IDB
Inter-American
Development Bank

Community Monitoring Improves Public Service Provision at Scale:

Experimental Evidence from a Child Development Program in Nicaragua

Social Protection and Health
Division

TECHNICAL
NOTE Nº
IDB-TN-2057

Florencia López Boo
Jane Leer
Akito Kamei

November 2020

Community Monitoring Improves Public Service Provision at Scale:

Experimental Evidence from a Child Development Program in Nicaragua

Florencia López Boo
Jane Leer
Akito Kamei

November 2020

IDB
Inter-American Development Bank

Cataloging-in-Publication data provided by the
Inter-American Development Bank
Felipe Herrera Library
López Boo, Florencia.
Community monitoring improves public service provision at scale: experimental evidence from a child development program in Nicaragua / Florencia López Boo, Jane Leer, Akito Kamei.
p. cm. — (IDB Technical Note ; 2057)
1. Child development-Government policy-Nicaragua. 2. Child care services-Government policy-Nicaragua. 3. Child welfare-Government policy-Nicaragua. 4. Child welfare-Nicaragua-Citizen participation. 5. Home-based family services-Nicaragua. I. Leer, Jane. II. Kamei, Akito. III. Inter-American Development Bank. Social Protection and Health Division. IV. Title. V. Series.
IDB-TN-2057

scl-sph@iadb.org

www.iadb.org/SocialProtection

Community Monitoring Improves Public Service Provision at Scale: Experimental Evidence from a Child Development Program in Nicaragua

Florencia Lopez Boo*, Jane Leer†, and Akito Kamei‡

November 20, 2020

Abstract

Expanding small-scale interventions without lowering quality and attenuating impact is a critical policy challenge. Community monitoring offers a low-cost quality assurance mechanism by making service providers accountable to local citizens, rather than distant administrators. This paper provides experimental evidence from a home visit parenting program implemented at scale by the Nicaraguan government, with two types of monitoring: (a) institutional monitoring; and (b) community monitoring. We find a positive intent-to-treat effect on child development, but only among groups randomly assigned to community monitoring. Our findings show promise for the use of community monitoring to ensure quality in large-scale government-run social programs.

Keywords: Early childhood development policy; Community monitoring; RCT; Latin America; Nicaragua

JEL Classification: I20, I24, J13

*Inter-American Development Bank, Washington, DC, US, florencial@iadb.org
†Duke University, jane.leer@duke.edu
‡University of Illinois at Urbana-Champaign, kamei2@illinois.edu

1 Introduction

A robust body of evidence demonstrates the value of investments made during the early years (Heckman, 2008; Richter, Daelmans, Lombardi, Heymann, Lopez Boo, Behrman, Lu, Lucas, Perez-Escamilla, Dua, et al., 2017). Expanding effective early childhood interventions is a global priority, as evidenced by recent multilateral commitments to early childhood development, such as Goal 4 of the 2015-2030 Sustainable Development Goals and the Group of 20 (G20) Initiative for Early Childhood Development. The key challenge going forward is not *whether* to invest in early childhood development, but *how* to cost-effectively expand evidence-based interventions in order to reach broader and more heterogeneous populations (Richter et al., 2017).

Home visits are one example of an evidence-based early childhood intervention (Berlinski and Schady, 2016). In home visits, trained home-visitors provide regular visits to families of young children with the goal of improving parents' capacity to support their children's development. The home visit theory of change relies on research showing the importance of warm, consistent, and frequent early stimulation in supporting early child development (Aboud and Yousafzai, 2015). Home visits tend to target vulnerable populations based on evidence showing socioeconomic gradients in the quality of the home environment (Lopez Boo, Cubides Mateus, Sorio, Garibotto, and Berón, 2019), which in turn is linked to the emergence of large socioeconomic disparities in cognitive, language, and social emotional development by at least as early as 18 months (Fernald, Marchman, and Weisleder, 2013).

A large and growing evidence base supports the potential of home vis-

its. Efficacy trials of home visit programs implemented in the United States and low and middle income countries (LMICs) have demonstrated positive short-term effects on early childhood development (e.g. Andrew, Attanasio, Fitzsimons, Grantham-McGregor, Meghir, and Rubio-Codina, 2018; Attanasio, Fernández, Fitzsimons, Grantham-McGregor, Meghir, and Rubio-Codina, 2014; Yousafzai, Rasheed, Rizvi, Armstrong, and Bhutta, 2014). There is also evidence of sustained effects, including improved educational achievement, better health, and less criminal behavior during adolescence (Olds, Henderson, Cole, Eckenrode, Kitzman, Luckey, Pettitt, Sidora, Morris, and Powers, 1998; Walker, Chang, Vera-Hernandez, and Grantham-McGregor, 2011), and improved labor market outcomes in adulthood (Gertler, Heckman, Pinto, Zanolini, Vermeersch, Walker, Chang, and Grantham-McGregor, 2014). However, the extant literature on home visits from LMICs relies primarily on pilot programs serving between 100 - 1,000 children.

We provide novel evidence of the effects of home visits delivered at scale. We evaluate the impact of a home visit parenting program implemented by the Nicaraguan Government, reaching 32,900 children, roughly 90% of the population of zero to three year-olds in rural municipalities. In addition, we compare two different approaches to ensuring quality at scale: "status-quo," institutional monitoring versus community monitoring. To do so, we designed an RCT with a control condition (no home visits) and two treatment arms. The first treatment arm received home visits with institutional monitoring, consisting of irregular and infrequent observations conducted by supervisors from the Ministry of the Family. On top of the institutional monitoring,

the second treatment arm leveraged Nicaragua's history of grass-roots civic participation (Arnove, 1981) to implement a community-led monitoring system (inspired in part by Björkman and Svensson, 2009).

We find a significant intent-to-treat (ITT) effect of 0.13 standard deviation (SD) on child development, but only in communities assigned to home visits with community monitoring. These effects are modest yet noteworthy considering there was substantial non-compliance with treatment assignment. Our local average treatment estimate (LATE) shows effects roughly twice as large: 0.28 SD for the child development. We also observe a positive ITT effect on children's behavior in both treatment groups (0.13 - 0.17 SD).

Our findings contribute to the literature in the following ways. First, we provide one of the first-ever experimental evaluations of a low-cost, publicly provided home visit program implemented at scale in a developing country. We show that a government-implemented home visit intervention reaching larger and more heterogeneous populations can be effective, but only in the presence of adequate accountability mechanisms. Second, and relatedly, we contribute to the question of how to ensure quality at scale, especially given limited institutional resources to monitor service delivery. We show that community-led monitoring offers potential gains that could be useful for similarly resource-constrained governments facing a tradeoff between quality and scale. Finally, we add to the literature on efficient public service delivery via community monitoring in LMICs, which to date has focused primarily on health, primary education, local infrastructure, and natural resource management (Fox, 2015; Waddington, Sonnenfeld, Finetti, Gaarder, John, and Stevenson, 2019). Our

findings suggest there is broad potential for community monitoring, even for interventions like home visits for which the potential benefits are not well understood by community members prior to intervention launch.

We begin by motivating our study within the literature on home visits and community monitoring. Section 3 discusses the setting of the home visit program and our experiment, sections 4 and 5 explain our empirical design, section 6 presents the results, and the last section concludes with policy and research directions.

2 Literature on Home Visit Programs and Community Monitoring

A robust literature documents the potential of home visits to positively impact child development. Recent evidence from meta-analyses of 18 home visit models in the United States suggest short-term effects on child development in the magnitude of 0.10 - 0.20 SD (e.g. Sama-Miller, Akers, Mraz-Esposito, Avellar, Paulsell, Del Grosso, et al., 2017). The best known and most well studied home visit programs in the United States are the Nurse-Family Partnership (NFP) and Early Head Start (EHS). There is evidence of longitudinal effects of both, namely on cognitive development at school entry (Bradley, Chazan-Cohen, and Raikes, 2009) and behavior at age 15 (Olds, Henderson, Kitzman, Eckenrode, Cole, and Tatelbaum, 1999; Olds, Kitzman, Cole, Robinson, Sidora, Luckey, Henderson, Hanks, Bondy, and Holmberg, 2004). One of the most influential studies of a home visiting program in a developing

country comes from Jamaica, where findings from an efficacy trial show positive effects of home visits during early childhood on subsequent cognitive, social-emotional, and health outcomes during middle childhood and adolescence (Baker-Henningham, Powell, Walker, and Grantham-McGregor, 2003; Grantham-McGregor, Powell, Walker, Chang, and Fletcher, 1994), leading to higher wages in adulthood (Gertler et al., 2014).

Home visits have since been piloted in Bangladesh, Brazil, Colombia, India, Jordan, and Pakistan, among other LMIC countries. The short term effects of these pilot programs are promising, with effect sizes on child development ranging from 0.20 - 0.70 SD (Aboud and Yousafzai, 2015). However, evidence on the extent to which home visits improve parenting behavior in LMICs is inconsistent; some studies have found large effects on parenting behavior (Aboud, Singla, Nahil, and Borisova, 2013), some find no effects (Chang, Grantham-McGregor, Powell, Vera-Hernández, Lopez Boo, Baker-Henningham, and Walker, 2015), but many do not measure parenting behavior at all. Currently, home visit programs operate at scale in several countries in Latin America and the Caribbean (LAC), including Brazil, Chile, Ecuador, and Peru (Leer and Lopez Boo, 2018), but existing evaluations of these programs use non-experimental methods. Our paper is therefore one of the first to provide robust causal estimates of a home visit program at scale.

Evidence suggests that the positive effects of home visits on child development and parenting behavior are conditional on home visit frequency, duration, content, and the quality of the home visitor-caregiver coaching relationship (Araujo, Dormal, and Rubio-Codina, 2018). Ensuring these four elements of

quality when going to scale is a key challenge. Large scale home visit programs in the United States frequently suffer from high program attrition, and observational data show that home visitors spend less than one fifth of the visit engaged in coaching interactions with caregivers (Vogel, Caronongan, Thomas, Bandel, Xue, Henke, Aikens, Boller, and Murphy, 2015). Similarly, observations of large-scale home visit programs in Latin America show that home visitors rarely devote sufficient time to language development, and engage primarily in child-focused activities, rather than providing caregivers with meaningful information regarding child development and early stimulation practice (Leer and Lopez Boo, 2018).

In light of these challenges, our main contribution, beyond estimating the impact of a government-run home visit program, relates to the use of community-led monitoring as an accountability mechanism to ensure the quality of service delivery at scale (Banerjee, Deaton, and Duflo, 2004; Björkman and Svensson, 2009; Olken, 2007). Community-led monitoring strategies aim to improve the quality of public service provision through two interrelated channels: (1) facilitating citizen engagement in service delivery, and (2) making service providers accountable to their local community, rather than to distant bureaucracies whose interests may not align with those of the community they serve (Ahmad, Commins, Devarajan, Filmer, Hammer, Pritchett, Reinikka, Shah, and Soucat, 2003).[1]

Positive effects of community monitoring have been observed in numer-

[1]Community monitoring is often described as a "short accountabilty route," linking citizens directly to service providers, in contrast to the "long route," whereby citizens only recourse to voice their (dis)content with service delivery is by voting (Ahmad et al., 2003).

ous health and primary education programs. For example, training school management committees to monitor teachers' performance has been shown to reduce teacher absenteeism and improve students' test scores (Duflo, Dupas, and Kremer, 2015), and community monitoring combined with information on staff performance has been linked to longer run improvements in health care delivery and health outcomes (Björkman Nyqvist, De Walque, and Svensson, 2017). However, community monitoring does not always work (Banerjee, Banerji, Duflo, Glennerster, and Khemani, 2010; Olken and Pande, 2012), and the extant literature highlights two key challenges: low rates of participation among community members and lack of relevant, actionable information regarding service delivery (Björkman and Svensson, 2009; Björkman Nyqvist et al., 2017).

Our intervention aims to address both of these challenges. To ensure participation, we work with pre-existing community associations in order to lend legitimacy and leverage pre-existing social networks. We also developed multiple written and oral feedback mechanisms to encourage participation from citizens with diverse preferences and skill sets. To ensure community members have relevant information regarding service delivery, the community-led monitoring strategy began with a series of information sessions covering what the intervention goals are in terms of quality and quantity of service delivery. This is key, because unlike other community monitoring experiments, which provided information about *outcomes* (e.g., score cards showing community-level infant mortality rates; Björkman and Svensson, 2009; Björkman Nyqvist et al., 2017), the information provided here focuses on the *quality* of service

delivery (e.g., the frequency and duration of home visits and the perceived quality of the home visitor-caregiver relationship).

This study is the first to examine community monitoring in the context of an early childhood development program. In doing so, we address an urgent gap in the early childhood policy literature regarding how to effectively monitor quality when going to scale. By relying on families themselves to provide frequent feedback on implementation quality, coupled with locally-managed feedback processes designed to instill prosocial incentives among home visitors, community-led monitoring may be a cost-effective way to expand early childhood interventions. Finally, whereas the evidence to date comes largely from pilot programs of community monitoring run by researchers, our study was part of a public program implemented by the Nicaraguan Government and reaching a large number of beneficiaries across the country.

