

HOW IS CHILD CARE QUALITY MEASURED?

A toolkit

FEBRUARY 2016

Florescia López Boo
María Caridad Araujo
Romina Tomé



Cataloging-in-Publication data provided by the Inter-American Development Bank
Felipe Herrera Library.

López Boo, Florencia.

How is child care quality measured? / Florencia López Boo, Maria Caridad Araujo, Romina Tomé.
p. cm.

Includes bibliographic references.

978-1-59782-222-0 (Paperback)

978-1-59782-223-7 (PDF)

1. Child development. 2. Child care. 3. Nursery schools. I. Araujo, Maria Caridad. II. Tomé,
Romina. III. Inter-American Development Bank. Social Protection and Health Division. IV. Title.

HQ778.5 .L67 2015 eng. ed.

IDB-BK-147

Editorial Design: Círculo Salvo Comunicación | circulosalvo.com

Keywords: child development, Early childhood development, ECD, child care, Nursery schools,
Quality measurement, Inter-American Development Bank, IDB, Social Protection

JEL codes: I10, I20, I30, I38, J13.

Copyright © 2016 Inter-American Development Bank. This work is licensed under a Creative Commons IGO 3.0 Attribution-NonCommercial-NoDerivatives (CC-IGO BY-NC-ND 3.0 IGO) license (<http://creativecommons.org/licenses/by-nc-nd/3.0/igo/legalcode>) and may be reproduced with attribution to the IDB and for any non-commercial purpose. No derivative work is allowed.

Any dispute related to the use of the works of the IDB that cannot be settled amicably shall be submitted to arbitration pursuant to the UNCITRAL rules. The use of the IDB's name for any purpose other than for attribution, and the use of IDB's logo shall be subject to a separate written license agreement between the IDB and the user and is not authorized as part of this CC-IGO license.

Note that link provided above includes additional terms and conditions of the license

The opinions expressed in this publication are those of the authors and do not necessarily reflect the views of the Inter-American Development Bank, its Board of Directors, or the countries they represent.



Inter-American Development Bank
Social Protection and Health Division

HOW IS **CHILD CARE QUALITY** MEASURED?

A toolkit

FEBRUARY **2016**

scl-sph@iadb.org
www.iadb.org/SocialProtection






TABLE OF CONTENTS

PREFACE	06
ACKNOWLEDGMENTS	08
USER'S GUIDE	09
Who should use this guide?	09
How to use this toolkit?	09
CHAPTER 1: WHY IS THE QUALITY OF CHILD CARE FOR YOUNG CHILDREN SO CRUCIAL?	10
1.1 Introduction	11
1.2 What Do We Mean by Quality? Critical Dimensions for Infant and Toddler Care	12
1.3 The Importance of Child Care Quality and its Measurement	14
Quality child care and child development in developed countries	14
Child care quality in Latin America and the Caribbean	16
CHAPTER 2: INTRODUCTION TO MEASURING CHILD CARE QUALITY	18
2.1 Introduction	19
2.2 Structural Variables	19
What are structural variables?	19
How are they measured?	19
What does the literature say?	19
2.3 Process Variables	21
What are process variables?	21
How are they measured?	21
What does the literature say?	21
2.4 How Do Governments Measure Child Care Quality?	22
CHAPTER 3: KEY DEFINITIONS FOR MEASUREMENT	24
3.1 Introduction	25
3.2 What is the Goal of measuring Child Care Quality?	25
3.3 What Is the Study group?	25
3.4 Measuring: the how	26
Observation vs. self-report	27
Filming	29

3.5 What Instruments Are Available?	29
3.5.1 ECERS-R and ITERS-R	30
Description	30
Experiences of use	31
3.5.2 CLASS Toddler	32
Description	32
Experiences of use	33
3.5.3 ORCE	34
Description	34
Experiences of use	34
3.5.4 CIS	35
Description	35
Experiences of use	35
3.5.5 Assessment Profile	35
Description	35
Experiences of use	36
3.5.6 MITRCC	36
Description	36
Experiences of use	37
3.5.7 CC-IT-HOME	37
Description	37
Experiences of use	37
3.5.8 Other instruments	38
3.6 The use of secondary surveys	39
3.7 How to Select an Instrument?	40
CHAPTER 4: IMPLEMENTATION	42
4.1 Introduction	43
4.2 Training	43
4.3 Piloting and Cultural Adaptation	44
4.4 Generation of Unique Identifiers in Databases	45
4.5 Critical Context Variables	45
4.6 Inter-rater Reliability	46
4.7 Following Protocols	47
General protocol	47
Interview/observation	47
Filming	48
Editing	49
Coding	50
CHAPTER 5: FINAL CONSIDERATIONS	52
TABLES	54
REFERENCES	75
ANNEX 1. VALIDITY OF THE CLASS PRE-K INSTRUMENT	83



PREFACE

Gaps in development become evident in the early years of life, which is why, in recent times, child development issues have been placed on the agenda of most countries in Latin America and the Caribbean.

In the past, governments mainly focused their efforts on expanding child care coverage for young children (available at child care centers often referred to as daycare centers, nursery schools and preschools); however, countries in the region now face the challenge of ensuring the quality of the child care services they offer, especially since the children who attend these centers do so during their first years of life, a critical period for skill development. Investment in quality child care is a social policy that favors equity by promoting opportunities for development from the beginning of life. Working to face this challenge demands, first and foremost, an understanding of what constitutes quality child care and what is required to provide it.

While there is no one formula for ensuring the quality of center-based child care, we know that it encompasses a number of easily

quantifiable elements including the provision of nutritious food, the child-caregiver ratio, the salary and education level of caregivers, continuing education for caregivers, as well as aspects that are more difficult to measure, such as the quality of interactions between caregivers and children.

One might be inclined to think that an average-quality service may be sufficient for poor families, given their rather precarious child care alternatives. For example, within highly vulnerable families, it is common practice to leave children home alone or in the care of another child, or to take them to the mother's workplace, where they receive no attention and may be exposed to risks. The literature concludes that the quality of care received by children attending child care centers—particularly in the case of poor or at-risk children—is crucial, because centers can meet children's needs in terms of nutrition, stimulation, hygiene, and affection that otherwise may not be met in the home. Thus, an investment in quality center-based child care gives these children a better foundation in the critical dimensions of development during early childhood.

In our region, as in others, there is not enough evidence about the specific critical and cost-effective actions needed to improve child care quality. In contrast with sectors such as primary education, in which the region has invested significant effort in collecting indicators on teacher quality (e.g., the Stallings Classroom Snapshot instrument), resources in schools (e.g., school censuses), and levels of student learning (e.g., the implementation of the Program for International Student Assessment [PISA]), governments still lack systematized and reliable information about child care services.

Contributing to this dearth of information is the lack of regular use of instruments that measure various aspects of quality at child care centers and the experiences that children have at these centers. This information is key to identifying the processes that contribute to improving the quality of centers in a cost-effective manner and also to defining and monitoring compliance with a set of quality standards at both publicly-funded and private centers. More importantly, there appears to be an increasing emphasis on child development indicators reflected in the development agenda following the progress report on the 2015 Millennium Development Goals. For this reason, having comparable data on the quality of child care centers will contribute to the creation of a measurable accountability mechanism for the investments made by countries.

To help fill this gap, and in an effort to operationalize some of the main findings of the Inter-American Development Bank's flagship 2015 publication, "The Early Years", the Bank's Division of Social Protection and Health presents this user-friendly tool that i) addresses in greater detail the definition of a quality service, ii) provides a menu of available tools for the measurement of the quality of child care centers serving infants and toddlers ages 0 to 3 years (36 months), and iii) reports on the implementation

process of these instruments in the region. In other words, it presents a theoretical description of the tools and a guide explaining where, how, and when to use each tool, based on a detailed approach with different dimensions to consider in order for the quality measurement to be successful.

This book contains several original contributions. First, it synthesizes both a theoretical and a practical approach to measuring the quality of center-based child care in a single publication. Second, it discusses, in detail, important aspects to consider when conducting measurements. Third, it systematizes the information about assessment tools so that the reader can easily make comparisons between their various dimensions, costs, and benefits. Lastly, it is also available in Spanish in order to facilitate Latin American readers' access.

This publication will serve as a guide for researchers and professionals interested in translating the discussion on improving child care quality into concrete actions and results.

I am sure you will enjoy it and find it useful.

Ferdinando Regalia

Chief of the Division of Social Protection and Health
Inter-American Development Bank



ACKNOWLEDGMENTS

Firstly, we would like to highlight the valuable contributions of the members of the IDB's advisory committee on child development issues, which met in January 2012 with the objective of guiding our team on the design of an applied agenda on the research of center-based care quality: Steve Barnett, Raquel Bernal, Margaret Burchinal, Patrice Engle, Sally Grantham McGregor, Jennifer LoCasale Crouch, Milagros Nores, Deborah Phillips, Maureen Samms Vaughan, Sergio Urzúa, Hiro Yoshikawa and Martha Zaslow.

Secondly, we would like to thank the anonymous referee for invaluable feedback, Sara Schodt and Milagros Nores for their ongoing contributions throughout the process of writing this document and for reading and commenting on the final draft. Thirdly, we wish to thank the company Habitus of Ecuador, with which we carried out our quality measurement project in Ecuador. We would also like to thank Guillermo Martínez for editing this document and Katie Metz for an excellent revision and translation.

Lastly, we gratefully acknowledge the institutional support received from Ferdinando Regalia, Chief of the IDB's Division of Social Protection and Health, Norbert Schady, Principal Economic Advisor for the Social Sector, as well as Héctor Salazar, Manager of the Social Sector, who helped us move forward with this project.



USER'S GUIDE

Who should use this guide?

This toolkit is designed to be a resource for researchers and technical staff of any discipline, working for governments and institutions interested in measuring and monitoring the quality of child care centers serving infants and toddlers ages 0 to 3 years (36 months).¹

How to use this toolkit?

This toolkit is organized into five chapters that sequentially present evidence to the reader about the importance of measuring child care quality, and then they guide her through the steps for measuring quality.

- **Chapter 1** explains **what** constitutes quality in child care services and **why** it should be measured;
- **Chapter 2** specifies **what to measure** in order to evaluate child care quality;
- **Chapter 3** discusses **the decisions** that should be evaluated **prior** to taking measurements, and it describes the different types of assessments and instruments;
- **Chapter 4** reflects on the **dimensions of measurement implementation**;
- **Chapter 5** presents some **conclusions and recommendations** for those planning to measure the quality of infant and toddler child care.

¹Throughout this guide, the terms “infants” and “toddlers” are used to refer to children under 36 months of age.



WHY IS THE QUALITY OF CHILD CARE FOR YOUNG CHILDREN SO CRUCIAL?

● CHAPTER **1**



1.1 / Introduction

Increasingly, countries are choosing to invest in early childhood development—focusing their efforts on children from the poorest households—as a strategy to improve educational outcomes for these children and their trajectories in adulthood (Engle et al., 2007; Engle et al., 2011). Specifically, children who grow up in an environment of poverty are exposed to multiple risk factors, including poor nutrition, precarious health conditions, dull learning environments, constant high levels of stress, and other environmental hazards. All of these factors affect their cognitive, socio-emotional, and physical development, and they slow the development of their potential during their school years and into adulthood (Fernald, Gertler and Neufeld, 2009; Grantham-McGregor et al., 2007). In Latin America and the Caribbean (LAC), it has been shown that children living in poverty, on average, achieve lower levels of cognitive development and drop out of school at higher rates than other

children, thereby dooming them to a lifetime of low-paying jobs and the perpetual cycle of poverty in which they were raised (Rolnick and Grunewald, 2007; Heckman, 2006; PREAL, 2006).

Policy interventions in early childhood are promising, as they have the potential to mitigate some of the effects of early deprivation. Evidence suggests that, in the case of child care centers, there is significant potential to improve the situation of the most vulnerable children, as long as quality services are provided. For this reason, it is essential that governments have tools that allow them to measure the quality of care provided by centers, focusing on those dimensions of quality that are critical to positive child development outcomes (Martinez-Beck in Zaslow et al., 2011; Peisner-Feinberg et al., 2001; Vandell and Wolfe, 2000; Lamb, 1996, 1998; Phillips, 1987; and Clarke-Stewart and Fein, 1983).

The need to provide quality services is crucial for the infant and toddler age group, i.e., children ages 0 to 3 years (36 months). This statement is supported by both biological evidence (the first 1,000 days of life, from conception, as a window of opportunity for



nutritional interventions) and the literature on child psychology (sensitive periods for attachment or language development). Remarkably little information exists on the differences in service quality parameters critical to the proper development of infants and toddlers, as compared to those required for preschoolers (Barros et al., 2011). Process variables,² which characterize the quality of child care routines and interactions between children and caregivers, become even more important for this age group. This is because even though young children need less in the way of structured content, they require more individualized care than older children.

On the other hand, internationally-developed instruments used to measure the quality of care for infants and toddlers cover different age ranges, which further hinders systematic and comparable measurements over time, even within the same “family” of instruments. For example, the Classroom Assessment Scoring System (CLASS) has two versions, one for infants up to 18 months of age and one for toddlers between 15 and 36 months of age. The Infant/Toddler Environment Rating Scale – Revised (ITERS-R) is used for children 0 to 30 months of age while the Early Childhood Environment Rating Scale – Revised (ECERS-R) is designed for children 30 to 60 months of age. The Child Care Infant/Toddler HOME (CC-IT-HOME) and the Missouri Infant/Toddler Responsive Caregiving Checklist (MITRCC) are used with children ages 0 to 36 months (a preschool version of the Child Care HOME is also available). All of these instruments are discussed in greater detail in section 3.5. In their desire to ensure quality, child care centers serving infants and toddlers may face complications in terms of identifying and measuring quality.

1.2 / What Do We Mean by Quality?

Critical Dimensions for Infant and Toddler Care

Part of the challenge of ensuring high-quality child care is that there is no one formula for achieving it. The general consensus, especially in early childhood, is that quality must be comprehensive, in other words, not only focused on education and health, but also on cognitive and emotional development, nutrition, and parenting (Halle, Whittaker and Anderson, 2010; Zill et al., 2003; NICHD, 1996). Kagan (2010) identifies critical areas in which the quality of child care should be evaluated, including the child’s relationships with other children and adults, lesson plans and activities, teaching, evaluation of the child’s progress, medical care, the profile of adult caregivers, family context, community relationships, the physical environment, the child care center coordinator’s leadership, and aspects related to center management.

Although a general consensus exists among experts in early childhood education and development regarding the importance of these areas (NAEYC, 2012; NICHD, 2000a,b), there is no single, universal way to describe how these elements—which together constitute quality—should be combined or evaluated and what weight should be assigned to each. Consequently, defining and measuring child care quality becomes a complex task. International studies on the supply of child care services during early childhood show that governments are capable of playing an important role in ensuring quality through their control and governance functions. For instance, both in

² Chapter 2 analyzes the two types of dimensions, process and structural, in detail. Process variables consist of the quality of interactions between the children and their caregivers and between the children themselves, as well as the activities they engage in while at the child care center. Structural variables generally identify the resources that facilitate those interactions: group size, the caregiver’s education, experience and salary, infrastructure and safety, curriculum, and materials.

the United States and Europe, those states and countries with the strictest licensing standards for the operation of child care centers and those who have established accreditation systems for their centers, are also the ones with the highest quality. In the US, it has also been found that centers that surpass state standards tend to provide higher quality care to children (Van Leer Foundation, 2011; Helburn, 1995). Howes et al. (1992) find that child care centers that regulate the child-caregiver ratio, or the number of children per adult, and group size, have a better quality of child care. In addition, they observed that exceeding the state standard for the optimal child-caregiver ratio in a classroom by just one child is associated with lower-quality care.

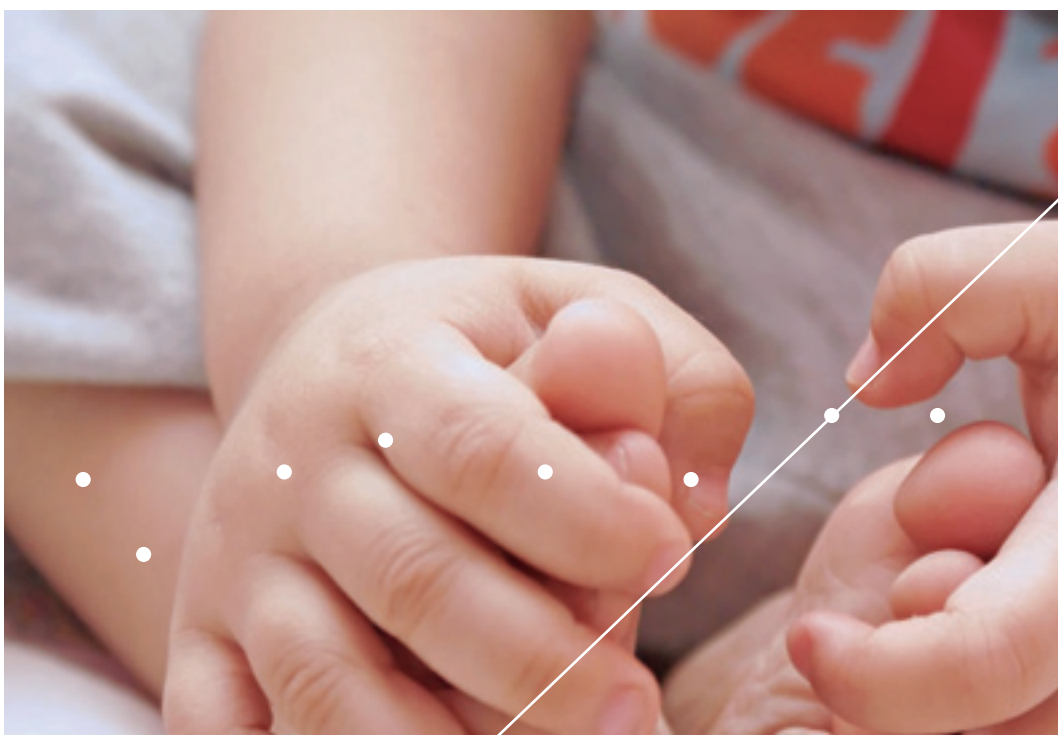
According to the opinion of experts consulted by the IDB (Barnett, 2012; IDB Advisory Committee, 2012), six critical elements should be used to define quality for the 0-3 years age group at child care centers in LAC:

1. The provision of nutritious food in optimal conditions of hygiene, sanitation and safety;
2. The child-caregiver ratio, which should not exceed six children per adult for the

12-36 months age group and should be even lower for infants;

3. The quality, frequency and intensity of interactions between caregivers and children;
4. A system that regularly monitors quality at all centers;
5. Training and professional development for caregivers and teachers;
6. Stimulating activities, play materials, and spaces.

Why would these critical dimensions differ from those relevant to the care of children over the age of 3? First and foremost, since infants and toddlers' immune systems are still developing, group care exposes them to greater risks. In addition, specialized literature documents that, prior to gaining full mobility, infants and toddlers have a limited ability to initiate interactions with their caregivers, other children, or their surroundings (Howes et al., 1992). For this reason, as well as in the early childhood attachment theories (Bowlby, 1969), process indicators such as caregivers' responsiveness and engagement with children play an even greater role in the development of infants and toddlers. As with older children, process variables are more closely associated with infants and toddlers' development than structural variables, although it is important to



note that improved infrastructure also allows for better processes.

Structural variables are more frequently measured in preschool and school environments. For example, the type and format of activities planned for infants and toddlers so that they develop behaviors that provide a solid cognitive, social and emotional foundation in life and that, over time, better prepare them for school. However, infant and toddler development depends more heavily upon the quality of children's physical and verbal interactions with their caregivers than that of older children (Gonzalez-Mena and Widmeyer Eyer, 2007; Raikes and Edwards, 2009). This is also related to the fact that younger children require more individualized attention.

So, while the six aforementioned critical elements for defining quality at child care centers may be relevant to all children ages 0 to 5, the second and third points seem much more relevant to the 0-3 years age group for two reasons: 1) infants and toddlers require lower child-caregiver ratios (fewer children per caregiver) and 2) the frequency and quality of interactions play a more important role in this group.

How might these critical dimensions differ in countries from different regions of the world? Unlike many studies conducted in the United States, analyses by Berlinski, Galiani and McEwan (2008) and Berlinski, Galiani and Gertler (2009) in LAC suggest that mere access to preschool education does have a positive impact on children's outcomes in primary school (although recent unpublished studies from Chile seem to show the opposite). Since the studies did not measure preschool quality, these results are consistent with two possible explanations: (i) that preschool services were actually high quality or (ii) that for children from very poor

families, even child care centers providing poor-quality care can produce positive results if the school environment is better than that found in the child's own home.³ It is not implied that the standards at child care centers in LAC are low, quite the contrary, as it is known that many of the poorest children in the region are at risk or vulnerable in their own homes. It is this reality, therefore, that demands a precise understanding of the quality of care provided at child care centers. In this sense, Halpern (1986) warns of the danger of generalizing U.S. studies to other contexts, since, in the absence of center-based child care, a significant number of children in LAC face much bleaker situations in terms of the quality of care in the home. These children live among poverty and malnutrition, without the social support networks that exist to help families in developed countries. These challenges suggest that early childhood programs in some LAC countries have a responsibility to address child development in a comprehensive manner, in other words, to take into account children's health and nutritional needs, as well as their cognitive and social development. Consequently, the instruments used to measure quality must also consider processes related to children's medical care and nutrition at centers.

1.3 / The Importance of Child Care Quality and its Measurement

Quality child care and child development in developed countries⁴

Although some of the empirical evidence shows that attendance at a child care center might have benefits, there is equally

³ The impact of child care on child development is related to the characteristics of the alternative form of care that the child would have had access to in the absence of the service (known as a counterfactual).

