

**Human Capital and Economic Opportunity:
A Global Working Group**

Working Paper Series

Working Paper No. 2011-007

**Cash Transfers, Behavioral Changes, and
Cognitive Development in Early Childhood:
Evidence from a Randomized Experiment**

Karen Macours

Norbert Schady

Renos Vakis

October, 2011

**Human Capital and Economic Opportunity Working Group
Economic Research Center
University of Chicago
1126 E. 59th Street
Chicago IL 60637
humcap@uchicago.edu**

Cash Transfers, Behavioral Changes, and Cognitive Development in Early Childhood: Evidence from a Randomized Experiment⁺

Karen Macours*, Norbert Schady**, and Renos Vakis***

July 2011

* Paris School of Economics and INRA; **Inter-American Development Bank; ***World Bank

⁺ We are grateful to Ximena Del Carpio, Fernando Galeana, and Patrick Premand for countless contributions during data collection and preparation, the program team at the Ministerio de la Familia and in particular Teresa Suazo for their collaboration during the design of the impact evaluation, and the Centro de Investigación de Estudios Rurales y Urbanos de Nicaragua (in particular Verónica Aguilera, Carold Herrera, Enoe Moncada, and the entire field team) for excellent data collection. Financial support for this research was received from the World Bank (ESSD trust funds, a RSB grant, as well as the Government of the Netherlands through the BNPP program), and BASIS (under the USAID Agreement No. EDH-A- 00-06-0003-00 awarded to the Assets and Market Access Collaborative Research Support Program). We thank Jere Behrman, Sally Grantham-McGreggor, Elizabeth King, John Maluccio, Christina Paxson, Elisabeth Sadoulet, Miguel Urquiola, seminar participants at the World Bank, the Inter-American Bank, FAO, the Institute for Fiscal Studies, Berkeley, Leuven, PSE, Paris 2, and Wisconsin-Madison, two anonymous referees and the editor Esther Duflo for providing us with very helpful comments. The views expressed in this paper are those of the authors and do not necessarily reflect those of the World Bank or any of its affiliated organizations. All errors and omissions are our own. Contacts: karenmacours@parisschoolofeconomics.eu, norberts@iadb.org and rvakis@worldbank.org

Abstract

Cash transfer programs have become extremely popular in the developing world. There is a large literature on the effects of these programs on schooling, health and nutrition, but relatively little is known about possible impacts on child development. This paper analyzes the impact of a cash transfer program on cognitive development in early childhood in rural Nicaragua. Identification is based on random assignment. We show that children in households assigned to receive benefits had significantly higher levels of development nine months after the program began. There is no fade-out of program effects two years after the program had ended and transfers were discontinued. We show that the changes in child development we observe are unlikely to be a result of the cash component of the program alone.

1. Introduction

Development in early childhood is an important predictor of success throughout life. In developed countries, children with low levels of cognitive development before they enter school have lower school achievement and earn lower wages (Currie and Thomas 2001; Case and Paxson 2008). In developing countries, low levels of cognitive development have been tied to poor performance in school in a number of settings (see Grantham-McGregor et al. 2007 for a review).

Evidence from the medical and economic literature suggests that outcomes in early childhood are malleable (Heckman 2006; Knudsen et al. 2006). Randomized trials in the US show that children who benefited from intensive preschool interventions have higher school attainment, better test scores, lower rates of criminality, and earn higher wages in adulthood (Currie 2001; Schweinhart 2005), although the impacts appear to be concentrated among girls (Anderson 2008). A well-known study from Jamaica shows that children randomly assigned to receive home-based early stimulation have substantial improvements in cognitive development and subsequent school performance (Grantham-McGregor et al. 1991 and 1997; Walker et al. 2000; Powell et al. 2004). Non-experimental evidence suggests that preschool attendance is associated with better school performance in Argentina (Berlinski et al. 2009) and Uruguay (Berlinski et al. 2008). There is also a large literature documenting the impacts of nutritional supplementation programs, including substantial evidence from randomized control trials (see Walker et al. 2007 for a review). In Guatemala, children exposed to a nutritional intervention have better reading comprehension and perform better on tests of cognitive development in adulthood, and earn higher wages (Maluccio et al. 2009; Hoddinott et al. 2008).

A reasonable amount of evidence is therefore available on how the cognitive development of young children responds to supply-side interventions, including access to preschool, or food supplementation programs. Much less is known about interventions that attempt to directly affect

the investments parents make in child development—either by relieving financial constraints, or by changing how resources are allocated within households.

This paper analyzes the impact of a cash transfer program on development in early childhood. The program, known as *Atención a Crisis*, made sizeable payments to poor households in rural areas in Nicaragua. There are a variety of reasons why one might expect a program like *Atención a Crisis* to improve development in early childhood. Children in better-off households generally have higher levels of development than those in poorer households in developing countries.¹ These associations may not be causal—rather, they may reflect a correlation between child development and parental wealth, parental behavior, or genetic endowments. However, if cash transfers like those made by *Atención a Crisis* allow households to spend more on nutritious foods, early stimulation, or health care, this may result in improvements in child development.

There are other features of the *Atención a Crisis* program that could result in improvements in child development. Beneficiaries were told that transfers were intended to improve the diversity and nutrient content of children’s diets and to buy school material. The social marketing of the program may have transmitted knowledge about child-rearing practices; it may also have affected how transfer income was used through a flypaper or labeling effect.² Such changes in behaviors could be further enhanced through social interactions with other program beneficiaries and peer pressure (Macours and Vakis 2009). Finally, *Atención a Crisis* transfers were made to women, and income controlled by women may be spent in a way that benefits children more than income that is controlled by men.³

¹ References include Paxson and Schady (2007) and Schady (2011) on Ecuador; Halpern et al. (1996) on Brazil; Ghuman et al. (2005) on the Philippines; see also Schady (2006) for a discussion.

² See Thaler (1999) for a general discussion. Fraker et al. (1995) presents evidence for the US, although these results have been challenged by Hoynes and Schanzenbach (2009). See also Kooreman (2000) for the Netherlands, Jacoby (2002) for the Philippines, and Islam and Hoddinott (2008) for Guatemala. Edmonds (2002) finds no evidence of labeling effects for child benefit income in Slovenia.

³ For example, Thomas (1994), Hoddinott and Haddad (1995), Doss (2006), and Schady and Rosero (2008) show that income controlled by women is associated with higher expenditures on food. Macours and Vakis (2010) show non-

A large number of studies have assessed the impact of cash transfers, conditional and unconditional, on health status, nutrition, and education.⁴ In contrast, we are aware of only two earlier papers on the impact of cash transfers on child development in developing countries. Fernald et al. (2008) suggest that larger transfers made by the PROGRESA program in Mexico resulted in better nutritional status, improved motor skills, and higher levels of cognitive development. However, the variation in the amount of cash that is used to identify these effects may be endogenous (Attanasio et al. 2010). Paxson and Schady (2010) use random assignment in the roll-out of the *Bono de Desarrollo Humano* (BDH) cash transfer program in Ecuador to analyze the effects on health and development of children between 3 and 6 years of age. They show that cash transfers resulted in an improvement of about 0.18 standard deviations in development among the poorest quartile of children in their sample, with no effects among somewhat less poor children.

Our analysis adds to the existing literature in a number of important ways. To the best of our knowledge, this is the first paper on the impact of cash transfers on child development in a developing country that uses data spanning the period before, during, and after the program ended. We show that children in households that were randomized into the *Atención a Crisis* program had significantly higher levels of development nine months after households started receiving transfers. Program effects of a similar magnitude are still apparent two years after *Atención a Crisis* had been discontinued and transfers had ended. Thus, there appears to be no fade-out of treatment effects among beneficiaries of the *Atención a Crisis* program, at least over the period covered in our study. This stands in contrast with the results from evaluations of a number of preschool programs in the United States (see Currie and Thomas 2000, and Garces, Thomas and Currie 2002 on Head Start,

experimental evidence on the positive impact of mother's seasonal migration on children's cognitive development that is consistent with this hypothesis. Lundberg et al. (1997) and Ward-Batts (2008) present quasi-experimental evidence from the United Kingdom to argue that income controlled by women is more likely to be spent on clothing for women and children than income controlled by men.

⁴ The literature is extensive—see Fiszbein and Schady (2009) for a review. Maluccio and Flores (2005) look at the effects of an earlier cash transfer program in Nicaragua.

and Heckman et al. 2010 on the Perry Preschool Program), the results of a randomized evaluation of a food supplementation program in Jamaica (Walker et al. 2000, 2005), and the results of the evaluation of PROGRESA on child height (Neufeld et al. 2005, and the discussion in Fiszbein and Schady 2009, pp. 148-50). On the other hand, a parenting program in Jamaica appeared to sustainably change behaviors, and there was no fade-out of program effects on child development (Walker et al. 2000, 2005).

Another important contribution of this paper is that it analyzes the extent to which changes in child development can be explained solely by the cash component of the *Atención a Crisis* program. We provide two pieces of evidence that strongly suggest that this is unlikely. First, the *Atención a Crisis* program randomly assigned a group of households to a variant of the basic treatment that included a substantially larger cash transfer. Relative to households in the basic treatment group, households that received the larger cash transfer had higher expenditure levels during and (in particular) after the program, but they did not have better child development outcomes.

Second, we analyze changes in a number of intermediate inputs into the production of child development, including the consumption of food, early stimulation, and the utilization of preventive health services. The changes in the use of these inputs among treated households, which persisted even after the program had ended, are inconsistent with a simple story of higher overall expenditure levels among *Atención a Crisis* beneficiaries. Hence, other program features, such as the social marketing that accompanied the transfers, or the fact that transfers were made to women, or both, are likely to be important in explaining the changes in child development we observe. In sum, then, our paper goes beyond Fernald et al. (2008) and Paxson and Schady (2010) in analyzing impacts during and after the intervention, in showing that the impact is due not just to the cash transfer, and in establishing impact on intermediate inputs, indicating the plausible underlying mechanisms.

The rest of the paper proceeds as follows. In section 2, we describe the *Atención a Crisis* pilot program, and the data, in particular the measures of cognitive development. Section 3 discusses methods. We present results in section 4. Specifically, section 4a presents the main results, 4b considers differences between variations of the treatment received by different households, and 4c presents evidence on the change in the use of various inputs into child development by *Atención a Crisis* beneficiaries. Section 5 concludes.

2. Program design, data, identification, and early childhood development outcomes

A. The “Atención a Crisis” pilot program

The *Atención a Crisis* pilot program was implemented between November 2005 and December 2006 by the Ministry of the Family in 6 municipalities in rural Nicaragua. We provide a detailed description of the program in Appendix 1. The program included a careful evaluation based on random assignment. Randomization was conducted as follows. First, among all communities in the 6 municipalities, 56 intervention and 50 control communities were randomly selected through a lottery. Second, baseline data were collected in both treatment and control communities. These data were used to define program eligibility based on a proxy means test. Around 10 percent of households (and only 5 percent of households with children under 6 years of age) in treatment and control communities were ineligible for the program because their estimated baseline expenditures, as determined by the proxy means, was above the pre-defined threshold. This process resulted in the identification of 3,002 households to participate in the program. A further 3.7 percent of households that had originally been deemed eligible by the proxy means were reclassified as ineligible after a process of consultation with community leaders, and a corresponding 3.7 percent that had originally been deemed ineligible were reclassified as eligible. To avoid any possibility of selection bias from these choices, we use the *original* eligibility as the intent-to-treat.

In communities randomly selected to participate in the *Atención a Crisis* program, the primary child caregiver (known as the “titular”), who in the vast majority of cases was a woman, was invited to a registration assembly where the program objectives and various components were explained. At the end of the assembly, a lottery took place in each community. Participation in the assemblies and lotteries was close to 100 percent. On the basis of this lottery, all eligible households within each community were assigned to one of three treatments.

Households in Group 1 were offered a cash transfer, paid to the “titular” every 2 months. For households with children ages 0-5 this transfer was in principle conditional on regular preventive health check-ups. However, in practice, this conditionality was not monitored, and households were not penalized for non-compliance. Households with children between 7 and 15 years old who had not finished primary school received an additional educational transfer, conditional on the school enrollment and regular attendance of those children. The education conditionality was monitored in practice. The basic *Atención a Crisis* intervention was modeled after an earlier CCT program in Nicaragua, the *Red de Protección Social* (RPS).⁵ On average, transfers made to this group represented 15 percent of per capita expenditures of the average recipient household in our sample over the year in which it was implemented.⁶ We refer to this treatment as the basic treatment.

Households in Group 2 received a cash transfer that was identical to that received by households in Group 1. In addition, they were offered a scholarship that allowed one of the household members to choose among a number of vocational training courses offered at the municipal headquarters. These household members also participated in labor market and business-

⁵ See Maluccio and Flores (2005) for the impacts on education, health and nutrition of the RPS program.

⁶ Households received a transfer of US \$ 145 if they had no children or only children younger than 7. In addition, households with children between 7 and 15 enrolled in primary school received US \$ 90 per household, and a further US \$ 25 per child.

skill training workshops organized in their own communities. We refer to this treatment as the training package.

Households in Group 3 received a cash transfer that was identical to that received by households in Group 1. In addition, they were offered a lump-sum payment to start a small non-agricultural activity. This lump-sum was conditional on the household developing a business development plan. It was paid out between the end of May and September 2006.⁷ The value of the lump-sum payment represented approximately 11 percent of per capita expenditures of the average recipient household over the year in which it was implemented. A household in Group 3 therefore was eligible for transfers equivalent to approximately 26 percent of annual expenditures. We refer to this treatment as the lump-sum payment package.

In addition, all beneficiaries of the *Atención a Crisis* program, regardless of the treatment they were assigned to, were exposed to repeated information and communication efforts by program staff during enrollment and pay-days. These stressed the importance of varied diets, health, and education, and were meant to change household investment and consumption patterns. Beneficiaries were also expected to attend regular meetings with local program promoters to talk about the objectives and conditionalities of the program.