3 Setting

Nicaragua is the second poorest country in Latin America and the Caribbean. The country is also among the top twenty countries with the highest income inequality in the world, with a Gini index of more than .50 (INIDE, 2014). The overall poverty rate (defined as annual consumption less than $500 USD per person) in 2014 was 30%, down from 43% in 2009 and 48% in 2003. Poverty rates are higher among households with children, with 35% of children younger than five living in poverty and 10% classified as extremely poor (INIDE, 2014).

There are two main government-provided early childhood development pro-

grams in Nicaragua. Child development centers (*centros de desarrollo infantil*, or CDI), implemented by the Ministry of Family, serve children aged roughly six months to four years old, free of charge, for three to four hours a day, five days a week. Public pre-schools, implemented by the Ministry of Education, enroll children aged three to six. At the beginning of our study (2013), CDIs were primarily concentrated in urban areas, and participation rates were low overall (11% of the eligible age group), but especially in rural areas, where less than 5% of parents reported sending their children to CDIs (INIDE, 2010). Pre-school enrollment was also low, at less than 40% overall and less than 30% among rural and low-income children aged three to six (INIDE, 2010). Further, after decades prioritizing expansion of coverage of early childhood education programs while largely ignoring the quality of services, most CDIs and pre-school centers lacked running water and electricity, and early childhood educators received minimal training, supervision, and remuneration. Private provision was also low (less than 4%; INIDE (2010)), and essentially non-existent outside of urban centers.

In response to these challenges, in 2011 the Government of Nicaragua passed the National Early Childhood Policy (*Política Nacional de Primera Infancia*), which aims to strengthen institutional support for early childhood development and raise awareness about the importance of the early years. The Home Visit Program, led by the Ministry of Family, was implemented under the umbrella of the National Early Childhood Policy.

4 Experimental Design

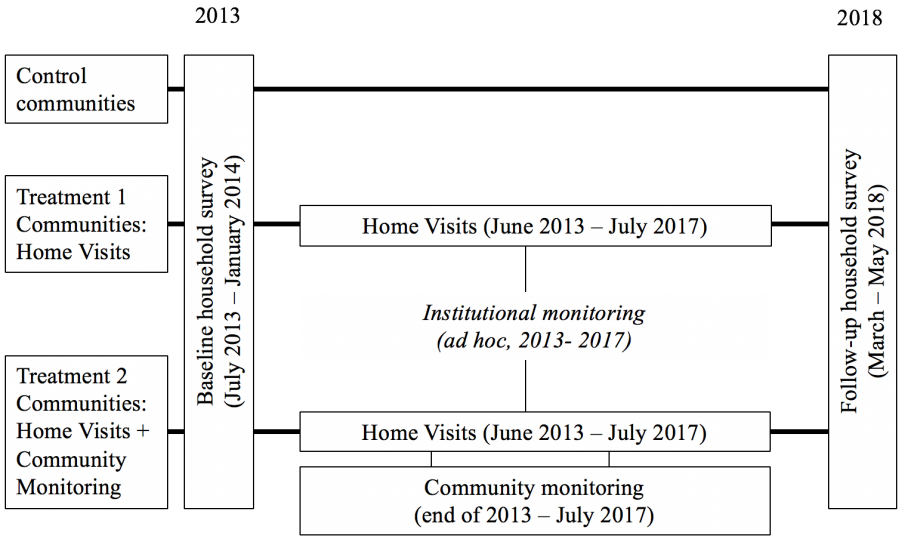
4.1 Overview

The Home Visit program targeted primarily rural areas, reaching a total of 32,900 children in close to 1,000 communities from 37 municipalities, selected based on need. Our experimental sample consists of a sub-sample of 210 communities from 33 municipalities.[2] Sample communities include 150 households on average, of which between 30-50% have children younger than three. The sociodemographic characteristics of our evaluation sample is comparable to the national average (Table A.1).

The 210 communities in the evaluation sample were randomly assigned to intervention and control conditions, with randomization stratified by municipality. Assignment to the control condition implied no home visits until after the study period. Treatment Arms 1 and 2 received home visits, and in Treatment Arm 2, communities were also supported to implement a community-based monitoring strategy, in addition to the standard institutional monitoring that both treatment arms received (described in Sections 4.2 and 4.3, respectively). Henceforth, we refer to Treatment Arm 1 as "Home Visits" and Treatment Arm 2 as "Home Visits + Community Monitoring." This design enables us to estimate the added benefit of community-based monitoring as a quality assurance mechanism. Figure 1 illustrates this design.

[2]See Appendix C for a description of the municipality and community selection process.

Figure 1: Intervention Timeline



4.2 Home Visits and Institutional Monitoring

The program design called for 60-minute visits, implemented fortnightly (every two weeks), at a cost of $100 per child, per year, for infants to three year olds. The Home Visit curriculum aims to stimulate language, cognitive and social emotional development using age-appropriate activities that were defined for each visit, based on the child's age (see Appendix B for details). Families received a booklet with text and pictorial descriptions of early childhood development milestones and parent-child activities, and a poster with pictures of different ways to promote early childhood development, but the program did not provide any learning materials. The visits included direct demonstration of activities by home visitors and time for caregivers to practice the activity with the target child and receive feedback from the home visitor. Home visitors also shared information regarding developmental milestones, empha-

sized the importance of consistent and warm caregiving practices, encouraged caregivers to avoid physical discipline practices, and shared information regarding relevant services in the community, such as healthy child check-ups and vaccination clinics.

The home visitors were community development workers, pre-school teachers, teaching assistants, or community health workers who were recruited from the participating communities. Some had post-secondary education and others had only completed secondary school. Each home visitor provided home visits to 18 - 25 families in addition to their regular duties as preschool teachers, community health workers, or administrative assistants for the Ministry of the Family. They were not paid extra for the home visits, although they did receive a small bonus (less than $40 a month, delivered sporadically) as well as a program hat, shirt, and backpack. Local level Ministry officials worked with the home visitors to identify and prioritize time for home visits - for example, in the afternoons after their shift at the pre-school.

After a one-week training period for home visitors, technical specialists at the municipal offices of the Ministry of the Family were responsible for ongoing supervision of home visit implementation. They were expected to observe each home visitor conducting home visits and provide feedback, at an intended frequency of at least four observations per year, per home visitor. However, in practice, there was much heterogeneity across communities in terms of the regularity and content of the feedback provided in the institutional monitoring visits (MIFAN, 2019). Most were only observed once or twice a year, and many specialists relied primarily on informal check-ins rather than home visit

observations. This irregularity was anticipated from the outset, for a number of reasons. First, the municipal-level specialists were responsible for overseeing all of the Ministry of the Family's activities in their respective municipality, meaning they had many other administrative and technical duties in addition to supervising home visitors. There were also financial constraints, as each monitoring visit entailed costs of travel to/from remote communities.

We refer to the institutional monitoring as "status quo" monitoring. We consider "status quo" monitoring to be comparable to the kind of institutional monitoring of social and education programs provided on average in many LMICs (Leer, Lopez Boo, Perez Expósito, and Powell, 2016), where a key challenge is how to ensure quality when monitoring costs are high and the frontline workers (i.e., home visitors) receive low wages, have other work responsibilities, and have minimal opportunities for professional development.

4.3 Community Monitoring

Community Monitoring offers a low-cost complement to weakly implemented "status quo" monitoring. The aim is to incentivize home visitors not only to actually conduct the home visits, but to do so with quality (i.e., adhere to the home visit design regarding frequency, duration, and visit content). Community monitoring also offered the possibility of increasing home visit take-up and persistence in the program by giving program beneficiaries a voice in program implementation.

In communities assigned to Home Visits + Community Monitoring (Treatment Arm 2), *Gabinetes de Familia, Comunidad, y Vida* (Family, Community,

and Life Committees) were responsible for leading the monitoring process. The *Gabinetes de Familia, Comunidad, y Vida* (*Gabinetes* hereafter) are established community associations that existed prior to the start of the home visit program and that were already meeting monthly.[3] *Gabinete* membership is voluntary and open to anyone in the community. Participants are typically teachers, health workers, and religious leaders, as well as adults and youth without any formal leadership position in the community.

Ministry officials led a full day in-person training for the *Gabinetes* that included information regarding the purpose, content, and process of both the home visits and the community monitoring strategy. The *Gabinetes* then organized a community assembly to inform the broader community, encourage participation in the home visits and the community monitoring process, and identify *jovénes promotores* (youth volunteers), one to two youth per community to support the community monitoring process. Families receiving home visits were provided with a survey every two months that was designed to elicit feedback on home visit frequency, duration, and content, as well as the quality of the home visitor/family interactions (see Figure B.1). A locked suggestion box was set up in a central location for families to deposit their feedback surveys and any other written feedback. Families were also encouraged to provide feedback verbally to the youth volunteers or *Gabinete* members.

Every two months, the *Gabinetes*, youth volunteers, and home visitors met to review the feedback received in the prior months. The *Gabinetes* and

[3]In about five communities, the *Gabinetes* were non-existent or not operational at the start of implementation. In these cases, a different pre-existing community association led the community monitoring, such as Parent/Teacher Committees)

youth volunteers then presented the feedback in community assemblies that were open to all participating families and home visitors. All feedback was intended to be presented anonymously, without comparing home visitors or calling them out individually. Each meeting concluded with an Action Plan, designed to highlight achievements to date as well as aspects for improvement and concrete next steps to improve home visits. The *Gabinetes* were expected to follow-up on the Action Plan agreements by providing support to families and home visitors as needed to meet the agreed upon goals.

Figure B.2 illustrates the community monitoring process. Importantly, we expect families receiving home visits in Treatment 2 communities to benefit from community monitoring regardless of their own participation in the feedback process or meetings. That is, community monitoring should serve as an accountability mechanism (via social pressure from community leaders and peers) to incentivize home visitors to improve the overall fidelity of implementation of home visits (we directly assess this hypothesis in Section 6.7).

4.4 Data

Data were collected from all 210 participating communities. Within each community, the sample was a simple random sample, drawn from the list of eligible households (i.e., all households with young children).[4] Baseline data were collected from July 2013 - January 2014. The response rate was high, 93% of the

[4]In participating communities, all households with children younger than six years old were identified via a census conducted immediately prior to baseline data collection by the survey team. We focus our analysis on the sample of children who were between six months and two years old at baseline, because our key outcome measure, the Denver-II, has an age cut-off of six years old and thus was not applied to children older than 5.99 years at follow-up. Moreover, children older than three were not eligible for fortnightly home visits.

originally randomly selected households agreed to participate in the baseline survey. Home visits and community monitoring were implemented from June 2013 through July 2017, and follow-up panel data were collected from March - May, 2018 (Figure 1). Children were between zero and one years old at baseline (mean = 5.2 months), and between four and five years old at follow-up (mean = 4.7 years). Our analytic dataset includes 1,324 children. Sample attrition from baseline to follow-up was less than 10%, and did not differ by treatment condition (Tables C.1 and C.2).

Questionnaires at baseline and follow-up collected information on household characteristics, child development, and parenting behavior, as follows:

Household characteristics. Data on household characteristics include the age, sex, education, and health status of all household members, as well as household assets, and household economic activity, collected from the household head.

Child development. We use two assessments that are widely used in the literature, have been previously adapted and validated in the Nicaraguan context, and include multiple domains: the Denver-II (Frankenburg, Dodds, Archer, Shapiro, and Bresnick, 1992) and the Peabody Picture Vocabulary Test (PPVT-III; Dunn and Dunn, 1981). From the Denver-II we derive measures of language and fine motor development via direct assessment (and some caregiver report, for very young children only). The PPVT-III measures receptive vocabulary via direct assessment. We compute a continuous score equal to the sum of the of the items successfully completed (in line with Rubio-Codina, Araujo, Attanasio, Muñoz, and Grantham-McGregor (2016)), and in all mod-

els we use the within-sample standardized score, calculated in two-month age intervals (i.e., we group children according to age and then compute z-scores within age bins of two months). We also report an aggregate measure of child development using factor analysis, constructed from the three direct assessment measures (Denver Language, Denver Fine Motor, and receptive vocabulary as measured by the PPVT).[5] Further details on the child development measures, including administration, scoring, and standardization methods are given in Appendix D.

Child Behavior Child behavior was measured via caregiver-report using an adapted version of Behavior Problem Index (BPI; Zill and Peterson (1986), with items derived from Achenbach and Edelbrock (1981)) (e.g., "does [your child] cry often?" "does she get distracted easily?"). We reverse code negative items so that higher scores indicate a better behavior, and then compute a continuous score equal to the sum of the of the items. In all models we use the within-sample standardized score, calculated in two-month age intervals. Child behavior was only measured at follow-up.