⁴ This section is based on the Project Concept Note (Schodt, 2012).

compelling evidence that suggests that if the goal is to achieve results in terms of the healthy development of children, then it is vital to ensure the quality of care at these centers (NICHD, 2000b). In fact, a negative association between low-quality care and child development has been documented, especially in social-emotional aspects such as aggression in children (Belsky, 1988; Howes et al., 1992). The question then is, what is the actual source of the problem, the child care centers themselves or the low-quality services they provide? If centers fail to provide the same or higher quality of care to children as compared to what they receive at home, the effects could be truly harmful to their health and development. Several studies yield mixed results in terms of the relationship between quality of care and children's social-emotional and behavioral development. While Love et al. (2003) and the NICHD study (2000a) find that center-based child care has positive outcomes on child development, the two studies analyzed in the next paragraph find the opposite.

A good example of the importance of assessing the quality of child care centers is the study by Baker, Gruber and Milligan (2008), which looks at the universal expansion of child care services subsidized by the government of Quebec for children up to age 5. The authors found that, in the short term, prolonged exposure to center-based child care was harmful, and specifically, it worsened "hyperactivity, inattention, aggression, motor and social skills, mental health, and propensity toward disease in children." In a long-term follow up, the authors then show these non-cognitive deficits persisted to school ages, and also that cohorts with increased child care access subsequently had worse health, lower life satisfaction, and higher crime rates later in life (Baker, Gruber and Milligan, 2015). In the same vein, Yamauchi and Leigh (2011) looked at the relationship between child care centers and the behavior of Australian preschoolers. They found that, in general, full-time attendance at a child care center

is negatively associated with children's behavior, especially for children from families with a strong emotional bond or high socioeconomic status. In both cases, it appears that attending these child care services (deemed by the authors to be of high quality) results in negative effects on behavior in young children; however, neither of the two studies explicitly addresses the measurement of quality.

Ample empirical evidence suggests that child care quality directly affects child outcomes in a number of dimensions. Specifically, participation in high-quality programs either has clear positive impacts on development, as demonstrated with experimental evaluations of programs such as the Abecedarian Project (Campbell and Ramey, 1995), Early Head Start (Love et al., 2005) and Perry Preschool (Schweinhart et al., 2005), or is associated with better health and social-emotional and cognitive development in the short term (Vandell and Wolfe, 2000; Phillips, 1987; NICHD, 2000a,b; Shonkoff and Phillips, 2000; Burchinal et al., 1997). While most available studies only consider children's outcomes in the short term, several recognized studies from the U.S., such as the Abecedarian Project, High/Scopes Perry Preschool, and Early Head Start, have also found significant positive long-term effects resulting from high-quality, well-designed, and well-implemented child care services. Less violent crime, fewer teenage pregnancies, less dependence on welfare programs, reduced drug use, and higher wages and rates of employment stand out from among the long-term impacts identified. These studies evaluated children from vulnerable households who received high-quality care during early childhood (NAEYC, 2012). Although the Abecedarian Project and High/Scopes Perry Preschool studies relied upon very small samples, they emphasize the need for quality care during early childhood in order to improve child development outcomes, and they suggest that low-quality interventions have failed to demonstrate

a positive association with child outcomes (Peisner-Feinberg et al., 2001). The long term evaluation of the at-scale Quebec study reinforces this message.

The “dose” of care—the frequency, intensity, and duration of exposure to the child care service—also appears to be an important aspect of quality that clearly affects child outcomes. In a meta-analysis of 34 preschool programs in the U.S., Nelson et al. (2003) found that those programs that offer at least 20 hours of care per week, and that provide more years of center based care produce greater impacts on cognitive development and academic achievement, suggesting that high-quality care for young children offers cumulative benefits.

Anywhere in the world, the ability of teachers and caregivers to interact effectively with young children depends, first and foremost, on the number of children in their care. Consequently, it can reasonably be assumed that the findings of the literature described in this section can be applied in LAC, where limitations in terms of human capital and funding translate into higher child-caregiver ratios than those seen in the U.S. This means that, in LAC, developing measurements of quality that go beyond structural aspects to really capture the quality of processes between caregivers and children is even more urgent.

Child care quality in Latin America and the Caribbean

Although there are experimental studies in the U.S. and other developed countries that clearly present the positive long-term effects of high-quality early childhood programs on academic progress, social skills and professional achievement, it is worth mentioning that this

type of longitudinal evidence has not been gathered for most of LAC.⁵

Other studies conducted in LAC evaluate the impact of preschool attendance: Berlinski, Galiani and Gertler (2009) in Argentina; Filp and Schiefelbein (1982) in Argentina, Bolivia, Chile and Colombia; McKay and McKay (1983) in Cali, Colombia; and Filp et al. (1983) in Chile. This research also documents moderately positive outcomes related to children’s development in elementary school and their performance on standardized tests. More recently, in Chile, Urzúa and Veramendi (2011) monitored the significant effort to expand child care coverage and found that participation in these programs improved cognitive development in children over age 2.

Using a quasi-experimental methodology, Arredondo et al. (2011) evaluated a Mexican program launched in 2007 called *Estancias infantiles para apoyar a madres trabajadoras* (Daycare Program to Assist Working Mothers), designed to promote labor force participation by mothers of children ages 1 to 4 through subsidies that cover the cost of child care outside the home. The evaluation method took advantage of the standard daycare enrollment process, with the treatment group formed by a sample of children receiving care and the control group formed by a sample of children on the waiting list. The authors found that, although access to subsidized child care increased mothers’ labor force participation, the children who attended daycare centers failed to realize any benefits in terms of their cognitive development or health.

In Brazil, Barros et al. (2011) analyzed data on child care cost and quality, along with development indicators and family

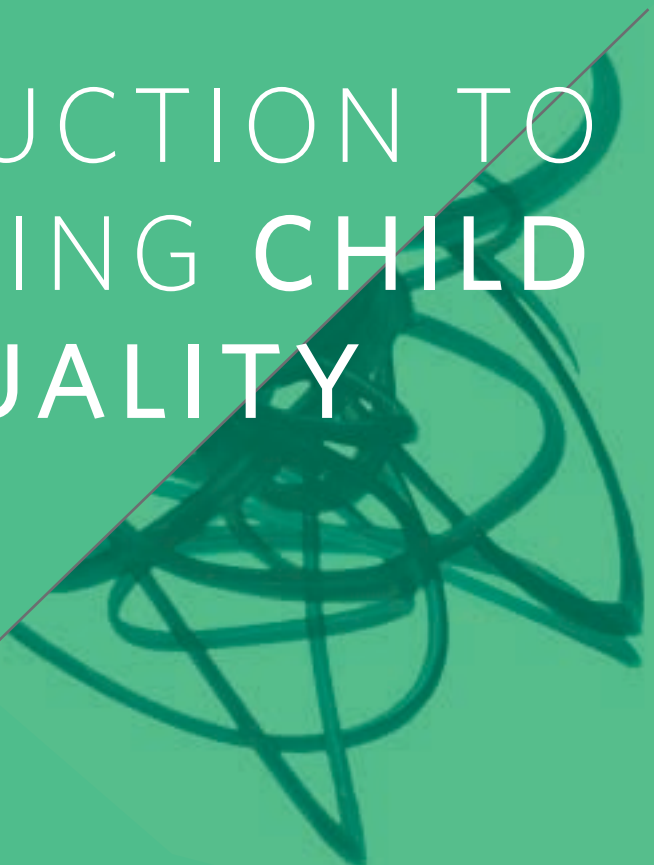
⁵ One exception, however, is the work of Grantham-McGregor et al. (1991, 1997), which evaluated a parenting intervention in Jamaica that included weekly home visits by community health workers. This program demonstrated positive outcomes in a wide range of social, health and educational dimensions up through the age of 22 in those children whose parents had benefited from the intervention (Walker et al., 2011). Similar results were also observed in the children of the children who originally took part (Walker et al., 2012). An adaptation of this pilot program in Colombia has also yielded positive results, although the nutritional supplements have not been shown to have an effect (Attanasio et al., 2012).

characteristics of children in the city of Sao Paulo. The authors found that it was process dimensions rather than structural ones that had a significant association with children's development outcomes. They also found that the cost-effectiveness ratio of improving the quality of care was almost 60 to 1; however, if the increase in quality were to focus on improving the process dimensions most strongly associated with the children's development, the authors estimate that quality could be achieved with a cost-effectiveness ratio almost 40 times higher.⁶ As for the "dose" of care—the frequency, intensity, and duration of exposure to the child care service—in Bolivia, Behrman, Cheng and Todd (2004) found that the greatest benefits are observed in children who begin attending daycare at age 2 and who participate for at least 12 to 24 months, with diminishing returns after that age.

Lastly, Rosero and Oosterbeck (2011) compare the impact of the two different types of child development programs that are implemented in Ecuador: home visits and child care centers. The study authors found that while home visits improved children's cognitive development and mothers' mental health, attendance at child care centers was detrimental to these two dimensions.

⁶ These results are similar to those from the "Cost, Quality, and Outcomes in Child Care Centers Study," which found that while child care is expensive across the board, the costs of high-quality services, though greater, are only marginally so when compared to those of poor-quality services (Helburn, 1995).

INTRODUCTION TO MEASURING CHILD CARE QUALITY



● CHAPTER 2



2.1 / Introduction

Measuring the quality of child care for infants and toddlers is a complex process. The first two sections of this chapter explicitly set forth what kind of variables are used to measure quality, what it is exactly that these variables measure, and what the literature says about how these variables relate to the development of young children. In addition, there is a section in this chapter that describes how some child care services in LAC measure and monitor the quality of their centers.

2.2 / Structural Variables

What are structural variables?

Structural variables determine the presence or absence of those resources that facilitate the interactions characteristic of an environment of care; they are also variables that can be more easily controlled (Thomason and La Paro, 2009). The authors have organized these variables into four major groups: center infrastructure, health and safety issues, the group of children, and the caregiver. Table 1 summarizes these variables and provides some examples. In terms of health and safety issues, some examples of structural variables include the following: hygiene, health and safety protocols; infectious disease protocol; and nutritional monitoring. With regard to children and caregivers, some of the structural indicators usually measured include group size, child-caregiver ratio (the number of children per adult), professional qualifications of the staff responsible for

the care and nurturing of the children (years of education, previous experience, and training), planning of daily activities, monitoring and supervision of caregivers, and staff salaries. Infrastructure variables include access to potable water, a waste disposal system, access to electricity and telephone service, the amount of physical space for children, the availability of materials, facilities protected against environmental hazards or other risks, and the characteristics of play areas and furnishings.

How are they measured?

These variables are measured using observation checklists, interviews or questionnaires. Some of the instruments reviewed in this toolkit that measure the structural variables at centers are ITERS-R, CC-IT-HOME, MITRCC, the Observational Record of the Caregiving Environment (ORCE), the Caregiver Interaction Scale (CIS), and the Assessment Profile. Supplemental ad hoc surveys can also be used to collect information on infrastructure variables and characteristics of the caregiver and the center coordinator. (Section 3.6 provides details about the three questionnaires used by the authors in Ecuador: the Structural Questionnaire, the Caregiver Survey, and the Coordinator Survey.) In some programs with sophisticated information systems, there are administrative databases that are regularly updated with this information for each center.

What does the literature say?

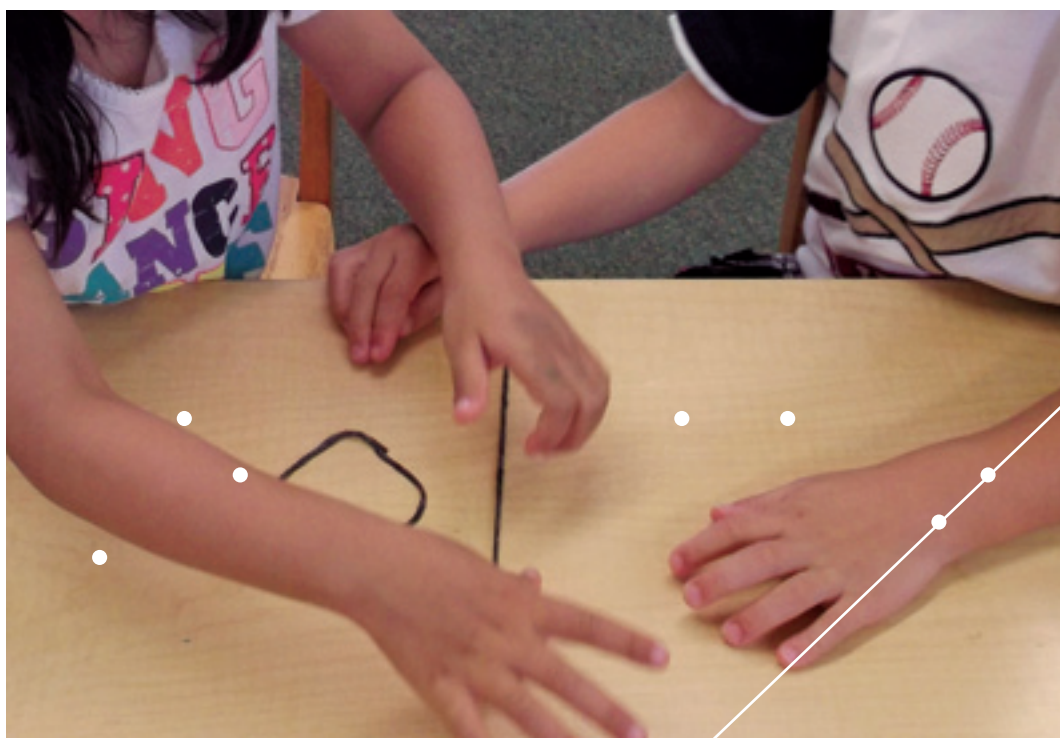
It has been found that structural variables are significantly correlated with child care quality (Howes et al., 1992). For example, when the child-caregiver ratio is higher, positive affective interactions occur less frequently. When a center has little or no materials or basic infrastructure, routine health and nutrition practices tend to be poor or nonexistent and activities are lower

in frequency and quality (Clarke-Stewart, Gruber and Fitzgerald, 1994; Howes, 1983; NICHD, 1996, 2000a; Phillips in Burchinal et al., 1997; Volling and Feagans, 1995). Furthermore, it has been shown that caregiver salaries are highly correlated with the quality of child care centers. Similarly, higher wages are associated with better outcomes for children (Kagan, 2010). In the same vein, the Study of Early Child Care conducted by the National Institute of Child Health and Human Development's (NICHD) Early Childhood Research Network found that caregivers with more years of education and those trained in early childhood education tend to use activities that are more appropriate and stimulating for the development of the children in their care (NICHD, 2000a; Vandell and Wolfe, 2000).

Several studies have found that there are significant relationships between structural quality variables and children's performance. Specifically, smaller class size, low child-caregiver ratios, and well-qualified caregivers are associated with improved quality of care as well as better development outcomes in children (NICHD, 1999). Furthermore, classrooms with more adults per child are associated with better language skills and a more sophisticated level of

conversation from the children, greater general knowledge, and more participation in positive cooperative interactions between children as compared to classrooms with fewer adults per child (Vandell and Wolf, 2000). Howes et al. (1992) suggest that these findings stem from the fact that caregivers who are responsible for smaller groups have the ability to provide children with more individualized care and attention.

Non-experimental research has shown that preschool children under the care of adults with better education and training in child development obtain significantly better results on standardized cognitive tests, concentrate more easily on tasks, and communicate better with adults and among themselves (Hausfather et al., 1997; Vernon-Feagans et al., 1997; Howes, 1997; Dunn, 1993; Clarke-Stewart et al., 2002). The large-scale National Institute of Child Health and Human Development (NICHD) study mentioned above echoed these findings. It is more feasible to provide quality child care when the child-caregiver ratio is lower, when caregivers have received specific training in child development, and when caregivers have developmentally-appropriate ideas about early child care (NICDH, 2000a).



On the other hand, there is empirical evidence to suggest that structural elements alone do not guarantee high-quality care and, therefore, using them as the sole representation in the measurement of child care quality may be inadequate and even dangerous (Hamre and Pianta, 2007). A meta-analysis of the effectiveness of over 300 interventions based on improving quality through changes to structural aspects, such as improving the professional qualifications of teachers and reducing class size across the board, from preschool through high school, found little evidence that these improvements made a significant impact on children's academic achievement or social skills (Hanushek, 2003; NICHD, 2002b). Another review of interventions that improve structural aspects suggests that increasing inputs or strengthening the qualifications of caregivers has no impact on children's development in their early years of schooling (Hanushek, 1998). With that said, it is important to reiterate that most of this data comes from classrooms of preschoolers and elementary-aged children in the United States. Structural elements such as the child-caregiver ratio or infrastructure, sanitation and hygiene conditions may be of greater importance in the case of children living in contexts of chronic malnutrition and extreme poverty or very young children, whose immune systems are still developing (Fernald, Gertler and Neufeld, 2009).

2.3 / Process Variables

What are process variables?

Process indicators tend to focus on dynamic aspects (Thomason and La Paro, 2009), such as the interactions between children and their caregivers, the relationships among the children themselves, the caregivers'

behavior, implementation of the curriculum and health and safety protocols, and the relationship between the caregivers and the children's families (table 1).

How are they measured?

The measurement of process indicators requires observing and coding child care routines at the center over a certain number of hours. Some of the instruments that focus entirely on process variables are the CLASS Toddler and the Routine Assessment Tools, piloted in Spanish in Ecuador by the IDB. Additionally, instruments such as the ITERS-R, CC-IT-HOME, MITRCC, ORCE, CIS and the Assessment Profile also measure some process variables (see chapter 3 for more details about measurement).

While observational instruments that characterize child care processes are a vital diagnostic tool for measuring quality, most of the currently-used validated instruments that focus on processes add overall scores that describe children's experiences in relation to their interactions in various dimensions, and they almost always include structural variables as well (NICHD, 2000a; Zaslow et al., 2006). This is the case, for example, with the ITERS-R (with some sub-items on materials used), the CC-IT-HOME (with a scale on learning materials), the Assessment Profile (with measures of safety practices in the classroom), and the MITRCC (with some items on materials). Nevertheless, there are some validated process measures, such as the CLASS, ORCE and CIS, which focus only on very specific process activities (e.g., language stimulation) performed by caregivers (Vandell and Wolfe, 2000).

What does the literature say?

Process indicators that measure the quality of child care centers are much more consistently related to the quality of care and children's developmental outcomes than are structural indicators (La Paro et al., 2004; Hamre and Pianta, 2007). A

considerable number of experimental and quasi-experimental studies have found that children in preschool classrooms or high-quality child care environments (as measured by the quality of interactions between the children themselves and between children and their caregivers) are not only able to initiate and engage in higher-order learning with their peers but also to achieve higher scores on academic achievement tests (Campbell and Ramey, 1995). These effects persist over time and can last into adulthood (Greenberg, Domitrovich and Bumbarger, 2001; Hamre and Pianta, 2007; Shonkoff and Phillips, 2000; Peisner-Feinberg et al., 2001; Burchinal et al., 1997; Helburn, 1995; Howes and Hamilton, 1993; Kisker, Hofferth, Phillips and Farquhar, 1991; Kontos and Wilcox-Herzog, 1997; NICHD, 2000a).

Process indicators are those that have been shown to relate to everyday experience, learning outcomes and child development. Pianta (2003) examines a nationally representative sample of 224 preschool classrooms in six U.S. states and concludes that the characteristics of classroom quality most directly related to children’s learning outcomes at school or in related learning environments are the type and quality of interactions between children and caregivers. The same process variables that are vital to ensuring quality care for infants and toddlers are also the most difficult to measure because they require expert interpretation and opinion in order to be measured. Child care process quality is achieved neither through guidelines developed by authorities nor through the exclusive use of particular policies (Howes and Smith, 1995). The existing empirical evidence in the U.S. suggests that it is much more difficult to produce high-quality care—as measured by process indicators—for infants and toddlers than for preschoolers. In the U.S., child care services provided for younger children are

generally of lower quality (Helburn, 1995; NICDH, 2000a) and much less regulated than programs for older children. This situation is due to the fact that in the face of almost non-existent publicly-funded child care, regulation tends to be basic, and there is no emphasis on a policy to guarantee quality care. In short, these market characteristics allow for the development of a low-quality private sector without strict controls.

In short, for the youngest children, quality is more strongly associated with process variables, precisely those that governments have most neglected in terms of measurement and priority setting (Howes et al., 1992).⁷

2.4 / How Do Governments Measure Child Care Quality?

In general, programs and/or governments regulate and measure the quality of child care for infants and toddlers through structural indicators such as basic infrastructure and the professional profile of caregivers and/or teachers (Howes et al., 1995); however, these types of indicators do not capture information about the processes that affect the child’s cognitive and emotional development as well as health. Neither they help on the identification of areas of improvement in these processes and the monitoring of changes over time (Schütz, Ursprung and Woessmann, 2008).

This section discusses the instruments used to monitor quality at three public child care services: one, part of a national program,

⁷ Nonetheless, the existing literature on infant and toddler child care suggests that structural components of quality can also be very important in the age range of 0 to 3 years, more so than for other age groups. For example, in a meta-analysis of the past three decades, Vandell and Wolfe (2000) find that both process and structural quality indicators are consistently associated with children’s outcomes in the short term.

in Mexico; one in the city of Buenos Aires (CABA), Argentina; and a third, part of a national program, in Ecuador. In Mexico, the Routine Monitoring Form is a monitoring instrument that must be completed on a bimonthly basis by program staff at each office of the Ministry of Social Development (SEDESOL) for all of the daycare facilities operating under its supervision. In CABA, the staff responsible for overseeing the program collects information on all of the centers every two months. In Ecuador, the instrument was completed by program staff twice and only for a sample of the centers.

The three countries routinely submit, with some regularity, their programs to a questionnaire or checklist that includes questions on the topics mentioned in Table 2. As can be seen in this table, safety issues are being monitored using the largest number of items (with 70% of items monitoring these issues in Ecuador and almost 58% of all issues in Mexico). In second place, monitoring of information about the staff occurs in Buenos Aires with 30% of the total items, while the conditions of the center property are monitored both in Mexico and Ecuador (12.6% of the items in Ecuador and 6.5% in Mexico). Equipment, orderliness and hygiene rank third in importance in Buenos Aires and in Mexico, while center information ranks third in Ecuador. It is important to note that Chile's National Board of Daycare Centers (JUNJI) also has a well-developed system of evaluation for its daycares and preschools.

