

Cash Transfers, Behavioral Changes, and the Cognitive Development of Young Children: Evidence from a Randomized Experiment⁺

Karen Macours*, Norbert Schady[°], and Renos Vakis[°]

April 2008

* Johns Hopkins University; [°] World Bank

⁺ We are grateful to Ximena Del Carpio 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, Carlos Obregón and the entire field team) for excellent data collection. Financial support for this research was received from ESSD trust funds, a RSB grant, as well as the Government of the Netherlands through the BNPP program. Seminar participants at the World Bank, the Inter-American Bank and the University of California, Berkeley provided 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: kmacours@jhu.edu, nschady@worldbank.org and rvakis@worldbank.org

Abstract

This paper analyzes the impact of a randomized cash transfer program on cognitive development in early childhood in rural Nicaragua. We first show that a very large sample of children in our sample are delayed along various dimensions, including language, memory, and social skills, and that these delays are related to household per capita expenditure levels. We then show that the program had significant effects on cognitive outcomes, especially language. Program impacts are larger for older pre-school aged children, who were also more likely to be delayed. We analyze how the program affected several inputs into early childhood development that have been identified as key risk factors in the literature. Households randomly assigned to receive transfers increased the consumption of nutritious foods and provided more cognitive stimulation to children than those assigned to the control group. Moreover, we show that these impacts are not due to the cash component of the program only. The expenditure patterns of beneficiary households are significantly different from those of control households, indicating that households do not spend the cash received through the program in the same way as other income.

I. Introduction

Cognitive 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 1999; Case and Paxson 2006). 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 had higher school attainment, better test scores, lower rates of criminality, and earned higher wages in adulthood (Currie 2001; Schweinhart 2005), although the impacts appear to be concentrated among girls (Anderson 2007). A well-known study from Jamaica shows that children randomly assigned to receive home-based early stimulation showed 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 attendance at nursery programs and preschool is associated with better school performance in Argentina (Berlinski et al. 2006), Uruguay (Berlinski et al. 2007), Colombia (Attanasio and Vera-Hernández 2004) and Bolivia (Behrman, Cheng, and Todd 2004). 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).

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 the cognitive development of their children—either by relieving financial constraints, or by changing how resources are allocated within households.

This paper therefore analyzes the impact of a cash transfer program on cognitive development outcomes in early childhood in rural areas of Nicaragua. The program, known as *Atención a Crisis*,

makes sizeable payments, equivalent to about 15 percent of per capita expenditures for the average recipient household. Households eligible for the program were randomly assigned to treatment and control groups. A follow-up survey collected data on both groups approximately 9 months after the treatment group started receiving transfers. Random assignment allows us to estimate program effects with very few assumptions.

There are a variety of reasons why one might expect a program like *Atención a Crisis* to have a positive effect on cognitive development in early childhood. Although the evidence is sparse, children in better-off households appear to have higher levels of cognitive development than those in poorer households in developing countries (see Paxson and Schady 2007, 2008 on Ecuador; Halpern et al. 1996 on Brazil; see also Schady 2006 for a discussion). These associations may not be causal—rather they may reflect a correlation between parental wealth, parental genetic endowments, and child cognitive development. 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 the cognitive development of children.

There are other features of the *Atención a Crisis* program that could result in improvements in cognitive development. *Atención a Crisis* transfers are made to women, and there is some evidence that income controlled by women is spent in a way that may benefit children more than income that is controlled by men. For example, Thomas (1990), Hoddinott and Haddad (1995), Doss (2005), and Schady and Rosero (2007) show that income controlled by women is associated with higher expenditures on food. Macours and Vakis (2007) show non-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) 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, although the conclusions from this study have been challenged by Hotchkiss (2005).

In order to remain eligible for *Atención a Crisis* transfers, parents in the program were meant to ensure that school-aged children enroll in school and attend regularly, and take preschool-aged children

for regular visits to health centers, where they are weighed, and receive vaccinations, micronutrients, or food supplements, as necessary.¹ In this sense, the program was similar in nature to other conditional cash transfer programs, such as the much-studied PROGRESA program in Mexico. The education and health conditions attached to the transfers imply changes in relative prices, which could result in shifts in expenditures towards human capital investments.

The *Atención a Crisis* program also included a social marketing campaign—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 transmit knowledge about child-rearing practices; it may also affect how transfer income is used through a flypaper or labeling effect—see Thaler (1999) for a general discussion, and Fraker et al. (1995) and Kooreman (2000) for evidence from the US and the Netherlands, respectively. Such changes in behaviors could be further enhanced through social interactions with other program beneficiaries and peer pressure (Macours and Vakis 2008).

Understanding the impact of a program like *Atención a Crisis* on cognitive development in early childhood is important for a number of reasons. One reason is the popularity of cash transfer programs in the developing world. A recent World Bank (2008) study estimates that at least 24 developing countries have a conditional cash transfer program in place, and many others have programs that transfer cash without conditions. In many cases, including in Mexico, Brazil, Ecuador, and South Africa, the cash transfer program is the biggest safety net program in the country.

A large number of studies have assessed the impact of cash transfers, conditional and unconditional, on health status, nutrition, and education.² In contrast, we are aware of only two earlier studies which estimate the impact of a cash transfer program on cognitive development. Fernald, Gertler,

¹ As we discuss below, there was an implementation glitch whereby compliance with the health condition was not monitored by the program.

² The literature is extensive. On education outcomes see, among others, Schultz (2004) and Behrman, Sengupta and Todd (2005) on Mexico; Schady and Araujo (2008) on Ecuador; Attanasio et al. (2005) on Colombia; Glewwe and Olinto (2004) on Honduras; Maluccio and Flores (2005) on Nicaragua; and Edmonds (2006) on South Africa. Macours and Vakis (2008) show education impacts of *Atención a Crisis*. On health outcomes see, among others, Gertler (2004), Behrman and Hoddinott (2005), and Rivera et al. (2004) on Mexico, Morris et al. (2004a) on Honduras; Morris et al. (2004b) on Brazil, and Duflo (2003) and Agüero et al. (2006) on South Africa. World Bank (2008) reviews the findings from these studies.

and Neufeld (2008) suggest that larger transfers made by the PROGRESA-Oportunidades program in Mexico resulted in better nutritional status, improved motor skills, and higher levels of cognitive development. However, the outcomes they study were collected too long after the initial control group in the study had been folded into the program for the authors to use the initial random assignment for identification. Paxson and Schady (2008) use random assignment in the roll-out of the *Bono de Desarrollo Humano* cash transfer program in Ecuador to analyze the effects on health and cognitive development of children between 3 and 6 years of age. They show that cash transfers resulted in an improvement of about 0.25 standard deviations in cognitive 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 variety of ways. First, and unlike earlier work on cash transfer programs, we collected data on measures of cognitive development for children as young as one month of age. We can therefore estimate the impact of the program among young children, and assess whether this impact is larger than among older children—as might be expected if early childhood is a very sensitive period for development, and if the potential for later catch-up is limited. Second, our data includes an extensive module on household per capita expenditures. As a result, we can see whether households randomly assigned to receive transfers increased spending on a variety of inputs into child cognitive development, such as the quantity and diversity of food, early stimulation, and health care. This allows us to provide better evidence on the transmission mechanisms from cash transfers to changes in cognitive development than has been the case in earlier papers. In particular, we can test whether *Atención a Crisis* transfers were used like other sources of income.

The rest of the paper proceeds as follows. In section 2, we describe the *Atención a Crisis* pilot program, our identification strategy, and the data, in particular the measures of cognitive development. Section 3 describes the frequency of early childhood development delays in our sample. We present the main results in the paper, including a disaggregation of program effects by age and gender, in section 4. In section 5 we discuss changes in intermediate inputs, and whether these can be explained entirely by the increase in overall expenditures due to the transfer. Section 6 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 Nicaragua.³ The pilot was targeted to 6 municipalities in rural Nicaragua that met two criteria—having been affected by a drought the previous year and having a high prevalence of extreme rural poverty based on the national poverty map. The beneficiaries of the pilot randomly received one of three packages: (i) a conditional cash transfer (CCT) conditional on children’s primary school and health service attendance; (ii) the CCT plus a scholarship that allowed one of the household members to choose among a number of vocational training courses offered in the municipal headquarters. These beneficiaries also participated in labor-market (and business-skill) training workshops organized in their own communities; and (iii) the CCT plus a productive investment grant, aimed at encouraging recipients to start a small non-agricultural activity with the goal of asset creation and income diversification. This grant was conditional on the household developing a business development plan.

The design of the CCT component of *Atención a Crisis* was modeled after the existing CCT program in Nicaragua, the *Red de Protección Social* (RPS).⁴ Women in beneficiary households receive sizable cash transfers every 2 months, averaging about 15 percent of per capita expenditures.⁵ The CCT component of the *Atención a Crisis* pilot differs from RPS mainly on its reliance on public health infrastructure, as opposed to the private health providers used in RPS. This led to some implementation problems. Specifically, the anticipated increase in health service supply did not occur, and children’s visits to the health centers were not monitored during the study period. On the other hand, the school enrollment and attendance requirement was carefully monitored (see Aguilera et al. 2006). The program included repeated information and communications during program enrollment and pay-days about the

³ For an extensive description of the program and evaluation design see Macours and Vakis (2005).

⁴ See Maluccio and Flores (2005) for the impacts on education, health and nutrition of the RPS program.

⁵ Households received a transfer of US \$ 145 even if they did not have children. However, households with children between 7 and 15 enrolled in primary school received in addition US \$ 90 per household, and an additional US \$ 25 per child.

importance of varied diets, health and education; these were meant to change household investment and consumption patterns.

The *Atención a Crisis* pilot included a careful evaluation design. Randomization was used to assign eligible households into one of four groups: control, CCT only, CCT plus vocational training, and CCT plus productive investment grant. This was done as follows. First, from the list of all communities in the 6 municipalities, 56 intervention and 50 control communities were randomly selected through a lottery. Second, baseline 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 7 years of age) in treatment and control communities were ineligible for the program because their estimated baseline consumption, 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.⁶

In communities randomly selected to participate in the *Atención a Crisis* program, the primary child caregiver, 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 in which the three packages described above were randomly allocated among the eligible households. Participation in the assemblies and lotteries was close to 100 percent. Note that, within treatment communities, one third of eligible households were assigned to each of the three treatment packages, and all of them received the CCT. In control communities, households did not receive any of the treatment packages.

B. Data

⁶ The weights in the proxy means were based on estimates from the national household data from 2001 (EMNV). Additional discussions with local leaders from each intervention community were conducted to identify possible exclusion or inclusions errors. Based on the discussions with leaders, 3.72 percent of all the households considered were re-assigned from non-eligible to eligible, and 3.65 percent from eligible to non-eligible. To avoid selection bias, we use the original eligibility, based on the same proxy means test for both treated and control households, as the intent-to-treat.

Baseline data for the evaluation were collected in April-May 2005. A follow-up survey, including a large number of tests to assess cognitive development, was collected in July-August 2006 (9 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 randomly assigned to the control group.

Attrition between the baseline and follow-up surveys was minimal. Only 1.3 percent of households interviewed at baseline, and 4.6 percent of children under the age of 7, could not be re-interviewed at follow-up. Attrition is uncorrelated with treatment—in a regression of attrited households on a dummy for treatment the coefficient is $-.004$, with a standard error of $.005$, and in a comparable regression for children the coefficient is $.004$, with a standard error of $.012$. Further, Appendix Table A1 shows that the baseline characteristics of the full sample of children and those that could be located at follow-up are very similar.