Program take-up was high. More than 95 percent of all households randomized into the three treatment groups signed up for the program and took up the basic cash transfer.⁸ A small fraction of those households, less than 5 percent, did not collect the full amount of the transfer they were eligible for because they had not complied with the school enrollment and attendance requirements. Take-up of the additional benefits offered to groups 2 and 3 was also high—89 percent for the

⁷ Households received US \$ 175 at the end of May, and an additional US \$ 25 in September, conditional on having started the nonagricultural activity that was planned.

⁸ The main reason households did not take-up the program was the fact that some originally eligible households were deemed ineligible by local leaders after the initial assignment—see above. A small number of households had also migrated out of the communities after baseline. In order to avoid any selection bias, we treat all of these households as *eligible*.

vocational training courses, and close to 100 percent for the lump-sum payment.⁹ Contamination of the control group was negligible (one household).

B. Data

Baseline data for the evaluation were collected in April-May 2005. A first follow-up survey was collected in July-August 2006, nine months after the households had started receiving payments. The sample includes the 3,002 eligible households in the treatment group, and a random sample of 1,019 eligible households in the communities that were assigned to the control group. A second follow-up survey, covering the same households as those included in the first follow-up, was collected between August 2008 and May 2009 (henceforth referred to as 2008). At this point, households had stopped receiving transfers for an average of two years.

Attrition over the study period was minimal, less than 1.3 percent in 2006 and 2.4 percent in 2008. Attrition is uncorrelated with treatment status, and does not differ across treatment packages. The baseline characteristics of the full sample of households and those that could be located at follow-up are very similar. We further discuss possible concerns regarding attrition and missing test data in Appendix 2.

All three surveys included comprehensive information on household socioeconomic status, including detailed expenditure modules,¹⁰ extensive information on child health and nutrition, including child height and weight, and one measure of child cognitive development, the TVIP. The TVIP is the Spanish-speaking version of the Peabody Picture Vocabulary Test (PPVT), a test of receptive vocabulary that can be applied to children 36 months and older (Dunn et al. 1986).

⁹ About 10 percent of the business development plans were initially turned down by the Ministry of the Family, which oversaw the program. These proposals were sent back to the households and virtually all of them developed a new plan, with the help of technical assistance (the few exceptions being households that had migrated out).

¹⁰ These modules were taken from the 2001 Nicaragua Living Standards Measurement Study (LSMS) survey. The expenditure module includes detailed information on various expenditure categories. For example, food expenditures include questions about 63 food items, and includes actual expenditures, home production, and food consumed outside the home.

Both follow-up surveys included a large number of tests to assess child development. Social-personal, language, fine motor, and gross motor skills for all children were assessed using the four sub-scales of the Denver Developmental Screening Test (Frankenberg and Dodds 1996). The Denver can be applied to children as young as one month of age. A slightly modified version of the Denver is used for child monitoring by the national early childhood stimulation program in Nicaragua, which suggests that the test is appropriate for the population we study.

For children age 36 months and older we applied five additional tests. The first of these is the TVIP, described above. We also use a short-term memory test from the McCarthy test battery, and a test of associative memory, drawn from the Woodcock-Johnson-Muñoz battery of cognitive abilities (Woodcock and Muñoz 1996; Schrank 2006; Schrank et al. 2005); the test of associative memory was only applied in the second follow-up survey. In both the first and second follow-up surveys we included a test of leg motor development from the McCarthy test battery (Boivin et al. 1995). The final test we use is the Behavior Problem Index (BPI), which is based on the caregiver's report of the frequency that a child displays each of 29 problematic behaviors, with responses coded as "never", "sometimes" and "often" (Baker and Mott 1989). We use the number of behavioral problems for which a caregiver answers "often".¹¹

All of the tests were carefully piloted in the field, and adjustments were made, as necessary. Many of these tests have been applied in similar populations in Latin America, including in the evaluations of cash transfer programs in Ecuador and Mexico (see Paxson and Schady 2010 and Fernald et al. 2008, respectively). An important advantage of the tests we use, with the exception of the BPI and a subset of items in the Denver, is that they provide observed, as opposed to parent-

¹¹ Unlike the other outcomes we study, behavioral problems do not necessarily indicate a delay, as there are no benchmarks or established ages at which they are predicted to decrease.

reported, measures of child development.¹² This substantially reduces concerns about reporting biases. Details of all of the tests we use are provided in Appendix 3.

The two follow-up surveys also include information on stimulation, birthweight, preventive health care, and caregivers' mental health. Mental health was measured using the Center for Epidemiological Studies Depression scale (CESD), a widely used measure of depression which consists of 20 questions on self-reported depression (Radloff 1977). Finally, caregivers' observed parenting behavior was registered through a shortened version of the HOME score, an index of 11 positive and negative behaviors that the enumerator observes during interviewing and testing (Bradley 1993; Paxson and Schady 2007, 2010).

Table 1 summarizes the baseline characteristics of households in our sample, focusing on socioeconomic status and child health. It shows that households and children are disadvantaged in a number of important ways. Expenditure levels are very low. Turning the local currency units (*Córdobas*) into US \$ shows that 81 percent of households in our sample have per capita expenditures that are below one dollar per capita per day. The mean years of schooling of mothers are 4 years, and 66 percent have not completed primary school. The mean years of schooling of fathers is equally low, and 72 percent have not completed primary school. Children in this sample have substantial health problems—27 percent are stunted (have height for their age that is more than two standard deviations below that of a reference population). Weight-for-height is not particularly low. The composition of food expenditures shows that a very high proportion of consumption consists of staples (59 percent), in particular tortilla, rice, and beans. Much smaller proportions of food consumption are animal products (16 percent) and, in particular, fruits and vegetables (5 percent). This suggests that lack of balance in diets, rather than insufficient overall caloric intake, may be part of the explanation for the nutritional deficiencies in this population.

¹² For the Denver subtests, there are no significant differences between children in the treatment and control groups in the likelihood that items were administered by direct observation rather than caregivers' report (see appendix 3).

Table 2 focuses on our measures of child development. It reports the fraction of children in the control group who are in the bottom 25 percent and, separately, the bottom 10 percent of the international distribution that was used to standardize a given test.¹³ The table shows that a very large fraction of children in our sample is delayed, although this varies considerably by outcome. The fraction of children who are behind for their age is largest for the measures of language—96 percent of children in our sample are in the lowest quartile of the distribution of the TVIP, and 84 percent have a score that places them in the lowest decile. Comparable numbers for the measure of language in the Denver test place 82 percent of children in the lowest quartile, and 60 percent in the lowest decile. A very large fraction of children in our sample is also delayed in memory—84 percent place in the lowest quartile of the test of short-term memory, and 58 percent in the lowest decile of the distribution used to standardize the test. In the case of the test of associative memory, 87 percent of children place in the lowest quartile, and 75 percent in the lowest decile.

These delays in language and memory are severe. For instance, the numbers for the TVIP imply that 85 percent of the children in our sample are at least 21 months delayed in receptive vocabulary. However, the implied delays are reasonably consistent with those observed among other populations with high poverty levels and low education in Latin America.¹⁴

Turning to other domains of child development, Table 2 shows that outcomes are somewhat better on the social-personal scale of the Denver—47 percent of children in the sample place in the lowest decile—and for fine motor skills—39 percent place in the lowest decile for this outcome.

¹³ For this purpose we use data from the first follow-up survey for all tests except for the test of associative memory, which was only collected in the second follow-up survey. Results are very similar if we use the second follow-up survey for all of these calculations.

¹⁴ Our analysis shares two tests with the results reported in Paxson and Schady (2007, 2010) and Schady (2011), namely the TVIP and the Woodcock-Johnson measure of associative memory. The average child in the sample from Ecuador places in the 11th percentile of the distribution of the TVIP, and in the 13th percentile of the test of associative memory. In our sample of children from Nicaragua, the average child places in the 6th percentile of the distribution of the TVIP and the 10th percentile of the test of associative memory. We note that the sample of children from Ecuador is considerably better off. 34 percent of households in the Ecuador study have consumption levels that are below one US dollar per capita per day, compared to 82 percent of households in this study. There are also marked differences in parental education, which is very robustly associated with performance on the cognitive tests—the average education of mothers in the Ecuador sample is 6.7, compared to 4.2 for the sample used in our paper.

Children in our sample perform even better in terms of gross motor skills: a much smaller fraction of children, 29 percent, place in the lowest decile of the distribution of the Denver, and 23 percent place in the lowest decile of the McCarthy leg motor scale. In addition to documenting the large fractions of children in our sample that are delayed, Table 2 shows that there are no obvious differences in delays between boys and girls. However, delays increase with child age for some outcomes.

It is more likely that cash transfers like those made by *Atención a Crisis* will result in improvements in cognitive development if there are socioeconomic gradients in these outcomes. Figure 1 presents nonparametric (Fan) regressions of each standardized outcome on log per capita expenditures among children in control communities (Fan and Gijbels 1996). The figure shows positive socioeconomic gradients in most measures of cognitive development. Gradients appear to be steepest for language (in particular, for the TVIP), height-for-age and weight-for-age.

Table 1 checks for balance between households randomly assigned to receive *Atención a Crisis* transfers and the control group (fourth column) and between households randomly assigned to the three treatment groups (basic treatment, training, lump-sum transfer – last two columns). The table shows that, by and large, random assignment equated the characteristics of households and children randomly assigned to different groups. Only one of 35 characteristics, whether the child received deworming drugs, is significantly different between treatment and control groups at the 5 percent level; for only one characteristic, the number of rooms in a house, can we reject the null of equal baseline means across the three treatment groups at the 5 percent level; and only for one characteristic, mother's education, can we reject the null of no differences between the basic treatment and lump-sum payment, which is the focus of the results we present on differences across treatment groups.

Although random assignment was successful, there are some small differences at baseline between households that were assigned to treatment and control groups. For example, children in the treatment group have somewhat lower height and weight than those in the control group. They are also less likely to have been weighed, and to have received vitamins or deworming drugs in the six months prior to the baseline survey. These differences suggest that it may be important to control for the baseline characteristics of households and children when estimating *Atención a Crisis* program effects on child development. We return to this point below.

3. Methods

We estimate child-level intent-to-treat regressions of the following form:

$$(1) \quad Y_k = \alpha_k T + \beta_k X + \varepsilon_k, \quad k=1 \dots K,$$

where Y_k is the k th outcome (out of 10 in the first follow-up survey, 11 in the second follow-up survey), T is a treatment indicator which takes on the value of one for children in communities that were randomly assigned to receive *Atención a Crisis* benefits, and X is a set of controls (including an intercept). To make it easier to draw comparisons across outcomes, we first convert each outcome into a within-sample z -score by subtracting the sample mean and dividing by the standard deviation of the control group.¹⁵ Also, we reverse the signs on the BPI, so that higher values correspond to “better” outcomes (as with the other outcomes). The coefficients on the treatment indicator therefore measure effect sizes in standard deviation units.

In one set of specifications, X includes only controls for the child’s age when the transfers started, in single-month intervals, and an indicator for the child’s gender. In another set of specifications, X also includes a number of baseline characteristics: age and gender of the household head, the years of schooling of the mother, the number of household members, the fraction of

¹⁵ We use the standard deviation of the control group in 2006 for both years, in order to be able to compare magnitudes across years.

members in five age categories, birth weight, height-for-age, weight-for-age, TVIP score, whether a child has been weighed, received deworming medicine, and vitamin A in the last six months, baseline community averages of height-for-age, weight-for-age, and TVIP score, and municipal fixed effects.¹⁶ Including these controls helps adjust for small baseline differences between treated and control groups, and may also make the estimated program effects more precise. Standard errors adjust for clustering at the community level.

In addition to estimating the effect for individual outcomes, we estimate the average treatment effect, across all outcome measures, and separately for the subsets of 6 cognitive and behavioral outcomes and 5 health and motor outcomes:

$$(2) \quad \bar{\alpha} = \frac{1}{K} \sum_{k=1}^K \hat{\alpha}_k$$

We estimate (1) or (2) by running seemingly unrelated regressions (SUR) for all (or a subset) of outcomes, and using the estimated variance-covariance matrix of the estimates to calculate the standard error of $\bar{\alpha}$ (see Kling et al. 2007; Duflo et al. 2008).

We also estimate intent-to-treat regressions that allow for separate effects for households that were randomly assigned to the three *Atención a Crisis* treatment packages:

$$(3) \quad Y_k = \gamma_k T_1 + \eta_k T_2 + \lambda_k T_3 + \beta_k X + \varepsilon_k, \quad k=1 \dots K,$$

where T_1 , T_2 and T_3 correspond to the basic treatment, the training package, and the lump-sum payment package, respectively. Finally, to tease out the role of higher expenditures on child development, we limit the sample to households assigned to either T_1 or T_3 , and run regressions of the following form:

$$(4) \quad Y_k = \theta_k T_3 + \beta_k X + \varepsilon_k, \quad k=1 \dots K,$$

¹⁶ In those cases where there are missing values for the covariates, we include the sample mean. However, our results are robust to including only covariates with very few missing values.

In this case, the coefficients θ_k are an estimate of the difference in outcomes between children in households assigned to the basic treatment and those that in addition were assigned to receive the lump-sum payment.

4. Results

A. Overall program effects

Our main results on the effect of the *Atención a Crisis* program on child health and development are reported in Tables 3 and 4. Table 3 focuses on program effects on individual outcomes in 2006 (upper panel) and 2008 (lower panel). In each case we include specifications that include controls for age and gender only (first row), and the extended set of controls described above (second row). All regressions are limited to children younger than six years of age at the time the transfers started (November 2005), as well as children born into these households since then.