KEY DEFINITIONS FOR MEASUREMENT

● CHAPTER **3**



3.1/ Introduction

As explained in the previous chapter, measuring the quality of child care for infants and toddlers involves key issues, including the selection of structural and process variables that measure quality and how these variables actually reflect what one is trying to measure.

This chapter will detail this process and explain how to conduct these measurements. In particular, it discusses the importance of defining the goal of the measurement and of identifying the study's unit of analysis (children in one group, groups of children, or centers). In addition, this chapter introduces a battery of instruments and proposes some guiding principles on how to select the most appropriate instrument.

3.2 / What is the Goal of Measuring Child Care Quality?

The first step to address is to concretely define the purpose of the measurement. The measurement of child care quality may be guided by a number of goals, including monitoring and/or follow-up of the status of children in the country, evaluation of the status of program beneficiaries, research on the impact of the child care service on certain indicators, accountability with regard to the use of public resources that fund the service, and continuous monitoring of process quality for improvement. The objective of the measurement will determine who, how, and what is to be measured.

It is important to highlight that, when the intention is to monitor compliance with minimum standards of quality and how this changes over time, the measurement should be closely tied to the standard set by the program. For example, an infant and toddler daycare program that includes neither the provision of nutritional supplements or micronutrients to children nor the tracking of their nutritional status in its model of care cannot be monitored on the basis of an indicator that describes the nutritional status of the children it serves.

3.3/ What Is the Study Group?

The second question of interest involves defining the target population to be measured. The measurement can be aimed at all of the children of a particular country, geographical region or ethnicity, as well as a group of children who benefit from a specific program. Once the population of interest is selected, researchers must decide whether it is necessary to observe the entire population or just a sample will suffice for the objectives of the assessment. For example, if a researcher's interest in the measurement of service quality lies in assessing the condition of the service provided to the country's child population, then it is important to measure a sample large enough to be representative of the child care centers serving relevant groups in that particular context (e.g., region, ethnic group or socioeconomic status). If, however, one wants to monitor the condition and development of the centers that participate in a program, then interest lies in measuring *each one* of the service providers. This could allow for the creation of a system with monitoring and tracking indicators. When the goal is diagnostic (for example, the identification of critical aspects of service quality improvement), it would probably be

necessary to select a representative sample of the centers offering the type of services that one is looking to improve and to invest in a set of instruments that would allow for a deep characterization of the current level of service quality at these particular centers. In sample surveys, sample size should be determined based on statistical power calculations for determining the minimum sample size required to statistically identify an effect of a given magnitude with an acceptable level of precision.

Once the target population of the assessment has been clearly identified, it is necessary to determine the unit of analysis at each center that will be studied. Is it the child care center? One or more groups of children attending that center and their caregiver(s)? Each of the children who attends it? Clearly, the goal of the study, together with the available budget, will guide this decision.

If child outcomes are to be measured in addition to the quality of the center, then the instruments described in this toolkit should be complemented by instruments for measuring child development. This is an example in which the study's unit of analysis is the child, and the quality of care he or she receives at the center constitutes an explanatory variable of child development. If, however, the study focuses exclusively on collecting quality measurements at centers, then the selection of a group and caregiver at each center would open up the possibility of a wider range of instruments. In the event that centers have more than one group, group selection should be conducted randomly, unless there are valid reasons for not doing so (e.g., the number of children in each class falls below the study requirement, children fall outside the study's target age range, etc.). When the goal of the quality measurement is to analyze differences in quality between different caregivers through fixed effects at the center, at least two groups per center must be selected.

3.4/ **Measuring: the how**

Collecting data on the quality of a child care facility can be accomplished through two types of mechanisms. The first consists of an evaluation based on observations made by a person trained to systematize the information obtained through an instrument. The second consists of an evaluation based on the report of an informant who is administered a survey or interview (e.g., it may be the child's mother, a center caregiver, or the center's director). Some measurement tools combine the two modes of data collection.

Measurements acquired through observation require that an observer spend a specific amount of time (usually more than an hour) with the children during their daily routine in order to collect information on the dimensions of care being evaluated and to systematize it according to the protocol used by the instrument. The observer must be trained in the use of the instrument and capable of capturing the activities, bonds, stimuli and interactions that must be reported, without becoming distracted by inconsequential details. In the case of data collected through self-report, a structured interview is conducted or a survey is administered to a qualified informant, using a list of questions on the dimensions of quality that the researcher wants to evaluate. Self-reports on children's behavior, or characteristics of the caregivers or the child care center can be included as well.

The assessment can be enhanced by the two data collection mechanisms. Implementing different forms of measurement for the same dimension will provide more data and greater precision than data obtained by just one method.

In turn, instruments that assess quality through observations of the daily routine can be used with direct observation by visiting the center or filming the experience for later coding. In this way, instead of simultaneously observing and scoring, one person takes responsibility for filming the routine of the child or group of children and then the footage is analyzed at another time. The results obtained through filming will depend on both the characteristics of the footage and the protocol established for how to film subjects, as well as those of the observer.

It is worth noting that, regardless of the type of study to be implemented, when selecting which instrument to use, it is necessary to explore the validity of that instrument for the population under study. This involves examining whether the instrument is appropriate for the population in terms of cultural relevance and language or if, instead, translation or adaptation of the assessment to the language or local context will be necessary (this issue is discussed in detail in chapter 4).

Observation vs. self-report

Observation demands more time, training and resources than the implementation of checklists or questionnaires based on the self-report of, for example, a caregiver or center coordinator. The main advantage of observational instruments is that they describe the experiences and interactions of children at the center. The presence of an observer at the center could, in some way, distort the interaction between caregiver and child (despite taking all necessary precautions to avoid interfering with activities, as detailed in chapter 4). However, it is considered that the longer the duration of the observer's visit at the child care center, the more difficult it will be for the caregiver to inhibit the tendencies and reactions that form part of her daily routine, especially after the first hour of the visit.

Observational assessments are more complex because the observer must be able to document, code or assign a score for each dimension, while maintaining an accurate, objective view. For this reason, personnel trained in this type of assessment are required, with accurate knowledge of the instrument to be administered (e.g., features, questions or items to be answered, time



needed for the assessment). In addition, it is necessary to ensure the reliability of the measurements collected. In other words, first, the results of the assessment must be similar, regardless of which observer conducted it, and second, the instrument must assign a consistent quality ranking to the child care center, even if the center is evaluated twice within a short period of time (e.g., a medium-quality child care center is consistently medium quality, whether it is evaluated today or in two weeks).

Observational measurements are time-intensive because they require being present during the children's experiences for a specific amount of time (usually more than one hour), during which the necessary data is obtained and coded. Instruments used for observational measurements also require a longer training period in order for observers to become familiar with the tools. In addition, a fair amount of practice is needed before starting field measurements to ensure reliability among observers. One risk is that these measurements may not be representative of a typical day. For example, the child-caregiver ratio can vary from one class to another, over the course of the day (children in full time vs. part-time attendance), on different days of the week, or in different seasons (Weber and Kariger, 2011). These measurements can also be more susceptible to problems such as measurement error and observer bias (Zaslow et al., 2006; NICHD, 1996). Constant, intensive training throughout the coding process and reliability exercises are measures that can be taken to mitigate these risks.

It is important to mention that, when using observational instruments (as opposed to those for self-reporting), it is vital for the time at which the observation is made to always be the same across all of the child care centers observed, given the very specific routines of young children. In other words, an observation made in the morning cannot be compared to an observation made in the afternoon, when the caregivers and children

are more tired. Other moments that form part of the daily routine and that reveal service quality are arrival time (including how children are greeted at the center), mealtimes and toileting. This requires the observer to arrive at the center before it opens. In order to observe mealtimes, the observer must plan to be present at breakfast and/or lunch. Naptime must also be considered, as there will be nothing to observe for about two hours; this means that, in general, observations are conducted between 8:00am and 1:00pm. This also means that an observer cannot cover more than one center per day.

In the case of self-reporting instruments, data is collected directly from the questions answered by the child's mother or caregiver; therefore, there is an advantage in terms of ease of administration, speed, lower cost, and a less-demanding training requirement. Self-report, however, can be subject to biases. For example, a mother has partial information about what happens at the center where she leaves her child; a caregiver is unwilling to share information that might reveal that the attention provided to the children in her care is not optimal; or a person with a low level of education may be unable to accurately report on the frequency with which she carries out activities with the child.

Empirical evidence shows that direct observation instruments used to evaluate service quality predict child outcomes to a significantly greater degree than interviews or checklists; therefore, observational measurements should be an integral part of any effort to ensure quality child care (Zaslow et al., 2006). Furthermore, as discussed in detail below, it is necessary to maintain reliability over time, and therefore, when surveys are conducted, reliability exercises are performed every so often to ensure the minimization of bias during this process.

Filming

As mentioned above, the type of data obtained through classroom observation or filming is characterized by a higher level of objectivity than that obtained by self-report assessments. Additionally, the advantage of filming is that the coder focuses on what is happening with the child, separating him from the activity in his surroundings (e.g., other groups of children that are not part of the sample or the activities of support staff), thereby allowing for a more accurate rating. The results, however, depend almost entirely on the manner in which the video was shot (i.e., whether the video focuses on the group of interest, audio and image quality, and a camera position that allows for proper observation of the children's activities). In other words, the quality of data obtained will depend on how well the visual and audio information from the daily routines is captured in the video. Moreover, there is always a risk that valuable information relevant to the measurement of process quality will take place off camera.

One of the main advantages of filming is that it allows for a more objective coding system, with different observers coding the video in multiple passes. In this way, if a mismatch occurs between the scores assigned by different observers, it is always possible to submit the video to evaluation by a third observer in order to identify the source of the discrepancy. An additional advantage of filming is that the video can be viewed as often as necessary and re-coded, or observers can code interactions that went unnoticed in the first pass of the video. It also allows for an analysis of the data in a more comfortable, controlled environment, eliminating the need for trained observers to travel to the child care centers. This can be an important factor, since many times the personnel most qualified to perform these observations have fewer incentives to travel to isolated areas and incur the personal costs that such travel represents.

Clearly, filming may be the more expensive option. While filming may be performed by the same person who later analyzes the video, it requires a greater number of work hours (filming, editing and then analysis). One option that is more cost-effective is to use two different people: one to shoot and edit the video (with a lower educational profile) and another to code (with a higher educational profile). The minimum equipment required for filming must be available, including cameras, microphones, software and computers to edit and play back the videos being analyzed. This approach also requires that personnel receive training on the correct procedures for using the camera, as well as training on the instrument.

Nonetheless, if the budget allows, direct observation and filming can be used together to obtain results with greater reliability.

3.5/ What Instruments Are Available?

The evidence reviewed thus far suggests that it is important to measure both structural and process elements in order to have a complete and reliable measurement of child care quality. As mentioned earlier, several instruments measure both aspects of quality. It is also important to note that instruments have been developed for assessing the quality of care and attention given to children in other settings, for example, at home with their parents or with a babysitter.


In the last two decades, the most commonly used instruments for measuring child care process quality, in U.S. child care centers are as follows:

- The Early Childhood Environment Rating Scale – Revised Edition (ECERS-R) and the Infant/Toddler Environment Rating Scale – Revised Edition (ITERS-R)

- The Classroom Assessment Scoring System (CLASS), which exists in several versions. The CLASS Pre-K and K-3rd. grade versions of the instrument have been used the longest. Versions for infants and toddlers, CLASS Infant and CLASS Toddler, respectively, have been developed more recently.⁸ CLASS Toddler is the version best suited to the age group considered in this guide.
- The Observational Record of the Caregiving Environment (ORCE)
- The Arnett Caregiver Interaction Scale (CIS)
- The Assessment Profile for Early Childhood Programs and the Assessment Profile for Family Child Care Homes (APECP/APFCCH)
- The Missouri Infant/Toddler Responsive Caregiving Checklist (MITRCC)
- The Child Care Infant/Toddler Home Observation for Measurement of the Environment (CC-IT-HOME)
- Feeding Scale, Nursing Child Assessment Satellite Training (NCAST), 2012⁹.

Before proceeding with the presentation of each tool, it is worth mentioning that the authors of this toolkit worked with some of the instruments mentioned above.¹⁰ In order to study the quality of the child care provided at *Centros Infantiles del Buen Vivir* (CIBV) in Ecuador in 2012 (Araujo et al., 2015), a battery of instruments was administered for the evaluation of both structural and process variables. The instruments administered at the 400 child care centers were the original Spanish version of the ITERS-R and translated versions of the CLASS Toddler, MITRCC and CC-IT-HOME. The Spanish-language translations were commissioned

from specialists in the field. ITERS-R training instructed observers to disregard stop rules, recommending instead that they score all items (this is the alternative scoring option that appears in the original Spanish version of the manual, page 6, Harms et al., 2006). Furthermore, it was necessary to adapt the MITRCC and the CC-IT-HOME (a process performed by NIEER) to the context of the study. Six items from the MITRCC that were not applicable in the Latin American context were removed, and one item was split into six items to make it easier to administer. The CC-IT-HOME was adapted for use in child care centers, resulting in the elimination of one item. The authors also designed and piloted daily routine scales in order to characterize some of the most important child care routines that occur at centers.

3.5.1/ **ECERS-R and ITERS-R**


Description

There is a family of instruments—the ECERS-R, ITERS-R and the FCCERS-R (Family Child Care Rating Scale)—that measures child care quality in different age groups (the ITERS-R for younger children and the ECERS-R for older children) or in different contexts (ECERS-R/ITERS-R for child care centers and FCCERS-R for in-home care).

The ECERS-R is designed for children between the ages of 2½ and 5 in center-based settings, while the ITERS-R is for

⁸ An upper elementary version for children in grades four through six has also been developed (CLASS Upper Elementary), and a version of the tool for secondary school classrooms is in development.

⁹ The authors adapted this instrument to create the “feeding assessment tool” presented in subsection 3.5.8. It is important to mention that this tool differs from the others in the sense that it is not meant to be a “global” quality instrument, but rather focus on a particular process, in this case a routine.

¹⁰ The questionnaires used by the authors in Ecuador were not included in order to keep the document to a reasonable length; however, they are available to interested readers. Please contact the authors at mcaraujo@iadb.org or florencial@iadb.org.

children ages 0 to 2½ years. They both evaluate seven dimensions. One of these dimensions focuses on the provision of space and furnishings, while the other six are process-related. One of the process dimensions specifically targets child-caregiver interactions. The instruments collect detailed information on center infrastructure, health and nutrition protocols, safety factors, the availability and accessibility of learning materials, the types of activities carried out during the day, etc. (Cryer et al., 2004a,b). The evaluation, consisting of observations and scoring, takes three hours to perform. Some of the items are completed through an interview with the caregivers. Each of the 43 items that compose the ITERS-R's seven dimensions is scored on a 7-point scale, with indicators for inadequate quality (1), minimal quality (3), good quality (5), and excellent quality (7). The 43 items that compose the ECERS-R are scored in the same way. Table 3 presents the ITERS-R's main features.

Experiences of use

The ECERS-R was widely used with a sample of children ages 2 to 4 at child care centers in the U.S. (Burchinal, Kainz and Cai, 2011). In LAC, four countries have implemented at least one of

the instruments from this family (see Table 3). The Colombian program *Hogares Comunitarios* was evaluated in 2007 using the FCCERS-R instrument to measure the quality of the services offered to children through the program. This program provides community-based child care, with a mother in the neighborhood caring for other women's children in her own home. The program has very broad coverage and has served the country's poorest children since 1986 (Bernal et al., 2009). One interesting finding of the assessment was that, despite the relative success of the program in terms of improving some cognitive and health outcomes in children, the quality of the services, as measured by the scores yielded on these scales on the structural and process aspects of the child care settings, was, on average, very low (a score of around 2 points). This finding gave way to a number of initiatives to improve service quality (Bernal et al., 2009). In addition, two projects using this instrument are under way in Colombia.

In 2009, a team of researchers used the ECERS-R and ITERS-R to assess the quality of child care and preschool services in Brazil. Given the heterogeneity in the quality of the child care evaluated in this study, the scale was adapted to the Brazilian context,



extending the range at the lower end of the scale to capture the variability in the lowest quality segment of the spectrum (Verdisco and Pérez Alfaro, 2010). The analysis allowed researchers to conclude that, although there is great variability in the quality of child care centers, it is the poorest children who generally receive lower-quality child care services (Verdisco and Pérez Alfaro, 2010).

The ECERS-R and ITERS-R were similarly adapted for use in Chile and were administered in infant and toddler classrooms at public and private daycare centers in order to test their validity in the Chilean context and to better understand the general quality of the early learning environments of children in that country (Herrera et al., 2005). The research team identified some psychometric characteristics of the instruments that were applicable to the Chilean context. Results from the administration of the ECERS-R and ITERS-R revealed that the quality of the centers evaluated was quite variable, and the average score across the entire sample was very low, particularly among children ages 0 to 3. Lastly, improvements in the quality of early childhood programs in Chile are closely linked to improved outcomes for children (Herrera et al., 2005). The authors of this toolkit administered the ITERS-R in Ecuador (Araujo et al., 2015), and the ITERS-R has also been used as a baseline for an impact assessment of the daycare service provided through Peru's nationwide program *Cuna Más* and is currently being considered in Uruguay.

3.5.2/ **Toddler CLASS**

Description

Of the instruments mentioned above, only CLASS focuses exclusively on the measurement of process variables. All the different versions of the CLASS instrument

—CLASS Pre-K (for child care settings with children ages 36 months until kindergarten age), CLASS Infant (children ages 0 to 18 months), and CLASS Toddler (children ages 15 to 36 months)—were designed to address the limitations of other instruments focused on structural variables. By focusing on process variables, these instruments assess “what teachers do with the materials they have and the interactions they have with children” (La Paro et al., 2004, p. 412).

It is worth highlighting that the dimensions that apply to the evaluation of interactions between teachers (or caregivers, in the case of CLASS Toddler) and children are similar throughout infant child care, preschool and elementary education; however, differences make themselves evident in the ways that these dimensions are manifested in each age group (La Paro et al., 2011). For this reason, the CLASS Toddler version adapts the descriptions of items to the specific circumstances found in toddler and preschool classrooms.

In the CLASS Toddler version interactions are scored on the basis of observations in eight dimensions contained within two domains: emotional and behavioral support and engaged support for learning. The first domain includes five dimensions that focus on the expression of emotions (both positive and negative) of caregivers and children, responsiveness, the caregiver's availability and sensitivity, the degree to which children's perspectives are taken into account and responsibility and independence are fostered, and support in the development of behavior regulation. The second domain is aimed at observing the caregiver in her role as a facilitator of activities designed to support children's learning and to promote language development, as well as the degree to which feedback is provided, based on comprehension and participation (La Paro et al., 2011).

A numerical rating from 1 to 7 is assigned to each dimension. The assigned score

depends on a series of assessments made by the observer, guided by the manual and knowledge gained during training so as to maintain objectivity and accuracy. The quality of the assessment is ensured through training and certification of personnel.

Thomason and La Paro (2009) analyze the validity of the CLASS Toddler on the basis of a study of 46 teachers in 30 different toddler and pre-school classrooms in the United States. The construct validity of the CLASS Toddler is based on comprehensive reviews of existing measurements, a review of the research on the unique aspects of early childhood development, and observations of infant/toddler and preschool settings. The instruments reviewed to establish the validity include the ITERS-R (Harms et al., 2006), the CIS (Arnett, 1989), and the ORCE used in the NICHD “Study of Early Child Care” (NICHD, 1996). The examples and indicators included in the adapted instrument reflect this review. Similar to the CLASS Pre-K (Pianta et al., 2008), examples are provided for the scoring of centers, with ratings of low (1-2), medium (3-5) or high (6-7). Additionally, the instrument was reviewed by an expert in infant/toddler development to ensure the validity of the adapted concepts.

Significant correlations were found between traditional measurements of quality in early childhood education, as well as the education level of the caregiver, group size, the child-caregiver ratio, the scores assigned (based on the inclusion of quality aspects in the program and the education level of caregivers) and dimensions. See annex 1 for more details on the CLASS Pre-K’s validity (Pianta et al., 2005). Table 4 presents the instrument’s salient characteristics.

Experiences of use

There is a meta-analysis of 20 studies published in professional journals that went through a process of external review by experts. These studies measure child care

quality and child development outcomes, using a sample of children ages 0 to 4 at child care centers in the United States during the last decade. Indeed, Burchinal, Kainz and Cai (2011) found that the CLASS Pre-K offered slightly more consistent associations with child development outcomes than the instrument that had been the most widely used to date, the ECERS-R. The individual items from the CLASS Pre-K that best predicted child outcomes were the dimensions of behavior management, productivity, positive climate, and negative climate. It should be noted that the CLASS Infant version is currently under development, so the data and evidence of its use in children under 18 months of age are limited (Teachstone, 2012c).

The U.S.’s well-known national Head Start (NHS) program currently uses the CLASS Pre-K and CLASS Toddler as tools for staff monitoring and professional development. This effort involves the training of over 200 NHS monitoring specialists from across the country, who use the CLASS to monitor quality and provide training on the tool to local program managers so that they, in turn, can monitor quality and set goals for improvement in their own programs (Teachstone, 2012a). The CLASS Pre-K primarily serves to provide a common vocabulary for educators, program managers and policymakers so that they can identify, discuss and monitor quality issues at various programs throughout the country. In the state of Virginia (U.S.), the CLASS Pre-K is being used as part of a statewide accreditation and monitoring system for early childhood education programs. The use of CLASS allows parents and policymakers to easily distinguish aspects related to the quality of the facilities. It also makes it possible to monitor changes in quality at the various programs and offers guidance to educators and caregivers on how to improve the quality of their services (Teachstone, 2012b).

