The Quality of Centros Infantiles del Buen Vivir in Ecuador

María Caridad Araujo
Florencia López-Boo
Rafael Novella
Sara Schodt
Romina Tomé

The Quality of Centros Infantiles del Buen Vivir in Ecuador

María Caridad Araujo
Florencia López-Boo
Rafael Novella
Sara Schodt
Romina Tomé

Inter-American
Development Bank

scl-sph@iadb.org

http://www.iadb.org/SocialProtection

The Quality of Centros Infantiles del Buen Vivir in Ecuador[1]

María Caridad Araujo, Florencia López-Boo, Rafael Novella, Sara Schodt and
Romina Tomé

Abstract

This study analyzes data collected in 2012 from a sample of public child care centers in
Ecuador. It aims to characterize different quality aspects of the services provided by these
centers to children under 3 years of age. The main finding is that the child care services
studied have a considerable margin for improvement in terms of structural and process
aspects. The quality challenges faced by these centers are consistent with those of similar
services in different countries throughout the region, such as Peru, Colombia, and Brazil.
The study also finds that while efforts undertaken in recent years in Ecuador to improve
service quality are aimed in the right direction, continuous, long-term interventions will be
required in order to make significant improvements in the quality indicators studied. The
most complex changes—but also the most necessary—are related to aspects of process
quality, particularly those connected to the training of child care staff and the skill-building
necessary to offer experiences that maximize well-being and promote learning in all areas of
child development.

JEL Classification: I00, I10, I20, I25, I30, I38, J13

Key Words: childcare services, childcare centers, quality, Ecuador, Centros Infantiles del
Buen Vivir, child development

---

1. Introduction

In this study, we present an analysis of information gathered in 2012 from a sample of 404 *Centros Infantiles del Buen Vivir* (CIBVs) in Ecuador, with the aim of characterizing different aspects of the quality of the services these centers provide to children under 3 years of age. The study was coordinated by the Ministries of Economic and Social Inclusion (MIES) and Social Development (MCDS) and implemented by the Inter-American Development Bank.

The CIBVs are the primary providers of public child care services in Ecuador. According to CIBV administrative data, in 2011 these centers served some 140,000 children at approximately 3,800 centers throughout the country (Araujo, López-Boo and Puyana, 2013). The service mainly operates under third-party agreements with local governments, community organizations, foundations, churches, etc., which receive a transfer of public resources to cover the service's operating costs. Some of these entities supplement the funds with their own resources. In 2012, the service began implementing several reforms in an effort to improve quality, including a review of quality standards, the hiring of a person with postsecondary education to fill the role of coordinator[2] at each center, and the introduction of various compliance verification processes to ensure minimum safety conditions at centers. Another reform implemented at the CIBVs during the study period was the transitioning of children 4 and 5 years of age to early childhood education services run by the Ministry of Education. Although the service had traditionally served children between 0 and 60 months of age, the new reform restricted the care provided by CIBVs to children between the ages of 45 days old and 36 months of age.

In order to evaluate different aspects of CIBV service quality across a sample of 404 centers, we administered a variety of tools that were designed to assess child care services for infants and toddlers and have been utilized in other countries in the region. We focused on measuring the quality of child care services for the youngest beneficiaries, particularly those under 36 months of age. To our knowledge, this study is the first to compare quality measures from such a varied set of instruments in the environment of a developing country, and therein lies its main contribution. In this paper, we present an analysis of the quality of Ecuadorian child care services through the lens of various tools, with a focus on different aspects of the quality of the services in question. The data reported here can serve as a

---

[2] According to the CIBV Procedures Manual – 2012, the coordinators are responsible for both the programming of educational activities at the center as well as the strengthening of the skills of caregivers and community support staff. In addition, they are the point of contact between the center and the MIES, and they are responsible for administrative and maintenance tasks.

baseline for the various CIBV quality improvement initiatives that have been undertaken in recent years.

Our main finding—according to all of the quality measures implemented—is that the CIBVs still have a considerable margin for improvement in terms of the structural and process aspects of their child care services. The quality challenges faced by Ecuador's CIBVs are consistent with those of similar services in different countries throughout the region, such as Peru, Colombia, and Brazil, as discussed in Berlinksi and Schady (2015), where scales like the ones used here were administered as part of a far-reaching reform process focused on improving quality. Using the most precise and rigorous tools used to measure service quality, we identified that few centers meet acceptable quality standards and that there is an important opportunity to improve the quality of these services. We also found that the efforts undertaken in recent years to improve quality are aimed in the right direction, although we acknowledge that significant improvements in quality indicators will require continuous, long-term interventions that achieve change beyond structural aspects and center infrastructure and equipment. Moreover, the most complex changes—but also the most necessary—are related to aspects of process quality, particularly those connected to the training of child care staff and the ability to offer experiences that maximize well-being and promote learning in all areas of child development.

This paper is organized into five parts. In the first section, we present a brief conceptual introduction to measuring the quality of child care services. In the second section, we describe the methodology used to collect the data reported in this study. In the third section, we describe the instruments used to measure quality and their scoring distributions. In the fourth section, we analyze the internal consistency of the quality measurements collected, summarize the association between them, and discuss some of the improvements that have been undertaken in recent years to increase quality at the CIBVs. Lastly, in the fifth section, we present our conclusions.

2. Measuring the Quality of Child Care for Infants and Toddlers

There is a consensus among experts about the importance of the quality of child care services, particularly those aimed at infants and toddlers. It has been documented that poor-quality child care services have neutral or even detrimental effects on child development (Baker, Gruber and Milligan, 2008; NICHD, 2000L). In the case of Ecuador, Rosero and Oosterbeek (2011) exploit a discontinuity in the formula used to allocate resources to fund various community-based child development programs. They identify that children who attend child care, as compared to those who do not, fail to make gains in social and motor

development and even demonstrate significant negative results in terms of nutritional status, cognition and vocabulary development. Furthermore, the mothers of children attending daycare more frequently experience symptoms of depression and are less receptive to their children's needs. In this context, understanding what goes on at child care centers and improving children's experiences become crucial. Nonetheless, it must be acknowledged that the starting point documented in the study by Rosero and Oosterbeek presents enormous challenges.

The specialized literature identifies two types of variables that describe child care quality, structural variables and process variables, both of which were measured in this study. Structural variables mainly encompass those factors that are easier to observe or report, such as basic infrastructure characteristics, staff educational profile and salaries, and child-caregiver ratios. Process variables describe the quality of interactions—whether between children and adults, among children, or between parents and center staff—and the routines and activities that take place at the center. Evidence has shown that process variables have an effect on children's cognitive and emotional development.[3] By their very nature, process variables present more complex measurement challenges because they require not only careful observation of interactions, routines and activities but also the design of a metric that allows them to be evaluated or converted into different quality levels in a way that captures the essential elements of these variables while maintaining consistency among different observers/raters.

In a meta-analysis that includes 34 studies, Vandell and Wolfe (2000) find that both structural and process variables are consistently interrelated. For example, when the child-caregiver ratio is higher, interactions occur less frequently. Other research provides evidence that caregivers with higher levels of education and specific training in early childhood education are more likely to engage the children in stimulating, age-appropriate activities. The structural variables most closely associated with processes, such as the number of children per group, the child-caregiver ratio, and caregivers' specialized training, are also the ones more correlated with the quality of the child care environment and child development outcomes. For example, the presence of more adults per child in the classroom allows for more complex verbal exchanges that encourage the development of children's verbal skills. More frequent cooperative and positive interactions between adults and children are also seen in these classrooms (Vandell and Wolf, 2000; Early et al., 2007; NICHD, 1999).

---

[3] An updated review of the literature on the measurement of child care services for infants and toddlers can be found in López-Boo et al. (2015).

In order to identify the main attributes of quality child care that matter to child development, Mashburn et al. (2008) compared information about the quality of early childhood education services in 11 U.S. states with child outcomes in the areas of language, social-emotional development and school readiness. The authors explored three different ways of defining quality: (a) the characteristics of each program's design and infrastructure (according to the standards outlined in the National Institute for Early Education Research's [NIEER] report *The State of Preschool*), (b) observations of structural and process quality (measured using the Early Childhood Environment Rating Scale, Revised Edition [ECERS-R]), and (c) observations of the quality of classroom interactions (measured using the Classroom Assessment Scoring System [CLASS]). The main finding of this study is that the quality of interactions is the measure most consistently related to child outcomes. Furthermore, there is evidence documenting that quality is crucial for the youngest children (Weilin et al., 2012). More specifically, a quality improvement of one standard deviation in the score on the Observational Record of the Caregiving Environment (ORCE) results in an increase of 0.15 standard deviations in the level of cognitive development in children at 24 months of age. Additionally, this study finds that if the focus on quality does not also extend to early childhood and preschool education, the impact of quality child care services for infants and toddlers disappears over time.

3. Methodology

In 2012, when data collection for this study was performed, the CIBVs were run by the Institute for Children and Families (INFA), which was later incorporated into the MIES. The population from which the sample in this study was selected, or the sampling frame, consisted of all child care centers in INFA's administrative databases in May 2012. That year, these databases contained information on a population of 3,575 centers, including data on the number of children enrolled and the number of community staff that worked at these centers. At that time, these centers accounted for about 92.4% of the Institute's CIBV. The remaining 7.6% consisted of centers whose information in INFA's systems was out of date when the sample was selected. It seems reasonable to assume that those centers that were not even registered in the INFA database are also those most likely to have lower levels of quality. Should this be the case, the levels of quality described in this study suffer from a positive bias.

The sample was stratified into two groups on the basis of child-caregiver ratios (high/low). For the purpose of sample stratification, child-caregiver ratios were calculated by dividing the total number of children, of any age, enrolled at the center by the total number of

adult caregivers. In other words, adults whose main duties revolve around food preparation were excluded from this calculation. The reason why this structural variable was chosen for stratification is that the specialized literature shows that child-caregiver ratios are strongly associated with the quality of care children receive. A low child-caregiver ratio is desirable for groups of infants and toddlers because it allows caregivers more time for individual interactions with the children. Moreover, a smaller group of children minimizes disease transmission, and the adults are better able to ensure the safety of the children in their care.

For sampling purposes, of the 3,575 centers, 16 were excluded because they did not have specific information about their caregivers. In addition, four centers located on the Galapagos Islands were eliminated from the study due to the costs to access them, resulting in a potential study population of 3,555 centers. At these centers, the median child-caregiver ratio based on administrative data corresponds to 9.2 children per adult. They were then divided into two groups, one with centers with high child-caregiver ratios, which ranged from 9.2 to 60 children per adult (1,779 centers), and the other with centers with low child-caregiver ratios, which ranged from 0.3 to 9.2 children per adult (1,776 centers). Two hundred centers were randomly chosen from each of these two groups.

In addition to this sample of 400 centers, 50 replacement centers were selected, balanced between the two strata, in anticipation of potential difficulties with data collection at some sites during the fieldwork phase. Indeed, the firm responsible for data collection had difficulty completing the study using the originally selected sample. For example, 32 centers were closed down in the span of time between sample selection and data collection (a three- to six-month span). This situation can be explained by the compliance verification process conducted by INFA in 2012 to ensure minimum safety conditions, which resulted in the closure of several centers. In addition, 10 centers were replaced because they were located in areas of difficult access. These cases were concentrated in the provinces of the Amazon region, in areas requiring expensive travel by air or water. One center was replaced because its directors refused to participate in the study. Most of the centers that were replaced were located in the coastal region. When making replacements, an effort was made to maintain the same regional diversity of the original sample, i.e., a substitute was chosen in the same canton and province as the center being replaced. For this reason, it was necessary to identify 20 additional replacements in the coastal region. In other words, the selection of centers to be used as replacements was random and restricted to the regions, provinces and cantons that corresponded to those centers that would not be visited.[4] Data collection was

---

[4] Sixty percent of the replacements belonged to the stratum of centers with child-caregiver ratios greater than the median and the remaining 40% belonged to the other stratum.

performed on a final sample of 404 centers. For the aforementioned reasons, this sample is likely to have a slight selection bias, since the sample excluded centers with incomplete information in the INFA databases, centers that were recently closed for violating INFA standards, and centers located in remote areas. In this sense, the quality measures presented herein may represent an upper bound of the actual distribution of quality in the universe of CIBVs at the time the study was conducted.