In addition to attrition because of failure to re-interview, 5 percent of the children did not do one or more tests. This was typically due to refusal to participate by extremely shy children, who were not willing to interact with the enumerators in a way that allowed the test to be conducted. Appendix Table A2 shows that the baseline characteristics of children who did all tests are very similar to those of other children located at follow-up. However, the share of children who refused to take at least one test is 2 percentage points lower in treated than in control communities, and this difference is significant at the 10 percent level. Appendix Table A3 shows that, on average, children who did not complete all tests did significantly worse on those tests they took than other children. This might indicate a possible treatment effect on the willingness for social interaction among the treated children, which is consistent with the results we present below. Because children who did not take a particular test are excluded from the sample when we consider the results of that test, we may therefore underestimate the treatment effects on cognitive development. However, given the small number of cases, the magnitude of the bias is likely to be small.

Program take-up in the treatment group was more than 95 percent, and contamination of the control was negligible (one household).⁷ As discussed above (footnote 6), the main reason households did not take-up the program was due to the fact that some originally eligible households were deemed ineligible by local leaders after the initial assignment. 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*.

Table 1 shows that random assignment effectively equated the characteristics of treatment and control groups: of the 34 variables that are summarized in the table, only one (the number of people aged 65 or older in a household) indicates a significant difference between the two groups at baseline. Table 1 also shows that households in our sample are disadvantaged in a number of important ways. The mean years of schooling of mothers are 4 years, and 67 percent have not completed primary school. The mean years of schooling of fathers is even lower, approximately 3.5 years, and 75 percent have not completed primary school. Children in this sample also appear to 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). Finally, expenditure levels are very low. Turning the local currency units in the table (*Córdobas*) into US \$ shows that 82 percent of households in our sample have per capita expenditures that are below one dollar per capita per day.

C. *Early childhood development indicators*

We focus on eight measures of early childhood development. Social-personal, language, fine motor, and gross motor skills for all children 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 which 75 percent of children of their age in the reference population can perform, the test falls back to easier tasks, up to the

⁷ A small fraction of 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.

point where tasks are reached that the child can perform.⁸ 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 her own, imitate others, etc. The language subtest covers recognition and use of sounds, words, sentences, etc. The fine motor skills subtest mainly relates to manual tasks such as drawing, playing with cubes, reaching for objects, etc. Finally, the gross motor tasks capture basic crawling, sitting, walking, as well as throwing, jumping, etc. 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 75 percent (or 90 percent, see below) of children of the same age in the reference population.⁹

For children age 36 months or older we applied four 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.¹⁰ 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 enumerator records the number of correct and incorrect responses, and the test stops when a child is making as many errors as she would be expected to make if she were randomly guessing.¹¹ We also use a short-term memory test and a leg motor test from the McCarthy test battery. In the memory test, the enumerator reads to the child increasingly long sequences of numbers, 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

⁸ Similarly, for children performing all tasks for their age group the test continues with more difficult tasks. For the children in our sample, however, this occurred very rarely.

⁹ 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).

¹⁰ See, for example, Paxson and Schady (2007, 2008), Umbel et al. (1992), Baydar and Brooks-Gunn (1991), Blau and Grossberg (1992), Rosenzweig and Wolpin (1994), and Fernald, Gertler, and Neufeld (2008).

¹¹ Before the tests 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.

on one foot.¹² 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”.¹³ 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.¹⁴

The Denver is designed for children between 0 and 6 years of age. For this study, the test was also applied to older children, given the substantial delays in cognitive development that exist in our sample (described in detail below). Our analysis focuses on children below 7 years at follow-up—0 to 83 months for the Denver, and 36-83 months for all other tests.¹⁵ We do not include the children 7 years or older in the analysis, both because of the age range of the Denver, and because the program requirement of primary school enrollment and attendance was binding for children age 7 and above.¹⁶

We also analyze impacts on intermediate outcomes that may be related to child cognitive development. The survey included an extensive expenditure module taken from the 2001 Nicaragua Living Standards Measurement Study (LSMS) survey. We focus on various expenditure categories, including food expenditures, which include actual expenditures, home production, and food consumed outside the home, and child food intake. The data also include information on stimulation, birthweight, children’s weight and height, 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 the HOME score, an index of 11 positive

¹² See Stoltzfus et al. (2001), Gertler and Fernald (2004), Fernald, Gertler, and Neufeld (2008), and Cogill et al. (1986) for other applications.

¹³ Recent applications of the BPI in Latin America include Fernald, Gertler, and Neufeld (2008); Paxson and Schady (2008).

¹⁴ 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.

¹⁵ Given that the program had been implemented for 9 months, the duration of exposure to the program is the same for all, including the youngest children, if one includes the time in-utero.

¹⁶ As discussed above, this implies that households with children 7 years or older received an extra amount of cash for school supplies.

and negative behaviors that the enumerator observes during interviewing and testing (Bradley 1993; Paxson and Schady 2007, 2008).

3. Delays in cognitive development in early childhood

We first describe cognitive development outcomes of children in our sample, focusing on the control group. Table 2 shows the fraction of children 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. The first point to note from the table is that a very large fraction of children in our sample appear to be delayed, although this varies considerably by outcome. The fraction of children who are behind for their age is largest for the measures of language—97 percent of children in our sample are in the lowest quartile of the distribution of the TVIP, and 85 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 61 percent in the lowest decile. A very large fraction of children in our sample are also delayed in short-term memory—85 percent place in the lowest quartile, and 61 percent in the lowest decile of the distribution used to standardize the test.

These delays in language and memory are very 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 in Latin America. For example, in the sample of children in Ecuador analyzed by Paxson and Schady (2007, 2008), which is considerably better off than the sample used for this paper, the average child places in the 11th percentile of the distribution of the TVIP.¹⁷ In the case of performance on memory, the tests used are not strictly comparable, but the average child in the Ecuador sample places in the 29th percentile of a test

¹⁷ 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 our study. There are also marked differences in parental education, which is very robustly associated with performance on the TVIP—the average education of mothers in the Ecuador sample is 6.7, compared to 4.1 for the sample used in our paper.

of short-term memory, and in the 13th percentile of a test of long-term memory from the Woodcock-Johnson battery of tests.

It is unlikely that the observed delays are due to cultural inappropriateness of the tests for children we study. 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 was standardized with a population of Mexican and Puerto Rican children, and the words are all part of standard vocabulary in Nicaragua. This was further verified during pre-testing.¹⁸ 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.

Turning to other domains of child development, Table 2 shows that outcomes are somewhat better on the social-personal scale of the Denver—46 percent of children in the sample place in the lowest decile—and for fine motor skills—40 percent place in the lowest decile for this outcome. 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 24 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 how these patterns vary with the age and gender of the child. In terms of gender, there appear to be no obvious differences—the fractions delayed are generally similar for boys and girls. However, there

¹⁸ 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.

does appear to be a pattern whereby delays increase with child age for some outcomes. In the case of language, the fraction of children who place in the lowest decile of the distribution of the language measure of the Denver increases from 48 percent for children ages 0-35 months, to 59 percent for children aged 36-59 months, and to 79 percent for children aged 60-83 months; comparable numbers for the TVIP, which can only be applied to children ages 36 months and older, are 70 percent for children aged 36-59 months, and 97 percent for children aged 60-83 months. Similar patterns can be observed for the social-personal test in the Denver, and for the McCarthy memory test. On the other hand, there are no clear age patterns in outcomes for the measures of fine and gross motor skills.

In sum, there are striking age patterns in some of the measures of cognitive development we use, with older children being more likely to be delayed than younger children. To some extent this is likely to be a result of the fact that older children have been exposed for longer to poor nutrition, inadequate stimulation, infectious disease, or other risk factors that lead to delayed development. However, some of the association may be mechanical because the number of tasks very young children are expected to fulfill is left-censored, and so the number of potential delays is also smaller. For the rest of the analysis, we therefore remove age-effects, by regressing outcomes of the children in the control group on a set of single month age dummies, and predicting the residuals from these regressions.¹⁹

We also turn every outcome into a z-score by subtracting the mean and dividing through by the standard deviation of the control group, after removing the age effect. Further, we reverse the sign for those outcomes in which negative values represent better performance. As a result, positive changes indicate improvements in performance for every outcome. Finally, for every outcome we remove 0.5 percent of observations with the highest value and 0.5 percent with the lowest value, as these largely appear to be cases of measurement error. As we show below, these adjustments do not affect the pattern of program effects we estimate.

¹⁹ For the Denver, we use the number of tasks for which a child places in the lowest quartile of the international distribution; for the TVIP and McCarthy, the raw scores on the test; and for the BPI, the number of behavioral problems that a child exhibits “often”, as described above.

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 the control group. The figure shows clear socioeconomic gradients in most measures of cognitive development. Gradients appear to be steepest for the language outcomes (in particular, for the TVIP). There is no socioeconomic gradient for the BPI, our measure of behavioral problems.

4. Program impacts on early childhood development outcomes

A. Overall program effects

We estimate a basic intent-to-treat regression of the following form:

$$(1) \quad Z_k = T\alpha_k + \varepsilon_k, \quad k=1 \dots K,$$

where Z_k is the k th age-adjusted z-score for a given outcome (out of 8), and T is an intent-to-treat indicator which takes on the value of one for eligible children in communities randomly assigned to receive transfers from the *Atención a Crisis* program.²⁰ Because of the standardization described above, all units are in standard deviations. Standard errors throughout the paper adjust for clustering at the community level.

Results are presented in Table 3. The table shows there are significant program effects for the social-personal and language measures of the Denver test, corresponding to impacts of 0.134 and 0.166 standard deviations, respectively, as well as for the TVIP test of receptive vocabulary, corresponding to an impact of 0.223 standard deviations. Note that these are domains where children had particularly large delays, and where socioeconomic gradients were steep. Program effects for all other outcomes are also

²⁰ Although the *Atención a Crisis* pilot included three different treatment packages we proceed as if these were a single intervention. This seems sensible for a variety of reasons. All three groups were exposed to the same information about the importance of better and more varied diets, and all households were subject to the same requirements in terms of school enrollment of school-aged children. Moreover, at the time of the follow-up survey the observed increase in consumption was similar for all three treatment packages (Macours and Vakis 2008). This is perhaps not surprising given that the vocational training courses had not yet started, and the beneficiaries of the productive investment package had received the investment grant only 2-3 months before. We also tested whether there were differences in program effects on child cognitive development across treatment packages and can never reject the null hypothesis of equal effects.

positive, but not significant, with particularly small coefficients for the measures of gross motor skills and leg motor skills. These are precisely the domains in which children in our sample were least likely to be delayed, and where socioeconomic gradients were generally less steep. The results in Table 3 are thus consistent with the *Atención a Crisis* program having an effect on those outcomes where deficits were largest, and the association between outcomes and per capita expenditures strongest.

We show that our findings are robust to a whole set of alternative specifications in Table 4. In particular, they are robust to adding the gender of child and education of parents as controls; to inclusion of the 1 percent largest outliers (in absolute value); to not removing age effects; to including the small fraction of ineligible children whose score on the proxy means placed them above the cut-off, in both treated and control communities; to narrowing the age range for the Denver so that it is only applied to children younger than 6 (rather than younger than 7, as in the main set of results), or to children age 10 months or older (which removes from the sample children with in utero exposure). In the specification that limits the sample to children younger than 6, the point estimate for the Denver language subtest is lower. This is consistent with results below which show that the largest treatment effects on language are concentrated among older children. Note, however, that the impacts we estimate are robust to exclusion of the 6-year olds who enrolled early in primary school, suggesting that the impacts are not primarily driven by this factor.

Table 4 also shows that results for the Denver are similar if we only include the number of tasks for which the child is in the lowest decile of the international distribution, rather than in the lowest quartile, as in our main set of results; or if we consider binary variables for children having one or more tasks in the lowest decile, or two or more tasks. Finally, the last row shows results based on the sample of children for whom the enumerator actually *observed* a child performing all appropriate tasks in the

Denver.²¹ If anything, these results suggest larger treatment effects on language and fine motor skills, and the impact on fine motor skills is now borderline significant.