Parenting measures. Data on parenting were collected via caregiver self-report and direct observation of the primary caregiver and child during the baseline and follow-up interview, using the Home Observation for Measurement of the Environment (HOME)(Caldwell, Bradley, et al., 1984). We include three sub-scales, based on the HOME Short-Frame: emotional support (e.g.,

[5]Our ITT and LATE findings are consistent using a measurement systems approach, in which we use the observed (but error prone) measures of child development to estimate a latent factor of child development, following Attanasio, Cattan, Fitzsimons, Meghir, and Rubio-Codina (2020); Cunha, Heckman, and Schennach (2010). Tables available upon request.

whether the caregiver responded verbally to the child during the interview), cognitive stimulation focusing on materials (e.g., presence of story books in the home), and cognitive stimulation focusing on activities (e.g., whether some one at home tells stories to the child). In line with evidence from numerous cultural contexts and socioeconomic groups showing a negative association between maternal depression and the quality of parent-child interactions (Wachs, Black, and Engle, 2009), we also use an adapted version of the Center for Epidemiologic Studies Depression Scale (CES-D) (Radloff, 1977), a screening test for depressive symptoms. In all estimates we use the with-in sample standardized score for the HOME and CES-D scores. We take the inverse of the CES-D score so that higher scores across all measures imply improved parenting/less maternal depression (see Appendix D for details).[6]

5 Empirical Strategy

To estimate the causal effect of Home Visits and Home Visits + Community Monitoring on child development and behavior we leverage random assignment of communities to treatment and control conditions. Our intent-to-treat estimation is as follows:

$$y_{ic2018} = \alpha + \beta_1 HomeVisits_c + \beta_2 HomeVisits\&Monitoring_c + y_{ic2013} + X_{ic} + \gamma_m + \epsilon_{ic} \quad (1)$$

where y_{ic2018} is the child development outcome for child (caregiver) i in commu-

[6]Lopez Boo (2019, mimeo) discusses the psychometric properties of Denver-II, PPVT-III, BPI, and CES-D in this impact evaluation.

nity c. The parameters of interest are the coefficients on the treatment indicators for assignment to Home Visits and assignment to Home Visits + Community Monitoring. X_{ic} is a vector of individual, household and community-level controls measured at baseline (child gender, child age in 2-month increments, mother's years of education, whether the biological mother is the primary caregiver, an index of household assets, the total number of household residents, and whether the community is semi-urban or rural).[7] Our preferred specification includes controls for precision, but our main findings are robust in unadjusted models. Because the intervention was stratified at the municipality level, we also include municipality fixed effects (γ_m), and we control for the baseline child development outcome (y_{ic2013}).[8] Standard errors are clustered at the community level.

Given random assignment, β_1 estimates the causal impact of residing in a community that was offered Home Visits, and β_2 estimates the causal impact of residing in a community that was offered Home Visits + Community Monitoring.

We also report the local average treatment effect (LATE) estimate. The LATE estimate uses community level assignment as an instrument to model the effect of actually receiving at least one home visit. These models are run separately for each of the two treatment arms. Findings from the LATE estimate apply only to the sub-sample of compliers, or individuals who were

[7]The asset index is created using principal component factor analysis, designed to be a proxy for household wealth.

[8]Receptive vocabulary (PPVT-III) and behavior (BPI) were only measured at follow-up. Thus, for these outcomes we use the baseline Denver-II language and composite child development score, respectively as a control for baseline developmental outcome.

induced to take-up the home visits because of their community-level assignment. By estimating the average effect even in the presence of substantial spillover, we argue that the ITT estimate more directly answers the question of *what is the impact of home visits implemented at scale?* in that it provides a more conservative estimate that is generalizable to our entire population of interest (i.e., all households with young children living in high poverty, rural or semi-rural communities in Nicaragua). The ITT is thus our preferred estimate, although the results of the LATE estimates are consistent with our main ITT results in terms of direction and significance of coefficients.

6 Results

6.1 Preintervention Differences

Table 1 presents baseline characteristics of children across treatment arms. Half of the children in our sample are male with an average age of 3 to 5 months. With respect to household characteristics, more than 90% of the children's primary caregiver is their biological mother, and mothers in our sample have on average seven to eight years of education, with almost a third reporting depressive symptoms. The majority of caregivers exhibited parental warmth during the interview (e.g., responded verbally to the child), but less than ten percent reported reading to their child in the past week.

Figure 2 shows mean within-sample age standardized child development scores at baseline, according to levels of household wealth (as measured by the household asset index). There are clear wealth gradients, even in our sample,

which by design targets the poorest communities in Nicaragua. This finding is consistent with previous evidence from Nicaragua and other countries in Latin America (e.g. Schady, Behrman, Araujo, Azuero, Bernal, Bravo, Lopez-Boo, Macours, Marshall, Paxson, and Vakis, 2015), and it is precisely these inequalities in early childhood development that motivate the home visit program.

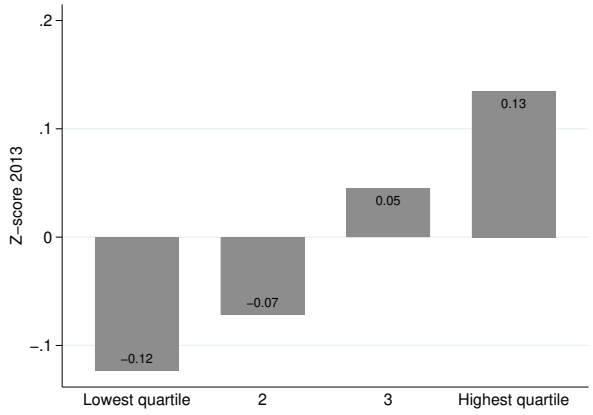
Note that baseline characteristics of all observed demographics, parenting behavior, and child development outcomes are balanced across treatment arms, with a few exceptions (Column (4) - (6)). Nonetheless, we control for all baseline characteristics in our estimates.

Table 1: Baseline Characteristics Across Treatment Arms (Baseline Balance)

	(1) Control	(2) Home Visits	(3) Home Visits +Community Monitoring	(4) Diff (2)-(1)	(5) Diff (3)-(1)	(6) Diff (3)-(2)
Child demographics						
Male	0.50	0.52	0.53	0.02	0.04	0.01
Age in years	0.33	0.38	0.38	0.05	0.05	0
Household characteristics						
Semi-urban	0.72	0.73	0.73	0	0	0
Household asset index	0.05	0.06	0.14	0.01	0.09	0.08
Biological mother is primary caregiver	0.93	0.93	0.92	0	-0.01	-0.01
Mother completed at least primary level	0.71	0.72	0.72	0.01	0.01	0
Mother years of education	7.80	7.56	7.90	-0.24	0.10	0.35
Parenting characteristics						
HOME Emotional Support (Z score)	0.01	0.02	0.12	0	0.10	0.10
Responded verbally to child	0.72	0.74	0.79	0.02	0.07	0.06
Praised child during the interview	0.65	0.64	0.66	-0.01	0.01	0.02
HOME Cognitive Stimulation (Z score)	-0.06	0.04	-0.05	0.10	0.01	-0.09
Someone reads to child	0.39	0.44	0.36	0.05	-0.03	-0.08
Read to child at least once a week	0.07	0.10	0.09	0.03	0.01	-0.01
Maternal depression (CES-D)	29.01	29.29	29.64	0.28	0.63	0.35
At risk of clinical depression	0.28	0.27	0.26	-0.01	-0.03	-0.02
Child development outcomes						
Child Development Composite	0.03	0.08	-0.09	0.04	-0.12	-0.17
Language (Z score)	0.09	0.10	-0.06	0.02	-0.14	-0.16
Fine Motor (Z score)	0.09	0.09	0.02	0	-0.07	-0.06
Observations	420	442	462	1325	1325	1325

Notes: Analysis restricted to panel dataset of children observed at baseline (2013) and follow-up (2018) (for this reason, the standardized development outcomes are not all equal to 0 for the overall sample). The household asset index is created using principal component factor analysis, designed to be a proxy for household wealth. The max score for Maternal depression (CES-D) is 36.

Figure 2: Wealth Gradients in Child Development at Baseline



Notes: Bars correspond to the mean of the child development composite indicator (z-score) at baseline for each quartile of household wealth (estimated using an index of household assets). N =1,319.

6.2 Home Visit Take Up and Participation in Community Monitoring

At follow-up, caregiver respondents were asked about home visit participation (i.e., "did you receive home visits?") and if so, the frequency and duration of the home visits, and how many months in total they received home visits. Caregivers were also asked about participation in community monitoring. Specifically, caregivers were asked whether or not they had attended a community meeting in which the quality of the home visits was discussed, and if so, how many total meetings they attended, and their perceptions regarding whether or not community monitoring improved home visit quality. We use these measures to estimate home visit take-up and participation in community monitoring, as well as the perceived effectiveness of the community monitoring,

as shown in Table 2.

As expected, families living in communities assigned to Home Visits (Treatment Arm 1) and Home Visits + Community Monitoring (Treatment Arm 2) were much more likely to receive home visits than families in communities assigned to the control condition, although compliance was imperfect. Approximately 68-69% of families in communities assigned to Home Visits and Home Visits + Monitoring received at least one home visit, compared to 30% in control communities. Comparing columns (2) and (3) in Table 2, the frequency, duration, and number of months of home visits appears to be slightly greater in communities assigned to community monitoring relative to communities that received "status quo" monitoring, although the differences are not significant by conventional standards. Multiplying self-reported number of months of home visits by reported frequency of home visits, we estimate the median number of total home visits at 24, for both treatment groups.

Non-compliance in the control group (i.e., receipt of home visits) appears to be partly explained by proximity to treatment communities. Control group participation in home visits is highest among households located close to treated communities (within 1 km; results available upon request). Baseline demographics, parenting, and child development characteristics do not predict home visit receipt.[9] Likewise, there is no evidence that non-compliance in the treatment groups (i.e., non-receipt of home visits) is related to baseline characteristics. Rather, we suspect that non-compliance in treatment communities is primarily explained by implementation challenges, such as home

[9] Assessed via logistic regressions of home visit receipt on demographic, parenting, and child development variables, tables available upon request.

visitor turnover which may have caused gaps in service delivery.

A small percentage (9%) of caregivers in communities assigned to the control condition reported attending at least one community monitoring meeting, compared to 22% in Home Visit communities and 33% in Home Visit + Community Monitoring communities. Non-compliance in this case also appears to be largely explained by proximity to Home Visit + Community Monitoring communities. Among households in Treatment 1 communities, participation in community monitoring is highest for those who live within 1 km of the nearest Treatment 2 community. Regional coordination in themes related to governance and program administration between *Gabinetes* is not infrequent, so it could be that members of the *Gabinetes* in Home Visit + Community Monitoring communities discussed the community monitoring mechanisms with *Gabinetes* or families in Home Visit communities. Alternatively, as described in Section 2.4, the *Gabinetes* are pre-existing associations that operate in most communities throughout the country and have a mandate to promote participation in government provided social services. It is conceivable that the *Gabinetes* in Home Visit communities decided on their own to implement some form of community monitoring. However, only communities assigned to the monitoring condition received training from the Ministry of the Family in community monitoring. Thus, the community monitoring that families in Home Visit-only communities reported participating in would not necessarily have followed the same structure in terms of frequency, organization, or content of the feedback processes.

Table 2: Program Participation by Treatment Assignment

	(1)	(2)	(3)	(4)	(5)	(6)
	Control	Home Visits	Home Visits +Community Monitoring	Diff (2)-(1)	Diff (3)-(1)	Diff (3)-(2)
Received home visits	0.30	0.68	0.69	0.37	0.39	0.02
Frequency of home visits						
Never received home visit	0.70	0.32	0.31	-0.37	-0.39	-0.02
Once a month	0.10	0.20	0.16	0.10	0.07	-0.04
Fortnightly	0.08	0.24	0.30	0.17	0.23	0.06
Once a week	0.09	0.17	0.15	0.08	0.06	-0.02
Duration of home visit						
Never received home visit	0.70	0.32	0.31	-0.37	-0.39	-0.02
Less than 20 mins	0.05	0.08	0.08	0.03	0.03	0
20-40 mins	0.15	0.31	0.26	0.17	0.11	-0.05
More than 40 mins	0.11	0.28	0.35	0.17	0.24	0.08
Number of months received home visits	3.43	10.90	11.51	7.47	8.08	0.61
Community Monitoring (CM)						
Attended CM Meetings	0.09	0.26	0.37	0.18	0.29	0.11
Number of CM Meetings Attended	0.27	1.47	2.12	1.20	1.85	0.65
CM makes the program						
Never attended	0.91	0.74	0.63	-0.18	-0.29	-0.11
Little better	0.02	0.07	0.06	0.05	0.04	0
No change	0	0.01	0.01	0.01	0.01	0
Much better	0.06	0.19	0.30	0.13	0.24	0.11
Observations	420	442	462	1325	1325	1325

Notes: Data on program participation come from caregiver self-report at follow-up.