Table 1 compares the characteristics of the centers in the study's sampling frame using the variables available in INFA's administrative databases. It also presents the values of these same variables for the centers in the sample, according to the data collected during the study's fieldwork phase. In both cases, the groups of centers with high and low child-caregiver ratios are compared, and the t-test results for differences between the means are reported (correcting standard errors for clustering at the canton level).

The table highlights that centers with high and low child-caregiver ratios differ on most of the variables for which information was available. The centers with the highest child-caregiver ratios are also those with the greatest number of children enrolled, for each age group and overall. Despite having the largest enrollment, centers with higher child-caregiver ratios employ less child care staff and more food service staff; therefore, no significant differences are observed between the two groups in terms of total staff. In terms of location, a greater percentage of centers with high child-caregiver ratios are found in urban parishes,[5] cantons containing provincial capitals, and the coastal region. In contrast, a lower percentage of centers with high child-caregiver ratios are located in mountainous areas. It is observed that the sample replicates the characteristics of the population from which it was drawn for all of the variables presented in table 1. Due to the lower number of observations, the significance of some differences is lost when comparing the population columns to the corresponding columns of the study sample, as a result of larger standard errors.

---

[5] In Ecuador, parishes are administrative subdivisions below cantons.

Table 1. Characteristics of the Center Population and Sample

	Population			Sample		
	Coef.≥ 9.2	Coef.< 9.2	Diff. (t-test)	Coef.≥9,2	Coef.< 9.2	Diff. (t-test)
Child-caregiver ratio	12.1	7.5	***	12.0	7.7	***
Total number of children	36.9	28.4	***	35.9	27.5	***
Children ages 0-1	1.4	1.3	*	1.3	1.2	
Children ages 1-2	5.7	4.7	***	5.8	4.8	***
Children ages 2-3	9.4	7.4	***	9.1	7.3	***
Children age 3 and older	20.4	15.0	***	19.8	14.2	***
Total number of staff	4.9	4.9		4.7	4.8	
Caregivers	3.3	3.8	***	3.2	3.6	***
Food service staff	1.6	1.2	***	1.5	1.2	***
Geographic location (%)						
Urban	53.3	45.4	***	49.0	38.0	*
Canton with provincial capital	41.8	38.4	*	38.0	42.0	
Coast	36.7	26.9	**	35.0	22.5	***
Mountains	50.4	59.3	***	53.5	63.5	**
Amazon region	12.9	13.7		11.5	14.0	
Type of operating entity (%)						
Municipality	25.7	32.5	**	25.0	37.0	**
Parish council	22.4	24.0		26.0	22.5	
Intl. NGO/committee/religious entity	11.7	8.2	**	9.0	10.0	
Provincial/central government	0.4	0.7	*	0.0	1.5	
Others	37.3	31.8		38.0	28.0	*
Observations	1,779	1,776		201	203	

Standard errors of the mean adjusted for clustering at the canton level
Significant differences at *** 99%, ** 95%,* 90%
Sources: INFA administrative databases (2012) and data collected by the authors

During the fieldwork phase, each of the centers in the sample was visited for a full day by a pair of researchers responsible for data collection. Since the staff at the CIBVs is predominantly female, it was requested that all members of the field research team be women in order to achieve greater rapport and trust between center personnel and the researchers.

In terms of the researchers' profile, each pair was required to include an experienced interviewer, who was responsible for conducting interviews and filming, and a researcher with post-secondary education in the field of child psychology or early childhood education, who was responsible for the administration of the most complex observational instruments. Each pair of researchers was assigned a field supervisor (male or female) who was trained on all of the instruments, and whose specific task during the data collection process, in

addition to logistical organization, was to administer a questionnaire on infrastructure variables. All of the instruments were administered in the same order and at the same time of day to ensure comparability across centers. Table 2 describes the team's work routine at each center.

Table 2. Work Routine During the Visit to Each Center

Time	Researcher 1	Researcher 2	Supervisor
7:30 a.m.	Arrival and preparation, selection of *group* to be studied		
8:00 a.m.-12:00 p.m.	Filming for CLASS coding Administration of MITRCC	ITERS-R	Structural quality questionnaire
12:00 p.m.-2:00 p.m.	Interview with caregiver (ITFI, Teacher Practices Survey, KIDI)	HOME	
2:00 p.m.-3:00 p.m.	Interview with coordinator (KIDI)		

The instruments administered are the Toddler Classroom Assessment Scoring System (CLASS), the Missouri Infant/Toddler Responsive Caregiving Checklist (MITRCC), Infant/Toddler Environment Rating Scale – Revised Edition (ITERS-R), the Infant-Toddler and Family Instrument (ITFI), the Knowledge of Infant Development Inventory (KIDI), and the Child Care Infant/Toddler Home Observation for Measurement of the Environment (HOME). These instruments are described in detail in section 4.

Upon arrival at the center, the researchers selected the *group* of children and the caregiver(s) who would be the subject of their observations throughout the day. In a school environment, the group to be studied would be the equivalent of a classroom with its teacher(s) and students. Since these child care services are organized differently from a school, for the remainder of the document, the word "group" is used to refer to the unit of analysis for this study, the caregivers and the group of children in their care, who do not always have a dedicated space for their activities and daily routine.

Group selection followed the same protocol at all of the centers visited. Since the main goal of this research was to study the quality of care for children under 36 months of age, the first step was to identify how many groups at the center had children who fell within this age range.[6] If the center only had one group of children in this age range, then that group constituted the study group. If there was more than one group of children in this age range, priority was given to the one in which all of the children fell within that range. For example, if there were two groups, one with children ages 0 to 24 months and the other with children ages 25 to 48 months, the former would be selected. When there were either multiple groups in which all of the children were under 36 months or no group in which all of the children were under 36 months, one group was selected at random; the researchers

---

[6] The age of the children considered for selection corresponded to their age at the beginning of the current school year. In other words, the reference group was composed of children who were 36 months old or younger at the beginning of the school year; therefore, it is possible for the study group to include children over the age of 36 months at the time of data collection.

assigned a number to each group and rolled a dice to choose. Once the selection was made, the researchers focused on the selected group for the rest of the day.

The fieldwork phase was conducted between August 5 and October 19, 2012. Although CIBV operational guidelines at the time of the study (CIBV Procedures and Operations Manual – 2012) dictated that children should be grouped by age,[7] one of the difficulties encountered during fieldwork was that this was virtually never the case; only two of the 404 centers visited (0.5%) followed the established guidelines for grouping children in all of their groups. Table 3 shows how many of the groups of children at the centers studied failed to meet these guidelines. This table relates to all of the groups of children at the centers in the sample, not just those included in the observations that form part of this study. At 93% of the centers, more than 50% of the groups of children did not conform to the age range established by the guidelines. Notably, at 283 centers (70%), none of the groups of children met the guidelines. For the purposes of this study, the fact that the children were not grouped by age caused the age composition of the groups studied to be more heterogeneous than planned. It is important to note that the instruments selected to measure child care quality were specifically designed for children *under* the age of 3. Since some of the groups in the sample included children over the age of 3, it is interesting that systematic changes were found in the quality measures in groups with older children. This is an initial finding of the quality assessment exercise for these services. In a child care setting like that of the CIBVs, where the staff members in charge of the children are not professionals, caregivers already have less ability to individualize the care provided to a homogeneous age group; a variety of ages make this task all the more complex. For this reason, in settings where children are grouped with peers of the same age, as provided for by the CIBV guidelines, it is expected that caregivers will be better able to facilitate the implementation of age-appropriate learning activities and high-quality interactions.

Table 3. Percentage of Center Groups that Do Not Meet the Guidelines for Age Grouping

% of groups that do not meet guidelines for age grouping	Number of centers
100	283
75-99	30
50-74	63
25-49	21
1-24	5
0	2
Total	404

---

[7] The study guidelines provided by INFA indicate that the children should be grouped into the following five age ranges: 3-12 months, 12-24 months, 24-36 months, 36-48 months and 48-59 months.

As a supplement to the information in table 3, figure 1 describes the age-range distribution of the children in every group at every center in the sample (i.e., not just the group that was observed). Each horizontal line relates to the percentage of groups containing children in a specific age range with respect to all of the groups in the sample. Some patterns stand out. First, as previously mentioned, the CIBVs generally group the children into very broad age ranges. Second, although the service seeks to prioritize the care of the youngest children, at the time of data collection, a significant percentage of the groups included children over the age of 3. With that said, it is important to note that if the service has historically provided care to children under the age of 6, the transition to a program that focuses solely on children ages 0 to 36 months involves significant changes in terms of physical infrastructure, such as the daily routine and educational activities and, by extension, the training of the staff responsible for the care of those children.

Figure 1: Distribution of Age Ranges Across Center Groups
(% with respect to all center groups in the sample)
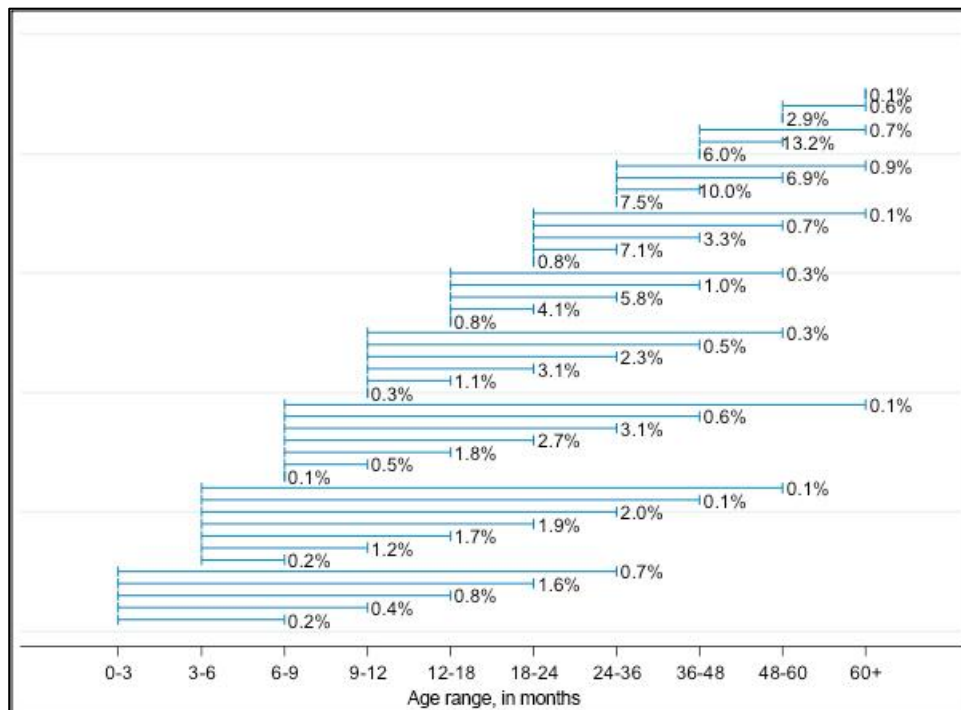
Since the children were not grouped by age, it was more difficult to identify which group to study and to ensure that it was entirely composed of children in the age range of interest for the study (0 to 36 months). Of the 404 groups that were selected and administered the quality measures, in 13.8% of cases, the group of children studied did not

fall entirely within the target age range. This is important to note because, as will be discussed later, the scores on instruments that measure process quality tend to be higher (indicative of higher quality) in groups of children with less age dispersion.

4. Quality Measures

The various instruments that were administered as part of this study to characterize the quality of child care services at the CIBVs are listed below:[8]

- Toddler Classroom Assessment Scoring System (CLASS)
- The Infant/Toddler Environment Rating Scale – Revised Edition (ITERS-R)
- The Child Care Infant/Toddler Home Observation for Measurement of the Environment (CC-IT-HOME)
- The Missouri Infant/Toddler Responsive Caregiving Checklist (MITRCC)
- The Infant-Toddler and Family Instrument (ITFI)
- The Knowledge of Infant Development Inventory (KIDI)
- The Teacher Practices Survey
- A questionnaire on center infrastructure, based on one used by INFA in years prior to the study

There were several reasons for choosing to administer multiple instruments that measure child care quality. First, different instruments take different approaches. For example, some instruments assess both structural and process aspects, while others focus solely on the latter. Another reason to administer multiple instruments was our interest in exploring the correlation among the various measures (i.e., their concurrent validity) in the context of developing countries, and more specifically, in Ecuador. To our knowledge, an analysis of this type that includes such a varied set of instruments has never before been performed. A third motivating factor was our interest in comparing the performance of instruments of more- and less-complex administration as it relates to the profile of the staff required to administer them, the observation time they require, and associated administration costs. This comparison informs the development of tools that allow for the continuous monitoring of the quality of CIBV services as part of the quality assurance process, with an emphasis on identifying areas that can be strengthened.