B. Heterogeneity of effects by child gender and age

We next test for heterogeneity of treatment effects by age and gender. To do this, we run variants of (1) which include interactions with treatment, as is standard:

$$(2) \quad Z_k = T\alpha_k + X\beta_k + (T * X)\delta_k + \varepsilon_k$$

where X is a child characteristic (for example, a dummy for girls). We then test for the significance of the differences in the δ_k coefficients.

There are a number of theoretical and empirical reasons why analyzing heterogeneity along these two dimensions is of particular interest. First, focusing on differences by gender, recall that Table 2 showed no significant differences between boys and girls in the fraction of children delayed. Further, socioeconomic gradients in cognitive outcomes do not differ by gender. We might therefore expect to see no differences in program effects for boys and girls. However, girls may benefit more from interventions in early childhood than boys (as suggested by Anderson 2007). Also, the *Atención a Crisis* transfers were given to women, and some research on health outcomes suggests that resources in the hands of women tend to benefit girls more than boys (see Duflo 2003 on South Africa, Thomas 1994 on results for Brazil, Ghana and the US).

Second, focusing on differences by age, recall that Table 2 showed that older children were more likely to be delayed in language, memory and personal-social skills and that the depth of these delays was larger for older children. Also, differences between children in poorer and less poor families become larger with the age of the child for some outcomes, notably TVIP scores. Both of these suggest that we might expect to see larger program effects among older children. On the other hand, there may be

²¹ This implies excluding 30 percent of scores for the language subtest, 6 percent for the fine motor subtest and 22 percent for the gross motor subtest. Note that observation-only is not possible for the social-personal Denver subtest, as most items in this task rely on caregiver responses.

“critical periods” in brain development at very young ages, and the potential for later catch-up may be limited. This is one reason why we might see larger effects of the program among the youngest children.

Results on program effects by gender and age are presented in Table 5. The upper panel of the table shows that there is no evidence that program impacts differ by gender—in some cases, the point estimates are larger for boys than for girls, and in other cases the opposite is true. In no case is the difference in program effects by gender significant at conventional levels.

The lower panel presents differences by child age, with children divided into three groups, corresponding to those age 0-35, 36-59, and 60-83 months, respectively. This analysis suggests there are important differences in impacts by child age—in particular, program effects for language outcomes in both the Denver and the TVIP are largest among the oldest children. For the language subscale of the Denver, the effect of receiving *Atención a Crisis* transfers rises from 0.056 (standard error: 0.060) for the youngest children, to 0.174 (standard error: 0.110) for children in the middle age group, to 0.300 (standard error: 0.088) for the oldest children; the difference in program effects between the youngest and oldest children is significant (p-value: 0.011). For the TVIP, the program effect is 0.054 (standard error: 0.052) for children between the ages of 36 and 59 months, and 0.362 (standard error: 0.118) for children age 60 months and older; this difference in program effects is once again highly significant (p-value: 0.005).

Turning to other outcomes, the point estimate on the treatment dummy for the social-personal Denver is largest for the middle age group, corresponding to children age 36-59 months. For this group, there is also a borderline significant effect on the BPI. As with language, this suggests some consistency within domains, as both the social-personal Denver and the BPI are likely to capture aspects of behavior, even if the BPI does not measure delays per se. However, the differences across age groups for the social-personal Denver test are not significant, and for the BPI it is only borderline significant. There is no apparent pattern of age effects for the McCarthy memory scale and for the measures of gross and fine

motor skills, and in no case are the estimated coefficients on the measure of exposure to the *Atención a Crisis* program significant.²²

The larger program effects on language outcomes among older children appear to be at odds with some other studies in the literature (see the review in Engle et al. 2007), and it is therefore important to consider this finding carefully. In principle, it is possible that the tests we use are more appropriate for older children, and that this makes it easier to identify program effects for this group. However, we do not think that this is the main reason for the results we estimate. The test-retest validity for the Denver language test is similar for older and younger children, and is very high for all three age groups—suggesting that the test consistently measures language skills and knowledge for both younger and older children.²³ Also, our results do not appear to be driven by differences in random measurement error across age groups. Table 5 shows that estimates of the *Atención a Crisis* program effects for younger children tend to have small coefficients, rather than large standard errors. All of this suggests that the fact that we find program effects on language among older kids, but not among younger kids, is not primarily a measurement problem. Rather, the pattern of results in Table 5 is consistent with the *Atención a Crisis* program having bigger effects among age groups where the margin for improvement was large and where language deficits are most correlated with overall expenditure levels, as described in Section 3 above.

That said, it is possible that we find larger program effects among older children because it takes longer for the benefits of the *Atención a Crisis* to become apparent among younger children. This is plausible given the very short amount of time children in our sample were exposed to the program—9 months. We expect to be able to provide a better answer to this question by returning to our sample in the

²² We also tested for heterogeneity by baseline per capita consumption and parental education levels. There is no consistent pattern of differences in program effects by these two measures of household socioeconomic status. This is quite different from the results in Paxson and Schady (2008) for Ecuador—the program effects they find are concentrated among children in the poorest quartile. However, as we note above, the sample of children in our study appears to be substantially worse off than the sample of children in Ecuador. Indeed, in terms of baseline expenditures, the mean control household in the Nicaragua sample has expenditures of US \$ 267 per person per year, almost identical to the value of US \$ 263 per person per year for the mean household in the *poorest quartile* in the sample for Ecuador.

²³ For the Denver language items, information provided in the test manual indicates that the average test-retest validity of the items tested is 92 percent for each of the three age groups we consider.

near future. In any event, our results make clear that there is potential for substantial catch-up in some domains of cognitive development among older preschool-aged children.

5. Can program effects be explained by increased income alone?

The results in section 4 make clear that the *Atención a Crisis* program had positive effects on a number of dimensions of child cognitive development. In this section we consider program effects on various “risk factors” that have been identified as important determinants of cognitive development in the literature (see the review by Walker et al. 2007). The first of these is inadequate nutrition, including poor maternal nutrition before birth. A large body of evidence suggests that poorly nourished children are at increased risk of poor cognitive development outcomes, and food supplementation programs have been shown to have positive effects on child development in a variety of settings, including Guatemala, Indonesia, and Jamaica, although positive effects in Jamaica were no longer apparent in the long run (Walker et al. 2005). We therefore assess whether families who benefited from the *Atención a Crisis* program spent more on food, and on different kinds of food.

The second risk factor we consider is inadequate stimulation—it is estimated that only between 10 percent and 41 percent of parents in developing countries provide cognitively stimulating materials to their children, and the fraction of parents involved in cognitively stimulating activities for their children is similarly low (Walker et al. 2007). Moreover, numerous studies have found that interventions that increase stimulation in early childhood result in improvements in child cognitive development. We therefore assess whether children in households randomly assigned to receive transfers received more stimulation.

Finally, lack of micro-nutrients, exposure to infectious disease and caregivers’ mental health are all thought to be important risk factors for child cognitive development (Walker et al. 2007; see also Sohr-Preston and Scaramella 2006, who review the relationship between maternal mental health and child cognitive and emotional outcomes in the US, especially among low-income children). We therefore consider whether the *Atención a Crisis* program had positive effects on the use of preventive health care

by children, and on measures of maternal depression, stress, and the parenting environment in children's homes.

Our approach is as follows: We first document program effects on these risk factors, and then analyze whether any observed program effects can be accounted for purely by the increase in overall expenditures among households that received *Atención a Crisis* transfers. Specifically, we test whether transfer income was spent in a similar way as other income, focusing on expenditures that could be relevant for the cognitive development of young children.

We begin by documenting how the program affected the levels of various inputs into child development in Table 6. To put the magnitude of effects into perspective, the table also includes the mean value of each variable for the control group. Table 6 shows that overall food expenditures increased among treated households, and expenditures on nutrient-rich food such as animal proteins, fruit and vegetables increased more than proportionally. Treatment effects on indicators of food intake of individual children under the age of 7 show a similar pattern.

Table 6 further shows that treated households have higher values for various indicators of early stimulation, such as the availability of books, and paper and pencil, as well as different behavior of caregivers—in treated households, children are more likely to be read to and to have stories told to them. There also a marginally significant impact on enrollment in a stimulation program or (pre)school.²⁴ Table 6 also shows program effects on a number of measures of preventive health care—treated children are more likely to have been weighed, and to have received vitamins, iron, and deworming drugs.²⁵ There is no program effect on child height or weight, although this may in part be a result of the fact that treated children had somewhat poorer outcomes than control children at baseline (differences were not significant); when we control for baseline characteristics, we find significant program effects on height

²⁴ While the schooling requirement was not binding for the children below 7 considered in this paper, this could point to a potential spillover effect.

²⁵ These self-reported measures of preventive health care were double-checked using a vaccination-health use card filled in by health care providers when children visit the health center. These cards were available for 87 percent of the children in the sample. Results are robust when only this 87 percent subsample is used, suggesting that results are not driven by reporting bias.

and weight among older children. We also consider the birth weight of recently born children (under 5 months) who would have benefited from the program for the majority of their time in utero. The point estimates are positive but not significant, likely due to the very limited sample size. Finally, in treated households there are improvements in the mental health of caregivers, as measured by the CESD depression scale, although these are only borderline significant. (Lower values correspond to better outcomes for the CESD.)²⁶

The findings in Table 6 are perhaps not surprising, given the magnitude of the transfer. We therefore turn to the question of whether the observed improvements in these risk factors are larger than what one would expect to see given the increase in overall expenditures that resulted from the program. Specifically, we use nonparametric regressions to compare expenditure and behavior patterns among households with similar overall expenditure levels in treated and control communities.

We first focus on overall food expenditures, and graph food Engel curves at follow-up for treated and control households in Figure 2. The curves show the familiar downward slope—better-off households tend to have lower food shares in expenditures, as predicted by Engel’s Law.²⁷ If transfers were spent like other sources of income, the *Atención a Crisis* program would move households along the Engel curve, and the food share would fall (even if the absolute amount spent on food increased). However, the figure shows that the food Engel curve for treated households is everywhere above that of control households. The lower panel of the figure graphs the difference between the two curves, and a 95 percent confidence interval based on bootstrapped standard errors (as in Deaton 1997; Kremer et al.

²⁶ When we break down these program effects on food, early stimulation, and preventive health care by age, larger impacts are generally found among older children—sometimes significantly so, as shown in Appendix Table A4. This is consistent with the fact that we observe larger program effects on cognitive development among older children, as shown in Table 5. Among other things, Appendix Table A4 shows an increase in enrollment in primary school among the older children. Yet, as discussed, the impacts on cognitive developments we estimate are robust to exclusion of those children.

²⁷ The increasing slope at the lowest level of expenditure is similar to results in other settings (Thomas 1986).

2004).²⁸ This panel shows that the difference between the two curves is significant for the bulk of the distribution of per capita expenditures.

We next turn to the composition of food expenditures in Figure 3. This figure shows that, at similar overall expenditure levels, treated households spent significantly less on staples (primarily beans, tortillas, and rice), and significantly more on animal proteins (chicken, meat, milk, eggs), as well as on fruits and vegetables. The catch-all “other food” category, which includes a variety of expenditures (mainly sweets, prepared food, vegetable oil and coffee), shows less of a clear pattern. However, among the better-off households in the sample, who had smaller decreases of expenditure shares on staples, the decrease in “other” food expenditures appears to be larger. Figure 3 thus shows that households diversified their diets and shifted towards higher quality sources of calories—see Hoddinott and Skoufias (2004) for similar evidence from the PROGRESA program in Mexico. Increased expenditures on animal proteins, fruits and vegetables may also have allowed households to acquire micronutrients, such as iron. This is important as a recent review (Walker et al. 2007) estimates that between 44 percent and 66 percent of children in developing countries suffer from anemia, and that iron deficiency is one of the main proximate causes for low levels of cognitive development among children in poor countries (see also Bobonis et al. 2007 for evidence from India).