In Santiago, Chile, the CLASS Pre-K is being used as a monitoring and professional

development tool (Yoshikawa et al., 2008). In 2008, the instrument was first administered by the pilot program *Un buen comienzo* (UBC) as part of a comprehensive professional development effort designed to improve the quality of preschool education and to improve children’s cognitive, social and health development as well as parental involvement in school. The CLASS Pre-K was used to monitor changes in quality throughout the course of the intervention, and it is currently administered at a number of additional pilot programs in other regions (Yoshikawa et al., 2008). In Ecuador, the CLASS K-3 has been widely administered to a sample of about 600 kindergarten and first-grade classrooms as part of a joint study by the Inter-American Development Bank and the Ecuadorian government on the quality of early childhood education. It is also being implemented in Peru (CLASS Toddler). The authors of this toolkit implemented the CLASS Toddler in Ecuador (Araujo et al., 2015).

3.5.3/ ORCE

Description

The ORCE has been used in a wide variety of early childhood settings with children between the ages of 0 and 60 months, and it has a strong focus on the quality of interactions between the child and the caregiver (or other adults present). This instrument seeks to characterize the interactions experienced individually by each one of the children in the classroom (Burchinal, Kainz and Cai, 2011; Vandell and Wolfe, 2000). It is important to highlight that the ORCE’s process quality section focuses on a specific child (the “target child”), while structural aspects are evaluated on the basis of the group. Ratings are made on a 4-point scale for each item. The total score is derived from the average score among the instrument’s

three domains for all observation cycles (Vandell and Wolfe, 2000).

In terms of the ORCE’s construct validity, the child care quality measurements that this instrument yields were positively related to structural variables such as the caregiver’s level of education and the child-caregiver ratio. With regard to predictive validity, these child care quality measurements were related to cognitive performance, language and social activity (Halle, Whittaker and Anderson, 2010). Table 5 presents the instrument’s salient characteristics.

Experiences of use

The ORCE was developed for a large-scale study on early childhood care conducted by the NICHD in the U.S. This study sought to better understand the relationship between early childhood care and child development outcomes (ages 0 to 3 years) in various types of early childhood settings (Burchinal, Kainz and Cai, 2011). This instrument was specifically designed to address two limitations of the ECERS-R and ITERS-R. Firstly, those instruments are too focused on structural characteristics, to the detriment of capturing the minute-by-minute experience and the proximity of the child to his caregiver. Secondly, the ECERS-R and ITERS-R were only appropriate for institutional settings such as child care centers (NICHD, 2009; Vandell and Wolfe, 2000), whereas the ORCE could be implemented in other child care settings such as the home. In their review of the literature, the authors found no references to the implementation of the ORCE outside the U.S., and it has not been implemented in Latin America.

3.5.4/ CIS

Description

The CIS consists of a reduced checklist that evaluates caregivers on their “sensitivity, harshness and detachment” in interactions with preschoolers in both home-based and center-based care. Scoring is based on the frequency with which certain caregiver behaviors are observed, focusing mainly on the structural characteristics of child care and the interactions between caregivers and children (Vandell and Wolfe, 2000; Burchinal, Kainz and Cai, 2011). Observers assign a score of 1 (not at all) to 4 (very much) for each of the 26 items in the scale’s four dimensions.

The scale does not address issues of curriculum or other classroom management strategies, such as grouping children according to their age or the flow of activities (Denham, Ji and Hamre, 2010). Table 6 presents the instrument’s salient characteristics.

Experiences of use

The CIS was implemented in Bermuda (Arnett, 1989) and in several studies at child care centers in Philadelphia (Jaeger and Funk, 2001) and other U.S. cities. This instrument has always shown child care quality levels below the minimums established by the standards of the program.

3.5.5/ **Assessment Profile****Description**

The Assessment Profile for Early Childhood Programs (APECP) is a tool that, like the others, evaluates the practices of a child care center or classroom. This instrument has a summative and a formative version. The summative evaluation assesses the learning environment and teaching practices in classrooms with children between the ages of 3 and 7, while the formative evaluation is used to improve early childhood programs. In addition to assessing dimensions such as classroom practices, the learning environment, planning, curriculum methods, interactions, and individualization, the formative evaluation also assesses administrative aspects, including facilities, food service, policies and procedures, staff, and professionalism, evaluation and development. This measure is broader than the summative evaluation, it is easy to use as part of the self-evaluation process for programs for young children (ages 0 to 10) since the items are simple to administer, and it also generates an extensive analysis and detailed recommendations for program improvement. Each of the items in the instrument’s five dimensions is scored on a yes/no basis, and the positive responses are tallied.

The instrument’s criterion validity was established by examining the relationship of the APECP with the ECERS (Harms and Clifford, 1980). In these studies, Wilkes (1989) found a significant correlation ($r = 0.64$, $p = 0.001$), and Abbott-Shim (1991) also found a significant correlation ($r = 0.74$, $p = 0.001$).

To test for construct validity, second-order factor analysis was performed to determine whether the APECP’s five scales formed a single quality construct in the classroom.

The results indicated that the measurements observed using these factors do come from a single underlying quality construct in the classroom (Abbott-Shim, Lambert and McCarty, 2000).

In turn, to establish content validity, the instrument was reviewed by a broad range of early childhood professionals and cross-referenced with accreditation criteria from the National Association for the Education of Young Children (NAEYC). Cross-referencing revealed marked consistency between the two measures, with a 100% match in the criteria (Abbott-Shim, Neel and Sibley, 2001). Table 7 presents the instrument’s salient characteristics.

Experiences of use

This scale was implemented in several U.S. states. Given their focus on aspects of children’s learning environment as well as interactions, Denham, Ji and Hamre (2010) identified the CLASS, ECERS-R, ITERS-R and Assessment Profile as tools that simultaneously evaluate the presence of critical elements for the healthy social-emotional and cognitive development of young children.

3.5.6/ MITRCC

Description

Another instrument that utilizes a checklist format is the Missouri Infant/Toddler Responsive Caregiving Checklist (MITRCC). It focuses on children under the age of 3 and can be used in center-based or home-based programs.

This instrument assesses the work of the caregiver: recognizing and responding to children’s cues; teaching children to anticipate routines and events; encouraging positive peer interactions and modeling empathy; encouraging children to recognize their emotions and solve their own problems; and stimulating fine and gross motor skill development and cognitive development. Additionally, the instrument includes the observation of structural variables with items, such as “caregivers organize the classroom so that children have an opportunity to observe their surroundings from more than one level,” “there are similar and different objects



available for children to compare and contrast,” and “there are materials to encourage pouring and dumping.” Each item is scored on a yes/no basis (the interaction must occur with 75% of the children who are awake and alert); the number of yes responses is divided by two in order to arrive at an overall score that falls between 0 and 10. Unlike the other instruments analyzed, the MITRCC’s items do not refer to the type of quality characteristics being evaluated but instead the areas of child development (social-emotional, physical and cognitive) being promoted through quality interactions. This is an important distinction because it underlies an assumption (or model) that different variables of child care quality contribute to different dimensions of child development. Table 8 presents the instrument’s salient characteristics.

Experiences of use

The MITRCC has been used in the U.S. (Missouri Head Start - State Collaboration Office, ZERO TO THREE), as well as in the evaluation of the *aeioTU* program in Colombia (NIEER and Bernal, work in progress). The authors of this toolkit also implemented it in Ecuador, and the results are available in Araujo et al. (2015) and in Lopez Boo and Tome (forthcoming, 2016).

3.5.7/ CC-IT-HOME

Description

For child care provided in the home setting, the HOME inventory is a quality measure that focuses on process variables but also evaluates the presence of structural variables.

An adapted version of the original HOME inventory (the Child Care Infant/Toddler HOME, CC-IT-HOME) is available for the evaluation of family child care in home settings, such as relative care or “baby sitters.” In Ecuador, the authors of this toolkit used a version of the CC-IT-HOME that was translated to Spanish and adapted by NIEER for use in center-based environments. The CC-IT-HOME employs a checklist approach to assess child care and includes the scoring of elements such as the safety of the environment, health practices, and the materials and activities provided.

The instrument combines observation with the information provided by the parents (or the caregiver in the center-based version). One particular feature of this inventory is that the unit of observation is not a group of children but rather the individual interaction between *one child and his caregiver*. In other words, use of this inventory requires that the observation focus on an individual child, with a detailed examination of the child care setting from that child’s perspective. This information is then generalized to all of the children in that setting. Each item is scored using a yes/no format, and the maximum total score is 42 points. The instrument is divided into six dimensions. Table 9 presents the instrument’s salient characteristics.

Experiences of use

The CC-IT-HOME has been widely used in the U.S. and the Caribbean. Clarke-Stewart et al. (2002) cite a study that shows that children who were cared for in higher-quality homes, according to the CC-IT-HOME, also scored higher on the Bayley Scales of Infant and Toddler Development¹¹ at age 2, and at age 3, they had significantly higher scores on school readiness as compared to their peers who received lower-

¹¹ There are currently three editions of the Bayley Scales of Infant Development: BSID-I, 1st edition; BSID-II, 2nd edition; BSID-III, 3rd edition. The Bayley Scales directly evaluate the development of children ages 1 to 42 months in a broad range of domains. The tool is used to diagnose developmental delays and to plan appropriate interventions, and it has been validated. The Bayley is considered the gold standard in measuring child development.

quality care in their homes. The authors of this toolkit implemented in Ecuador an adapted version of the CC-IT-HOME for use in child care centers (Araujo et al., 2015).

3.5.8/ Other instruments

Routine assessment tools

The authors of this toolkit piloted a set of observational scales in Ecuador, which assess, through observation and coding, the characteristics of the main child care routines at the center. The goal of this effort was to examine whether it was possible to simplify the observational instruments used to assess process quality, with the aim of creating measures that can be implemented at a lower cost (in terms of training, time, and staff profile) while still providing data on these dimensions of quality. The authors identified key moments in which it was possible for a caregiver to demonstrate individualized interactions with sensitivity toward the children in their care. The key moments in the daily routine that were selected for evaluation by these instruments include the following: (a) mealtimes; (b) diapering or toileting, depending on age; and (c) the management of disruptive situations throughout the day such as one or more crying children; episodes of verbal, physical or psychological abuse; or the disconnection of a child from his surroundings and the people around him.

These rating scales for child care routines have been used to measure the quality of interactions between parents and children and caregivers and children in a variety of contexts. These scales assess interactions and communication between children and parents/caregivers, during either feeding (for children ages 0 to 12 months) or teaching (for children between the ages of 0 and 36

months). These scales have the advantage that they can be administered relatively quickly; the feeding scale only requires the time span of a meal (NCAST, 2012).

The feeding assessment tool piloted in Ecuador was adapted by the IDB team from the Feeding Scale (NCAST, 2012) administered in the U.S. Basically, the frequency and duration of certain reactions (both positive and negative) by the caregiver were observed during the meal. The original scale has parameters for internal, external and predictive validity plus adequate reliability (NCAST, 2012). The instrument piloted by the authors to characterize interactions during diapering or toileting is based on the feeding assessment tool, and it focuses on observing and assessing the caregiver's reactions during this particular event, as is done during a meal.

The third instrument piloted in Ecuador, which focused on observing moments of crying, abuse, or disconnection, is based on several existing instruments: the Infant-Toddler and Family Instrument (Provence and Apfel, 2001), the Family Care Instrument (FCI), and the Regional Project on Child Development Indicators (PRIDI) instrument (Engle, Cueto, Ortiz and Verdisco, 2011). The instrument requires researchers to note the frequency and duration of any of these three events (crying, caregiver abuse or abuse among children, and disconnection) over a 40-minute period and to observe the actions of the caregiver when these events occur. Using a minimal number of observational items or questions, the instrument aims to assess the manner in which the caregiver disciplines the child or reacts to a stressful situation. One difficulty posed by this instrument is that it may require a longer observation period in order to actually witness a situation of this nature. In quality child care settings, one can imagine that circumstances of disconnection or abuse are quite rare, and the instrument focuses solely on characterizing the reactions of adults during instances of crying.

All of the routine assessment tools originated in and were refined with input from the IDB Advisory Committee (2012). The authors of this toolkit piloted these instruments in Ecuador in 2012. The analysis of the implementation and functionality of these scales is in process at the time of writing this guide. Table 10 presents the instrument's salient characteristics.

New instruments

Other quality measures worth mentioning for their relatively frequent use over the last two decades are the Program for Infant/Toddler Care Program Assessment Rating Scale (PITC-PARS) and the Preschool Program Quality Assessment (PQA). These are both validated instruments that, to varying degrees, measure process and structural elements of infant and toddler care, and they have been implemented to improve, monitor or give accreditation to a program for research/evaluation purposes at U.S. child care centers (Denham, Ji and Hamre, 2010; Halle, Whittaker and Anderson, 2010). One new measure, the Quality of Caregiver-Child Interaction for Infants and Toddlers (Q-CCITT), is currently in development and being piloted at 400 U.S. child care centers. Funded by the Office of Head Start, the Q-CCITT promises to be a psychometric indicator sufficiently broad as to make comparisons among a variety of childcare settings where parents are not present. The introduction of this new instrument seeks to overcome the limitations of most other instruments that are setting-, age- and language-specific (Halle et al., 2011). The Q-CCITT is unique in its specific focus on process variables, its application in different child care settings (the home, child care centers, classrooms divided by age or multi-age classrooms, with a focus on children ages 0 to 3), and its coverage of classrooms serving children with

disabilities and those whose mother tongue is different from the language used at the child care center (Halle et al., 2011).

3.6/ The use of secondary surveys

This section documents the specific experience of the authors in the Ecuador study (Araujo et al., 2015). The authors of this toolkit supplemented their study in Ecuador with surveys that were directly administered to the caregiver of the group of children (40-minute interview) and the coordinator of the center under study (one-hour interview).¹² In both cases, demographic information was collected from the respondent, and the researcher inquired about her training, experience, continuing education and salary. In addition, both caregivers and coordinators responded to the Knowledge of Infant Development Inventory (KIDI) (MacPhee, 1981), which seeks to identify how much adults know about different normal behaviors for each age group for children under 5.

The interview with the coordinator also served to collect general information about the center's characteristics, its administrative organization and its staff, namely: data on all staff working at the center (level of education, experience, years worked at the center, pay); information on the center's organization (group composition, organization, child-caregiver ratios); parent-teacher communication and interaction; and the inclusion of children from ethnic minorities or with disabilities.

Lastly, a questionnaire focused on structural elements was administered. An instrument focused entirely on the observation of

¹² The center coordinator is responsible for leading, coordinating and promoting a learning community aimed at children. This person is responsible for proposing a program of activities tailored to the interests and educational needs of the children, as well as resolving any disputes that arise, with the goal of achieving an appropriate environment.

structural variables was developed on the basis of a structural survey by Ecuador's Institute for Children and Families (INFA) and checklists from different Latin American countries. On the basis of a 40-minute observation supplemented by a 20-minute interview with the center coordinator, data on the child care center's basic services, infrastructure, classroom and play area vulnerabilities, possible risks for children, and existing capacity was collected.

It should be mentioned that it was important to include these three additional instruments in the battery of quality measures, because they contain valuable information on aspects of child care quality that are critical to the well-being of infants and toddlers. For example, the information about the average age of the group of children served (from the coordinator questionnaire) allowed the authors to examine whether quality measures were better in groups of older or younger children. Furthermore, using the caregiver and infrastructure questionnaires, data for center subgroups could be disaggregated by quality of physical infrastructure or caregivers' level of education, for example. Table 11 shows the most salient characteristics of these three additional questionnaires.

3.7/ How to Select an Instrument?

There are many child care quality measures, each with its own distinctive features. How does one choose the right instrument for each assessment? The answer depends on various parameters over and above

levels of complexity and implementation costs. The researcher or policymaker must give priority to one of these parameters, depending on whether the goal is to measure monitoring, tracking and/or continuous improvement processes for child care quality. Table 12 presents a summary of the reviewed instruments, which facilitates their comparison in terms of measurement and administration.

Table 13 presents a comparison of four essential elements of the quality measures—the number of dimensions, minimum administration time, availability of official training, and total cost, which mostly refers to training costs—in the first four main columns. It also presents four other dimensions that may be useful for decision-making: LAC implementation/adaptation of the instrument, availability of a Spanish-language translation of the instrument, validation of the instrument, and educational profile of the interviewer.

For example, the administration time for the CC-IT-HOME is one hour, placing it first in the ranking, followed by the routine assessment tools, the ORCE and the CIS at one-and-a-half hours, with the rest of the instruments requiring more than one-and-a-half hours of observation. The instruments with the longest administration times are the ITERS-R, the MITRCC and the APECP, all at more than three hours.

The total cost ranking, which is partly related to costs stemming from the licensing and use of intellectual property, is an important consideration.¹³ In this dimension, the least expensive instruments are the routine assessment tools and the MITRCC. It is important to note that no cost data is available for the CIS or the ORCE, and for the APECP, only material costs are reported, which is why this instrument ranks

¹³ Licensing costs also mean high transaction costs for the use of instruments.

¹⁴ It is important to note that the developers of the ORCE will generally lend the instrument since it was developed as part of an assessment for the U.S. government, but researchers must ask to borrow it.

among the least expensive.¹⁴ Next in the ranking are the ITERS-R (US\$22.90), the CC-IT-HOME (US\$40.30), and lastly, the CLASS Toddler (US\$902.90). The high cost of the latter is explained by the fact that it is the only instrument that requires mandatory training. It is worth emphasizing that these costs are only representative of a few implementation experiences, and they will depend a great deal on the context in which the instruments are implemented, the size of the list of child care centers to be evaluated, and the experience of the individual implementing them.

IMPLEMENTATION

CHAPTER 4



4.1/ Introduction

After reflecting on the appropriate instrument for measuring quality, one must consider other important questions involved in the implementation of a quality assessment. What type of profile should the individuals responsible for collecting information on child care quality have? Who will train the team that will administer the instruments, and how? How long will the pilot run? What are the context variables that must be taken into account during implementation and the potential changes that must be anticipated? What are the protocols to be followed in the field? What protocol will be established for inter-rater reliability in those instruments that require coded observation? How often should the observers re-establish reliability? These and other issues are addressed in this section.

4.2/ Training

Observers and interviewers must be properly trained by a team of trainers. Every individual should receive the same instruction. Trainers must be professionals with an extensive knowledge of child development and child care settings, and they must have previous experience with and comprehensive knowledge of the instrument on which they are providing training. In order to provide training on certain instruments, such as the CLASS, trainers are required to be certified by the company that distributes the instrument. This means that trainers have received specific training from that organization and that they have passed a series of evaluations that demonstrate their knowledge of the

instrument (the certification must be updated on a regular basis).

Training of the data collection team begins with an exhaustive review of the items and scoring methods of the test to be administered. It should also include all aspects of the assessment situation: contact and rapport with center staff, introduction of the tool, reading of the instructions, administration of the items, and recording of the responses. It is very important for the team to have training in fields related to child development, psychology or education and to have some experience assessing children or institutions. More importantly still is for the team to become solidly familiarized with and trained on the instruments to be implemented, the items they contain, and the concepts they present before entering the field. By the same token, the team must be skilled in the administration of the instruments (forms, questionnaires, etc.) to be used during fieldwork. In this way, the team will be able to gather accurate information and devote its attention to observation of the child care setting and data collection, while systematizing the information on the questionnaire in a quick, error-free manner.

Effective training combines academic explanations with many examples, case studies and practical activities that simulate situations in which researchers may find themselves in the field. Trainers also discuss how to handle these types of situations. During training, videos and photos of classrooms similar to those to be studied can be used to illustrate situations that occur in child care centers and to encourage discussion of certain concepts or topics. These types of exercises ensure that the researchers create a shared understanding of complex concepts such as the elements which characterize a warm, sensitive interaction between the child and caregiver and how these elements are observed in the cultural context in which the study will be conducted.

4.3/ **Piloting and Cultural Adaptation**

Culture has a significant impact on the way in which concepts are expressed, as well as their relevance and, in some cases, people's ability to understand them. For this reason, when quality measures are administered in a context different from the one for which they were originally designed or validated, it may be necessary to adapt them to the local language and culture. Specifically, it is necessary to evaluate the linguistic, functional, cultural and metric equivalence of the new version as compared to the original (Peña, 2007).

There are several techniques used to carry out a cultural translation. One of these techniques consists of translating the original version to the target language and then, using a second translator (or group of translators), translating the text back into the source language. The translation's accuracy is then evaluated by comparing the original version to the back-translation (van de Vijver and Tanzer, 2004). In parallel, a bilingual, bicultural native speaker should be involved from the beginning of the process or at least in the revision stage (Peña, 2007; van Widenfelt et al., 2005). Ideally, the team of translators should have knowledge not only of the two languages and cultures but also of the instrument's subject matter. Although it is desirable for a translation to remain as faithful as possible to the original text, translators need to be able to critically evaluate the text and identify situations in which a literal translation would be inappropriate in order to avoid losing the meaning of the content (van Widenfelt et al., 2005). Despite controls that can be implemented when translating from one language to another, it is possible that the version obtained may still be incongruous in its meaning and requires modification in

order for the instrument to prompt similar target behavior as originally designed (Peña, 2007).

In addition, the cultural adaptation process may involve changing the item order or modifying or removing (either partially or completely) items that are not applicable or culturally appropriate (van Widenfelt et al., 2005); however, the adaptation achieved using the last two alternatives produces results that are not directly comparable with the original version of the instrument (van de Vijver and Tanzer, 2004). For example, if the order of administration of the subscales or certain items has been changed in the adapted version of instrument, then researchers will be unable to compare the scores on the adapted instrument with the scores obtained through other studies using the original instrument. It is important to bear in mind that some of the publishers that own the rights to these instruments have specific policies regarding translation and adaptation.