The results in Table 3 are generally consistent with positive *Atención a Crisis* effects on child health and development. More than three-quarters (33 out of 42) of the coefficients are positive, and almost one-half of those that are positive (15 out of 33) are significant at the 10 percent level or higher. There are no significant negative coefficients. The evidence in favor of positive program effects is stronger in those specifications that include the extended set of controls than in those that only include controls for child age and gender. This likely reflects a small degree of imbalance between treatment and control at baseline, as seen in Table 1. In the case of the regressions of child height and weight, where the baseline imbalance was apparent, all of the coefficients are negative with the basic set of controls, but positive with the extended set of controls.

Table 4 reports the average effect across all outcomes, and separately for cognitive and socio-emotional development (the two language tests, the two memory tests, the two behavioral

tests) and health and motor development (the measures of gross motor, leg motor, fine motor, height and weight). The upper panel reports the mean effect sizes in 2006 and the lower panel in 2008, as before.

The first two rows in each panel correspond to the specifications in Table 3. In the specification with extended controls, households randomized into the *Atención a Crisis* program had outcomes that were 0.09 standard deviation higher than households randomized into the control group in 2006, and 0.08 standard deviations higher in 2008. In both years, the p-values for the mean effect sizes are below 0.01. For the cognitive and socio-emotional outcomes the program effects are 0.12 standard deviations in 2006 and 0.08 standard deviations in 2008. For the health and motor outcomes the program effects are 0.05 standard deviations in 2006 and 0.07 standard deviations in 2008.

Other rows in the table provide three important robustness checks on our main results. The Denver and the BPI tests are based, in part, on parents' reports about their children's development. It is conceivable that parents randomly assigned into the *Atención a Crisis* program were more likely to over-report the development of their children because they thought that this is what enumerators expected to hear (although it is unclear why this would affect the results for 2008, two years after the program had ended). It is also possible that the program made parents better able to detect delays in child development, in which case the treatment effects we estimate could be biased down. To check for these kinds of effects, we recalculated the averages but excluded the Denver and BPI. Excluding tests that are partly parent-reported does not have a substantive effect on our results—the mean effect size for the remaining outcomes is 0.12 standard deviations in 2006 and 0.06 standard deviations in 2008, both of which are highly significant. Thus, it does not appear that the positive program effects we estimate are a result of systematic misreporting by *Atención a Crisis* beneficiaries.

One difficulty in comparing the magnitude of the effects in 2006 and 2008 is that new children are born into the sample. Also, baseline children can age into tests that can only be applied to those 36 months or older, or age out of the Denver when they turn seven years of age or older. Therefore, the composition of the sample changes between 2006 and 2008. Moreover, the 2008 survey included an additional memory test. To see how this could affect our results, we report results that exclude the associative memory test, and are estimated over a “restricted” sample of children who took a given test in both years.¹⁷ The results for this smaller sample are very similar to those for the full sample: They suggest average *Atención a Crisis* program effects of 0.07 standard deviations in 2006 and 0.08 standard deviations in 2008.

Our main specification includes all children of relevant ages who were living in the sample of households randomized into a particular group at baseline, plus all additional children that were born to baseline household members. This implies that the person who received the cash transfer (the “titular”) was not always the mother of the child in our sample. Moreover, in a small number of cases, households split between baseline and the first or second follow-up, so the titular might no longer be living with the children we study. As a final robustness check, we restrict the sample to include only children of the titular at baseline (excluding children of other household members) and still living with the titular at the time of the follow-up surveys. Again, these results are similar to those from the larger sample—the mean effect size across all outcomes is 0.07 standard deviations in 2006 and 2008. This suggests that the program effects we estimate are not primarily a result of any possible effects of the *Atención a Crisis* program on household formation or dissolution.¹⁸

¹⁷ This also implies that the duration of exposure to the program is the same for all the children in this restricted sample, including the youngest children, if one includes the time *in-utero* and given that the 2006 follow-up was conducted 9 months after the start of the transfers.

¹⁸ The *Atención a Crisis* program effects we estimate are also robust to accounting in alternative ways for the relationship between the child and the titular, the main caregiver and the mother, to removing outliers, and to different ways of coding the tests. Results for families of outcomes are also similar when estimating the impact on the average of the standardized test scores, instead of using SUR (see Kling et al. 2007). We also tested for heterogeneity by child age

In sum, the results in Table 4 make clear that the *Atención a Crisis* program improved the health and development of children in beneficiary households. There is no evidence that the positive program effects we estimate are a result of systematic misreporting by parents. There is no apparent fade-out of program effects two years after the program ended, and the persistence of program effects cannot be explained by compositional changes in the sample.

B. Disaggregated effects by treatment package

An important question is whether the changes in child outcomes we observe can plausibly be explained by the income effect of the transfer alone. To answer this question we first estimate the impact of the *Atención a Crisis* program on the log of total per capita expenditures. These results are in Table 5. In the first column of the table, we report the results from a specification for the program as a whole, without differentiating by treatment package. The second through fourth columns separately estimate the effect of the basic treatment, the basic treatment plus training grant, and the basic treatment plus lump-sum transfer.

The results in Table 5 make clear that the *Atención a Crisis* program had large effects on household per capita expenditures in 2006. The specification for the full sample, including the extended set of controls, shows that households randomly assigned to the basic treatment increased their expenditures by 28 log points.¹⁹ The coefficient on households that received the basic treatment plus the lump-sum payment implies an increase in per capita expenditures of 33 log points. The relatively small difference in total expenditures between households assigned to receive only the basic treatment and those assigned to also receive the lump-sum payment can be explained

and gender. Program effects are generally somewhat larger for children who were older at baseline and for girls. These results are available from the authors upon request.

¹⁹ This increase in expenditures is substantially larger than the magnitude of the transfer: On average, households in this group received a transfer of US \$20 per month, but increased their expenditures (a large share of which is food expenditures, with recall of the last two weeks) by almost US \$35. The transfers were made somewhat irregularly, and the two transfers prior to the survey had occurred in a period of 6 weeks (instead of 2 months), including one just prior to the survey, which could explain the large effect on per capita consumption.

by the timing of the payment. The largest share of the lump-sum payment was made at the end of May, and the first follow-up survey was collected between July and August of that year.²⁰ The small increase in expenditures in households that received the lump-sum payment is also consistent with households investing (part of) the additional transfer in income-generating activities, as was intended.

Results for 2008, in the second row of the table, show that households that received the lump-sum payment continue to have higher per capita expenditures than those in the control group, about 8.8 log points. In contrast, the effect of the basic treatment on per capita expenditures is very small, about 2.2 log points, and is not significantly different from zero. This is not surprising given that the program had ended, and transfers had been discontinued, for approximately two years. An F-test rejects the null of equal coefficients for the basic treatment and the lump-sum payment in both 2006 and 2008.

Table 5 shows that households randomly assigned to the lump-sum payment had significantly higher consumption levels than those assigned to the basic treatment, most clearly in 2008. We therefore next compare child development outcomes for these two groups of households by estimating equation (4). These results are reported in Table 6. They show no evidence of better child development outcomes among households that received the lump-sum payment, relative to those that only received the basic treatment.

On the basis of the values in the table we conducted a simple back of the envelope calculation. Households that received the lump-sum payment had per capita expenditures that were 5 log points higher than those that received only the basic treatment in 2006, and 6.6 log points higher in 2008. If the program effects on expenditures for 2007, when there was no survey, are reasonably similar to those for 2006 and 2008, then households assigned to receive lump-sum

²⁰ This timing also implies that all groups likely had similar levels of consumption for the first 7 months of the transfers.

payments had cumulative per capita expenditures roughly 17 log points higher than those assigned to the basic treatment over the three-year period between 2006 and 2008. In 2006, households assigned to receive the basic treatment had per capita expenditure levels that were 28 log points higher than those assigned to the control group, and child development outcomes that were 0.088 standard deviations higher. Conservatively, we would therefore expect that children in households assigned to the lump-sum payment would have child development outcomes that are 0.053 standard deviations $[(17/28)*0.088]$ higher than those assigned to receive the basic treatment. In fact, this value falls outside the 90 percent confidence interval (-0.032 to 0.046) for the effect of the lump-sum payment, relative to the basic treatment. Similarly, for the family of cognitive development outcomes, we would expect that children in households assigned to the lump-sum payment would have child development outcomes that are 0.072 standard deviations $[(17/28)*0.118]$ higher than those assigned to receive the basic treatment. This value falls outside the 99 percent confidence interval (-0.074 to 0.059) for the effect of the lump-sum payment relative to the basic treatment.

In sum, the higher expenditure levels of households that randomly received the lump-sum treatment do not appear to have resulted in better child development outcomes, especially in terms of cognitive development. It is possible that this is a result of convexity in the relationship between outcomes and expenditures—although Figure 1 shows no evidence of such non-linearities for most outcomes. More likely, perhaps, the results suggest that something other than (or in addition to) the cash explains the *Atención a Crisis* treatment effects on child development we observe.

One limitation of the comparison between households randomly assigned to the basic treatment and the lump-sum payment is the fact that the latter were expected to start a small business. In particular, one concern is that starting a small business may itself have an effect on child development. The lump-sum payment is therefore not a clean measure of the possible effects of the additional cash. To assess the extent to which these concerns are important, Table 7 compares

the economic activity of mothers, patterns of work and time use, maternal mental health, and the home environment between households randomly assigned to the basic treatment and the lump-sum payment.

As expected, Table 7 shows that mothers assigned to the lump-sum payment spent fewer days in wage work than those assigned to the basic treatment, and more days in self-employment. In total, mothers assigned to the lump-sum payment worked 33 more days in 2006 (from a control group mean of 71 days), but there is no significant difference in the total number of days worked in 2008. There is no evidence that mothers assigned to the lump-sum payment spent fewer hours taking care of their children than those that received the basic treatment, no matter whether we consider hours that were devoted only to caregiving or also hours of caregiving while working. Mothers assigned to the lump-sum payment were as likely to read or tell stories to their children. Finally, there is no evidence that the lump-sum transfer had an effect on the mental health of mothers or on the quality of the home environment. In sum, the lump-sum payment does not appear to have had any obvious, negative effects on the amount or quality of the time that mothers spent with their children. Thus, the absence of better child development outcomes for households in the lump-sum transfer group, in spite of the larger transfers they received and the higher overall levels of expenditures, cannot easily be explained by other changes that could have had a deleterious effect on child development.

C. Changes in the use of intermediate inputs

We next analyze *Atención a Crisis* program effects on a number of “risk factors” that have been identified as important determinants of child development in the literature—namely, expenditures on food, availability of micro-nutrients, inadequate stimulation, exposure to infectious disease and caregivers’ mental health (see the review by Walker et al. 2007).

Table 8 reports the effects of the *Atención a Crisis* program on various measures of these risk factors. We include both estimates of changes in individual outcomes, and averages across families of inputs (the latter, in standard deviation units, as before). The top panel of the table reports results for 2006, and the bottom panel for 2008. The first two columns focus on the impact of the *Atención a Crisis* program, without distinguishing between treatment packages, while the last column focuses only on the effects of the basic treatment, relative to the control group.

The first column in Table 8 shows that the *Atención a Crisis* program had a substantial effect on the use of various inputs into child development. In 2006, households randomly assigned to the program changed the composition of food expenditures—spending a lower fraction on staples, and higher fractions on animal proteins, fruits and vegetables;²¹ treated households had substantial increases in various measures of child stimulation—they were more likely to tell stories, sing to, or read to their children, and to have pen, paper and toys for children in the house; children in households randomly assigned to the *Atención a Crisis* program were also more likely to have been weighed, received iron, vitamins or deworming medicine, and they spent fewer days in bed. The magnitude of the changes is substantial. For example, the mean increase in stimulation is 0.26 standard deviations, and the mean increase in health inputs is 0.13 standard deviations.

In Nicaragua, as elsewhere, wealthier households generally spend more on relatively expensive sources of calories (animal proteins and fresh fruits and vegetables, rather than staples); provide more inputs for child stimulation (books, toys); and make more use of preventive health services. At first blush, then, the overall program effects for 2006 may not be surprising, given that the *Atención a Crisis* program made substantial cash transfers. The remaining results in Table 8

²¹ We also investigated whether treated households report a higher number of days that children consumed specific food items, including tortillas, milk, meat, eggs, fruits and vegetables in the last week. These results are consistent with those in Table 8 for 2006, but differences between treated and control households are no longer significant in 2008. We note, however, that there is considerably less variability in these measures of reported intake than in the measures of household expenditures used for Table 8.

investigate whether the effects of the program on the use of various inputs into child development are consistent with an explanation that focuses on the cash transfer alone.

The second column of the table includes controls for the log of total per capita expenditures, and its square. Controlling for the higher total expenditures of the *Atención a Crisis* beneficiaries has only a modest effect on the estimated coefficients. For example, the mean increase in stimulation among treated households from these regressions for 2006 is 0.20 standard deviations (rather than 0.26 standard deviations in the regressions that do not control for total expenditure levels), while the increase in health inputs is 0.12 standard deviations (rather than 0.13 standard deviations). It does not seem that the higher use of inputs into child development by *Atención a Crisis* households can easily be explained by their higher overall expenditure levels alone.²²

An important caveat for these estimates is that total expenditures are themselves determined by the *Atención a Crisis* program, which could bias the regression coefficients. The remaining results in the table attempt to deal with this concern. Recall from Table 5 that households assigned to receive only the basic treatment did not have higher expenditures than those in the control group in 2008—the coefficient in a regression of the log of total per capita expenditures for these households is 0.022 (with a standard error of 0.028). Nevertheless, these households continue to show significant differences in the use of inputs into child development. On average, households that were randomly assigned the basic treatment had a 0.12 standard deviation increase in stimulation, and a 0.08 standard deviation increase in health inputs, relative to those in the control group, in 2008. Households assigned to the basic treatment also continued to devote a higher fraction of food expenditures to animal proteins and a lower fraction to staples. These effects cannot easily be explained by any contemporaneous income effect of the transfer. Rather, they suggest that

²² We also conducted a similar analysis non-parametrically by running Fan regressions of the nutrition, stimulation, health and environment inputs as a function of the log of total per capita expenditures, separately for households in the *Atención a Crisis* treatment and control groups. These results are very similar in character to those in Table 8, and are available from the authors upon request.

the *Atención a Crisis* program had an effect on behavior, and that some of these behavioral changes were still apparent two years after the program had been discontinued.