6.3 Quality of Home Visit Implementation

Our findings regarding program take-up (Table 2) show that community monitoring did not meaningfully increase the quantity of home visits, but did community monitoring improve the *quality* of home visits? To answer this question, we leverage a unique dataset of in-situ observations of home visits. Rarely do early childhood development programs in LMIC collect data on program implementation like this (**?**). The observations were conducted by technical specialists from the Ministry of the Family as part of the institutional ("status quo") monitoring in both treatment arms. The data represent the total universe of observations collected and recorded by technical specialists throughout the duration of the study, from 2013 - 2017. The data include a total of 2,130 observations from 44 (of 70) Home Visit communities,

and 2,783 observations from 51 (of 70) Home Visit + Community Monitoring communities (see Appendix E). The fact that more than 30% of treatment communities are not included in these data demonstrates the challenges of institutional monitoring in this setting, and hence, the rational for community monitoring. It also means that these data are not representative of the full sample of visits conducted as part of the Home Visit Program Program, which is a limitation of this analysis.

Home visit quality was measured using a check-list tool designed to be completed immediately following each in-situ observation, based on Leer et al. (2016). The checklist includes four categories: the quality of the home visitor-caregiver relationship (e.g. "home visitor encourages caregiver questions"), visit materials and preparation (e.g. "home visitor had prepared in advance for the visit"), visit content (e.g. "home visitor models activities for families"), and the extent to which the home visitor adhered to program design by focusing on demonstrating activities to the caregiver. Each category includes one to four items, scored on a four-point scale. We calculate scores for each category by taking the sum of the corresponding items (or the sum of all items for the overall category), and standardizing by year.

We regress home visit quality on community assignment to Home Visits + Community Monitoring.[10] As shown in Table 3 (Panel A), we find that home visits were significantly higher quality in communities assigned to Home Visits + Community Monitoring, relative to communities that just received

[10]The regression from which we extract our main parameter of interest, β_1, is:

$$y_{ijc} = \alpha + \beta_1 HomeVisits\&CommunityMonitoring_c + \delta_t + \gamma_k + \epsilon_{ijc} \quad (2)$$

institutional monitoring. The overall quality of home visits was 0.12 SD higher in communities assigned to community monitoring, equivalent to about two additional items with high scores on the home visit observational check-list. However, we interpret this estimate with some caution given the fact that our sample of home visits might be selective as it does not include all treatment communities, nor participating families and home visitors.

To further examine the role of community monitoring in improving home visit quality, we use data from the caregiver interview at follow-up to construct two measures of participation in monitoring at the community level. First, we estimate the number of meetings held per community, using caregivers' self-report data on number of meetings they attended. Second, we estimate the perceived role of monitoring in improving the quality of home visits at the community level by taking the average of caregivers' reports in each community (scored on a Likert scale, ranging from 1=never attended to 4=much better). Restricting our analysis to Treatment 2 communities, we then regress home visit quality on these two measures of participation in community monitoring.[11] As shown in Table 3 (Panel B), the number of monitoring meetings

where y_{ijc} is the quality of home visit i (in year-standardized z-scores) conducted by home visitor j in community c. The parameter of interest is β_1, the coefficient on the treatment indicator for assignment to Home Visits + Community Monitoring. We include year fixed effects (δ_t) and technical supervisor fixed effects (γ_k). Standard errors are clustered by home visitor to account for non-independence of errors between observations of the same home-visitor.

[11] The regression model is as follows:

$$y_{ijc} = \alpha + \beta_1 No.of Monitoring Meetings_c + + \beta_2 Perceived Role of Monitoring_c + \delta_t + \gamma_k + \epsilon_{ijc} \quad (3)$$

where y_{ijc} is the quality of home visit i (in year-standardized z-scores) conducted by home visitor j in community c, with year and technical supervisor fixed effects (δ_t and γ_k), and standard errors clustered by home visitor.

held per community significantly significantly predicts all domains of visit quality. The coefficients on the perceived role of monitoring in improving visits are large and positive, but only significant for whether or not the home visitor focused on the caregiver. We take these findings as encouraging, if suggestive, evidence that community monitoring served as an effective quality assurance mechanism.

6.4 Child Development Outcomes

Table 4 present our main estimates (See Figure F.1 for the coefficient plot). We find no evidence of effect of home visits with "status-quo" institutional monitoring on child development, but there is a positive effect of home visits with community monitoring on child development and child behavior (ITT 0.13 and 0.17 SD, respectively). The impacts are driven by expressive language and fine motor skills in the case of the child development composite.

The magnitude of the effects on child development, language and fine motor development are consistent with the short-term ITT effects found in home visit programs implemented at scale in the U.S. (e.g. Love, Chazan-Cohen, Raikes, and Brooks-Gunn, 2013). Understandably, the impacts we find here are substantially smaller than the short term effects found in effectiveness trials in LMICs, which tend to be between 0.20 - 0.70 SD (Aboud and Yousafzai, 2015). Indeed, the United States literature shows that as child care centers from Head Start were expanded, the impacts were dramatically attenuated (Duncan and Magnuson, 2013).

Table 3: Relation Between Community Monitoring and Home Visit Quality

Panel A
Assignment to Community Monitoring and Home Visit Quality

	Overall quality	Home visitor - caregiver relationship	Visit materials and preparation	Visit content	Focuses on caregiver
Community Monitoring	0.12	0.13	0.09	0.13	0.09
	(0.06)	(0.06)	(0.05)	(0.06)	(0.05)
Controls	Y	Y	Y	Y	Y
N	4,913	4,913	4,913	4,913	4,913
Adjusted R^2	0.32	0.31	0.27	0.28	0.22

Panel B
Community-Level Participation in Community Monitoring and Home Visit Quality

	Overall quality	Home visitor - caregiver relationship	Visit materials and preparation	Visit content	Focuses on caregiver
Number of monitoring meetings held per community	0.05	0.04	0.04	0.05	0.05
	(0.02)	(0.03)	(0.02)	(0.02)	(0.02)
Perceived role of monitoring in improving home visits	0.35	0.28	0.34	0.28	0.52
	(0.25)	(0.23)	(0.24)	(0.24)	(0.24)
Controls	Y	Y	Y	Y	Y
N	2,717	2,717	2,717	2,717	2,717
Adjusted R^2	0.32	0.30	0.27	0.29	0.21

Note: Outcomes, in columns, come from in-situ observations of home visits, collected using a checklist of home visit quality by trained specialists from the Ministry of the Family and presented as z-scores standardized by year. Panel A shows the result of regressing home visit quality outcomes using observations collected from communities assigned to both treatment arms on assignment to Treatment 2 (Home Visits + Community Monitoring). Analyses in Panel B are restricted to the observations conducted in communities assigned to Treatment 2 (Home Visits + Community Monitoring). We regress home visit quality outcomes on two community-level measures of participation in community monitoring, constructed from caregiver reports at follow-up: (1) the average number of monitoring meetings caregivers reported participating in, and (2) the perceived role of community monitoring in improving home visit quality (1, community monitoring does not improve home visits, to 3 community monitoring makes the home visits much better), averaged by community. Controls in Panel A and B consist of year and supervisor fixed effects. Standard errors in parentheses are clustered by home visitor.

Table 4: Estimated Effect of Home Visits on Child Development

	Intent-to-Treat Effect		**Local Average Treatment Effect**	
	(1)		(2)	(3)
	Home Visits	**Home Visits+ Community Monitoring**	**Home Visits**	**+Home Visits Community Monitoring**
Child Development	**0.10**	**0.13**	**0.22**	**0.28**
	(0.07)	**(0.07)**	**(0.18)**	**(0.14)**
N		*1,285*	*838*	*860*
Adjusted R^2		0.27	0.29	0.29
P-Val: HV v HV+CM		0.62		
Language	**0.09**	**0.11**	**0.25**	**0.25**
	(0.07)	**(0.06)**	**(0.18)**	**(0.12)**
RW*P*-value	[0.29]	[0.05]	[0.36]	[0.11]
N		*1,324*	*862*	*882*
Adjusted R^2		0.12	0.12	0.13
P-Val: HV v HV+CM		0.68		
Fine Motor	**0.03**	**0.13**	**0.12**	**0.27**
	(0.05)	**(0.05)**	**(0.15)**	**(0.10)**
RW*P*-value	[0.86]	[0.05]	[0.65]	[0.11]
N		*1,324*	*862*	*882*
Adjusted R^2		0.11	0.11	0.09
P-Val: HV v HV+CM		0.02		
Receptive Vocabulary	**0.02**	**-0.04**	**0.01**	**-0.04**
	(0.05)	**(0.05)**	**(0.16)**	**(0.18)**
RW*P*-value	[0.86]	[0.56]	[0.94]	[0.81]
N		*1,290*	*840*	*865*
Adjusted R^2		0.38	0.38	0.38
P-Val: HV v HV+CM		0.18		
Behavior	**0.13**	**0.17**	**0.30**	**0.35**
	(0.07)	**(0.08)**	**(0.22)**	**(0.18)**
RW*P*-value	[0.29]	[0.05]	[0.41]	[0.11]
N		*1,319*	*840*	*877*
Adjusted R^2		0.04	0.02	0.01
P-Val: HV v HV+CM		0.47		

Note: Dependent variables are in rows, and independent variables are in columns. Table reports coefficients with standard errors in parentheses clustered at the community level, and Romano-Wolf step-down *P*-values in brackets below. All outcomes have been internally standardized for age in 2-month increments and are expressed in standard deviation units, with higher scores indicating improved development /better behavior. Specification (1) is the intent-to-treat estimate. Specifications (2) and (3) are two separate models used to estimate the local average treatment effect (LATE) of receiving a home visit in a community assigned to just home visits (Specification 2) or receiving a home visit in a community assigned to home visits with monitoring (Specification 3), using random assignment at the community level as an instrument. Specification (2) is run on the sub-sample of children in communities assigned to control and Home Visits. Specification (3) is run on the sub-sample of children in communities assigned to control and Home Visits + Community Monitoring. All specifications include municipality fixed effects and the following controls: child gender, age in 2-month increments, baseline development scores, mother's years of education, whether or not the child's primary caregiver is the biological mother, household asset index, number of household residents, and an indicator equal to 1 for semi-urban communities, 0 for rural. Romano Wolf Step-down procedures applied to the block of outcomes (language, fine motor, receptive vocabulary, and behavior), with 200 repetitions.

The ITT effects are notable considering the presence of substantial non-compliance with treatment assignment, and we note that our LATE estimates are roughly two times the magnitude of our ITT estimates. For families who were induced to participate in home visits with community monitoring via the randomized community-level assignment, the effect of actually receiving home visits on child development and child behavior are large and significant, 0.28 SD and 0.35 SD. The LATE effect sizes are in fact comparable to the effects of effectiveness studies in LAC (Attanasio et al., 2014).

We fail to detect an effect of either treatment arm on receptive vocabulary. The lack of a finding on receptive vocabulary is somewhat inconsistent considering we did find an effect on the language domain of the Denver-II. However, the PPVT-III scale focuses exclusively on receptive vocabulary, while the Denver-II includes foundational language skills that may be more aligned with the kinds of early stimulation activities promoted by home visitors in this program (e.g., ability to string together words, basic verbalizations, receptive vocabulary, and ability to solve basic cognitive tasks according to verbal instructions).

Lastly, we find positive effects on child behavior in both treatment arms. Since behavior is caregiver reported, these effects could include a mix of improvements in child behavior and improvements in caregivers' ability to effectively respond to children's problem behaviors.

The use of Romano Wolf P- values to control for multiple hypotheses—as is standard in the literature (Carneiro and Ginja, 2014)—did not alter our

findings.[12] Indeed, the ITT effect of Treatment Arm 2 (Home Visits + Community Monitoring) on language, fine motor and behavior are both statistically significant at $p < 0.10$ with, while the LATE estimates are significant at $p < 0.15$ with Romano Wolf P- values.

6.5 Heterogeneity by Age

Considering the existence of sensitive periods for child development (Black, Walker, Fernald, Andersen, DiGirolamo, Lu, McCoy, Fink, Shawar, Shiffman, et al., 2017; Nelson III and Gabard-Durnam, 2020), we examine heterogeneous effects by child age. Given our main finding of treatment effects for Treatment Arm 2 only, we focus our heterogeneity analysis on Treatment Arm 2 (relative to control communities). Figure 3 depicts the magnitude and significance of the treatment effect for three different age groups: the youngest group consists of children aged 4.5 - 4.9 years at follow-up (circle), the middle group includes children aged 5 - 5.4 years at follow-up (triangle), and the older group includes children aged 5.5 - 5.9 at follow-up (square). We find a significant relationship between assignment to Home Visits + Community Monitoring and child's age at follow-up, indicating stronger effects on the child development composite for younger children (see Table F.1 in the appendix for further details). Overall, treatment effects were statistically significant for the youngest two age groups (age 4.5 - 5.4), whereas the treatment effect was insignificant for children who were between 5.5 - 5.9 at follow-up. We also observe suggestive evidence of

[12]Romano Wolf step-down procedures were applied to the block of child development outcomes (language, fine motor, PPVT-III, and behavior). The results from different choices of repetition (100, 200, 500, and 1,000) are all significant at $p<0.10$, we show the results for 200 repetitions.