The main characteristics of the instruments administered and the distribution of scores resulting from the administration of each instrument to the centers in the CIBV sample are briefly discussed below. In each case, the scores are distributed among three groups of

---

[8] A detailed description of the first four instruments and their characteristics is presented in López-Boo et al. (2015).

equal size, in an attempt to characterize the range of quality as low, mid or high. A summary of the scores from all of the instruments administered is systematized in annex 1. In addition, annex 2 summarizes the measures of internal consistency.

Toddler CLASS

Toddler CLASS is an observational instrument developed to assess the interaction between a child and caregiver at a child care center. It focuses exclusively on process variables and is designed for groups of children between the ages of 15 and 36 months.

The Toddler CLASS explores eight dimensions of process quality: a) positive climate, b) negative climate, c) teacher sensitivity, d) regard for child perspectives, e) behavior guidance, f) facilitation of learning and development, g) quality of feedback, and h) language modeling. These dimensions, in turn, are grouped into two domains. Emotional and behavioral support, which encompasses the first five dimensions, is mainly focused on the expression of the adult's emotions—both positive and negative—toward the children, as well as the caregiver's responsiveness, availability, sensitivity, effort to take into account the children's perspectives, fostering of independence, and support for the development of behavior regulation. The second domain, engaged support for learning, is composed of the remaining three dimensions; it focuses on assessing the role of the caregiver as a facilitator of activities that guide children's learning and promote their cognitive and linguistic development, as well as her ability to provide feedback to the children and encourage their participation (La Paro, Hamre and Pianta, 2011).

The CLASS coder assigns a score of 1 to 7 for each of the dimensions. The higher the score, the better the quality of the processes observed in the classroom. More specifically, scores correspond to the following ranges of quality: 1 to 2, low; 3 to 5, mid; and 6 to 7, high.

In this study, CLASS administration was performed using classroom video footage of a four-hour day, from which four 20-minute segments were extracted. The field research team was solely responsible for shooting the videos that were subsequently evaluated by a team of certified CLASS coders. The video footage and the data from observational and survey instruments administered in the field were obtained from the same groups on the same days.

Of all of the instruments utilized in this study, it was the administration of CLASS that followed the most rigorous reliability protocol. The team of CLASS coders received intensive

training on the instrument. Specifically, CLASS coders completed eight days of training and practice, followed by seven days during which they worked on group coding exercises that enabled them to calibrate their coding. In addition, the coders were accompanied throughout the entire coding process by a certified CLASS trainer, with whom they performed daily group coding practice as part of a continuous training process. These exercises also served to ensure that inter-rater reliability was maintained during the process. Each video segment was coded twice by two different people who were randomly assigned to the task. If a mismatch occurred between the scores assigned by the two coders,[9] a third person was asked to evaluate the segment. In some cases, the trainer served as the third coder during group sessions. The score used in the analysis is the average of the two ratings with the smallest discrepancy between them.

Figure 2 shows the distribution of CLASS scores by domain. It can be seen that the mean score for emotional and behavioral support is 3.6 (SD=0.5)[10] on a 7-point scale (left-hand panel), i.e., mid-range quality. Virtually the entire distribution is concentrated in the middle third of possible scores for this instrument. For its part, not only does the mean score for engaged support for learning fall in the low range—1.6 (SD=0.4) on a 7-point scale (right-hand panel)—but almost the entire distribution is concentrated in the bottom third of possible scores for this instrument.

Figure 2. Scoring Distribution for the Domains of Toddler CLASS
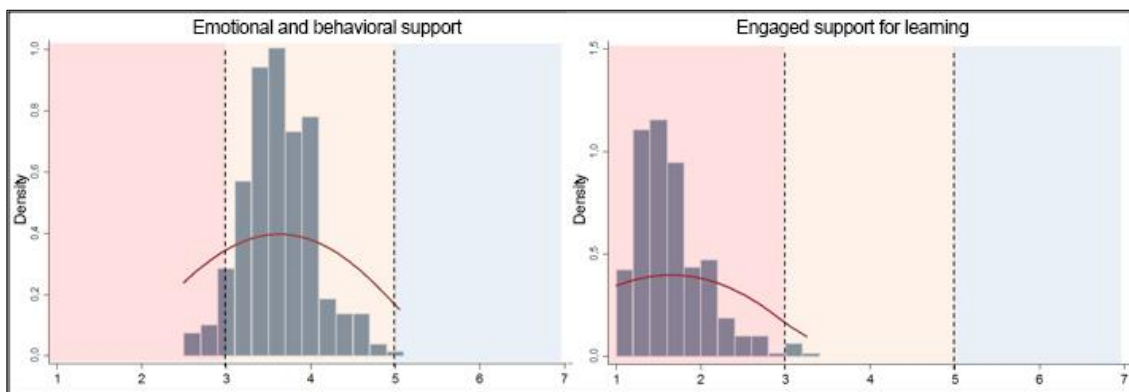
Figure 3 shows the results for the dimensions that make up the emotional and behavioral support domain. The mean scores were as follows: 3.3 (SD=0.6) for positive climate (mid-range quality); 6.6 (SD=0.4) for negative climate (high quality); 3.4 (SD=0.6) for

---

[9] The discrepancy threshold for a third coding was two points for dimensions with less variability (negative climate, regard for child perspectives, quality of feedback, and language modeling) and three points for dimensions with more variability (positive climate, teacher sensitivity, behavior guidance, facilitation of learning and development).
[10] From here onward, SD is used to refer to the standard deviation of each score.

teacher sensitivity (mid-range quality); 2.0 (SD=0.3) for regard for child perspectives (low quality); and 2.9 (SD=0.5) for behavior guidance (low quality, close to mid-range).

It is noted that the dimensions of negative climate and regard for child perspectives display a high concentration of high and low scores, respectively, with little variability in the rest of the area of distribution. In other words, 91.8% of centers had a score of 6 or more for negative climate (suggesting that negative climate is not an issue at most of these centers, since the higher the score, the lower the negative climate in the group), while 98.8% had a score of 3 or less for regard for child perspectives, providing evidence that this dimension is practically absent in the interactions observed at the CIBVs. The graphs in figure 3 show that the scores for the other three dimensions have a higher level of dispersion.

Figure 3. Scores for the Dimensions of Emotional and Behavioral Support
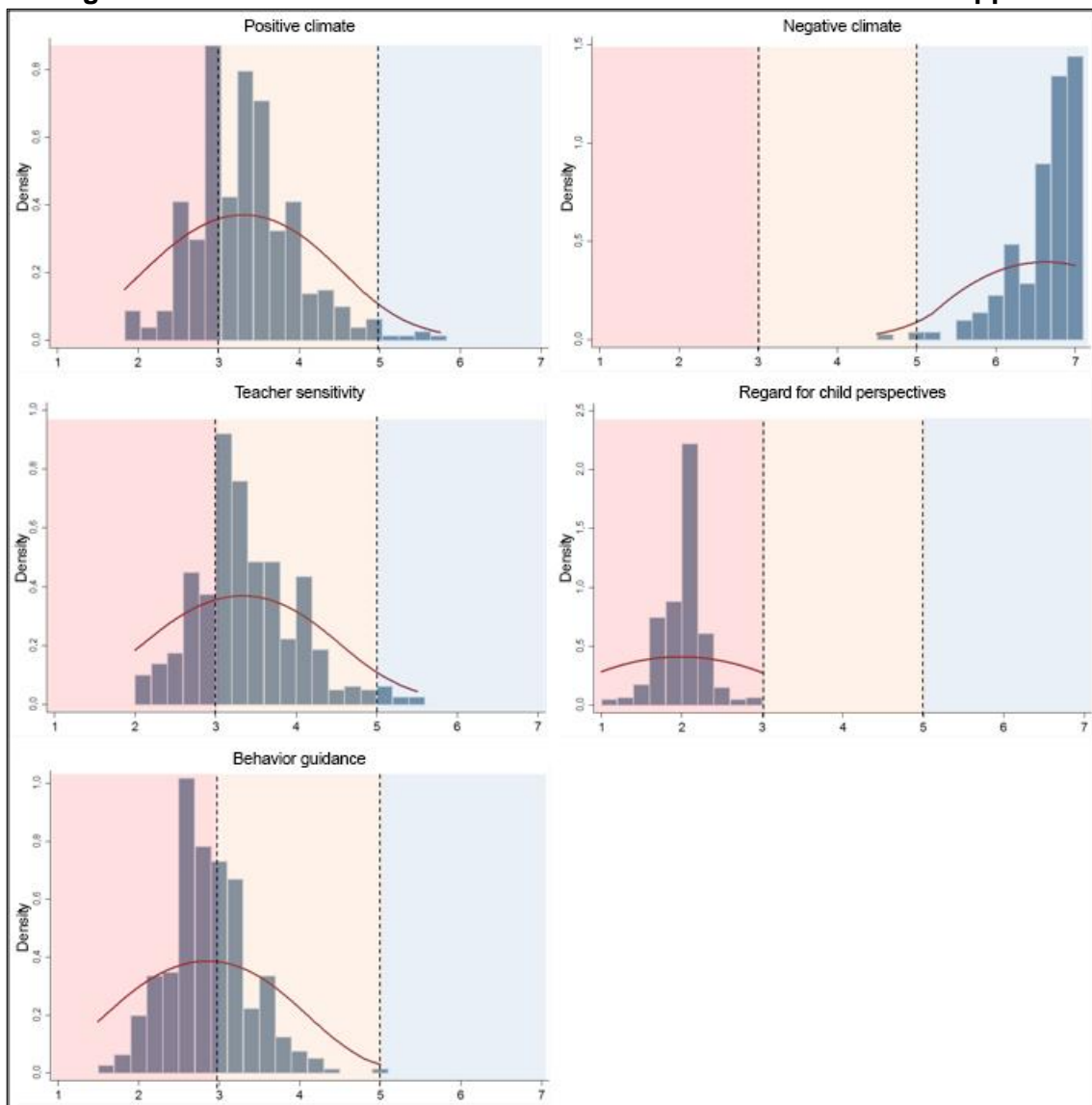
Figure 4 presents the distribution of Toddler CLASS scores for the dimensions of the engaged support for learning domain. The scoring distribution for the three dimensions is concentrated in the bottom third of possible scores. The mean scores were 2.1 (SD=0.5) for facilitation of learning and development, 1.3 (SD=0.3) for quality of feedback, and 1.6 (SD=0.5) for language modeling. The CIBV mean falls within the low-quality range on all three subscales. Given the large gaps in cognitive development among children of lower socioeconomic status that have been documented in Ecuador and that widen after the age of 3 (Schady et al., 2015), the results of these dimensions from Toddler CLASS are particularly significant. For example, such a low score on the dimension of language modeling, critical to the acquisition and development of language that occur precisely in the first 3 years of life, suggests that the CIBVs face an enormous challenge in this area.

Another finding from the analysis of the scoring distribution on Toddler CLASS for the CIBVs is that even better-quality centers (those in the 90th percentile of the scoring distribution) barely attain mid-range quality on the dimensions of the emotional and behavioral support domain, and they achieve low levels of quality for the dimensions of the engaged support for learning domain. This suggests a systemic lack of best practices related to process quality, making the task of developing these types of skills in the child care staff at daycare centers even more complex.

It is worth mentioning that, as shown in figure 1, groups in which the majority of children were under 15 months of age were rare in this sample; however, in those few groups with a significant number of children in that age group, the coding of this version of CLASS proved more difficult. This is due to the fact that, as previously mentioned, the instrument is designed for groups of children between the ages of 15 and 36 months old. Fortunately, less than 5% of the groups analyzed were composed of children under 15 months of age.