Lastly, it is necessary to analyze the interpretations and responses obtained for each of the items in the survey section of the instrument, as well as their degree of difficulty in the new language. What does the caregiver think is being asked? What does this question mean to the caregiver? Piloting an instrument with a small group of people whose characteristics are similar to the population of interest is critical for obtaining useful information and assessing whether additional modifications are necessary (van Widenfelt et al., 2005). Pilot testing allows researchers to anticipate the difficulties that could arise during the data collection phase, offering the possibility to correct trouble spots prior to full implementation (Litwin, 1995). Qualitative research tools, such as focus groups, are another type of resource that can help validate the adapted instruments. Researchers can convene small groups with characteristics similar to the study population to discuss how well the items on the instrument reflect the reality

of the child care centers with which they are familiar.

In summary, the appropriate steps for cultural adaptation include a faithful translation of all materials and underlying concepts, adaptation of the assessments' content and process of administration to the local context, and thorough pilot testing that allows for adjustments and subsequent re-evaluation of the instruments (Fernald, Kariger et al., 2009). Furthermore, in countries where financial and human resources are very limited, instruments may be shortened or simplified to reduce costs and implementation time. It is worth remembering that the more an instrument is modified with respect to the original, the less comparable the results.

4.4/ Generation of Unique Identifiers in Databases

It is also very important to generate unique identifiers for each unit studied (geographical unit, child care center, caregiver, coordinator, group of children, child) that will allow data from one level of analysis to be easily merged with data from another level. For example, it is necessary to know whether child "1" belongs to group "A" or "B" at center "1" or "2"; otherwise, when it comes time to analyze the data, it will be impossible to identify the potential explanatory variables of quality being measured.

Similarly, using identifiers such as ID card numbers, birth certificate numbers, municipal or provincial administrative code numbers, or unique identifiers used by the child care center in connection with its mother institution will also facilitate the later consolidation of the various databases with other administrative data. While it is necessary to be cautious with

regard to data access and confidentiality, recent literature has shown that the use of administrative data consolidated with ad hoc surveys can provide richer analysis.

4.5/ Critical Context Variables

It is extremely important to have an extensive understanding of the context in which the quality measures will be implemented in order for the assessment to run smoothly and to avoid difficulties in the field. A list of key elements to research before designing and piloting the questionnaires is presented below:

1. Location of child care centers (urban, rural, remote);
2. child care frequency (number of days per week);
3. operating hours and number of sessions per day;
4. center routines (arrival/greeting, snack, lunch, nap, late snack and/or dinner; and departure); see if caregivers have free time during the day to complete a survey;
5. organizational structure (investigate if centers are local, municipal, provincial and/or national; investigate how the operational, administrative and financial aspects of the center are organized);
6. structure of human resources (one or more center coordinators, one or more caregivers per group of children, group size, child-caregiver ratios depending on age range, presence of part-time support staff);
7. structure of the groups served at the child care center (organization of classrooms by age groups; frequency of care for children of different ages, e.g., a small number of children under the age of 1, or various age groups represented);

8. spaces where different aspects of the children’s routine occur (indoor spaces, outdoor spaces, classrooms, cafeteria, etc.);
9. size of the centers in the population to be studied (distribution of the number of children, caregivers and groups);
10. language of the service’s user population and of center staff.

In addition, researchers must consider whether the aforementioned variables will vary significantly among different regions or between urban and rural areas.

4.6/ **Inter-rater Reliability**

Inter-rater reliability quantifies the similarity of quality scores assigned by different observers using the same instrument in the same study (Gwet, 2008). More specifically, inter-rater reliability means that different trained observers can make accurate and consistent judgments about the caregiver’s performance, the dimensions captured by process variables, and other evidence of child care quality. This type of reliability is important to ensure that different observers score each situation in the same way, thus reducing the potential bias resulting from subjective personal assessment (Fernald, Kariger et al., 2009).

High inter-rater reliability must be established before entering the field, and it is necessary to maintain it throughout the data collection phase using periodic comparison, calibration or re-certification of observers (APPR, 2012). With that said, observers must begin establishing inter-rater reliability during training, with clear objectives on how to gather, analyze and report the evidence observed at child care centers. Working as a team, observers can view sample videos

showing the daily routine at some centers and discuss the criteria used, evidence gathered, and decisions made with regard to the scoring of the items on each test.

How do researchers test inter-rater reliability? Researchers must select a gold-standard observer—an individual highly-trained and efficient in the administration of the instrument—She will then assess the same situation as the team of observers. The responses of each team member are then compared to the responses of the gold-standard observer to ensure a correlation of at least 0.8, meaning 80% agreement on the items assessed (Fernald, Kariger et al., 2009). This requirement that observers demonstrate agreement with a standard—and that the instrument be specific enough to enable this agreement—is a fundamental property of any assessment tool (Pianta, 2011).

Test-retest reliability is used to assess an instrument’s ability to assign a consistent quality ranking to a child care center on two occasions over a period of time (e.g., a medium-quality child care center is consistently medium quality, whether it is evaluated today or in two weeks). The idea is to prove that the quality assessment is consistent over the short term, i.e., the score given the first time was not a product of circumstance but rather a true reflection of the quality of that center. With test-retest reliability, items are expected to have at least 80% agreement. It is very important to document this type of exercise as part of the preparation process for data collection. Shadow scoring should also be carried out to ensure that observer reliability is consistent over time.

4.7/ Following Protocols

This section describes some of the protocols that must be defined for fieldwork involving data collection of this nature.

General protocol

This protocol contains essential information to be given to the child care center when the study is introduced. It is necessary to provide the child care center with the following information: who the researcher is, whom she works for, the purpose of the study and who commissioned it, what activities will be performed and by/with whom, how long she will be at the center, what support resources she needs during her visit, and what documentation or spaces she will need to verify or examine. It is also necessary to convey that there is a precise protocol for working with children. Furthermore, once all questions raised about the study have been answered, the representative of the child care center to be studied must give her informed consent to participate in the research. When the quality assessment is entrusted to a third party (e.g., a survey research firm), it is common for the entity that oversees child care services to provide a letter explaining the nature, purpose and characteristics of the study to the child care providers and requesting their participation and collaboration with the researchers. It is also important to explain whether the data will be anonymous, who has access to it, and the impact this information may have on center operations.

The second important issue when arriving at the center is the selection of the group of children to be studied. Since groups at a given center are usually organized by age, making this decision involves compromise.

The researcher can select a) *all* of the groups at the center that fall within the target age (0 to 36 months, for example), which can be an expensive option, or b) *just one* group from all of those groups who fall completely within the target age. When more than one group meets the characteristics required by the study, the selection of study subjects should be performed randomly to ensure external validity in the children/caregiver/groups/classrooms that will be observed. For example, a number can be assigned to each group that falls completely within the age range, and then the number rolled on a die can be used to determine which group to observe. Ideally, the center does not know which classroom, class, or group of children will be observed that day.

It may also be important to collect demographic information (name, age, ethnicity) for each of the children in the group to be studied, as well as relevant group variables (size, the number of caregivers, the age range of the children served, etc.), although some instruments already include some of these questions. This is important not only for group selection but also for the disaggregation of data for analysis and interpretation.

Interview/observation

When one visits a child care center with the aim of conducting an interview or observation, it is important to remember that the researchers are interested in learning about the daily routine that is followed at the center. For this reason, it is desirable to minimize the effects that the presence of observers/videographers/interviewers may have on the activities that normally take place at the center and the way in which the staff and children behave on a daily basis. Discretion should be exercised during the visit, and observers should work to build relationships of trust and respect with center staff. It is recommended that observers invest all the time necessary at the start of the project in getting to know and establishing

relationships with the staff, so that the school day flows smoothly and with little interruption.

The observer must situate herself in a place where she will not disrupt the classroom environment, so as not to interfere with routine activities while observing. It is also recommended that she avoid interacting with the children and interrupting the caregiver or other staff members. While it is necessary for the observer to be located somewhere where she can comfortably record her observations, it is preferable not to use the furniture intended for the children or their caregiver.

When interviewing subjects, it is recommended that the interviewer use a proper and respectful tone and style. The interviewer must feel comfortable in her role and be able to improvise and adapt to different situations. The idea is to generate a fluid exchange in which the interviewer is capable of 1) letting the conversation deviate from the main topic and then politely redirecting the discussion, 2) recognizing when an issue requires further inquiry (e.g., if she detects inconsistencies between two responses), 3) delving deeper into issues when necessary, or 4) pausing the interview for a few minutes (e.g., to pay attention to the children). The main objective is to avoid questions that may be perceived as threatening or judgmental, as this would lead to a defensive and biased attitude from the staff, yielding responses perceived as “correct” but not necessarily true.

It is essential to remember that, regardless of the context or the response, both the interviewer and the observer must remain objective and understanding and maintain a neutral attitude that demonstrates neither approval nor disapproval of the situation. In other words, the interviewer must not express opinions or make suggestions and

must maintain a neutral tone throughout her visit, even when the respondent asks her opinion about a specific situation. This attitude must be transmitted to all center staff in order to prevent staff from offering the answers they consider to be “correct”. An example of this situation is when the caregiver asks the interviewer or observer for her opinion about the information being conveyed. In this case, the interviewer’s response should leave no doubt about the nature of her role (obtaining and systematizing data) and the importance of the answers provided by the caregiver, which are not used to rate her performance or that of the center but instead to understand the condition of quality in the sample or total population being studied.

If the attitude of child care staff is one of disapproval or discomfort, even after following the aforementioned recommendations, the most prudent option is to offer to reschedule the visit for another time.

Filming

The main purpose of conducting an assessment based on video observation is to obtain information on the activities and types of actions and interactions that take place at a child care center; therefore, it is essential that the entire audiovisual recording process be performed in the most accurate manner possible. Key aspects to obtaining a recording that adequately captures the visual and audio information of a typical day are listed below.

General recommendations for the videographer¹⁵

- Arrive at the child care center before the school day starts, with enough time to communicate with the caregiver, learn which areas will be used by the children that

¹⁵ This protocol is based on the CLASS Toddler video recording protocol (Reyes Ugalde and Schodt, 2011). The authors’ study used the same protocol outlined in the CLASS.

day, and set up the equipment in the room where filming will take place.

- Explain to the caregiver that the idea is that the video be as natural as possible and that she should not worry about making introductions or offering explanations about the videographer's presence.
- Have a friendly and open attitude to keep the children and caregivers from feeling stressed or nervous.
- Do not intervene, talk to the children or help with activities. The videographer can sit in a chair so as not to interfere too much with the class, but he should not use the children's work tables or chairs. Avoid using the space where the caregiver and children interact.
- Store personal belongings out of the reach of children and in a place where they do not occupy classroom space. Try not to bring anything other than video equipment to the classroom.
- Once at the child care center, the videographer must turn off all electronic devices, including cell phones, for the duration of the visit.

Recommendations on filming

- Use the setting that allows the video camera to record for as long as possible (lowest video quality), rather than recording in high resolution. This minimizes the risk that the video camera will stop recording prematurely due to a lack of space on the memory card.
- Make sure that the video camera's rechargeable battery is fully charged. The video camera should be set up as close as possible to the wall at the back of the classroom in order to have a full view of the room (for example, when using CLASS) or

of the entire group of children (when using MITRCC).¹⁶ Avoid rigging up complicated or unstable structures to support the video camera, as they can be dangerous. Always mount the video camera on a tripod to keep the image stable, thus facilitating observation of the video and subsequent analysis. Carefully position the camera so as to avoid shooting into the sun. It is very important to avoid backlighting because it has a serious impact on the quality and clarity of the footage.

- Make sure that the video camera is turned on and that the lens cap has been removed. It is advisable to test the sound and video quality before recording.
- Constantly check the video camera to make sure it is recording.
- Do not stop recording during breaks/recess and do not create new files, because the subsequent work of coding and analysis becomes more confusing and complex if the footage is contained in multiple files.
- Once filming begins, the videographer must not speak to the caregiver or the children. If they speak to him, he can politely respond that he can resume the conversation at the end of the day.

Editing

Editing is a costly and important step in the process of working with videos. The fundamental methodological aspects of this process include the following: 1) editors must be familiar with the editing protocol,¹⁷ 2) footage that does not comply with the protocol must be re-edited, 3) back-ups of all material must be saved for re-editing whenever necessary, and 4) all footage must be properly organized (using file names, the center's unique identifier, etc.).

¹⁶ When filming for an assessment that uses an instrument whose subject is a child (as in the case of the CC-IT-HOME, for example), the camera should only focus on the caregiver and the child.

¹⁷ The editing protocol was not included in order to keep the document to a reasonable length; however, it is available to interested readers. Please contact the authors at mcaraujo@iadb.org, florencial@iadb.org.

Coding

The coding team works in an established office space. For the analysis of videos (i.e., for coding), each coder requires a computer with the ability to read .mov, .mp4, and .avi files and DVDs, a pair of headphones, and a permanent, quiet workstation.

Coders can work flexible hours. It is recommended that coders work no more than four hours per day (plus a break), as coding is a task that requires a very high level of concentration. To ensure the quality and validity of the assessment, coding requires that two different individuals evaluate each of the videos and at least one subsample in order to achieve adequate inter-rater reliability. Videos must be randomly assigned to coders in order to prevent them from evaluating the same videos at the same time or in collusion. When there are numerous discrepancies between the two coders on the same segment of video, a third analysis can be performed by an individual or group of coders, or the trainer or supervisor of the coders can get involved.



FINAL CONSIDERATIONS

CHAPTER 5



This toolkit has shown that the measurement of child care quality for infants and toddlers involves key issues, including the selection of variables that measure quality, how the instrument actually reflect what one is trying to measure, critical elements in this context, and researchers' professional qualifications, just to name a few.

As a final reflection, this toolkit proposes to identify three stages to consider when seeking to measure the quality of child care for infants and toddlers.

1. Determine the purpose of the measurement. Researchers may seek to measure quality for the following reasons:

- a. to improve and expand child care services;
- b. to monitor programs;
- c. to perform impact evaluations;
- d. to diagnose service quality;
- e. to redesign the curriculum.

2. Determine the type of measurement to be performed.

a. Structural variables. These variables determine the presence or absence of those resources that facilitate the interactions characteristic of an environment of care. They can be more easily controlled, and they are organized into four major groups: center infrastructure, health and safety issues, the group of children, and the caregiver. These variables are useful for measurements with the aforementioned purposes.

b. Process variables. Process indicators tend to focus on dynamic aspects, such as the interactions between children and their caregivers, the relationships among the children themselves, caregivers' behavior, and the implementation of the curriculum and health and safety protocols. These variables are useful when the goal of the measurement fits the situations described

in points (a), (c), (d) and (e) above. When the stated purpose of the measurement is monitoring, it is recommended that more simple quality measures be considered.

3. Select the instrument(s) to be administered according to the following criteria:

- a. the purpose of the measurement;
- b. the validity of the instrument;
- c. the age and characteristics of the group of infants and toddlers to be studied;
- d. whether to perform the assessment using direct observation (or filming), self-report, or a combination of the two;
- e. dimensions of quality to be measured;
- f. context (center-based, community-based or home-based; small or large groups);
- g. cultural adaptation to the context of the country where the measurement will be performed;
- h. budget for the measurement/study;
- i. the use of copyrighted instruments;
- j. administration time;
- k. training time (and cost);
- l. profile required of observers.
- m. the availability of professional staff who can conduct the necessary training (which will involve training time and costs different from those for the observers).

Table 1: Structural and Process Variables , some examples

	Structural Variables	Process Variables
Health & Safety	<ul style="list-style-type: none"> Public health measures, health and safety procedures, documents 	<ul style="list-style-type: none"> Observed health and safety practices. The caregiver helps the children follow safety rules and explains the rationale behind these rules.
Groups of Children	<ul style="list-style-type: none"> Group size Child-caregiver ratio 	<ul style="list-style-type: none"> Children interact with each other for much of the day. The caregiver helps the children empathize with their peers; she explains children's actions, intentions and feelings to other children. The caregiver interrupts a negative interaction between children and helps them understand the effects of their actions on others.
Caregiver	<ul style="list-style-type: none"> Score: years of education, training in child development, previous experience and professional development Lesson planning Caregiver supervision Salary 	<p>Caregiver Behavior:</p> <ul style="list-style-type: none"> Caregivers are attentive to all children, even while working with an individual child. How the caregiver responds when a child cries; the caregiver does not express annoyance or hostility toward the child. How many times the caregiver uses abrupt movements when feeding a child, complains about his behavior, or has a threatening attitude. Caregivers greet/say goodbye to each child and his parent during arrival and departure times. Caregivers react quickly to solve problems. <p>Child-caregiver interactions:</p> <ul style="list-style-type: none"> Caregiver uses a variety of simple words to communicate with the children. Caregiver talks about many different topics with the children, asks them simple questions, and/or expands on the children's ideas with other words and ideas. Caregiver does not reprimand, criticize or punish the child. Caregiver encourages the children to dance, clap or sing together. Caregiver hugs or kisses the child at least once per day. <p>Curriculum implementation:</p> <ul style="list-style-type: none"> Caregiver is flexible with regard to lesson plans and activities, he selects most classroom activities taking consideration children's preferences. Caregivers introduce concepts of relational correspondence, more-less-the same, or cause and effect during teachable moments. Naptime is optional, and there are activities for children who do not sleep. There is free play for much of the day. Children and caregivers play together with building blocks.
Infrastructure	<ul style="list-style-type: none"> Access to potable water Waste disposal Electricity Telephone service Physical space per child Materials Protected facilities, play areas 	

Table 2: Comparison of Quality Measures Used to Monitor Child Care Services in Ecuador, the City of Buenos Aires, and Mexico (relative weight of each item, %)

Categories	Ecuador	City of Buenos Aires	Mexico	Mexico - Annex	Examples of Information Collected for Each Category
Supervision data	0.0%	1.2%	1.8%	0.0%	Date of supervision, supervisor in charge, status of the center (open or closed)
Center information	7.2%	3.9%	2.5%	4.5%	Location (state/province, county, city/town, district/neighborhood), address, phone number, director of the center, center opening date, modality, hours of operation
Condition of the property	12.6%	3.1%	6.5%	15.0%	Condition and materials of ceilings, walls, floors, bathrooms, the electrical system, and furnishings; condition and other uses of the property
Staff information	4.0%	29.8%	14.0%	0.0%	Gender, length of service at the center, level of education, training, availability of staff health records (director, teachers, aides, kitchen staff, etc.)
Population served	5.4%	8.1%	2.5%	8.5%	Number of boys and girls served, by classrooms or age groups, with or without financial support
Safety	70.4%	5.0%	57.5%	4.5%	Condition of the lighting, ventilation, stairs and furnishings; waste management; presence of chemicals; presence of safety elements such as fire extinguishers and a first aid kit; fire and disaster drills; emergency information
Facilities, orderliness and cleanliness	0.0%	5.8%	5.5%	5.5%	Cleanliness of bathrooms and kitchens, bathroom exclusively for children, fixed toilets and sinks, furniture free from defects that may pose a risk to children (sharp corners, splinters, loose or rusty parts)
Attendance record	0.0%	0.0%	2.3%	0.0%	Existence of a complete attendance record (dates, names, arrival and departure times)
Documentation for the center and the children	0.0%	5.4%	4.5%	15.5%	Existence of internal policies, meeting minutes and lesson plans; existence of a copy of the birth certificate, current photo, health certificate, immunization record, home address, and phone number of a contact person for each child
Nutrition assistance, educational activities and growth monitoring	0.4%	36.0%	1.3%	0.0%	Children receive breakfast, lunch and/or a snack, compliance with the menu and portion size; content that is presented through stimulating and educational activities; workshops with families; anthropometric monitoring
Electricity and water costs	0.0%	0.0%	0.0%	5.0%	Water and electricity rates, changes in the rate since the center first opened
Inspections by external agencies	0.0%	0.0%	0.0%	33.0%	Supervision by personnel at Ministries or Departments that do not manage the program
Other aspects	0.4%	1.6%	1.8%	8.5%	Observation of verbal, physical or psychological abuse, staff smoking and/or drinking alcoholic beverages, supervision of children at all times (even when they are sleeping); service contingent upon payment
Total items	223	258	400	200	

Table 3: Characteristics of the ITERS-R

Instrument		The Infant/Toddler Environment Rating Scale – Revised Edition (ITERS-R)
Design	Authors/Source	Harms, Cryer and Clifford (2006)
	Country of Design	USA
	Country of Implementation	Brazil - Verdisco and Pérez Alfaro (2010) Chile - Herrera et al. (2005) Colombia - Bernal et al. (2009) Ecuador - Araujo et al. (2015) USA - Harms, Cryer and Clifford (2006), among others
	Reliability and Validity	Cryer et al. (2004a) analyze the validity and confiability with a field-testing performed with groups of programs in North Carolina (USA), based on 90 observations with two paired observations each in 45 group settings. Measures of reliability: <ul style="list-style-type: none"> • Inter-rater reliability: <ul style="list-style-type: none"> ▪ Indicator reliability: Of a total of 39 items, there was agreement on 91.6% of all scores given by the raters. If only the first six subscales (32 items) are considered, then the observer agreement was 90.27%. Only one item, safety practices, had indicator agreement of less than 80%. ▪ Item reliability: The agreement between pairs of observers was calculated within 1 point on the seven-point scale. Across the 32 child-related items, agreement was obtained 83% of the time, and for the full 39 items, there was agreement in 85% of the cases. The mean weighted Kappa for the first 32 items was 0.55 and for the full 39-item scale it was 0.58. Only two items had a weighted Kappa below 0.40; therefore, the authors made minor changes to all items with a weighted Kappa below 0.50 to improve their reliability without changing the basic content. • Overall agreement: The intraclass correlation was 0.92. • Internal consistency: Overall, the scale has a high level of internal consistency, with a Cronbach's alpha of 0.93; however, the authors recommend caution when using the Space and Furnishings and Personal Care Routines subscales. Furthermore, they recommend using the Program Structure subscale excluding item 32 unless most programs being assessed include children with disabilities (Cryer et al., 2004a).
Measurement	Type of Assessment	Direct observation and reporting.
	Dimensions	1. Space and Furnishings (5 items) 2. Personal Care Routines (6 items) 3. Listening and Talking (3 items) 4. Activities (10 items) 5. Interaction (4 items) 6. Program Structure (4 items) 7. Parent and Staff (7 items)
	Scoring	Items are rated on a 7-point scale. The score for each subscale is computed by averaging the scores on the respective items per scale.
	Administration Time	3 hours and 30 minutes (minimum). At least 3 hours of observation and coding plus 20-30 minutes of reporting
	Interviewer Requirements	Specific requirements: <ul style="list-style-type: none"> • Complete training and achieve the desired level of reliability for coding General requirements: <ul style="list-style-type: none"> • Good interpersonal skills • Strong verbal communication skills (clear and articulate) • Affinity for children • Relevant post-secondary education (e.g., psychology, early childhood education, elementary education) • Experience working with children • Field experience in qualitative work • Good negotiation and problem solving skills • Ability to solve unforeseen problems

Table 3: Characteristics of the ITERS-R (continued)

Instrument		The Infant/Toddler Environment Rating Scale – Revised Edition (ITERS-R)
Administration	Where?	Center.
	To Whom?	Setting (e.g., classroom), group of children, primary caregiver, and aide (when present).
	Age Range	0-30 months.
Requirements	Training	To formally use the scale, Harms, Cryer and Clifford (2006) recommend participating in training led by an experienced ITERS-R trainer; it is not necessary to participate in official training. The coders/observers who will use the scale must participate in at least two practice observations in a classroom with a small group of observers and compare scores within the group. ¹
	Materials	Manual and score sheets. Table and chair for reporting.
	Others	Achieving the desired level of reliability within groups of coders/observers may require additional practice observations in the field. Additional materials for this purpose, such as training videos, are available through Teachers College Press. ² Furthermore, Cryer et al. (2004a) presents information and photos that are useful for understanding the scale or interpreting observations in order to improve inter-rater reliability.
Cost	Training	Formal training is not required. ³
	Materials	USD\$22.90 for the manual. ⁴ Score sheets are freely accessible and of no cost. ⁵
	Staff Required	Trainer for coders/field observers and coders/field observers.
	Others	<ol style="list-style-type: none"> 1. A person with a high level of education (post-secondary) and specialization in child development could be trained in ITERS-R and CC-IT-HOME at the same time in order to reduce the length of training. 2. The Family Day Care Rating Scale (FDCRS) is a version of the ITERS applicable to family-based child care situations. This scale was implemented in Colombia (Bernal et al., 2009).

