IDB WORKING PAPER SERIES Nº IDB-WP-779

Child Care Quality and Child Development

Maria Caridad Araujo
Marta Dormal
Norbert Schady

Inter-American Development Bank
Social Protection and Health Division

February 2017

Child Care Quality and Child Development

Maria Caridad Araujo
Marta Dormal
Norbert Schady

February 2017

Cataloging-in-Publication data provided by the
Inter-American Development Bank
Felipe Herrera Library
Araujo, Maria Caridad
Child Care Quality and Child Development / Maria Araujo, Marta Dormal, Norbert Schady.
p. cm. — (IDB Working Paper Series ; 779)
1.Child care-Evaluation. 2. Child development-Evaluation. 3. Child developmentTesting.
I. Dormal, Marta. II. Schady, Norbert. III. Inter-American Development Bank.
Social Protection and Health Division. IV. Title. V. Series.
IDB-WP-779

http://www.iadb.org

scl-sph@iadb.org

www.iadb.org/socialprotection

Child Care Quality and Child Development

Maria C. Araujo*
Marta Dormal*
Norbert Schady*

Abstract

Development in early childhood predicts schooling and labor market outcomes in adulthood. In this paper we use a fixed effects identification strategy to assess how differences in the quality of child care affect the communication, fine motor, and problem solving skills of infants and toddlers. We show that children have significantly better development outcomes in classrooms with more experienced caregivers, and classrooms with caregivers who demonstrate higher-quality interactions with children. There is substantial heterogeneity in the effects of caregiver quality on child development. Parents either cannot observe, or do not value, the quality of care.

JEL codes: I00, I10, I20, I25, I30, I38, J13.

Key words: child care, quality, child development, CLASS, ASQ-3.

This work would not have been possible without the support of the technical staff of Cuna Mas in Peru, and of the children, families, caregivers and supervisors of all the centers studied. Romina Tomé and María Adelaida Martínez Cabrera provided excellent research assistance at different stages of the project. Sara Schodt had a key role in the training and supervision of the team of CLASS coders. We thank Pedro Carneiro, Jennifer LoCasale-Crouch, Costas Meghir, Robert Pianta, Marta Rubio-Codina, Hiro Yoshikawa and participants at the Latin American Economic Association (LACEA) meeting held in Medellín, Colombia, on November 10-12, 2016 for their comments.

*All authors are with the Inter-American Development Bank. Araujo: mcaraujo@iadb.org; Dormal: mdormal@iadb.org; Schady: norberts@iadb.org.

1. Introduction

A large literature from multiple disciplines indicates that development in early childhood has long-lasting consequences (Almond and Currie 2010; Shonkoff and Phillips 2000). Research from neurology shows that the brain is highly plastic at early ages, and is very sensitive to environmental enrichment (Nelson and Sheridan 2011). In long-term panels, young children with better language, cognitive, motor, and socio-emotional development have better outcomes in adulthood (Case and Paxson 2008; Currie and Thomas 2001; Moffitt et al. 2011). Plausibly exogenous improvements in maternal education (Carneiro et al. 2013), early nutrition (Hoddinott et al. 2008; Maluccio et al. 2009), the quality of the home environment (Eckenrode et al. 2010; Gertler et al. 2014; Olds et al. 1998), preschool attendance (Berlinski et al. 2008; Campbell et al. 2002, 2014; Heckman et al. 2010), and kindergarten classroom quality (Chetty et al. 2011) have been shown to result in higher school achievement, better health status, lower levels of criminal behavior, and better labor market outcomes in developed and developing countries.

Many children, especially among the poor, have deep deficits in development at young ages. In the United States, poor children lag behind their better-off counterparts in test scores by one standard deviation or more at an early age (Carneiro and Heckman 2003; Duncan and Magnuson 2013). In the developing world, more than 200 million children are estimated not to reach their development potential (Grantham-McGregor et al. 2007). More than half of poor children in many Latin American countries are a year or more delayed in terms of their language development by the time they begin formal schooling, and these gaps remain as children progress through the school system (Schady et al. 2015).

Child development occurs as young children interact with their environment. A large proportion of children in the United States and other developed countries attend preschool or child care.[1] Child care coverage has also been increasing in many developing countries.[2] In

[1] There is no sharp distinction between "child care" and "preschool" in the literature. Sometimes, the term child care is used to refer to care for infants and toddlers (children under the age of 36 months of age), while preschool is used for somewhat older children (ages 3 and 4 years). Preschool is usually offered during a 4-hour morning shift. Child care is generally offered for longer hours as it is meant to facilitate parental work. In this paper, in discussing programs from Latin America, we refer as "child care" to services provided by agencies outside the Ministry of Education or by private providers, frequently without a clear curriculum or child development goals, with low-skilled, non-professional staff, and generally (but not always) targeted at younger children. We refer as "preschool" to services that are provided by the Ministry of Education (or regulated by it, when provided by the private sector), generally for older children (3-5 years of age), that have professional educators and a curriculum (frequently designed to facilitate the transition to formal schooling). By this definition, the program we study in this paper, Cuna Mas, is child care rather than preschool.

[2] Publicly financed child care services reach more than 3.1 million children through over 114,000 providers in Latin America and the Caribbean, according to a study of 36 of the largest child care programs in the region (Araujo et al. 2013).

Chile and Brazil, for example, the proportion of children 3 years of age or younger who are in some form of child care doubled in the last decade, and in Ecuador it increased six fold (Berlinski and Schady 2015).

Research from the United States suggests that high-quality child care and preschool can have large benefits for children, especially those in poor households. This is particularly apparent in small-scale pilots of model programs like the Perry Preschool Program in Ypsilanti, Michigan (Heckman et al. 2010; Schweinhart et al. 2005) and the Abecedarian Program in Chapel Hill, North Carolina (Campbell et al. 2002, 2014). There is also evidence of benefits from Head Start, the nationwide program which reaches almost one million low-income children in the United States.[3] However, child care has also been shown to have negative effects on some children (as in Baker et al. 2008, 2015, and Kottelenberg and Lehrer 2016, on the effects of subsidized child care in Quebec; and Havnes and Mogstad 2015 on universal child care in Norway). There are a variety of possible explanations for these seemingly contradictory findings, but one is that there are large differences in both the quality of care and the quality of home environments that young children are exposed to.[4] There are also challenges implementing high-quality programs at-scale (as in the Quebecoise and Norwegian experiences).

Much less is known about the effects of child care on child outcomes in developing countries, in particular for infants and toddlers.[5] Behrman et al. (2004) and Bernal and Fernandez (2013) evaluate the effects of community-based care on child development in Bolivia and Colombia, respectively. Their results suggest a positive effect of child care among somewhat older children (roughly 4 years of age or older), with no effects (or negative effects) among younger children. Rosero and Oosterbeek (2011) find that child care attendance has negative effects on cognitive and language development of children between the ages of 3 and 5 years in Ecuador; mothers of children who attend child care are also less likely to provide responsive parenting.[6]

[3] The literature on the effects of Head Start is large. Important references include Carneiro and Ginja (2014), Currie and Thomas (1995), Deming (2009), Garces et al. (2002), Kline and Walters (2016), Ludwig and Miller (2007), and Puma et al. (2010; 2012). Currie (2001), Duncan and Magnuson (2013), and Ludwig and Phillips (2007) discuss the evidence.

[4] Note that what matters is the quality of child care *relative* to the counterfactual home environment that children would have been in if they had not been in child care. This counterfactual environment is generally better for children from high-socioeconomic status households than for those in low-socioeconomic status households, which may explain why positive impacts of child care are more often found among poorer children.

[5] Infants are children younger than 12 months, and toddlers are children between the ages of 12 and 36 months.

[6] There is also some work on the effects of formal preschool attendance among older children in Latin America. Berlinski et al. (2009) study the effects of a large preschool construction program in Argentina. The program varied in timing and intensity across provinces, which the authors use for identification. They find that children who attended preschool at age 4 years had higher test scores and fewer behavioral problems in third grade. In a separate paper using data from Uruguay, Berlinski et al. (2008) show that children who attended preschool were 27 percent more likely to still be in school at age 15 than other children who did not attend preschool.

In this paper, we study how the quality of child care affects the development of children 6-24 months of age using data from Peru, a middle-income country. The first two years of life are a period in which the brain exhibits great plasticity and sensitivity to environmental influences (Fox et al. 2010; Grantham-McGregor et al. 2007; Shonkoff and Phillips 2000). We are aware of only two earlier papers that explicitly look at the relationship between quality of care and child development in this critical age range in a developing country. Bernal et al. (2015) carefully evaluate a program that moved children from community care to care in large centers (generally covering 150 or more children each) in Colombia. They show that the move had substantial costs, both in terms of the infrastructure that was constructed, and because the centers included additional professional staff, who were paid higher salaries. However, quality did not improve, and neither did measures of child development. Also in Colombia, Bernal (2015) evaluates the effect of an in-service training program for the community mothers who work as caregivers in child care. She finds that the program increased the quality of care, and improved the cognitive and socio-emotional development of children under the age of 3 years.

As we discuss below, the quality of child care can be measured in a variety of ways. In our analysis, we focus on the frequency and quality of interactions that young children have with their caregivers, as measured by a classroom observation tool known as the Toddler Classroom Assessment Scoring System (CLASS hereafter; La Paro et al. 2012; Pianta et al. 2007). The CLASS correlates strongly with child development in the United States (Bandel et al. 2014; Mashburn et al. 2008; Pianta et al. 2016). In addition, we test whether children who have caregivers with more experience or more years of completed schooling have better outcomes.

To assess the effects of child care quality on child development, we use data from a sample of 291 child care centers. Our identification strategy is based on within-center, cross-classroom comparisons of quality and child outcomes. We show that a child assigned to a caregiver who has one standard deviation higher CLASS scores has 0.07 standard deviation better development outcomes. Moreover, we argue that the association between CLASS scores and child development we report is likely to be downward-biased by measurement error in the CLASS, probably by a factor of two or more.

We find no evidence that caregivers with higher education levels are more effective in our sample. However, an additional year of caregiver experience is associated with 0.03 standard deviations higher child development outcomes. We also note that experience and the CLASS score are uncorrelated in our data, which indicates that more experienced

caregivers must carry out activities that foster child development that are not picked up by our measure of caregiver-child interactions.

Our analysis carefully considers possible sources of heterogeneity in the relationship between caregiver quality and child development. To this effect, in addition to the results that focus on our aggregate measure of child development, we separately estimate effects on language, fine motor skills, and problem solving; we break down the CLASS score into its two domains—emotional and behavioral support, and engaged support for learning; we test for non-linearities in the effect of caregiver experience on child development; and we analyze the effects of quality on the distribution (not just the mean) of child development. The most important result of this analysis is that the effect of having a caregiver with a higher CLASS score is concentrated at the bottom of the distribution of child development, while the effect of having a caregiver with more experience is particularly beneficial for children at the top.

The center fixed effects strategy we use is appealing because it sweeps out all time-invariant characteristics of centers and the population that they serve (for example, differences in socioeconomic status across neighborhoods). Nevertheless, unobserved differences in children or caregivers across classrooms *within* a center are a potential threat to identification.