Food expenditures are measured at the household rather than the child level. It is therefore conceivable, although unlikely, that these changes in expenditure patterns would not affect the children in our sample. As a robustness check, we use the data we collected on the frequency of intake of specific food items (for example, tortilla, fruit juice, milk, meat, vegetables) of each child in the last 7 days. This analysis, unreported but available from the authors upon request, shows a very similar pattern as that in Figure 3, confirming that the shifts in food expenditure patterns among treated households affected the

²⁸ Fan regressions were estimated with bandwidth of 0.99 and results are not sensitive to the choice of bandwidth. The significance of the difference was established using bootstrapped standard errors, clustered by community. The 1 percent highest and lowest values for log per capita expenditures were trimmed to avoid empty cells when bootstrapping due the low density of observations in the tails of the distribution.

children in our sample directly. Hence, the program seems to have resulted in a shift towards more diversified diets and more nutrient-rich food for young children.

We next turn to other broad expenditure categories in Figure 4. The figure shows that the share of expenditures on schooling increased significantly for all but the richest households, while the share of expenditures on health, housing, and other non-food items generally decreased. These decreases are significant for a large range of consumption levels. Decreased expenditure shares on health might indicate an improvement of the overall health status of the beneficiary households.

Figure 5 presents a comparable analysis for child stimulation. In particular, we consider whether there are changes in the relationship between household expenditures and various inputs, such as drawing and reading material, story-telling and reading by parents to their children. Figure 6 shows that at similar expenditure levels, children in treated households are more likely to have access to pen, paper and books, and parents spent more time reading to them. With the exception of story-telling, the upward shift in the curves is significant for a large range of expenditure levels. Similar results are apparent for other stimulation inputs (unreported but available from the authors upon request).

A less clear pattern emerges for indicators of preventive health care and caregivers' mental health. While the upper panel of Figure 6 suggests some upward shifts of the curves, and a downward shift for depression, differences between treated and control groups are smaller and less often significant. In fact, the slopes of the curves that relate preventive health care indicators to per capita expenditures are small, and are often not significantly different from zero. In this population, overall expenditures are a poor predictor of use of preventive health care, and the program does not appear to have changed this in an appreciable way.²⁹

In sum, the *Atención a Crisis* program changed household expenditure patterns in important ways. At follow-up, treated households spend a higher proportion of their expenditures on food, their

²⁹ In Nicaragua, as in many other developing countries, preventive health care visits are nominally free—which could explain the relative flat socioeconomic gradients; it also helps explain why treated households decreased the share of expenditures on health without appreciable changes in the use of preventive health care at a given overall consumption level.

children have more diversified and nutrient-rich diets, more access to material that can stimulate their cognitive development, and better preventive health care indicators. Caregivers in treated households also seem to allocate more time to reading to their children, and have better mental health outcomes. The changes in food expenditure patterns and stimulation indicators are larger than what would have been expected from the increase in overall expenditures that resulted from the program, which suggests that there were behavioral changes. These changes could be a result of the social marketing of the program, the fact that transfers were made to women, or possibly other program design features. We cannot disentangle where these changes originate from nor can we directly identify the impacts of these intermediate inputs on cognitive development. Nevertheless, the observed patterns are consistent with the hypothesis that changes in the allocation of resources by households help explain how cognitive development outcomes had improved significantly among treated children after only 9 months.

A final indication that the program's impact on early childhood outcomes is not only a result of changes in overall expenditures is suggested by the results on the reduction of behavioral problems among the 3-4 year olds. The BPI is not related to overall expenditures in our data. If anything, the slope in Figure 2 appears to be slightly negative. The borderline significant program effect on the BPI, in spite of the fact that there are no socioeconomic gradients for this outcome, suggests that the impacts we observe are likely to be related to changes in the attitudes and behaviors of caregivers.

7. Conclusion

This paper documents substantial delays in cognitive development among young children in rural areas in Nicaragua, a low-income country. These delays are likely to have serious implications for the success of these children as adults. Indeed, a variety of theories of skill formation suggest that investments in schooling and other dimensions of human capital will have only low productivity if children do not have adequate levels of cognitive and social skills at early ages (for example, Cunha et al. 2005).

The paper then estimates the impact of a cash transfer program on a large set of measures of cognitive development. We show that the program had a substantial positive impact on both personal-social and language development after only 9 months. Program effects on language outcomes are largest among older children, suggesting that there is substantial potential for catch-up in this domain.

Furthermore, the positive impacts we find suggest that gains in early childhood development outcomes, which have not been widely studied in the economic literature on developing countries, should be taken into account when assessing the benefits of cash transfer programs.

We show that households who benefited from transfers increased expenditures on some critical risk factors for cognitive development in early childhood. Specifically, households spend more on nutrient-rich foods. They also appear to have provided more early stimulation to their children, and to have made more use of preventive health care. Changes in the pattern of food expenditures and in stimulation inputs are larger than what one would expect to see if the program were simply moving children along the curves that relate these inputs to overall expenditures—clear evidence that the program affected how households allocate their budget. Thus, in Nicaragua, a dollar is not always a dollar (or, rather, a *Córdoba* is not always a *Córdoba*). More research is needed to understand what features of program design, including the gender of the recipients or the social marketing of the program, resulted in substantial improvements in cognitive development in a very short period of time.

More generally, the evidence in this paper illustrates that gains in early childhood development can result from interventions that focus on investments made by parents to reduce risk factors for cognitive development. Further research is needed to understand how these policies or programs can best complement more traditional attempts to improve the access to or the quality of supply side interventions.

References

- Aguero, J., M. Carter, and I. Woolard. 2006. "The Impact of Unconditional Cash Transfers on Nutrition: The South African Child Support Grant." Unpublished manuscript, University of California at Riverside.
- 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. 2007. "Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects.", forthcoming, *The Journal of the American Statistical Association*.
- Attanasio, O., and M. Vera-Hernández. 2004. "Medium- and Long-Run Effects of Nutrition and Child Care: Evaluation of a Community Nursery Program in Rural Colombia." Working Paper EWP04/06, Institute for Fiscal Studies, University College, London.
- Attanasio, O., E. Battistin, E. Fitzsimons, A. Mesnard, and M. Vera-Hernández. 2005. "How Effective are Conditional Cash Transfers? Evidence from Colombia." Unpublished manuscript, The Institute for Fiscal Studies, London, UK.
- 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., and J. Hoddinott. 2005. "Programme Evaluation with Unobserved Heterogeneity and Selective Implementation: The Mexican PROGRESA Impact on Child Nutrition." *Oxford Bulletin of Economics and Statistics* 67(4): 547-69.
- Behrman, J., P. Sengupta, and P. Todd. 2005. "Progressing through Progres: An Impact Assessment of a School Subsidy in Mexico." *Economic Development and Cultural Change* 54(1) 237-75.
- Behrman, J., Y. Cheng, and P. Todd. 2004. "Evaluating Preschool Programs When Length of Exposure to the Program Varies: A Nonparametric Approach." *Review of Economics and Statistics* 86(1): 108-32.
- Berlinski, S., S. Galiani, and P. Gertler. 2006. "The Effect of Pre-Primary Education on Primary School Performance." Working Paper W06/04, Institute for Fiscal Studies, University College, London.
- Berlinski, S., S. Galiani, and M. Manacorda. 2007. "Giving Children a Better Start in Life: Preschool Attendance and School-Age Profiles." Forthcoming, *Journal of Public Economics*.
- 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.
- Bobonis, G. J., E. Miguel, and C. Puri-Sharama. 2007. "Anemia and School Participation." *Journal of Human Resources* 41(4): 692-721.
- 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: 437-490.
- Case, A., and C. Paxson. 2006. "Stature and Status: Height, Ability, and Labor Market Outcomes." NBER Working Paper No.12466. Cambridge, MA.
- 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: 3162-65.

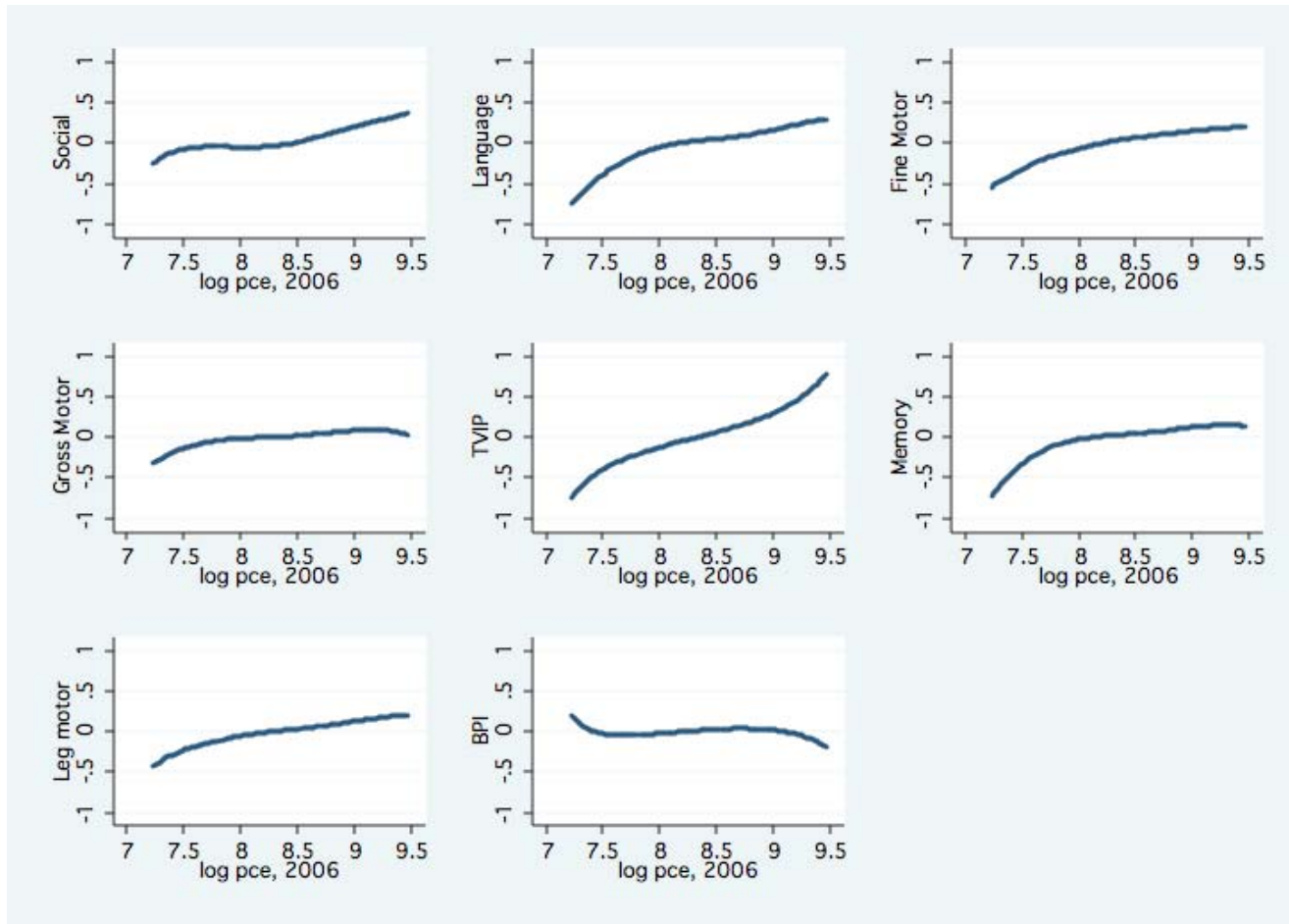
- Cogill, S.R., Caplan, H.L., Alexandra, H., Robson, K.M., and Kumar, R. 1986 "Impact of Maternal Postnatal Depression on Cognitive Development of Young Children." *British Medical Journal* 292(6529): 1165-67.
- Cunha, F., J. Heckman, L. Lochner, D. Masterov. 2005. "Interpreting the Evidence on Life Cycle Skill Formation." 2005. *NBER Working Paper* 11331, Cambridge, MA.
- Currie, J. 2001. "Early Childhood Education Programs." *Journal of Economic Perspectives* 15(2): 213-38.
- Currie, J., and D. Thomas. 1999. "Early Test Scores, Socioeconomic Status and Future Outcomes." NBER Working Paper No. 6943. Cambridge, MA.
- Deaton, A., 1997. *The Analysis of Household Surveys: A Microeconometric Approach to Development Policy*. Johns Hopkins University Press, Baltimore.
- 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: 262-67.
- Duflo, E. 2003. "Grandmothers and Granddaughters: Old-Age Pensions and Intrahousehold Allocation in South Africa." *World Bank Economic Review* 17(1): 1-25.
- Edmonds, E.V. 2006. "Child Labor and Schooling Responses to Anticipated Income in South Africa." *Journal of Development Economics* 81:386-414
- Engle, P., M. Black, J. Behrman, M. Cabral de Mello, P. Gertler, L. Kapiriri, R. Martorell, M. E. Young, and the International Child Development Steering Group. 2007. "Strategies to Avoid the Loss of Developmental Potential in More than 200 Million Children in the Developing World." *Lancet* 369: 229-42.
- 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: 828-837.
- 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.
- Gertler, P. 2004. "Do Conditional Cash Transfers Improve Child Health? Evidence from PROGRESA's Controlled Randomised Experiment." *American Economic Review* 94(2): 331-336.
- Gertler, P. and L. Fernald. 2004. "The Medium Term Impact of *Oportunidades* on Child Development in Rural Areas." Unpublished manuscript, University of California at Berkeley.
- Glewwe, P., and P. Olinto. 2004. "Evaluating the Impact of Conditional Cash Transfers on Schooling: An Experimental Analysis of Honduras. PRAF Program." Not Published.
- 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: 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." *Lancet* 369: 60-70.
- Grantham-McGregor, S.M., and C. Ani. 2001. "A Review of Studies on the Effect of Iron Deficiency on Cognitive Development in Children." *Journal of Nutrition* 131: 649S-668S.

