The Effects of the Transition from Home-based Childcare to Childcare Centers on Children’s Health and Development in Colombia

Abstract

Colombia’s national early childhood strategy launched in 2011 aimed at improving the quality of childcare services offered to socio-economically vulnerable children, and included the possibility that children and their childcare providers could transfer from non-parental family daycare units to large childcare centers in urban areas. This study seeks to understand whether the offer to transfer from one program to the other had an impact on child cognitive and socioemotional development, and nutrition, using a cluster-randomized control trial with a sample of 2,767 children between the ages of 6 and 60 months located in 14 cities in Colombia. The results indicate a negative effect of this initiative on cognitive development, a positive effect on nutrition and no statistically significant effect of the intervention on socioemotional development. We also explored the extent to which these impacts might be explained by differences in the quality of both services during the transition, and report that quality indicators are low in both programs but are significantly worse in centers compared to community nurseries.

Keywords: Colombia, child development, home-based childcare, center-based childcare, poverty, nutrition
The Effects of the Transition from Home-based Childcare to Childcare Centers on Children’s Health and Development in Colombia

An overwhelming 250 million children younger than 5 in developing countries have been identified as being at risk of not reaching their full developmental potential because they live in poverty and face the psychosocial and material challenges associated with economic adversity (Black et al., 2017). These children have higher probabilities of underperforming at school and, thus, of reaching adulthood with lower earnings capacity, higher probability of risky behaviors and lower overall quality of life (Britto et al., 2017).

The period between birth and 5 years of age has been identified as critical to alter the long-term developmental trajectories of individuals (Cunha, Heckman, Lochner & Masterov, 2006; Engle et al., 2007; Heckman, 2006; Yoshikawa et al., 2013). Although there is increasing agreement about the value of early interventions, there is less consensus about the specific types of interventions that are most effective. In addition, most of the available evidence is based on a few rigorous small randomized trials for developed countries while developing nations usually have limited resources and capacity to conduct longitudinal or randomized control trial evaluations of large-scale programs (Britto, Yoshikawa & Boller, 2011).

While the initial focus in low- and middle-income countries (LMIC) was on increasing access to early childhood education (ECE), the emphasis has now shifted to quality, particularly where enrollment rates are high (Yoshikawa et al., 2015). However, there is little evidence on how to increase quality effectively at scale (Britto et al., 2017). Several LMIC initially expanded early education services through low-cost home-based childcare (Nores & Barnett, 2010). Often these home-based interventions are contrasted with considerably more expensive center-based ones,
under the presumption that the latter are of higher quality than the former. This study provides a
direct and rigorous comparison between these two types of childcare programs.

**Background and Early Childhood Policy in Colombia**

In Colombia, 2.8 million children younger than 6 years of age live in poverty, 14% of these
poor children are stunted and their scores in receptive language are one standard deviation below
those of their peers in higher SES households (Bernal & Quintero, 2014). Among socio-
economically vulnerable children 0-6 years old, enrollment in public ECE programs ranged from
20% to 40% for most of the period since the late 1980s (Bernal & Camacho, 2011). All public
ECE programs are targeted to children and families in poverty (representing close to 65% of the
total population). More than half of the provision was through small home-based community
nurseries (*Hogares Comunitarios de Bienestar* or HCBs henceforth) since the 1980s. HCBs are
non-parental family childcare units where care is provided in the childcare provider’s own home.
HCBs are run by one woman in the community known as “communitarian mother” (MC). Each
HCB serves 12-15 children between ages 6 months and 6 years from poor households in full-day
schedules, and provide 70% of daily nutritional requirements. MCs typically have a high school
degree (67.3%) or some higher education (17%), with a smaller percentage having only elementary
education (15%). MCs are required to attend a 40 hour pre-service training. The average cost of
the program is 440 US dollars (USD) per child per year.

Bernal and Fernández (2013) report that exposure to the HCB program for more than 15
months had positive effects on child cognitive and socio-emotional development. Similarly,
Attanasio, Di Maro and Vera-Hernández (2013) found that HCBs improve children’s nutritional
status. However, Bernal et al. (2009) reported deficiencies in the quality of HCBs: the average
Family Day Care Rating Scale (FDCRS) scores (Harms & Clifford, 1989) were about 3.08 (scores
range from 1 = insufficient quality to 7 = optimal quality) in a sample of 400 HCBs nationwide, with a standard deviation of 0.89. Using similar instruments, a variety of ECE programs in the Latin-American region have been shown to exhibit similar low levels of quality (Araujo, López Boo, Novella, Schodt, & Tomé, 2015; Berlinski & Schady, 2015; and Verdisco & Pérez Alfaro, 2010), in countries such as Brazil, Ecuador and Perú.

It is worth noting that despite the quality issues of HCBs reported in the evaluation study, Bernal and Fernández (2013) and Attanasio et al. (2013) estimated positive program effects. The former study compared treated children with high exposure to the program with treated children with low exposure to the program. The latter paper used an instrumental variable approach, using variation in the supply of program slots and distance to the HCB as instruments for program participation. The identification strategy used in both studies reflects the fact that treated children were found to be systematically different from similar children residing in the same neighborhoods who did not participate in the program. In particular, participating children were more socioeconomically vulnerable and with significant developmental risks. Both papers found that extended exposure to the program significantly improved the condition of very disadvantaged children, despite quality issues, probably, because of very poor learning environments at home.

Around 2011, nearly 1.2 million children were being served by community-based programs (enrollment of about 25%-30% of eligible populations), mainly HCBs, while only about 130 thousand children attended public center-based care. For that reason, very little was known about the quality and impacts of center-based childcare at the time. Apart from these public services provided through the social protection system, specifically the Colombian Family Welfare Agency (Instituto Colombiano de Bienestar Familiar or ICBF henceforth), there was also a supply of private preschools targeting socioeconomically vulnerable populations. In 2010, close to 60% of
all poor children attending some ECE program attended public programs provided by ICBF or local governments, while the other 40% was provided privately. Between 2010 and 2013 some public schools in the formal education sector, especially in big cities, introduced one preschool grade. The latter accounted for 6% of total ECE enrollment of low SES populations in 2013. These services were provided in schools offering primary and secondary education with teachers from the formal education sector. To our knowledge, there is no data regarding the quality of these services, which have continued expanding since 2013.

On the other hand, Maldonado-Carreño and Votruba-Drzal (2014) reported that the quality of private ECE for socioeconomically vulnerable populations was not significantly better than that found for public programs as the ones under evaluation in this study. The authors used the Classroom Assessment Scoring System –CLASS– (Pianta, LaParo & Hamre, 2008), which evaluates the quality of interactions among teachers and children in classrooms. In particular, it measures the emotional, organizational, and instructional supports provided by teachers. Scores are completed using a 7-point scale, with the low range being a score of 1-2, middle range 3-5 and the high range 6-7. The authors reported that only the emotional support domain scores above four, while the classroom organization and instructional support domains score between 2 and 3.

Starting in 2011, the national government decided to prioritize early childhood investment as a result of increased evidence about the importance of pre-primary education programs for development and equity. In 2011, the government launched the national early childhood strategy De Cero a Siempre (From Zero to Forever, DCAS) aimed at increasing access and, most importantly, improving the quality of early childhood services provided to poor children. The objective was to deliver high-quality and comprehensive early childhood services for 1.2 million children under the age of 6 in poverty.
The main hypothesis of the DCAS theory of change is that given the holistic nature of early childhood development, it is critical to provide fully integrated / comprehensive early childhood services to successfully promote early development (Comisión Inter-Sectorial para la Primera Infancia [CIPI], 2013). The Colombian Board for Early Childhood defined a service or program as fully integrated if it could provide concurrently: childcare, health and nutrition, early education, recreation and the exercise of the child’s rights as a citizen. In light of this approach, HCBs were not deemed as fully integrated services because of the way that they operated at the time of the introduction of DCAS. For this reason, one of the initial pillars of the DCAS strategy was to offer children the possibility to transfer from HCBs to large childcare centers (Centros de Desarrollo Infantil or CDIs henceforth) in urban areas. The strategy was based on the assumption that CDIs could provide fully integrated services, which we discuss in detail below.

