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

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

September 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. We thank Jere Behrman, Sally Grantham-McGreggor, John Maluccio, Christina Paxson, Elisabeth Sadoulet, Miguel Urquiola, seminar participants at the World Bank, the Inter-American Bank, the Institute for Fiscal Studies, Berkeley, Leuven, and Wisconsin-Madison for providing us with very helpful comments. The views expressed in this paper are those of the authors and do not necessarily reflect those of the World Bank or any of its affiliated organizations. All errors and omissions are our own. Contacts: kmacours@jhu.edu, nschady@worldbank.org and rvakis@worldbank.org

Cash Transfers, Behavioral Changes, and Cognitive Development in Early Childhood:

Evidence from a Randomized Experiment

Abstract

In many developing countries, cash transfer programs are an important component of the social safety net. A large number of studies have assessed the impact of such programs on consumption poverty, health status, nutrition, and education. Much less is known about the extent, if any, to which cash transfers also improve the cognitive and socio-emotional development of young children. This is important because a variety of theories of skill formation suggest that investments in schooling and other dimensions of human capital will have lower returns if children do not have adequate levels of cognitive and social skills at early ages. This paper analyzes the impact of a randomized cash transfer program on cognitive development in early childhood in rural Nicaragua. It shows that the program had significant effects on cognitive outcomes, especially language. Impacts are larger for older pre-school aged children, who are also more likely to be delayed. The program increased intake of nutrient-rich foods, early stimulation, and use of preventive health care—all of which have been identified as risk factors for development in early childhood. Households increased expenditures on these inputs more than can be accounted for by the increases in cash income only, suggesting that the program changed parents' behavior. The findings suggest that gains in early childhood development outcomes should be taken into account when assessing the benefits of cash transfer programs in developing countries. More broadly, the paper illustrates that gains in early childhood development can result from interventions that facilitate investments made by parents to reduce risk factors for cognitive development.

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 have higher school attainment, better test scores, lower rates of criminality, and earn higher wages in adulthood (Currie 2001; Schweinhart 2005), although the impacts appear to be concentrated among girls (Anderson 2007). A well-known study from Jamaica shows that children randomly assigned to receive home-based early stimulation have substantial improvements in cognitive development and subsequent school performance (Grantham-McGregor et al. 1991 and 1997; Walker et al. 2000; Powell et al. 2004). Non-experimental evidence suggests that 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 et al. 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). In Guatemala, children exposed to a nutritional intervention have better reading comprehension and perform better on tests of cognitive development in adulthood, and earn higher wages (Maluccio et al. 2008; Hoddinott et al. 2008).

A reasonable amount of evidence is therefore available on how the cognitive development of young children responds to supply side interventions, including access to preschool or food supplementation programs. Much less is known about interventions that attempt to directly affect the investments parents make in the cognitive development of their children—either by relieving financial constraints, or by changing how resources are allocated within households.

This paper analyzes the impact of a cash transfer program on 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. Children in better-off households appear to have higher levels of cognitive development than those in poorer households in developing countries.[1] 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. 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.[2] 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 have transmitted knowledge about child-rearing

---

[1] References include Paxson and Schady (2007, 2008) on Ecuador; Halpern et al. (1996) on Brazil; Ghuman et al. (2005) on the Philippines; see also Schady (2006) for a discussion.

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

practices; it may also have affected how transfer income was used through a flypaper or labeling effect.[3] Such changes in behaviors could be further enhanced through social interactions with other program beneficiaries and peer pressure (Macours and Vakis 2008). Finally, *Atención a Crisis* transfers are made to women, and there is some evidence that income controlled by women is spent in a way that benefits children more than income that is controlled by men.[4]

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 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 (Fiszbein and Schady 2008). 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.[5] In contrast, we are aware of only two earlier studies which estimate the impact of a cash transfer program on cognitive development. Fernald et al. (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

[3] See Thaler (1999) for a general discussion. Fraker et al. (1995) presents evidence for the US, Kooreman (2000) for the Netherlands, Jacoby (2002) for the Philippines, and Islam and Hoddinott (2008) for Guatemala. An exception is Edmonds (2002) who finds no evidence of labeling effects for child benefit income in Slovenia.