We provide three pieces of evidence that suggest that our identification strategy is reasonable, in particular for the CLASS. First, our estimates are very insensitive to the addition of a large number of controls. Second, we make use of the fact that our data include an assessment of caregiver quality by parents (on a 1-4, Likert-like scale), and center supervisors (who place caregivers into one of three performance categories, which, in turn, determine pay). We show that neither parents nor supervisors give higher ratings to caregivers with better CLASS scores, suggesting that they are not aware of, or do not value, the quality of caregiver-child interactions as measured by the CLASS. We take this as evidence that the CLASS effects we estimate are very unlikely to be biased by purposeful sorting of children to caregivers. Finally, we show that the coefficients in our regressions of child development on the CLASS are very close to others reported in the literature. This suggests that the magnitude of the effects we estimate is plausible.

The rest of the paper proceeds as follows. In Section 2 we briefly review the literature on child care quality. Section 3 describes our data and the Peruvian setting. We discuss our identification strategy in Section 4, and present results in Section 5. Section 6 concludes.

2. Measuring quality in child care services

Child care quality is a multi-faceted concept. Frequently, a distinction is made between elements of structure and process. Structural quality refers to resources that can facilitate the interactions that are necessary in a learning environment. It includes measures of the quality of the infrastructure, the availability of learning materials and a curriculum, the qualifications and experience of caregivers, and child-to-caregiver ratios. Structural quality is relatively easy to regulate and to measure.

Process quality, on the other hand, focuses on more subtle dimensions of quality, such as the way in which the curriculum is implemented and the frequency, type, and nature of interactions that occur between children, between children and caregivers, and between parents and caregivers. Measuring process variables is more complicated and lengthy, because it requires the reliable observation and coding of these interactions. However, process quality has been shown to predict child outcomes in a variety of contexts.[7]

Process and structural aspects of quality are often related. When child-to-caregiver ratios are high, positive interactions are less frequent; when few or no materials or basic infrastructure are present, health and nutrition routines tend to be of low quality, and activities are fewer in number and poorer in quality. Better-educated caregivers and those specifically trained in early childhood education may provide more developmentally appropriate and stimulating activities (NICHD 2000a, 2000b; Vandell and Wolfe 2000). Structural variables such as staff wages have also been shown to predict other aspects of child care quality (Whitebook et al. 2001).

Research from the United States has documented associations between the quality of care, and child development and learning. Ruzek et al. (2014) use propensity score matching to show that higher-quality care for toddlers is associated with higher levels of cognitive development at 24 months of age. Peisner-Feinberg et al. (2001) show that two aspects of process quality in preschool, classroom practices and the closeness of teacher-child relationships, predict skills through elementary school.

Mashburn et al. (2008) compare data on preschool quality from 11 US States, and explore the association between program quality and language, academic, and social development. The authors use three measures of quality: features of program design and infrastructure following standards from the National Institute for Early Education Research indicators from The State of Preschool, observations of overall classroom quality measured

[7] Important references include La Paro et al. (2004); Pianta (2003); Pianta et al. (2016); Thomason and La Paro (2009); Vandell and Wolfe (2000).

by the Early Childhood Environment Rating Scale (ECERS; Harms and Clifford 1980; Harms, Clifford, and Cryer 1998), and observations of interactions measured by the CLASS. They conclude that the quality of teacher-child interaction is most consistently associated with child developmental outcomes.

Another issue that has received attention is whether the effects of better quality child care on child development are sustained over time. Vandell et al. (2010) report that attending high-quality child care during the early years is associated with higher cognitive academic achievement at age 15, and fewer externalizing behaviors.[8] Li et al. (2013) show that a one standard deviation increase in the quality of care received by infants and toddlers (measured by the Observational Record of the Caregiving Environment, ORCE; NICHD 1996) is associated with a short-run increase of 0.15 standard deviations in cognitive development. However, unless children also attend high-quality preschool at ages 4-5 years, the benefits of higher-quality care at earlier ages fade out quickly; see also Garces et al. (2002) on how Head Start may interact with the quality of schooling that children attend later on.

In sum, the literature from the US suggests that children who attend child care or preschool of higher quality, in particular better process quality, have better outcomes. In practice, however, quality is very variable, and few children are consistently exposed to high-quality early education (Pianta et al. 2016). Much less is known about the quality of the child care children receive in developing countries, including in Latin America. What little evidence is available suggests that quality is frequently low, in particular in those dimensions of quality that are most directly linked with child outcomes (Araujo et al. 2015; Berlinski and Schady 2015).

3. Setting and data

A. Setting

We study the effects of child care quality for children between 6 and 24 months of age in Peru, a middle-income country in South America. In the last decade, Peru has exhibited high rates of economic growth—5.9 percent per year, the highest rate in Latin America—and

[8] A distinction in the literature on child psychology and psychiatry is often made between "internalizing" and "externalizing" problem behaviors (Achenbach 1978; Liu 2006). Children with internalizing problems are withdrawn, anxious, or depressed, while those with externalizing problems tend to be hyperactive, disruptive, or aggressive. Externalizing behaviors in early childhood have been shown to predict juvenile delinquency, adult crime, and violence.

substantial reductions in poverty—the proportion of the population living below the World Bank poverty line of US $3.10 per capita per day fell from 27 to 9 percent (World Bank 2016).

Peru has also seen substantial improvements in various indicators of child wellbeing. Between 2000 and 2015, chronic malnutrition (stunting, or low height-for-age) of children under the age of 5 years fell from 31 percent to 14 percent, and infant mortality fell from 30 per 1,000 to 13 per 1,000 live births. There have also been increases in school enrollment, although learning outcomes continue to be poor.[9]

Our paper studies the *Programa Nacional Cuna Mas* (Cuna Mas, hereafter). In urban areas, Cuna Mas offers child care services, geographically targeted to districts with a high concentration of poverty, and available for children 6-36 months of age. Child care is provided in community centers.[10] Children are attended by caregivers (not professional educators) hired from within the community. Each classroom has one caregiver. Caregivers do not hold a formal employment relation with Cuna Mas; they receive a monthly stipend from the program for their services, which is often supplemented by copayments from parents. Caregivers are meant to have completed secondary school, although this is not always the case in practice. There is a professional educator who acts as a supervisor in charge of 12-15 caregivers, and of the centers where they work. In our sample, each professional educator supervises 8 centers, on average.

B. Data

Using administrative data, we took a random sample of 301 community child care centers which had at least two classrooms. These centers cover much of the country—they are drawn from 301 localities (*centros poblados*) in 137 different municipalities (*distritos*) and 22 Peruvian departments.[11]

[9] Peru had the lowest scores of any participating country on the international PISA test of math achievement of 15-year olds in 2014 and, although it experienced the largest improvement in the region in the 2015 PISA test, it still scored 66th out of 72 participant countries (OECD 2014, 2016). On a math test applied in 15 Latin American countries in 2013, 40 percent of third graders in Peru had scores that placed them in the lowest performance category—better than the regional average (47 percent), but substantially worse than high-performing countries like Chile (15 percent) and Costa Rica (23 percent) (UNESCO 2015).

[10] In addition to community centers, Cuna Mas operates two other child care modalities. In one, the service is provided in a family home. This modality is in the process of being discontinued. In another, child care is provided in infrastructure constructed for this purpose (*Centros Infantiles de Atención Integral,* CIAIs, or Comprehensive Early Childhood Centers). Our focus in this study is only on child care provided in community centers. In September 2015, community centers served 60 percent of children in Cuna Mas child care centers, compared to 33 percent in family homes, and 7 percent in CIAIs. These numbers are based on administrative data (direct communication with Cuna Mas in December 2015).

[11] According to administrative data, there were 584 Cuna Mas community centers that had two or more classrooms at the time the sample was drawn. The sample frame excluded one region of the country (known as VRAE) where Cuna Mas staff determined that it would not be possible to conduct the study because of accessibility and safety concerns.

Data were collected between November 2013 and January 2014. Enumerators visited the centers and drew up a roster of children in each classroom. In practice, in 10 of the 301 centers in the original sample there was only one classroom, or only one classroom with children between 6 and 24 months of age (the age range we consider in the study). These centers were dropped from the sample. Just over half (159) of the remaining centers had exactly two classrooms. When there were three or more classrooms, two were chosen at random. Our final sample therefore includes 582 classrooms in 291 centers.

Data on center, classroom, and caregiver characteristics were collected at the center, and data on households and children were collected in children's homes. There were a total of 4,058 children in the 582 classrooms in the sample. We attempted to contact all of the 2,324 children who were within the age range (6-24 months) of the study in these classrooms. We successfully completed the child development assessment for 2,198 children in 2,173 households (94.6 percent of those we attempted to contact).[12]

Table 1 presents descriptive statistics of the children in the sample and their families; the unit of observation is always the child. Cuna Mas locates its child care services mostly in urban areas, and 88 percent of households in the sample live in an urban area. Just under half of the children in the sample are girls.

The socioeconomic status of households in our sample appears to be similar to other households in urban areas in Peru. Average maternal education is 10 years (which is also the average for adult women in urban Peru, according to the 2014 *Encuesta Nacional de Hogares*, ENAHO, a multi-purpose household survey regularly carried out by the Peruvian National Institute of Statistics, INEI). Seventy-eight percent of households have piped water and 64 percent have access to the sewerage system (compared to 85 percent and 79 percent, respectively, for urban Peru). Sixteen percent of mothers in the survey report being of indigenous or of Afro-Peruvian descent (compared to 25 percent in urban areas in the nationwide survey). Seventy-eight percent of the children in the sample live with both parents.

Proximity and child age seem to be important determinants of the demand for child care services in Peru. Families report living, on average, 10.59 minutes walking distance from the center. Older children are more likely to attend child care. At the time of the developmental assessment, 21 percent of children in the study were 6-12 months old, 36 percent were 13-18 months old, and the remaining 44 percent were 19-24 months old.

[12] The remaining children had either incomplete or inconsistent information (24 cases), were absent when visited by the enumerator (101 cases), or refused to participate (1 case).

The main outcome we use in the paper is the Ages and Stages Questionnaire screener (third version, henceforth ASQ; Squires et al. 2009). We applied three of the five scales in the ASQ: communication, problem solving and fine motor. We did not apply the gross motor and personal-social scales, primarily because of time and resource constraints. The ASQ includes a mix of questions that are answered by a child's mother and others that are recorded by an enumerator through direct observation. The test has been applied in many developing countries, including in Mexico (Angeles et al. 2011), Colombia (Bernal 2015), Mozambique (Martinez et al. 2012), and in a study covering four countries, including Peru (Fernald et al. 2012). We used a team of Peruvian psychologists to pilot the ASQ, and to make changes as needed. Further details on the test and its application are given in Appendix A.

To measure the quality of caregiver-child interactions, we used the Toddler Classroom Assessment Scoring System (CLASS hereafter; La Paro et al. 2012 ; Pianta et al. 2007). The CLASS is an observational instrument for use with children 15-36 months of age.[13] It assesses the quality of caregiver-child interactions in two domains: emotional and behavioral support, and engaged support for learning. The first of these domains has five dimensions: positive climate, negative climate, teacher sensitivity, regard for child perspectives, and behavior guidance. Negative climate is reverse-coded, so that higher scores indicate less negative climate in a classroom. The second domain, in turn, has three dimensions: facilitation of learning and development, quality of feedback, and language modeling. CLASS is scored on a scale from 1-7. Scores between 1 and 2 are considered low quality, scores between 3 and 5 are considered medium quality, and scores between 6 and 7 are thought of as high quality.