5. Conclusion

In many developing countries, young children suffer from profound delays in cognitive development. These delays have serious implications for the success of these children as adults. A variety of theories of skill formation suggest that investments in schooling and other dimensions of human capital will have low returns if children do not have adequate levels of cognitive and social skills at early ages (for example, Cunha et al. 2006). Understanding the causes of deficits in early childhood and identifying interventions that can help address them are important priorities for research.

This paper uses a randomized evaluation to assess the impact of a cash transfer program on a large set of measures of child development in Nicaragua, a low-income country. The identification is straightforward—it is based on random assignment, with almost perfect compliance, and remarkably low levels of attrition over three survey waves. We show that a program that transferred cash to women improved child development. Remarkably, there was no fade-out of impacts two years after the program was ended and transfers discontinued. This stands in contrast with evaluations of a number of interventions in both developed and developing countries.

The magnitude of the effects we estimate is modest, but not trivial. One way of putting the magnitude in context is by comparing it with differences in outcomes between children of mothers with more or less schooling. In the control group, every year of maternal schooling is associated with 0.05 standard deviations better child development, on average. The program effects we estimate are therefore equivalent to comparing children with mothers with one-and-a-half more or less years of schooling—a substantial amount, given the control group average of four years of

schooling. Another way of putting the magnitude in context is by comparing it with the impacts of interventions on child development estimated elsewhere. Paxson and Schady (2010) estimate that the BDH unconditional transfer program in Ecuador had an insignificant impact of 0.05 standard deviations on cognitive and socio-emotional development, and an effect of 0.18 standard deviations for the poorest quartile of children. Berlinski et al. (2009) report an effect size of 0.23 standard deviations for the impact of one year of preschool for children age 3-5 years of age on learning outcomes in Argentina. Behrman et al. (2004) report an impact of 3-4 percent of the mean for a preschool program in Bolivia. All of these estimates refer to cognitive outcomes, and to children 36 months and older. *Atención a Crisis* program effects on cognitive outcomes (language and memory) for these older children are 0.19 standard deviations in 2006, and 0.20 in 2008. Our estimates are therefore very close in magnitude to those that have been reported from other settings in Latin America.

Households who benefited from transfers increased expenditures on critical inputs into child development. They spent more on nutrient-rich foods, provided more early stimulation to their children, and made more use of preventive health care. Changes in the use of these inputs are larger than what one would expect to see if the program were simply moving children along the curves that relate inputs to overall expenditures. Thus, the program appears to have resulted in behavioral changes. Some of these behavioral changes persisted after the program ended, although the differences in input use between the *Atención a Crisis* treatment and control groups are generally smaller than when the program was still operating. It is therefore not clear whether the persistence of better child development outcomes among *Atención a Crisis* beneficiaries is a result of the one-time jump in outcomes that took place while the program was operating, or behavioral changes that continued after the program ended. We note, however, that the fact that fade-out of impacts appears

to occur for many different early childhood programs suggests that the behavioral changes among *Atención a Crisis* beneficiaries are likely to be important.

The *Atención a Crisis* program randomized three treatment variations. One of the treatment groups had significantly higher per capita expenditure both during the program and after the program ended. We find no evidence that child development outcomes are better for these households. Thus, in Nicaragua, a dollar is not always a dollar (or, rather, a *Córdoba* is not always a *Córdoba*). Something other than, or in addition to, the cash appears to be important. The social marketing that accompanied the transfers, or the fact that transfers were made to women, or both, could be part of the explanation. A better understanding of what features of this and other cash transfer programs account for improvements in child development is an important priority for future research.

References

- Aguilera, V., X. Del Carpio, C. Herrera, K. Macours, M. Enoe Moncada, C. Obregón, and R. Vakis, 2006. “Estudio Cualitativo del Componente de Atención a Crisis (CAC) del Ministerio de la Familia, Nicaragua”, mimeo, Johns Hopkins University.
- Anderson, M.A. 2008. “Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects.” *The Journal of the American Statistical Association*, 103(484): 1481-1495.
- Attanasio, O., C. Meghir, and N. Schady. 2010. “Mexico’s Conditional Cash Transfer Programme.” *The Lancet* 375 (9719): 980.
- Baker, P.C. and F.L Mott. 1988. *NLSY Child Handbook 1989: A Guide & Resource Document for the National Longitudinal Survey of Youth 1986 Child Data*. Columbus, Ohio: Center for Human Resource Research, The Ohio State University.
- Baydar, N. and J. Brooks-Gunn. 1991. “Effects of Maternal Employment and Child-Care Arrangements on Preschoolers’ Cognitive and Behavioral Outcomes: Evidence from the Children of the National Longitudinal Survey.” *Development Psychology* 27(6): 932-45.
- Behrman, J., Y. Cheng, and P.E. Todd. 2004. “Evaluating Preschool Programs When Length of Exposure to the Program Varies: A Nonparametric Approach.” *Review of Economics and Statistics* 86(1): 108-32.
- Berlinski, S., S. Galiani, and P. Gertler. 2009. “The Effect of Pre-Primary Education on Primary School Performance.” *Journal of Public Economics* 93(1-2): 219-34.
- Berlinski, S., S. Galiani, and M. Manacorda. 2008. “Giving Children a Better Start in Life: Preschool Attendance and School-Age Profiles.” *Journal of Public Economics* 92(5-6): 1416-40.
- Blau, F. D., and A. J. Grossberg. 1992. “Maternal Labor Supply and Children’s Cognitive Development.” *Review of Economics and Statistics* 74(3): 474-81.
- Boivin, M. J., Green, S. D., Davies, A. G., Giordani, B., Mokili, J. K., & Cutting, W. A. (1995). A preliminary evaluation of the cognitive and motor effects of pediatric HIV infection in Zairian children. *Health Psychology* 14(1): 13-21.
- Bradley, R. H. 1993. “Children’s Home Environments, Health, Behavior, and Intervention Efforts: A Review Using the HOME Inventory as a Marker Measure.” *Genetic, Social, and General Psychology Monographs* 119(4): 437-490.
- Case, A., and C. Paxson. 2008. “Stature and Status: Height, Ability, and Labor Market Outcomes.” *Journal of Political Economy* 116(3): 499-532.
- Cheung, Y.B, Yip, P.S.F, and Kalrberg, J.P.E. 2001. “Fetal Growth, Early Postnatal Growth and Motor Development in Pakistani Infants.” *International Journal of Epidemiology* 30: 66-72.

- Choudhury, N., and Gorman, K.S. 2003. "Subclinical Prenatal Iodine Deficiency Negatively Affects Infant Development in Northern China." *The Journal of Nutrition* 133(10): 3162-65.
- Cogill, S.R., Caplan, H.L., Alexandra, H., Robson, K.M., and Kumar, R. 1986 "Impact of Maternal Postnatal Depression on Cognitive Development of Young Children." *British Medical Journal* 292(6529): 1165-67.
- Cunha, F., J. Heckman, L. Lochner, D. Masterov. 2006. "Interpreting the Evidence on Life Cycle Skill Formation." In *Eric A. Hanushek and Finis Welch (eds). Handbook of the Economics of Education, 1, pp 697-812.*
- Currie, J. 2001. "Early Childhood Education Programs." *Journal of Economic Perspectives* 15(2): 213-38.
- Currie, J., and D. Thomas. 2001. "Early Test Scores, School Quality and SES: Longrun Effects on Wage and Employment Outcomes." In *Worker Wellbeing in a Changing Labor Market.* S.W. Polacheck, ed., pp. 103-132, Amsterdam; London and New York: Elsevier
- Currie, J., and D. Thomas. 2000. "School Quality and the Longer-Term Effects of Head Start." *Journal of Human Resources* 35(4): 755-74.
- Dewey, K.G., Cohen, R.J., Brown, K.H., Rivera, L.L. 2001. "Effects of Exclusive Breastfeeding for Four versus Six Months on Maternal Nutritional Status and Infant Motor Development: Results of Two Randomised Trials in Honduras." *The Journal of Nutrition* 131(2): 262-67.
- Doss, C. 2006. "The Effects of Intrahousehold Property Ownership on Expenditure Patterns in Ghana." *Journal of African Economies* 15(1): 149-80.
- Duflo, E., R. Glennerster and M. Kremer. 2008. "Using Randomization in Development Economics: A Toolkit." In *Handbook of Development Economics*, T.P. Schultz and J.A. Strauss, eds., Vol. 4, pp. 3895-3962, Elsevier.
- Dunn, L.M., E.R. Padilla, D.E. Lugo, and L.M. Dunn. 1986. *Test de Vocabulario en Imágenes Peabody.* Circle Pines, Minnesota: American Guidance Service.
- Edmonds, E. 2002. "Reconsidering the Labeling Effect for Child Benefits: Evidence from a Transition Economy." *Economics Letters* 76(3): 303-09.
- Fan, J. and I. Gijbels 1996. *Local Polynomial Modelling and Its Applications.* New York: Chapman & Hall.
- Fernald, L., P. Gertler, and L. Neufeld. 2008. "Role of Cash in Conditional Cash Transfer Programmes for Child Health, Growth, and Development: An Analysis of Mexico's *Oportunidades.*" *The Lancet* 371(9615): 828-837.
- Fiszbein, A., and N. Schady. 2009. *Conditional Cash Transfers: Reducing Present and Future Poverty.* Washington, D.C.: World Bank.

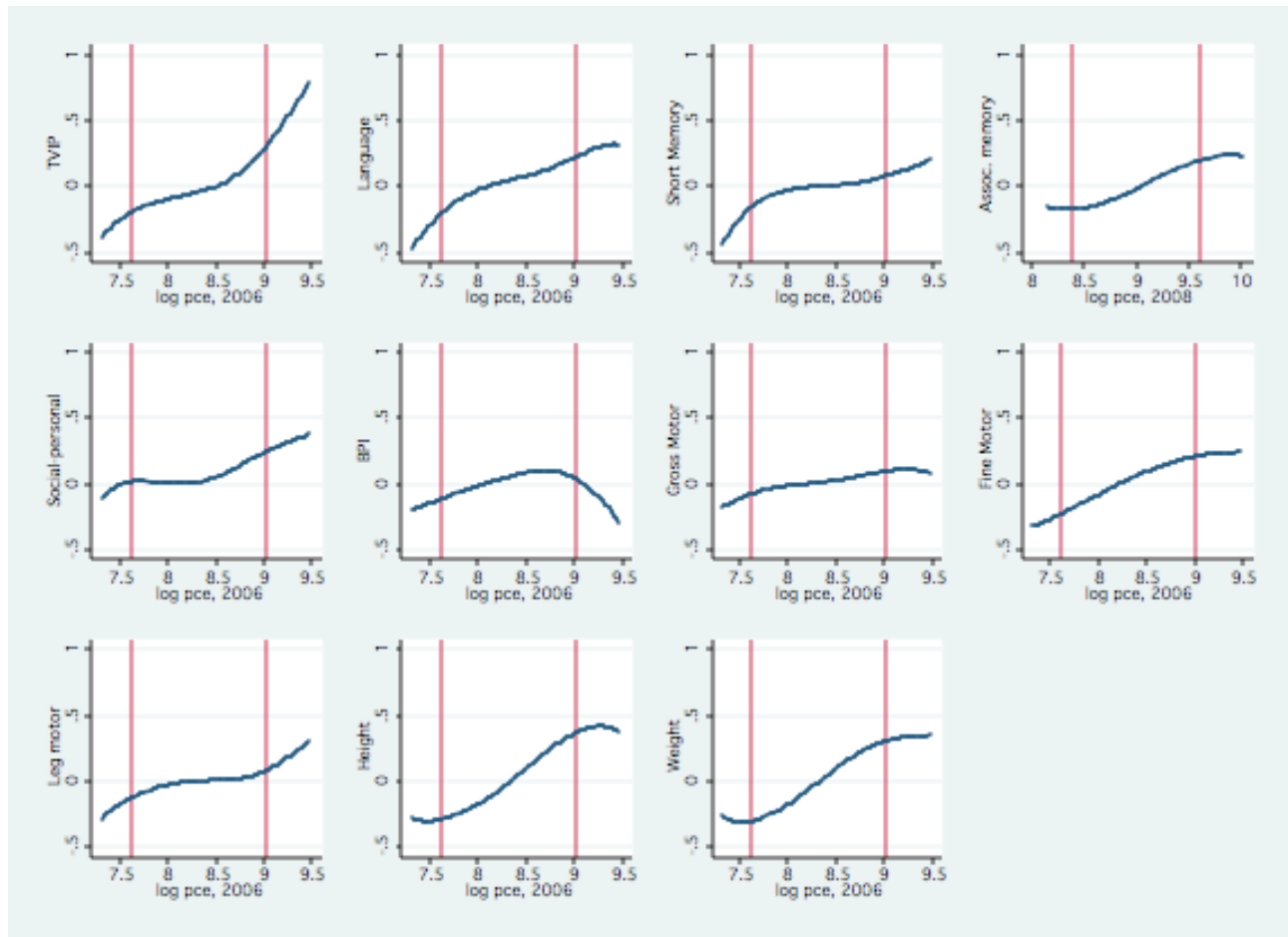
- Fraker, T. M., A. P. Martini, J. C. Ohls. 1995. "The Effect of Food Stamp Cashout on Food Expenditures: An Assessment of the Findings from Four Demonstrations." *Journal of Human Resources* 30(4): 633-49.
- Frankenberg, W.K., and J.B. Dodds. 1996. The Denver Developmental Screening Test. Denver, CO: University of Colorado Medical Center.
- Garces, E., D. Thomas and J. Currie. 2002. "Longer-Term Effects of Head Start." *American Economic Review* 92(4): 999-1012.
- Ghuman, S., J. Behrman, J. B. Borja, S. Gultiano and E. King. 2005. "Family Background, Service Providers, and Early Childhood Development in the Philippines: Proxies and Interactions." *Economic Development and Cultural Change* 54(1): 129-164.
- Grantham-McGregor, S. M., S. P. Walker, S. M. Chang and C. A. Powell. 1997. "Effects of Early Childhood Supplementation With and Without Stimulation on Later Development in Stunted Jamaican Children." *American Journal of Clinical Nutrition* 66(2): 247-253.
- Grantham-McGregor, S. M., Y. B. Cheung, S. Cueto, P. Glewwe, L. Richter, B. Strupp, and the International Child Development Steering Group. 2007. "Developmental Potential in the First 5 Years for Children in Developing Countries." *The Lancet* 369(9555): 60-70.
- Grantham-McGregor, S.M., C.A. Powell, S.P. Walker and J.H. Himes, 1991. "Nutritional Supplementation, Psychosocial Stimulation, and Mental Development of Stunted Children: The Jamaican Study." *The Lancet* 338(8758): 1-5.
- Halpern, R., and others. 1996. "Desenvolvimento neuropsicomotor aos 12 meses de idade em uma coorte de base populacional no Sul do Brasil: diferenciais conforme peso ao nascer e renda familiar." *Cad. Saúde Pùb.* 12(suppl. 1): 73-78.
- Heckman, J. J. 2006. "Skill Formation and the Economics of Investing in Disadvantaged Children." *Science* 312(5782): 1900-1902.
- Heckman, J.J., S.H. Moon, R. Pinto, P. Savelyev, and A. Yavitz. 2010. "The Rate of Return to the HighScope Perry Preschool Program." *Journal of Public Economics* 94(1-2): 114-28.
- Hoddinott, J., and L. Haddad. 1995. "Does Female Income Share Influence Household Expenditures? Evidence from Cote d'Ivoire." *Oxford Bulletin of Economics and Statistics* 57(1): 77-96.
- Hoddinott, J., J. A. Maluccio, J. Behrman, R. Flores and R. Martorell. 2008. "Effect of a nutrition intervention during early childhood on economic productivity in Guatemalan adults." *The Lancet* 371(9610): 411-416.
- Hoynes, H., Schanzenbach, D.W. 2009. "Consumption Responses to In-Kind Transfers: Evidence from the Introduction of the Food Stamp Program." *American Economic Journal: Applied Economics* 1(4): 109-39.