In this study, we investigated whether the offer to transfer from HCBs to CDIs had impacts on child cognitive and socioemotional development, and nutrition. We also sought to understand some mediational pathways that could potentially explain these effects. In particular, we studied the differences in quality between the two programs and discuss how these could have mediated the estimated impacts. With this objective, in 2010, we designed and implemented a clustered-randomized trial to evaluate the impacts of the offer to transfer from HCBs to CDIs in Colombia.

The Intervention

CDIs are different from HCBs in a variety of ways. First, each CDI serves close to 300 children between the ages of 6 months and 5 years of age in buildings specifically designed for early education (compared with 12-15 children being cared for in the provider’s own household in HCBs) with specialized areas such as playground and refectory. An appropriate infrastructure is often thought of as a minimum requirement for quality in terms of suitable physical spaces and
basic safety features. CDI slots were first offered to children who were in randomly selected nearby HCBs. If these children did not fill all available vacancies, the remainder slots were offered to eligible children residing in the community. MCs with at least a high-school degree could transfer to CDIs as classroom teachers.

Second, CDIs are organized in same-age classrooms, while HCBs serve children from 6 months to 5 years of age in the same physical space. There is mixed evidence regarding the benefits of age-specific preschool classrooms with respect to mixed-age classrooms. Moller, Forbes-Jones and Hightower (2008) argued that same-age classrooms allow teachers to target their curriculum more specifically for a particular age range, which allows them to focus on what is developmentally appropriate for that group. On the other hand, Guo, Tompkings, Justice and Petscher (2016) found positive effects of classroom variation in age composition on children’s vocabulary gains for younger but not for older children. Similarly, Bandura (1986) showed that in age-mixed classrooms, young children learn by emulating and observing older children, and Derscheid (1997) provided evidence that older children practice prosocial behaviors by supporting and mentoring younger children. However, Goldman & Chaillé (1984) found no evidence that in mixed-age classrooms, older children engage in nurturing and mentoring relationships with younger children that might enhance child development as compared to same-age groups.

Third, in addition to teachers, CDIs also hire housekeeping and administrative staff (e.g., kitchen, cleaning, and center director) as well as a nutritionist and a psychologist, allowing teachers to devote more time to classroom activities as compared to MCs in HCBs who have to cook, clean and run their own service unit, in addition to caring for the children. At the same time, teachers have to comply with operational guidelines (administrative, not pedagogical) regarding preparation of materials, documentation and child assessments, reports to parents, safety in the
classroom, etc. These procedures are not as structured and certainly not well monitored in HCBs. Andrew et al. (2016), who assessed the effects of the presence of the interdisciplinary group of professionals in childcare centers on children’s development, reported that only about 38% of centers fully complied with these staffing guidelines. The main tasks of the nutritionists include administration of all processes related to food provided to children in centers, revision of menus, supervision of portions served, monitoring of special cases, anthropometric follow-up, training kitchen personnel and working with families on better nutrition and health-related habits. On the other hand, psychologists are supposed to support parents on all issues within the socio-emotional domain, handle domestic violence cases, diagnose and prepare plans for cases requiring follow-up, in particular, those in which developmental lags have been identified (Andrew et al., 2016).

Fourth, the child-teacher ratio is worse in CDIs than in HCBs (25:1 in CDI and 12:1 in HCB). The literature reveals that programs with class sizes of at most 20 children can be shown to produce large gains for disadvantaged children (Barnett & Boocock, 1998). More generally, a large body of literature points to the links between staff–child ratios, program quality, and child development. In particular, it has been shown that with a ratio of about 10 children per adult, children have greater opportunities for learning (Bowman, Donovan & Burns, 2001).

Fifth, both CDIs and HCBs provide approximately 70% of daily nutritional needs. In both cases, a morning and afternoon snack is provided, as well as lunch. However, servings might be different in CDIs with respect to HCBs due to the presence of a professional nutritionist in CDIs. There is no specific curricular guideline neither in CDIs nor in HCBs. The Board for Early Childhood has emphasized the principle of curricular freedom, and national standards are intentionally broad. That means that both, teachers in CDIs and MCs in HCBs, are expected to adapt the learning standards to their own classrooms and contexts (CIPI, 2013). Teacher training
and coaching strategies are not common in ECE programs in Colombia and vary significantly across providers for both CDIs and HCBs. Yoshikawa, Weiland and Brooks-Gunn (2016) presented recent evidence about the positive effects of the use of a structured preschool curriculum focused on a particular set of skills incorporating engaging, play-based activities on childcare quality and child development. Furthermore, the authors emphasized that a developmentally focused curricula, combined with intensive in-service training or coaching for teachers, has been shown to improve the quality of preschool instruction.

It is important to note that all public ECE services (including CDIs and HCBs) have comprehensive operational and technical guidelines (referents) that relate mostly to administrative guidelines and certain structural service parameters such as the number of children per square meter, characteristics of physical areas, teachers’ qualifications, food handling, bookkeeping, etc., but do not provide specific pedagogical directives.

The annual cost per child in a CDI is much higher than in a HCB, at 1,600 USD (excluding the cost of the infrastructure). Both programs are run by the ICBF. DCAS has enrolled approximately 235 thousand children younger than 6 years in CDIs since 2011, out of which at least half were children who were transferred out of HCBs into CDIs.

The theory of change behind the introduction of CDIs indicates that, on one hand, age-specific grouping of children, better infrastructure, more time available for teachers to spend in classroom activities, and staffing of centers with professionals could improve teaching practices and routines in the classroom, as well as improve parental practices at home through a more close relationship with the psychologist and nutritionist. If this were the case, the changes could in turn, improve cognitive and socio-emotional development of children. One could also expect the presence of a nutritionist and specialized kitchen personnel in centers to be related with improved
food handling and content of food servings, improved nutritional habits, and better monitoring of children’s nutritional status, and thus, improvements in children’s nutrition and health.

On the other hand, while more time is given to teachers in CDIs to spend with children (vs. other tasks such as food preparation), it is also possible that children in HCBs benefit from the smaller adult-to-child ratio. Also, CDI teachers might feel less accountable to parents than MCs because CDI teachers are part of a larger team, and might care for different children every year. In sum, there are both positive and negative potential effects of transitioning from HCBs to CDISs and hence, the overall magnitude and direction of the impact cannot be clearly determined \textit{ex ante}.

\textbf{On the Quality of Early Education Services}

In LMICs there have been few comprehensive educational interventions as the expansion of early childhood services occurred through less costly alternatives such as home-based childcare or conditional cash transfer programs (Nores & Barnett, 2010). As a result, there is very little empirical evidence about the effects of more comprehensive, potentially higher quality, but also more expensive early education programs. This study contributes to the understanding of the effects of center-based care as result of a specific offer to transfer out of home-based services and into childcare centers.

The evidence resulting from developed countries, especially the United States, based on small randomized control trials of intensive, high-quality early educational interventions finds positive effects on child development and suggests that the extent to which children benefit from center-based childcare can vary depending on its quality (Britto et al., 2011; NICHD, 2000; Cleveland & Krashinsky, 2013). Quality might have different dimensions, including structural and process quality. Structural quality refers to features such as class size, child–adult ratio, teacher
qualifications, and physical environment, while process quality is associated with features such as teacher–child interaction and the environment in which children learn (Yoshikawa et al., 2015).

There seems to be an agreement that measures of process quality are critical aspects that promote child development (NICHD, 2000; Yoshikawa et al., 2016). The evidence from developed countries suggests that childcare of good quality has positive strong effects on children’s development and health that are longlasting and sizable relative to its cost (Cunha et al., 2006; Camilli, Vargas-Baron, Ryan & Barnett, 2010). There is still little evidence on center-based childcare programs on a large scale. Engle et al. (2011) reviewed research on center-based early childcare in LMICs and concluded that most effects on children’s cognitive functioning, school readiness, and school performance were observed in cases of high-quality programs. Similarly, Britto et al. (2017) reviewed the evidence on the effects of early education programs in developing countries, and concluded that regardless of the type of program, quality is a key predictor of its effectiveness, in particular, factors such as positive interactions, individualized attention and positive emotional climate. This study contributes to this literature by comparing the impacts of two different programs on a large scale, which are different in terms of components of ECE quality.