To calculate the CLASS, the two classrooms in each center were filmed for four hours each, always at the same time of day to ensure comparability. The resulting video was cut into 20-minute segments. For each classroom, four segments were selected and coded twice, by two different coders assigned at random.[14] Classroom scores are averaged across the two coders and the four segments. Detailed information on the reliability of CLASS in our sample and other properties of the instrument are discussed in Appendix B.

[13] At the time this study was planned, there was no version of the CLASS to measure the quality of child care for children younger than 15 months of age. We corresponded with the creators of the CLASS, who recommended we use the Toddler CLASS even though about one-third of the children in the sample were younger than 15 months (Authors' correspondence with Jennifer LoCasale-Crouch and Robert Pianta, July 2013).

[14] On rare occasions when the scores between the two coders differed by more than a pre-established amount, the segment in question was coded by a third coder.

Table 2 summarizes the mean characteristics of centers and caregivers in our sample. Centers in the sample had an average of 2.64 classrooms and 18 children. The average class size in the classrooms in the sample is 8.68. For our analysis, we focus only on those younger than 24 months of age. Our sample of 291 centers includes 184 center coordinators, who split their time between multiple centers. They have completed 16 years of schooling, on average, and have 1 year of experience working at the center.[15]

Caregivers have lower levels of education than coordinators (a mean of 10 years of schooling). On average, they have 2 years of experience, although there is considerable variation: almost half (46 percent) are in their first year working at the center. Nineteen percent of caregivers are indigenous or Afro-Peruvian. Interestingly, both in terms of ethnicity and education levels, caregivers have a similar profile to mothers. This is not surprising as it is local women from the community who work in Cuna Mas centers. CLASS scores in the emotional and behavioral support domain are generally in the medium quality range, 3.95 on average. Scores in the engaged support for learning domain are low, with an average score of 1.81.

Our estimates focus on the within-center, across-classroom difference in CLASS scores, caregiver experience, and caregiver education. On average, this within-center difference is 0.85 standard deviations for CLASS scores, 2 years for caregiver experience, and 2.6 years for caregiver education. Simple decompositions show that 31 percent of the variation in the CLASS, 36 percent of the variation in experience, and 34 percent of the variation in caregiver education occurs within centers. This is the variability we exploit for identification.

4. Identification strategy

Our main results are based on OLS regressions of the following form:

$$(1)\quad Y_{ihkc} = \alpha_c + \beta \mathbf{X}_{ihkc} + \theta Q_{kc} + \varepsilon_{ihkc},$$

where Y_{ihkc} is the development of child i in household h, classroom k of center c, as measured by the ASQ; α_c is a set of center fixed effects; $\mathbf{X}_{ihkc}$ is a set of controls for children

[15] Years of experience working at a center are low because, at the time the data were collected, Cuna Mas had recently hired center coordinators with higher educational credentials in all of its centers.

and households, described in greater detail below; Q_{kc} is a set of classroom characteristics, including the CLASS, experience, and education of caregivers; and ε_{ihkc} is the error term.

The parameter of interest is θ, which measures the extent to which child development is higher in classrooms in which caregivers have higher CLASS scores, more education, or more years of experience. Both the ASQ and the CLASS scores have been standardized so they have mean zero and unit standard deviation. In the case of the ASQ, we remove age effects before converting the scores to z-scores using the non-parametric methods proposed in Rubio-Codina et al. (2016). Education and experience are in years. Standard errors are clustered at the center level.

In our first set of results, we focus on a child's total ASQ score, which is given by the simple average of her scores on the communication, fine motor, and problem-solving scales. We then consider whether there is heterogeneity in the relationship between caregiver quality and child outcomes. For this purpose, we first estimate equation (1), separately for the different ASQ scales and by CLASS domain (emotional and behavioral support, and engaged support for learning). To take account of multiple hypotheses testing, we report whether the coefficients in these regressions are significant when we use the step-down procedure in Romano and Wolf (2005).[16]

Another interesting dimension of heterogeneity is possible non-linearity in the returns to experience. The literature on teachers in the United States shows that teachers get substantially better in their first year on the job, but returns to experience flatten out thereafter (see Jacob 2007; Staiger and Rockoff 2010, and the references therein). To test whether there is such a pattern in our data, we consider alternative definitions of "experienced" caregivers—those with 1 or more years of experience, 2 or more years, and so on, up to 7 or more years of experience. We then run regressions of child development on these alternative definitions of experienced caregivers. If the returns to experience plateau after a certain point, we would expect the coefficients for "experienced" caregivers to be similar for alternative cutoffs of experience beyond this point.[17]

Finally, we analyze whether the effects of caregiver quality (as measured by the CLASS or experience) are especially large at a particular point in the distribution. This is important as a number of papers have found that access to child care can benefit some

[16] In the regressions in which the explanatory variable is experience or the CLASS (first two rows of Table 4) we correct standard errors for the testing of 3 hypotheses. In the regressions in which the explanatory variable is one of the two domains of the CLASS (third and fourth row of Table 4) standard errors are corrected for the testing of 6 hypotheses.

[17] For example, if caregivers with 3 years of experience are no more effective than those who have only 2 years on the job, we would expect that the coefficient in a regression of child development on a dummy for experienced caregivers is similar when experienced caregivers are defined as having 2 or more years of experience or 3 or more years of experience.

children more than others, or benefit some children and harm others. For example, Havnes and Mogstad (2015) argue that universal access to child care in Norway benefited children from poor households, but had negative effects on children from better-off families. Kottelenberg and Lehrer (2016) show that the provision of child care services in Quebec had positive effects on child outcomes for the poorest children, but had negative impacts for children between the 10^{th} and 50^{th} percentiles of the distribution.

To analyze the distributional effects of differences in caregiver quality, we proceed in two ways. First, within each center, we identify the caregiver with the higher (or lower) CLASS, as well as with more (or less) experience. We then calculate ASQ scores of children at the 5^{th}, 10^{th} … 95^{th} percentiles of the distribution in each of these four groups, and take the difference in scores at each ventile between children with caregivers with high and low CLASS, and between those with high and low experience. In our second approach, we generate dummy variables for children below the 10^{th}, 25^{th} and 50^{th} percentiles of the distribution of ASQ scores, and above the 50^{th}, 75^{th} and 90^{th} percentiles, and run regressions of each of these variables, in turn, on a caregiver's CLASS score or experience.

As we discuss above, our identification strategy assumes that there is no purposeful sorting of children with unobservable characteristics that are correlated with child development to high- or low-quality caregivers. We cannot test this assumption directly. It is very unlikely, however, that children would be matched to caregivers of different quality if parents and supervisors do not observe, or do not value, quality. We therefore analyze how parents and supervisors rate caregivers.

For the analysis of parents, we make use of a question in the household survey in which mothers were asked to rate their child's caregiver as "very good", "good", "bad", or "very bad". In practice, 31 percent of mothers gave caregivers a rating of "very good", and another 68 percent judged them to be "good". Because less than 2 percent of mothers rated caregivers "bad" or "very bad", in our analysis we simply generate a dummy variable that takes on the value of one if a mother rated the caregiver "very good", zero otherwise. We then regress this variable on the CLASS, years of experience, education, and the mean ASQ score in a classroom, including center fixed effects.

For the analysis of supervisors, we make use of data that indicates whether a supervisor assigned a given caregiver to one of three performance categories, A, B, or C. Caregivers in performance category A were paid 300 Peruvian soles per month, while those

in categories B and C were paid 330 and 360 Soles, respectively.[18] In practice, 45 percent of caregivers were assigned to category A, 7 percent were assigned to category B, and 48 percent to category C. Given this distribution, we generate a dummy variable that takes on the value of one if a caregiver was assigned to performance category C, zero otherwise. As with the analysis of parents, we then regress this variable on the CLASS, years of experience, education, and the mean ASQ score in a classroom, including supervisor fixed effects.

Throughout, we report the results from two specifications. In one specification we control only for fixed effects (center fixed effects in regressions in which the unit of observation is the child or her mother, supervisor fixed effects in those that analyze the performance category assigned to caregivers). In the second specification, we add controls for classroom composition (the number of children, the proportion of female, and the mean age). When the unit of observation is the child or her mother, we also add controls for child gender, household demographics (the education, age, and ethnicity of the mother, whether both parents live at home, and the number of household members), the number of assets in the household (assets include refrigerator, gas stove, washing machine, iron, blender, TV, DVD, computer, stereo, cell phone, and cable), the distance from the household to the center (in minutes), and variables that measure the quality of housing (whether the household has piped water inside the home, and separately, is connected to the sewerage system, the number of bedrooms, and whether the house has dirt floors).[19] A comparison of the coefficients in the regressions with and without controls is a general robustness test. In addition, if observed and unobserved determinants of child development are correlated, it is an indication of the extent to which our results may be biased by unobservables (Altonji et al. 2005).

5. Results

A. Main results

Table 3 reports the results from estimating regression (1) for alternative measures of caregiver quality: the CLASS, caregiver experience, and caregiver education. The dependent

[18] In December 2013, 1 US$ was equivalent to 2.8 Peruvian Soles.
[19] Missing values were imputed and replaced by the sample median (for continuous variables) or mode (for binary variables) in 79 cases where one or more of the household-level variables with missing data. In these cases in the regressions we also include a dummy variable equal to one when these data were replaced.

variable is a child's total ASQ score. The first set of regressions (columns 1 and 2) of the table focuses on the CLASS. It shows that children in classrooms with caregivers who have one standard deviation higher CLASS scores have 0.07 standard deviations higher ASQ scores. The difference in the parameter estimates between the specification with and without controls is only 0.001 standard deviations. We take this as strong evidence that children of different characteristics are not sorted to caregivers with better or worse CLASS scores.

Is the magnitude of the effects we estimate plausible? To answer this question, we compare our estimates with others reported in the literature. NICHD Early Child Care Research Network and Duncan (2003) report that in the United States, a one standard deviation increase in the quality of child care received by children 6-24 months of age (with quality measured by the ORCE) is associated with increases in cognitive development of 0.08 standard deviations at 54 months of age.

Araujo et al. (2016) analyze how the quality of teacher-child interactions in kindergarten classrooms (children 5-6 years of age), as measured by the CLASS, predicts the language, math, and executive function scores of children at the end of kindergarten in Ecuador, a country that neighbors Peru. The identification strategy is based on the random assignment of children to classrooms within schools, with a compliance rate of 98.5 percent. Based on this unusually clean identification, Araujo et al. (2016) find that a one standard deviation increase in the CLASS raises test scores by between 0.08 standard deviations (when using the once-lagged CLASS for the teachers in their sample) and 0.06 standard deviations (when using the contemporaneous CLASS).[20]

In sum, other papers that have looked at the association between quality of care for similarly-aged children in a very different context (the United States), or older children (5-6 years, rather than 6-24 months) in a context that is similar to ours, using the same instrument to measure quality (the CLASS), report effect sizes that are very close to those we estimate in this paper.