- Islam, M. and J. Hoddinott. 2009. "Evidence of Intra-Household Flypaper Effects from a Nutrition Intervention in rural Guatemala." *Economic Development and Cultural Change* 57(2): 215-38.
- Jacoby, H. 2002. "Is there an Intrahousehold 'Flypaper Effect'? Evidence from a School Feeding Programme." *Economic Journal* 112: 196-221.
- Kling, J., J. Liebman, and L. Katz. 2007. "Experimental Analysis of Neighborhood Effects." *Econometrica* 75(1): 83-119.
- Knudsen, E.L., J.J. Heckman, J.L. Cameron and J.P. Shonkoff. 2006. "Economic, neurobiological and behavioral perspectives on building America's workforce." *Proceedings of the National Academy of Sciences* 103(27): 10155-10162.
- Kooreman, P. 2000. "The Labeling Effect of a Child Benefit System." *American Economic Review* 90(3): 571-83.
- Lundberg, S., R. A. Pollack, T. J. Wales. 1997. "Do Husbands and Wives Pool their Resources? Evidence from the United Kingdom Child Benefit." *Journal of Human Resources* 32(3): 463-80.
- Macours, K., and R. Vakis, 2010. "Seasonal Migration and Early Childhood Development", *World Development* 38(6): 857-869.
- Macours, K., and R. Vakis, 2009. "Changing Households' Investments and Aspirations through Social Interactions: Evidence from a randomised transfer program in a low-income country", World Bank Policy Research Working Paper 5137, World Bank: Washington D.C.
- Maluccio, J.A., and R. Flores. 2005. "Impact Evaluation of a Conditional Cash Transfer Program: The Nicaraguan *Red de Protección Social*." *Research Report* 141, International Food Policy Research Institute, Washington, D.C.
- Maluccio, J., J. Hoddinott, J. Behrman, R. Martorell, A. Quisumbing and A. Stein. 2009. "The Impact of Improving Nutrition During Early Childhood on Education among Guatemalan Adults", *The Economic Journal* 119(537):734-763
- Neufeld, L. M., D. Sotres-Álvarez, P. Gertler, L. Tonlentino Mayo, J. Jiménez Ruiz, L. Fernald, S. Villalpando, T. Shamah, and J. Rivera Dommarco. 2005. "Impacto de Oportunidades en el Crecimiento y Estado Nutricional de Niños en Zonas Rurales." In *Evaluación Externa de Impacto del Programa Oportunidades 2004: Alimentación*, ed. B. Hernández Prado and M. Hernández Ávila. Cuernavaca, Morales, México: Instituto Nacional de Salud Pública.
- Oberhelman, R., E. Guerrero, M. L. Fernandez, M. Silio, D. Mercado, N. Comiskey, G. Ihenacho, and R. Mera. 1998. "Correlations between Intestinal Parasitosis, Physical Growth, and Psychomotor Development among Infants and Children from Rural Nicaragua." *American Journal of Tropical Medical Hygiene* 58(4): 47-75.
- Paxson, C., and N. Schady. 2007. "Cognitive Development Among Young Children in Ecuador: The Role of Health, Wealth and Parenting." *Journal of Human Resources* 42(1): 49-84.

- Paxson, C., and N. Schady. 2010. "Does Money Matter? The Effects of Cash Transfers on Child Health and Development in Rural Ecuador." *Economic Development and Cultural Change* 59(1): 187-230.
- Powell, C., H. Baker-Henningham, S. Walker, and S. Grantham-McGregor. 2004. "Feasibility of Integrating Early Stimulation into Primary Care for Undernourished Jamaican Children: Cluster Randomized Controlled Trial." *British Medical Journal* 329(7457): 89.
- 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.
- Rosenzweig, M., and K. I. Wolpin. 1994. "Are there Increasing Returns to the Intergenerational Production of Human Capital? Maternal Schooling and Child Intellectual Achievement." *Journal of Human Resources* 29(2): 670-93.
- Schady, N. 2006. "Early Childhood Development in Latin America and the Caribbean." *Economía* 6(2): 185-213.
- Schady, N. 2011. "Parental Education, Vocabulary, and Cognitive Development in Early Childhood: Longitudinal Evidence from Ecuador." Forthcoming, *American Journal of Public Health*.
- Schady, N., and J. Rosero. 2008. "Are Cash Transfers Made to Women Spent like Other Sources of Income?" *Economics Letters* 101(3):246-284.
- Schrank, F.A., K.S. McGrew, M.L. Ruef, C.G. Alvarado, A.F. Muñoz-Sandoval, and R.W. Woodcock. 2005. "Bateria III Woodcock-Muñoz: Assessment Service Bulletin Number 1, Overview and Technical Supplement." Chicago: Riverside Publishing
- Schrank, F.A.. 2006. "Specification of the Cognitive Processes Involved in the Performance on the Woodcock-Johnson III." Woodcock-Johnson III Normative Update: Assessment Service Bulletin Number 7, Chicago: Riverside Publishing
- Schweinhart, L.J. 2005. "The High/Scope Perry Preschool Study through Age 40: Summary, Conclusions, and Frequently Asked Questions." Ypsilanti, Mich.: High/Scope Press.
- Stoltzfus, R.J., J.D. Kvalsvig, H.M. Chwaya, et al. 2001. "Effects of Iron Supplementation and Anthelmintic Treatment on Motor and Language Development of Pre-school Children in Zanzibar: Double Blind, Placebo Controlled Study." *British Medical Journal* 323(7326): 1389-93.
- Thaler, R.H. 1999. "Mental Accounting Matters." *Journal of Behavioral Decision Making* 12: 183-206.
- Thomas, D. 1994. "Like Father, Like Son, Like Mother, Like Daughter: Parental Resources and Child Height." *Journal of Human Resources* 29(4): 950-88.
- Umbel, V. M., B. Z. Pearson, M. C. Fernandez and D. K. Oller. 1992. "Measuring Bilingual Children's Receptive Vocabularies." *Child Development*. 63(4): 1012-20.

- Walker, S.P., S.M. Grantham-McGregor, C.A. Powell, S.M. Chang. 2000. "Effects of Growth Restriction in Early Childhood on Growth, IQ, and Cognition at Age 11 to 12 Years and the Benefits of Nutritional Supplementation and Psychological Stimulation." *Journal of Pediatrics* 137(1): 36-41.
- Walker, S.P., S.M. Chang, C.A. Powell, S.M. Grantham-McGregor. 2005. "Effects of Early Childhood Psychosocial Stimulation and Nutritional Supplementation on Cognition and Education in Growth-Stunted Jamaican Children: Prospective Cohort Study." *The Lancet* 366(9499): 1804-07.
- Walker, S.P., T. Wachs, J. Meeks Gardner, B. Lozoff, G. Wasserman, E. Pollit, J. Carter. 2007. "Child Development: Risk Factors for Adverse Outcomes in Developing Countries." *The Lancet* 369(9556): 145-57.
- Ward-Batts, J. 2008. "Out of the Wallet and Into the Purse: Using Micro Data to Test Income Pooling." *Journal of Human Resources* 43(2): 325-51.
- Woodcock, R.W. and A.F. Muñoz-Sandoval. 1996. *Batería Woodcock-Muñoz Pruebas de Aprovechamiento-Revisada*. Chicago: Riverside Publishing.

Figure 1: Socio-economic gradients of test outcomes in control communities



Note: Outcomes for 2006, except associative memory which is for 2008. All outcomes are standardized by subtracting the mean and dividing by the standard deviation of the control group. Sample includes children under 6 years old when the transfers started and all children born in sample households since. For the Denver (personal, language, fine motor, gross motor), the sample includes children upto 83 months; for the TVIP (receptive language), McCarthy (memory, leg motor), Woodcock-Munoz (associative memory) and BPI the sample includes children 36 months or older; height-for-age and weight-for-age is for all children. For the Denver test, calculations are based on the number of delays, for TVIP, McCarthy and Woodcock-Munoz calculations are based on raw test scores. Vertical lines are included at 10th and 90th percentiles of log per capita expenditures in control communities. Fan regressions with bandwidth of 0.99. 2.5% highest and lowest outliers of log(pce) trimmed from graph.

Table 1: Baseline characteristics and randomization checks

	N	Control	Treatment	P-value diff. T-C	P-value diff. T1=T2=T3	P-value diff. T1=T3
Child-specific characteristics						
All children						
male	4245	0.49	0.50	0.376	0.912	0.727
age in months when transfers started	4245	22	21	0.194	0.465	0.488
mother lived in household at baseline	4245	0.95	0.97	0.183	0.699	0.632
# years education mother	4005	4.21	4.05	0.557	0.075*	0.025**
# years education father	4007	3.88	3.81	0.773	0.572	0.877
Children age 3-6 at baseline						
tvip (vocabulary recognition) test score	1066	5.37	6.23	0.207	0.396	0.290
Children age 0-5 at baseline						
weight-for-age z-score	2377	-0.88	-1.06	0.094*	0.510	0.466
height-for-age z-score	2368	-1.08	-1.27	0.109	0.081*	0.673
weight-for-height z-score	2383	-0.16	-0.18	0.799	0.829	0.724
birth weight	2415	6.76	6.75	0.947	0.340	0.193
weighed in last 6 months	2503	0.93	0.90	0.178	0.698	0.817
Received vitamins in last 6 months	2503	0.75	0.68	0.070*	0.541	0.276
Received deworming drugs in last 6 months	2503	0.59	0.51	0.036**	0.578	0.319
Household-level characteristics						
male household head	2407	0.84	0.85	0.539	0.397	0.215
household size	2407	6.05	5.90	0.344	0.732	0.446
# hh members 0-5 years old	2407	1.06	1.04	0.705	0.686	0.655
# hh members 5-14 years old	2407	1.69	1.70	0.954	0.382	0.627
# hh members 15-24 years old	2407	1.21	1.17	0.515	0.601	0.853
# hh members 25-64 years old	2407	1.88	1.84	0.473	0.423	0.205
# hh members more than 65 years old	2407	0.18	0.13	0.061*	0.757	0.625
Number of rooms in the house	2407	1.63	1.57	0.498	0.040**	0.387
Time to school (minutes)	2407	0.31	0.26	0.149	0.062*	0.683
Time to health center (minutes)	2407	1.28	1.17	0.493	0.968	0.802
Time to municipal headquarters (minutes)	2407	1.69	1.58	0.523	0.940	0.763
Owens toilet/latrine	2407	0.75	0.72	0.461	0.827	0.887
Access to water	2407	0.11	0.13	0.646	0.104	0.151
Access to electricity	2407	0.36	0.38	0.790	0.633	0.780
Own land	2407	0.64	0.63	0.731	0.829	0.543
Total Consumption per capita (cordobas)	2407	4723	4635	0.809	0.841	0.586
Food consumption per capita (cordobas)	2407	3333	3110	0.408	0.632	0.364
Proportion of food in total expenditures	2407	0.70	0.68	0.132	0.376	0.165
Proportion of staples in all food exp.	2399	0.59	0.58	0.871	0.389	0.172
Proportion of animal proteins in all food exp.	2399	0.16	0.16	0.868	0.372	0.479
Proportion of fruit and vegetables in all food exp.	2399	0.05	0.05	0.573	0.375	0.168

Note: P-values based on standard errors clustered by community. Data for all children and household-level characteristics are based on all children in 2008 sample with at least one test result that were either younger than 6 years when transfers started or born to baseline household members since the baseline. Data for children age 3-6, respectively 0-5, at baseline does not include children born after the baseline.