[4] For example, Thomas (1990), Hoddinott and Haddad (1995), Doss (2005), and Schady and Rosero (2008) 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) and Ward-Batts (2008) present quasi-experimental evidence from the United Kingdom to argue that income controlled by women is more likely to be spent on clothing for women and children than income controlled by men.

[5] The literature is extensive. On education outcomes see, among others, Schultz (2004) and Behrman et al. (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. Fiszbein and Schady (2008) review the findings from these studies.

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 development for children as young as one month of age. We can therefore estimate the impact of the program among young children, which is important 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 6 municipalities in rural Nicaragua.[6] The beneficiaries of the pilot randomly

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

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).[7] Women in beneficiary households receive sizable cash transfers every 2 months, averaging about 15 percent of per capita expenditures.[8] 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 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

---

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

[8] 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.

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.[9]

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

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 two surveys was minimal, less than 1.3 percent. Attrition is uncorrelated with treatment status, and the baseline characteristics of the full sample of children and those that could be located at follow-up are very similar—see Appendix 1.

Program take-up in the treatment group was more than 95 percent, and contamination of the control was negligible (one household).[10] As discussed above (footnote 9), 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

[9] 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.

[10] 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.

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 households randomly assigned to receive *Atención a Crisis* transfers and households in the control group: 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 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). The Denver can be applied to children as young as one month of age. A slightly modified version of the Denver is used for child monitoring by the national early childhood stimulation program in Nicaragua, which suggests that the test is appropriate for the population we study.

For children age 36-83 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. We also use a short-term memory test and a leg motor test from the McCarthy test battery. 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". 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.[11] All of the tests we use were carefully piloted in the field, and adjustments were made, as necessary. Details of all of the tests we use are provided in Appendix 2.

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. The data also include information on child food intake, 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 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 is delayed, although this varies considerably by outcome. The fraction of children who are behind for their age is largest for the measures of language—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

[11] 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.

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 is 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 with high poverty levels and low education in Latin America.[12]

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 that there appear to be no obvious differences in delays between boys and girls, but 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, but not for the measures of fine and gross motor skills. In sum, there are striking age patterns in some outcomes,

[12] In the sample of children in Ecuador analyzed by Paxson and Schady (2007, 2008) the average child places in the 11th percentile of the distribution of the TVIP. 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 of short-term memory, and in the 13th percentile of a test of long-term memory from the Woodcock-Johnson battery of tests. However, the sample of children from Ecuador in Paxson and Schady is considerably better off than our sample of children from Nicaragua. 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.

with older children being more likely to be delayed than younger children. This may 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.

For the rest of the analysis, we 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.[13] 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.

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

(1) $$Z_k = T_1\alpha_k + T_2\beta_k + T_3\chi_k + \varepsilon_k, \quad k=1 \ldots K,$$

---

[13] 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.

where Z_k is the *k*th z-score for a given outcome (out of 8), and T_1, T_2, and T_3 are intent-to-treat indicators for households randomly assigned to the three treatment packages in the *Atención a Crisis* program. Because of the standardization described above, all units are in standard deviations. Standard errors throughout the paper adjust for clustering at the community level.

As we show, the coefficients on T_1, T_2, and T_3 are very close in magnitude, and we can never reject the null hypothesis that they are equal. We therefore next focus on a specification that treats the three packages as if they were a single *Atención a Crisis* program:

(2) $Z_k = T\phi_k + \varepsilon_k$, k=1…K.

Results for these specifications are presented in Table 3, for the three separate treatment packages (upper panel) and for the specification with a single treatment (lower panel). The table shows there are significant program effects for the social-personal and language measures of the Denver test, corresponding to impacts of 0.13 and 0.17 standard deviations, respectively, as well as for the TVIP test of receptive vocabulary, corresponding to an impact of 0.22 standard deviations. Treatment effects for the three packages are very similar throughout. Although this may seem surprising, we note that the observed increase in overall consumption levels at the time of the follow-up survey was similar for households assigned to the three treatment packages (Macours and Vakis 2008). At this point, 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. Also, 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.