Figure 4. Scores for the Dimensions of Engaged Support for Learning
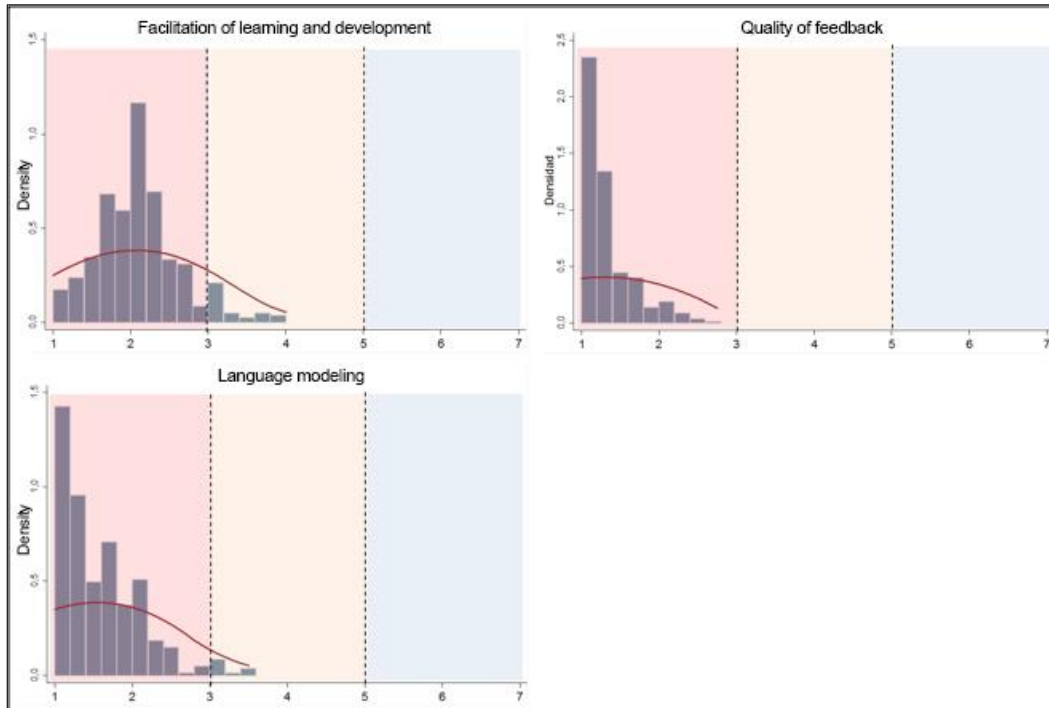
One question that arose when analyzing these scores was how to compare them to those of other Toddler CLASS assessments performed at child care centers elsewhere. This instrument was administered in Peru as part of an impact evaluation of the improvements to the daycare centers in the national program *Cuna Más*, a service similar to that of the CIBVs, which serves children between the ages of 6 and 36 months. At the time of measurement, the presence of older children was also identified. Toddler CLASS scores from Peruvian child care centers are of a similar magnitude to those from Ecuador's CIBVs (Berlinski and Schady, 2015), confirming that poor process quality is a shared feature of public child care services for infants and toddlers in the two countries. Another point of comparison is offered by the use of the instrument in the United States, as part of an evaluation of child care services provided by the national Early Head Start program, aimed at children under the age of 3. Consistent with the observations made in Ecuador and Peru, the greatest challenges in the United States in terms of service quality for this age group also lie in the area of engaged support for learning. In the U.S. context, it is observed that Toddler CLASS scores fall in somewhat higher ranges, specifically in the upper mid-range for emotional and behavioral support and the lower mid-range for engaged support for learning (Vogel et al., 2015).

ITERS-R

The ITERS-R is an instrument that measures process quality in the classroom by documenting aspects related to the quality of interactions (between adults, between adults

and children, and between children and activities in the environment) as well as the quality of the resources that support those interactions, such as spaces, routines, materials, etc. (Cryer et al., 2004). The ITERS-R combines observation items with interview items that are administered to center staff. This instrument consists of 39 items, spanning seven dimensions: a) space and furnishings, b) personal care routines, c) listening and talking, d) activities, e) interaction, f) program structure, and g) parents and staff. The ITERS-R has probably been administered more than any other instrument to measure the quality of child care for infants and toddlers in Latin America and other countries around the world. In the region of Latin America and the Caribbean, it has been administered in Bolivia, Brazil, Chile, Colombia, Mexico and Peru.[11]

Each of the ITERS-R subscales is scored on a 7-point scale, with quality indicators rated as (1-2) inadequate, (3-4) minimal, (5-6) good, or (7) excellent.

The ITERS-R requires a minimum of three and a half hours of direct classroom observation and an interview with the observed staff member. Given the complexity of the instrument, the ITERS-R was administered by the most qualified researcher, with post-secondary education in the field of psychology or education.

The ITERS-R is designed to be administered gradually, meaning that once a classroom fails to comply with a subset of consecutive indicators within a given item, it receives the score corresponding to the highest indicator attained and the observer moves to the next item, without checking if the classroom complies with any of the indicators *above* that level of quality. In other words, the ITERS-R protocol does not require evaluation of the indicators above this stop point. Nevertheless, since the ITERS-R had never before been administered in Ecuador, for the present study it was decided to score all indicators at all levels, ignoring the stop rule dictated by the instrument's guidelines, in order to test to what extent the graduality assumptions implicit in the scale's design fit with the reality of Ecuadorian centers. Our preliminary analysis shows that, by dispensing with the stop rule, valuable information was recovered about the quality of the centers that would have otherwise been lost. Specifically, ITERS-R scores calculated without respecting the instrument's stop rules are compared with those that would have resulted if the stop rules had been applied, resulting in the identification of individual items for which it is possible to acquire additional information about center quality when scoring the instrument with the alternate protocol.

The total ITERS-R score for the CIBVs (calculated as the average of the scores for the seven subscales) reveals inadequate levels of quality, with a mean of 2.1 (SD=0.5). Even when including the better-quality centers in the 90th percentile of the distribution, a

---

[11] In some cases, the ITERS-R was administered together with another instrument from the same family, the ECERS-R, which has a similar structure and is used in child care settings for slightly older children, ages 30 to 60 months.

minimal level of quality is barely achieved (an ITERS-R score of 3.1). Figure 5 shows the scoring distribution for the ITERS-R by subscale.

A glance at the ITERS-R subscales reveals that, across the board, quality levels are consistent with what is described in the previous paragraph. With just one exception, the mean scores for all of the subscales are clustered in the inadequate quality range. The mean scores on the ITERS-R subscales are 2.1 (SD=0.6) for space and furnishings, 1.7 (SD=0.5) for personal care routines, 2.5 (SD=1, 2) for activities, 1.5 (SD=0.5) for listening and talking, 3.3 (SD=1.3) for interaction, 2.6 (SD=1.3) for program structure, and 2.0 (SD=0.6) for parents and staff. Three of the ITERS-R subscales—personal care routines, activities, and parents and staff—are surprising not only in that the mean scores are so low but also for the fact that not even the best centers in the sample (those in the 90th percentile of the distribution) manage a score of 3, placing them within the minimal quality range as well. For the other three subscales—space and furnishings, listening and talking, and program structure—the distributions show a slightly greater dispersion, and the centers in the 90th percentile obtained scores of between 3 and 4.3, i.e., they achieved somewhat better quality levels, although still within the minimal quality range as interpreted by this instrument.

The interaction subscale is the only one that achieves a mean score beyond the range of inadequate quality and within the group of minimal quality levels. It is also the only subscale for which the best centers in the sample (those in the 90th percentile of the distribution) achieve a score of 5, a level of quality considered good. The results from the interaction subscale are counterintuitive if they are compared to those reported in the previous section on the CLASS, an instrument that focuses exclusively on process quality. Compared to other ITERS-R subscales, the interaction subscale entails complex coding. Since Toddler CLASS is administered using a higher standard of quality than the ITERS-R (double coding) and focuses exclusively on processes, we feel that it probably reflects this aspect of quality more accurately.

Figure 5. Scoring Distribution for the ITERS-R by Subscale
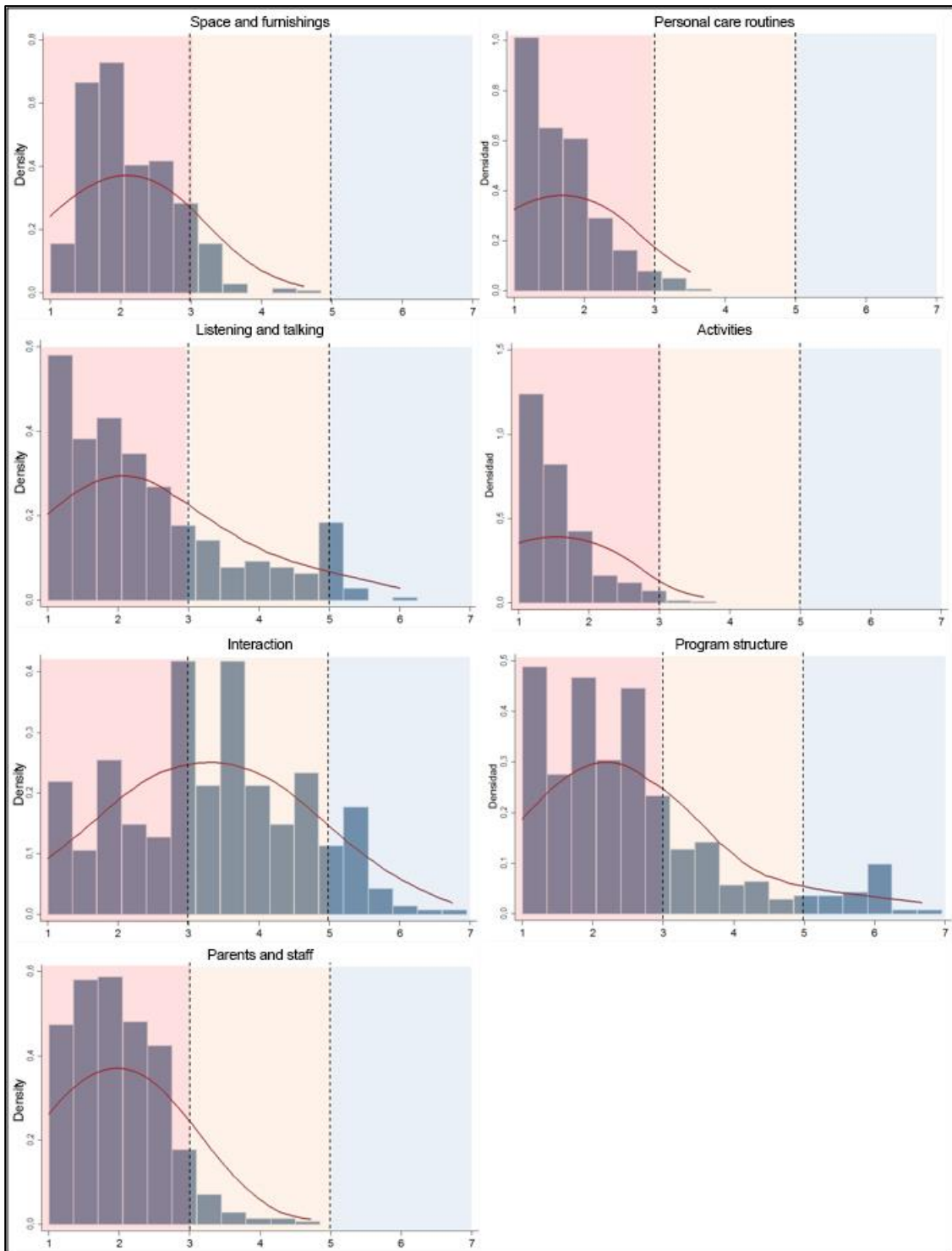
Since ITERS-R has been administered in several countries within the region, it is possible to compare the scores of Ecuador's CIBVs with those of other public child care programs for infants and toddlers that have characteristics similar to the Ecuadorian program. In general, CIBV scores are similar to those observed in the Colombian Family Welfare Institute's programs *Centros de Desarrollo Infantil* and *Hogares Infantiles*. With

regard to the data available for Colombia, measurements were collected in classrooms that serve older children and children under the age of 2. In general, it was found that the scores for classrooms with children 2 years and older were slightly lower than those for classrooms with younger children. CIBV scores are slightly lower than those of the daycare centers in Peru's *Cuna Más* program (which serves children ages 6 to 36 months) but higher than those of Bolivia's *Crecer Bien para Vivir Bien* program (which mostly provides care to children ages 0 to 3, and where, as in the case of *Cuna Más*, the presence of slightly older children is also sometimes observed in classrooms, Berlinski and Schady, 2015).