The results in Araujo et al. (2016) are also useful because they include a careful discussion of how measurement error in the CLASS could affect the estimated coefficients. The authors show that coder error in the CLASS (differences between two coders coding the same video segment) is small. However, there is substantial error in the CLASS associated with the fact that teachers are only observed for a single day. Because they have both the lagged and the contemporaneous CLASS scores for the same teachers, Araujo et al. (2016)

[20] The CLASS in our study was coded by the same group of coders, and supervised by the same master coder, as in Araujo et al. (2016).

can purge their estimates of measurement error by regressing child test scores on the contemporaneous CLASS instrumented with the lagged CLASS. The coefficient on the CLASS in these regressions is 0.18—2.5 to 3 times their simple OLS estimate. If measurement error introduces attenuation bias of the same magnitude to the estimates we report, this would imply that the true effect of the behaviors measured by the CLASS on child development in our sample of child care centers in Peru would be on the order of 0.17-0.21 standard deviations.

We next turn to caregiver experience. The second set of regressions (columns 3 and 4) of Table 3 shows that, on average, an additional year of caregiver experience is associated with 0.03 standard deviations higher child development, with or without the additional controls. Columns 5 and 6 of Table 3 show that caregiver education is not associated with child development in our sample. The final columns (7 and 8) include all three caregiver characteristics in the regression. The coefficients on each of the caregiver characteristics are essentially unchanged. This indicates that more experienced caregivers improve child development in ways that are not captured by caregiver-child interactions, as measured by the CLASS. [21] While our data does not allow us to identify what these factors might be, a possibility is that more experienced caregivers are able to offer better-quality and more age-appropriate contents to the children in their classrooms. CLASS is focused on interactions and does not assess contents aspects of quality.

In sum, Table 3 shows that caregiver experience and the CLASS (but not caregiver education) are robustly associated with within-center differences in child outcomes, no matter whether we include a long list of controls in the regression or not, and no matter whether we include each of the caregiver attributes on their own or together.

B. Heterogeneity

To begin our analysis of heterogeneity, we first explore the associations between quality and the different ASQ scales (communication, fine motor, and problem solving skills). This analysis also breaks down the CLASS into scores on the two domains (emotional and behavioral support, and engaged support for learning).

Table 4 reports the coefficients from separate regressions of a given outcome variable (for example, scores on the ASQ communication scale) on one measure of caregiver

[21] In a regression of the CLASS on caregiver years of experience, including center fixed effects, the coefficient is 0.004, with a standard error of 0.021.

quality (for example, her score on the CLASS engaged support for learning domain). The table shows that caregiver experience and the CLASS are associated with improvements in different domains of child development. In the case of experience, the biggest effects are found for the communication and fine motor scales of the ASQ. These results are significant at the 10 percent level or higher even after we control for multiple hypotheses testing. In the case of the CLASS, the biggest effects are found on the ASQ problem solving scale. Once again, these results are significant at the 10 percent level or higher even after we control for multiple hypotheses testing. When the CLASS is broken down by domain, the effects are broadly similar for the emotional and behavioral support and the engaged support for learning domains.

Next, we analyze whether the returns to experience in our sample of caregivers are linear. In Figure 1, we plot the coefficients from separate regressions of child development on a dummy variable for experienced caregivers, when the cutoff between experienced and inexperienced caregivers is given by different values of experience. The figure shows that the coefficients increase monotonically from left to right as the cutoff for experienced caregivers increases from 1 to 6 years, and appear to plateau thereafter. For example, when an experienced caregiver is defined as having at least 1 year of experience, the coefficient is a statistically insignificant 0.08 standard deviations; when the cutoff is set at 3 years of experience, it is an insignificant 0.13 standard deviations; and when the cutoffs are set at 5 and 7 years, the coefficients are 0.21, and 0.24 standard deviations, respectively. This shows that the returns to experience among caregivers in Peru rise over a longer horizon than is the case for school teachers in the United States.

Finally, we analyze whether the effects of caregiver quality (as measured by the CLASS or experience) are especially large at a particular point in the distribution. Figure 2 plots the difference in ASQ scores between children with caregivers who have high and low CLASS, and between those with high and low experience, at every ventile of the distribution.[22] The line that corresponds to the CLASS clearly shows that the largest effects of having a high-CLASS caregiver are found at the bottom of the distribution. For example, in the high-CLASS sample, children at the 10th percentile of the distribution have an ASQ score that is 0.18 standard deviations higher than children at the 10th percentile of the distribution in the low-CLASS sample. Differences at higher ventiles are smaller. On the other hand, the line that corresponds to experience shows that the largest effects of having a more

[22] For example, the first (leftmost) point in the line corresponding to the CLASS is given by the difference between the ASQ scores of children at the 5th percentile of the distributions for high- and low-CLASS caregivers.

experienced caregiver are concentrated at the top of the distribution. In the high-experience sample, children at the 90th percentile of the distribution have an ASQ score that is 0.15 standard deviations higher than children at the 90th percentile of the distribution in the low-experience sample. Differences at lower ventiles are smaller.

We complement the figure with the results of regressions of dummy variables for children below the 10th, 25th and 50th percentiles of the distribution of ASQ scores, and above the 50th, 75th and 90th percentiles, on a caregiver's CLASS score or experience. As is the case in Figure 2, the results in Table 5 show different patterns. Having a more experienced caregiver appears to benefit most children, but has the largest effects at the top of the distribution: an additional year of caregiver experience increases the proportion of children above the 90th percentile of the distribution of ASQ scores by 1 percentage point. On the other hand, the effect of better caregiver-child interactions is concentrated at the bottom of the distribution: caregivers with 1 standard deviation higher CLASS scores reduce the proportion of children below the 10th percentile of the distribution by 3.1 percentage points.

In sum, Figure 2 and Table 5 suggest that the mean impacts in Table 3 miss important distributional effects, consistent with the results in Havnes and Mogstad (2015) for Norway, and Kottelenberg and Lehrer (2016) for Quebec. In our case, however, these distributional effects are different for our two measures of caregiver quality—the CLASS, and experience.

C. Do mothers and center supervisors observe and value quality?

Panel I of Table 6 reports the results of regressions of how mothers rate caregivers. There is weak evidence that mothers value caregiver experience: an additional year of caregiver experience increases the probability that a caregiver is judged to be "very good" by about 1 percentage point. However, the magnitude of the coefficient is very small, and is only marginally significant. There is no evidence that mothers give higher scores to caregivers with higher CLASS scores, those who have more education, or those in whose classrooms average ASQ scores are higher.

Panel II focuses on the performance category that caregivers are assigned to by their supervisors. These results very clearly indicate that center supervisors reward seniority. An additional year of experience increases the probability that a caregiver is assigned to performance category C by 8 percentage points. As with parents, there is no evidence that

supervisors reward caregivers with higher education, better CLASS scores, or higher child development.

In sum, with the exception of a small, marginally significant effect for experience, we find little evidence that mothers give higher ratings to better caregivers. This is of substantive interest: if parents are unable to observe quality, or do not value quality, they are unlikely to demand higher-quality child care services.[23] Meanwhile, the results for supervisors strongly suggest that supervisors simply reward seniority, regardless of the interactions that caregivers have with children or how effective they are in producing more child development. This, too, is of substantive interest because it suggests that in Peru the performance evaluation conducted by supervisors is largely a mechanical exercise.

6. Conclusion

An increasing number of children in both developed and developing countries are in child care. This occurs at an age when the brain is very plastic and affected by environmental influences. The quality of child care is likely to be an important determinant of the benefits, if any, of child care attendance. However, very little is known to date about child care quality in developing countries.

In this paper, we analyze how differences in quality across caregivers within a center affect the development of infants and toddlers in Peru, a middle-income country. We show that children assigned to more experienced caregivers, and caregivers that display better interactions with children, have higher development outcomes. Better caregiver-child interactions are particularly beneficial for children at the bottom of the distribution of ASQ scores. More experienced caregivers have substantial positive effects at the top of this distribution.

Our findings add to a very small literature on the effects of quality in child care in developing countries. Moreover, they have clear policy implications. In both developed and developing countries, staff working at child care centers generally have lower qualifications, receive less pay, and have higher levels of turnover than is the case with teachers of somewhat older children. In the United States, for example, the annual turnover of child care workers is between 25 and 40 percent (Porter 2012 and the references therein). In Latin

[23] The difficulty that parents have observing childcare quality is a point made elsewhere in the literature, including by Blau (2001), and Blau and Currie (2006).

America, early childhood educators have lower remunerations than, and different evaluations and career paths from, primary school teachers (Kagan et al. 2015).

High levels of turnover prevent caregivers from acquiring experience, and keep children from forming stable, secure attachments (Love et al. 2003; Raikes 1993; Scarr et al. 1994). A long literature in child development and psychology has shown that the capacity of very young children to form secure attachments is critical for their development (in particular, socio-emotional development), and that children who do not form secure attachments with adults have worse outcomes in school, and higher levels of clinical depression and criminal activity in adulthood (Ainsworth and Bell 1970; Bowlby 1969; Rutter and the English and Romanian Adoptees (ERA) Study Team 1998; Shonkoff and Phillips 2000). Turnover of caregivers in our sample is very high—46 percent are in their first year on the job. In developing countries like the one we analyze in this paper, there are likely to be substantial benefits to policies that seek to reduce turnover among caregivers in child care. These policies could include professionalization of the workforce, and competitive compensation.

Our paper also shows that the quality of the interactions between caregivers and children in our sample is very low in the engaged support for learning domain of the CLASS. This is the domain (and the equivalent instructional support domain of the CLASS for somewhat older children) that is most strongly associated with cognitive development in early childhood, and with subsequent performance on tests among school-aged children (Burchinal et al. 2008, 2010; Hamre and Pianta 2005; Mashburn et al. 2008). Some pilots of in-service programs for teachers of young children that have focused on teacher-child interactions have shown promise in the United States (Bierman et al. 2008; Downer et al. 2013; Hamre et al. 2010). In-service training for low-skilled community mothers engaged in child care has also been found to be effective in Colombia (Bernal 2015). Further experimentation and careful evaluation of innovative forms of pre-service or in-service training for caregivers is likely to be important in settings like the one we study in this paper.

More generally, our results suggest that in developing countries in which the coverage of child care is high or growing, more attention should be given to programs that seek to retain effective caregivers, and increase their capacity to engage in frequent, high-quality interactions with the children they care for.

References

Achenbach, Thomas M.. 1978. “The Child Behavior Profile: I. Boys aged 6-11.” *Journal of Consulting and Clinical Psychology* 46(3): 478-88.

Ainsworth, Mary D. Salter, and Silvia M. Bell. 1970. “Attachment, Exploration, and Separation: *Illustrated* by the Behavior of One-Year Olds in a Strange Situation.” *Child Development* 41(1): 49-67.

Almond, Douglas, and Janet Currie. 2011. “Human Capital Development before Age Five.” In *Orley* Ashenfelter and David Card, eds., *Handbook of Labor Economics* (pp. 1315-486). North Holland: Amsterdam.

Altonji, Joseph, Todd E. Elder, and Christoper R. Taber. 2005. “Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools.” *Journal of Political Economy* 113(1): 151-84.