- 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." *Lancet* 338: 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.
- Hoddinott, J., and E. Skoufias, 2004, "The Impact of PROGRESA on Food Consumption," *Economic Development and Cultural Change* 53(1): 37-61.
- 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.
- Kremer, Michael, Edward Miguel and Rebecca Thornton. 2004. "Incentives to Learn", NBER Working Paper No. 10971. Cambridge, MA
- Lozoff, B., E. Jimenez, J. Hagen, et al. 2000. "Poorer Behavioral and Development Outcome More Than 10 Years after Treatment for Iron Deficiency in Infancy." *Pediatrics* 105(4): 1-11.
- 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, 2007. "Seasonal Migration and Early Childhood Development", *Social Protection Discussion Paper* 0702, World Bank.
- Macours, K., and R. Vakis, 2008. "Changing households' investments and aspirations through social interactions: Evidence from a randomised transfer program in a low-income country", mimeo, Johns Hopkins University and World Bank.
- Macours, K., and R. Vakis, 2005. "Weather risk and poverty in Nicaragua: Expanding risk management options for the poor. Pilot program objectives and impact evaluation design, mimeo, Johns Hopkins University and World Bank.
- Maluccio, J., and R. Flores. 2004. "Impact Evaluation of a Conditional Cash Transfer Program: The Nicaraguan Red de Protección Social." Unpublished manuscript, Food and Nutrition Division, International Food Policy Research Institute, Washington, D.C.
- Morris, S., R. Flores, P. Olinto, and J.M. Medina. 2004. "Monetary Incentives in Primary Health Care and Effects on Use and Coverage of Preventative Health Care Interventions in Rural Honduras: Cluster Randomized Trial." *Lancet* 364: 2030-2037.
- Morris, S., P. Olinto, R. Flores, E. Nilson, and A. Figueiró. 2004b. "Conditional Cash Transfers Are Associated with a Small Reduction in the Rate of Weight Gain of Preschool Children in Northeast Brazil." *Journal of Nutrition* 134(9): 2336-42.
- 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. 2008. "Does Money Matter? The Effects of Cash Transfers on Child Health and Development in Rural Ecuador." Unpublished manuscript, Princeton University and The World Bank.

- 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: 385-401.
- Rivera, J. A., D. Sotres-Alvarez, J. P. Habicht, T. Shamah and S. Villalpando. 2004. "Impact of the Mexican Program for Education, Health, and Nutrition (Progresa) on Rates of Growth and Anemia in Infants and Young Children: a Randomised Effectiveness study." *Journal of the American Medical Association* 291(21): 2563-70.
- 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., and C. Araujo. 2008. "Cash Transfers, Conditions, and School Enrollment in Ecuador." Forthcoming, *Economía* 8(1).
- Schady, N., and J. Rosero. 2007. "Do Cash Transfers to Women Affect the Composition of Expenditures? Evidence on Food Engel Curves in Rural Ecuador." *World Bank Policy Research Working Paper* 4282.
- Schultz, T. 2004. "School Subsidies for the Poor: Evaluating the Mexican Progresa Poverty Program." *Journal of Development Economics* 74: 199-250.
- Schweinhart, L.J. 2005. "The High/Scope Perry Preschool Study through Age 40: Summary, Conclusions, and Frequently Asked Questions." Ypsilanti, Mich.: High/Scope Press.
- Sohr-Preston, S. L. and L. V. Scaramella. 2006. "Implications of Timing of Maternal Depressive Symptoms for Early Cognitive and Language Development." *Clinical Child and Family Psychology Review* 9(1): 65-83.
- 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: 1-8.
- Thomas, D. 1986. "The Food Share as a Welfare measure." Ph.D. dissertation, Department of Economics, Princeton University.
- 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: 1012-20.
- Walker, S.P., and others. 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: 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." *Lancet* 369(9556): 145-57.

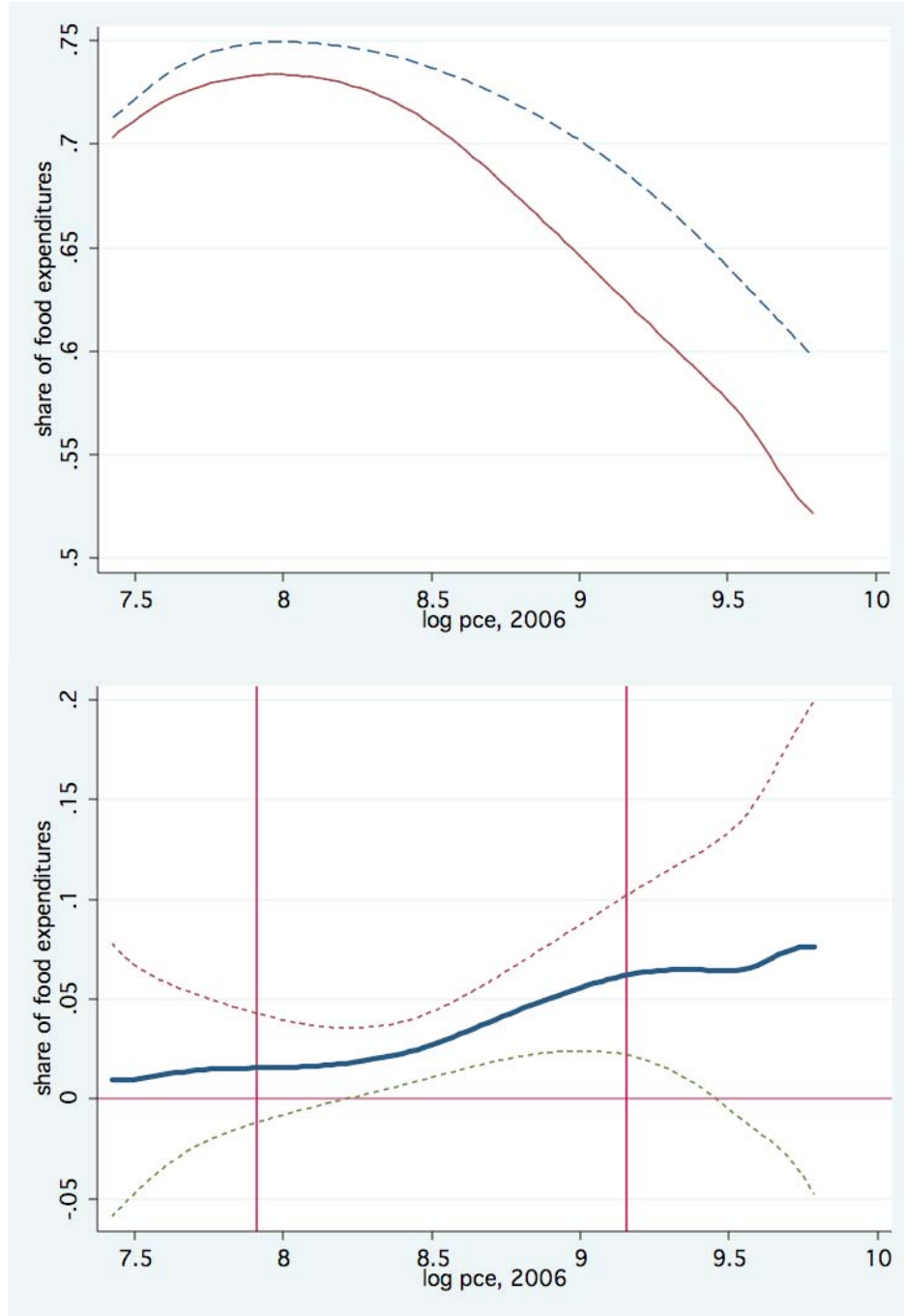
The World Bank. 2008. *Conditional Cash Transfers for Attacking Present and Future Poverty*.
Forthcoming, World Bank, Washington D.C.

