Early Childhood Research Quarterly, 23*, 27–50.

Hu, L.-T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives alternatives. *Structural Equation Modeling, 6*, 1–55.

Ishimine, K & Wilson, R. (2009). Centre-based child care quality in urban Australia. *Australasian Journal of Early Childhood, 34*(2), 19-28.

Ishimine, K.; Wilson, R. & Evans, D. (2010). Quality of Australian childcare and children's social skills.*International Journal of Early Years Education, 18*(2), 159 -175.

Isley, B. J. (2000). *Tamil Nadu Early Childhood Environment Rating Scale (TECERS).* Madras, India: M.S. Swaminathan Research Foundation

Leyva, D; Weiland, C.; Barata, M.; Yoshikawa, H.; Snow, C.,; Treviño, E, and Rolla, A. (2015). Teacher-Child Interactions in Chile and Their Associations with Prekindergarten Outcomes. *Child Development.*

Klibanoff, R.S.; Levine, S.C.; Huttenlocher, J.; Vasilyeva. M., & Hedges, L. V. (2006). Preschool children's mathematical knowledge: The effect of teacher "math talk". *Developmental Psychology, 42*(1):59-69.

Li, K.; Ying Hu, Y.; Pan, Y; Qin, J, & Fan, X. (2014). Quarterly Chinese Early Childhood Environment Rating Scale (trial) (CECERS): A validity study. *Early Childhood Research Quarterly 29,* 268–282

Maldonado-Carreño, C., & Votruba-Drzal, E. (2014). Early childhood development and preschool quality in Bogota, Colombia. Unpublished manuscript, Universidad de Los Andes, Bogota, Colombia.

Malmberg, L., Mwaura, P., & Sylva, K. (2010). Effects of a preschool intervention on cognitive development among East-African preschool children: A flexibly time-coded growth model. *Early Childhood Research Quarterly, 26*, 124-133

Mathers, S., Linskey, F., Seddon, J. and Sylva, K. (2007). Using quality rating scales for professional development: experiences from the UK. *International Journal of Early Years Education, 15*(3): 261-274.

Mathiesen, M. E., Merino, J., Herrera, M. E., Castro, G., & Rodríguez, C. (2011). Validación del funcionamiento de la escala ECERS-R en algunas regiones de Chile. *Estudios Pedagógicos, 2*, 147-160.

Moore, A. C., Akhter, S., & Aboud, F. E. (2008). Evaluating an improved quality preschool program in rural Bangladesh. *International Journal of Educational Development, 28*, 118-131.

Munton, A. G., Rowland, L., Mooney, A., & Lera, M. J. (1997). Using the Early Childhood Environment Rating Scale (ECERS) to evaluate quality of nursery provision in England: Some data concerning reliability. Educational Research, 39, 99–104.

Muñoz-Sandoval, A.F., Woodock, R.W., McGrew, K.S., Mather, N., and Schrank, F.A. (2005). *Bateria III Woodcock-Muñoz*. Itasca, IL: Riverside Publishers.

Murnane, R. J., & Willett, J. B. (2010). *Methods matter: Improving causal inference in educational and social science research*. Oxford University Press.

Muthén, L. K.., & Muthén, B. O. (2012). Mplus User's Guide (Sixth Edition). Los Angeles, CA: Muthén & Muthén.

Myers, R. (1992) "Early Childhood Development Programs in Latin America: Toward a Definition of an Investment Strategy". Washington, D.C. The World Bank, Latin American and the Caribbean Region, Human Resources Division.

Myers, R.G (2006). In search for quality programs of early childhood care and education. Background paper for Education for All, Global Monitoring Report 2005. Paris: UNESCO. Available at www.unesco.org/education/gmr_download/references_2005.pdf

Peisner-Feinberg, E. S., & Burchinal, M. R. (1997). Relations between preschool children's experiences and concurrent development: The Cost, Quality, and Outcome Study. *Merrill Palmer Quarterly, 43*(3), 451-477.

Peisner-Feinberg, E., Burchinal, M., Clifford, R., Culkin, M., Howes, C., Kagan, S., et al. (2001).The relation of preschool child-care quality to children's cognitive and social development trajectories through second grade. *Child Development, 72*, 1534-1553.

Peisner-Feinberg, E. S., Burchinal, M. R., Clifford, R. M., Culkin, M.L., Howes, C., Kagan, S. L., Yazejian, N., Byler, P., Rustici, J., & Zelazo, J. (1999). *The children of the Cost, Oualitv. and Outcomes. Study go to school: Technical report*. Chapel Hill: University of North Carolina at Chapel Hill, Frank Porter Graham Child Development Center.

Pena, E. D. (2007). Lost in translation: Methodological considerations in cross-cultural research. Child development, 78(4), 1255-1264.

Sireci, S. G., & Berberoglu, G. (2000). Using bilingual respondents to evaluate translatedadapted items. Applied Measurement in Education, 13, 229 – 248.

Smith-Donald, R., Raver, C. C., Hayes, T., & Richardson, B. (2007). Preliminary construct and concurrent validity of the Preschool Self-Regulation Assessment (PSRA) for field-based research. *Early Childhood Research Quarterly, 22*, 173-187.