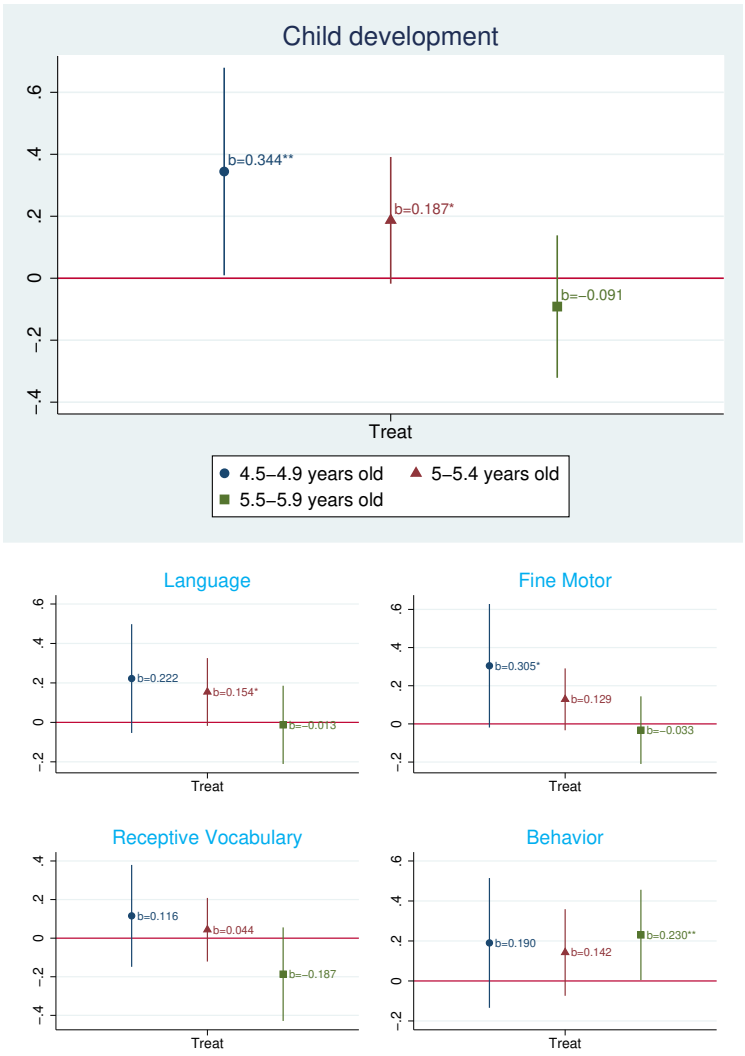
a similar heterogenous treatment effect by age for the other developmental domains, as shown in Figure 3.

These findings suggest that children who were around one to two years of age at the start of the intervention (4.50-4.9 at follow-up) benefited 2.5 times as much as older children. Since there were no clear differences by age in terms of the total number of months home visits received (tables available upon request), we interpret this finding as consistent with the literature showing larger effects of early stimulation interventions for younger children (Behrman, King, Armecin, Duazo, Ghuman, Gultiano, and Lee, 2006; Wint and Janssens, 2008), likely based on sensitive periods for neurodevelopment, due to heightened brain plasticity (Black et al., 2017; Nelson III and Gabard-Durnam, 2020).

6.6 Parenting Behavior

Table 5 shows that while home visits with community monitoring had a positive effect on children's development and behavior, there were no impacts on any of our parenting behavior measures nor maternal depression. This is a puzzling finding considering the role of parental investments in determining child development, and prior evidence that the effects of home visits on child development are fully explained by increases in parental investments (Attanasio et al., 2020). However, our findings are consistent with a recent meta-analysis that found only modest effects of home visits on parenting behavior (Sama-Miller et al., 2017), and there are other examples of interventions that have positively impacted child development without any observed impact on

Figure 3: Effects on Child Development Outcomes by Age Group (6 Months Increment) in Communities Assigned to Home Visits + Community Monitoring



Notes: Sample of analysis is children in control communities and Home Visit + Community Monitoring communities. Age refers to age at follow-up. Significance level at, $^{*}p < .10$,** $p < .05$,*** $p < .01$. Treatment effect for each age group in 6 months increments are estimated using subsample analysis (Blue circle: 4.5-4.9 years old, Red triangle: 5-5.4 years old, and Green square: 5.5 years old).

parenting behavior (Chang et al., 2015).

6.7 Robustness Checks

We examine the robustness of our main results in two ways. First, we show our findings are robust to various specifications of controls. Second, we examine the extent to which the benefits of community monitoring diffused to all community members in Treatment 2 regardless of whether or not they participated in the monitoring meetings.

Our main specification controls for individual and community-level characteristics as well as baseline child development outcome measures. However, as shown in Table G.2, our results are consistent using various specification approaches. Column (1) is the most naive estimate with no control except municipality fixed effects.[13] Column (2) controls for baseline child characteristics. The coefficient of treatment assignment is similar to our main specification (Column 3).

Next, we test the hypothesis that community monitoring improves home visit quality for all children in Treatment 2 communities, regardless of whether or not the caregiver participated in community monitoring meetings. To do so, we regress child development outcomes on Treatment 2 assignment and an indicator variable equal to 1 if the caregiver reported participating in monitoring meetings (restricting the sample to treatment 2 and control communities only).

[13]The base estimation controls for municipality as randomization is stratified at municipality level.

Table 5: Effect of Home Visits on Parenting Behavior

	Intent-to-Treat Effect		**Local Average Treatment Effect**	
	(1)		(2)	(3)
	Home Visits	**Home Visits+ Community Monitoring**	**Home Visits**	**+Home Visits Community Monitoring**
Emotional Support	**-0.00**	**0.01**	**0.11**	**-0.06**
	(0.08)	**(0.08)**	**(0.21)**	**(0.17)**
RWP-value	[0.99]	[0.90]	[0.90]	[0.91]
N		*1,314*	*855*	*874*
Adjusted R^2		0.14	-0.05	-0.04
P-Val: HV v HV+CM		0.88		
Cognitive Stimulation:	**0.07**	**0.06**	**0.11**	**0.15**
Materials	**(0.08)**	**(0.07)**	**(0.21)**	**(0.19)**
RWP-value	[0.62]	[0.73]	[0.90]	[0.74]
N		*1,182*	*768*	*790*
Adjusted R^2		0.27	-0.06	-0.07
P-Val: HV v HV+CM		0.90		
Cognitive Stimulation:	**0.09**	**0.06**	**0.22**	**0.13**
Activities	**(0.07)**	**(0.07)**	**(0.18)**	**(0.16)**
RWP-value	[0.49]	[0.73]	[0.59]	[0.74]
N		*1,181*	*767*	*789*
Adjusted R^2		0.17	-0.06	-0.04
P-Val: HV v HV+CM		0.61		
No Maternal	**0.00**	**-0.03**	**0.01**	**-0.06**
Depression (CES-D)	**(0.06)**	**(0.06)**	**(0.18)**	**(0.12)**
RWP-value	[0.99]	[0.82]	[0.93]	[0.91]
N		*1,310*	*855*	*870*
Adjusted R^2		0.09	0.02	-0.02
P-Val: HV v HV+CM		0.54		

Note: Dependent variables are in rows, and independent variables are in columns. Table reports coefficients with standard errors in parentheses clustered at the community level, and Romano-Wolf step-down P-values in brackets below, when applicable. All outcomes are internally standardized and expressed in standard deviation units, with higher scores indicating better parenting /less depression. Specification (1) is the intent-to-treat (ITT) estimate. Responsive parenting and harsh discipline practices are direct observations, corresponding to the caregiver's interactions with the target child during the interview, and are thus modeled using the child-level sample. CES-D corresponds is measured via caregiver self-report, and is thus measured using the caregiver dataset. Results indicate the effect of residing in communities randomly assigned to Treatment Arm 1 (Home Visits) or Treatment Arm 2 (Home visits + Community Monitoring). Specifications (2) and (3) are two separate models used to estimate the local average treatment effect (LATE) of receiving a home visit in a community assigned to just Home Visits (Specification 2) or receiving a home visit in a community assigned to Home Visits with Community Monitoring (Specification 3), using random assignment at the community level as an instrument. Specification (2) is run on the sub-sample of children/ caregivers in communities assigned to control and Home Visits. Specification (3) is run on the sub-sample of children/ caregivers in communities assigned to control and Home Visits + Community Monitoring. All specifications include municipality fixed effects and the following controls: mother's years of education, whether or not the child's primary caregiver is the biological mother, baseline parenting behaviors, household asset index, number of household residents, and an indicator equal to 1 for semi-urban communities, 0 for rural. Romano Wolf Step-down procedures applied to the block of parenting outcomes (emotional support, both measures of cognitive stimulation, and CES-D), with 200 repetitions.

We find that our results are mostly robust in terms of magnitude, but the significance disappears for the child development composite score and language domain, which may be due to lack of statistical power as we restrict our analysis to the Treatment 2 and control groups (Table G.3). We interpret this as suggestive evidence in support of the hypothesis that the effects of community monitoring are not conditional on whether or not the caregiver attends meetings – i.e., community monitoring construed broad benefits for the full community.

7 Conclusion

We estimate the effects of a large-scale, low-cost ($100/child/year) home visit parenting program implemented by the Nicaraguan Government. Four years after program launch, we find a positive ITT effect of 0.13 SD on child development, but only in communities assigned to home visits with community monitoring. There was also a positive ITT effect of both treatment arms on caregiver-reported child behavior, which likely includes some mix of actual improvements in child behavior and caregivers' ability to respond effectively to their child's behavior. The magnitude of our ITT findings on child development are consistent with estimates from large-scale home visit programs in the United States (e.g. Sama-Miller et al., 2017). Given noncompliance across treatment arms, we also examine LATE effects and find effects on child development and behavior roughly twice as large as the ITT estimate. The impacts are much stronger among younger children, consistent with the evi-

dence of critical periods in child development (Black et al., 2017; Nelson III and Gabard-Durnam, 2020).

We find no evidence that the intervention improved parenting behavior or reduced maternal depression. This finding is consistent with the somewhat mixed evidence to date regarding the effects of home visits on parenting outcomes. In this case, it may be that home visits with community monitoring affected a unique set of parenting skills, not captured by the standard measures of parenting that are used in the early childhood literature and that we used here (i.e., the HOME). By bringing parents together for bimonthly meetings, community monitoring may have not only ensured the quality of home visits but also strengthened parental self-efficacy, connected parents to social resources in their community, and supported the development of strong family peer networks. Such an interpretation makes sense considering our finding that community monitoring was necessary for impact on child development. It may be worth identifying and developing new tools to incorporate different aspects of parenting behaviors, especially for interventions that involve innovations such as community monitoring.

Using a unique dataset of home visit observations, we show that home visits were higher quality in communities assigned to community monitoring. Relative to home visits with "status-quo" monitoring, home visits with community monitoring achieved higher quality ratings on all domains—including higher quality home visitor-caregiver relationships and visit content—two aspects of quality that are especially difficult to ensure at scale (Leer and Lopez Boo, 2018). In addition, in communities where participation in monitoring was

higher, home visit quality was higher. Conditional on assignment to community monitoring, the number of monitoring meetings held per community was positively associated with the quality of home visits. These findings are consistent with prior evidence that community monitoring works best when citizens have real authority (Fox, 2015). Community monitoring in this case empowered citizens to directly influence home visit quality, vis-a-vis their participation in the development of bimonthly Action Plans. In addition, because the home visitors were not financially rewarded for high performance, our findings add to the evidence that pro-social incentives can serve as an effective accountability mechanism for service providers (Ashraf, Bandiera, and Lee, 2015).

However, whereas prior studies have focused on researcher-led community monitoring pilots, we show that community monitoring can be effective within the context of a large-scale government-run program. Our study is also the first to use community monitoring for an early childhood development program. Home visits provide a particularly unique case study for community monitoring. Unlike health care or basic education, both of which enjoy strong citizen demand and some a priori notion of what service delivery *should* look like, home visits in most LMIC contexts are new to communities and not necessarily in high demand before they are launched.

It is worth noting that because both treatment arms received "status-quo" monitoring, we cannot say to what extent community monitoring would work in the complete absence of institutional monitoring. However, such a scenario seems unlikely. Governments generally have some incentive to conduct

their own monitoring, and the key question is not *whether* to monitor, but rather *how to provide effective monitoring*, especially to remote communities in resource constrained settings.

This paper is among the first to provide experimental evidence of a government-run home visit program implemented at scale in a high poverty country. Our findings highlight the importance of implementation quality for impact on child development, and demonstrate that community monitoring can be a cost effective way to ensure quality when expanding to the national level. However, community monitoring may have influenced child development through a different set of parenting mechanisms than those typically measured in the context of early childhood interventions in LMICs. This finding problematizes the underlying assumptions regarding the home visit theory of change that will be critical to explore going forward, as countries across the globe seek to develop high quality early childhood development services.