As can be inferred from the earlier description of the intervention, most differences between CDIs and HCBs are related to features of structural quality: adult-child ratios, group sizes and their age composition, infrastructure and basic safety, and qualified personnel. Features of process quality such as a developmentally focused curriculum, teacher pedagogical and classroom organization skills, and warm and responsive environments (Yoshikawa et al., 2016) are not explicitly part of the transfer from HCB to CDI. Thus, we are careful in measuring process quality at endline and our statistical analysis allows for the possibility of positive or negative treatment effects on child development.
Methods

Study Design

We evaluate the impact of the offer to transfer from HCBs to CDIs using a cluster-randomized effectiveness trial in 14 cities in Colombia involving 15 CDIs. These cities include the entire pipeline of centers under construction at the time that the evaluation started. The initiative to build a center came from each of the municipalities, as the infrastructure was funded directly by the local government -while operating expenses of centers are funded at the national level-. The cities are medium to large, with reasonable institutional and financial capacity. In particular, seven of the 14 cities had between half a million and 1.2 million inhabitants at the time of the study, and the other seven cities had between 112 thousand and 270 thousand inhabitants.

The study took place between November 2010 and November 2012. For each CDI in the sample, we identified 28 HCBs located in the vicinity. Up to 5 HCBs of these were directly offered to transfer to the CDI because of extreme poverty conditions or close proximity. From the remaining HCBs, we randomly selected 20 to make part of the study. We then randomly assigned 15 of these to receive an offer to transfer to a CDI while the remaining five did not receive this offer. The former constitute our treatment group (intent-to-treat group), while the latter are our control group. Each cluster (one HCB) had an average of 9.2 children aged 6–60 months at baseline. We excluded children older than 60 months at baseline because they would have been served by the intervention for a very short period before leaving for elementary school. The ICBF informed HCB units whether they received an offer to transfer to the closest CDI or not. However, the transition was voluntary. Parents of beneficiary children were supposed to be encouraged by their MC to transfer and were not directly contacted by ICBF officers or the research team.
Baseline data collection occurred between November 2010 and May 2011. Follow-up data collection took place in two different stages to accommodate the budgetary constraints of the study funder. The first stage took place in November and December 2011 in six out of 15 CDIs in the sample. The second stage took place from September to November 2012 in the remaining nine CDIs. Different opening dates of CDIs and different timings of follow-up data collection implied three different groups of CDIs in terms of how long children could have been exposed to the treatment. In a first group of six CDIs, nine to ten months elapsed between the opening of centers and follow-up data collection. In a second group of six CDIs, this period was closer to 19 months. Finally, in a third group of three CDIs six to eight months elapsed. Differences in center opening dates occurred due to delays in their construction associated with “La Niña” rain season in 2011.

**Study sample**

This protocol implied a total sample of $N = 2,767$ children younger than 60 months of age and 300 HCBs assessed at baseline. To determine the sample size for the trial, we assumed an average intervention effect of 0.20 SD of a standardized score, based on Nores and Barnett (2010) who reported effect sizes of about 0.25 SD for continuous outcomes in cognition for early education programs in the region. We estimated the conditional intracluster correlation to be 0.0135, using height $Z$-scores based on data from Bernal and Fernández (2013). Finally, we assumed a 10% attrition rate between baseline and follow-up. Power calculations yielded a 90% power (with $\alpha = 0.05$) for a sample of nine children per HCB for a total of 300 HCBs in an array of 75% treated and 25% controls. In sum, 225 HCBs (2,067 children) were assigned to receive an offer to transfer to CDIs, and 75 HCBs (700 children) did not receive the offer to transfer (see Figure 1).

At follow-up, 2,455 children and their families were re-interviewed (89%). Using a linear probability model for the likelihood of attrition, we found that random assignment to treatment
ITT) does not predict attrition, but household poverty and parental absence do. We controlled for both in the econometric analyses. We also observed that treated and control attritors were not different at baseline except for household poverty with the former being poorer than the latter.

We observed significant non-compliance with the study’s protocol. Close to 22% of the children of HCBs that received an offer to transfer to a CDI were effectively enrolled in a CDI at follow-up, and 16% of children of HCBs that did not receive an offer to transfer were enrolled in a CDI at follow-up. The former fraction is statistically significantly higher than the latter. We found that non-compliance of treated children was higher for older children, children of mothers with lower education, and children in larger households. In particular, 39% of children younger than three assigned to the treatment were enrolled in CDI at follow-up while only 14% of children older than three were. More than half of non-complying children did not transfer to CDIs because they were too close to reaching the maximum age for participation. Also, 17% had transferred to pre-primary education in public schools by the time centers opened, and 16% reported that the center was too far. Albeit the fact that the transfer from HCB to CDI required a 1 kilometer proximity criterion, this distance was in most cases larger that the distance between households and HCBs which is, on average, less than a 10-minute walk (Bernal et al., 2009).

In terms of compliance of MCs, close to 75% of MCs assigned to treatment transitioned to CDI, while 60% of MCs assigned to control did not transfer. On one hand, by moving to centers, MCs were expected to change from working in their own home to working in the CDI center, from being their own boss to having an on-site boss, and from having flexibility in their daily activities to having specific guidelines about how their time should be allocated, including having a formal schedule and specific responsibilities that were closely monitored by center directors. On the other hand, the transfer might have had positive features such as being released from administrative
duties, and cooking and cleaning responsibilities. In the absence of a wage change associated to
the transfer from HCB to CDI, a fraction of MCs did not agree to the transfer, possibly because
the disadvantages weighed more than the advantages.

We found that 22% of non-complying MCs did not want to transfer, 11% were requested
not to transfer by parents, 27% reported personal problems, 16% mentioned health-related issues,
and 11% had not completed high-school. At follow-up, 89% of all teachers working in centers had
previously worked as MCs. We found that MCs who chose not to move to centers were, on
average, older, had more work experience and were less educated than both, MCs who chose to
transition to centers, and CDI teachers who had not worked as MCs. 22% of MCs that did not
transition to CDI centers had completed a degree in higher education, while close to 55% of both,
MCs in centers and new teachers, had.

In terms of enrollment in CDI, 26% of the children who were still eligible to be enrolled at
follow-up were indeed enrolled in CDI. Of those children not enrolled in CDI, 25% were enrolled
in an HCB, 15% in other public early education program, 41% attended pre-school in public or
private schools and 20% were not enrolled in any early education program.

**Data collection procedures**

Informed consent was obtained from all subjects in this study. Ethics Committees at
participating institutions approved the study’s protocol in 2010. Parent questionnaires on
household’s socio-demographic characteristics, children’s socio-emotional outcomes and
cognitive and motor outcomes (by the Ages & Stages questionnaire) were collected by parental
report at HCBs at baseline and HCBs/CDIs at follow-up. In case children did no longer attend
early education services at follow-up these were collected at the child’s household. Data on child’s
enrollment status at follow-up were collected from administrative records and parental report.
Children were assessed by Woodcock-Muñoz III in HCBs or CDIs and mothers were invited to be present. Tester held degrees in psychology and had four weeks’ training, including practice sessions with children of the target age groups. The inter-rater reliability (intracluster correlation) was above 0.9 on each subscale of the Woodcock-Muñoz III. When it was not possible to assess children in service units, they were assessed at their own homes trying to guarantee similar testing conditions. Classroom quality measures by ECERS-R, ITERS-R and FDCRS (details below) were completed by psychologists who observed CDI classrooms and HCBs for at least half of a school day. Assessors were trained to be unobtrusive, to the extent possible, in the classroom or HCB while observing and recording. The team of six psychologists was trained for three weeks and assessed for reliability in two centers and 10 HCBs in Bogotá, which were not in the study sample. The inter-reliability was above 0.8 on all subscales of the three instruments.