Table 2: Frequency of delay in control communities compared to international norm in 2006

	Children 0-83 months old				Children 36-83 months old				All children	
	Denver			Gross motor	TVIP	WJ	Mccarthy		Weight	Height
	Social	Language	Fine Motor		Receptive language	Associative memory	Short memory	Leg motor		
Child is in lowest 25% of international distribution										
All	0.65	0.82	0.60	0.46	0.96	0.87	0.84	0.40	0.56	0.64
Child is in lowest 10% of international distribution										
All	0.47	0.60	0.39	0.29	0.84	0.75	0.58	0.23	0.37	0.47
Boys	0.49	0.63	0.41	0.30	0.83	0.72	0.57	0.23	0.38	0.49
Girls	0.44	0.57	0.37	0.29	0.85	0.77	0.59	0.24	0.36	0.45
0-35 months	0.30	0.48	0.28	0.41					0.34	0.41
36-59 months	0.48	0.59	0.53	0.27	0.70	0.78	0.56	0.21	0.37	0.54
60-83 months	0.68	0.77	0.41	0.14	0.98	0.75	0.61	0.25	0.41	0.49

Note: All tests from 2006, except WJ associative memory from 2008. To calculate delays, international standardized scores were calculated for each test. For the Denver, which consists of various tasks, each of which is age standardized, children are categorized in the lowest 25% (resp 10%) if they are in the lowest 25% (10%) for at least one of the tasks.

Table 3: Impacts on individual tests in 2006 and 2008

	Cognitive and socio-emotional outcomes						Health and motor development outcomes				
	TVIP	Language	Short memory	Assoc. memory	Social-personal	Behavior	Gross motor	Fine motor	Leg motor	Height-for-age	Weight-for-age
2006: All children											
Age&gender controls	0.201*** (0.075)	0.108* (0.055)	0.087 (0.056)		0.114** (0.050)	-0.007 (0.088)	-0.031 (0.058)	0.024 (0.064)	0.023 (0.092)	-0.063 (0.091)	-0.061 (0.081)
Extended controls	0.228*** (0.062)	0.139*** (0.050)	0.156*** (0.044)		0.130*** (0.047)	-0.048 (0.084)	-0.006 (0.046)	0.038 (0.063)	0.130* (0.076)	0.072** (0.034)	0.036 (0.037)
N	1817	3287	1827		3307	1620	3253	3265	1838	3082	3082
2008: All children											
Age&gender controls	0.104 (0.100)	0.060 (0.056)	0.0789 (0.050)	0.073 (0.062)	0.056 (0.052)	0.016 (0.060)	0.056 (0.064)	0.099* (0.051)	-0.036 (0.046)	-0.096 (0.094)	-0.065 (0.082)
Extended controls	0.094 (0.078)	0.093** (0.045)	0.086* (0.044)	0.105** (0.046)	0.098** (0.046)	0.021 (0.063)	0.102 (0.064)	0.156*** (0.039)	0.006 (0.034)	0.045 (0.031)	0.029 (0.043)
N	2990	3095	3011	3015	3097	2863	3080	3085	1881	4185	4185

Note: Coefficients and standard errors (in parentheses). Standard errors adjust for clustering at the community level. Controls include individual-level controls (dummies for child gender and month dummies for child age, the years of schooling of the mother, baseline height-for-age, weight-for-age, TVIP score, and birthweight) household-level controls (age and gender of the household head, the number of household members, the fraction of members in five age categories) and community-level controls (baseline community averages of the height-for-age, weight-for-age, TVIP score, participation to weight controls and vitamin and deworming intake, and municipal fixed effects). Variations in sample size across tests are mainly driven by the fact that different tests apply to different age groups. Within age groups, it is also due to a limited number of missing observations (see appendix 2 for details).

Table 4: Impact on early childhood development outcomes : Mean effect size by family of outcome

	All outcomes		Cognitive and socio-emotional outcomes		Health and motor development		N. Obs
<u>2006</u>							
Age and gender controls only	0.0395	(0.046)	0.1007**	(0.040)	-0.0217	(0.058)	N=3326
Extended controls	0.0876***	(0.028)	0.1211***	(0.028)	0.0541	(0.035)	N=3326
Excluding caregiver-reported tests	0.1246***	(0.026)	0.1921***	(0.035)	0.0795**	(0.032)	N=3305
Sample same tests 2006-2008	0.0706**	(0.030)	0.0978***	(0.031)	0.0434	(0.040)	N=3149
Mother is titular	0.0697**	(0.032)	0.1019***	(0.030)	0.0375	(0.041)	N=2423
<u>2008</u>							
Age and gender controls only	0.0314	(0.043)	0.0646	(0.043)	-0.0085	(0.049)	N=4245
Extended controls	0.0758***	(0.025)	0.0827***	(0.029)	0.0674***	(0.026)	N=4245
Excluding caregiver-reported tests	0.0607**	(0.026)	0.0949**	(0.042)	0.0265	(0.026)	N=4228
Sample same tests 2006-2008	0.0755***	(0.029)	0.0964**	(0.043)	0.0546*	(0.028)	N=3149
Mother is titular	0.0651**	(0.026)	0.0736**	(0.030)	0.0548*	(0.030)	N=2917

Note: *** p<0.01, ** p<0.05, * p<0.1. Coefficients for index of family of outcomes (estimated with SUR following Kling, et al., 2007); standard errors (in parentheses) adjusted for clustering at the community level. See table 3 for info controls. "Excluding caregiver reported tests" excludes all tests that in part are reported by the caregiver (Denver and BPI). Sample same tests 2006-2008 only includes children who took a given test twice. "Mother is titular" restricts sample to children whose mother was the recipient of the cash in the household.

Table 5: Impact on household-level per capita consumption, by treatment

	3 TREATMENT PACKAGES		BY TREATMENT PACKAGE				F-test	t-test
	(1)	(2)	Basic	Training	Lum sum payment	3 packages	Basic vs Grant	
						P-value	P-value	
<u>2006</u>								
No controls	0.293*** (0.051)	0.287*** (0.057)	0.278*** (0.052)	0.314*** (0.050)		0.315	0.391	
Extended controls	0.299*** (0.028)	0.281*** (0.032)	0.285*** (0.031)	0.331*** (0.031)		0.061	0.083	
<u>2008</u>								
No controls	0.030 (0.037)	0.008 (0.043)	0.024 (0.041)	0.054 (0.037)		0.235	0.109	
Extended controls	0.054** (0.023)	0.022 (0.028)	0.048 (0.030)	0.088*** (0.026)		0.044	0.012	

Note: *** p<0.01, ** p<0.05, * p<0.1. Standard errors (in parentheses) adjusted for clustering at the community level. Controls include baseline log per capita consumption and household and community-level controls as defined in table 3. The number of households is 2212 for 2006 and 2561 for 2008. This is higher than in table 1, as split-off households with sample children are included.

Table 6: Differences in early childhood development outcomes with lump sum payment package versus basic package

	All outcomes		Cognitive and socio-emotional outcomes		Health and motor development		N. Obs
<u>2006</u>							
Age and gender controls only	-0.0047	(0.026)	-0.0127	(0.030)	0.0034	(0.029)	1625
Extended controls	0.0183	(0.023)	0.0160	(0.028)	0.0205	(0.023)	1625
<u>2008</u>							
Age and gender controls only	-0.0242	(0.026)	-0.0385	(0.026)	-0.0070	(0.033)	2114
Extended controls	0.0072	(0.024)	-0.0079	(0.026)	0.0253	(0.029)	2114

Note: *** p<0.01, ** p<0.05, * p<0.1. Coefficients for index of family of outcomes (estimated with SUR following Kling, et al., 2007); standard errors (in parentheses) adjusted for clustering at the community level. Controls and categories as defined in Table 3.

Table 7: Differences between mothers in households with lump sum payment package versus basic package

	2006			2008		
	mean basic	coef	se	mean basic	coef	se
<u>Economic Activity mother:</u>						
Number of days in year work in :						
agricultural wage work	3.41	1.638	(2.073)	8.92	-3.875**	(1.817)
non agricultural wage work	17.04	-5.394	(4.345)	16.26	-8.038**	(3.217)
non agricultural self employment	31.86	41.02***	(6.923)	55.11	25.83***	(7.171)
professional wage job	14.10	-3.399	(3.504)	12.92	-2.172	(3.129)
of which days in seasonal migration	6.26	0.412	(1.726)	20.61	-3.932*	(2.281)
TOTAL days	70.77	32.57***	(9.295)	107.20	9.883	(8.137)
<u>Time mother allocates to child</u>						
tells stories to child	0.65	0.031	(0.030)	0.68	0.007	(0.022)
read stories to child	0.13	0.033	(0.024)	0.08	0.002	(0.015)
number of hours reading per week	0.30	0.114	(0.110)	0.18	0.046	(0.035)
Total number of hours caregiving per day		na		6.31	0.170	(0.147)
Number of hours uniquely caregiving per day		na		2.87	-0.009	(0.100)
Number of hours caregiving while working per day		na		3.45	0.179*	(0.099)
<u>Environment</u>						
CESD depression scale	10.96	0.324	(0.672)	13.75	-0.345	(0.601)
home	3.76	0.103	(0.182)	3.83	0.151	(0.142)

Note: *** p<0.01, ** p<0.05, * p<0.1. Standard errors adjusted for clustering at the community level. Estimations for economic activity of the mother and for environmental factors control for mother's education, household-level and community level baseline characteristics as defined in table 3. Estimations for time mother allocates to child (estimated at the child-level), in addition have controls for child age and gender. The sample only includes mothers in household eligible for the basic treatment and the lump sum payment package. The sample for estimations for economic activity are mothers that were household members at baseline at follow-up: 1073 mothers in 2006 (of 1527 sample children) and 1260 mothers in 2008 (of 1994 sample children). Info on time allocated to specific child is only available if mother is main caregiver: available for 1019 mothers of 1458 children in 2006 and 1163 mothers of 1854 children in 2008. Depression scale and HOME for mothers who are main caregivers: available for 937 mothers in 2006 and 1151 mothers in 2008.

Table 8: Impact on intermediary inputs

2006	Mean control	3 TREATMENT PACKAGES TOGETHER ind, hh and com controls				BASIC PACKAGE	
		(1) coef	se	add. control for log(pce) and log(pce)^2 (2) coef	se	(3) coef	se
<u>Nutrition:</u>							
% food in total expenditures	0.707	0.009	(0.008)	0.010	(0.008)	0.005	(0.009)
% staples in total food expenditures	0.567	-0.093***	(0.011)	-0.066***	(0.011)	-0.087***	(0.013)
% animal proteins in total food exp.	0.155	0.087***	(0.008)	0.060***	(0.008)	0.082***	(0.009)
% fruit and vegetables in total food exp.	0.073	0.029***	(0.005)	0.020***	(0.005)	0.028***	(0.006)
INDEX (of standardized outcomes)		0.422***	(0.036)	0.299***	(0.036)	0.393***	(0.043)
<u>Stimulus</u>							
has pen and paper in house	0.682	0.113***	(0.026)	0.102***	(0.026)	0.107***	(0.028)
somebody tells stories/sings to child	0.527	0.124***	-0.029	0.080***	(0.030)	0.088***	-0.033
Number of hours read to per week	0.134	0.295***	(0.061)	0.266***	(0.062)	0.145**	(0.059)
has toy in house	0.271	0.068***	(0.023)	0.038	(0.026)	0.059*	(0.030)
INDEX (of standardized outcomes)		0.258***	(0.037)	0.204***	(0.039)	0.183***	(0.040)
<u>Health</u>							
weighed	0.735	0.044**	(0.018)	0.035**	(0.018)	0.050***	(0.019)
got vitamins or iron	0.750	0.082***	(0.017)	0.072***	(0.018)	0.097***	(0.020)
got deworming drugs	0.567	0.059***	(0.021)	0.036	(0.023)	0.043*	(0.026)
number of days sick in bed (last month)	0.623	-0.327***	(0.123)	-0.425***	(0.144)	-0.357***	(0.133)
INDEX (of standardized outcomes)		0.131***	(0.022)	0.117***	(0.024)	0.138***	(0.024)
<u>Environment</u>							
CESD depression scale	11.88	-0.605	(0.749)	-0.328	(0.779)	-0.480	(0.696)
HOME scale	4.018	-0.265	(0.291)	-0.088	(0.284)	-0.204	(0.308)
INDEX (of standardized outcomes)		0.079	(0.075)	0.032	(0.075)	0.061	(0.072)
ALL RISK FACTORS : INDEX (standardized outcomes)		0.222***	(0.021)	0.163***	(0.022)	0.194***	(0.024)