References

Aboud, F. E., D. R. Singla, M. I. Nahil, and I. Borisova (2013). Effectiveness of a parenting program in Bangladesh to address early childhood health, growth and development. *Social Science & Medicine 97*, 250–258.

Aboud, F. E. and A. K. Yousafzai (2015). Global health and development in early childhood. *Annual Review of Psychology 66*, 433–457.

Achenbach, T. M. and C. S. Edelbrock (1981). Behavioral problems and competencies reported by parents of normal and disturbed children aged four through sixteen. *Monographs of the Society for Research in Child Development*, 1–82.

Ahmad, J., S. Commins, S. Devarajan, D. Filmer, J. Hammer, L. Pritchett, R. Reinikka, S. Shah, and A. Soucat (2003). World Development Report 2004: Making services work for poor people. Technical report, The World Bank.

Andrew, A., O. Attanasio, E. Fitzsimons, S. M. Grantham-McGregor, C. Meghir, and M. Rubio-Codina (2018). Impacts 2 years after a scalable early childhood development intervention to increase psychosocial stimulation in the home: A follow-up of a cluster randomised controlled trial in Colombia. *PLoS Medicine 15*(4).

Araujo, C., M. Dormal, and M. Rubio-Codina (2018). Quality of parenting programs and child development outcomes: The case of Peru's Cuna Mas. Technical Report IDB-WP-951.

Arnove, R. F. (1981). The Nicaraguan national literacy crusade of 1980. *Comparative Education Review 25*(2), 244–260.

Ashraf, N., O. Bandiera, and S. S. Lee (2015). Do-gooders and go-getters: career incentives, selection, and performance in public service delivery. *Harvard Business School*.

Attanasio, O., S. Cattan, E. Fitzsimons, C. Meghir, and M. Rubio-Codina (2020). Estimating the production function for human capital: results from a randomized controlled trial in Colombia. *American Economic Review 110*(1), 48–85.

Attanasio, O., C. Fernández, E. O. Fitzsimons, S. M. Grantham-McGregor, C. Meghir, and M. Rubio-Codina (2014). Using the infrastructure of a conditional cash transfer program to deliver a scalable integrated early child development program in Colombia: Cluster randomized controlled trial. *BMJ (Online) 349*(September), 1–12.

Baker-Henningham, H., C. Powell, S. Walker, and S. M. Grantham-McGregor (2003). Mothers of undernourished jamaican children have poorer psychosocial functioning and this is associated with stimulation provided in the home. *European Journal of Clinical Nutrition 57*(6), 786.

Banerjee, A. V., R. Banerji, E. Duflo, R. Glennerster, and S. Khemani (2010). Pitfalls of participatory programs: Evidence from a randomized evaluation in education in India. *American Economic Journal: Economic Policy 2*(1), 1–30.

Banerjee, A. V., A. Deaton, and E. Duflo (2004). Wealth, health, and health services in rural Rajasthan. *American Economic Review 94*(2), 326–330.

Behrman, J. R., E. M. King, G. Armecin, P. Duazo, S. Ghuman, S. Gultiano, and N. Lee (2006). *Early childhood development through an integrated*

program: Evidence from the Philippines. The World Bank.

Berlinski, S. and N. Schady (Eds.) (2016). *The Early Years, Child Well-Being and the Role of Public Policy*. Inter-American Development Bank.

Björkman, M. and J. Svensson (2009). Power to the people: evidence from a randomized field experiment on community-based monitoring in uganda. *The Quarterly Journal of Economics 124*(2), 735–769.

Björkman Nyqvist, M., D. De Walque, and J. Svensson (2017). Experimental evidence on the long-run impact of community-based monitoring. *American Economic Journal: Applied Economics 9*(1), 33–69.

Black, M. M., S. P. Walker, L. C. Fernald, C. T. Andersen, A. M. DiGirolamo, C. Lu, D. C. McCoy, G. Fink, Y. R. Shawar, J. Shiffman, et al. (2017). Early childhood development coming of age: science through the life course. *The Lancet 389*(10064), 77–90.

Bradley, R. H., R. Chazan-Cohen, and H. Raikes (2009). The impact of Early Head Start on school readiness: New looks. *Early Education and Development 20*(6), 883–892.

Caldwell, B. M., R. H. Bradley, et al. (1984). *Home observation for measurement of the environment*. University of Arkansas at Little Rock Little Rock.

Carneiro, P. and R. 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.

Chang, S. M., S. M. Grantham-McGregor, C. A. Powell, M. Vera-Hernández, F. Lopez Boo, H. Baker-Henningham, and S. P. Walker (2015). Integrat-

ing a parenting intervention with routine primary health care: a cluster randomized trial. *Pediatrics 136*(2), 272–280.

Cunha, F. and J. J. Heckman (2008). Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation. *Journal of Human Resources 43*(4), 738–782.

Cunha, F., J. J. Heckman, and S. M. Schennach (2010). Estimating the Technology of Cognitive and Noncognitive Skill Formation. *Econometrica 78*(3), 883–931.

Duflo, E., P. Dupas, and M. Kremer (2015). School governance, teacher incentives, and pupil-teacher ratios: Experimental evidence from Kenyan primary schools. *Journal of Public Economics 123*, 92–110.

Duncan, G. J. and K. Magnuson (2013). Investing in preschool programs. *Journal of Economic Perspectives 27*(2), 109–32.

Dunn, L. M. and L. M. Dunn (1981). *Peabody Picture Vocabulary Test Revised*. American Guidance Service, Incorporated.

Fernald, A., V. A. Marchman, and A. Weisleder (2013). SES differences in language processing skill and vocabulary are evident at 18 months. *Developmental Science 16*(2), 234–248.

Fox, J. A. (2015). Social accountability: What does the evidence really say? *World Development 72*, 346–361.

Frankenburg, W. K., J. Dodds, P. Archer, H. Shapiro, and B. Bresnick (1992). The Denver II: a major revision and restandardization of the Denver Developmental Screening Test. *Pediatrics 89*(1), 91–97.

Gertler, P., J. Heckman, R. Pinto, A. Zanolini, C. Vermeersch, S. Walker,

S. M. Chang, and S. M. Grantham-McGregor (2014). Labor market returns to an early childhood stimulation intervention in jamaica. *Science 344*(6187), 998–1001.

Grantham-McGregor, S. M., C. Powell, S. Walker, S. Chang, and P. Fletcher (1994). The long-term follow-up of severely malnourished children who participated in an intervention program. *Child development 65*(2), 428–439.

Heckman, J. J. (2008). Schools, skills, and synapses. *Economic Inquiry 46*(3), 289–324.

INIDE (2010). Estudio Nacional de las Condiciones de Vida y Primera Versión del índice de Bienestar de la Niñez y la Adolescencia Nicaraguense. Technical report, Gobierno de Nicaragua.

INIDE (2014). Encuesta de Medición de Nivel de Vida 2014. Technical report, Gobierno de Nicaragua.

Leer, J. and F. Lopez Boo (2018). Assessing the quality of home visit parenting programs in latin america and the caribbean. *Early Child Development and Care 189*(13), 2183–2196.

Leer, J., F. Lopez Boo, A. Perez Expósito, and C. Powell (2016). A snapshot on the quality of seven home visit parenting programs in Latin america and the Caribbean.

Lim, G. Y., W. W. Tam, Y. Lu, C. S. Ho, M. W. Zhang, and R. C. Ho (2018). Prevalence of depression in the community from 30 countries between 1994 and 2014. *Scientific Reports 8*(1), 1–10.

Lopez Boo, F., M. Cubides Mateus, R. Sorio, G. Garibotto, and C. Berón

(2019). Measuring the quality of the home environment of young children in Uruguay: Socioeconomic gradients in the HOME inventory.

Love, J., R. Chazan-Cohen, H. Raikes, and J. Brooks-Gunn (2013, feb). What makes a difference: Early Head Start evaluation findings in a developmental context. *Monographs of the Society for Research in Child Development 78*(1), vii–viii, 1–173.

MIFAN (2019). Informe de implementación paippi. Technical report, Gobierno de Nicaragua.

Nelson III, C. A. and L. J. Gabard-Durnam (2020). Early adversity and critical periods: Neurodevelopmental consequences of violating the expectable environment. *Trends in Neurosciences 43*(3), 133–143.

Olds, D. L., C. R. Henderson, R. Cole, J. Eckenrode, H. Kitzman, D. Luckey, L. Pettitt, K. Sidora, P. Morris, and J. Powers (1998). Long-term effects of nurse home visitation on children's criminal and antisocial behavior: 15-year follow-up of a randomized controlled trial. *JAMA 280*(14), 1238–1244.

Olds, D. L., C. R. Henderson, H. Kitzman, J. Eckenrode, R. Cole, and R. Tatelbaum (1999). Prenatal and infancy home visitation by nurses: recent findings. *Future of Children 9*(1), 44–65 22p.

Olds, D. L., H. Kitzman, R. Cole, J. Robinson, K. Sidora, D. W. Luckey, C. R. Henderson, C. Hanks, J. Bondy, and J. Holmberg (2004). Effects of Nurse Home-Visiting on Maternal Life Course and Child Development: Age 6 Follow-Up Results of a Randomized Trial. *Pediatrics 114*(6), 1550–1559.

Olken, B. A. (2007). Monitoring corruption: evidence from a field experiment

in indonesia. *Journal of Political Economy 115*(2), 200–249.

Olken, B. A. and R. Pande (2012). Corruption in Developing Countries. *Annual Review of Economics 4*(1), 479–509.

Radloff, L. S. (1977). The CES-D scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement 1*(3), 385–401.

Richter, L. M., B. Daelmans, J. Lombardi, J. Heymann, F. Lopez Boo, J. R. Behrman, C. Lu, J. E. Lucas, R. Perez-Escamilla, T. Dua, et al. (2017). Investing in the foundation of sustainable development: pathways to scale up for early childhood development. *The Lancet 389*(10064), 103–118.

Rubio-Codina, M., M. C. Araujo, O. Attanasio, P. Muñoz, and S. M. Grantham-McGregor (2016). Concurrent validity and feasibility of short tests currently used to measure early childhood development in large scale studies. *PLoS One 11*(8).

Sama-Miller, E., L. Akers, A. Mraz-Esposito, S. Avellar, D. Paulsell, P. Del Grosso, et al. (2017). Home visiting programs: Reviewing evidence of effectiveness (brief). Technical report, Mathematica Policy Research.

Schady, N., J. Behrman, M. C. Araujo, R. Azuero, R. Bernal, D. Bravo, F. Lopez-Boo, K. Macours, D. Marshall, C. Paxson, and R. J. Vakis (2015). Wealth gradients in early childhood cognitive development in five Latin American countries. *Journal of Human Resources 50*(2), 446–463.

Vogel, C. A., P. Caronongan, J. Thomas, E. Bandel, Y. Xue, J. Henke, N. Aikens, K. Boller, and L. Murphy (2015). Toddlers in early head start: A portrait of 2-year-olds, their families, and the programs serving them. volume 1: Age 2 report. opre report 2015-10. *Administration for Children & Families*.

Wachs, T. D., M. M. Black, and P. L. Engle (2009). Maternal depression: a global threat to children's health, development, and behavior and to human rights. *Child Development Perspectives 3*(1), 51–59.

Waddington, H., A. Sonnenfeld, J. Finetti, M. Gaarder, D. John, and J. Stevenson (2019). Does incorporating participation and accountability improve development outcomes? meta-analysis and framework synthesis. *3ie Systemic Review 43*.

Walker, S. P., S. M. Chang, M. Vera-Hernandez, and S. Grantham-McGregor (2011). Early Childhood Stimulation Benefits Adult Competence and Reduces Violent Behavior. *Pediatrics 127*(5), 849–857.

Wint, E. and W. Janssens (2008). *Impact Study of the Roving Caregiver Programme: Research findings*.

Yousafzai, A. K., M. A. Rasheed, A. Rizvi, R. Armstrong, and Z. A. Bhutta (2014). Effect of integrated responsive stimulation and nutrition interventions in the Lady Health Worker programme in Pakistan on child development, growth, and health outcomes: A cluster-randomised factorial effectiveness trial. *The Lancet 384*(9950), 1282–1293.

Zill, N. and J. L. Peterson (1986). Behavior Problems Index. *Washington, DC: Child Trends*.

8 Acknowledgements

The authors thank the team of specialists from the Inter-American Development Bank, led by Emma Sanchez-Monin, Meri Helleranta, and Leonardo Pinzon, who oversaw the implementation of the program in Nicaragua, as well as technical staff and officials at MIFAN, for their invaluable support. We thank FIDEG-Nicaragua and NORC at the University of Chicago for data collection, Deiby Mayaris Cubides Mateus for research assistance, and Pablo Celhay for technical guidance and feedback on initial drafts. Anna Gassman-Pines, Lisa Gennetian, Sarah Reynolds, Frances Aboud, Joan Lombardi, Raquel Bernal, Jakob Svensson, and Hiro Yoshikawa also provided invaluable comments on early drafts, as did participants at the Duke University Sanford School of Public Policy Graduate Student Research Seminar (November 2019). All errors and omissions are our own.