**Measures**

**Household demographic and socio-economic status** data were collected at baseline and follow-up. The survey included information on household characteristics: basic utilities and characteristics of the home; income; durable goods ownership; parental education; presence of the father; employment status; marital status; household size and composition, and parental practices. Using principal-component analysis, we constructed a wealth index on a subset of household characteristics. We then used this index to classify households into wealth quintiles.

**Nutrition.** We collected information on height, weight and arm circumference at baseline and follow-up following World Health Organization (WHO) standards. Based on these measures, we constructed several nutritional indicators (WHO Multicentre Growth Reference Study Group, 2006/7). No reference for mid-upper-arm circumference for 5-19 year olds was included in the
WHO 2007 revision. For this reason, this variable was used only for a subsample of children younger than five at follow-up or not used at all.

**Health.** We measured health by: (1) blood haemoglobin by a finger prick assay with HemoCue Hb 201+ devices for all children at follow-up and we construct the fraction of anemic children as the percentage of children under age 5 with haemoglobin levels below 100 g/l and below 110 g/l for children older than 5; and (2) parasitic examination in stool sample, for a random subsample of 75% of all children at follow-up only due to budgetary constraints.

**Cognitive development.** (1) We used the Ages & Stages Questionnaire (ASQ), third edition, for all children at baseline and follow-up. The ASQ is a questionnaire focusing on cognitive development and the identification of children at risk of cognitive developmental difficulties. It includes fine and gross motor, communication, and problem solving. The ASQ has been used for early development assessments in LMICs (Rubio-Codina, Araujo, Attanasio, Muñoz, & Grantham-McGregor, 2015). The Cronbach’s $\alpha$ is 0.87 for total ASQ scores (by subscale, the Cronbach’s $\alpha$ was 0.61 for communication, 0.67 for problem resolution, 0.76 for gross motor and 0.70 for fine motor). We report raw scores in the descriptive statistics and use nonparametrically age-standardized scores in the statistical analysis. It is worth noting that the ASQ is generally regarded as a screener test and not considered a gold-standard measure of child development. For this reason, we are careful in the interpretation of these results. (2) The Woodcock–Muñoz III Tests of Achievement (WM-III) for a random subsample of 75% children older than 30 months of age at follow-up. The WM-III is a set of individually administered tests of children’s early literacy and mathematical skills and knowledge, and we use subtests 1, 5, 6, 7, 10, and 14, to measure general verbal ability and receptive language, associative memory, executive functioning, and mathematical reasoning (Muñoz-Sandoval, Woodock, McGrew,
Mather & Schrank, 2005). The WM-III has been used to evaluate effects of early childhood interventions in Latin-American contexts (Fernald, Kariger, Engle, & Raikes, 2009). We report age-standardized scores using the Compuscore software. We computed composites directly in the field and do not have individual item scores available. However, we found that all subcales positively and significantly correlate with maternal education and household wealth.

**Socio-emotional development.** We used the ASQ for the socio-emotional domain (ASQ:SE) (Squires, Bricker & Twombly, 2009) for all children at baseline and follow-up. The ASQ:SE focuses on socio-emotional development and the identification of children at risk of social-emotional difficulties. It includes self-regulation, compliance, communication, adaptive functioning, autonomy, affect, and interaction. The ASQ:SE has been used for early development assessments in LMICs (Handal, Lozoff, Breilh & Harlow, 2007; Heo, Squires & Yovanoff, 2007). The Cronbach’s α for total ASQ:SE scores is 0.70 but there is significant variation by subscale (self-regulation=0.50, compliance=0.49, communication=0.34, adaptive function=0.21, autonomy=0.13, affect=0.25 and interaction=0.36). We report raw scores in the descriptive statistics and use nonparametrically age-standardized scores in the statistical analysis. Again, we note that ASQ:SE is generally regarded as a screener test and not considered a gold-standard measure of child development.

**Class environment and activities.** We used three of the Environment Rating Scales (ERS) to measure classroom quality for the different childcare programs assessed in this study. The ERS scales provide a global measure of preschool classroom quality with complimentary scales assessing different segments of the childcare spectrum. In particular, the Early Childhood Environmental Rating Scale - Revised (ECERS-R; Harms, Clifford & Cryer, 1998) is designed for preschool aged children in center-based care; the Infants and Toddlers Environmental Rating Scale
- Revised (ITERS-R; Harms, Cryer & Clifford, 2003) for infants and toddlers in center based care; and the Family Daycare Rating Scale (FDCRS; Harms & Clifford, 1989) provides a global measure of environmental quality for children in home-based childcare settings such as HCBs. The instruments cover a broad range of quality considerations from safety to teacher-child interaction to parent involvement based on checklists that range from 39 to 43 items.

These measures have been used extensively in the field and have well-established validity and reliability. ERS scales have been used in a wide range of countries with different cultures and economic contexts. The ERS scales have shown predictive validity to child gains across cognitive domains (Burchinal et al., 2000) and social-emotional domains (Sylva et al., 2006).

ECERS-R, ITERS-R and FDCRS were completed by psychologists who observed CDI classrooms and HCBs for at least half of a school day. Each assessor rates each of 43 items in ECERS, 39 in ITERS and 40 in FDCRS. All items are presented as a 7-point scale with descriptions of what is required under four quality levels: 1 (inadequate), 3 (minimal), 5 (good) and 7 (excellent). Each description consists of one or more numbered indicators that must be scored by observation and evaluation of a classroom. The total number of items in each scale is organized into seven subscales that provide a practical and conceptual organization for the items. The following are the seven elements of preschool quality considered by ERS subscales: physical environment; basic care; pedagogical activities and curriculum; interaction; schedule and program structure; and parent and staff education. Most of the items examine the quality of what children actually experience in the program. A small subset of items looks into the quality of provisions for the adults involved in the program, i.e., parents and staff.

We collected ECERS for 119 classrooms in all CDIs at follow-up, ITERS for 37 classrooms in all CDIs at follow-up, and FDCRS for 54 HCBs still running at follow-up. These
instruments were not collected at baseline. The ECERS, ITERS and FDCRS capture the same constructs to the extent that is possible. However, it is clear that they measure specific features in each type of setting. No studies are available that compare the three scales. According to the test developers, scores from these tests are often treated as if they are comparable, particularly in the U.S. literature.

The ERS measures have been widely used in countries in the region similar to Colombia, such as Ecuador, Peru, Bolivia, Chile, and Brazil (Araujo et al., 2015; Berlinksy & Schady, 2015; Verdisco & Pérez Alfaro, 2010). In this study, ERS scores are positively correlated with the care provider’s years of experience and years of schooling, and negatively correlated with the classroom’s adult–child ratio.

**Statistical Analysis**

The hypothesis that we tested was whether the offer to transfer from HCB to CDI had a positive or negative average impact on child development and nutrition. Treatment was randomly assigned given the cluster-randomized effectiveness trial design. We estimate intent-to-treat effects (ITT) by Ordinary Least Squares using the following specification:

\[
A_{i,h,t} = \lambda_1 + \lambda_2 ITT_{i,h} + \lambda_3 A_{i,h,t-1} + \lambda_4 X_{i,h,t-1} + \epsilon_{i,h,t}
\]  

(1)

$ITT_{i,h}$ is a binary variable that equals 1 if child $i$ from HCB $h$ was randomly assigned to treatment, i.e., offered to transfer from an HCB to a center, and 0 otherwise. $A_{i,h,t}$ is an outcome variable for child $i$ from HCB $h$ in period $t$, $A_{i,h,t-1}$ is the same outcome for child $i$ in HCB $h$ but measured at baseline (or one in the same developmental domain if it was not measured at baseline), and $X_{i,h,t-1}$ is a vector of baseline control variables that includes child’s gender, age and age squared in months, birth order, maternal years of schooling, indicators for whether the father and the mother are
present, female-headed household, employment status of the head of household, household size, household wealth quintile, number of children in the household, the mean of the outcome variable in the HCB at baseline, years of schooling and working experience of the MC in the HCB at baseline, and city fixed effects. $\epsilon_{i,h,t}$ reflect unobserved factors that influence child outcomes.