¹ One trainer and 15 coders/observers participated in the IDB study in Ecuador (Araujo et al., 2013). The training consisted of a five-day train-the-trainer program at NIEER (New Jersey, USA), with one day of theory and four days of practice to establish reliability. Field coders/observers received two days of training.

² <http://www.tpress.com/ERS.html>

³ The authors offer training at the University of North Carolina, Chapel Hill. Visit http://ersi.info/training_live.html for more information.

⁴ Prices were retrieved on December 28, 2015 from <http://www.tpress.com/ERS.html>.

⁵ Available at <http://www.tpress.com/pdfs/iterss.pdf>.

Table 4: Characteristics of the CLASS Toddler

Instrument		The Classroom Assessment Scoring System (CLASS) Toddler
Design	Authors/Source	Pianta, La Paro and Hamre (2008)
	Country of Design	USA
	Country of Implementation	USA Ecuador (version translated to Spanish) - Araujo et al. (2015)
	Reliability and Validity	Thomason and La Paro (2009) analyze the construct validity of the CLASS Toddler on the basis of a study of 46 teachers in 30 different pre-school classrooms in the U.S. The construct validity of the CLASS Toddler is based on comprehensive reviews of existing measures, a review of the research on the unique aspects of early childhood development, and observations of preschool settings. The instruments reviewed include the ITERS (Harms et al., 2006), the CIS (Arnett, 1989), and the ORCE (NICHD, 1996). The examples and indicators included in the adapted instrument reflect this review. Similar to the CLASS Pre-K (Pianta et al., 2005), examples are provided for the scoring of centers, with ratings of low (1-2), medium (3-5) or high (6-7). Additionally, the instrument was reviewed by an expert in infant/toddler development to ensure the validity of the adapted concepts. No additional information on the instrument's validity was found (see annex 1 of this toolkit for information on the validity of the CLASS Pre-K, on which the CLASS Toddler is based). Significant correlations were found between traditional measurements of quality in early childhood education, such as the education level of the caregiver, group size, and the child-caregiver ratio, and the scores assigned (based on the inclusion of quality aspects in the program and the education level of caregivers) and dimensions.
Measurement	Type of Assessment	Observation (direct or video-based).
	Domains and Dimensions	<ol style="list-style-type: none"> 1. Emotional and behavioral support <ol style="list-style-type: none"> 1.1. Positive climate (1 item) 1.2. Negative climate (1 item) 1.3. Teacher sensitivity (1 item) 1.4. Regard for child perspectives (1 item) 1.5. Behavior guidance (1 item) 2. Engaged support for learning <ol style="list-style-type: none"> 2.1. Facilitation of learning and development (1 item) 2.2. Quality of feedback (1 item) 2.3. Language modeling (1 item)
	Scoring	Dimensions are rated on a 7-point scale.
	Administration Time	2 hours (minimum). Four 20-minute observation cycles and 10 minutes for coding.
	Interviewer Requirements	<p>Specific requirements:</p> <ul style="list-style-type: none"> • Complete training and pass a reliability test • Complete a daily validity test <p>General requirements:</p> <ul style="list-style-type: none"> • Good interpersonal skills • Strong verbal communication skills (clear and articulate) • Affinity for children • Relevant post-secondary education (e.g., psychology, early childhood education, elementary education) • Experience working with children • Field experience in qualitative work • Good negotiation and problem solving skills • Ability to solve unforeseen problems

Table 4 (continued)		
Instrument		The Classroom Assessment Scoring System (CLASS) Toddler
Administration	Where?	Center
	To Whom?	Group of children, primary caregiver, and aide (when present).
	Age range	15-36 months.
Requirements	Training	Each coder/observer is required to be certified on the instrument by participating in a two-day training and passing a reliability test. To pass this test, observers must achieve 80% agreement (within one point of the consensus rating) across five training videos, with at least two scores per dimension correct across the five videos (no variance in the rating). Training can be provided directly by Teachstone or by a trainer certified through Teachstone, who must have obtained certification through the three-day training program within the last year. ¹ In order to maintain validity, coders/observers undergo daily reliability checks before beginning their work. ²
	Materials	Manual, score sheets, training videos, projection screen.
Cost	Training	The Teachstone train-the-trainer program costs USD\$4,500. Alternatively, Teachstone training for coders/observers is USD\$850.30. ³
	Materials	USD\$49.90 for the manual and USD\$15 for the score sheets. ⁴
	Staff Required	Coders/field observers and the trainer for coders/field observers.
Others		The instrument can be coded in the field or with a video recording. For more details on video coding, see section 3.4 of this toolkit.

¹ For more information on CLASS certification for observers and trainers, see www.classobservation.com.

² One trainer and 15 coders/observers participated in the IDB study performed in Ecuador (Araujo et al., 2013). Total training time was approximately three weeks. It consisted of three days for the official Teachstone train-the-trainer program, five days to prepare training videos and other materials, and five days for the training for field observers/coders.

³ Prices were retrieved on December 28, 2015. They correspond to training in the U.S. and do not include travel expenses. <http://store.teachstone.org/>.

⁴ Prices were retrieved on December 28, 2015 from <http://store.teachstone.org/>. Score sheets are available, in English, in packs of 5 booklets. Each booklet contains six observation sheets and one scoring summary sheet. The questionnaire translated to Spanish and used by the authors in Ecuador is available to interested readers.

Table 5: Characteristics of the ORCE

Instrument		The Observational Record of the Caregiving Environment (ORCE)
Design	Authors/Source	NICHD Early Child Care Research Network (1996)
	Country of Design	USA
	Country of Implementation	USA (2002, 2010)
	Reliability and Validity	<p>Reliability information for the ORCE is based on a sample of 1,364 families from 10 U.S. sites (NICHD Study of Early Child Care and Youth Development), with children born at 24 hospitals</p> <ul style="list-style-type: none"> The Behavior Scales produced Pearson correlations ranging from 0.41 to 0.99 (with most estimates above 0.80), 0.08 to 0.99, and 0.34 to 0.97, for the 6-month, 36-month and 54-month scales, respectively. Reliability data was not reported for each variable on the 15-month scale, and acceptable levels of reliability were calculated for the variables on the 24-month scale (with some exceptions). The qualitative ratings revealed Pearson correlations above 0.80 for the 6-month scales, a range of 0.20 to 0.85 for the 15-month scales, a range of 0.47 to 0.76 for the 24-month scales, a range of 0.57 to 0.93 for the 36-month scales, and estimates between 0.20 and 0.83 for the 54-month scales. <p>(Halle, Whittaker and Anderson, 2010).</p>
Measurement	Type of Assessment	Observation (direct or video).
	Domains and Dimensions	<p>The instrument has four versions (6, 15, 24, 36, and 54 months), one for each time point when data was collected (the 24- and 36-month versions are the same). Each version is divided into the same three domains; however, there are some differences in the dimensions. The 6-month scales are presented below (see Halle, Whittaker and Anderson [2010] for the other scales derived from this one).</p> <ol style="list-style-type: none"> Behavior Scales <ol style="list-style-type: none"> 1.1 Positive and negative affect (3 items) 1.2 Language-focused interaction (3 items) 1.3 Stimulation (2 items) 1.4 Behavior management (5 items) 1.5 Child's activity (6 items) 1.6 Child's interaction with other children (2 items) Qualitative Ratings <ol style="list-style-type: none"> 2.1 Caregiver notes (8 items) 2.2 Child notes (5 items) Structural Variables <ol style="list-style-type: none"> 3.1 Ratio (1 item) 3.2 Group size (1 item) 3.3 Numbers of children (1 item) 3.4 Numbers of adults available (1 item) 3.5 Proportion of observation completed outdoors (1 item) 3.6 Amount of time caregiver is involved with child (1 item) 3.7 Age mix of the group (1 item)
	Scoring	Items are rated on a 4-point scale. The overall rating is derived from the average among the three domains for all observation cycles (Vandell and Wolfe, 2000).
	Administration Time	90 minutes (minimum), following the NICHD Study of Early Child Care Phase I Instrument Document (2004) in Halle, Whittaker and Anderson (2010). Two to four 44-minute cycles. Each cycle includes three 10-minute intervals of observation, interspersed with 2-minute periods of note-taking, followed by a 10-minute interval of observation and note-taking focused on global qualities of behavior. Vandell and Wolfe (2000) mention that a minimum of four observation cycles distributed over a period of two days is required.

Measurement	Interviewer requirements	<p>Specific requirements:</p> <ul style="list-style-type: none"> • Complete training and achieve the desired level of reliability for coding <p>General requirements:</p> <ul style="list-style-type: none"> • Good interpersonal skills • Strong verbal communication skills (clear and articulate) • Affinity for children • Relevant post-secondary education (e.g., psychology, early childhood education, elementary education) • Experience working with children • Field experience in qualitative work • Good negotiation and problem solving skills • Ability to solve unforeseen problems
Administration	Where?	Non-maternal care environment (center, home).
	To Whom?	A child and a caregiver or other adult present.
	Age Range	6, 15, 24, 36 and 54 months.
Requirements	Training	Coders/observers must be trained on this instrument, including methods of administration, in order to achieve reliability. ¹
	Materials	Manual and score sheets.
Cost	Training	N/A
	Materials	N/A
	Staff Required	Coders/field observers and the trainer for coders/field observers.
	Others	The instrument can be coded in the field or with a video recording. For more details on video coding, see section 3.4 of this toolkit.

¹ The NICHD Study of Early Child Care required coders/observers to achieve at least 90% reliability to be certified, and they were re-tested every four months to maintain reliability, Bradley et al. (2003) in Halle, Whittaker and Anderson (2010).

Table 6: Characteristics of the CIS

Instrument		The Arnett Caregiver Interaction Scale (CIS)
Design	Authors/Source	Arnett (1989)
	Country of Design	USA
	Country of Implementation	Bermuda - Arnett (1989) USA - Jaeger and Funk (2001)
	Reliability and Validity	Jaeger and Funk (2001) report the following: <ul style="list-style-type: none"> • Inter-rater reliability: coefficients in a range from 0.75 to 0.97 between certified observers and trainees. They report coefficients of 0.81 and higher for the sensitivity, punitiveness, and detachment subscales. • Internal consistency: Cronbach's alphas (Layzer et al., 1993) of 0.91 and 0.90 were found for warmth/responsiveness and harshness, respectively.
Measurement	Type of Assessment	Observation (direct or video-based).
	Dimensions	1. Sensitivity (10 items) 2. Harshness (8 items) 3. Detachment (4 items) 4. Permissiveness (4 items)
	Scoring	Items are rated on a 4-point scale.
	Administration Time	90 minutes. Two observation cycles of 45 minutes, on separate occasions (Vandell and Wolfe, 2000).
	Interviewer Requirements	Specific requirements: <ul style="list-style-type: none"> • Complete training and achieve the desired level of reliability for coding General requirements: <ul style="list-style-type: none"> • Good interpersonal skills • Strong verbal communication skills (clear and articulate) • Affinity for children • Relevant post-secondary education (e.g., psychology, early childhood education, elementary education) • Experience working with children • Field experience in qualitative work • Good negotiation and problem solving skills • Ability to solve unforeseen problems
	Where?	Non-maternal care environment (center, home).
Administration	To Whom?	Group of children, primary caregiver, and aide (when present).
	Age Range	36-60 months.
	Training	Theoretical and practical training are required. Prior to conducting independent observations, observers must achieve reliability with the trainer. In order to be a certified CIS observer, one must achieve a reliability level of 70% on two consecutive visits (Jaeger and Funk, 2001).
Requirements	Materials	Manual and score sheets.
	Others	Two different observers (Vandell and Wolfe, 2000) are required.
	Training	N/A
Cost	Materials	The manual and score sheets are freely accessible and of no cost. ¹
	Staff Required	Coders/field observers and the trainer for coders/field observers.
	Others	The instrument can be coded in the field or with a video recording. For more details on video coding, see section 3.4 of this toolkit.

¹ Available at http://www.eec.state.ma.us/docs1/qris/20110121_arnett_scale.pdf

Table 7: Characteristics of the Assessment Profile

Table 7: Characteristics of the Assessment Profile	
Instrument	The Assessment Profile for Early Childhood Programs (APECP)
Design	Authors/Source Abbott-Shim, Neel and Sibley (2001)
	Country of Design USA
	Country of Implementation USA (NICHD, 2000a)
	Reliability and Validity <ul style="list-style-type: none"> • Inter-rater reliability: For both the summative and formative versions, the reliability between a trainer and observer is consistently reported with a mean of 93% to 95% agreement, with a range of 83% to 99% agreement (Abbott-Shim, Lambert, and McCarty, 2000). Other studies have reported similar levels of inter-rater reliability. • Internal consistency: the reliability coefficients for the five dimensions (Learning Environment, Scheduling, Curriculum, Interacting, and Individualizing) range from 0.79 to 0.98 for the Kuder-Richardson 20 and from 0.81 to 0.98 for the Spearman-Brown. The IRT-based reliabilities for the five scales range from 0.83 to 0.91 (Abbott-Shim, Neel and Sibley, 1992).
Measurement	Type of Assessment Direct observation and reporting.
	Domains and Dimensions <p>Summative measure:</p> <ol style="list-style-type: none"> 1. Classroom practices <ol style="list-style-type: none"> 1.1 Learning Environment (12 items) 1.2 Scheduling (12 items) 1.3 Curriculum Methods (12 items) 1.4 Interacting (12 items) 1.5 Individualizing (12 items) <p>Formative measure¹:</p> <ol style="list-style-type: none"> 1. Classroom practices <ol style="list-style-type: none"> 1.1 Safety (109 items) 1.2 Learning Environment (73 items) 1.3 Scheduling (34 items) 1.4 Curriculum Methods (49 items) 1.5 Interacting (61 items) 1.6 Individualizing (25 items) 2. Administrative practices <ol style="list-style-type: none"> 2.1 Physical facilities (68 items) 2.2 Food service (45 items) 2.3 Program management (63 items) 2.4 Personnel (38 items) 2.5 Program development (31 items)
	Scoring Each item is scored using a yes/no format.
	Administration Time Two or three classes/groups can be evaluated per day. Classrooms should be observed in the morning, with record reviews and teacher interviews conducted in the afternoon. The administrative component of the formative assessment requires four to six hours. Therefore, assuming an eight-hour work day, it is estimated that the administration time for the summative assessment is three hours, while the formative assessment requires seven to nine hours.
	Interviewer Requirements <p>Specific requirements:</p> <ul style="list-style-type: none"> • Complete training and achieve the desired level of reliability for coding <p>General requirements:</p> <ul style="list-style-type: none"> • Good interpersonal skills • Strong verbal communication skills (clear and articulate) • Affinity for children • Relevant post-secondary education (e.g., psychology, early childhood education, elementary education) • Experience working with children • Field experience in qualitative work • Good negotiation and problem solving skills • Ability to solve unforeseen problems

Table 7 (continued)

Administration	Where?	Center.
	To Whom?	Group of children, primary caregiver, and aide (when present).
	Age Range	Summative measure: classes/groups with children ages 3-7. Formative measure: infants (0-12 months), toddlers (12-26 months), preschoolers (3-5 years) and school-age children (5-10 years) in each class/group.
Requirements	Training	Training is necessary to establish inter-rater reliability. It involves a review of data collection criteria and methods and on-site practical observation, record reviews and reports. General training takes two to three days (Halle, Whittaker and Anderson, 2010).
	Materials	Manual and score sheets. Table and chair for reporting.
Cost	Training	N/A
	Materials	Summative measure: USD\$50 for the manual and USD\$25 for the score sheets. ² Formative measure: USD\$50 for the manual and USD\$120 for the packet of score sheets for each domain and age range (administrative practices and classroom practices for infants, toddlers, preschoolers and school-age children). ³
	Staff Required	Coders/field observers and the trainer for coders/field observers.

¹ The number of items in each dimension varies depending on the age group observed. The maximum number possible is given in parentheses.

² All prices were retrieved on December 28, 2015 from <http://www.qassist.com/pages/research-and-evaluation>.

³ Score sheets for each age group can be purchased separately for USD\$25.

Table 8: Characteristics of the MITRCC

Instrument		The Missouri Infant/Toddler Responsive Caregiving Checklist (MITRCC)
Design	Authors/Source	MU Center for Family Policy & Research (2003)
	Country of Design	USA
	Country of Implementation	USA Colombia (version translated and adapted to Spanish) Ecuador (version translated and adapted to Spanish) - Araujo et al. (2015)
	Reliability and Validity	Thomburg (2009) uses a sample of 99 children with the version of MITRCC; and report an alpha coefficient for MITRCC = 0.85, and a correlation between ITERS and MITRCC = 0.69.
Measurement	Type of Assessment	Observation (direct or video-based).
	Items	20 items.
	Scoring	Each item is scored using a yes/no format. The number of yes responses (between 0 and 20) is divided by two in order to arrive at an overall score that falls between 0 and 10.
	Administration Time	3 hours and 30 minutes (minimum). At least three hours of observation plus 30 minutes of coding.
	Interviewer Requirements	<p>Specific requirements:</p> <ul style="list-style-type: none"> Complete training and achieve the desired level of reliability for coding <p>General requirements:</p> <ul style="list-style-type: none"> Good interpersonal skills Strong verbal communication skills (clear and articulate) Affinity for children Relevant post-secondary education (e.g., psychology, early childhood education, elementary education) Experience working with children Field experience in qualitative work Good negotiation and problem solving skills Ability to solve unforeseen problems
	Where?	Center.
	To Whom?	Group of children, primary caregiver, and aide (when present).
Age Range	0-36 months.	
Requirements	Training	The completion of formal training with a certified trainer is not required; however, it is recommended that coders/observers participate in training and achieve reliability in the coding process. Although the manual provides some examples, it may be necessary (and helpful) to review additional examples as well as sample videos. ¹
	Materials	Manual, score sheets, projection screen (optional).
Cost	Training	Formal training is not required.
	Materials	The manual and score sheets are freely accessible and of no cost. ²
	Staff Required	Coders/field observers and the trainer for coders/field observers.
Others	The instrument can be coded in the field or with a video recording. For more details on video coding, see section 3.4 of this toolkit.	

¹ One trainer and 15 coders/observers participated in the IDB study performed in Ecuador (Araujo et al., 2013). The total training time was two days (in the field).

² Available at https://www.openinitiative.org/content/pdfs/MoNotes/IT_Checklist_Notes.pdf. The questionnaire translated to Spanish, adapted and used by the authors in Ecuador is available to interested readers.