HOME

HOME is a quality measure that primarily focuses on process variables, although it does also evaluate some structural variables. A minimum of one hour for observation and reporting is required for the administration of this instrument. The HOME is broken into six subscales: a) learning materials, b) caregiver responsivity, c) acceptance, d) organization, e) caregiver involvement, and f) variety of stimulation (Bradley, Caldwell and Corwyn, 2003). Each item is scored using a binary format (no=0, yes=1), and the resulting subscale rating is the sum of the individual item scores. The ratings for the different subscales are combined to obtain a maximum total score of 42.

For this study, the researchers administered a version of HOME designed for children ages 0 to 3 that was translated to Spanish and adapted for use at child care centers.[12] Like the ITERS-R, the HOME is an instrument that combines observation with an interview; however, in contrast to the ITERS-R and Toddler CLASS, the unit of analysis for the adapted version of HOME is not the group (i.e., the caregiver and her group of children). Instead, the HOME evaluates the quality of the relationship between a specific child within the group and the caregiver.

The focal child, or the child to be observed for this study, was randomly selected without revealing to the caregiver that her interactions with that particular child were being evaluated. Specifically, the observer was asked to list the names of all of the children in the study group (in any order) and to choose the fifth name on the list. As with the ITERS-R, the HOME was administered by the most-qualified researcher, with college-level training in psychology or early childhood education. Ninety-nine percent of the children administered the HOME were under 36 months of age, i.e., they fell within the appropriate age range for this instrument.

Figure 6 shows the scoring distribution for the HOME by subscale. The mean scores on the various HOME subscales were as follows: 4.5 out of 9.0 (SD=2.1) for learning

---

[12] Specifically, one of the items from the acceptance subscale (item 18) was removed, as it was not applicable to the context of the CIBVs.

materials, 6.7 out of 11.0 (SD=2.5) for caregiver responsivity, 5.7 out of 6.0 (SD=0.6) for acceptance, 2.8 out of 6.0 (SD=1.0) for organization, 3.6 out of 6.0 (SD=1.8) for caregiver involvement, and 1.4 out of 4.0 (SD=0.8) for variety of stimulation. The mean score for the acceptance subscale was higher than for the other subscales, falling in the top third of the quality distribution. The acceptance subscale evaluates whether the caregiver accepts the child's behaviors with patience and tolerance (even when those behaviors involve noise or clutter in the environment) and to what extent she manages to avoid using unnecessary restrictions or punishment in the interaction.

Figure 6. Scoring Distribution for HOME
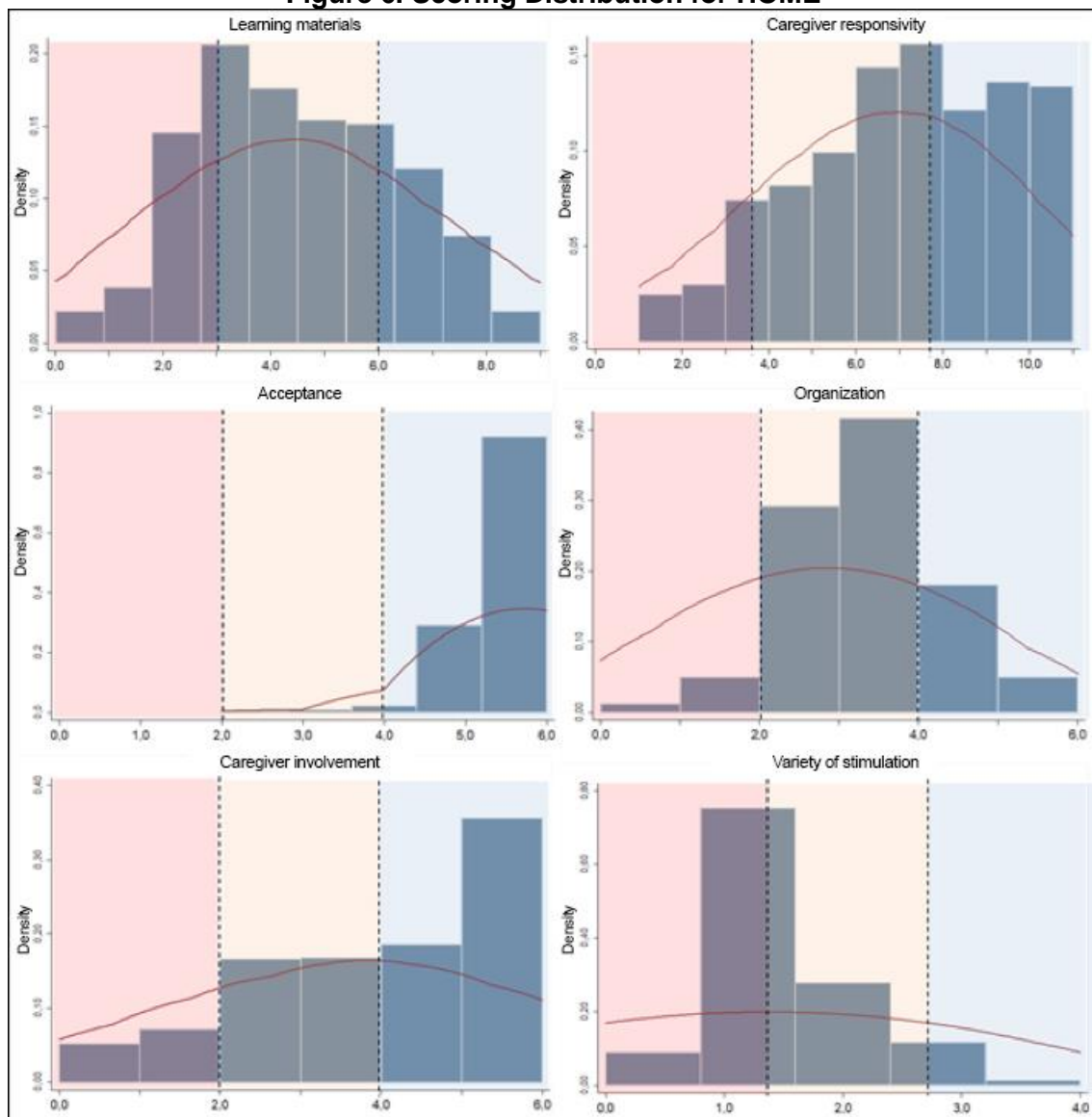
The mean scores on four of the other five HOME subscales were concentrated in the middle third of the quality distribution. The only exception is the variety of stimulation subscale, which stands out for its concentration of scores in the bottom third of the distribution. This subscale assesses the inclusion of people or events in the daily routine that

create variety without disrupting the child's life. These results suggest that children's routines at the CIBVs have little in the way of variety.

MITRCC

The MITRCC is an observational instrument (MU Center for Family Policy & Research, 2003), whose administration requires a minimum of three hours in the classroom. Although it primarily focuses on process quality, it also measures some structural elements. The observation period is expected to afford each child a sufficient amount of time to express a sense of well-being, belonging, exploration and communication. It is also expected that children will have had a chance to engage in a wide variety of interactions and experiences during the period of observation. This instrument focuses on assessing whether a certain percentage of children in the group are engaging in desirable interactions and activities. Each item is scored using a binary format, with a score of one assigned if the attribute is observed in at least 75% of children in the group and zero if it is not observed.

For this study, the original instrument was translated and culturally-adapted for use in the Ecuadorian context.[13] The maximum total score that can be obtained on this instrument is 20. Unlike the ITERS-R and HOME, the MITRCC was administered to the same group of children studied and the caregiver by the researcher-interviewer.[14]

Figure 7 shows the scoring distribution for the MITRCC. One can appreciate that the scores are quite evenly distributed in the bottom two thirds, with a somewhat higher concentration of values around the mean of 8.76 out of 20 (SD=4.7). Scores greater than 16 occur less frequently.

---

[13] Specifically, six items that were not applicable to the context of the CIBVs were removed. Additionally, another item was disaggregated to facilitate its administration.
[14] From a quality-monitoring perspective, the question arises of just how much impact the educational profile of the observer has on her assessment of quality. Specifically, a key question with regard to the cost-effectiveness of quality monitoring is if it is necessary to hire observers with a higher level of education to more accurately capture important aspects of quality. Tomé and López-Boo (2015) discuss this issue and conclude that there are systematic differences between the scores assigned by observers with different educational backgrounds. On average, less educated observers tend to assign higher scores (indicative of higher quality), particularly on items involving more complex observation.

Figure 7. Scoring Distribution for the MITRCC
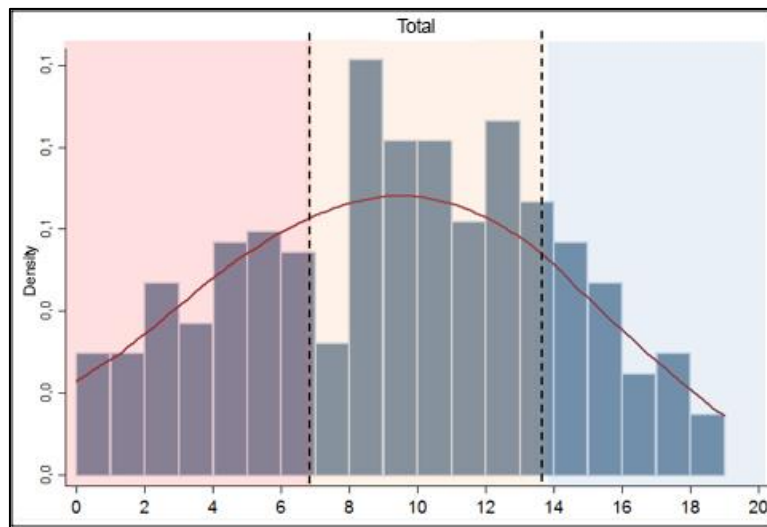
ITFI

The ITFI focuses on process variables and serves as a screening tool to identify poor-quality child care environments that require intervention (Provence and Apfel, 2001). It combines interviewing and observation. The observation focuses on the reactions exhibited by the adult caregiver when questioned about specific topics. As with the HOME, the questions on this instrument are asked in reference to a specific child in the group. In the U.S., the ITFI is used by social workers to evaluate the child care environment for children ages 6 to 36 months old.

For this study, 21 items were selected from the IFTI, three of which were further divided into sub-items. Each item and sub-item receives a score of one if answered satisfactorily, with a maximum score of 34 for the entire scale. The main goal of the selected items is to assess the extent to which the caregiver knows the child, his interests and his family.

The ITFI was administered by the researcher-interviewer. The questions on this instrument are administered in reference to a *different* child from the one observed for the HOME. To ensure random selection of the child to be studied, and in light of the fact that the researchers had already spent several hours in the classroom by the time the ITFI was administered, they were instructed to administer the ITFI on the basis of the last child to have his diaper changed or to have gone to the bathroom prior to the start of the administration of the questionnaire. All of the children selected for the ITFI fell within the age range for this questionnaire.

As seen in figure 8, the ITFI scores display a relatively wide range of variability and are skewed to the left. In other words, the distribution is concentrated in the top two thirds of the range of possible scores. The mean score is 23.6 out of a total of 34 (SD=4.4). This shows that at most of the centers visited, the caregivers seem to know the children in their care and their families. This finding is consistent with our expectations, since the staff members working at the CIBVs are usually from the same neighborhood or community where the center is located and where the families that use the service reside. Interviews with the caregivers revealed that 35% have worked three years or more at that center.