$\lambda_2$ captures the impact of ITT on the outcome, that is, the causal effect of having been offered the chance to move from HCB to CDI, regardless of whether individuals complied with the offer. To assess how the offer to transfer affected parental childcare choices we also estimated equation (1) using as dependent variable the type of early education program used at follow-up.

We used two-tailed tests in our statistical analysis and cluster standard errors at the HCB level to account for the clustering of children by HCB at baseline. We also adjusted the $p$-values of estimated effects to correct for multiple hypotheses testing, using the step-down procedure proposed by Romano and Wolf (2005) – RW. In particular, the RW-adjusted $p$-values take into account the fact that within each developmental domain we tested multiple hypotheses by using several outcome measures.

We also estimated treatment-on-treated (TOT) effects. That is, ITT effects adjusted by the take-up rate, which take into account the fact that not all children took up the offer to transfer from HCB to CDI. We estimated TOT effects using an instrumental variable approach. In particular, we used random assignment (ITT) as an instrument for program compliance. Compliance is defined as a binary variable that equals 1 if child $i$ from HCB $h$ is enrolled in CDI at follow-up (i.e. registered in CDI rosters, not actual daily attendance), and 0 otherwise. Enrollment is directly obtained from CDI administrative records. This procedure provides an unbiased estimate of the program’s effects among those who actually made the transition from HCB to CDI. The ITT indicator is a valid instrumental variable as it was randomly assigned and it significantly explains
CDI enrollment. A regression of enrollment in CDI on random assignment yields a statistically significant positive coefficient with an F-test of 9.96. Unfortunately we do not have information about daily attendance and can only measure enrollment.

The main limitation of this study is that compliance with the random assignment to transfer from HCB to CDI was imperfect. Although this issue does not invalidate the experimental design, as attendance to a CDI is significantly more likely for the treated than the control group, it might imply that the study could be underpowered to detect statistically significant program effects. In addition, the requirement to adjust $p$-values to take into account of multiple hypotheses testing exacerbates the problem of insufficient power, especially given our rich list of developmental outcomes. Although there is no perfect solution for this issue, we tried to alleviate this concern by aggregating outcomes into a few developmental domains rather than using all developmental measures individually. By doing this, we could estimate the impacts more precisely.

To aggregate the outcomes, we conducted an exploratory factor analysis on all developmental outcomes available at follow-up, and constructed a dedicated measurement system. In particular, we included four ASQ subscales (communication, problem resolution, gross and fine motor), five WM-III subscales (general verbal ability, associative memory, executive function, mathematical reasoning and receptive language), seven ASQ:SE subscales (compliance, communication, adaptive functioning, autonomy, affect and interaction), two anthropometric measures (heigh and weight for age) and two health measures (anemia and parasites). We excluded from this exploratory analysis the arm-circumference because WHO standards are not available for children older than five years of age. The dedicated measurement system required that (i) factors selected had eigenvalues higher than one, (ii) items that heavily loaded in more than one factor were excluded, and (iii) items with factor loadings lower than 0.2 were also excluded.
This analysis yielded the following results. First, the exploratory analysis identified four main developmental factors: cognitive development by ASQ, cognitive development by WM, socio emotional development by ASQ:SE (excluding autonomy which did not load into any factor) and a nutrition / health factor which included height-for-age and weight-for-age. Anemia and parasites did not load into any factor so these were excluded from the construction of developmental factors. Given that overweight and obesity are not very prevalent in this sample (see descriptive statistics below) we assumed that a higher nutrition factor was always better. We estimated equation (1) using these four factors as dependent variables. However, we also report impacts by individual developmental outcomes as a robustness check.

We also investigated the possibility of heterogeneous impacts by key baseline characteristics. In particular, we looked into differential effects by child’s gender and age, and potential duration of exposure to the program. The latter is proxied by the amount of rainfall in a given city during the time elapsed between baseline and center opening as construction delays occurred due to a climatic phenomenon known as *La Niña*. Thus, while duration of exposure to the program might be endogenous (and we do not have data on actualy daily attendance), the amount of rainfall is plausibly exogenous to unobserved characteristics of children.

Finally, we also explored the extent to which these impacts might be explained by differences in the quality of both services during the transition. To this aim, we present differences in classroom quality between children served in CDIs and children served in HCBs at follow-up using the ERS scales.

**Results**

In this section, we present descriptive statistics on the sample, and estimates of the effects of the offer to transfer from HCB to CDI on child development and health, as well as, the effects
of the offer on the type of early education program children attended and the differences in quality of care provided to them at follow-up.

**Descriptive Statistics**

Table 1 provides summary statistics at baseline by treatment group. We included sociodemographic characteristics of children, as well as developmental outcomes. Children were, on average, 44 months of age, with about 30% of the households being single-headed, and 1.5 children younger than five per household. Maternal average schooling was 10 years, and 38% of children did not live with their fathers. Average height-for-age and weight-for-age are 1.5 and 0.5 SD below the mean of the reference population, respectively. Nutritional status in our sample is worse than that of comparable children in Colombia’s national longitudinal survey (ELCA, 2013). In particular, stunting is about 17% in our sample compared to 13% for the lowest urban SES in Colombia and 10% nationwide.

We did not collect WM-III, haemoglobin or stool sample at baseline so these are not reported in Table 1. At follow-up, 28% of children were anemic and 53% had parasites. In sum, children in this sample had a high developmental risk. Table 1 also provides confirmation of baseline balance between study groups. Overall, there are no statistically significant differences between ITT groups at baseline, and the differences observed are not systematically in favor of one of the two groups. Baseline characteristics do not significantly predict ITT.

**Results on the choice of early education service**

In Table 2, we present the ITT impacts of the offer to transfer from HCB to CDI on the ECE program that the child was enrolled in at follow-up. The dependent variable was defined as a categorical variable indicating if child $i$ was enrolled in CDI, HCB, or other early education
program at follow-up, or was not enrolled in any early education program, based on parental reports. We estimated it by using a multinomial regression model.

The results indicate that being offered the chance to transfer from HCB to CDI did significantly increase CDI enrollment at follow-up by 13 percentage points and decreased HCB enrollment at follow-up by 11 percentage points. The offer did not have any statistically significant effects on attendance to other programs or not using childcare. These results indicate that while the study might be underpowered due to noncompliance, the offer to transfer from one service to another did have the expected effects on the choice of childcare program.

**Results on child development and health**

Table 3 reports the ITT impacts on the four developmental outcomes described earlier: cognitive development by ASQ, cognitive development by WM, socio-emotional development, and nutrition (left panel), as well as TOT estimates on these four outcomes (right panel). The results indicate that the offer to transfer from HCB to CDI had a negative and statistically significant effect on cognitive development by ASQ of 0.11 standard deviations (SD) at 95% confidence level. At the same time, the offer had a positive effect on nutrition of 0.05 SD at 95% confidence level. Similar results can be observed for TOT effects reported in the right panel of Table 1. There is a negative effect of the treatment-on-the-treated on cognitive development by ASQ of 1.5 SD and a positive effect of 0.6 SD on nutrition. Both are statistically significant at 90% confidence. The effect on the WM cognitive factor exhibits a negative sign but is not statistically significant. Similarly, the effect of the offer and the effect of the TOT on socioemotional development is positive (because it decreases behavioral problems) but is not statistically significant.

In Appendix 1, we report the same results for individual developmental measures instead of factors. As can be observed, we report only a statistically significant increase in the probability
of wasting at 95% confidence level. All other ITT and TOT effects are statistically insignificant. The direction of the effects coincide for most part with those presented in Table 3. However, due to the low power of the study and the proliferation of measures, which required adjustment for multiple hypotheses testing, these effects are imprecisely estimated.