Figure 1: Socioeconomic gradients in cognitive development



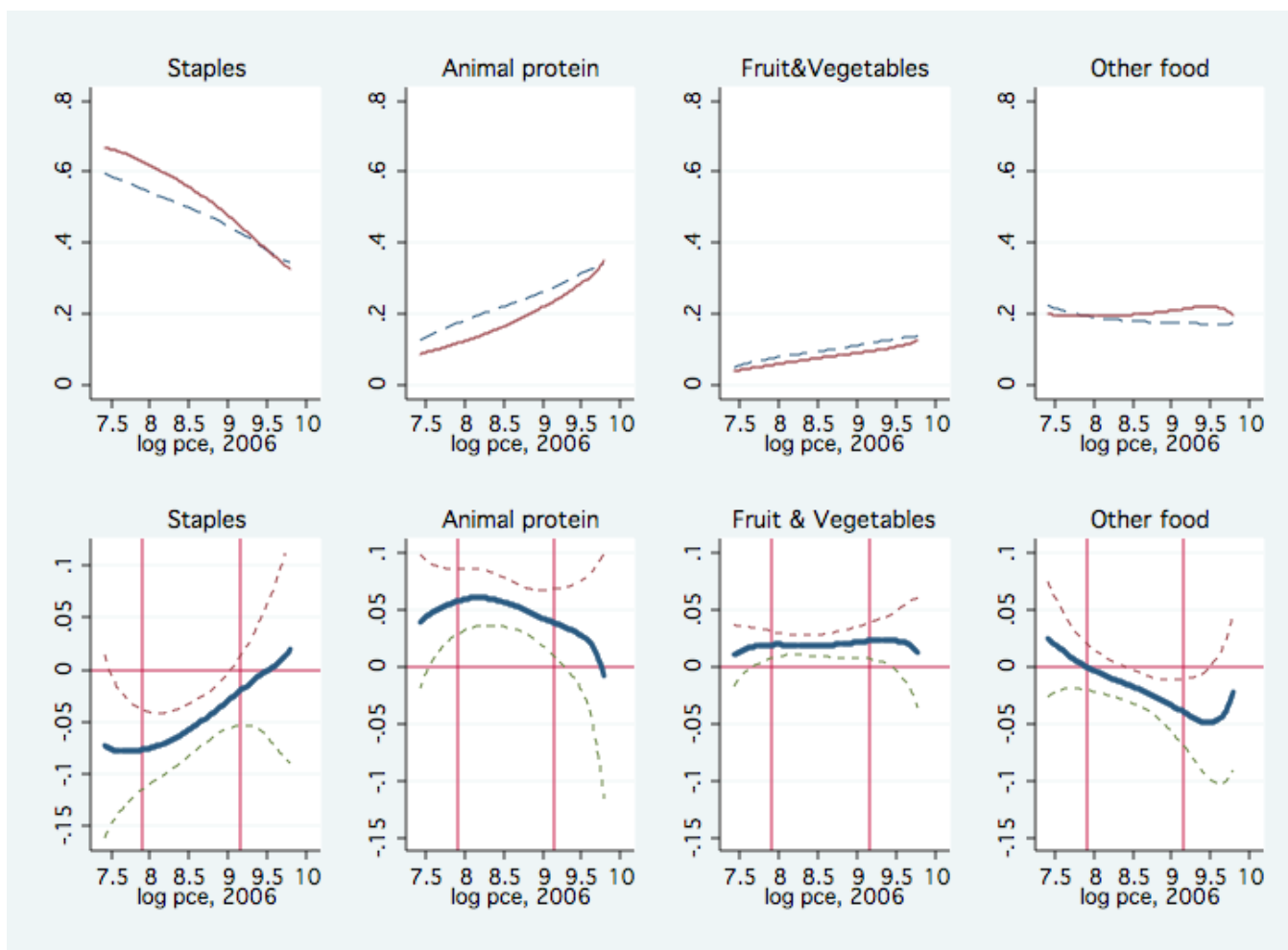
Note: Outcomes for the control group only. Age trends have been removed from all outcomes. All outcomes are then standardized by subtracting the mean and dividing by the standard deviation of the control group. For the Denver (personal, language, fine motor, gross motor), the sample includes children between the ages of 0 and 83 months; for the TVIP (receptive language), McCarthy (memory and leg motor), and BPI the sample includes children between the ages of 36 and 83 months. For the Denver test, the dependent variables are defined in terms of the number of delays plus cautions. Fan regressions with bandwidth of .99. Outliers trimmed at 1% of highest and lowest values for Fan. 2.5% highest and lowest outliers of log(pce) trimmed from graph.

Figure 2: Food Engel curves



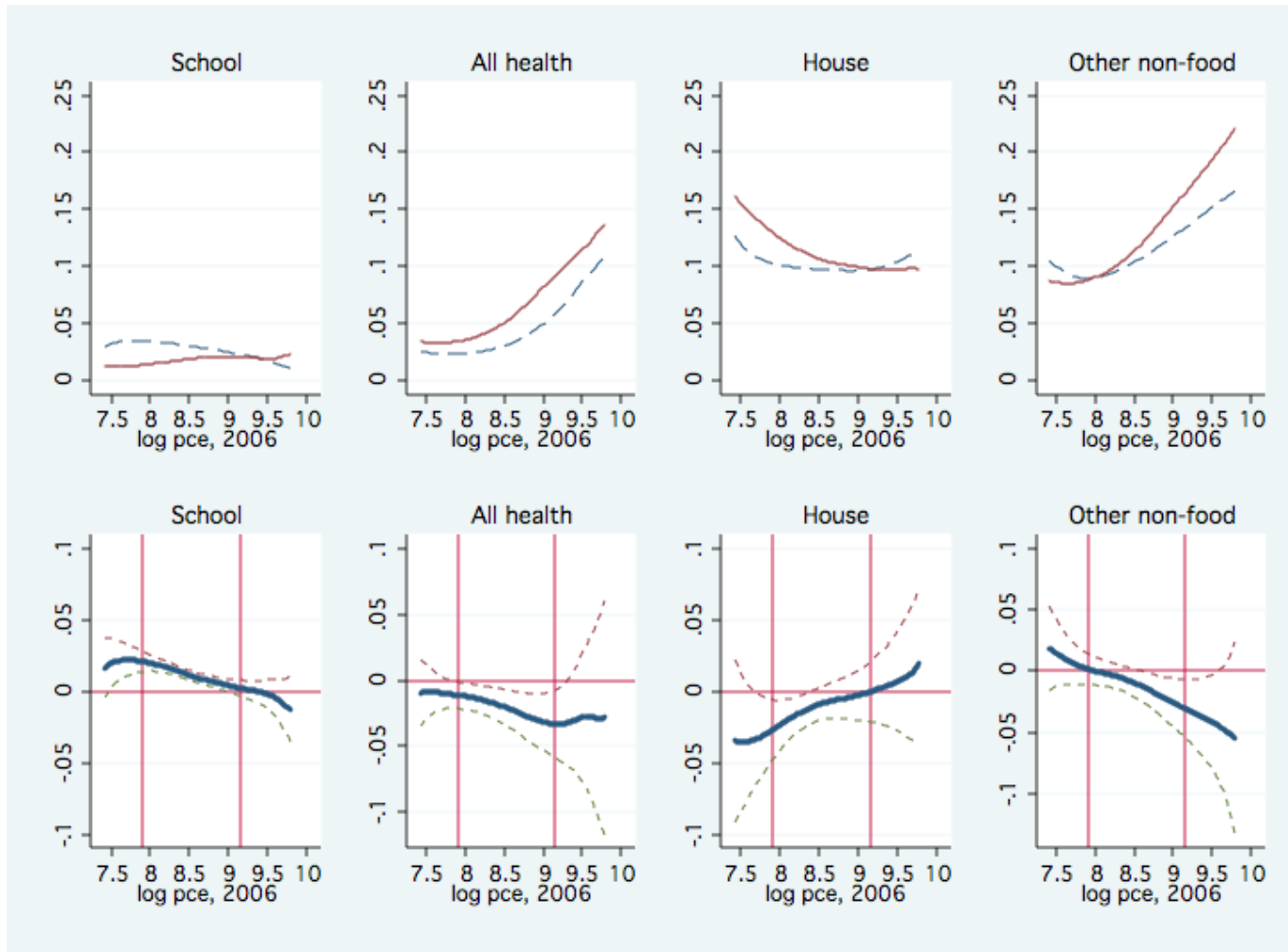
Note: In upper panel the dashed line correspond to treatment, the solid line to control. In lower panel, the bold line corresponds to the difference between treatment and control, and the bounds of the 95% confidence interval are indicated in short dashes. Vertical lines are included at 10th and 90th percentiles of log per capita expenditures. Fan regressions with bandwidth of .99. Outliers trimmed at 1% highest and lowest values for Fan.

Figure 3: Composition of food expenditures: Share of different types of food expenditure in total food expenditure



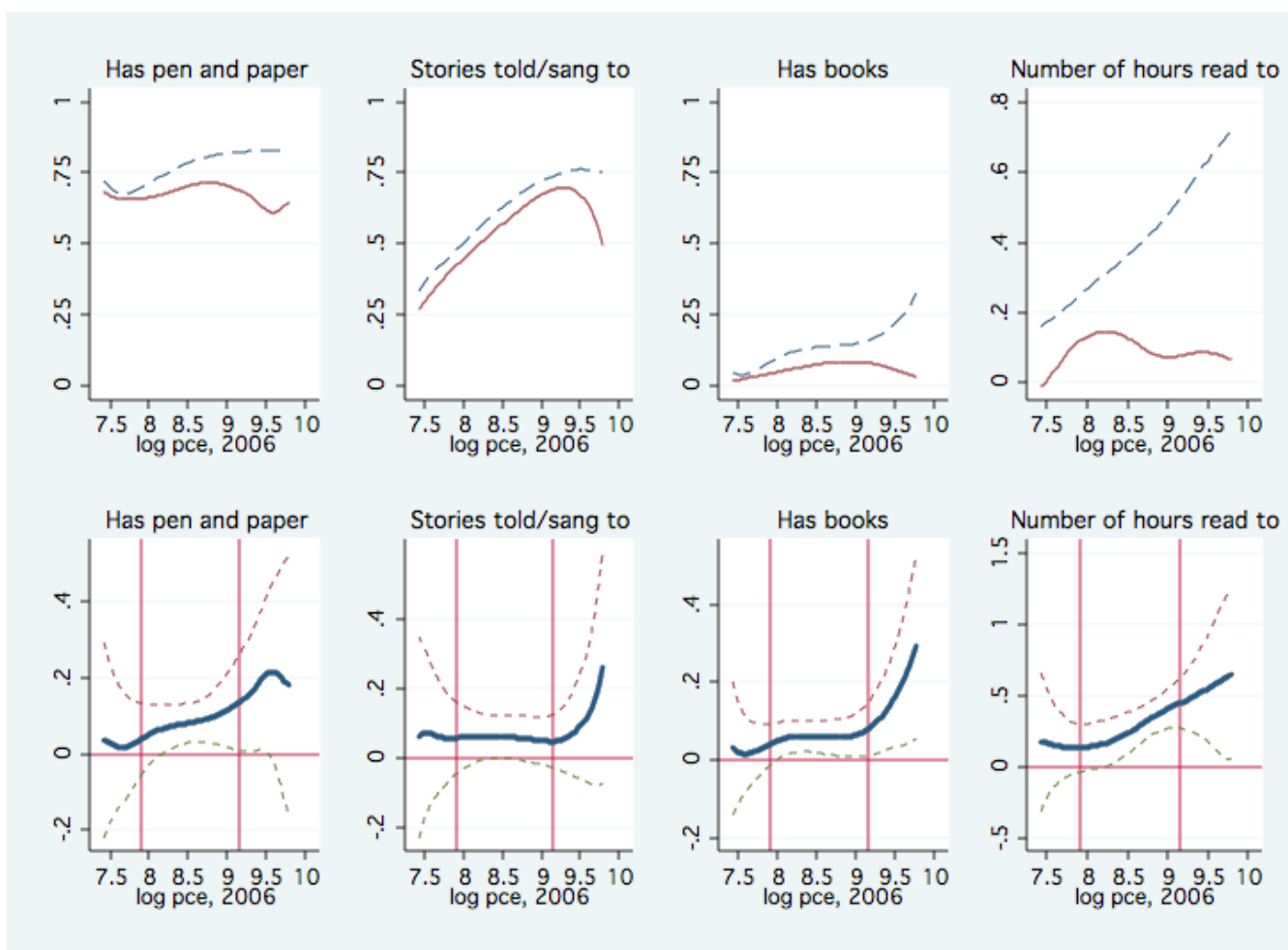
Note: In upper panel dashed lines correspond to treatment, solid lines to control. In lower panel, the bold lines corresponds to the difference between treatment and control, and the bounds of the 95% confidence intervals are indicated in short dashes, reflecting bootstrapped standard errors, clustered by community. Vertical lines are included at 10th and 90th percentiles of log per capita expenditures. Fan regressions with bandwidth of 0.99. Outliers trimmed at 1% highest and lowest values for Fan.