Perlman, M., Zellman, G. L., & Le, V. (2004). Examining the psychometric properties of the Early Childhood Environment Rating Scale-Revised (ECERS–R). *Early Childhood Research Quarterly, 19*, 398- 412.

Phillipsen, L. C., Burchinal, M. R., Howes, C., & Cryer, D. (1997). The prediction of process quality from structural features of child care. *Early Childhood Research Quarterly, 12*, 281-303.

Rao, N. (2010) Preschool Quality and the Development of Children from Economically Disadvantaged Families in India. *Early Education and Development, 21*(2), 167-185.

Rao, N. & Pearson, E. (2007). An evaluation of early childhood care and education programmes in Cambodia. UNICEF, Geneva. Available at http://www. unicef. org/evaldatabase/files/CBD_early_childhoodcare_evaluation. pdf

Rentzou, K. (2010). Using the ACEI Global Guidelines Assessment to Evaluate the Quality of Early Child Care in Greek Settings.*Early Childhood Education Journal, 38(*1), 75-80.

Rubio, M.; Pinzón, L. & Gutiérrez, M. (2010). Atención integral a la primera infancia en Colombia: estrategia de país 2011-2014. *Nota Técnica # 244, Banco Interamericano de Desarrollo.*

Rubio-Codina, M., Araujo, M. C., Attanasio, O., Muñoz, P., and Grantham-McGregor, S. (2015) Comparisson of instruments to measure early childhood development on children 6-24 months. Unpublished Manuscript.

Sakai, L. M., Whitebook, M., Wishard, A., & Howes, C. (2003). Evaluating the EarlyChildhood Environment Rating Scale (ECERS): Assessing differences betweenthe first and revised editions. *Early Childhood Research Quality, 18*(4), 427–445.

Scarr, S., Eisenberg, M., & Deater-Deckard, K. (1994). Measurement of quality in child care centers. *Early Childhood Research Quarterly, 9,* 131–151.

Schrank, F. (2011). Woodcock-Johnson III Tests of Cognitive Abilities. Davis, A. (Ed.) Handbook of Pediatric Neuropsychology. Springer, NY.

Squires, J., Bricker, D., Heo, K., and Twombly, E. (2002). Identification of social-emotional problems in young children using a parent-completed screening measure. *Early Childhood Research Quarterly, 16*(4), 405-419.

Sylva, K., Siraj-Blatchford, I., Taggart, B., Sammons, P., Melhuish, E., Elliot, K., & Totsia,V. (2006). Capturing quality in early childhood through environmental ratingscales. *Early Childhood Research Quarterly, 21*(1), 76–92.

Tabachnick, B. G., & Fidell, L. S. (2001). *Using Multivariate Statistics*. Boston: Allyn and Bacon.

Tietze, W.; Cryer, D.; Bairrao, J.; Palacios, J., & Wetzel, G. (1996). Comparisons of observed process quality in early child care and education programs in five countries. *Early Childhood Research Quarterly, 11* (4), 447-475.

Treviño, E., Toledo, G., & Gempp, R. (2012). *¿Cómo son las prácticas docentes en aulas preescolares? Un análisis factorial para aulas chilenas.* Presentedin Colloquium Mide-UC. Santiago de Chile: Centro de Políticas Comparadas de Educación. Avaliable in http://mideuc.cl/wp-content/uploads/2012/10/practicas-docentes-en-aulas-preescolares.pdf

Unesco. (2012). *Education for All Global Monitoring Report*. *Youth and skills.* Available at *http://en.unesco.org/gem-report/report/2012/youth-and-skills-putting-education-work#sthash.buVD89K9.dpbs*

Unesco. (2015). *Education for All Global Monitoring Report, 2000-2015*. *Achievements and challenges*. Available at http://unesdoc.unesco.org/images/0023/002324/232435s.pdf

Villalon, M., Suzuki, E., Herrera, M. O., & Mathiesen, M. E. (2002). Quality of Chilean earlychildhood education from an international perspective. International Journal of Early Years Education, 10(1), 49-59.

WHO, Multicentre Growth Reference Study Group (2006). WHO Child Growth Standards: Length/height-for-age, weight-for-age, weight-for-length, weight-for-height and body mass index-for-age: Methods and development. Geneva: World Health Organization; pp 312. Available at http://www.who.int/childgrowth/publications/en/

WHO, Multicentre Growth Reference Study Group (2007). WHO Child Growth Standards: Head circumference-for-age, arm circumference-for-age, triceps skinfold-for-age and subscapular skinfold-for-age: Methods and development. Geneva: World Health Organization.

Wortham, S. (2013). Looking at Early Childhood Programs from a Global Perspective. In: *Common Characteristics and Unique Qualities in Preschool Programs.* Susan Wortham (Ed). San Antonio, TX: Springer.

Yoshikawa, H. & Kabay, S. (2014). The Evidence Base on Early Childhood Care and Education in Global Contexts. Background Paper for the UNESCO 2015 Education for All Global Monitoring Report.