In sum, the results indicate positive program effects on nutrition and suggest negative effects on cognitive development. The effect on nutrition is small representing about 5% of a standard deviation. This result might be due to the presence of a professional nutritionist in CDIs who is in charge of overseeing all processes related to food provision in centers (purchases, reviews of menus, revision of portions), as well as anthropometric follow-up and monitoring of children at nutritional risk, and involvement with families to improve nutrition and health habits. While both programs provided the same fraction of daily nutritional requirements, it is possible that contents were better in CDI than in HCB, and that there was better monitoring of the process of ensuring food quality and overseeing feeding practices in the centers. Unfortunately we do not have data on these processes to corroborate these hypotheses.

On the other hand, the results suggest negative effects on cognitive development. We are cautious in interpreting this finding as the ASQ is generally regarded as a screener test and not considered a gold-standard measure of child development. The cognitive factors included dimensions such as language, problem resolution, mathematical reasoning and memory. These abilities are probably more closely associated with teacher–child interaction and the learning environment in the classroom than nutrition is. The negative or null effects on cognition, might be indicative of the fact that the transition from CDI to HCB did not explicitly change process quality features such as teacher pedagogical and classroom organization skills, a developmentally focused curriculum, and a language-rich and responsive environment (Yoshikawa et al., 2016). In other
words, the mere transfer to a different physical space, the age-grouping of children and the presence of more qualified staff was not enough to guarantee significant changes in the learning environment at CDIs with respect to HCBs. We explored this hypothesis further by analyzing data on classroom quality (see below).

We also looked into the effects for specific groups of children. In particular, we analyzed heterogeneous effects on developmental factors by child’s age and sex, and the amount of rainfall in the municipality of residence. These results are not shown but are available upon request. We did not find any differences by child’s age (neither by ITT nor by TOT). We observed a negative ITT effect on cognitive development by ASQ that was higher for boys (-0.13 SD) than for girls (-0.08 SD) but the difference was not statistically significant. No statistically significant differences were observed by TOT.

Children in this study could have been exposed to the intervention for different lengths of time mostly due to delays in the construction of the centers associated with the La Niña climatic phenomenon. We used the amount of rainfall in the child’s town of residence between baseline and the opening of the center as a proxy of the duration of exposure to the intervention. We did not observe any differences in program impacts depending on how long children could have been exposed to the intervention. This is important because, being a new initiative, CDIs might have required some time to optimize procedures. It is possible that the transition period required to achieve an optimized model was longer than the duration of the evaluation study.

**Differences in childcare quality by program**

In Table 4, we present differences in classroom quality as measured in 113 CDI classrooms of children older than two years of age by ECERS, 36 CDI classrooms of infants and toddlers by ITERS and 54 HCBs by FDCRS, at follow-up. In panel I, we show total scores computed by
averaging over all items (Cronbach’s $\alpha = 0.76$ for ECERS, 0.52 for ITERS and 0.81 for FDCRS), as well as total scores split in two categories: the space & furnishings subscale (Cronbach’s $\alpha = 0.40$ for FDCRS but undetermined for ECERS and ITERS due to a negative covariance among items) and all the other subscales excluding space & furnishings (Cronbach’s $\alpha = 0.76$ for ECERS, 0.52 for ITERS and 0.79 for FDCRS). The Space & Furnishings subscale includes items that characterize the indoor space, the furniture, the room arrangement, the space and equipment for motor gross play, and the space available for child privacy (i.e., items related to the infrastructure). The rest of the subscales are more closely related to what children actually experience in the program. In panel II, we show scores for all the seven subscales available in the ERS scales. In particular, we included the personal care routines subscale (Cronbach’s $\alpha$ lower than 0.2 in all cases), language and reasoning (Cronbach’s $\alpha$ between 0.4 and 0.55), activities (Cronbach’s $\alpha$ is 0.33 for ECERS/ITERS and 0.75 for FDCRS), interaction (Cronbach’s $\alpha = 0.74$ for ECERS, 0.44 for ITERS and 0.24 for FDCRS), program structure (Cronbach’s $\alpha$ lower than 0.4 for ECERS/ITERS - the scale is not available in FDCRS), and parents and staff (Cronbach’s $\alpha = 0.63$ for ECERS, 0.46 for ITERS and 0.53 for FDCRS).

It is important to note that the ERS scales are not generally thought to be accurate measures of relevant dimensions of process quality such as positive interactions, individualized attention and positive emotional climate, and the evidence is mixed regarding their power to predict child development (Mashburn et al., 2008; Burchinal et al., 2000; Sylva et al., 2006). The internal reliability measures reported here are not very high either. As it is uncommon to have this type of data available in evaluations of large scale programs, we chose to present these results as suggestive evidence that can explain the program impacts reported above, however we acknowledge the potential limitations of the quality indicators presented.
Overall and perhaps surprisingly, quality scores were significantly higher in HCB (total score=2.19) than in CDI classrooms (total score= 1.75). Both had inadequate quality levels by this measure. The computation of the t-test for the difference between CDI and HCB adjusted for CDI clusters. When we analyzed the Space & Furnishings subscale separately from the rest, we found that structural quality (as measured by these items) is significantly higher in CDI than in HCB, as expected. The difference is statistically significant at 95% confidence level. However, the average of all items excluding the Space & Furnishing subscale exhibited significantly higher quality in HCB than in CDI. The difference of about 0.56 scale-points (or 1.3 SD of FDCRS) is large. When looking into individual subscales, we observed statistically significant differences in favor of HCB in all but: space & furnishings and interactions. Other than that, HCBs scored significantly better in personal care routines, language and reasoning, activities, and parents and staff.

One important issue to bear in mind when interpreting these results, is the fact that CDIs were newly developed centers at the time of the evaluation study. As with any new program, the optimization of procedures can take time. It is possible that these results reflect a transition period and that quality might have improved as the program established. However, the results still suggest that during this transition period, the differences in quality, and particularly, differences in quality aspects different from infrastructure, might have had a role in explaining the negative and null impacts on children’s cognitive and socio emotional development, respectively.

**Discussion and Conclusion**

To evaluate the impacts of the offer to transfer from small community nurseries to large centers on childcare quality and child development and nutrition, we randomly assigned HCBs to receive an offer to transfer to CDIs in 14 cities in Colombia. Despite of imperfect compliance with the random assignment, the results suggest negative and significant effects of both the offer to
transfer from HCB to CDI and the TOT on children’s cognitive development, as well as positive program ITT and TOT effects on nutrition. No significant effects were found on socioemotional development.

The significant improvement in nutrition might be associated with the presence of a professional nutritionist in CDIs, who is in charge of all processes related to food provided to children in centers. On the other hand, the abilities considered in the cognitive factors such as language, mathematical reasoning and problem resolution are plausibly more closely associated with teacher–child interactions and the learning environment in the classroom than nutritional status. As changes associated with the transfer from HCB to CDI related more explicitly to physical infrastructure, age-specific classrooms, and more qualified staff, it was not clear that process quality (e.g., teacher pedagogical strategies and responsive learning environments) improved. Comparisons of classroom quality by ERS scales provided some evidence in favor of this hypothesis, as these showed that infrastructure is, in fact, better in CDIs than in HCBs, while all other aspects more closely related to what children actually experience in the program exhibited higher scores in HCBs than in CDIs. It is important to mention that CDIs did have comprehensive operational and technical guidelines (or referents). The results point to the fact that those guidelines did not provide pedagogical or operational directives of enough specificity to drive process quality improvements as intended.

In understanding these results, it is important to keep in mind a few facts. First, the ERS scale might not be the best way to measure relevant dimensions of process quality such as positive interactions, individualized attention and positive emotional climate. Second, as CDIs were newly developed centers at the time of the evaluation study and new programs might take time to achieve optimal procedures, we might be measuring both, program effects and program quality during the
transition period. In this sense, it is crucial to keep monitoring the quality of CDIs over time, to explore if process quality improves after the new program has had some time to get established. Finally, this study did not address the predictive power of ERS scales on child development and the literature is inconclusive in this sense. It is clear, then, that further research is needed to assess more systematically the effects of different elements of ECE quality on child development.