Table 9: Characteristics of the CC-IT-HOME

Instrument	The Child Care Infant/Toddler Home Observation for Measurement of the Environment (CC-IT-HOME)
Design	Authors/Source Bradley, Caldwell and Corwyn (2003)
	Country of Design USA
	Country of Implementation USA - NICHD (1996, 2000b) ¹ , Clarke-Stewart et al. (2002) Colombia (version translated and adapted to Spanish by NIEER) Ecuador (version translated and adapted to Spanish by NIEER) - Araujo et al. (2015)
	Reliability and Validity The NICHD Study of Early Child Care analyzed a sample of 53 paired observations. Measures of reliability: <ul style="list-style-type: none"> • Inter-rater reliability: there was a very high level of agreement on all of the items. At the 24-month data collection, an analysis of the pairs of scores revealed Pearson correlations ($r = 0.94$) and a Winer correlation ($r = 0.97$) that were both very high. At the 54-month data collection, the Pearson correlations ($r = 0.98$) and the Winer correlation ($r = 0.99$) were still very high (Halle, Whittaker and Anderson, 2010; Bradley et al., 2003, p. 301). • Internal consistency: in general, the scale has a high level of internal consistency, with a Cronbach's alpha of 0.81 (NICHD, 1996).
Measurement	Type of Assessment Direct observation and reporting.
	Dimensions <ol style="list-style-type: none"> 1. Caregiver responsivity (11 items) 2. Acceptance (7 items) 3. Organization (6 items) 4. Learning materials (9 items) 5. Caregiver involvement (6 items) 6. Variety of stimulation (4 items)
	Scoring Each item is scored using a yes/no format, with a total score between 0 and 43.
	Administration Time 1 hour of observation, reporting and coding.
	Interviewer Requirements <p>Specific requirements:</p> <ul style="list-style-type: none"> • Complete the training or practice suggested by the authors of the instrument and achieve the desired level of reliability for coding <p>General requirements:</p> <ul style="list-style-type: none"> • Good interpersonal skills • Strong verbal communication skills (clear and articulate) • Affinity for children • Relevant post-secondary education (e.g., psychology, early childhood education, elementary education) • Experience working with children • Field experience in qualitative work • Good negotiation and problem solving skills • Ability to solve unforeseen problems
	Where? Non-maternal care environment (babysitter or family relatives). In Colombia and Ecuador the instrument was adapted to be used in center-based care
Administration	To Whom? A child and a caregiver.
	Age Range 0-36 months.

Table 9 (continued)	
Instrument	The Child Care Infant/Toddler Home Observation for Measurement of the Environment (CC-IT-HOME)
Requirements	Training Participation in training conducted by certified trainers is not required to use the instrument. Nonetheless, the authors do recommend it, even arguing that it is the norm prior to using the instrument. If the observer cannot complete training, the following steps are essential: 1) carefully read and reread the manual; 2) perform a visit and an interview accompanied by a trained observer, independently code the observations, and compare the results; 3) perform five more visits, critically analyzing the technique used; and 4) repeat step 2. The goal is to achieve 90% agreement between the observers (Caldwell and Bradley, 2003). The authors of the instrument recommend coding sample videos as additional practice to achieve the desired level of reliability. (Caldwell and Bradley, 2003). ²
	Materials Manual and score sheets. Table and chair for reporting.
Cost	Training Formal training is not required.
	Materials USD\$40 for the manual and USD\$15 for the score sheets. ³
	Staff Required Coders/field observers and the trainer for coders/field observers.
Others	If both the CC-IT-HOME and the ITERS-R are being used, a person with a high level of education and specialization in child development could be trained on both instruments simultaneously in order to reduce the length of training.

¹ These studies relate to the IT-HOME, a previous version of the CC-IT-HOME.

² One trainer and 15 coders/observers participated in the IDB study performed in Ecuador (Araujo et al., 2015). The training consisted of a four-day train-the-trainer program at NIEER, with a half day of theory and three-and-a-half days of practice to establish reliability. Coders/observers then received four days of training in the field.

For the NICHD Study of Early Child Care, observers completed a half day of training, followed by practice with the instrument to achieve a reliability level of 90% on the coding of videos. With regard to this study, Bradley et al. (2003) suggest that such intensive training is not generally necessary to achieve the desired reliability.

³ Prices were retrieved on December 28, 2015 from <http://fhdri.clas.asu.edu/home/contact.html>. Score sheets are available in packs of 50 booklets. The questionnaire translated to Spanish, adapted and used by the authors in Ecuador is available to interested readers.

Table 10: Characteristics of the Routines Assessment Tool

Instrument		Routines
Design	Authors/Source	IDB (2012)
	Country of Implementation	Ecuador - Araujo et al. (2015)
	Reliability and Validity	Not available, however, the Feeding Scale (NCAST) has parameters for internal, external and predictive validity plus adequate reliability.
Measurement	Type of Assessment	Direct observation.
	Domains and Dimensions	<ol style="list-style-type: none"> 1. Feeding, based on the Feeding Scale (NCAST, 2012) that was administered in the U.S. <ol style="list-style-type: none"> 1.1 Details of the routine such as duration and the presence of others (8 items) 1.2 Interaction between the child and caregiver during the meal (20 items) 2. Diapering/Toileting <ol style="list-style-type: none"> 2.1 Interaction between the child and caregiver during diapering or toileting (22 items) 3. Crying, abuse, or disconnection, based on the Infant-Toddler and Family Instrument (Provence and Apfel, 2001), Family Care Instrument (FCI), and the PRIDI (Engle, Cueto, Ortiz and Verdisco [2011] and IDB Advisory Committee, January 2012) <ol style="list-style-type: none"> 3.1 State of the child before, during and after crying, abuse or disconnection (16 items) 3.2 Caregiver reaction to crying, abuse or disconnection (5 items)
	Scoring	Each item is scored using a yes/no format
	Administration Time	90 minutes. 40 minutes for the feeding dimension, 10 minutes for the diapering/toileting dimension, and 40 minutes for the crying/abuse/disconnection dimension.
	Interviewer Requirements	<p>General requirements:</p> <ul style="list-style-type: none"> • Good interpersonal skills • Strong verbal communication skills (clear and articulate) • Affinity for children • Relevant post-secondary education (e.g., psychology, early childhood education, elementary education) • Experience working with children • Field experience in qualitative work • Good negotiation and problem solving skills • Ability to solve unforeseen problems
	Administration	Where?
To Whom?		A caregiver and a child, depending on the situation of the observed routine (a child who does NOT eat independently; a child who does NOT go to the bathroom independently; a child who cries, has been abused, or who disconnects from his surroundings, the caregiver or other children).
Age Range		Depending on the dimension: 0-24 months for feeding, 0-30 months for diapering/toileting, 0-60 months for crying/abuse/disconnection.
Requirements	Training	Two days of training are required.
	Materials	Manual and score sheets.
Cost	Training	Formal training is not required.
	Materials	Manual and score sheets are freely accessible and of no cost (consult the authors of this toolkit).
	Staff Required	Coders/field observers and the trainer for coders/field observers.

Table 11: Characteristics of the Caregiver/Coordinator/Structural Questionnaire

Instrument		Caregiver, Coordinator and Structural Questionnaire	
Design	Authors/Source	IDB (2012)	
	Country of Implementation	Ecuador - Araujo et al. (2015)	
	Reliability and Validity	N/A	
Measurement	Type of Assessment	Direct observation and reporting.	
	Sections	<ol style="list-style-type: none"> 1. Caregiver <ol style="list-style-type: none"> 1.1 Caregiver data (28 items) 1.2 Knowledge of Infant Development Inventory (KIDI) (58 items) 1.3 Child care practices inventory, based on an instrument designed in the U.S. to operationalize National Association for the Education of Young Children (NAEYC) guidelines (23 items) 1.4 The caregiver-child relationship, based on the Infant-Toddler and Family Instrument (ITFI, 2001) (11 items) 2. Coordinator <ol style="list-style-type: none"> 2.1 Child care staff data (12 items) 2.2 Coordinator data (10 items) 2.3 Child care center data (20 items) 2.4 Knowledge of Infant Development Inventory (KIDI) (58 items) 2.5 Discipline inventory (2 items) 3. Structural <ol style="list-style-type: none"> 3.1 General information (9 items) 3.2 Basic services (7 items) 3.3 Infrastructure (4 items) 3.4 Classroom, play and other areas vulnerabilities (68 items) 3.5 Possible risks for children (11 items) 3.6 Existing capacity (6 items) 3.7 Other (1 item) 	
	Scoring	Open-ended or pre-coded responses.	
	Administration Time	2 hours and 40 minutes. 40 minutes for the caregiver report, 1 hour for the coordinator report, 40 minutes of center observation, and 20 minutes for the center report.	
	Interviewer Requirements	General requirements: <ul style="list-style-type: none"> • Good interpersonal skills • Strong verbal communication skills (clear and articulate) • Affinity for children • Relevant post-secondary education (e.g., psychology, early childhood education, elementary education) • Experience working with children • Field experience in qualitative work • Good negotiation and problem solving skills • Ability to solve unforeseen problems 	
	Administration	Where?	Center.
		To Whom?	Caregiver, coordinator, or center, depending on the dimension.
Requirements	Duration of Training	2 days.	
	Materials	Manual and score sheets. Table and chair.	
Cost	Training	Formal training is not required.	
	Materials	Manual and score sheets are freely accessible and of no cost (consult the authors of this toolkit).	
	Staff Required	Field interviewers and trainer for field interviewers.	

Table 12: Summary of the Instruments

Instruments	Measurement				
	Domains ¹	Number of Dimensions	Dimensions/Subscales ¹	Scoring of Each Item	Minimum Time
CLASS Toddler	1. Emotional and behavioral support (P) 2. Engaged support for learning (P)	8	1.1 Positive climate (P) 1.2 Negative climate (P) 1.3 Teacher sensitivity (P) 1.4 Regard for child perspectives (P) 1.5 Behavior guidance (P) 2.1 Facilitation of learning and development (P) 2.2 Quality of feedback (P) 2.3 Language modeling (P)	1-7	2 hours
ITERS-R	N/A	7	1. Space and furnishings (S) 2. Personal care routines (P) 3. Listening and talking (P) 4. Activities (P) 5. Interaction (P) 6. Program structure (P) 7. Parent and staff (P)	1-7	3 hours and 30 minutes
CC-IT-HOME	N/A	6	1. Caregiver responsivity (P) 2. Acceptance (P) 3. Organization (P) 4. Learning materials (S) 5. Caregiver involvement (P) 6. Variety of stimulation (P)	yes/no	1 hour
MITRCC	N/A	N/A	N/A	yes/no	3 hours and 30 minutes
Routines	1. Feeding (P) 2. Diapering/toileting (P) 3. Crying/abuse/disconnection (P)	5	1.1 Details of the routine (P) 1.2 Interaction between the child and caregiver during the meal (P) 2.1 Interaction between the child and caregiver during diapering/toileting (P) 3.1 State of the child (P) 3.2 Caregiver reaction (P)	yes/no	90 minutes

Table 12 (continued)														
Instruments	Where Can It Be Administered?	To Whom?	Administration						Human Capital			Total Cost (in USD) ³		
			What ages? ²				How?		Minimum Interviewer Requirements					
			Children 0-15 Months	Children 15-36 Months	Children 36-60 Months	Children 60 Months and Up	Direct Observation	Report	Video-based Observation	None	Secondary		Post-secondary	
CLASS Toddler	Center	Group of children, primary caregiver, and aide (when present)		x				x		x				902.90
ITERS-R	Center	Setting, group of children, primary caregiver, and aide (when present)	x	x				x	x				x	22.90
CC-IT-HOME	Non-maternal care setting	A child and a caregiver	x	x				x	x				x	40.30
MITRCC	Center	Group of children, primary caregiver, and aide (when present)	x	x				x		x			x	0.00
Routines	Center	A child and a caregiver	x	x	x			x					x	0.00

Table 12 (continued)

Instruments	Measurement				
	Domains ¹	Number of Dimensions	Dimensions/Subscales ¹	Scoring of Each Item	Minimum Time
ORCE	1. Behavior scales (P) 2. Qualitative ratings (P) 3. Structural variables (S)	15 ⁴	1.1 Positive and negative affect ⁵ (P) 1.2 Language-focused interaction (P) 1.3 Stimulation (P) 1.4 Behavior management (P) 1.5 Child's activity (P) 1.6 Child's interaction with other children (P) 2.1 Caregiver notes (P) 2.2 Child notes (P) 3.1 Ratio (S) 3.2 Group size (S) 3.3 Numbers of children (S) 3.4 Numbers of adults available (S) 3.5 Proportion of observation completed outdoors (S) 3.6 Amount of time caregiver is involved with child 3.7 Age mix of the group (S)	1-4	90 minutes
CIS	N/A	4	1. Sensitivity (P) 2. Harshness (P) 3. Detachment (P) 4. Permissiveness (P)	1-4	90 minutes ⁶
Assessment Profile - Summative Measure	1. Classroom practices (P)(S)	5	1.1 Learning environment (S) 1.2 Scheduling (S) 1.3 Curriculum methods (S) 1.4 Interacting (P) 1.5 Individualizing (P)	yes/no	3 hours ⁷
Assessment Profile - Formative Measure	1. Classroom practices (P) (S) 2. Administrative practices (S)	11	1.1 Safety (S) 1.2 Learning environment (S) 1.3 Scheduling (S) 1.4 Curriculum methods (S) 1.5 Interacting (P) 1.6 Individualizing (P) 2.1 Physical facilities (S) 2.2 Food service (S) 2.3 Program management (S) 2.4 Personnel (S) 2.5 Program development (S)	yes/no	7 hours

¹ This table considers the construct of domain or dimension as defined by the manual for each instrument. The type of variable is noted in parentheses: process (P) or structural (S).

² The proposed age ranges do not represent the exact ranges for which the following instruments are applicable: ITERS-R, Feeding/Diapering/Toileting, ORCE and the Assessment Profile - Formative. For further details, see the summary table for each of the instruments.

³ Corresponds to the cost of materials. The only exception is CLASS Toddler, for which the cost of mandatory official training is included.

⁴ The dimensions associated with the domains "Behavior scales" and "Qualitative ratings" correspond to the version for children 6 months of age. The versions for 15, 24, 36 and 54 months include additional or different dimensions.

Table 12 (continued)													
Instruments	Where Can It Be Administered?	To Whom?	Administration						Human Capital		Total Cost (in USD) ³		
			What ages? ²				How?		Minimum Interviewer Requirements				
			Children 0-15 Months	Children 15-36 Months	Children 36-60 Months	Children 60 Months and Up	Direct Observation	Report	Video-based Observation	None		Secondary	Post-secondary
ORCE	Non-maternal care setting	A child and a caregiver or other adult present	x	x	x		x					x	N/A
CIS	Non-maternal care setting	Group of children, primary caregiver, and aide (when present)			x		x					x	N/A
Assessment Profile - Summative Measure	Center	Group of children, primary caregiver, and aide (when present)				x	x	x				x	43.00
Assessment Profile - Formative Measure	Center	Group of children, primary caregiver, and aide (when present)	x	x	x	x	x	x				x	N/A

⁵ The instrument has four versions (6, 15, 24, 36, and 54 months), one for each time point when data was collected (the 24- and 36-month versions are the same). Each version is divided into the same three domains; however, there are differences among the dimensions. The 6-month scales are presented here (see Halle, Whittaker and Anderson [2010] for the other scales derived from this one).

⁶ These 90 minutes must be broken into two separate observation cycles of 45 minutes each (Vandell and Wolfe, 2010).

⁷ Two or three groups are observed/reported on in one day. A work day is considered to be eight hours.

Table 13: Comparison of Instruments

Instruments	Number of Dimensions	Minimum Administration Time		Total cost ¹		Official Training	Administered in/adapted to a country in LAC	Spanish Translation	Validity ²	Educational level of the interviewer
		Time	Order (from least to greatest amount of time)	Cost (in USD)	Order (from lowest to highest cost)					
CLASS Toddler	8	2 hours	3	902.9	6	Yes	Yes	Yes	Yes	Post-secondary
ITERS-R	7	3 hours and 30 minutes	5	22.9	2	No	Yes	Yes ³	Yes	Post-secondary
CC-IT-HOME	6	1 hour	1	40.3	3	No	Yes	Yes	Yes	Post-secondary
MITRCC	N/A	3 hours and 30 minutes	5	0	1	No	Yes	Yes	Yes	Post-secondary
Routines	3	1 hour and 30 minutes	2	0	1	No	Yes	Yes	N/A	Secondary
ORCE	3	1 hour and 30 minutes	2	N/A	N/A	No	No	No	Yes	Post-secondary
CIS	4	1 hour and 30 minutes	2	N/A	N/A	No	No	No	Yes	Post-secondary
Assessment Profile - Summative Measure	5	3 hours ⁴	4	75	4	No	No	No	Yes	Post-secondary
Assessment Profile - Formative Measure	11	7 hours	6	170	5	No	No	No	Yes	Post-secondary

¹ Corresponds to the cost of materials per observer/coder. The only exception is CLASS Toddler, for which the cost of mandatory official training for observers is included.

² Corresponds to the existence of studies on the validity and reliability of the instrument.

³ Published by Teachers College Press.

⁴ Two or three groups are observed/reported on in one day. A work day is considered to be eight hours.

References

- Abbott-Shim, M. (1991). *Quality care: A global assessment*. Unpublished manuscript. Georgia State University.
- Abbott-Shim, M., Lambert, R. and McCarty, F. (2000). Structural model of Head Start classroom quality. *Early Childhood Research Quarterly*. 15(1): 115-134.
- Abbott-Shim, M., Neel, J. and Sibley, A. (2001). *Assessment Profile for Early Childhood Programs - Research Edition II: Technical Manual*. Atlanta, GA: Quality Counts, Inc.
- Annual Professional Performance Review (APPR). (2012). <http://goo.gl/orIqwB>
- Araujo, M.C., López Boo, F. and Puyana, J.M. (2012). "Overview of Early Childhood Development Services in Latin America and the Caribbean." IDB-MG-149, Division of Social Protection and Health, Inter-American Development Bank.
- Araujo, M.C., López Boo, F., Novella R., Schodt S., and Tomé, R. (2013). "La calidad de los Centros Infantiles del Buen Vivir en Ecuador." Resumen de políticas DB-PB-248.
- Arnett, J. (1989). Caregivers in day-care centers: Does training matter? *Journal of Applied Developmental Psychology*. 10: 541-522.
- Arredondo, S.A., Del Pilar Torres Pereda, M., Kenefick, S., Neufeld Lynnette, M., De la Peza, P.G. and Valencia, A. (2011). "Evaluación de impacto del programa de estancias infantiles para apoyar a madres trabajadoras." CIEE.
- Attanasio, O., Duque, J. F., Fitzsimmons, E., Grantham-McGregor, S., Meghir, C. and Rubio-Codina, M. (2012). Early Childhood Development Pilot in Colombia: Home Visiting & Nutrition. Center for the Evaluation of Development Policies at IFS and UCL, London. Accessed April 25, 2012. <http://www.theigc.org/wp-content/uploads/2014/08/ecd-pilot-in-colombia-for-india-igc.10.pdf>
- 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.
- Barnett, W. (2012). "What is Quality?" Presentation at IDB Advisory Committee, January 2012, Washington, DC.
- Barros, R., de Carvalho, M., Franco, S., Mendonca, R. and Rosalem, A. (2011). "A Short-Term Cost Effectiveness Evaluation of Better-Quality Daycare Centers." Inter-American Development Bank.
- Behrman, J., Cheng, Y. and P. Todd. (2004). "Evaluating Preschool Programs When Length of Exposure to the Program Varies: A Nonparametric Approach." *Review of Economics and Statistics*. 86(1): 108-132.
- Belsky, J. (1988). "The effects of infant day care reconsidered." *Early Childhood Research Quarterly*. 3: 237-272.
- Berlinski, S., Galiani, S. and Gertler, P. (2009). "The effect of pre-primary education on primary school performance." *Journal of Public Economics*. 93:219-234.
- Berlinski, S., Galiani, S. and MacEwan, P. (2008). "Pre-school and maternal labor outcomes: evidence from a regression discontinuity design." IFS Working Paper. W09/05.

- 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.
- Bernal, R., Fernández, C., Flórez, C., Gaviria, A., Ocampo, P., Samper, B. and Sánchez, F. (2009). *Evaluación de impacto del Programa hogares comunitarios de bienestar del ICBF*. CEDE Documents, ISSN 1657-7191.
- Bowlby, J. (1969). Attachment. Attachment and loss: Vol. 1. Loss. New York: Basic Books.
- 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.
- Burchinal, M., Campbell, F., Bryant, D., Wasik, B. and Ramey, C. (1997). "Early intervention and mediating processes in cognitive performance of children of low-income African American families." *Child Development*. 68: 935-954.
- Burchinal, M., Kainz, K. and Cai, Y. (2011). "How well do our measures of quality predict child outcomes?" In Zaslow et al. Eds. *Quality Measurement in Early Childhood Settings* Brookes Publishing, Washington, DC.
- Caldwell, B., and Bradley, R. (2003). Home Observation for Measurement of the Environment: Administration Manual. Little Rock, AR: University of Arkansas.
- Campbell, F. and Ramey, C. (1995). "Cognitive and school outcomes for high-risk African-American students at middle adolescence: Positive effects of early intervention." *American Educational Research Journal*. 32: 743-772.
- Clarke-Stewart, K. and Fein, G. (1983). "Early Childhood Programs." In P. H. Mussen, Ed. (M. M. Haith and J. J. Campos, series Eds.). *Handbook of Child Psychology, Volume II: Infancy and Developmental Psychobiology*. New York: John Wiley and Sons.
- Clarke-Stewart, K., Gruber, C. and Fitzgerald, L. (1994). *Children at Home and in Day Care*. Hillsdale, NJ: Erlbaum.
- Clarke-Stewart, K., Vandell, D., Burchinal, M., O'Brien, M. and McCartney, K. (2002). "Do regulable features of child-care homes affect children's development?" *Early Childhood Research Quarterly*. 17:52-86.
- Cryer, T., Harms, T. and Riley, C. (2004a). "All About the ITERS-R." Lewisville, NC: PACT House Publishing.
- Cryer, T., Harms, T. and Riley, C. (2004b). "All About the ECERS-R." Lewisville, NC: PACT House Publishing.
- Denham, S., Ji, P. and Hamre, B. (2010). *Compendium of Preschool Through Elementary School Social-Emotional Learning and Associated Assessment Measures*. University of Illinois, Chicago.
- Dunn, L. (1993). "Proximal and Distal Features of Day Care Quality and Children's Development." *Early Childhood Research Quarterly*. 8: 167-192.
- Engle, P., Black, M., Behrman, J., Cabral de Mello, M., Gertler, P., Kapirini, L., Martrell, R. and Eming Young, M. (2007). "Strategies to avoid the loss of developmental potential in more than 200 million children in the developing world." *The Lancet*. 369:229-242.
- Engle, P., Cueto, S., Ortíz, M.E. and Verdisco, A. (2011). "Programa Regional de Indicadores de Desarrollo Infantil (PRIDI). Marco Conceptual." Inter-American Development Bank.