Angeles, Gustavo, Paola Gadsen, Sebastian Galiani, Paul Gertler, Andrea Herrera, Patricia Kariger, and Enrique Seira. 2011. “Evaluación de Impacto del Programa Estancia Infantiles para Apoyar a Madres Trabajadoras. Informe Final de la Evaluación de Impacto.” *Instituto Nacional de Salud Pública*. México. Available at http://www.2006-2012.sedesol.gob.mx/work/models/SEDESOL/EvaluacionProgramasSociales/Evaluacion_Impacto/EI_PEI_2011/Inf_Final_PEI.pdf

Araujo, M. Caridad, Pedro Carneiro, Yyannu Cruz-Aguayo, and Norbert Schady. 2016. “Teacher Quality and Learning Outcomes in Kindergarten.” *Quarterly Journal of Economics* 131(3): 1415-53.

Araujo, M.Caridad, Florencia Lopez-Boo, Rafael Novella, Sarah Schodt, and Romina Tomé. 2015. “The Quality of Centros Infantiles del Buen Vivir in Ecuador.” IADB Policy Brief No.IDB-PB-248. DOI: http://dx.doi.org/10.18235/0000184#sthash.4aTsjRAx.dpuf

Araujo, M. Caridad, Florencia Lopez Boo, and Juan Manuel Puyana. 2013. *Overview of Early Childhood Development Services in Latin America and the Caribbean*. Washington, D.C: Inter-American Development Bank. Available at http://www.iadb.org/es/temas/proteccion-social/desarrollo-infantil-temprano,1929.html.

Baker, Michael, Jonathan Gruber, and Kevin Milligan. 2008. “Universal Child Care, Maternal Labor Supply, and Family Well-Being.” *Journal of Political Economy* 116(4): 709–45.

----------. 2015. “Non-Cognitive Deficits and Young Adult Outcomes: The Long-Run Impacts of a Universal Child Care Program.” NBER Working Paper 21571.

Bandel, Eileen, Nikki Aikens, Cheri A. Vogel, Kimberly Boller, and Lauren Murphy. 2014. “Observed Quality and Psychometric Properties of the CLASS-T in the Early Head Start Family and Child Experiences Survey.” OPRE Technical Brief 2014-34, Mathematica Policy Research. Available at: https://www.acf.hhs.gov/sites/default/files/opre/baby_faces_class_t_final_final_r.pdf

Behrman, Jere R., Yingmei Cheng, and Petra E. Todd. 2004. “Evaluating Preschool Programs When Length of Exposure to the Program Varies: A Nonparametric Approach.” *Review of Economics and Statistics* 86(1): 108–32.

Berlinski, Samuel, Sebastian Galiani, and Paul Gertler. 2009. “The Effect of Pre-Primary Education on Primary School Performance.” *Journal of Public Economics* 93(1–2) February: 219–34.

Berlinski, Samuel, Sebastian Galiani, and Marco Manacorda. 2008. “Giving Children a Better Start: Preschool Attendance and School-Age Profiles.” *Journal of Public Economics* 92(5–6) June: 1416–40.

Berlinski, Samuel, and Norbert Schady. 2015. *The Early Years: Child Well-Being and the Role of Public Policy*. New York: Palgrave Macmillan.

Bernal, Raquel. 2015. “The Impact of a Vocational Education Program for Childcare Providers on Children’s Well-Being.” *Economics of Education Review* 48: 165-83.

Bernal, Raquel, and Camila Fernández. 2013. “Subsidized Childcare and Child Development in Colombia: Effects of Hogares Comunitarios de Bienestar as a Function of Timing and Length of Exposure.” *Social Science and Medicine* 97: 241–49.

Bernal, Raquel, Orazio Attanasio, Ximena Peña, and Marcos Vera-Hernández. 2015. “The Effects of the Transition from Home-Based Childcare to Center-Based Childcare in Colombia.” Universidad de los Andes, Bogotá, and Institute for Fiscal Studies, London. Unpublished.

Bierman, Karen, Celine Domitrovich, Robert Nix, Scott Gest, Janet Welsh, Mark Greenberg, Clancy Blair, Keith Nelson, and Sukhdeep Gill. 2008. “Promoting Academic and Social-Emotional School Readiness: The Head Start REDI Program.” *Child Development* 79: 1802–17.

Blau, David. 2001. *The Child Care Problem: An Economic Analysis.* New York: Russel Sage Foundation.

Blau, David, and Janet Currie. 2006. “Pre-School, Day Care, and After-School Care: Who’s Minding the Kids?” In Eric Hanushek and Finnis Welch, eds., *Handbook of the Economics of Education*, Volume 2. Amsterdam: North Holland.

Bowlby, John. 1969. *Attachment and Loss, Vol. I: Attachment*. New York: Basic Books.

Burchinal, Margaret, Carollee Howes, Robert Pianta, Donna Bryant, Dianne Early, Richard Clifford, and Oscar Barbarin. 2008. “Predicting Child Outcomes at the End of Kindergarten from the Quality of Pre-Kindergarten Teacher-Child Interactions and Instruction.” *Applied Developmental Science* 12(3): 140-53.

Burchinal, Margaret, Nathan Vandergrift, Robert Pianta, and Andrew Mashburn. 2010. “Threshold Analysis of Association between Child Care Quality and Child Outcomes for Low-Income Children in Pre-Kindergarten Programs.” *Early Childhood Research Quarterly* 25(2): 166-76.

Campbell, Frances, Gabriela Conti, James J. Heckman, Seong Hyeok Moon, Rodrigo Pinto, Elizabeth Pungello, and Yi Pan. 2014. “Early Childhood Investments Substantially Boost Adult Health.” *Science* 343(6178): 1478–85.

Campbell, Frances, Craig T. Ramey, Elizabeth Pungello, Joseph Sparling, and Shari Miller-Johnson. 2002. “Early Childhood Education: Young Adult Outcomes from the Abecedarian Project.” *Applied Developmental Science* 6(1): 42–57.

Carneiro, Pedro, and Rita Ginja. 2014. “Long-Term Impacts of Compensatory Preschool on Health and Behavior: Evidence from Head Start.” *American Economic Journal: Economic Policy* 6(4): 135-73.

Carneiro, Pedro and James J. Heckman. 2003. “Human Capital Policy.” In J. Heckman and A. Krueger, eds., *Inequality in America: What Role for Human Capital Policies?* Cambridge, MA: MIT Press.

Carneiro, Pedro, Costas Meghir, and Matthias Parey. 2013. “Maternal Education, Home Environments, and the Development of Children and Adolescents.” *Journal of the European Economic Association* 11: 123-60.

Case, Anne, and Christina Paxson. 2008. “Stature and Status: Height, Ability, and Labor Market Outcomes.” *Journal of Political Economy* 116(3): 499-532.

Chetty, Raj, John N. Friedman, Nathaniel Hilger, Emmanuel Saez, Diane Whitmore Schanzenbach, and Danny Yagan. 2011. “How Does your Kindergarten Classroom Affect your Earnings? Evidence from Project STAR.” *Quarterly Journal of Economics* 126(4): 1593-1660.

Currie, Janet. 2001. ‘‘Early Childhood Education Programs.’’ *Journal of Economic Perspectives* 15(2): 213–238.

Currie, Janet, and Duncan Thomas. 1995. “Does Head Start Make a Difference?” *American Economic Review* 85(3): 341-64.

----------. 2001. “Early Test Scores, Socioeconomic Status, and Future Outcomes.” *Research in Labor Economics* 20: 103-32.

Deming, David. 2009. “Early Childhood Intervention and Life-Cycle Skill Development: Evidence from Head Start.” *American Economic Journal: Applied Economics* 1(3): 111-34.

Downer, Jason, Robert Pianta, Margaret Burchinal, Samuel Field, Bridget Hamre, Jennifer LoCasale-Crouch, Carollee Howes, Karen La Paro, and Catherine Scott-Little. 2013. ‘‘Coaching and Coursework Focused on Teacher- Child Interactions during Language-Literacy Instruction: Effects on Teacher Outcomes and Children’s Classroom Engagement.’’ Unpublished paper, University of Virginia.

Duncan, Greg J., and Katherine Magnuson. 2013. “Investing in Preschool Programs.” *Journal of Economic Perspectives* 27(2): 109-32.

Eckenrode, John, Mary I. Campa, Dennis W. Luckey, Charles R. Henderson, Robert Cole, Harriet Kitzman, Elizabeth Anson, Kimberly Sidora-Arcoleo, Jane Powers, and David Olds. 2010. “Long-Term Effects of Prenatal and Infancy Nurse Home Visitation on the Life Course of Youths: 19-Year Follow-Up of a Randomized Trial.” *Archives of Pediatrics and Adolescent*

Medicine 164(1): 9-15.

Fernald, Lia C.H., Patricia Kariger, Meslissa Hidrobo, and Paul Gertler. 2012. "Socioeconomic Gradients in Child Development in Very Young Children: Evidence from India, Indonesia, Peru, and Senegal." *Proceedings of the National Academy of Sciences* 109(2): 17273-80.

Fox, Sharon E., Pat Levitt, and Charles A. Nelson. 2010. "How the Timing and Quality of Early Experiences Influence the Development of Brain Architecture." *Child Development* 81(1): 28-40.

Garces, Eliana, Janet Currie, and Duncan Thomas. 2002. "Longer-Term Effects of Head Start." *American Economic Review* 92(4): 999-1012.

Gertler, Paul, James Heckman, Rodrigo Pinto, Adriana Zanolini, Christel Vermeersch, Susan Walker, Susan Chang, and Sally Grantham-McGregor. 2014. "Labor Market Returns to an Early Childhood Stimulation Intervention in Jamaica." *Science* 344(6187): 998-1001.

Grantham-McGregor, Sally, Yin Bun Cheung, Santiago Cueto, Paul Glewwe, Linda Richter, Barbara Strupp, and the International Child Development Steering Group. 2007. "Developmental Potential in the First Five Years for Children in Developing Countries." *Lancet* 369(9555): 60–70.

Harms, T., and R. M. Clifford. 1980. Early Childhood Environment Rating Scale. New York: Teachers College Press.

Harms, T., R. M. Clifford, and D. Cryer. 1998. Early Childhood Environment Rating Scale—Revised. New York: Teachers College Press.

Hamre, Bridget K., and Robert Pianta. 2005. "Can Instructional and Emotional Support in the First-Grade Classroom Make a Difference for Children at Risk of School Failure?" *Child Development* 76(5): 949-67.

Hamre, Bridget K., Laura M. Justice, Robert Pianta, Carolyn Kilday, Beverly Sweeney, Jason Downer, and Allison Leach. 2010. "Implementation Fidelity of MyTeachingPartner Literacy and Language Activities: Association with Preschoolers' Language and Literacy Growth." *Early Childhood Research Quarterly* 25(3): 329–347.

Havnes, Tarjei, and Magne Mogstad. 2015. "Is Universal Child Care Leveling the Playing Field?" *Journal of Public Economics* 127(C): 100-14.

Heckman, James, Seong Hyeok Moon, Rodrigo Pinto, Peter A. Savelyev, and Adam Yavitz. 2010. "The Rate of Return to the HighScope Perry Preschool Program." *Journal of Public Economics* 94(1-2): 114-28.

Hoddinott, John, John A. Maluccio, Jere R. Behrman, Rafael Flores, and Reynaldo Martorell. 2008. "Effect of a Nutrition Intervention during Early Childhood on Economic Productivity in Guatemalan Adults." *Lancet* 371(9610): 411-16.

Jacob, Brian. 2007. "The Challenges of Staffing Urban Schools with Effective Teachers." *The Future of Children* 17(1): 129-54.