Figure 4: Share of expenditures of different non-food items



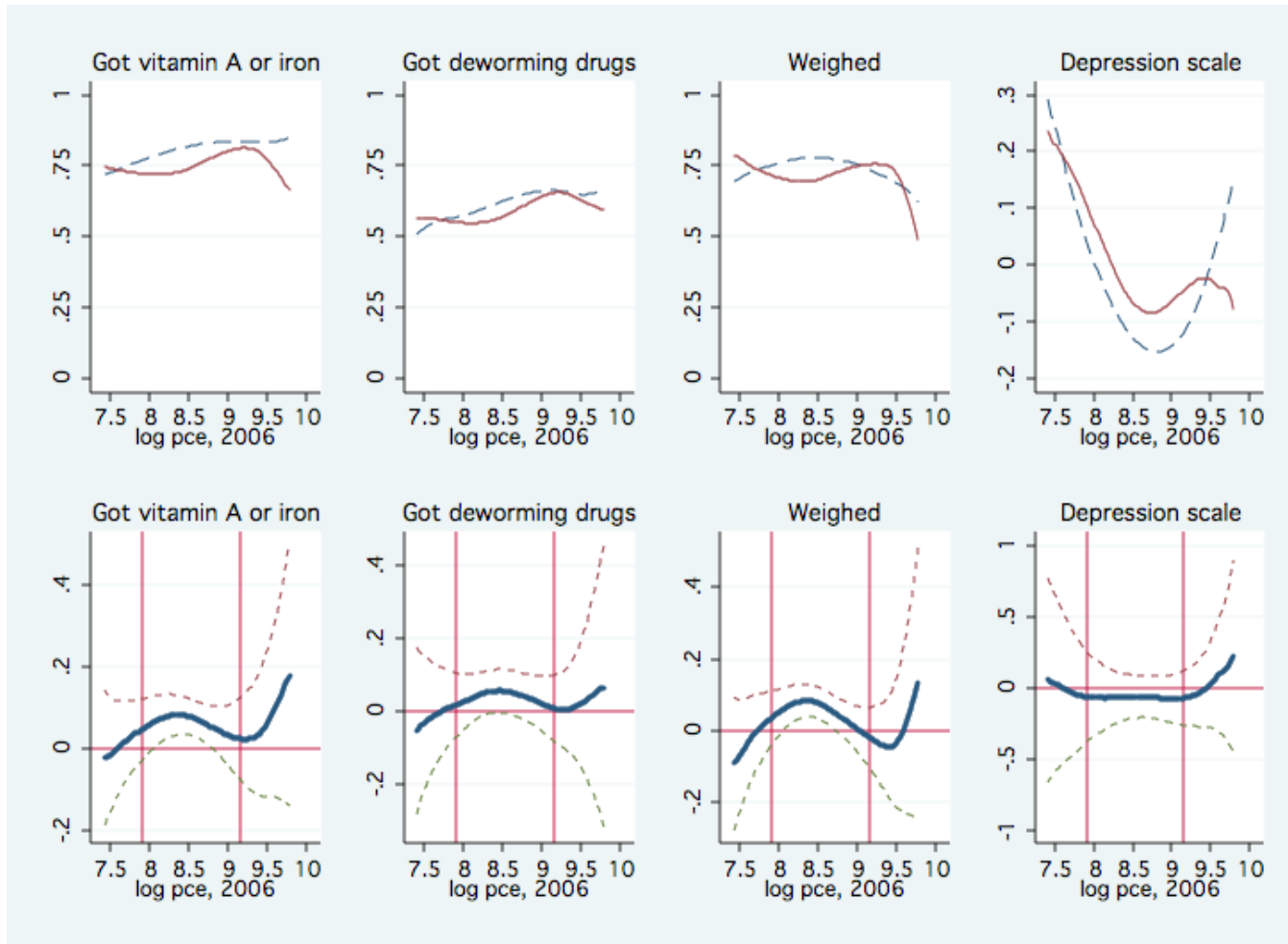
Note: In upper panel dashed lines correspond to treatment, solid lines to control. In lower panel, the bold lines correspond to the difference between treatment and control, and the bounds of the 95% confidence intervals are indicated in short dashes, reflecting bootstrapped standard errors, clustered by community. Vertical lines are included at 10th and 90th percentiles of log per capita expenditures. Fan regressions with bandwidth of 0.99. Outliers trimmed at 1% highest and lowest values for Fan.

Figure 5: Relationship between consumption levels and inputs for early childhood stimulation



Note: In upper panel dashed lines correspond to treatment, solid lines to control. In lower panel, the bold lines correspond to the difference between treatment and control, and the bounds of the 95% confidence intervals are indicated in short dashes, reflecting bootstrapped standard errors, clustered by community. Vertical lines are included at 10th and 90th percentiles of log per capita expenditures. Fan regressions with bandwidth of 0.99. Outliers trimmed at 1% highest and lowest values for Fan.

Figure 6: Relationship between consumption levels, preventive health care, and caregiver mental health



Note: In upper panel dashed lines correspond to treatment, solid lines to control. In lower panel, the bold lines corresponds to the difference between treatment and control, and the bounds of the 95% confidence intervals are indicated in short dashes, reflecting bootstrapped standard errors, clustered by community. Vertical lines are included at 10th and 90th percentiles of log per capita expenditures. Fan regressions with bandwidth of 0.99. Outliers trimmed at 1% highest and lowest values for Fan.

Table 1: Program info and baseline characteristics of control and treatment children

	Nr. of obs	Mean control	Mean treatment	P-value difference
Program info				
Take-up rate (contamination in control)		0.1	95.3	
Total cash transfer as a share of overall consumption		-	15	
Child-specific characteristics				
All Children				
Male	3505	0.52	0.50	0.388
Age in months	3506	44.02	43.56	0.587
0-35 months old	3506	0.40	0.40	0.983
36-59 months old	3506	0.27	0.28	0.603
60-83 months old	3506	0.33	0.32	0.643
Years education mother	3481	4.01	4.07	0.825
Years education father	3218	3.48	3.57	0.681
Children 0-5 at baseline				
Weight-for-age z-score	2365	-0.91	-1.06	0.168
Height-for-age z-score	2355	-1.10	-1.28	0.122
Weight-for-height z-score	2369	-0.19	-0.17	0.764
Consulted doctor if sick or diarea	1599	0.72	0.76	0.290
Weighed in last 6 months	2488	0.93	0.90	0.154
Received vitamins in last 6 months	2488	0.74	0.68	0.117
Received deworming drugs in last 6 months	2488	0.58	0.51	0.076
Household characteristics				
Male household head	2270	0.84	0.86	0.459
Household size	2270	6.14	6.05	0.621
Number of children under 5	2270	1.14	1.13	0.867
Number of children between 5 and 14	2270	1.72	1.73	0.927
Number of persons between 15 and 24	2270	1.19	1.17	0.862
Number of persons between 25 and 64	2270	1.88	1.86	0.714
Number of people 65 and older	2270	0.19	0.13	0.050
Number of men	2270	3.02	2.97	0.658
Number of women	2270	3.12	3.07	0.633
Rooms in the house	2270	1.63	1.58	0.539
Distance to school (hours)	2270	0.31	0.26	0.149
Distance to health center (hours)	2270	1.28	1.18	0.542
Distance to municipal headquarters (hours)	2270	1.72	1.58	0.438
Owns toilet/latrine	2270	0.73	0.71	0.603
Water in house	2270	0.11	0.12	0.744
Access to electricity	2270	0.35	0.38	0.687
Own land	2270	0.65	0.63	0.591
Total income per capita (cordobas)	2270	3794	3811	0.956
Total consumption per capita (cordobas)	2270	4547	4562	0.964
Total food consumption per capita (cordobas)	2270	3180	3063	0.595

Note: P-values based on standard errors clustered by community

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

	Children 0-83 months old				Children 36-83 months old		
	Denver				TVIP	Mccarthy	
	Social	Language	Fine Motor	Gross motor	Receptive language	Memory	Leg motor
Child is in lowest 25% of international distribution for at least one task.							
All	0.64	0.82	0.60	0.44	0.97	0.85	0.41
Child is in lowest 10% of international distribution for at least one task.							
All	0.46	0.61	0.40	0.29	0.85	0.61	0.24
Boys	0.48	0.65	0.42	0.29	0.84	0.60	0.24
Girls	0.44	0.58	0.37	0.28	0.86	0.61	0.25
0-35 months	0.30	0.48	0.28	0.41			
36-59 months	0.48	0.59	0.53	0.27	0.70	0.56	0.21
60-83 months	0.65	0.79	0.43	0.14	0.97	0.65	0.27

Note: Calculations for TVIP and McCarthy based on information about distribution in the international reference population (Mexico and Puerto Rico for TVIP, US for Mccarthy. Denver results reflect the direct outcomes of the Denver subtests, which considers whether tasks are in the lowest quartile or decile compared to a US reference population

Table 3: Main reduced-form treatment effects

	Denver Test			Gross Motor	TVIP	McCarthy		BPI
	Social	Language	Fine Motor		Receptive Language	Memory	Leg Motor	Behavior Problems
Treated	.134** (.059)	.166*** (.060)	.099 (.068)	.007 (.066)	.223*** (.078)	.092 (.072)	.002 (.109)	.053 (.063)
N	3454	3432	3420	3405	1971	1985	1993	2091

Note: The table reports coefficients and standard errors corrected for clustering (in parentheses). Age trends have been removed from all outcomes. All outcomes are then standardized by subtracting the mean and dividing by the standard deviation of the control group. For the Denver, the sample includes children between the ages of 0 and 83 months; for the TVIP, McCarthy, and BPI the sample includes children between the ages of 36 and 83 months. For the Denver test, the dependent variables are defined in terms of the number of delays plus cautions. Standard errors correct for clustering at the community level.

Table 4: Robustness checks

	Denver Test			
	Social	Language	Fine Motor	Gross Motor
Basic specification	.134** (.059)	.166*** (.060)	.099 (.068)	.007 (.066)
Including additional controls	.120** (.057)	.144*** (.052)	.090 (.070)	.015 (.067)
Without dropping outliers	.144** (.056)	.162*** (.058)	.076 (.066)	-.034 (.068)
Without removing age effects	.122** (.058)	.152*** (.056)	.085 (.067)	.010 (.063)
Using narrower age range for Denver	.145** (.061)	.122* (.063)	.067 (.070)	.004 (.074)
Including non-eligible	.125** (.060)	.148** (.060)	.089 (.069)	-0.003 (.066)
Excluding children with early enrollment in primary school	.149** (.060)	.138** (.061)	.077 (.068)	0.003 (.071)
Only children 10-83 months old	.133** (.063)	.187*** (.065)	.101 (.072)	.018 (.070)
Number of tasks in bottom decile (not quartile)	.120** (.059)	.147*** (.059)	.084 (.063)	.002 (.055)
At least one task in bottom decile	.059** (.027)	.031 (.027)	.000 (.030)	-.007 (.025)
At least two tasks in bottom decile	.041* (.021)	.056** (.026)	.034 (.022)	-.004 (.015)
Using only observed (not reported) outcomes		.199*** (.063)	.130* (.069)	.040 (.072)
	TVIP	McCarthy		BPI
	Receptive Language	Memory	Leg Motor	Behavior Problems
Basic specification	.223*** (.078)	.092 (.072)	.002 (.109)	.053 (.063)
Including additional controls	.189*** (.071)	.071 (.066)	.013 (.111)	.071 (.065)
Without dropping outliers	.241*** (.078)	.087 (.069)	.006 (.107)	.074 (.056)
Without removing age effects	.166*** (.070)	.052 (.060)	.001 (.093)	.032 (.065)
Including non-eligible	.185** (.074)	.081 (.071)	.001 (.110)	.060 (.066)
Excluding children with early enrollment in primary school	.186*** (.069)	.070 (.070)	.014 (.110)	.077 (.062)

Note: The table reports coefficients and standard errors corrected for clustering (in parentheses). Also, see notes at foot of table 3.