2008	Mean control	3 TREATMENT PACKAGES TOGETHER ind, hh and com controls				BASIC PACKAGE	
		(1) coef	se	add. control for log(pce) and log(pce)^2 (2) coef	se	(3) coef	se
<u>Nutrition:</u>							
% food in total expenditures	0.719	-0.008	(0.009)	-0.007	(0.008)	-0.008	(0.009)
% staples in total food expenditures	0.589	-0.025***	(0.009)	-0.019**	(0.008)	-0.026***	(0.009)
% animal proteins in total food exp.	0.161	0.022***	(0.008)	0.017**	(0.007)	0.024***	(0.009)
% fruit and vegetables in total food exp.	0.064	0.009*	(0.005)	0.007	(0.005)	0.009*	(0.005)
INDEX (of standardized outcomes)		0.096***	(0.035)	0.074**	(0.035)	0.101***	(0.037)
<u>Stimulus</u>							
has pen and paper in house	0.824	0.039**	(0.017)	0.037**	(0.017)	0.044*	(0.025)
somebody tells stories/sings to child	0.600	0.066***	(0.023)	0.062***	(0.022)	0.067**	(0.027)
Number of hours read to per week	0.191	0.039	(0.035)	0.032	(0.035)	0.006	(0.040)
has toy in house	0.849	0.081***	(0.028)	0.078***	(0.028)	0.079**	(0.033)
INDEX (of standardized outcomes)		0.129***	(0.034)	0.121***	(0.033)	0.120***	(0.043)
<u>Health</u>							
weighed	0.646	0.007	(0.025)	0.005	(0.025)	0.003	(0.027)
got vitamins or iron	0.558	0.079***	(0.025)	0.076***	(0.025)	0.064**	(0.029)
got deworming drugs	0.547	0.070***	(0.022)	0.069***	(0.022)	0.066***	(0.024)
number of days sick in bed (last month)	0.669	-0.047	(0.107)	-0.053	(0.107)	-0.101	(0.116)
INDEX (of standardized outcomes)		0.084***	(0.024)	0.082***	(0.024)	0.078***	(0.025)
<u>Environment</u>							
CESD depression scale	14.00	0.027	(0.640)	0.036	(0.640)	-0.039	(0.785)
HOME scale	4.072	-0.081	(0.120)	-0.078	(0.119)	-0.128	(0.135)
INDEX (of standardized outcomes)		0.017	(0.042)	0.016	(0.041)	0.031	(0.047)
ALL RISK FACTORS : INDEX (standardized outcomes)		0.081***	(0.018)	0.073***	(0.017)	0.083***	(0.021)

Note: Coefficients and standard errors (in parentheses). Standard errors adjust for clustering at the community level. Negative sign on CESD and Home scale indicate improvement. Coefficients for index of family of outcomes calculated following Kling, et al. (2007) and giving Nutrition, Stimulus, Health and Environment equal weight. Individual, household and community-level controls as defined in table 3. Results for full sample of children with at least one test. N=3326 for 2006; N=4245 for 2008.

Appendix 1: Design of *Atención a Crisis* and its evaluation

Further information on the *Atención a Crisis* interventions

From November 2005 until December 2006, the Ministry of the family in Nicaragua (MIFAMILIA) implemented the *Atención a Crisis* pilot program in 6 municipalities in the northern part of the country. The municipalities were selected for their extreme levels of poverty and because they had been affected by a severe drought in the previous year. The program had two objectives. First, it aimed to serve as a short-run safety net by providing cash transfers to reduce the need for adverse coping mechanisms, such as taking children out of school or reductions in food consumption. Second, the program intended to promote long run upward mobility and poverty reduction by enhancing households' asset base and income diversification capacity.

A total of 3002 households were selected to participate in the program. These households were allocated one of three different packages through a participatory lottery:

- (i) A conditional cash transfer (CCT).
- (ii) A conditional cash transfer plus a scholarship that allowed one of the household members to participate in a vocational training course.
- (iii) A conditional cash transfer plus a productive investment grant, aimed at encouraging recipients to start a small non-agricultural activity.

All selected beneficiary households received the basic CCT. All households, including those without children, received bi-monthly payments, adding up to a total transfer of US \$ 145 during the year of the program. Households with children between 7 and 15 enrolled in and attending primary school received an additional US \$ 90 per household, and an additional US \$ 25 per child (with all amounts referring to the total transfer received over the year), conditional on school enrollment and attendance. The school enrollment and attendance requirement was carefully monitored by the ministry, through data received from the primary school teachers. The program

was also meant to include a condition that required that households with children under the age of 6 would take these children to health centers for preventive health check-ups. However, due to implementation problems, these visits to the health centers were not monitored by the program (Aguilera et al. 2006).

In addition to the CCT, one third of the beneficiary households also received a scholarship that allowed one of the adult household members (preferably a member between 15 and 25 years of age) to choose among a number of vocational training courses offered in the municipal headquarters. In addition to covering the costs of the training, the program also compensated the participants for lost wages while in training, up to 6 months (US \$15 per month). The scholarship was conditional on regular attendance at the course. The courses aimed at providing participants with new skills for income diversification outside of subsistence farming. These beneficiaries were also offered labor-market and business-skill training workshops organized in their own communities.

Finally, another third of the beneficiary households received, in addition to the CCT, a grant for productive investments aimed at encouraging recipients to start a small non-agricultural business activity with the goal of asset creation and income diversification. This grant was conditional on the household developing a business development plan, outlining the objectives of the business and proposed investments in new livestock or non-agricultural income-generating activities. Beneficiaries received technical assistance to make a business plan and also participated in business-skills training workshops organized in their own communities.

Due to implementation delays, the vocational training courses had not started at the moment the data of the 2006 follow-up survey were collected. They took place in the fall of 2006. At the time of the 2006 survey, the difference between the vocational training beneficiaries and those of the basic CCT package was hence limited, though they might have had different expectations about

future skills, about related future income and/or expectations about compensation for the time spent in training. The beneficiaries of the productive investment package, on the other hand, had received the largest amount of benefits: at the end of May 2006 they had received \$175 to invest in a nonagricultural activity. The remaining \$25 was to be paid on the next payment day (after survey completion). In addition they had received technical assistance to select the activity and develop a business plan, help which they were still receiving during the 2006 follow-up survey.

All beneficiaries of the *Atención a Crisis* program, regardless of the treatment they were assigned to, were exposed to repeated information and communication efforts by program staff during enrollment and pay-days. These stressed the importance of varied diets, health, and education, and were meant to change household investment and consumption patterns. Program participants were also required to participate in a number of local events and talks ranging from discussions on nutrition practices to workshops on business development and labor market skills. A subset of beneficiary women were selected during the registration assemblies to serve as *promotoras* or leaders of small groups of beneficiary women (approximately 10 per group) in order to further enhance information flows and motivation and to encourage compliance with the various program requirements and conditionalities. Specifically, the *promotoras* were expected to frequently meet with the beneficiaries in their groups to talk about the objectives and the conditionalities of the program.

Further information on the impact evaluation design

The program was targeted to 6 municipalities in the Northwest of Nicaragua.²³ From the list of all communities in the 6 municipalities, 56 intervention and 50 control communities were

²³ The budget for the pilot allowed targeting 3000 households for a one-year period, which was much smaller than the population of the 6 municipalities. The program was therefore allocated randomly. Households were notified that funding of the project implied that the program would last 1 year, and would only cover the treatment communities.

randomly selected through a lottery to which the mayors of the 6 municipalities were invited to attend and participate.²⁴ Baseline data on household assets and household composition were then used to define program eligibility, resulting in the identification of 3002 households to participate in the program. This amounts to more than 90 percent of the households in treatment communities, and includes 95 percent of households with children under 6 years old. The eligibility criteria were determined using the proxy means methodology developed for the RPS and based on the national household data from 2001 (EMNV). Additional discussions with local leaders from each intervention community were conducted to identify possible exclusion or inclusion errors. Based on the discussions with leaders, 3.7 percent of all the households considered were re-assigned from non-eligible to eligible, and 3.7 percent from eligible to non-eligible. To avoid any possible selection bias resulting from the re-assignment by the leaders, the results we present use eligibility by the proxy means as the intent-to-treat (without taking into account the reclassification by the community leaders).

From each eligible household, the female household member who was reported as the primary caregiver was then invited to a registration assembly.²⁵ During the assemblies, the program objectives and its various components were explained, women were asked to enroll in the program, and *promotoras* were selected on a voluntary basis. Among the intent-to-treat households,

Households in the control communities did not receive any program benefits. They were notified that if there was a decision to scale up the program after the initial year, the control communities would be incorporated. People in the treatment communities understood the program was only to last for a year, and people in the control communities knew that there was a possibility they might receive the program the next year. They also knew it was likely to depend on the result of the national elections that were to be organized at the end of that year. In that election, the government changed and the project was not scaled up.

²⁴ Before the lottery, all communities in the 6 municipalities were grouped in pairs based on similarity in road access and microclimate. Through the lottery, one community of each pair was selected as a treatment community, the other as control. In case of uneven number of communities, a “pair” consisted of the largest community and the combination of the two other communities. The identification of communities and community pairs was based on maps and discussions with municipality technical personnel. Communities tend to be geographically separated from each other, which reduces the potential for spillover effects from the treatment to the control communities.

²⁵ In the few cases where there was no adult female in the household, an adult male was selected as the program recipient.

enrollment in the program was about 95%.²⁶ At the end of each assembly, all the beneficiaries participated in a lottery process through which the three packages described above were randomly allocated among the eligible households. Specifically, each beneficiary was asked to randomly draw a ball with 1 of 3 colors from a black, nontransparent bag. For each assembly, the bags contained an equal number of balls of each color, and the total number of balls matched the total number of beneficiaries in the assembly. At the end of the day each color was matched to an intervention package through another lottery attended by all beneficiaries from the community. Hence, at the moment households signed up for the program, they did not know which of the three interventions package they would end up receiving.²⁷ Take-up of the CCT component was the same for all packages. Take up of the additional packages was also high. Among households enrolled in the program, 89 percent of the households eligible for the vocational training grant enrolled a household member in a course. Take-up of the productive investment grant among eligible households in the program was almost 100 percent. About 10 percent of the business development plans had initially been refused by the Ministry of the Family, but these were sent back to the households and virtually all of them developed a new plan, with the help of technical assistance (with the few exceptions being households that migrated out after the registration assemblies).

Given the objectives of the program, the outcomes that the impact evaluation as a whole focuses on are related to household investments in human capital and productive assets. Among human capital outcomes, there was a strong interest on early childhood development outcomes, the focus of this paper.²⁸ This was motivated by anecdotal evidence of early childhood delays in the region of study during baseline pretesting, and given the lack of evidence on CCT effects on ECD

²⁶ Of the 5 percent that did not enroll, the majority are households that had been reclassified by the leaders. The remainder are households that had migrated out of the treatment communities by the start of the program, and (very few) cases of households that refused the program.

²⁷ Due to the transparency of the process, the lottery process was widely perceived as fair. Participation by the invited beneficiaries to the assemblies and lotteries was near 100%.

²⁸ See Macours and Vakis (2009) for short-term impacts on other outcome variables.

outcomes. With this objective, a cognitive test (TVIP) was added to the baseline survey, and data on anthropometrics were collected. In both follow-up surveys a more extensive set of tests was conducted (see appendix 3).

Data were collected at baseline on all households in the treatment communities, and for a random sample of households in the control group. The sample size in the control communities was chosen to be equal to one-third of the population in the treatment, in order to obtain a control group of equal size to each of the treatment groups. The proxy means test used to identify eligible households in the treatment communities was also used to identify households that would have been eligible in the control communities.

Appendix 2: Attrition

Attrition can potentially introduce serious biases into the estimation of program effects. In this study, attrition between the baseline and follow-up surveys was minimal. Only 1.3 percent and 2.4 percent of households interviewed at baseline could not be re-interviewed in 2006 and 2008, respectively. Among children less than 6 years old when the transfers started and living in the sample households at baseline, 4.7 percent of children could not be re-interviewed in the 2006 follow-up, and 3.1 percent in the 2008 follow-up. The low attrition rates are a result of repeat visits to recover temporary absence and extensive tracking of migrants. Migrant households and children were interviewed and tested in their new locations.

Attrition is uncorrelated with treatment—in a regression of attrited households on a dummy for treatment, the coefficient is -0.004 in 2006 (with a standard error of .005), and 0.003 in 2008 (with a standard error of 0.006). In a comparable regression for children the coefficient for 2006 is 0.005, with a standard error of 0.013, and for 2008 the coefficient is 0.0004 with a standard error of 0.009. Similar results are obtained when considering each of the treatment packages separately, and when comparing between treatment packages. Appendix Table A1 shows that the baseline characteristics of the full sample of households and those that could be located at follow-ups are very similar.

Aside from attrition because of failure to re-interview, 6 percent of the children did not do one or more of the tests that rely on their active participation (Denver sub-scores, TVIP, memory, leg motor) in 2006. This was typically due to refusal to participate by extremely shy children, who were not willing to interact with the test administrators in a way that allowed the test to be conducted. This is of potential concern if the treatment affected the willingness of the children to participate. Appendix Table A2 shows that the baseline characteristics of children who did all tests

are very similar to those of all children located at follow-up. The share of children who refused to take at least one test is lower in treated than in control communities in 2006, but the difference is not significant (the coefficient is 0.014 and the standard error 0.012). There are also no significant difference between the basic package and the lump sum payment package (coefficient -0.009, standard error 0.012).

In addition, there are some other children for whom the full set of outcome variables is not available, either because of missing anthropometric measures, which was typically due to logistical problems with the anthropometric material (6 percent of children in 2006) or because the behavioral problem index is missing, typically because the caregiver could not be located by the test administrator (8 percent of children in 2006). There is no significant difference between treatment and control in the likelihood that the full set of outcome variables is available (the coefficient is 0.014 and standard error 0.019).

In 2008, field teams did more repeat visits and attempts to convince the children to participate in case of refusals and to recover missing information for anthropometrics or the behavioral index. As a result, the share of children who refused to participate in one or more tests is only 4 percent, and there is no significant difference between treatment and control communities (the coefficient is -0.003, and the standard error 0.009), nor between the basic and the lump-sum package (coefficient -0.004, standard error 0.009). Similarly, there are fewer children for whom anthropometric measures are missing (1 percent), or for whom the BPI is missing (an additional 4 percent). Here too there are no significant difference between treatment and control on the likelihood that the full set of outcome variables is available or not (the coefficient is as 0.004 and standard error 0.014).