Figure 8. Scoring Distribution for the ITFI
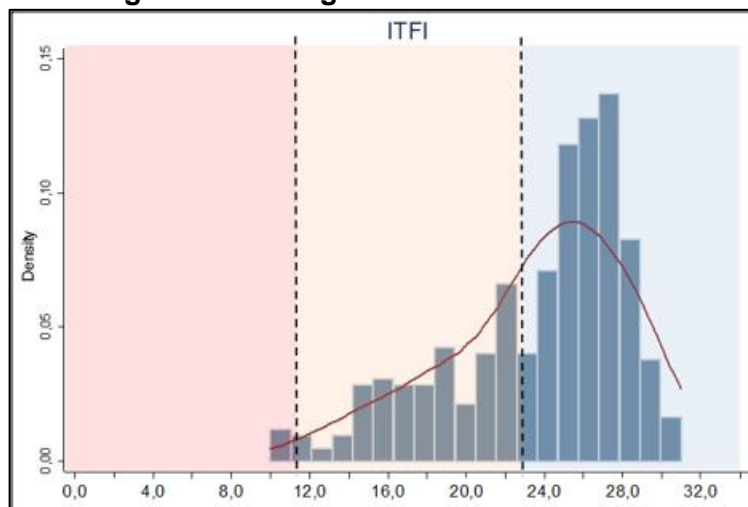
KIDI

The KIDI consists of 58 items that evaluate an individual's knowledge of child development processes and norms for children ages 0 to 5 (MacPhee, 2002). The respondent can choose from three possible answers: agree, disagree, and not sure. The respondent receives one point for each correct answer. The total score is calculated as the sum of correct answers for all 58 questions on the scale. The researcher-interviewer administered this inventory to the caregiver of the group of children studied and the center coordinator.

Figure 9. Scoring Distribution for the KIDI
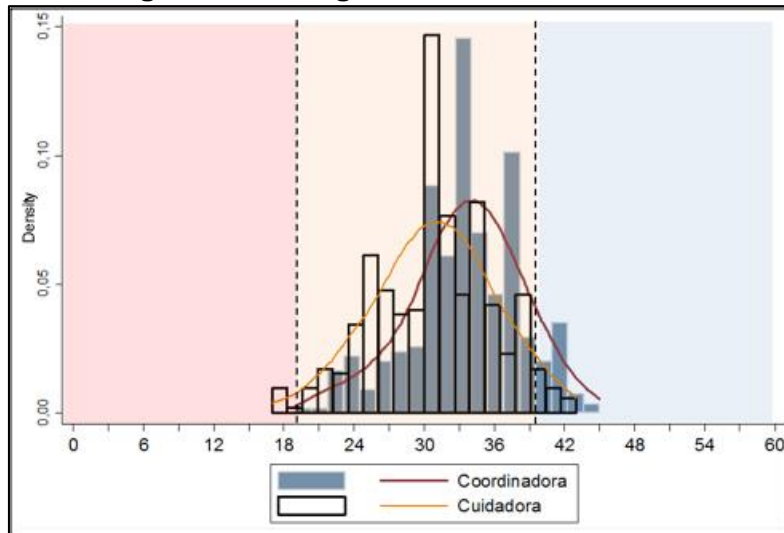
Figure 9 shows the scoring distribution for the KIDI for both caregivers and coordinators. Almost all of the KIDI scores for caregivers and coordinators are concentrated in the middle third of the distribution; the number of scores falling in the top third of the distribution is minuscule. The mean score on the KIDI for a caregiver, 30.8 (SD=4.9), is about three points below that of a coordinator, 33.5 (SD=4.6). Although small in magnitude, the difference is statistically significant at 99%. As a point of reference, if the respondents had randomly selected their answers, one could expect an average of 29 correct responses. In other words, the caregivers' and coordinators' scores are so low that they only exceed that level by a few points, despite the fact that, as shown in table 4, the coordinators have a higher level of education than the caregivers. Consistent with CIBV guidelines and reforms implemented in early 2012, coordinators have a college education, while caregivers are only required to have completed high school (a requirement which, on average, is not met).

Table 4. Education Level and Knowledge of Child Development

	Coordinators	Caregivers	Diff. (t-test)
Average score on KIDI	33.5	30.8	***
Years of education	16.3	11.1	***

Significant differences at *** 99%

Teacher Practices Survey

The Teacher Practices Survey is based on an instrument designed in the U.S. to operationalize the guidelines of the National Association for the Education of Young Children or NAEYC (Kyung-Ran, 2005). The original instrument focuses not only on teacher practices but also beliefs. In this study, the researcher-interviewer administered the survey through an interview with the caregiver in charge of the study group. The interview asks questions about the frequency with which the caregiver engages the children in activities from a list appropriate for three- to five-year-olds. It is important to note that since this scale was designed for use in classrooms of older children, some adjustments were necessary in order to use it in this study. The original scale consisted of 30 items, but in this case, only 23 items were administered, omitting those that were clearly focused on school readiness activities and, therefore, too advanced for the children who form the core of this study.[15] The items that were administered focused on activities relevant to children 18 months and older, e.g., building with blocks, singing and dancing, tracing or drawing, playing with puzzles, or playing games that allow them to show affection.

The scale offers four response options representing four different frequencies. The total score is derived from the sum of the frequencies, with 1 = almost never, 2 = sometimes, 3 = regularly, and 4 = always. The minimum and maximum scores that could be obtained on the questionnaire were 23 and 92, respectively.

Figure 10 shows the scoring distribution for the Teacher Practices Survey. The mean score was 54.5 (SD=7.7), with the highest concentration of scores around the mean. This suggests that, on a rather frequent basis, caregivers for the groups studied conducted learning activities from the scale that are not necessarily appropriate for the youngest children in the age range of this study.

---

[15] Specifically, items 12, 13, 15, 26 and 30 were eliminated, which included activities such as filling in worksheets, practicing handwriting, using flashcards with letters or words, solving math problems, and completing activities involving multiple subjects. Items 10 and 20 were also eliminated, as they were not considered relevant to the CIBV context during pilot testing. These items made reference to the use of purchased phonics materials, such as audio books, and the use of incentives or prizes to encourage children's participation.

27

Figure 10. Scoring Distribution for the Teacher Practices Survey
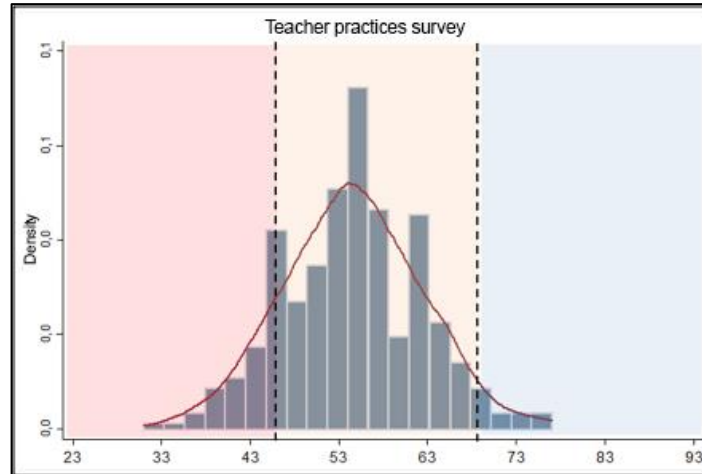
Structural Quality Questionnaire

Lastly, as a supplement to the data collected through the aforementioned instruments, the supervisor who accompanied the two researchers on their visits to the center administered a detailed questionnaire on the availability and condition of the center's physical infrastructure as well as other structural variables associated with the quality of the child care service. The questionnaire was designed by the researchers, incorporating elements of the monitoring worksheet from INFA and other similar services in the region. In order to aggregate the variables collected in this questionnaire, a structural quality index (table 5) was constructed. The index takes into account infrastructure variables including center access to clean drinking water, a sewage system, metered electricity, good natural lighting, no broken windows, adequate furnishings for the number of children, and the presence of a dining area, classroom, storage area, fencing and a safe play area for the children. Other variables included in the index are related to characteristics of the service offered at the center, the coordinators and the caregivers. This subgroup includes the percentage of children waiting to enroll due to a lack of available slots, the center's child-caregiver ratio, the caregiver's and coordinator's years of experience working with children under 60 months, the percentage of coordinators with child care duties, and the caregiver's education level.

Table 5: Variables Considered in the Structural Quality Index

	Obs.	Mean	SD
Percentage of centers with infrastructure			
Drinking water	403	0.56	0.50
Sewage system	403	0.59	0.49
Electricity, with meter	400	0.63	0.48
Good natural lighting*	403	0.74	0.44
No broken windows	403	0.90	0.30
Adequate furnishings for the children	403	0.65	0.48
Dining space	403	0.71	0.45
Nursery space	403	0.66	0.47
Storage space	403	0.37	0.48
Fencing	403	0.65	0.48
Safe playground equipment	403	0.46	0.50
Characteristics of the service and its staff			
% of children waiting to enroll but no slots available	403	90.01	16.28
Child-caregiver ratio at center	403	9.46	1.80
Coordinator			
Years of experience working with children < 60 months old	403	3.32	3.74
Does not have child care duties	399	0.15	0.35
Group caregiver			
Years of experience working with children < 60 months old	403	2.93	3.95
Years of education	403	11.10	2.81

*Natural lighting is considered "good" when windows allow plenty of natural light to enter.

Figure 11 shows the distribution for the structural quality index. The index was constructed using principal component analysis (see annex 3 for more information on the construction of the index). The index is normalized to have a mean of 0 and a standard deviation of 1. It can be observed that the centers on the median of the quality distribution are worse than the average, located 0.2 standard deviations to the left of the mean. If the range of feasible values is divided into thirds, as was done for the other quality measures, it is noted that the greatest number of centers for the sample is found in the bottom third, i.e., in the lowest levels of structural quality.

Figure 11. Distribution for the Structural Quality Index
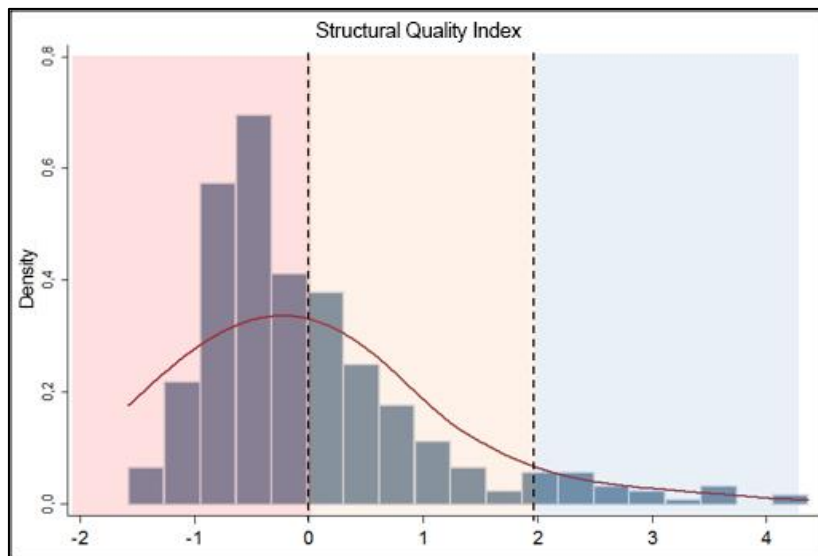
5. Analysis

Consistency among quality measures

Two types of internal consistency indicators were calculated for each instrument in the study that included scales (Toddler CLASS, ITERS-R and HOME). In the psychometric literature, it is standard practice to report the internal consistency, with a view to documenting the quality of the data collected, particularly when working with multi-item scales (McCrae et al., 2011). The two measures of internal consistency that are reported in annex 2 are a) the correlation between each scale and the total score of the instrument and b) the Cronbach alphas. The correlation between each scale and the total score of the instrument indicates the extent to which each of the instrument's *scales* is associated with the total quality measurement captured by that instrument. The Cronbach alphas represent the degree to which the *items* within each scale appear to measure the same concept. Generally, the literature considers a Cronbach alpha in the range of 0.6 to 0.7 to be a reasonable measure of internal consistency.

Table A2.1 shows the internal consistency of the Toddler CLASS. The Cronbach alphas for the two domains are both greater than 0.85 and the value for the full scale is 0.91, demonstrating that there is good consistency. Table A2.2 shows the internal consistency of the ITERS-R. The Cronbach alpha is close to or greater than 0.6 for all of the subscales except personal care routines. As with the CLASS, the internal consistency of the full ITERS-R scale is high at 0.87. The correlation coefficients between scales and the total score of

ITERS-R are also high. Lastly, the internal consistency of the HOME is shown in Table A2.3. The internal consistency levels for the HOME's subscales are variable; for the full scale, the value is similar to that of the ITERS-R.