Table 3 makes obvious that program effects are concentrated in social-personal skills and, especially, in language development. Note that these are domains where children had particularly large delays, and where socioeconomic gradients were steeper. Program effects for all other outcomes are also 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.

How large are the estimated program effects? One way to put the magnitudes in context is by comparing them with the depth of the delays. The median child in our sample is 28 months delayed on the TVIP. Turning the results in Table 3 into program effects on the number of months delayed suggests that the *Atención a Crisis* program allowed children to make up approximately 1.5 months delay on the TVIP. This is a modest effect relative to the depth of the delays, although it is worth remembering that households had only received transfers for a short period of time.

We conducted a large number of robustness tests to these results. Table 4 shows that our findings are robust to controlling for age and gender in the regression, as opposed to removing age effects first; to adding the education of parents as extra controls; to inclusion of the 1 percent largest outliers (in absolute value); and to not removing age effects at all. The are also robust 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 years of age (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).[14] 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 program effects that are based on the sample of children for whom the enumerator actually *observed* a child performing all appropriate tasks in the Denver, rather than those that are based on the sample of children where some tasks

[14] In the specification that limits the sample to children age 5 and younger, the point estimate for the Denver language subtest is about 30 percent lower. This is consistent with results below which show that the largest treatment effects on language are concentrated among older children. However, 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.

were observed and others were reported by parents, suggest larger treatment effects on language and fine motor skills than those reported in Table 3.[15]

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 (2) which include interactions with treatment, as is standard:

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

where *X* is a child characteristic (for example, an indicator variable 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 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 "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.

---

[15] 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 sub-test, as most items in this task rely on caregiver responses.

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.[16] For the language subscale of the Denver, the effect of receiving *Atención a Crisis* transfers rises from 0.06 (standard error: 0.06) for the youngest children, to 0.17 (standard error: 0.11) for children in the middle age group, to 0.20 (standard error: 0.09) for the oldest children; the difference in program effects between the youngest and oldest children is significant (p-value: 0.01). For the TVIP, the program effect is 0.05 (standard error: 0.05) for children between the ages of 36 and 59 months, and 0.36 (standard error: 0.12) for children age 60 months and older; this difference in program effects is once again significant (p-value: 0.005).

How large are the program effects on language among the oldest children in the sample? Once again turning the coefficients in Table 4 for children age 60 months and older into program effects on the number of months delayed is informative. This suggests that the *Atención a Crisis* program allowed the oldest group of treated children in our sample to make up approximately 2.4 months delay on the TVIP.

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 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. (For the BPI, the P-value of the F-test is 0.05) . There is no apparent pattern of age effects

[16] This result is not sensitive to the method used for age standardization. We obtained very similar results when age is controlled for by including a series of age-month dummies, instead of removing the age effects based on the estimated trend in the control group.

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.[17]

The fact that program effects on language outcomes are larger among older children is somewhat surprising given a consensus view that very early childhood is a particularly important period for development, and it is therefore important to consider this finding carefully. There are a number of possible explanations for this finding. First, it may be 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 difference in program effects across age groups we estimate.[18] Second, it is possible that it takes longer for the benefits of the *Atención a Crisis* to become apparent among younger children—see Behrman et al. (2005), Armecin et al. (2008) and, in particular, Behrman and King (2008) for a thoughtful discussion of considerations of timing and duration of exposure in explaining the impact of programs, including child development programs, on outcomes. Third, it may be that the *Atención a Crisis* transfers were used in such a way that they particularly benefited older children, and we present some evidence below that is consistent with this interpretation. 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. This stands in contrast with most studies of program effects on nutritional outcomes—a well established finding in this literature is that, after 2 to 3 years of age, the potential for catch-up in height is quite limited (Martorell et al. 1994; Martorell 1999; Victora et al. 2008).

[17] 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.

[18] 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, suggesting that the test consistently measures language skills and knowledge for both younger and older children. 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.

5. Transmission mechanisms

The results in section 4 make clear that the *Atención a Crisis* program had positive effects on a number of dimensions of child development. We now 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 and the health status of children, and on measures of maternal depression 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.