Appendices

A External Validity

Table A.1: Home Visit Sample and National Comparison

	(1) Home-visit population	(2) National population	(3) Min	(4) Max
Household				
No. of children younger than 2 years old	1.09	1.07	1	4
Access to piped water	0.63	0.57	0	1
Improved sanitation	0.66	0.59	0	1
Electricity	0.80	0.74	0	1
Household head				
Female	0.34	0.32	0	1
Characteristics of children younger than 2				
Lives with biological mother	0.99	0.98	0	1
Lives with biological father	0.73	0.73	0	1
Age of biological mother	25.56	25.78	13	80
Age of biological father	31	31	15	90
Mother years of education	7.75	6.73	0	17
Father years of education	6.98	6.31	0	17
Observations	1322	3292	4614	4614

Sources: Home visit baseline data (2013) and Nicaragua DHS 2011.

Notes: The DHS data applies sample weight for national average, and the analysis is retricted to households with children younger than 6 years old. Access to piped water includes piped water inside the house or compound. Improved sanitation refers to toilet and latrine (excusado o letrina) in the house. The DHS data were collected 4-5 years prior to the Home Visit baseline sample, which may explain why the Home Visit sample appears to be somewhat more advantaged than the national comparison in terms of household characteristics and years of education.

The selection process for the Home Visit communities was as follows: First, the 37 municipalities with the highest rates of extreme poverty were selected using the Government's official measure of Unmet Basic Needs. The Unmet

Basic Needs index, or *Índice de Necesidades Básicas Insatisfechas* is measured at the household level, and refers to the number of basic needs not met (of 5). All municipalities with an index of 0.20 or higher (indicating that households on average experienced at least two of five unmet basic needs) were selected. Within these municipalities, 1,402 communities were identified as eligible for the Home Visit Parenting program, using the same threshold of Unmet Basic Needs (Communities are defined according to Nicaraguan census designations, they can be considered akin to small villages or towns). The 210 communities, in 33 municipalities, were selected for the experimental evaluation based on the following criteria: (1) communities had to have at least 35 children younger than six, (2) the Ministry of the Family had to have some presence in the community (e.g., existing community development workers who could be trained to be home visitors), and (3) community selection was stratified at the municipal level.

B Intervention Description

The home visit curriculum focuses on language, cognitive, affective, and socio-emotional development, and emphasizes the importance of affectionate, consistent, warm caregiving and respect for children's agency and autonomy. The curriculum was developed by local early childhood development practitioners and pedagogical experts, drawing primarily from the Government's early childhood education strategy, with some inspiration from other home visit programs in the LAC region. Each visit included the following sequence of activities:

- (i) reflection and review of activities from the previous visit,
- (ii) the provision of age appropriate information regarding developmental milestones and processes,
- (iii) a demonstration from the home visitor of each early stimulation activities with the child for the caregiver to observe (roughly 1-5 activities per visit, depending on the age and developmental strengths/ needs of the child),
- (iv) time for the caregiver to implement each activity with the child herself, in the presence of the home visitor, and
- (v) time for the caregiver and home visitor to reflect on the activities, discuss, and commit to ways the caregiver will continue these activities on her own with the child in the weeks leading up to the next visit.

Home visitors were expected to have a schedule planned out for each visit, with activities and objectives clearly defined.

Home visitor training was conducted in a cascade manner. First, central authorities from the Ministry of the Family were trained by a team of local and international experts during a week long diploma course in early childhood development and home visit methodology. The central authorities then rolled out this training to technical specialists at the municipal level, and the technical specialists trained the home visitors.

Figure B.1: Home Visit Poster

Notes: Families in Treatment Arm 2 (Home Visits + Community Monitoring) received this poster. The top portion, *Educamos, cuidamos y amamos a nuestros hijos e hijas* (Educate, care for, and love our sons and daughters) includes pictures and text description of ways families can support their young children's healthy development. The bottom pink panel, *Tu opinión nos importa!* (Your opinion matters to us!) is the feedback survey for community monitoring. Families were encouraged to record their satisfaction with the home visits in terms of visit duration, material, content, and quality of the interaction with the home visitor. Every two months, youth volunteers provided families with a new feedback survey. Families in Treatment Arm 1 (Home Visits) received just the top portion of this poster, without the feedback survey.

Figure B.2: Community Monitoring Poster

Notes: In addition to the poster shown above (Figure B.1), families in Treatment Arm 2 (Home Visits + Community Monitoring) received this poster, which describes the community monitoring processes, as follows: (A) fill out the feedback survey and give it to the youth volunteer, or (B) give your feedback verbally to members of the *Gabinete*, and (C) meet every two months with the *Gabinete*, youth volunteer, home visitors, and families to review feedback, answer questions and concerns, and develop a Community Action Plan.

C Sample Attrition

The full baseline sample consists of 9,700 children between the ages of six months and six years. At follow-up, data collection was limited to children who were younger than 36 months at baseline, because children older than 36 months at baseline would not have received home visits.[14] Based on this criteria, 5,145 children were identified as eligible (i.e., younger than 36 months at baseline). The survey team randomly selected 4,200 children of those 5,145 (stratified by community) to comprise the follow-up sampling frame. 945 additional children were randomly selected as replacements (stratified by community), such that if an originally sampled child was not home or unavailable after four attempts, was deceased, was sick, refused the interview, or had moved to a community outside of the study sample, s/he could be replaced by a child from the replacement list who was also included in the baseline sample. At follow-up, complete assessments were collected from 3,808 children (3,297 of which were from the original list of 4,200 children, and 512 of which were from the list of randomly selected replacements). This represents a 9.3% attrition rate.[15] However, our analysis of the probability of attrition between treatment groups (Table C.1) and based on observable characteristics at baseline C.2) is limited to 139 "attritors" who we can identify as attritors based on their unique ID. Unfortunately, the survey team did not provide the unique identi-

[14]The Ministry of Family had initially considered conducting fortnightly home visits for children up to age six, but in response to budget and logistical constraints, this idea was abandoned before program launch and only children up to age three years old ended up being eligible for fortnightly home visits.

[15]In some communities it was not possible to complete the sample (i.e., there were not enough replacement children available), hence the total sample is less than 4,200.

fiers for the full list of 4,200 children and 945 replacement children, so we are unable to identify these children in the baseline dataset.

Table C.1: Probability of Attrition Based on Treatment Status

	Attritor
T1 (Home Visits)	0.01
	(0.01)
T2 (Home Visits and Community Monitoring)	0.01
	(0.01)
Observations	3,947
Adjusted R^2	-0.00
Mean Attrition Rate	0.04

Regressions of attrition variable on an indicator for treatment assignment (control group is the reference group). Attrition status is identified for 3,947 children interviewed at baseline and matched to follow-up information regarding completion of the follow-up interview. Of those, 139 are classified as attritors (not interviewed at follow-up). Standard errors clustered at the community level in parentheses, $^{*}p < .10,^{**}p < .05,^{***}p < .01$.

Table C.2: Relation Between Observable Baseline Characteristics and Attrition Status, by Treatment Group

	Male	Age in years	Semi-urban	Household asset index	Biological mother is primary caregiver	Mother's Education (years)	Caregiver depression (CES-D)
Attritor	0.00	-0.12	0.01	-0.13	-0.19	-1.54	-1.01
	(0.07)	(0.15)	(0.08)	(0.14)	(0.07)	(0.67)	(1.03)
Attritor*T1	0.05	-0.07	0.11	0.12	0.14	0.87	1.47
	(0.10)	(0.19)	(0.11)	(0.23)	(0.09)	(0.97)	(1.69)
Attritor*T2	-0.06	-0.09	0.04	-0.00	0.16	0.80	-0.22
	(0.10)	(0.20)	(0.10)	(0.18)	(0.08)	(0.94)	(1.57)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3,947	3,947	3,947	3,721	3,752	3,749	3,745
Baseline Mean	0.51	1.02	0.71	0.03	0.90	6.98	7.15

	Composite child development (z-score)	Social Indiv (z-score)	Language (z-score)	Fine Motor (z-score)	Emotional Support (z-score)	Cognitive Stimulation (z-score)
Attritor	0.19	0.07	0.18	0.11	0.40	-0.16
	(0.20)	(0.16)	(0.23)	(0.16)	(0.40)	(0.21)
Attritor*T1	-0.45	-0.20	-0.48	-0.21	-0.93	0.01
	(0.31)	(0.25)	(0.30)	(0.24)	(0.53)	(0.26)
Attritor*T2	-0.41	-0.04	-0.50	-0.25	-0.21	0.15
	(0.29)	(0.25)	(0.29)	(0.24)	(0.46)	(0.26)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3,391	3,719	3,556	3,408	2,885	2,686
Baseline Mean	-0.00	0.03	0.03	0.01	8.90	1.77

Regressions of dependent variable on an indicator variable for attritors, and the interaction between attritors and treatments (treatment dummies controlled). T1 refers to communities assigned to Home Visits, T2 refers to communities assigned to Home Visits and Community Monitoring. Attrition status is identified for 3,947 children interviewed at baseline and matched to follow-up information regarding completion of the follow-up interview. Of those, 139 are classified as attritors (not interviewed at follow-up). Standard errors clustered at the community level in parentheses.

D Description of Child Development and Parenting Measures

Baseline data were collected by a team of 50 enumerators, all of whom had prior experience collecting household survey data. Baseline enumerator training was conducted over 20 days and included two practice tests in the field as well as a refresher training. Follow-up data were collected by a team of 28 enumerators, all of whom had prior experience applying the Denver, PPVT, and /or similar child development instruments. Training at follow-up was conducted over 12 days, and included two practice tests in the field (in communities not included in the experimental evaluation). The same survey firm led both rounds of data collection.

Child Development Measures

Denver II
(collected at baseline and follow-up)
Source: Frankenburg, Dodds, Archer, Shapiro and Bresnick (1992).
Description: The Denver II measures fine motor, gross motor, social personal and language. Consistent with prior literature showing socioeconomic gradients in gross motor are minimal, and the fact that the home visit curriculum did not include gross motor exercises, we do not include gross motor in our analysis. We also exclude social personal at follow-up given administration errors (erroneous interpretation of several items that led to a binomial distribution, which factor analysis showed was minimally related to the underlying construct of child development used in the composite indicator). The Denver-II includes 125 items total, but the total number of items depends on the child's age and ability. The Denver II is designed for children aged 1 month to 5.99 years.
Administration: The Denver items are scored based on direct assessment of the child's performance in a series of tasks, administered by the enumerator with support from the primary caregiver.

For young children only, some items are caregiver-reported. The first item to administer is determined by the child's age in months, and administration continues item by item until a child fails or refuses to answer three consecutive items.
Scoring: We compute a continuous score equal to the sum of the of the items successfully completed (in line with Rubio-Codina et al. (2016), we assume that the child can successfully complete all of the items that precede the starting item according to the child's age group). We then standardize scores within our sample in two-month age intervals (i.e., we group children according to age and then compute z-scores within age bins of two months).[16]

PPVT-III
(collected at follow-up only)
Source: Dunn and Dunn (1981)
Description: The Peabody Picture Vocabulary Test-III (PPVT) measures receptive vocabulary. The words used in the PPVT were adapted to the Nicaraguan context with the permission of the publisher. Similar to the Denver II, the are192 items total, but the total number of items administered depends on the child's age and ability. The PPVT is designed for children aged 2.5 years and older.
Administration: The PPVT is administered via direct assessment. Children are shown a series of images and asked to point to the image as the assessor reads the corresponding word aloud.
Scoring: We compute a continuous score equal to the sum of words successfully identified, and standardize within two-month age intervals. We exclude from our analysis any cases with fewer than 100 children per two-month interval.

Behavioral Problem Index
(collected at follow-up only)
Source: Adapted from Zill and Peterson (1986), with items derived from Achenbach and Edelbrock (1981).
Description: The Behavioral Problem Index (BPI) consists of 25 caregiver-reported items that measure a combination of "problem behaviors" (internalizing and externalizing) behaviors as well as positive behaviors (e.g. "considers others' feelings," and "offers help when others are hurt"). The BPI is intended for use with children aged three to eight years old.

[16]For all three child development measures (Denver II, PPVT-III, and BPI), we exclude any cases with fewer than 100 children per two-month interval from our analysis.

Administration: Enumerators read aloud each item and caregivers say whether the behavior is often, sometimes, or never true of the target child's behavior in the last six months.
Scoring: Before scoring, individual items are recoded such that a score of "1 (never)" becomes "0" and a score of "2" (sometimes) and '3" (often) becomes 1. We then reverse code negative (problem) items and compute a continuous score, with higher scores indicating "better" behavior, and standardize within two-month age intervals.