Table 6 summarizes the correlation among the total scores of the main instruments administered.[16] It is worth repeating that each quality measure focuses on different aspects and variables, so they are not expected to perfectly correlate with one another. In fact, table 6 shows that the only pair of instruments that exhibits a high level of correlation is the ITERS-HOME. For its part, the MITRCC is moderately correlated with the ITERS and the HOME. This finding stands out since these are the three instruments that combine both structural and process aspects of quality measurement. The correlation among the ITFI and the three previously mentioned instruments is low. For their part, the levels of correlation among the KIDI, the Teacher Practices Survey, the Structural Quality Index and all of the other quality measures are very low.

One aspect worth mentioning is the correlation between the Toddler CLASS and the other instruments administered. As already discussed, the Toddler CLASS focuses exclusively on process quality. In addition, it is the most costly measurement tool and the one with the highest standards of reliability. It stands out that the levels of correlation for the CLASS fluctuate between low (with the ITERS-R, HOME, MITRCC and Teacher Practices Survey) and very low (with the other measures). This finding suggests that, at the aggregate level, the less costly instruments are not capturing the important process quality aspects that the CLASS measures. This partly explains why CLASS is the only instrument that exclusively looks at issues related to process quality. Nonetheless, with a view toward being able to operationalize indicators that can be used at scale and whose administration does not involve inordinate costs, the ideal would be to identify subscales or specific items that are highly correlated with the quality measured by CLASS. An accompanying study presents a more careful analysis of the correlations among the scores on the full scales, their dimensions—particularly those that measure overlapping aspects—and even specific items.

For reference purposes, annex 1 consolidates the scores from all of the instruments discussed in this paper.

---

[16] A correlation is considered very high in the range of 0.80 to 1, high in the range of 0.6 to 0.8, moderate in the range of 0.4 to 0.6, low in the range of 0.2 to 0.4, and very low when less than 0.2.

Using the same data, Berlinski and Schady (2015) present a multivariate analysis to explore whether there are conditional associations (a causal interpretation) between the two main quality measures analyzed in this study (Toddler CLASS and ITERS-R) and some structural variables typically used as proxies for child care service quality (caregivers' level of education and experience and child-caregiver ratios).

This analysis shows that only some structural variables are associated with better process quality. For example, Toddler CLASS and ITERS-R scores are between 0.26 and 0.30 standard deviations higher in those groups where the caregiver has completed high school (compared to those in which she has not attained this level of education). In addition, ITERS-R scores are lower in groups with higher child-caregiver ratios; with each additional child, scores fall by 0.05 standard deviations. Interestingly, this association was not observed between CLASS scores and child-caregiver ratios. Lastly, the caregiver's years of experience working with children are not associated with a significant improvement in quality.

Table 6. Correlations across Quality Measures

	CLASS	ITERS-R	HOME	MITRCC	ITFI
CLASS	1				
ITERS-R	0.3351***	1			
HOME	0.3551***	0.8005***	1		
MITRCC	0.3160***	0.4894***	0.5338***	1	
ITFI	0.1662***	0.3654***	0.3780***	0.3708***	1
KIDI - Coordinator	0.1075**	0.0944*	0.0401	0.0405	0.0486
KIDI - Caregiver	0.1108**	0.1239***	0.1956***	0.1103**	0.0144
Teacher Practices	0.3329***	0.1701***	0.1582***	0.1700***	0.0380
Structural quality	0.1056**	0.2368***	0.1837***	0.1711***	0.1710***

	KIDI - Coordinator	KIDI - Caregiver	Teacher practices	Structural quality
KIDI - Coordinator	1			
KIDI - Caregiver	0.2535***	1		
Teacher Practices	0.1163***	0.0168	1	
Structural quality	0.1519***	0.0572	0.1560***	1

Efforts to improve CIBV quality

As mentioned above, during the data collection period for this study and in subsequent years, the MIES mobilized funding and undertook a series of steps to improve CIBV service quality. In this section, we focus on three of these initiatives: the hiring of coordinators with post-secondary education; the inspection of centers and subsequent termination of agreements with those that failed to meet standards; and the transfer of children over the age of 3 to early childhood education services.

The first area of improvement to be discussed is the hiring of staff with post-secondary education to serve as coordinators at each of the centers. Coordinators are responsible for educational activities, but they also perform administrative and maintenance tasks at the centers. It is noteworthy that, at the time of data collection, the hiring of professionals as coordinators had been implemented at almost all of the centers studied. In 2012, 88% of the coordinators at the centers researched already met this requirement, and 99% had at least one year of post-secondary education. Nevertheless, and despite the fact that the CIBVs largely employ coordinators with post-secondary education, service quality ranks consistently low. What is also surprising is that the scores obtained by coordinators on the KIDI are low and barely higher than those of their peers, the caregivers, whose position requires a lower level of education. These findings point to the need to strengthen the selection and training (initial and ongoing) of all staff members—both coordinators and caregivers alike—in order to provide them with relevant, specific knowledge and skills.

A second measure implemented as part of efforts to improve quality focuses on inspections by MIES personnel to ensure compliance with a series of standards, especially in the areas of infrastructure and safety. As a result of these inspections, agreements with some of the centers' implementing units have been suspended. The suspension of these agreements means that the MIES no longer funds the operation of the centers, most likely leading to their closure; however, the suspension of an agreement does not necessarily mean that a center will shut down, because it might obtain other sources of funding that will allow it to continue operating.

Annex 4 presents a comparison of the quality measures between those centers that were still operating under an agreement with the MIES three years after data collection (in April 2015) and those that were no longer listed in the administrative records. It is worth noting that when attempting to merge the 2012 sample with MIES administrative records for the centers in April 2015, 238 of the 404 schools in the sample (about 60%) were no longer

33

found. Authorities attribute the absence of records to the fact that these centers no longer have an agreement with the MIES.

The centers that in 2015 still have an agreement with the MIES were in 2012, on average, the largest in terms of the number of children served (37 vs. 24 children), had been operating longer (11.4 vs. 9.6 years), and were more frequently located in urban areas (57% vs. 30% of the centers). In addition, centers that remain under agreement with the MIES are those that, in 2012, organized their groups in such a way as to achieve a smaller age gap between the children. They also more frequently charged families a co-payment for use of the service (55% vs. 44%).

In terms of service quality, the centers whose agreements with the MIES had terminated in 2015 were those that in 2012 had, on average, lower levels of quality. Although these quality differences are of modest magnitude, they are statistically significant. They are consistent in terms of variables that measure process quality as well as those that focus on structural aspects. The scores on the Toddler CLASS, ITERS-R, HOME, MITRCC and Structural Quality Index are, on average, slightly higher for centers that remain under agreement with the MIES; however, these same centers continue to provide poor quality care and, therefore, have significant room for improvement in terms of their services.

The third improvement to be discussed here consists of an effort to consolidate the focus of the CIBVs, serving only children under the age of 3, while shifting the responsibility for the remaining early childhood education services to the Ministry of Education. In principle, the direct result of this action would be that the CIBVs could focus their service on infants and toddlers and ensure more homogeneous groups in terms of age composition. The data suggests that this change could result in an improvement in service quality. Annex 5 presents quality differences between groups with greater and lesser homogeneity in terms of age composition. More specifically, the chart compares quality levels in groups of children with an age gap of up to 12 months and groups with age gaps of 12 months or more. Service quality, process quality and structural quality are slightly higher in those centers with smaller age gaps between the children receiving the service. These differences are not always statistically significant.

Although it appears that the proposed reforms are a step in the right direction in terms of improving quality, increasing the quality of this type of service is a complex, long-term process; therefore, it is essential to continue and intensify these efforts, identifying

quality improvement actions that have the potential for the greatest impact on the welfare of children and the quality of service.

6. Conclusions

In this study, we present an exhaustive analysis of different aspects of the quality of a sample of 404 CIBVs, which provide child care services to Ecuadorian children. We used instruments of varying methodological complexity to measure elements of both structural and process quality. All of the quality measures studied consistently agreed that the child care services provided by the CIBVs have ample room for improvement, in terms of structural aspects and, in particular, dimensions of process quality. Several of the most rigorous and commonly applied measures in these types of studies, such as the Toddler CLASS and ITERS-R, describe the scores of nearly all the centers studied as concentrated in the lower levels of the quality distribution. The poor performance of both caregivers and center coordinators on the KIDI, a scale that measures knowledge of child development, also stands out. The quality challenges identified at Ecuador's CIBVs are consistent with those of services in other countries throughout the region, where these types of scales were administered (Peru, in the case of the CLASS; Bolivia, Brazil, Colombia, and Peru, in the case of the ITERS, Berlinksi and Schady, 2015).

In addition, we found that the correlations among the total scores of the instruments for measuring child care quality were low, particularly between the low-cost instruments and the more specialized and complex measures that focus on the most important aspects such as process quality at child care centers. This suggests that the ability to operationalize quality measures that can regularly be administered at scale in a cost-effective manner requires a closer look at the data to identify if stronger associations do indeed exist between these instruments at lower aggregation levels, i.e., at the subscale or item level.

Efforts to improve CIBV service quality in recent years, specifically, the hiring of staff with post-secondary education for the role of center coordinator, the inspection of centers and subsequent termination of agreements with those that fail to meet standards, and the transfer of children over the age of 3 to early childhood education services appear to be a step in the right direction in terms of improving quality. Nevertheless, achieving significant and sustainable improvement on different aspects of child care quality is not an easy task and requires systematic, long-term efforts. With this in mind, it could be said that Ecuador's CIBVs are in the midst of the initial stage of a process to improve and strengthen their quality. Beyond the aspects of program quality that can be identified and strengthened, we

recommend that the issue of quality be considered from a systematic perspective that takes into account all of the following elements:

1. Sector governance and opportunities for horizontal and vertical coordination between the various stakeholders involved (central government, ministries, regional governments, private sector, civil society organizations, communities and families)

2. Financing needs, the importance of long-term budgetary commitments, and the existence of administrative processes that allow resources to reach providers in a timely manner

3. The importance of establishing continuous quality improvement systems that include the development of standards (with regard to the services to be offered by providers, the outcomes to be obtained by the children, and the competencies to be developed by caregivers); timely measurement (service quality and child outcomes); monitoring, follow-up and reporting systems; and improvement implementation and evaluation processes

4. The urgency of training personnel with the necessary skills to provide the care that children require during this crucial period, while considering salary and professional development incentives to retain staff and enable further growth over time

In this sense, the reforms undertaken and the openness to rich and diverse measurements of service quality on different dimensions constitute a solid foundation for the continuous quality improvement process. The data presented in this study can also serve as a baseline for monitoring the quality of child care services and the results of the investment in their improvement over time.

References

Araujo, M.C., López-Boo, F. and Puyana, J.M. 2013. *Panorama sobre los servicios de desarrollo infantil temprano en América Latina y el Caribe,* Inter-American Development Bank.

Baker, M., Gruber, J. and Milligan, K. 2008. "Universal Childcare, Maternal Labor Supply, and Family Well-Being." *Journal of Political Economy.* 116(4):709–45.

Berlinski, S. and Schady, N. eds. 2015. *The Early Years: Child Well-being and the Role of Public Policy.* Development in the Americas Series. New York: Palgrave MacMillan, and Washington, DC: Inter-American Development Bank.

Bradley, R., Caldwell, B. and Corwyn, R. 2003. The Child Care HOME Inventories: Assessing the quality of family child care homes. *Early Childhood Research Quarterly.* 18: 294-309.

Cryer, T., Harms, T. and Riley, C. 2004. "All About the ITERS-R." Lewisville, NC: PACT House Publishing.

Early, D., Maxwell, K. and Burchinal, M. et al. 2007. "Teacher's education, classroom quality and young children's academic skills: results from seven studies of preschool programs." *Child Development.* March/April 2007, Vol 78. Number 2, pgs. 558-80.

Kyung-Ran, K. 2005. *Teacher Beliefs and Practices Survey: Operationalizing the 1997 NAEYC Guidelines.* Dissertation submitted to Louisiana State University.

La Paro, K., Hamre, B. and Pianta, R. 2011. *Classroom Assessment Scoring System. Toddler Manual.* Preliminary draft prior to publication. Teachstone.

López-Boo, F., Araujo M.C. and Tomé, R. 2015. *¿Cómo se mide la calidad de los servicios de cuidado? Guía de herramientas para la medición de calidad de centros de cuidado de bebés y niños pequeños*, Washington DC, Inter-American Development Bank.