Kagan, Sharon Lynn, M. Caridad Araujo, Analia Jaimovich, and Yyannú Cruz-Aguayo. 2015.

"Understanding Systems Theory and Thinking: Early Childhood Education in Latin America and the Caribbean." In Anne Farrell, Sharon Lynn Kagan, and Kay M. Tisdall, eds., *SAGE Handbook of Early Childhood Research* (pp. 163-84). Sage: London.

Kline, Patrick, and Christopher R. Walters. 2016. "Evaluating Public Programs with Close Substitutes: The Case of Head Start." *Quarterly Journal of Economics* 131(4): 1795-1848.

Kottelenberg, Michael J., and Steven F. Lehrer. 2016. "Targeted or Universal Coverage? Assessing Heterogeneity in the Effects of Universal Childcare." Forthcoming, *Journal of Labor Economics*.

La Paro, Karen M., Bridget K. Hamre, and Robert Pianta. 2012. *Classroom Assessment Scoring System (CLASS) Manual, Toddler.* Baltimore, MD: Paul H. Brookes.

La Paro, Karen M., Robert Pianta, and Megan Stuhlman. 2004. "The Classroom Assessment Scoring System: Findings from the Prekindergarten Year." *The Elementary School Journal* 104(5): 409-26.

Li, Weilin, George Farkas, Greg J. Duncan, Margaret Burchinal, and Deborah L. Vandell. 2013. "Timing of High-Quality Child Care and Cognitive, Language, and Preacademic Development." *Developmental Psychology* 49(8): 1440-51.

Liu, Jianghong. 2004. "Childhood Externalizing Behavior: Theory and Implications." *Journal of Child and Adolescent Psychiatric Nursing* 17(3): 93-103.

Love, John M., Linda Harrison, Abraham Sagi-Schwartz, Marinus H. Van Ijzendorrn, Christine Ross, Judy A. Ungerer, Helen Raikes, Christy Brady-Smith, Kimberly Boller, Jeanne Brooks-Gunn, Jill Constantine, Ellen Eliason Kisker, Diane Paulsell, and Rachel Chazan-Cohen. 2003. "Child Care Quality Matters: How Conclusions May Vary With Context." *Child Development* 74(4): 1021-33.

Ludwig, Jens, and Douglas Miller. 2007. ''Does Head Start Improve Children's Life Chances? Evidence from a Regression Discontinuity Design.'' *Quarterly Journal of Economics* 122(1): 159–208.

Ludwig, Jens, and Deborah Phillips. 2007. ''The Benefits and Costs of Head Start.'' NBER Working Paper 12973.

Maluccio, John A., John Hoddinott, Jere R. Behrman, Reynaldo Martorell, Agnes R. Quisumbing, and Aryeh D. Stein. 2009. "The Impact of Improving Nutrition During Early Childhood on Education among Guatemalan Adults." *Economic Journal* 119(537): 734-63.

Martínez, Sebastian, Sophie Naudeau, and Vitor Pereira, V. 2012. "The Promise of Preschool in Africa: a Randomized Evaluation of Early Childhood Development in Rural Mozambique". Unpublished paper, The World Bank and Save the Children. Available at http://siteresources.worldbank.org/INTAFRICA/Resources/The_Promise_of_Preschool_in_Africa_ECD_REPORT.pdf.

Mashburn, Andrew J., Robert Pianta, Bridget K. Hamre, Jason T. Downer, Oscar A. Barbarin, Donna Bryant, Margaret Burchinal, Diane M. Early, and Carollee Howes. 2008. "Measures of Classroom Quality in Prekindergarten and Children's Development of Academic, Language, and Social Skills." *Child Development* 79(3): 732-49.

Moffitt, Terrie, Louise Arseneault, Daniel Belsky, Nigel Dickson, Robert Hancox, HonaLee Harrington, Renate Houts, Richie Poulton, Brent Roberts, Stephen Ross, Malcolm Sears, W. Murray Thomson, and Avshalom Caspi. 2011. "A Gradient of Childhood Self-Control Predicts Health, Wealth, and Public Safety." *Proceedings of the National Academy of Sciences* 108(7): 2693-98.

Nelson, Charles A., and Margaret A. Sheridan. 2011. "Lessons from Neuroscience Research for Understanding Causal Links between Family and Neighborhood Characteristics and Educational Outcomes." In Greg J. Duncan and Richard J. Murnane, eds., *Whither Opportunity: Rising Inequality, Schools, and Children Life Chances* (pp. 27-46). New York: Russell Sage Foundation.

National Institute of Child Health and Human Development (NICHD), Early Child Care Research Network. 2000a. "Characteristics and Quality of Child Care for Toddlers and Preschoolers." *Applied Developmental Science* 4(3): 116-135.

----------. 2000b. "The Relation of Child Care to Cognitive and Language Development." *Child Development* 71(4): 960-80.

National Institute of Child Health and Human Development Early (NICHD) Child Care Research Network, and Greg J. Duncan. 2003. "Modeling the Impacts of Child Care Quality on Children's Preschool Cognitive Development." *Child Development* 74(5): 1454-75.

National Institute of Child Health and Human Development Early (NICHD) Child Care Research Network. 1996. "Observational Record of the Caregiving Environment" *Early Childhood Research Quarterly* 11(3): 296-306.

Organisation for Economic Co-operation and Development (OECD). 2014. PISA 2012: *Results in Focus: What 15-Year-Olds Know and What They Can Do With What They Know*. Available at http://www.oecd.org/pisa/keyfindings/pisa-2012-results-overview.pdf. Accessed on November 25, 2016.

----------. 2016. PISA 2015: Results in Focus. Available at http://www.oecd.org/pisa/pisa-2015-results-in-focus.pdf. Accessed on December 21, 2016.

Olds, David, Charles R. Henderson, Robert Cole, Harriet Kitzman, Dennis W. Luckey, Lisa Pettitt, Kimberly Sidora, Pamela Morris, and Jane Powers. 1998. "Long-Term Effects of Nurse Home Visitation on Children's Criminal and Antisocial Behavior." *Journal of the American Medical Association* 280(14): 1238-44.

Peisner-Feinberg, Ellen S., Margaret Burchinal, Richard M. Clifford, Mary L. Culkin, Carolle Howes, Sharon Lynn Kagan, and Noreen Yazejian. 2001. "The Relation of Preschool Child-Care Quality to Children's Cognitive and Social Developmental Trajectories through Second Grade." *Child Development* 72(5): 1534-53.

Pianta, Robert. 2003. *Experiences in p-3 Classrooms: The Implications of Observational Research for Re-Designing Early Education.* New York: Foundation for Child Development.

Pianta, Robert, Jason Downer, and Bridget Hamre. 2016. "Quality in Early Education Classrooms: Definitions, Gaps, and Systems." *Future of Children* 26(2): 119-37.

Pianta, Robert, Karen M. La Paro and Bridget K. Hamre. 2007. *Classroom Assessment*

Scoring System—CLASS. Baltimore: Brookes.

Porter, Noriko. 2012. “High Turnover among Early Childhood Educators in the United States.” Available at http://www.childresearch.net/projects/ecec/2012_04.html.

Puma, Michael, Stephen Bell, Ronna Cook, and Camilla Heid. 2010. *Head Start Impact Study: Final Report*. Washington, DC: U.S. Department of Health and Services, Administration for Children and Families.

Puma, Michael, Stephen Bell, and Camilla Heid. 2012. *Third Grade Follow-Up to the Head Start Impact Study*. Washington, DC: U.S. Department of Health and Human Services, Administration for Children and Families.

Raikes, Helen. 1993. “Relationship Duration in Infant Care: Time with a High-Ability Teacher and Infant-Teacher Attachment. *Early Childhood Research Quarterly* 8(3): 309-25.

Romano, Joseph P., and Michael Wolf. 2005. “Stepwise Multiple Testing as Formalized Data Snooping.” *Econometrica* 73(4): 1237-82.

Rosero, José, and Hessel Oosterbeek. 2011. “Trade-offs between Different Early Childhood Interventions: Evidence from Ecuador.” Discussion Paper No. 11-102/3, Faculty of Economics and Business, University of Amsterdam, and Tinbergen Institute, Amsterdam.

Rubio-Codina, Marta, M. Caridad Araujo, Orazio Attanasio, Pablo Muñoz, and Sally Grantham-McGregor. 2016. “Concurrent Validity and Feasibility of Short Tests Currently Used to Measure Early Childhood Development in Large Scale Studies.” *PlosOne.* Published: August 22, 2016. http://dx.doi.org/10.1371/journal.pone.0160962.

Rutter, Michael, and the English and Romanian Adoptees (ERA) Study Team. 1998. “Developmental Catch-Up, and Deficit, Following Adoption After Severe Global Early Privation.” *Journal of Child Psychology and Psychiatry* 39(4): 465-76.

Ruzek, Erik, Margaret Burchinal, George Farkas, and Greg J. Duncan. 2014. “The Quality of Toddler Child Care and Cognitive Skills at 24 Months: Propensity Score Analysis Results from the ECLS-B”. *Early Childhood Research Quarterly* 29(1): 12-21.

Scarr, Sandra, Marlene Eisenberg, and Kirby Deater-Deckard. 1994. “Measurement of Quality in Child Care Centers.” *Early Childhood Research Quarterly* 9(2): 131-51.

Schady, Norbert Jere R. Behrman, M. Caridad Araujo, Rodrigo Azuero, Raquel Bernal, David Bravo, Florencia López Boo, Karen Macours, Daniella Marshall, Christina Paxson, and Renos Vakis. 2015. “Wealth Gradients in Early Childhood Cognitive Development in Five Latin American Countries.” *Journal of Human Resources* 50(2): 446–63.

Schweinhart, Lawrence J., Jeanne Montie, Zongping Xiang, W. Steven Barnett, Clive R. Belfield, and Milagros Nores. 2005. *Lifetime Effects: The High/Scope Perry Preschool Study Through Age 40.* Ypsilanti, MI: High/Scope Press.

Shonkoff, Jack P. and Deborah A. Phillips. 2000. *From Neurons to Neighborhoods: The Science of Early Childhood Development*. Washington, D.C.: National Academy Press.

Squires Jane, Diane Bricker, Elizabeth Twombly, Robert Nickel, Jantina Clifford, Kimberly Murphy, Robert Hoselton, LaWanda Potter, Linda Mounts, and Jane Farrell. 2009. *Ages &*

Stages English Questionnaires, Third Edition (ASQ-3): A Parent-Completed, Child-Monitoring System. Baltimore, MD: Paul H. Brookes Publishing Co.

Staiger, Douglas O., and Jonah E. Rockoff. 2010. “Searching for Effective Teachers with Imperfect Information.” *Journal of Economic Perspectives* 24(3): 97-118.

Thomason, Amy C, and Karen M. La Paro. 2009. “Measuring the Quality of Teacher-Child Interactions in Toddler Child Care.” *Early Education and Development*. 20(2): 285-304.

United Nations Educational, Scientific and Cultural Organization (UNESCO). 2015. *Informe de Resultados TERCE (Tercer Estudio Regional Comparativo y Explicativo)*. Available at http://unesdoc.unesco.org/images/0024/002435/243532S.pdf. Accessed on November 25, 2016.