We estimate aggregate child development via factor analysis with oblique quartimin rotation, using the within-sample age standardized scores for each Denver domain and receptive vocabulary (PPVT). Our findings are also consistent using a measurement systems approach (Attanasio et al., 2020; Cunha and Heckman, 2008; Cunha et al., 2010) (tables available upon request).

Parenting Measures

Home Observation for Measurement of the Environment (HOME) (collected at baseline and follow-up)
Source: Adapted from Caldwell et al. (1984).
Description: We used nine items from the HOME scale based on interviewer observation, focusing on the quality of parent-child interactions.
Administration: Enumerators completed the HOME scale on their own at the conclusion of each interview, the questions were in reference to the behavior of the caregiver towards the sample child throughout the interview.
Scoring: Each item was scored in a binary (yes/no) fashion (e.g., "during the interview, did the caregiver respond to the child's verbalizations?").

Center for Epidemiological Studies Depression Scale (CES-D) (collected at baseline and follow-up) *Source*: Radloff (1977).
Description: We used a 12 item version of the CES-D that had been previously validated in Latin American contexts. The CES-D asks caregivers to rate how many days in the past week (7 days) they experienced symptoms associated with depression, such as loss of appetite, trouble sleeping, feeling lonely or hopeless.

Administration: Enumerators read aloud each symptom and caregivers responded with the number of days they felt that way.
Scoring: Per the CES-D guidelines, the number of days are categorized into a Likert scale ranging from 0 to 3, where 0 = rarely or none of the time (less than 1 day in the past week), 1 = some or a little of the time (1-2 days), 2 = occasionally or a moderate amount of time (3-4 days), and 3 = most of the time (5-7 days). We then sum all items to create an index score ranging from 0 (no depressive symptoms) to 36. We classify caregivers as at risk of clinical depression if they have a score of 10 or above, for illustrative purposes only, as reported in Table 1. Our finding that 28% of caregivers were at risk of clinical depression at baseline is in line with estimates of the global prevalence of clinical depression being higher among women (14%, relative to 13% overall) and in countries with a medium human development index (24%)(Lim, Tam, Lu, Ho, Zhang, and Ho, 2018). For all analyses, we use the within sample standardized score of the CES-D index score, reverse scored so that higher scores indicate less depressive symptoms.

E Home Visit Observation Tool and Dataset

Home visit quality observations were conducted by technical specialists from the municipal offices of the Ministry of the Family as part of the "status-quo" institutional monitoring, using a home visit observation rubric based on based on Leer et al. (2016) and adapted by the Ministry of Family. Table E.1 describes the items used for each category.

The observations were scored on paper in the field by municipal specialists and subsequently digitized by staff at the Ministry of the Family. The dataset includes a total of 4,913 unique observations collected over the roughly four years of program implementation (see Table E.2). 44 (of 70) Home Visit communities and 51 (of 70) Home Visit + Community Monitoring communities are represented. There are a total of 466 unique home visitors and 86 technical specialists represented in the dataset. On average, each home visitor was observed on 18 different occasions.

Data on the quality of home visit implementation is rare, especially in a developing country context. To the best of our knowledge, this study is the first to use observational data on home visit quality to complement experimental impact estimates on child and parent outcomes in a high poverty country.

These strengths not withstanding, there are several limitations to note. Fist, it is possible that more observations were collected but not digitized. Further, given non-random timing of home visit observations, potential non-random selection by municipal technical specialists of home visitors and families to observe, and the fact that we do not observe home visits in all treatment communities, the home visit dataset is not representative of the universe of

home visits conducted throughout program implementation.

As described in Table 3, we use the home visit observation dataset to estimate (1) the intent to treat effect of assignment to Home Visits + Community Monitoring on the quality of home visit implementation, and to examine (2) the extent to which greater engagement in community monitoring processes (as reported by caregivers) predicts improved home visit quality. We consider this analysis to be suggestive evidence only. Considering the data limitations described above, we consider this suggestive evidence.

Table E.1: Home Visit Observation Rubric

Home-visitor-caregiver relationship	**Visit materials and preparation**
(1) HV asks about child's advances	(1) HV had prepared for visit
(2) HV encourages caregiver questions and opinions	(2) HV encourages use of readily available materials
(3) HV praises caregiver	(3) HV uses the program booklet throughout the visit
(4) HV involves caregiver in visit activities	
(5) HV has strong communication with caregiver	
(6) HV asks caregiver for feedback on the visit	
(7) HV plans activities with caregiver for weeks between visits	
Visit Content	**Focus on caregiver**
(1) HV encourages use of community services	(1) Visit focuses on the caregiver
(2) HV models activities for families	
(3) HV explains activities in relation to child development	
(4) HV encourages play, and use of cognitive stimulation activities	
(5) HV promotes caregiver-child affection and warmth	
(6) HV emphasizes ways caregivers can support language development	
(7) Visit activities are appropriate for child's age	

Note: Each item is scored on a 4-point Likert scale. We sum all items for each category, and then standardize scores by year.

Table E.2: Number of home visit observations per year, by treatment assignment

	2014	2015	2016	2017	Total
Treatment Assignment					
Home Visits	79	679	973	399	2,130
Home Visits + Community Monitoring	204	947	1,193	518	2,783
Total	235	1,626	2,166	917	4,913

Note: Each observation represents one home visit that was observed and scored by a municipal technical specialist.

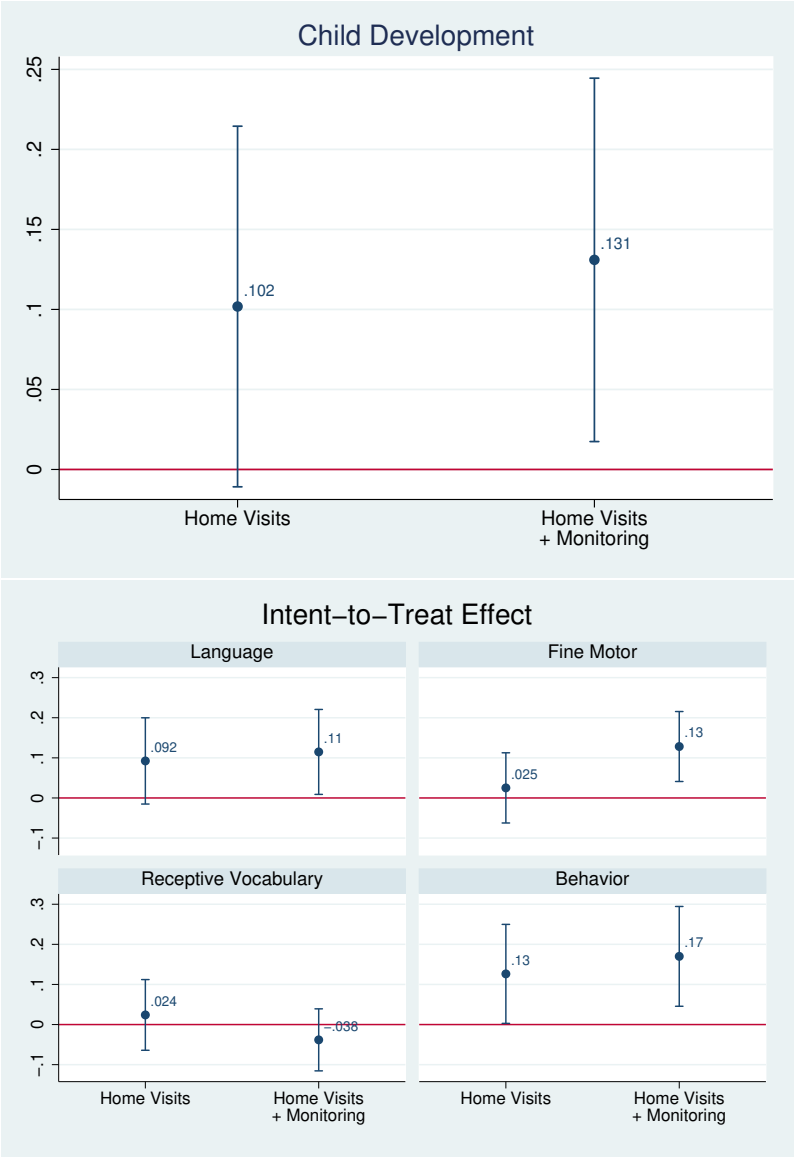
F Treatment Effects on Child Development: Additional Analyses

Table F.1: ITT Effect of Home Visits with Community Monitoring, by Age

	Child Development Composite	Language	Fine Motor	PPVT	Behavior
Treat 2	0.25*	0.18	0.24**	0.07	0.23
	(0.13)	(0.12)	(0.12)	(0.12)	(0.14)
Treat 2 × 5-5.4 Years old	-0.04	0.02	-0.09	-0.04	-0.14
	(0.16)	(0.13)	(0.14)	(0.15)	(0.15)
Treat 2 × 5.5-5.9 Years old	-0.34**	-0.24*	-0.25*	-0.21	-0.02
	(0.15)	(0.12)	(0.13)	(0.17)	(0.16)
Municipality FE	Yes	Yes	Yes	Yes	Yes
Control	Yes	Yes	Yes	Yes	Yes
Controls for Baseline Score	Yes	Yes	Yes	Yes	Yes
Observations	860	882	882	865	877
Adjusted R^2	0.29	0.13	0.11	0.38	0.02

Notes: Sample of analysis is children in control/home visit+community monitoring communities. The base of the interaction term is 4.5-4.9 Years old. Standard errors clustered at the community level in parentheses, $^{*}p < .10$,$^{**}p < .05$,$^{***}p < .01$.

Figure F.1: Estimated Effect of Home Visits on Child Development



Notes: Figure reports coefficients with standard errors. The top panel reports the intent-to-treat estimate of composite index. Results indicate the effect of residing in communities randomly assigned to Treatment Arm 1 (Home Visits) or Treatment Arm 2 (Home visits + Monitoring). The bottom panel report the similar estimate but with outcomes of language, fine motor, PPVT (vocabulary) and behavior index. All specifications include municipality fixed effects and the following controls: child gender, age in 2-month increments, baseline development scores, mother's years of education, whether or not the child's primary caregiver is the biological mother, household asset index, number of household residents, and an indicator equal to 1 for semi-urban communities, 0 for rural.

G Tables in Robustness Section

Table G.2: Intent-to-Treat Effect of Home Visits on Child Development, Specification

	(1)	(2)	(3)
Home Visits	0.09	0.10	0.10
	(0.10)	(0.07)	(0.07)
Home Visits + Community Monitoring	0.14	0.13*	0.13*
	(0.09)	(0.07)	(0.07)
Baseline Child Development			0.07***
			(0.02)
Municipality FE	Yes	Yes	Yes
Control	No	Yes	Yes
Observations	1,290	1,290	1,285
Adjusted R^2	0.09	0.27	0.27
Control Group Mean (2018)	0.02	0.02	0.02
F Stats: T1 vs T2	0.45	0.18	0.24
P-Val: T1 vs T2	0.50	0.67	0.62

Standard errors clustered at the community level in parentheses, $^{*}p < .10$,$^{**}p < .05$,$^{***}p < .01$. Controls at the individual level include child gender, age in 2 month increments, mother's years of education, and whether or not the child's primary caregiver is the biological mother. Household level controls include an index of household assets and the number of household residents, both measured at baseline. At the community level we include a binary indicator equal to 1 for semi-urban communities, 0 for rural.

Table G.3: Intent-to-Treat Effect Home Visits on Child Development, Controlling for Caregiver Participation in Community Monitoring Meetings

	Child Development	Language	Fine Motor	Receptive Vocabulary	Behavior
Assignment to Home Visits + Community Monitoring	0.10	0.09	0.13**	-0.04	0.12
	(0.08)	(0.07)	(0.06)	(0.06)	(0.09)
Participated in Community Monitoring	0.08	0.07	-0.02	0.07	0.11
	(0.07)	(0.06)	(0.07)	(0.07)	(0.11)
Municipality FE	Yes	Yes	Yes	Yes	Yes
Control	Yes	Yes	Yes	Yes	Yes
Controls for Baseline Score	Yes	Yes	Yes	Yes	Yes
Observations	860	882	882	865	877
Adjusted R^2	0.28	0.13	0.11	0.38	0.02

Note: Analysis restricted to children residing in communities assigned to Treatment 2 (Home Visits + Community Monitoring) and the control condition. Participation in community monitoring is an indicator variable equal to 1 if the caregiver reported attending community meetings to discuss home visit quality. Age refers to age at follow-up. Standard errors clustered at the community level in parentheses, $^{*}p < .10$,$^{**}p < .05$,$^{***}p < .01$. All outcomes are internally standardized for age in 2-month increments and are expressed in standard deviation units, with higher scores indicating improved development /better behavior. Controls include child gender, age in 2-month increments, baseline development scores, mother's years of education, whether or not the child's primary caregiver is the biological mother, household asset index, number of household residents, and an indicator equal to 1 for semi-urban communities, 0 for rural.