MacPhee, D. 2002. *Knowledge of Infant Development Inventory Survey of Child Care Experiences Manual.* Mimeo.

Mashburn, A., Pianta, R., Hamre, B., Downer, J., Barbarin, O., Bryant D., Burchinal, M., Early, D. and Howes, C. 2008. "Measures of classroom quality in prekindergarten and children's development of academic, language, and social skills." *Child Development*, May/June 2008, vol. 79, number 3, pgs. 732-749.

McCrae, R.R., Kurtz, J.E., Yamagata, S. and Terracciano, A. 2011. "Internal Consistency, Retest Reliability, and their Implications for Personality Scale Validity." *Personality and social psychology review : an official journal of the Society for Personality and Social Psychology, Inc.* 2011;15(1):28-50.

MU Center for Family Policy & Research. 2003. Missouri Infant/Toddler Responsive Caregiving Checklist. Available at https://www.openinitiative.org/content/pdfs/MoNotes/IT_Checklist_Notes.pdf

NICHD. 1999. "Child Outcomes when Child Care Center Classes meet Recommended Standards for Quality." *American Journal of Public Health*. 89(7): 1072-1077.

NICHD. 2000. "Characteristics and quality of child care for toddlers and preschoolers." *Applied Developmental Science*, 4, 116-135.

Provence, S. and Apfel, N. 2001. "Infant-toddler and Family Instrument (ITFI)." Paul H. Brookes Publishing Co. Baltimore.

Rosero, J. and Oosterbeek, H. 2011. "Trade-offs between Different Early Childhood Interventions: Evidence from Ecuador." Tinbergen Institute Discussion Paper, No. 11-102/3. Available at http://www.econstor.eu/bitstream/10419/87225/1/11-102.pdf.

Schady, N., Behrman, J., Araujo, M.C., Azuero, R., Bernal, R., Bravo, D., López-Boo, F., Macours, K., Marshall, D., Paxson, C. and Vakis, R. 2015. "Wealth Gradients in Early Childhood Development in Five Latin American Countries." Forthcoming, *Journal of Human Resources*.

Tomé, R. and López-Boo, F. 2015. "Calidad de Centros de Cuidad o: ¿Importa el perfil del observador? Evidencia de la aplicación del MITRCC en Ecuador." Inter-American Development Bank. *Unpublished manuscript.*

Vandell, D.L. and Wolfe, B. 2000. "Child care quality: Does it matter and does it need to be improved?" Report prepared for the U.S. Department of Health and Human Services, Office for Planning and Evaluation. *Child Development*, 63, 938–949.

Vogel, C., Caronongan, P., Thomas, J., Bandel, E., Xue, Y., Henke, J., Aikens, N., Boller, K. and Murphy, L. 2015. "Toddlers in Early Head Start: A Portrait of 2-Year-Olds, Their Families, and the Programs Serving Them. Volume I: Age 2 Report." Washington, DC, U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation.

Weilin, L., Farkas, G., Duncan, G., Burchinal, M. and Lowe Vandell, D. 2012. "Timing of high-quality child care and cognitive, language, and preacademic development." *Developmental Psychology*. November 5. doi: 10.1037/a0030613.

Annex 1. Statistical Summary of the Instruments

	Mean	SD	90th Percentile	Possible Range of Scores	Obs.
CLASS					
Total	**2.88**	**0.42**	**3.34**	**[1-7]**	**403**
Emotional and behavioral support	3.62	0.45	4.13	[1-7]	403
Positive climate	3.34	0.62	4.13	[1-7]	401
Negative climate	6.61	0.44	7.00	[1-7]	402
Teacher sensitivity	3.36	0.63	4.13	[1-7]	401
Regard for child perspectives	1.98	0.28	2.38	[1-7]	402
Behavior guidance	2.85	0.50	3.50	[1-7]	402
Engaged support for learning	1.64	0.42	2.17	[1-7]	403
Facilitation of learning and development	2.08	0.53	2.75	[1-7]	402
Quality of feedback	1.30	0.33	1.75	[1-7]	401
Language modeling	1.56	0.51	2.25	[1-7]	402
ITERS-R					
Total	**2.08**	**0.53**	**2.83**	**[1-7]**	**403**
Space and furnishings	2.11	0.62	3.00	[1-7]	403
Personal care routines	1.69	0.54	2.50	[1-7]	403
Listening and talking	2.48	1.18	4.33	[1-7]	403
Activities	1.54	0.47	2.22	[1-7]	403
Interaction	3.30	1.26	5.00	[1-7]	403
Program structure	2.57	1.26	4.33	[1-7]	403
Parents and staff	2.00	0.64	2.83	[1-7]	403
HOME					
Total	**24.69**	**6.07**	**33.00**	**[0-42]**	**403**
Caregiver responsivity	6.68	2.52	10.00	[0-11]	403
Acceptance	5.69	0.59	6.00	[0-6]	403
Organization	2.86	0.99	4.00	[0-6]	403
Learning materials	4.47	2.09	7.00	[0-9]	403
Caregiver involvement	3.62	1.76	6.00	[0-6]	403
Variety of stimulation	1.37	0.79	3.00	[0-4]	403
MITRCC					
Total	8.76	4.65	15.00	[0-20]	**403**
ITFI					
Total	**23.53**	**4.37**	**28.00**	**[0-34]**	**403**
KIDI - Coordinator					
Total	**33.55**	**4.65**	**39.00**	**[0-58]**	**396**
KIDI - Caregiver					
Total	**30.80**	**4.97**	**37.00**	**[0-58]**	**403**
Teacher Practices Survey					
Total	**54.49**	**7.69**	**64.00**	**[0-92]**	**403**
Structural quality					
Total	**0.00**	**1.00**	**1.22**	**[-2-5]**	**348**

Annex 2. Internal Consistency

Table A2.1. Internal Consistency of CLASS

	Correlation	Alpha
Emotional and behavioral support	0.97*	0.86
Positive climate	0.93*	
Negative climate	0.50*	
Teacher sensitivity	0.92*	
Regard for child perspectives	0.59*	
Behavior guidance	0.89*	
Engaged support for learning	0.91*	0.90
Facilitation of learning and development	0.86*	
Quality of feedback	0.80*	
Language modeling	0.81*	
Full scale		0.91

*Significant correlations at 99%

Table A2.2. Internal Consistency of ITERS-R

	Correlation	Alpha
Space and furnishings (items 1-5)	0.63*	0.61
Personal care routines (items 6-11)	0.69*	0.49
Listening and talking (items 12-14)	0.76*	0.61
Activities (items 15-24)**	0.65*	0.63
Interaction (items 25-28)	0.82*	0.77
Program structure (items 29-32)**	0.76*	0.76
Parents and staff (items 33-39)**	0.63*	0.57
Full scale (items 1-39)		0.87
Child-related items (items 1-32)		0,87

*Significant correlations at 99%
**The correlation and alphas do not factor in items 23, 32 and 36, since these apply to less than 15%, 5% and 11% of the centers, respectively.

Table A2.3. Internal Consistency of HOME

	Correlation	Alpha
Caregiver responsivity (11 items)	0.83*	0.74
Acceptance (6 items)	0.32*	0.51
Organization (9 items)	0.56*	0.26
Learning materials (9 items)	0.74*	0.66
Caregiver involvement (6 items)	0.74*	0.66
Variety of stimulation (4 items)	0.53*	0.35
Full scale (42 items)		0.83

*Significant correlations at 99%

Annex 3. Construction of the Structural Quality Index

To construct the structural quality index, variables relating to center infrastructure and services, as well as some characteristics of the coordinators and caregivers, were considered; however, all of the variables available in the survey questionnaire were not included in the index. In order for variables to be included in the index, they had to meet two criteria. The first was their variability, with characteristics present in less than 10% of the centers in the sample being discarded. The second criterion was that the variables needed to have made a significant contribution to the component. The first component was chosen to calculate the final wealth index. Table A3.1 presents the weight of each variable in the generated index, or the correlation between each variable and the component.

Table A3.1. Rotated Component Matrix

Variable	Component
Drinking water	0.43
Sewage system	0.42
Electricity, with meter	0.34
Good natural lighting	0.15
No broken windows	0.25
Adequate furnishings for the children	0.13
Dining area	0.09
Nursery area	0.08
Storage area	0.33
Fencing	0.29
Safe playground equipment	0.22
% of children waiting to enroll but no slots available	0.19
Child-caregiver ratio at center	-0.14
Coordinator's years of experience working with children < 60 months old	0.21
Coordinator does not have child care duties	0.08
Caregiver's years of experience working with children < 60 months old	0.08
Caregiver's years of education	0.23

Extraction method: principal component analysis. Rotation method: Varimax with Kaiser normalization. Eigenvalue =2.88. Proportion of variance explained=0.17.

Annex 4: Comparison between Centers Whose Operators Had an Agreement with MIES in April 2015 and Those Without

Table A4.1. Quality of the Centers with and without an Agreement with MIES in April 2015

	Centers with an agreement	Centers without an agreement	Diff. (t-test)
CLASS			
Total	2.92	2.82	**
Emotional and behavioral support	3.65	3.57	**
Engaged support for learning	1.69	1.57	***
ITERS-R			
Total	2.15	1.99	**
Space and furnishings	2.17	2.01	**
Personal care routines	1.76	1.59	***
Listening and talking	2.53	2.41	
Activities	1.56	1.51	
Interaction	3.41	3.14	*
Program structure	2.69	2.39	**
Parents and staff	2.06	1.92	**
HOME			
Total	25.44	23.60	***
Caregiver responsivity	6.92	6.33	**
Acceptance	5.72	5.64	
Organization	2.89	2.80	
Learning materials	4.73	4.10	***
Caregiver involvement	3.73	3.47	
Variety of stimulation	1.44	1.27	*
MITRCC			
Total	9.37	8.09	***
ITFI			
Total	23.79	23.11	
KIDI - Coordinator			
Total	34.17	32.61	***
KIDI - Caregiver			
Total	30.90	30.53	
Teacher Practices Survey			
Total	54.92	53.65	
Structural quality			
Total	0.15	-0.23	***
Observations	**166**	**238**	

Standard errors of the mean adjusted for clustering at the canton level

Significant differences at *** 99%, ** 95%,* 90%

Table A4.2. Other Characteristics of the Centers with and without a MIES Agreement in April 2015

	Centers with an agreement	Centers without an agreement	Diff. (t-test)
Monthly salary, caregiver (US$)	222.59	216.35	*
Monthly salary, coordinator (US$)	533.80	542.33	
Total number of children at the center	36.78	23.64	***
Number of years center has been in operation	11.43	9.61	*
% of families that make a payment for services	0.55	0.44	*
% of centers in urban areas	0.57	0.30	***
Age gap between children in the study group (months)	15.72	18.70	***

Standard errors of the mean adjusted for clustering at the canton level

Significant differences at *** 99%, ** 95%,* 90%

Annex 5. Quality and Homogeneity of the Age Groups

Table A5.1. Comparison of Groups of Children with Age Gaps of up to 12 months and 12 months or more

	Gap ≤ 12 mo.	Gap > 12 mo.	Diff. (t-test)
CLASS			
Total	2.86	2.88	
Emotional and behavioral support	3.61	3.62	
Engaged support for learning	1.63	1.65	
ITERS			
Total	2.17	2.05	*
Space and furnishings	2.19	2.08	
Personal care routines	1.76	1.66	
Listening and talking	2.50	2.49	
Activities	1.54	1.53	
Interaction	3.55	3.20	***
Program structure	2.80	2.48	**
Parents and staff	2.07	1.97	
HOME			
Total	26.27	23.98	***
Caregiver responsivity	7.27	6.45	***
Acceptance	5.66	5.70	
Organization	3.03	2.78	**
Learning materials	4.80	4.31	**
Caregiver involvement	4.07	3.42	***
Variety of stimulation	1.45	1.32	
MITRCC			
Total	9.38	8.60	
ITFI			
Total	23.93	23.32	
KIDI - Coordinator			
Total	33.69	33.41	
KIDI - Caregiver			
Total	31.66	30.29	***
Teacher Practices Survey			
Total	53.60	54.57	
Structural quality			
Total	0.19	-0.12	***
Observations	**120**	**277**	

Standard errors of the mean adjusted for clustering at the canton level

Significant differences at *** 99%, ** 95%,* 90%