Vandell, Deborah L. and Barbara Wolfe. 2000. *Child Care Quality: Does It Matter and Does It Need to Be Improved?* Washington, D.C: US Department of Health and Human Services, Office for Planning and Evaluation.

Vandell, Deborah L., Jay Belsky, Margaret Burchinal, Nathan Vandergrift, Laurence Steinberg, and NICHD Early Child Care. 2010. “Do Effects of Early Child Care Extend to Age 15 Years? Results from the NICHD Study of Early Child Care and Youth Development.” *Child Development* 81(3): 737-56.

Whitebook, Marcy, Laura Sakai, Emily Gerber, and Carollee Howes. 2001. “Then and Now: Changes in Child Care Staffing, 1994-2000”. *Washington, DC and Berkeley, CA: Center for the Child Care Workforce and the Institute of Industrial Relations, University of California.*

World Bank. 2016. “World Development Indicators.” Available at http://databank.worldbank.org/data/reports.aspx?source=world-development-indicators. Accessed on November 25, 2016.

Table 1: Summary Statistics for Children and their Families

	Mean/ Proportion	Standard Deviation	N
Child age in months	17.04	4.85	2,198
Proportion of children aged 6-12 months	0.21	0.41	2,198
Proportion of children 13-18 months old	0.36	0.48	2,198
Proportion of children 19-24 months old	0.44	0.50	2,198
Proportion of female	0.47	0.50	2,198
Mother's age	28.15	6.96	2,175
Mother's years of schooling	10.05	3.27	2,175
Proportion of indigenous mothers	0.16	0.36	2,153
Number of household members	4.75	1.86	2,196
Proportion of households with both parents	0.78	0.42	2,198
Number of bedrooms in home	1.92	1.13	2,188
Household has piped water	0.78	0.41	2,198
Household is connected to sewerage system	0.64	0.48	2,198
Main material of floors is earth	0.32	0.47	2,198
Number of assets (0-11)	5.09	2.63	2,198
Distance to child care center (minutes)	10.59	15.18	2,182
Proportion urban	0.88	0.32	2,198

Note: The summary statistics refer to the 2,198 children 6-24 months of age who attend the 291 centers in the estimation sample. Assets include refrigerator, gas stove, washing machine, iron, blender, TV, DVD, computer, stereo, cell phone, and cable.

Table 2: Summary Statistics for Caregivers, Center Coordinators, and Child Care Centers

	Mean/ Proportion	Standard Deviation	N
Caregiver characteristics			
Total CLASS score	3.15	0.33	582
CLASS-emotional and behavioral support	3.95	0.36	582
CLASS-engaged support for learning	1.81	0.34	582
Years of experience in center	2.05	2.86	582
Years of completed schooling	10.01	3.03	581
Proportion of indigenous	0.19	0.39	576
Supervisor characteristics			
Years of experience in center	1.01	2.36	184
Years of completed schooling	15.84	1.17	179
Classroom and center characteristics			
Number of children per classroom	8.68	2.28	582
Number of children attending the center	18.20	6.09	291
Number of classrooms	2.64	0.80	291
Center has electricity	0.67	0.47	291
Center has piped water	0.74	0.44	291
Center is connected to sewerage system	0.65	0.48	291

Note: The summary statistics refer to the 582 caregivers, 184 supervisors, and 291 centers in the estimation sample. The number of children per classroom refers to the total number of children (including children older than 24 months of age and, therefore, not included in our analysis) who attend the 582 classrooms in the sample. The total number of classrooms in the center refers to all classrooms in a given center, including the two classrooms that are part of our analysis sample.

Table 3: Classroom Quality and Child Development Outcomes

	Total ASQ score							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Total CLASS	0.070**	0.069**					0.071**	0.073**
	(0.034)	(0.033)					(0.034)	(0.033)
Experience			0.030***	0.030***			0.030***	0.029***
			(0.011)	(0.011)			(0.012)	(0.011)
Education					0.002	0.001	0.004	0.001
					(0.012)	(0.012)	(0.012)	(0.012)
Controls	N	Y	N	Y	N	Y	N	Y

Note: Sample size is 2,198 in all specifications. The dependent variable is the ASQ score (mean zero, unit standard deviation). The CLASS score has been normalized to have mean zero and unit standard deviation, experience is years working at the child care center, and education is the years of completed schooling. All regressions include center fixed effects. Odd-numbered specifications include no controls. In the even-numbered columns, controls include child gender; household demographics (the education, age, and ethnicity of the mother; whether both parents live at home; the number of household members); the number of assets in the household (including refrigerator, gas stove, washing machine, iron, blender, TV, DVD, computer, stereo, cell phone, and cable); the distance of the household to the center (in minutes), variables that measure the quality of the home (whether the household has piped water inside the home, and separately, is connected to the sewerage system; the number of bedrooms; whether the house has dirt floors); and controls for the composition of the classroom (number of children, the proportion of female, and the mean age). Standard errors are clustered at the center level. *, **, ***, significant at the 10 percent, 5 percent and 1 percent levels, respectively.

Table 4: Heterogeneity of Quality Effects

	Total ASQ		ASQ Communication		ASQ Fine Motor		ASQ Problem Solving	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Experience	0.030***	0.030***	0.021*	0.024**[1]	0.029**[2]	0.027**[1]	0.018	0.016
	(0.011)	(0.011)	(0.011)	(0.010)	(0.012)	(0.011)	(0.012)	(0.012)
Total CLASS	0.070**	0.069**	0.021	0.028	0.049	0.050	0.091**[2]	0.080**[1]
	(0.034)	(0.033)	(0.036)	(0.035)	(0.035)	(0.034)	(0.036)	(0.035)
CLASS-emotional and behavioral support	0.054	0.056*	0.002	0.009	0.044	0.047	0.079**	0.073**
	(0.034)	(0.033)	(0.034)	(0.034)	(0.036)	(0.035)	(0.035)	(0.034)
CLASS-engaged support for learning	0.081**	0.077**	0.050	0.057	0.045	0.044	0.089**[1]	0.072**
	(0.036)	(0.037)	(0.041)	(0.039)	(0.038)	(0.037)	(0.036)	(0.036)
Controls	N	Y	N	Y	N	Y	N	Y

Note: Each cell corresponds to a different regression (32 separate regressions). Sample size is 2,198 in all specifications. The dependent variables and the CLASS score have been normalized to have mean zero and unit standard deviation. Experience is years working at the child care center. All regressions include center fixed effects. Odd-numbered specifications include no controls. The even-numbered columns include the controls listed at the foot of Table 3. Standard errors clustered at the center level are reported in parentheses. *, **, ***, significant at the 10 percent, 5 percent and 1 percent level, respectively, when standard errors are not corrected for multiple hypothesis testing.[1], [2], at the 10 percent and 5 percent level, respectively, when standard errors are corrected for multiple hypothesis testing using the step-down procedure in Romano and Wolf (2005).

Table 5: Distributional Effects of Differences in Classroom Quality

	Total CLASS		Experience	
	(1)	(2)	(1)	(2)
Mean effect	0.070**	0.069**	0.030***	0.030***
	(0.034)	(0.033)	(0.011)	(0.011)
Fixed-effects regressions for dummy variables for scores above (below) different points in the distribution				
Below 10th percentile	-0.031***	-0.031***	-0.007*	-0.007*
	(0.011)	(0.011)	(0.004)	(0.004)
Below 25th percentile	-0.007	-0.007	-0.008	-0.006
	(0.016)	(0.016)	(0.005)	(0.005)
Below 50th percentile	-0.034**	-0.036**	-0.012**	-0.013**
	(0.017)	(0.018)	(0.006)	(0.006)
Above 50th percentile	0.034**	0.036**	0.012**	0.013**
	(0.017)	(0.018)	(0.006)	(0.006)
Above 75th percentile	0.024	0.024	0.009*	0.009*
	(0.016)	(0.015)	(0.005)	(0.005)
Above 90th percentile	0.004	0.003	0.009*	0.010**
	(0.011)	(0.011)	(0.004)	(0.004)
Controls	N	Y	N	Y

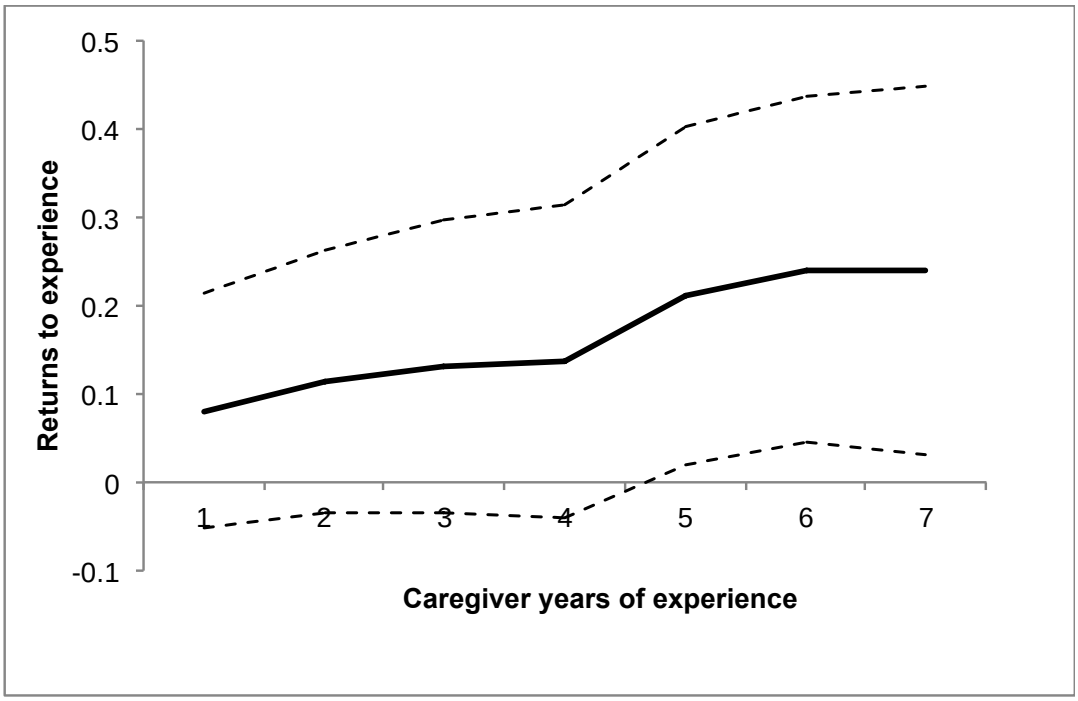
Note: Sample size is 2,198 in all specifications. The dependent variable is given by dummy variables that take on the value of one if a child has a total ASQ score below the 10th, 25th, and 50th percentiles, respectively, or above the 50th, 75th, and 90th percentiles, respectively. The CLASS score has been normalized to have mean zero and unit standard deviation, and experience is years working at the child care center. All regressions include center fixed effects. See the note to Table 3 for the list of controls. Standard errors are clustered at the center level. *, **, ***, significant at the 10 percent, 5 percent and 1 percent level respectively.

Table 6: What Characteristics of Caregivers Are Valued by Parents and Supervisors?