- Engle, P., Fernald, L., Alderman, H., Behrman, J., O’Gara, C., Yousafzai, A., et al. (2011). “Strategies for reducing inequalities and improving developmental outcomes for young children in low and middle income countries.” *The Lancet*. 378(9799):1339-1353.
- Fernald, L., Gertler, P. and Neufeld, L. (2009). “10 year effect of Oportunidades, Mexico’s conditional cash transfer programme, on child growth, cognition, language, and behavior: a longitudinal follow-up study.” *The Lancet*. 374:1997-2005.
- Fernald, L., Kariger, P., Engle, P. and Raikes, A. (2009). “Examining Early Child Development in Low-Income Countries: A toolkit for the Assessment of Children in the First Five Years of life.” The World Bank.
- Filp, J. and Scheifelbein, E. (1982). “Efecto de la educación preescolar en el rendimiento de primer grado de primaria: El estudio UMBRAL en Argentina, Bolivia, Colombia y Chile.” *Revista Latinoamericana de Estadística Educación*. 12: 9-41.
- Filp, J., Donoso, S., Cardemil, C., Dieguez, E., Torres, J. and Sheifelbein, E. (1983). “Relationship between preprimary and grade one primary education in state schools in Chile.” In K. King and R. Meyers, Eds. *Preventing School Failure: The Relationship Between Preschool and Primary Education*. Ottawa: IDRC.
- Fundación Van Leer. (2012). *Early Childhood Matters*.
- Grantham-McGregor, S., Cheung, Y. B., Cueto, S., Glewwe, P., Richter, L. and Strupp, B. (2007). “Developmental potential in the first 5 years for children in developing countries.” *The Lancet*. 369:60–70.
- Grantham-McGregor, S., Powell, C., Walker, S., et al. (1991). “Nutritional Supplementation, Psychosocial Stimulation, and Mental Development of Stunted Children: The Jamaican Study.” *The Lancet*. 338(8758): 1-5.
- Grantham-McGregor, S., Walker, S., Chang, S., et al. (1997). “Effects of Early Childhood Supplementation With and Without Stimulation on Later Development in Stunted Jamaican Children.” *American Journal of Clinical Nutrition*. 66: 247-253.
- Greenberg, M., Domitrovich, C. and Bumbarger, B. (2001). “The prevention of mental disorders in school-aged children: Current state of the field.” *Prevention & Treatment*, 4, Article 1.
- Greenberg, M., Weissberg, R., O’Brien, M., Zins, J., Fredericks, L., Resnik, H. and Elias, M. (2003). “School-based prevention: Promoting positive social development through social and emotional learning.” *American Psychologist*. 58(6/7): 466-474.
- Gonzalez-Mena, J. and Widmeyer Eyer, D. (2007). *Infants, Toddlers, and Caregivers: A Curriculum of Respectful, Responsive Care and Education*. McGraw-Hill Education.
- Gwet, K. (2008). “Computing inter-rater reliability and its variance in the presence of high agreement.” *British Journal of Mathematical and Statistical Psychology*. 61: 29–48: The British Psychological Society.
- Halle, T., Anderson, R., Blasberg, A., Chrisler, A. and Simkin, S. (2011). *Quality of Caregiver-Child Interactions for Infants and Toddlers (Q-CCIT): A Review of the Literature*, OPRE 2011 – 25. Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.
- Halle, T., Vick Whittaker, J.E. and Anderson, R. (2010). *Quality in Early Childhood Care and Education Settings: A Compendium of Measures, Second Edition*. Washington, DC: Child Trends. Prepared by Child Trends for the Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

- Halpern, R. (1986). "Effects of Early Childhood Intervention on Primary School Progress in Latin America." *Comparative Education Review*. 30 (2): 193-215. The University of Chicago Press on behalf of the Comparative and International Education Society. Accessed March 12, 2012. <http://www.jstor.org/stable/1188528>
- Hamre, B. and Pianta, R. (2007). Learning opportunities in preschool and early elementary classrooms. In R.C. Pianta, M.J. Cox and K.L. Snow, Eds. *School readiness & the transition to kindergarten in the era of accountability*. 49-83. Baltimore, MD: Paul H. Brookes Publishing Co.
- Hanushek, E. (1998). "Conclusion and controversies about the effectiveness of school resources." *Economic Policy Review*, Federal Reserve Bank of New York. 4(1): 11-28.
- Hanushek, E. (2003). "The failure of input based schooling policies." *Economic Journal*. 113: 64-98.
- Harms, T., and Clifford, R. (1980). *The early childhood environment rating scale*. New York: Teachers College Press.
- Harms, T., Cryer, T. and Clifford, R. (2006). *Infant/Toddler Environment Rating Scale Revised Edition – ITERS-R*, Teachers College Press.
- Hausfather, A., Toharia, A., LaRoche, C. and Engelsmann, F. (1997). "Effects of Age of Entry, Day-Care Quality, and Family Characteristics on Preschool Behavior." *Journal of Child Psychology and Psychiatry and Allied Disciplines*. 38: 441-448.
- Heckman, J. (2006). "Skill Formation and the Economics of Investing in Disadvantaged Children." *Science*. 312(5782): 1900-1902.
- Helburn, S. editor. (1995). *Cost, Quality and Child Outcomes in Child Care Centers: Technical Report*. Denver, CO: Dept of Economics, Center for Research in Economic and Social Policy, University of Colorado.
- Herrera, M., Mathiesen, M., Merino, J. and Recart, I. (2005). "Learning contexts for young children in Chile: process quality assessment in preschool centres." *International Journal of Early Years Education*. 13 (1): 13-27.
- Howes, C. (1983). "Caregiver Behavior in Center and Family Day Care." *Journal of Applied Developmental Psychology*. 4: 99-107.
- Howes, C. (1997). "Children's Experiences in Center-Based Child Care as a Function of Teacher background and adult:child ratio." *Merrill-Palmer Quarterly: Journal of Developmental Psychology*. 43(3): 404-425.
- Howes, C. and Hamilton, C. (1993). "The changing experience of child care: Changes in teachers and in teacher-child relationships and children's social competence with peers." *Early Childhood Research Quarterly*. 8: 15-32.
- Howes, C., Phillips, D. and Whitebook, M. (1992). "Thresholds of quality: Implications for the social development of children in center-based child care." *Child Development*. 63: 449-460.
- Howes, C. and Smith, E. (1995). "Relations among Child Care Quality, Teacher Behavior, Children's Play Activities, Emotional Security, and Cognitive Activity in Child Care." *Early Childhood Research Quarterly*. 10: 381-404.

- IDB - Advisory Committee on Quality of Center-Based Care for Infants and Toddlers (2012). *Measuring Quality of Center-Based Care for Infants and Toddlers in Latin America and the Caribbean, Summary of Main Agreements*, Inter-American Development Bank, Washington, DC, 20 January 2012.
- Jaeger, E. and Funk, S. (2001). *The Philadelphia Child Care Quality Study: An Examination of Quality in Selected Early Education and Care Settings*. Philadelphia, PA: Saint Joseph's University.
- Kagan, S.L. (2010). "Promoting ECD Quality: Policy Lessons for Central America." Presentation at the IDB, November 2010, El Salvador.
- Kisker, E., Hofferth, S., Phillips, D. and Farquhar, E. (1991). *A Profile of Child Care Settings: Early Education and Care in 1990, Volume 1*. Princeton, NJ: Mathematica Policy Research.
- Kontos, S. and Wilcox-Herzog, A. (1997). "Influences on Children's Competence in Early Childhood Classrooms." *Early Childhood Research Quarterly*. 12: 247-262.
- La Paro, K., Pianta, R. and Stuhlman, M. (2004). "The classroom assessment scoring system: Findings from the prekindergarten year." *Elementary School Journal*. 104: 409-426.
- La Paro, K., Hamre, B. and Pianta, R. (2011). *Classroom Assessment Scoring System. Toddler Manual*. Pre-publication advance review draft. Teachstone.
- Lamb, M. (1996). "Effects of Non-Parental Child Care on Child Development: An Update." *The Canadian Journal of Psychiatry*. 41: 330-342.
- Lamb, M. (1998). "Nonparental Child Care: Context, Quality, Correlates, and Consequences." In W. Damon, I. E. Sigel, and K. A. Renninger, Eds. *Handbook of Child Psychology. Vol. 4: Child Psychology in Practice*. New York: John Wiley.
- Layzer, J., Goodson, B., and Moss, M. (1993). *Observational Study of Early Childhood Programs, Final Report, Volume I: Life in Preschool*. Cambridge, MA: Abt Associates, Inc.
- Litwin, M. (1995). *How to measure survey reliability and validity*. The Survey Kit, Vol 7. SAGE Publications.
- Love, J., Harrison, L., Sagi-Schwartz, A., van Ijzendoorn, M., Ross, C. and Ungerer, J. (2003). "Child care quality matters: How conclusions may vary with context." *Child Development*. 74: 969-1226.
- Love, J., Kisker, E., Ross, C., Raikes H., Constantine, J., Boller, K., Brooks-Gunn, J., Chazan-Cohen, R., Tarullo, L., Brady-Smith, C., Fuligni, A., Schochet, P., Paulsell, D. and Vogel, C. (2005). "The effectiveness of early head start for 3-year-old children and their parents: lessons for policy and programs." *Dev Psychol*. 41(6):885-901.
- MacPhee, D. (1981). *Knowledge of Infant Development Inventory. Survey of child care experiences - Manual*.
- Martinez-Beck, I. (2011). "Introduction: Why Strengthening the Measurement of quality in Early Childhood Settings Has Taken on New Importance". In Zaslow et al., Eds. *Quality Measurement in Early Childhood Settings*. Paul H. Brookes Publishing Co.
- McKay, A. and McKay, H. (1983). "Primary School Progress after Preschool Experience: Troublesome Issues in the Conduct of Follow-Up Research and Findings from the Cali, Colombia, Study." In M. Herrera and C. Super. *School Performance and Physical Growth of Underprivileged Children: Results of the Bogota Project at Seven Years*, Report to the World Bank, Cambridge, MA: Harvard School of Public Health.

- MU Center for Family Policy & Research. (2003). Missouri Infant/Toddler Responsive Caregiving Checklist. <http://goo.gl/RyjAqy>
- National Association for the Education of Young Children (NAEYC). (2012). Accessed March 22, 2012. <http://www.naeyc.org/>
- NCAST. (2012). Parent-Child Interaction (PCI) Feeding & Teaching Scales. Accessed April 18, 2012. <http://ncast.org/index.cfm?category=2>
- Nelson, G., Westhues, A. and MacLeod, J. (2003). "A meta-analysis of longitudinal research on preschool prevention programs for children." *Prevention & Treatment*. 6 (1). Retrieved from https://www.researchgate.net/publication/232570779_A_Meta-Analysis_of_Longitudinal_Research_on_Preschool_Prevention_Programs_for_Children
- NICHD Early Child Care Research Network. (1996). "Characteristics of Infant Child Care: Factors Contributing to Positive Caregiving." *Early Childhood Research Quarterly*. 11: 269-306.
- NICHD Early Child Care Research Network. (1999). "Child Outcomes when Child Care Center Classes meet Recommended Standards for Quality." *American Journal of Public Health*. 89(7): 1072-1077.
- NICHD Early Child Care Research Network. (2000a). "Characteristics and quality of child care for toddlers and preschoolers." *Applied Developmental Science*. 4: 116-135.
- NICHD Early Child Care Research Network. (2000b). "The relation of child care to cognitive and language development." *Child Development*. 71: 960-980.
- NICHD Early Child Care Research Network (2002a). "Early child care and children's development prior to school entry: Results from the NICHD Study of Early Child Care." *American Education Research Journal*. 39:133-164.
- NICHD Early Child Care Research Network. (2002b). "The relation of global first- grade classroom environment to structural classroom features and teacher and student behaviors." *Elementary School Journal*. 102(5): 367-387.
- NICHD Early Child Care Research Network. (2009). "Study of Early Child Care and Youth Development Phase I Instrument Document Observational Record of the Caregiving Environment (ORCE): Behavior Scales, Qualitative Scales, and Observed Structural Variables for 6, 15, 24, & 36 months."
- Office of Head Start. (2008). Head Start. Accessed November 10, 2009. <http://www.acf.hhs.gov/programs/ohs/>
- Peisner-Feinberg, E., Burchinal, R., Clifford, R., Culkin, M., Howes, C., Kagan, S. and Yazejian, N. (2001). "The Relation of Preschool Child-Care Quality to Children's Cognitive and Social Developmental Trajectories through Second Grade." *Child Development*. 72(5): 1534-1553.
- Peña, E. (2007). "Lost in translation: Methodological considerations in cross-cultural research." *Child Development*. 78(4): 1255-1264.
- Phillips, D., editor. (1987). *Quality in Child Care: What Does Research Tell Us?* Washington, DC: National Association for the Education of Young Children.
- Phillipsen, L. C., Burchinal, M. R., Howes, C. and Cryer, D. (1997). "The prediction of process quality from structural features of child care". *Early Childhood Research Quarterly*. 12: 281-303.

- Pianta, R. (2003). "Experiences in p-3 classrooms: The implications of observational research or re-designing early education." *Foundation for Child Development*, New York.
- Pianta, R. (2011). *Teaching Children Well: New Evidence-Based Approaches to Teacher Professional Development and Training*. Center for American Progress.
- Pianta, R., Howes, C., Burchinal, M., Bryant, D., Clifford, R., Early, D. and Barbarin, O. (2005). Features of pre-kindergarten programs, classrooms, and teachers: Do they predict observed classroom quality and child-teacher interactions? *Applied Developmental Science*. 9(3): 144–159.
- Pianta, R. C., La Paro, K. M. and Hamre, B. K. (2008). Classroom assessment scoring system, manual, K–3. Baltimore, MD: Brookes Publishing.
- PREAL. (2006a). *Quantity Without Quality: A Report Card on Latin America*. Accessed March 11, 2012. <http://elibrary.kiu.ac.ug:8080/jspui/bitstream/1/1465/1/Quantity%20without%20Quality%20-%20Education%20in%20Latin%20America.pdf>
- Provence, S. and Apfel, N. (2001). "Infant-toddler and Family Instrument (ITFI)." Paul H. Brookes Publishing Co. Baltimore.
- Raikes, H. and Edwards, C. (2009). Extending the dance in infant and toddler caregiving. Baltimore, MD: Brookes Publishing Company.
- Reyes Ugalde, J. and Schodt, S. (2011). "Manual de registro de video UBC 2011." Working paper.
- Rolnick, A. and Grunewald, R. (2007). "Early Intervention on a Large Scale." Accessed March 28, 2012. <https://www.minneapolisfed.org/publications/special-studies/early-childhood-development/early-intervention-on-a-large-scale>
- Rosero, J. and Oosterbeck, H. (2011). "Trade-offs between different early childhood interventions: Evidence from Ecuador." IADB Unpublished Manuscript.
- Schweinhardt, L., Montie, J., Xiang, Z., Barnett, W., Belfield, C. and Nores, M. (2005). "Lifetime effects: The High/Scope Perry Preschool study through age 40." Monographs of the High/Scope Educational Research Foundation, 14. Ypsilanti, MI: High/Scope Educational Research Foundation.
- Schodt, S. (2012). Nota conceptual del proyecto de diagnóstico de calidad de los CIBV.
- Shonkoff, J. and Phillips, D., editors. (2000). *From Neurons to Neighbourhoods: The Science of Early Childhood Development*. National Academy Press, Washington, DC.
- Teachstone. (2012a). The Head Start Professional Development and CLASS Initiative: Ensuring Strong Linkages between Program Monitoring and Professional Development. Accessed April 10, 2012. <http://www.vbgrowsmart.com/providers/Documents/CLASSImplementationGuide.pdf>
- Teachstone. (2012b). Virginia's Star Quality Initiative: Careful Planning and Infrastructure Development. Accessed April 10, 2012. <http://www.vbgrowsmart.com/providers/Documents/CLASSImplementationGuide.pdf>
- Teachstone. (2012c). About the CLASS. Accessed April 11, 2012. www.teachstone.com
- Thomason, A. and La Paro, K. (2009). "Measuring the quality of teacher-child interactions in toddler child care." *Early Education and Development*. 20(2):285-304.
- Urzúa, S. and Veramendi, G. (2011). "The impact of Out-of-Home Childcare Centers on Early Childhood Development." Inter-American Development Bank. Unpublished manuscript.

- Vandell, D. and Wolfe, B. (2000). "Child care quality: Does it matter and does it need to be improved?" Reported prepared for the US Department of Health and Human Services, Office for Planning and Evaluation. *Child Development*. 63: 938–949.
- Van de Vijver, F. and Tanzer, N. (2004). "Bias and equivalence in cross-cultural assessment: An overview." *Revue européenne de psychologie appliquée*. Elsevier. 54: 119–13.
- Van Widenfelt, B., Treffers, P., de Beurs, E., Siebelink, B. and Koudijs, E. (2005). "Translation and cross-cultural adaptation of assessment instruments used in psychological research with children and families." *Clinical Child and Family Psychology Review*. 8(2):135-147.
- Verdisco, A. and Pérez Alfaro, M. (2010). "Measuring education quality in Brazil." Inter-American Development Bank. <https://publications.iadb.org/handle/11319/3100?locale-attribute=en>
- Vernon-Feagans, L., Emanuel, D. and Blood, I. (1997). "The Effect of Otitis Media and Quality Daycare on Children's Language Development." *Journal of Applied Developmental Psychology*. 18: 395–409.
- Volling, B. and Feagans, L. (1995). "Infant day care and children's social competence." *Infant Behavior and Development*. 18:177–188.
- Walker, S., Chang, S., Vera-Hernández, M. and Grantham-McGregor, S. (2011). "Early Childhood Stimulation Benefits Adult Competence and Reduces Violent Behavior." *Pediatrics*. 127(5): 849 -857 (doi: 10.1542/peds.2010-2231): <http://pediatrics.aapublications.org/content/127/5/849>
- Walker, S., Chang, S., Grantham-McGregor, S., Osmond, C. and López Boo, F. (2012). "Do early childhood experiences affect development in the next generation: Preliminary findings from the Jamaica intergenerational study." Research presentation at "Early Childhood Development and Human Capital Accumulation" conference at UCL, London, June 25-26, 2012.
- Weber, K. and Kariger, P. (2011). "Instrumentos para la evaluación de calidad de centros y desarrollo infantil." Presentation at the IDB, December 2011, Nicaragua.
- Wilkes, D. (1989). "Administration, classroom program, sponsorship: Are these indices of quality care in day care centers." (Doctoral Dissertation, Georgia State University, 1989). *Dissertation Abstracts International* 50, AAI8922912.
- Yamauchi, C. and Leigh, A. (2011). Which children benefit from non-parental care? *Economics of Education Review*. 30(6): 1468-1490, ISSN 0272-7757.
- Yoshikawa, H., Barata, M., Rolla, A., da Silva, C., Ayoub, C., Arbour, M. and Snow, C. (2008). "Un Buen Comienzo, an initiative to improve preschool education in Chile: Data from the initial year of implementation." Santiago, Chile: Unicef.
- Zaslow, M., Halle, T., Martin, L., Cabrera, N., Calkins, J., Pitzer L., and Geyelin, N. (2006). "Child Outcome Measures in the Study of Child Care Quality." *Eval Rev*; 30; 577.
- Zill, N., Resnick, G., Kim, K., O'Donnell, K., Sorongon, A., McKey, R. H. et al. (2003). "Head Start FACES 2000: A whole-child perspective on program performance." Washington, DC: U.S. Department of Health and Human Services, Administration on Children, Youth, and Families.
- Zins, J., Bloodworth, M., Weissberg, R. and Walberg, H. (2007). "The scientific base linking social and emotional learning to school success." *Journal of Educational and Psychological Consultation*, 17: 191-210.

Annex 1. Validity of the CLASS Pre-K Instrument (based on Pianta et al., 2005)

Criterion validity

The domains of emotional support, classroom organization and instructional support on the CLASS correlate with teacher reports on depression and adult-centered attitudes. Specifically, classrooms with lower scores on these dimensions were associated with higher levels of depression in teachers, while those with lower classroom organization and instructional support scores reported a greater number of teachers holding more adult-centered attitudes.

Concurrent validity

In a comparison of the CLASS with the Early Childhood Environment Rating Scale – Revised Edition (ECERS-R), classrooms with a higher score on the CLASS also rated higher on ECERS-R interactions (with a range of correlations from 0.45 to 0.63). The correlation between the overall score on the CLASS and the provisions for learning factor (furnishings and materials) on the ECERS-R was only moderate, ranging from 0.33 to 0.36 (Pianta et al., 2005).

Not surprisingly, the correlation between CLASS instructional support and the time spent on literacy and mathematics as measured by the Snapshot (an observational instrument that measures the quality and variety of activities that the teacher offers the students in the class, by evaluating the percentage of time spent on various activities; it is used to indicate the extent to which the teacher implements the curriculum [see Pianta et al., 2005]) was low (but significant), since the CLASS measures the quality of activities rather than the quantity.

Predictive validity

The results of a study conducted in several U.S. states offer evidence that the quality of teaching, as measured by the CLASS, is associated with the performance of children at the end of their preschool experience, as well as improved performance throughout preschool (Howes et al., 1992). The CLASS emotional support domain is associated with an increase in children's expressive and receptive language scores, as well as decreases in behavior problems reported by the teacher (Howes et al., in press).

Content validity

CLASS dimensions are based on observations of teacher-student classroom interactions. Dimensions were identified through an extensive review of the constructs validated by other classroom observation instruments, such as those used in the child care literature or research on primary education or effective teaching practices. This literature review was supplemented by pilot testing and consultations with focus groups.

ISBN: 978-1-59782-223-7



9 781597 822237