Table 5: Heterogeneity of treatment effects by child gender and age

	Denver Test			Gross Motor	TVIP	McCarthy		BPI
	Social	Language	Fine Motor		Receptive Language	Memory	Leg Motor	Behavior Problems
Heterogeneity by child gender								
T*Boy	.188** (.077)	.206** (.092)	.125 (.083)	-.023 (.073)	.233** (.095)	.068 (.095)	-.073 (.112)	.041 (.080)
T*Girl	.077 (.068)	.121** (.056)	.072 (.075)	.042 (.088)	.219** (.096)	.114 (.082)	.081 (.134)	.064 (.075)
Boy	-.141* (.074)	-.170* (.088)	-.090 (.072)	.082 (.081)	.060 (.090)	.005 (.089)	.111 (.102)	.003 (.068)
F-test	0.193	0.374	0.520	0.489	0.901	0.659	0.184	0.799
Heterogeneity by child age								
T*AGE1	.079 (.076)	.056 (.060)	.028 (.076)	-.016 (.093)				
T*AGE2	.234*** (.088)	.174 (.110)	.144 (.103)	.005 (.109)	.054 (.052)	.111 (.079)	-.032 (.131)	.151* (.084)
T*AGE3	.120 (.105)	.300*** (.088)	.152 (.113)	.040 (.066)	.362*** (.118)	.077 (.093)	.031 (.107)	-.030 (.078)
AGE2	.064 (.098)	.075 (.095)	.008 (.076)	.042 (.105)				
AGE3	.002 (.109)	.038 (.075)	.023 (.114)	.013 (.077)	.013 (.088)	.044 (.077)	.015 (.079)	.033 (.081)
F-test 1	0.183	0.293	0.200	0.861				
F-test 2	0.306	0.265	0.953	0.733	0.005	0.733	0.501	0.076
F-test 3	0.743	0.011	0.340	0.565				

Note: The table reports coefficients and standard errors corrected for clustering (in parentheses). Age trends have been removed from all outcomes. All outcomes are then standardized by subtracting the mean and dividing by the standard deviation of the control group. For the Denver, the sample includes children between the ages of 0 and 83 months; for the TVIP, McCarthy memory test, and the BPI the sample includes children between the ages of 36 and 83 months. For the Denver test, the dependent variables are defined in terms of the number of delays plus cautions. Standard errors correct for clustering at the community level. AGE1 refers to children aged 0-35 months, AGE2 refers to children aged 36-59 months, and AGE3 refers to children aged 60-83 months. The F-test in the middle panel reports the p-value on a test of equality of the coefficients on (T*girl) and (T*boy). The F-tests in the lower panel reports the p-value on the following tests: F-test 1 tests of equality of the coefficients on (T*AGE1) and (T*AGE2), F-test 2 tests of equality of the coefficients on (T*AGE2) and (T*AGE3), and F-test 3 tests of equality of the coefficients on (T*AGE1) and (T*AGE2).

Table 6: Treatment effects on intermediate inputs

Variable	Obs	Mean control	Treatment effect	s.e.
Hh-level food consumption per capita (logs)				
Total food	2265	8.028	0.310***	(0.04)
Staples	2270	7.214	0.195***	(0.05)
Animal protein	2270	5.488	1.071***	(0.12)
Fruit and vegetables	2270	4.580	1.005***	(0.12)
Child food intake (nr days in last week)				
Tortilla	3499	5.950	0.043	(0.10)
Milk	3499	1.580	1.141***	(0.23)
Meat	3497	0.564	0.764***	(0.08)
Eggs	3499	1.594	1.258***	(0.14)
Fruit	3498	2.552	0.452**	(0.20)
Vegetables	3499	1.468	0.713***	(0.21)
Stimulus				
Got toy in last 6 months	3445	0.271	0.068**	(0.03)
Has pen and paper to draw	3480	0.690	0.101***	(0.03)
Has books	3490	0.073	0.067***	(0.02)
Somebody tells stories	3494	0.520	0.125***	(0.04)
Somebody reads to child	3491	0.080	0.081***	(0.02)
Nr of hours read to	3476	0.134	0.257***	(0.06)
Enrolled in stimulus program or (pre)school	3497	0.503	0.064*	(0.04)
Early enrollment in primary school	3506	0.096	0.025*	(0.01)
Preventive health care and health outcomes				
Birth weight in kg (0-5 monts old)	174	2.987	0.161	(0.13)
Waz-score	3137	-0.958	-0.052	(0.10)
Haz-score	3136	-1.243	-0.052	(0.12)
Consulted doctor if sick or diarea	2568	0.730	0.057**	(0.03)
Weighed in last 6 months	3473	0.705	0.063***	(0.02)
Received vitamin A or iron in last 6 months	3490	0.734	0.086***	(0.02)
Received deworming drugs in last 6 months	3489	0.566	0.066***	(0.03)
Caregivers				
CESD depression scale	1855	0	-0.113*	(0.06)
HOME scale	1939	0	0.0138	(0.11)

NOTE: Variables for caregivers are standardized using mean and standard deviation of the control group. Standard errors corrected for clustering at community level.

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

	Mean all	Mean all re- interviewed
Household characteristics		
Male household head	0.82	0.82
Household size	5.28	5.30
Number of children under 5	0.65	0.65
Number of children between 5 and 14	1.53	1.54
Number of persons between 15 and 24	1.07	1.08
Number of persons between 25 and 64	1.77	1.78
Number of people 65 and older	0.24	0.24
Number of men	2.65	2.67
Number of women	2.62	2.63
Rooms in the house	1.62	1.63
Distance to school (hours)	0.27	0.27
Distance to health center (hours)	1.14	1.14
Distance to municipal headquarters (hours)	1.54	1.54
Owens toilet/latrine	0.76	0.76
Water in house	0.13	0.13
Acces to electricity	0.40	0.40
Own land	0.66	0.66
Total income per capita (cordobas)	4742	4725
Total consumption per capita (cordobas)	5443	5424
Total food consumption per capita (cordobas)	3537	3527
Nr of observations	4021	3969

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

	All		Did all tests	
	Nr. of observations	mean	Nr. of observations	mean
Child-specific characteristics				
All Children				
Male	3505	0.50	3325	0.50
Age in months	3506	43.67	3326	43.27
0-35 months old	3506	0.40	3326	0.42
36-59 months old	3506	0.27	3326	0.26
60-83 months old	3506	0.32	3326	0.32
Years education mother	3481	4.06	3303	4.09
Years education father	3218	3.55	3057	3.57
Children 0-5 at baseline				
Weight-for-age z-score	2365	-1.02	2219	-0.99
Height-for-age z-score	2355	-1.23	2211	-1.20
Weight-for-height z-score	2369	-0.17	2224	-0.15
Consulted doctor if sick or diarrhea	1599	0.75	1494	0.75
Weighted in last 6 months	2488	0.91	2329	0.91
Received vitamins in last 6 months	2488	0.69	2329	0.69
Received deworming drugs in last 6 months	2488	0.53	2329	0.52
Household characteristics				
Male household head	2270	0.86	2197	0.86
Household size	2270	6.07	2197	6.09
Number of children under 5	2270	1.14	2197	1.14
Number of children between 5 and 14	2270	1.73	2197	1.74
Number of persons between 15 and 24	2270	1.18	2197	1.18
Number of persons between 25 and 64	2270	1.87	2197	1.87
Number of people 65 and older	2270	0.15	2197	0.15
Number of men	2270	2.99	2197	2.99
Number of women	2270	3.08	2197	3.09
Rooms in the house	2270	1.59	2197	1.59
Distance to school (hours)	2270	0.27	2197	0.28
Distance to health center (hours)	2270	1.20	2197	1.20
Distance to municipal headquarters (hours)	2270	1.62	2197	1.63
Owens toilet/latrine	2270	0.72	2197	0.72
Water in house	2270	0.12	2197	0.12
Access to electricity	2270	0.37	2197	0.37
Own land	2270	0.63	2197	0.64
Total income per capita (cordobas)	2270	3806	2197	3812
Total consumption per capita (cordobas)	2270	4558	2197	4553
Total food consumption per capita (cordobas)	2270	3093	2197	3092

NOTE: children 0-83 months

Appendix Table A3: Test outcomes for children who did all tests versus those that missed some but not all

Outcome	Did all tests		Did not do all tests		P-value difference
	Obs	Mean	Obs	Mean	
Denver: personal	3302	0.116	152	-0.214	0.0023
Denver: language	3305	0.139	127	-0.245	0.0021
Denver: fine motor	3312	0.084	108	-0.202	0.0233
Denver: gross motor	3312	0.028	93	-0.780	0.0000
TVIP	1879	0.176	92	0.009	0.1330
McCarthy-memory	1924	0.079	61	-0.255	0.0149
McCarthy-leg motor	1921	0.019	72	-0.456	0.0011
BPI	1923	0.032	151	-0.013	0.6240

Note: Standard errors clustered at community level

Appendix Table A4: Treatment effects on intermediate inputs: youngest and oldest age groups

	Children 12-35 months				Children 60-83 months				P-value difference young versus old
	Obs	Mean control	Treatment effect	s.e.	Obs	Mean control	Treatment effect	s.e.	
Child food intake (nr days in last week)									
Tortilla	945	6.19	0.109	(0.17)	1130	6.65	0.202*	(0.11)	0.546
Bread	945	2.70	1.094***	(0.31)	1129	2.27	1.834***	(0.27)	0.003
Milk	945	2.33	0.796**	(0.31)	1130	1.01	1.578***	(0.21)	0.006
Meat	945	0.63	0.700***	(0.10)	1129	0.54	0.939***	(0.10)	0.019
Eggs	945	1.82	1.166***	(0.18)	1130	1.66	1.522***	(0.21)	0.107
Fruit	945	2.37	0.924***	(0.22)	1130	2.88	0.561*	(0.29)	0.179
Vegetables	945	1.49	0.832***	(0.22)	1130	1.65	0.781***	(0.29)	0.842
Fruit juice	943	3.07	1.060***	(0.25)	1128	2.30	1.532***	(0.23)	0.074
Stimulus									
Got toy in last 6 months	936	0.33	0.079**	(0.04)	1116	0.24	0.051	(0.04)	0.534
Has pen and paper to draw	942	0.60	0.159***	(0.04)	1125	0.83	0.082***	(0.03)	0.083
Has books	945	0.07	0.038	(0.03)	1126	0.07	0.088***	(0.02)	0.080
Somebody tells stories	946	0.63	0.093*	(0.05)	1127	0.41	0.158***	(0.04)	0.170
Somebody reads to child	945	0.08	0.037	(0.02)	1128	0.10	0.128***	(0.03)	0.004
Nr of hours read to	940	0.09	0.193***	(0.07)	1122	0.19	0.259***	(0.09)	0.511
Enrolled in stimulus program or (pre)school	944	0.18	0.110*	(0.06)	1131	0.68	0.070**	(0.03)	0.531
Early enrollment in primary school		-	-	-	1132	0.29	0.079**	(0.04)	-
Preventive health care and health outcomes									
waz-score	934	-1.11	-0.114	(0.13)	799	-1.28	0.046	(0.11)	0.186
haz-score	933	-1.27	-0.116	(0.12)	798	-1.56	-0.013	(0.13)	0.360
Consulted doctor if sick or diarea	756	0.82	0.027	(0.03)	809	0.62	0.105**	(0.04)	0.089
Weighed in last 6 months	941	0.97	-0.004	(0.02)	1115	0.26	0.138***	(0.04)	0.001
Received vitamin A or iron in last 6 months	945	0.86	0.090***	(0.03)	1123	0.63	0.167***	(0.04)	0.108
Received deworming drugs in last 6 months	945	0.48	0.079*	(0.04)	1122	0.67	0.136***	(0.04)	0.264

NOTE: Standard errors corrected for clustering at community level.