Given that the differences between treatment and control are small and not significant, and given that we control for baseline differences, any resulting selection bias is likely to be small. This

is further confirmed by a robustness check in which only the sample of children that have completed all the tests for their age group is used. The results from this robustness check are very similar to the main results we report—an average program effect of 0.08 standard deviations in 2006 (0.1 standard deviations for cognitive-social emotional outcomes, and 0.05 for health and motor development) and an average program effect of 0.07 standard deviations in 2008 (0.08 standard deviations for cognitive-social emotional outcomes and 0.07 for health and motor development).

Appendix 3: Measurement of early childhood development tests and intermediate inputs

Early childhood development tests

We focus on eleven measures of early childhood development. Height and weight were measured using standardized anthropometric material and procedures for all children. Height-for-age and weight-for-age were calculated using the international norms.

Social-personal, language, fine motor, and gross motor skills for children between 0 and 83 months old were assessed using the four sub-tests of the Denver Developmental Screening Test (Frankenberg and Dodds 1996). For each subtest, the child is asked to perform a number of age-specific tasks. When children fail to perform a task that 75 percent of children of their age in the reference population can perform, the test falls back to easier tasks, up to the point where tasks are reached that the child can perform.²⁹ In case certain behaviors or tasks cannot be observed, the caregiver is asked about the ability of the child to perform them. The social-personal subtest mainly consists of behavior that the caregiver is asked about, such as social interactions, the ability of a child to dress and eat on their own, imitate others, etc. The language subtest covers recognition and use of sounds, words, sentences, etc. The fine motor skills subtest mainly relates to observations of manual tasks such as drawing, playing with cubes, reaching for objects, etc. Finally, the gross motor tasks capture observations of basic crawling, sitting, walking, as well as throwing, jumping, etc. For the language, fine motor and gross motor subtests, most items are scored through direct observation of whether the child can perform the task. A relatively small share (15 out of 39 for language, 1 out of 29 for fine motor, and 5 out of 32 for gross motor) can be scored by asking the caregiver about the child's ability for that specific task, in case the item cannot be directly observed. In the sample, 40 (37) percent of language tests, 8 (4) percent of fine motor tests and 29 (23) percent of gross

²⁹ Similarly, for children performing all tasks for their age group the test continues with more difficult tasks. For the children in our sample this occurred very rarely.

motor tests have at least one item administered through the caregiver's report in 2006 (2008). Importantly, neither in 2006 nor in 2008 is there a significant difference between treatment and control on the likelihood that the Denver items were obtained uniquely through direct observation, or in the number of items obtained through self-reporting (P-values of the significance tests range from 0.23 to 0.97). Hence, caregivers in the treatment are not more likely to provide answers on behalf of their children, further reducing concerns regarding potential differences in reporting. Moreover, as discussed in section 4, our findings are robust to excluding the Denver subtests from the estimations.

The Denver scores are based on the number of tasks a child fails to perform, when these tasks can be carried out by more than 90 percent of children of the same age in the reference population.³⁰ The Denver is designed for children between 0 and 6 years of age. For this study, the test was also applied to somewhat older children, given the substantial delays in cognitive development that exist in our sample (described in detail in the paper).³¹ For children age 36 months or older we applied five additional tests. The first of these is the TVIP, the Spanish-speaking version of the Peabody Picture Vocabulary Test (PPVT), a test of receptive vocabulary that has been widely used in developed and developing countries.³² Children are shown a series of slides with four pictures each (for example, the first slide has a picture of a flashlight, a boat, a basket, and a hot-air balloon), and are asked to point at a given object stated by the enumerator (for example, "boat"). Test items gradually become more difficult. The test administrator records the number of correct and incorrect responses, and the test stops when a child is making as many errors as she

³⁰ The Denver has been used in other studies of early childhood development in developing countries, including in Nicaragua (Oberhelman et al. 1998). Other applications in developing countries include Halpern et al. (1996); Cheung et al. (2001); Choudhury and Gorman (2003); and Dewey et al (2001).

³¹ Results are robust when only including the Denver test results for children up to 72 months of age (to reflect the age range the Denver was originally designed for). They are also robust to alternative ways of scoring the Denver test (accounting for the number of delays and cautions, or alternatively using a binary indicator on whether the child has at least one, or alternatively two, delays).

³² See, for example, Paxson and Schady (2007, 2010), Schady (2011), Umbel et al. (1992), Baydar and Brooks-Gunn (1991), Blau and Grossberg (1992), Rosenzweig and Wolpin (1994), and Fernald, Gertler, and Neufeld (2008).

would be expected to make if she were randomly guessing.³³ We also use a short-term memory test and a leg motor test from the McCarthy test battery. In the memory test, the test administrator reads increasingly long sequences of numbers to the child, and asks the child to repeat them. The leg motor test measures the ability of children to execute six predetermined tasks—for example, walking on tiptoes or backwards, and standing on one foot.³⁴ In 2008 an associative memory test of the Woodcock-Muñoz test battery was also used.³⁵ In this test, children are introduced to an increasing number of space creatures with nonsensical names, and are asked to identify the different figures on a page that shows many of them at the same time. The final test we use is the Behavior Problem Index (BPI), which is based on the caregiver’s report of the frequency that a child displays each of 29 problematic behaviors, with responses coded as “never”, “sometimes” and “often”.³⁶ We use the number of behavioral problems for which a caregiver answers “often”. Unlike the other outcomes we study, behavioral problems do not necessarily indicate a delay, as there are no benchmarks or established ages at which they are predicted to decrease.³⁷

Our analysis focuses on children below 6 years old when the transfers started. This means that all of them are below 7 years at the 2006 follow-up. The program requirement of primary school enrollment and attendance was binding for children age 7 and above. None of the children in our sample are bilingual—an obvious concern with tests that measure language ability. All of the tests were carefully pre-tested in the field and a handful of items that appeared to be culturally inappropriate were amended. The TVIP is standardized with a population of Mexican and Puerto

³³ Before the test starts, the enumerator explains the test with the help of a few example slides. She proceeds to the actual test slides only once the child has demonstrated understanding of the test.

³⁴ See Stoltzfus et al. (2001), Gertler and Fernald (2004), Fernald, Gertler, and Neufeld (2008), and Cogill et al. (1986) for other applications.

³⁵ The associative memory tests was added for comparability with other studies on ECD outcomes in Latin American countries, in particular Paxson and Schady (2010), Schady (2011) and Fernald, Gertler, and Neufeld (2008).

³⁶ Recent applications of the BPI in Latin America include Fernald, Gertler, and Neufeld (2008); Paxson and Schady (2010).

³⁷ There is some overlap between the BPI and the social-personal behaviors measured in the Denver. For instance, the Denver personal-social subtest has a number of items that relate to social interactions; and the BPI also has questions about whether or how the child interacts with others.

Rican children, and the words are all part of the vocabulary in Nicaragua. This was further verified during pre-testing.³⁸ In the case of the Denver test of language, there also does not appear to be an obvious concern with cultural appropriateness—the test measures whether infants can utter various sounds and, for older children, whether they can identify and name simple concepts, such as body parts. Additional evidence that the Denver is appropriate for our study population is provided by the fact that the national early childhood stimulation program in Nicaragua uses a slightly modified version of this test for child monitoring. For all these reasons, it seems unlikely that the observed delays are due to possible cultural inappropriateness of the tests for children we study.

All tests were administered by a specially trained team of female-only test administrators. They were selected for their background (training as psychologists, social workers, or similar) and for their ability to establish good contact with small children. During the training, high emphasis was placed on gaining the confidence of the children before starting the test administration and on the standardized application of each of the tests. Moreover, data collection and test administration was organized in such a way that the test administrators would maintain a balance between the number of children visited in treatment and control communities, and visits to treatment and control were also balanced in time, to avoid any seasonal differences. Consistent with this approach, the results are robust to controls for the identity of the test administrator and for the month the tests was taken.

Measurement of intermediate inputs

³⁸ Paxson and Schady (2007) show that in rural Ecuador children whose mothers or fathers have completed secondary schooling have average scores that place them in the 50th percentile of the test, indicating they perform as well as the international reference population, even if many others in the same setting also have very large delays. In our sample, only a very small number of parents (5.3 percent of mothers, and 4.5 percent of fathers) have completed secondary school, preventing us from carrying out a similar calculation. Nevertheless, it is telling that the results of the Denver reported in Oberhelman et al. (1998) for a Nicaraguan population that has much higher education levels show much smaller delays.

As in most surveys, measures of input use are obtained by caregiver reports (for child food intake and stimulus indicators), or by reports by the household head or his spouse (for household food consumption and health indicators). Given the social marketing component of the intervention, this raises potential concern regarding reporting bias, as informants in the treatment could, in principle, be more likely to provide answers that they believe are socially desirable. The social marketing did not include messages on early childhood stimulation, making it less likely that we would observe reporting bias on those indicators. Also, while reporting error might be of concern for the measurement of other intermediate inputs during the intervention (in 2006), such concerns a priori seem to be much less for the results for 2008, when the intervention had concluded for approximately two years. Moreover, for at least one measure, the self-reported measures of growth check-ups, we can double-check the information using a vaccination-health use card filled in by health care providers when children visit the health center. Enumerators directly observed the presence of those cards. In 2006, the share of children with cards is higher in the treatment group, consistent with the fact that the health use cards are given to parents during growth check-ups. Among children with cards, however, there is no difference between treatment and control groups in the extent to which the information derived from the cards is consistent with answers provided by respondents in the interview. Hence, the finding that a higher share of children went to growth controls in the treatment is unlikely to be due to overreporting of desirable behavior by treatment parents. For 2008, children in the treatment are no longer more likely to be weighed and, consistently, they are also no longer significantly more likely to have vaccination-health use cards. In 2008, too, the level of concordance between parental reports and the cards is not different between treatment and control groups.

Table A1: Baseline characteristics of all eligible households compared to eligible households tracked back at follow-ups

	Mean all	Mean all reinterviewed	
		2006	2008
Household characteristics			
male household head	0.82	0.82	0.83
household size	5.28	5.30	5.33
# hh members 0-5 years old	0.65	0.65	0.65
# hh members 5-14 years old	1.53	1.54	1.55
# hh members 15-24 years old	1.07	1.08	1.08
# hh members 25-64 years old	1.77	1.78	1.79
# hh members more than 65 years old	0.24	0.24	0.24
Number of rooms in the house	1.62	1.63	1.63
Time to school (minutes)	0.27	0.27	0.27
Time to health center (minutes)	1.14	1.14	1.14
Time to municipal headquarters (minutes)	1.54	1.54	1.54
Owns toilet/latrine	0.76	0.76	0.76
Access to water	0.13	0.13	0.13
Access to electricity	0.40	0.40	0.40
Own land	0.66	0.66	0.66
Total Consumption per capita (cordobas)	5443	5424	5410
Food consumption per capita (cordobas)	3537	3527	3524
Number of households	4021	3969	3923

Note: Sample includes all eligible households in treatment and control

Table A2: Baseline characteristics of eligible children who did all age-relevant tests compared to all eligible children

	2006		2008	
	All	Did all tests	All	Did all tests
Child-specific characteristics				
All children				
male	0.50	0.50	0.50	0.50
age in months when transfers started	32.80	32.45	21.10	21.38
mother lived in household at baseline	0.95	0.95	0.96	0.96
# years education mother	3.98	4.01	4.08	4.08
# years education father	3.65	3.67	3.82	3.83
Children age 3-6 at baseline				
tvip (vocabulary recognition) test score	6.01	6.05	6.00	6.02
Children age 0-5 at baseline				
weight-for-age z-score	-1.02	-0.98	-1.01	-1.01
height-for-age z-score	-1.23	-1.19	-1.22	-1.22
weight-for-height z-score	-0.16	-0.14	-0.17	-0.17
birth weight	6.79	6.81	6.79	6.79
weighed in last 6 months	0.91	0.91	0.91	0.91
Received vitamins in last 6 months	0.69	0.69	0.69	0.70
Received deworming drugs in last 6 months	0.53	0.52	0.53	0.54
Household-level characteristics				
male household head	0.85	0.86	0.85	0.85
household size	6.09	6.10	6.19	6.20
# hh members 0-5 years old	1.16	1.17	1.08	1.09
# hh members 5-14 years old	1.72	1.72	1.76	1.76
# hh members 15-24 years old	1.18	1.18	1.27	1.27
# hh members 25-64 years old	1.87	1.86	1.90	1.91
# hh members more than 65 years old	0.15	0.15	0.16	0.15
Number of rooms in the house	1.59	1.58	1.60	1.61
Time to school (minutes)	0.27	0.28	0.27	0.27
Time to health center (minutes)	1.20	1.21	1.20	1.20
Time to municipal headquarters (minutes)	1.63	1.63	1.61	1.62
Owns toilet/latrine	0.72	0.72	0.73	0.73
Access to water	0.12	0.12	0.12	0.12
Access to electricity	0.37	0.37	0.37	0.37
Own land	0.63	0.64	0.64	0.65
Total Consumption per capita (cordobas)	4548	4547	4561	4567
Food consumption per capita (cordobas)	3095	3098	3102	3111
Proportion of food in total expenditures	0.69	0.69	0.69	0.69
Proportion of staples in all food exp.	0.59	0.58	0.59	0.59
Proportion of animal proteins in all food exp.	0.16	0.16	0.16	0.16
Proportion of fruit and vegetables in all food exp.	0.05	0.05	0.05	0.05
Number of children	3333	3135	4316	4153
Number of households	2221	2137	2604	2539

Note: Sample is based on all children in each year visited by the test administrator that were either younger than 6 years when transfers started or born to baseline household members since the baseline.