This study faced other important limitations. First, it suffered from significant compliance issues, which reduced the statistical power of the study. This was related to the fact that the transition from HCB to CDI was voluntary and many MCs and families did not take up this offer. The issue of low compliance to CDI enrollment is an important finding in and of itself. It poses the question about how to more systematically assess these barriers to program access, understand how these affect attendance and adherence, and design strategies in order to overcome those barriers. For example, 16% of the non-compliance of treated children reported that the CDI was too far, even though CDI centers were, at most, within one kilometer from HCBs. It is important that policy design take distance and other convenience features into consideration as these could potentially impact not only enrollment but also attendance to the program.

To address the loss of power associated with noncompliance and the strict statistical power requirements implied multiple hypotheses testing on a rich list of developmental outcomes, we aggregated measures into four developmental factors to estimate the main impacts. Results are consistent with the analyses of individual measures, but do provide more precision.

Second, it is also possible that developmental outcomes might have been collected with measurement error, particularly those that are reported by parents (such as health, ASQ and ASQ:SE), and that these specific measures lack enough sensitivity due to the fact that these are commonly used as a screener and not as a gold-standard measure of child development. However,
these measures have been complemented with more reliable direct measures such as the Woodcock-Muñoz and anthropometric measures. Third, we did not collect daily attendance data, which prevents us from studying more rigourously the effects of actual duration of exposure to the program. We have tried to overcome this issue by presenting heterogeneous effects by rainfall which proxies for delays in center opening.

The results of this study show that although modernization of early childhood services is desirable, it is critical that the transition is done gradually, and that quality in some specific dimensions, related to processes, is specifically targeted and monitored. The findings in this study and the comparative costs of the new centers and the existing HCBs suggest that the strategy should be thought through carefully. Alternatives might have to be tested, including improvements to HCBs and training of MCs (Bernal, 2015). Different pedagogical models should also be assessed and adult-child ratios carefully considered. Our conclusions point to the fact that the offer to transition to center-based care, with the specific characteristics with which it occurred in Colombia, did not have positive impacts on children. However, these results do not necessarily imply that center-based childcare cannot have positive effects, if implemented with the characteristics required to guarantee key elements of structural and process quality (Britto et al., 2011).

Some examples include the results presented by Nores, Bernal and Barnett (2016) who assess a high-quality childcare center-based service in Colombia. The aeioTu program is characterized by high quality components such as high qualifications requirements for staff, pre- and in-service training, strong support staff and services, child monitoring and information system, family and community participation, and the use of a structured developmentally-oriented curriculum. The authors reported strong effects on language and cognitive after only 8 months of
program attendance. Similarly, Attanasio, Baker-Henningham, Bernal, Meghir, and Rubio-Codina (2017) reported positive effects on children’s cognition and language as a result of the implementation of a structured early stimulation curriculum in addition to training and coaching of paraprofessional personnel in a parenting program in rural areas in Colombia. Both are public programs targeted to poor households.

Thus, it is important that in implementing strategies to improve child care services for vulnerable children at scale, key factors such as the provision of teacher pre-service and in-service training, children assessment and monitoring, and strong curricular background are kept as a priority.
References


Encuesta Longitudinal Colombiana de la Universidad de los Andes (ELCA, 2013).  
https://encuestalongitudinal.uniandes.edu.co/


Figure 1. Study Design

Source: Consort Flow Chart. Own elaboration.
Table 1. Descriptive statistics at baseline

<table>
<thead>
<tr>
<th>Sociodemographic</th>
<th>$ITT = 0$</th>
<th>$ITT = 1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Child's age in months</td>
<td>44.15 (10.32)</td>
<td>44.16 (10.6)</td>
</tr>
<tr>
<td>Child's gender (male)</td>
<td>0.51 (0.50)</td>
<td>0.51 (0.50)</td>
</tr>
<tr>
<td>Maternal yrs of schooling</td>
<td>10.08 (3.11)</td>
<td>9.90 (3.07)</td>
</tr>
<tr>
<td>Father present</td>
<td>0.62 (0.49)</td>
<td>0.64 (0.48)</td>
</tr>
<tr>
<td>Mother present</td>
<td>0.95 (0.22)</td>
<td>0.94 (0.24)</td>
</tr>
<tr>
<td>Female head of household</td>
<td>0.30 (0.46)</td>
<td>0.29 (0.45)</td>
</tr>
<tr>
<td>Household head working</td>
<td>0.88 (0.33)</td>
<td>0.86 (0.34)</td>
</tr>
<tr>
<td>Household size</td>
<td>5.19 (2.14)</td>
<td>5.23 (2.18)</td>
</tr>
<tr>
<td>0-5 children at home</td>
<td>1.52 (0.70)</td>
<td>1.54 (0.72)</td>
</tr>
<tr>
<td>Wealth quintile 1</td>
<td>0.18 (0.39)</td>
<td>0.21 (0.40)</td>
</tr>
<tr>
<td>Wealth quintile 5</td>
<td>0.20 (0.40)</td>
<td>0.19 (0.39)</td>
</tr>
<tr>
<td>No. of observations</td>
<td>700</td>
<td>2,067</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Outcomes</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Cognitive development</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASQ communication</td>
<td>42.5 (21.8)</td>
<td>42.8 (21.7)</td>
</tr>
<tr>
<td>ASQ problems resolution</td>
<td>41.5 (20.2)</td>
<td>41.8 (20.2)</td>
</tr>
<tr>
<td>ASQ gross motor</td>
<td>40.9 (20.7)</td>
<td>40.5 (20.3)</td>
</tr>
<tr>
<td>ASQ fine motor</td>
<td>41.2 (20.0)</td>
<td>41.2 (18.6)</td>
</tr>
<tr>
<td>No. of observations</td>
<td>700</td>
<td>2,066</td>
</tr>
</tbody>
</table>

| Socio-emotional development           |                 |                 |
| ASQ:SE total                          | 68.9 (26.3)     | 67.6 (26.7)     |
| No. of observations                   | 700             | 2,066           |

| Nutrition                             |                 |                 |
| Weight for age Z-score                | -0.47 (0.93)    | -0.48 (0.95)    |
| Length for age Z-score                | -1.02 (1.02)    | -1.06 (1.00)    |
| Weight for length Z-score             | 0.20 (0.92)     | 0.20 (0.92)     |
| Arm circumference-for-age Z-score     | 0.10 (0.84)     | 0.08 (0.81)     |
| Underweight                           | 0.04 (0.20)     | 0.04 (0.21)     |
| Wasting                               | 0.00 (0.07)     | 0.01 (0.09)     |
| Stunting                              | 0.17 (0.37)     | 0.16 (0.37)     |
| Overweight                            | 0.02 (0.14)     | 0.03 (0.16)     |
| Obesity                               | 0.01 (0.11)     | 0.01 (0.07)     |
| No. of observations                   | 605             | 1,700           |

Note: ***$p<0.01$; **$p<0.05$; *$p<0.1$. Cluster-robust standard errors.
Table 2. Intent-to-treat effect on type of childcare enrollment

<table>
<thead>
<tr>
<th>Enrollment by type of program at follow-up</th>
<th>ITT estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>CDI</td>
<td>0.132***</td>
</tr>
<tr>
<td></td>
<td>(0.04)</td>
</tr>
<tr>
<td>HCB</td>
<td>-0.114***</td>
</tr>
<tr>
<td></td>
<td>(0.04)</td>
</tr>
<tr>
<td>Other early education service</td>
<td>-0.027</td>
</tr>
<tr>
<td></td>
<td>(0.04)</td>
</tr>
<tr>
<td>Not enrolled</td>
<td>0.010</td>
</tr>
<tr>
<td></td>
<td>(0.02)</td>
</tr>
<tr>
<td>No. of observations</td>
<td>2,455</td>
</tr>
<tr>
<td>Wald chi2 (105)</td>
<td>1.258</td>
</tr>
<tr>
<td>Prob &gt; chi2</td>
<td>0.0000</td>
</tr>
<tr>
<td>Pseudo R2</td>
<td>0.3345</td>
</tr>
</tbody>
</table>

Note: ***p<0.01; **p<0.05; *p<0.1. Model estimated by multinomial logit. Marginal effects reported. Cluster robust standard errors in parentheses. The specification includes the same covariates as in equation (1).