	(1)	(2)	(3)	(4)	(5)
Panel I: Parents					
Total CLASS	0.012				0.012
	(0.020)				(0.021)
Experience		0.010*			0.010*
		(0.005)			(0.005)
Education			-0.007		-0.006
			(0.007)		(0.007)
Average total ASQ score				0.004	-0.003
				(0.028)	(0.028)
Panel II: Supervisors					
Total CLASS	-0.002				-0.004
	(0.029)				(0.027)
Experience		0.080***			0.080***
		(0.011)			(0.011)
Education			-0.018**		-0.010
			(0.008)		(0.008)
Average total ASQ score				0.020	-0.015
				(0.047)	(0.039)

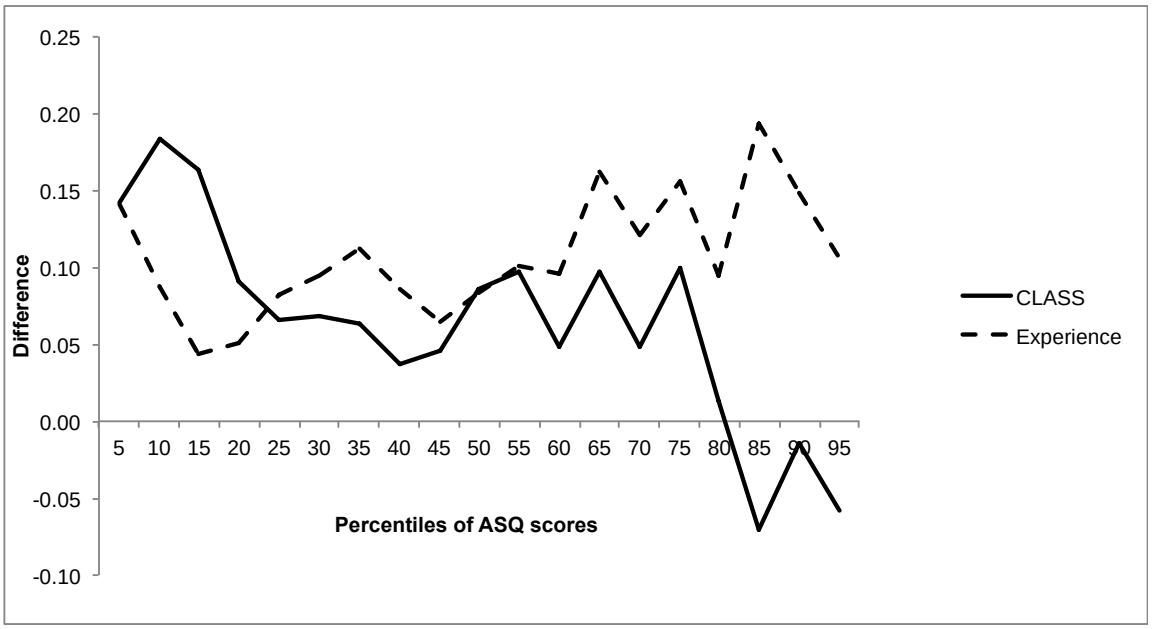
Note: Sample size is 2,122 in all regressions for parents (one for each parent), 572 in all regressions for supervisors (one for each caregiver for which data on ranking is available). The dependent variable in Panel I is a dummy variable equal to 1 if the mother rated the caregiver as "very good", 0 otherwise; and in Panel II a dummy variable equal to 1 if caregivers were assigned the highest performance category (out of 3) by the supervisor. All regressions for parents include center fixed effects, and the controls at the base of Table 3; all regressions for supervisors include supervisor fixed effects, as well as controls for the total number of children in each classroom, the proportion of them that are girls, and the mean classroom age. Standard errors are clustered at the center level in the regressions for parents, at the supervisor level in the regressions for supervisors. *, **, ***, significant at the 10 percent, 5 percent and 1 percent level, respectively.

Figure 1: Returns to Experience, Alternative Definitions of "Experienced" and "Inexperienced" Caregivers



Note: the figure graphs the coefficients and 95 percent confidence intervals from separate regressions of child development on alternative definitions of "experienced" and "inexperienced" caregivers. For example, the first (leftmost) point corresponds to a regression in which an experienced caregiver is defined as having at least 1 year of experience, the second point to a regression in which an experienced caregiver is defined as having at least 2 years of experience, and so on. All regressions include the controls listed at the foot of Table 3 and center fixed effects. Standard errors are clustered at the center level.

Figure 2: Distributional Effects of Caregiver CLASS Scores and Experience



Note: To generate the figure, we proceeded as follows. First, we calculated the mean CLASS score and the mean experience of caregivers in each center. Second, we calculated whether every caregiver has CLASS scores above or below the center mean ("high CLASS" and "low CLASS", respectively) and experience above or below the center mean ("high experience" and "low experience", respectively). Third, we calculated ASQ scores of children at the 5^{th}, 10^{th} … 95^{th} percentiles of the distribution in each of these four groups (high and low CLASS, high and low experience). Fourth, we took the difference in scores at each ventile between high CLASS and low CLASS, and between high experience and low experience. The figure graphs these differences.

Appendix A: The Ages and Stages Questionnaire (ASQ)

The Ages and Stages Questionnaires screener (third version, ASQ henceforth; Squires et al. 2009) was administered to measure child development in the children of the sample. The ASQ consists of age-specific questionnaires, a total of 10 for the 6-24 months-old range. While the ASQ was originally designed as a questionnaire to be completed by the mother, for this study it was administered by an interviewer to the child in the presence of his or her mother or main caregiver in the home. The Spanish version of the test was modified by local psychologists to adjust it to the Peruvian context. Moreover, in order to establish homogeneous administration protocols across the sample, the team of psychologists defined specific items that interviewers were required to administer directly to the child, while others could be collected by maternal report. As a result of this adjustment, across all age-specific questionnaires, the majority of items of the fine motor and problem solving scales were administered directly to children, while the majority of items in the communication scale were collected by maternal report. More precisely, across all age-specific questionnaires, only 6.7 percent of all items in the fine motor scale and 10 percent of all items in the problem solving scale had to be collected primarily by maternal report. In turn, 71 percent in the communication scale had to be collected by maternal report. Careful piloting of the instruments ensured that the recommended administration protocol could actually be followed during data collection.

Each ASQ age-specific questionnaire includes six items per developmental area. Every item receives a score of 10 if the child completes it, 5 if she does it sometimes, and 0 if she does not do it, for a maximum of 60 points per scale. Given that the test was designed as a screener, and in order to reduce the number of children that reach the test ceiling (and therefore increase score variability in the upper end of the distribution), for children who completed all six items in a scale correctly, we also administered the first three items of the questionnaire corresponding to the subsequent age, following Rubio-Codina et al. (2016). This adjustment allowed for children who otherwise would have reached the test ceiling (10-11 percent of the sample depending on the scale) to be tested on three additional, more difficult items. As a result of this adjustment, ASQ raw scores are over a total of 90 points.

The psychologists who adjusted the language and administration of ASQ also carried out its test-retest in a sample of 26 children, with an average difference of two weeks between the first and the second administration. Considering 9 items per developmental area, the intra-child correlation (ICC) obtained was 0.80-0.85.

The ASQ was administered by a team of 23 interviewers, selected out of 36 who participated in the training. Interviewers had post-secondary education, but no prior experience administering the ASQ. They had all worked before as survey enumerators. Interviewer training was conducted by one of the psychologists who adapted the ASQ. It included 5 days of classroom training and a minimum of 10-15 practice administrations, some of them supervised by the trainer. All ASQ administrations were conducted in Spanish. While our survey was short and did not collect data on the language spoken at home as this was an urban sample and only 15 percent of the household-heads interviewed self-identified as indigenous based on their culture and customs.

Table A1 shows the internal consistency of the instrument using two types of indicators: the correlation between each scale and the overall score of the instrument, and Cronbach's alphas. [24,25] The Cronbach's alphas were calculated separately for each questionnaire. The figures reported refer to the average of the ten alphas (i.e., one per questionnaire) for each developmental area. Cronbach's α for the ASQ were within a reasonable range for the three areas –α=0.644 for communication, α=0.567 for fine motor, α=0.592 for problem solving–, and good for the total (i.e., the items of all three areas together) with α=0.751. The correlation coefficients are all high.

Table A1. Internal consistency of the ASQ-3

	Correlation	Alpha
Communication	0.718***	0.644
Fine Motor	0.744***	0.567
Problem Solving	0.750***	0.592
Full Scale		0.751

Note: Cronbach's alphas and Pearson correlation coefficients between each developmental area and the total score. *, **, ***, significant at the 10 percent, 5 percent and 1 percent respectively.

[24] Generally, the literature considers a measure of internal consistency to be reasonable when it falls within the range of 0.60 to 0.70 for both indicators.

[25] Correlations are considered very high when in the range of 0.80 to 1, high in the range of 0.60 to 0.80, moderate in the range of 0.40 to 0.60, low in the range of 0.20 to 0.40, and very low when less than 0.20.

Appendix B: The Toddler Classroom Assessment Scoring System (CLASS)

Process quality in the classroom was measured using the Toddler Classroom Assessment Scoring System (CLASS hereafter; La Paro et al. 2012; Pianta et al. 2007) coded from four 20-minute videos recorded over the course of a normal day at the center. The CLASS is an observational instrument for use with children 15-36 months of age. It assesses the quality of teacher-child interactions, in two domains: emotional and behavioral support, and engaged support for learning.

Each video segment was coded twice, by two different coders assigned at random. A team of six CLASS-certified observers did the coding. Coders carried out daily group exercises supervised by a CLASS-certified trainer to stay reliable over time. When scale scores differed by more than 2 points (1 point in scales of less variability), a third coding was carried out, and the score for the segment corresponds to the average of the two with the smallest discrepancy. Classroom scores are averaged for the four segments. CLASS scores were computed for the 582 classrooms in the sample.

Ninety-eight percent of videos were recorded in Spanish; of the remainder, 8 videos had fragments in Spanish and an indigenous language and 2 videos were exclusively in an indigenous language. These 10 videos correspond to 8 different centers. Coders were all native Spanish speakers and they did not speak indigenous languages.

Table B1 presents the internal consistency of the instrument using the same indicators as for the ASQ in Appendix A (i.e., the correlation between each subscale and the overall score of the instrument, and Cronbach's alphas). Overall, the instrument shows excellent consistency with both indicators. The correlations are all in the very high range (i.e., above 0.80), with two exceptions: regard for child's perspectives in the high range (ICC=0.73), and negative climate in the moderate range (ICC=0.53). The lower correlation for negative climate as compared to other subscales could be due to its low variability in the sample. Cronbach's α were good: α=0.87 for emotional and behavioral support, α=0.85 for engaged support for learning, and α=0.91 for the full scale.

Table B1. Internal Consistency of CLASS

	Correlation	Alpha
Emotional and Behavioral Support	**0.959*** **	**0.870**
1. Positive Climate	0.916***	
2. Negative Climate	0.530***	
3. Teacher Sensitivity	0.931***	
4. Regard for Child Perspectives	0.734***	
5. Behavior Guidance	0.846***	
Engaged Support for Learning	**0.858*** **	**0.853**
6. Facilitation of Learning and Development	0.881***	
7. Quality of Feedback	0.852***	
8. Language Modeling	0.897***	
Full Scale		**0.906**

Note: Cronbach's alphas for CLASS dimensions and domains and Pearson correlation coefficients between CLASS dimensions and domains with the total score. *, **, ***, significant at the 10 percent, 5 percent and 1 percent level respectively.