Table 3. Intent-to-treat and treatment-on-treated effects on children's developmental factors

<table>
<thead>
<tr>
<th>Outcomes</th>
<th>ITT estimates(^a)</th>
<th>TOT estimates(^b)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Effect</td>
</tr>
<tr>
<td>ASQ cognitive factor</td>
<td>2,435</td>
<td>-0.11</td>
</tr>
<tr>
<td>WM cognitive factor</td>
<td>1,905</td>
<td>-0.01</td>
</tr>
<tr>
<td>ASQ socio-emotional factor</td>
<td>2,429</td>
<td>-0.02</td>
</tr>
<tr>
<td>Nutrition / health factor</td>
<td>2,159</td>
<td>0.05</td>
</tr>
</tbody>
</table>

Note: ***p<0.01; **p<0.05; *p<0.1. Cluster robust standard errors in parentheses. The covariates included are listed below equation (1).
\(^a\) OLS estimates.
\(^b\) Two-stage least squares estimator; uses random assignment as an instrument to predict program compliance.
<table>
<thead>
<tr>
<th>Quality scores</th>
<th>CDI ECERS</th>
<th>CDI ITERS</th>
<th>CDI TOTAL (A)</th>
<th>HCBs FDCRS</th>
<th>HCBs DIFF CDI vs HCB (A) - (B)</th>
<th>Diff p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>I. Total scale score</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Space &amp; Furnishings subscale</td>
<td>1.64(0.53)</td>
<td>1.89(0.29)</td>
<td>1.70(0.51)</td>
<td>1.52(0.45)</td>
<td>0.18(0.45)</td>
<td>0.011***</td>
</tr>
<tr>
<td>Score excl. Space &amp; Furnishings</td>
<td>1.74(0.59)</td>
<td>1.84(0.28)</td>
<td>1.76(0.61)</td>
<td>2.32(0.46)</td>
<td>0.56(0.08)</td>
<td>0.000***</td>
</tr>
<tr>
<td><strong>II. Subscales</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Space and Furnishings</td>
<td>1.64(0.53)</td>
<td>1.89(0.29)</td>
<td>1.70(0.51)</td>
<td>1.52(0.45)</td>
<td>0.18(0.45)</td>
<td>0.011***</td>
</tr>
<tr>
<td>Personal Care Routines</td>
<td>1.38(0.56)</td>
<td>1.21(0.20)</td>
<td>1.34(0.52)</td>
<td>1.67(0.36)</td>
<td>-0.33(0.06)</td>
<td>0.000***</td>
</tr>
<tr>
<td>Language and Reasoning</td>
<td>1.70(0.82)</td>
<td>1.85(0.63)</td>
<td>1.74(0.89)</td>
<td>2.17(0.70)</td>
<td>-0.43(0.12)</td>
<td>0.000***</td>
</tr>
<tr>
<td>Activities</td>
<td>1.49(0.51)</td>
<td>1.47(0.35)</td>
<td>1.48(0.52)</td>
<td>2.27(0.62)</td>
<td>-0.79(0.09)</td>
<td>0.000***</td>
</tr>
<tr>
<td>Interaction</td>
<td>2.07(1.05)</td>
<td>2.17(0.71)</td>
<td>2.09(1.13)</td>
<td>2.11(0.67)</td>
<td>-0.01(0.13)</td>
<td>0.916</td>
</tr>
<tr>
<td>Program Structure</td>
<td>1.59(0.92)</td>
<td>2.07(0.60)</td>
<td>1.70(1.00)</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Parents and Staff</td>
<td>2.20(1.63)</td>
<td>2.27(0.94)</td>
<td>2.21(1.80)</td>
<td>3.39(1.04)</td>
<td>-1.18(0.20)</td>
<td>0.000***</td>
</tr>
</tbody>
</table>

Note: ***p<0.01; **p<0.05; *p<0.1. Standard deviations of means and standard errors of differences in parentheses. The t-test for the differences between CDI and HCB adjusted for CDI clusters. Standard deviations of means in columns (2) and (3) adjusted for CDI clusters.

ECERS: Early Childhood Environmental Rating Scale (Harms, Clifford, and Cryer, 1998)
FDCRS: Family Day Care Rating Scale (Harms and Clifford, 1989)
For total scores or subscales, the score was calculated as an average rating of all items.
### Appendix 1. Intent-to-treat and treatment-on-treated effects on children’ outcomes

<table>
<thead>
<tr>
<th>Outcomes</th>
<th>ITT estimates&lt;sup&gt;a&lt;/sup&gt;</th>
<th>TOT estimates&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Effect</td>
</tr>
<tr>
<td><strong>Cognitive development</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASQ communication</td>
<td>2,448</td>
<td>-0.06</td>
</tr>
<tr>
<td>ASQ problems resolution</td>
<td>2,436</td>
<td>-0.06</td>
</tr>
<tr>
<td>ASQ gross motor</td>
<td>2,446</td>
<td>-0.11</td>
</tr>
<tr>
<td>ASQ fine motor</td>
<td>2,446</td>
<td>-0.10</td>
</tr>
<tr>
<td>WM general verbal ability</td>
<td>2,277</td>
<td>-0.22</td>
</tr>
<tr>
<td>WM associative memory</td>
<td>2,039</td>
<td>-0.19</td>
</tr>
<tr>
<td>WM executive function</td>
<td>2,095</td>
<td>0.15</td>
</tr>
<tr>
<td>WM mathematical reasoning</td>
<td>2,250</td>
<td>-0.69</td>
</tr>
<tr>
<td>WM receptive language</td>
<td>2,270</td>
<td>0.18</td>
</tr>
<tr>
<td><strong>Socio-emotional development</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASQ:SE total</td>
<td>2,429</td>
<td>-0.03</td>
</tr>
<tr>
<td><strong>Nutrition</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weight for age z-score</td>
<td>2,169</td>
<td>0.03</td>
</tr>
<tr>
<td>Length for age z-score</td>
<td>2,162</td>
<td>0.06</td>
</tr>
<tr>
<td>Weight for length z-score</td>
<td>1,141</td>
<td>-0.09</td>
</tr>
<tr>
<td>Arm circumference-for-age z-score</td>
<td>1,156</td>
<td>-0.03</td>
</tr>
<tr>
<td>Underweight</td>
<td>2,169</td>
<td>0.01</td>
</tr>
<tr>
<td>Wasting</td>
<td>1,141</td>
<td>0.01</td>
</tr>
<tr>
<td>Stunting</td>
<td>2,162</td>
<td>0.02</td>
</tr>
<tr>
<td>Overweight</td>
<td>2,066</td>
<td>0.00</td>
</tr>
<tr>
<td>Obesity</td>
<td>2,066</td>
<td>0.01</td>
</tr>
<tr>
<td><strong>Health</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anemic</td>
<td>2,113</td>
<td>-0.01</td>
</tr>
<tr>
<td>Parasites</td>
<td>2,143</td>
<td>0.01</td>
</tr>
</tbody>
</table>

Note: ***p<0.01; **p<0.05; *p<0.1. Cluster robust standard errors in parentheses.
The covariates included are listed below equation (1).

<sup>a</sup> OLS estimates.
<sup>b</sup> Two-stage least squares estimator; uses random assignment as an instrument to predict program compliance.
<sup>c</sup> The p-value obtained from the RW procedure for multiple hypotheses testing, within a developmental domain. Total scores are not included in the RW procedure (hence it is shown as NA).