A. *Treatment effects on intermediate inputs*

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. For inputs related to household food consumption, and parenting and mental health, the sample size corresponds to the number of households. For inputs related to child food intake, stimulation, and health status, the sample size corresponds to the number of children.[19]

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

Table 6 further shows that treated households have higher values for various indicators of early stimulation: books, and paper and pencil are more likely to be available for children, children are more likely to be read and to have stories told to them, and there is a marginally significant impact on early enrollment in primary school.[20]

Finally, Table 6 shows program effects on a number of measures of preventive health care—treated children are more likely to have had a growth check-up, and to have received vitamins, iron, and deworming drugs.[21] These children are also more likely to see improvements in mother-reported health status, are less

---

[19] As before, these regressions include all children between 0 and 83 months, and hence include those children that spent part of the time during which the treatment was available in utero. Because the margin for changes in inputs for these very young children might be limited, we also estimated all regressions excluding children from 0 to 9 months old and obtained very similar results.

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

[21] These self-reported measures of growth check-ups 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

likely to have been in bed because of illness, and are more likely to have been taken to a doctor if they were ill. On the other hand, there is no program effect on child anthropometric measures, or on 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. (For this last measure, the point estimate is positive but not significant, but sample sizes are very small.) 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.)

Table 7 presents a comparable analysis broken down by child age. We focus on inputs that are child-specific—for example, child food intake (rather than expenditures on food items, which cannot be disaggregated for individual children), the likelihood that a child received a toy, or was read to (rather than the availability of books, or paper and pencil for children), and measures of child health. The table shows that for a number of food items, including tortilla, milk, meat and eggs, the program effects on food intake are larger among older children—and in the case of milk, meat, and eggs the differences are significant. Treatment effects on the probability that a child has been read to are also significantly larger among older children. Finally, for older children there are larger impacts on the probability that they received vitamin A or iron, de-worming medication, or had a growth check-up, compared to younger children. These results might be driven by the fact that there is more room for changes in behavioral patterns for older children—while diets of very young children are likely to be similar for all households (milk), variation in the nutritional value of diets is larger among older children. Nevertheless, regardless of the reasons why changes are larger for older children, the results in Table 7 are consistent with the program effects we observe on child development: Older children randomly assigned to receive transfers had both larger changes in a number of key inputs into child development than younger children, and also saw larger changes in cognitive outcomes.

---

in the sample. Results are robust when only this 87 percent subsample is used, suggesting that results are not driven by reporting bias.

B. Engel curve analysis

The findings in Table 6 are perhaps not surprising, given the magnitude of the transfer. We therefore next 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.[22] 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. 2004).[23] 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. Figure 3 thus shows that households diversified their diets and shifted towards higher quality sources of calories. 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

[22] The increasing slope at the lowest level of expenditure is similar to results in other settings (Thomas 1986).
[23] Fan regressions were estimated with bandwidth of 1 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 2.5 percent highest and lowest values for log per capita expenditures were trimmed, because of the low density of observations in the tails of the distribution.

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).

Figure 4 presents the results for expenditures on other broad categories. 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. Smaller expenditure shares on health are consistent with the improved health status of households that received the *Atención a Crisis* transfers.

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 5 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). The pattern that emerges for indicators of preventive health care and general health status is similar though somewhat less clear. While the upper panel of Figure 6 suggests some upward shifts of the curves the differences between treated and control groups are less often significant.[24]

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 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 heath outcomes. Many of these changes are larger among older children—precisely the age group for which we estimated the largest program effects on language development. The changes in food expenditure patterns and stimulation

[24] In Nicaragua, as in many other developing countries, preventive health care visits are nominally free. This could explain the relative flat socioeconomic gradients.

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 among treated households.

6. Conclusion

In many developing countries, young children suffer from profound delays in cognitive development. 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). Understanding the causes for deficits in early childhood, and identifying interventions that can help address them are very important priorities for research.

This paper uses a randomized evaluation design to estimate the impact of a cash transfer program on a large set of measures of cognitive development among young children in Nicaragua, a low income country. 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 larger among older children, suggesting that there is substantial potential for catch-up in this domain. Furthermore, the positive impacts 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 spent 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, the social marketing of the program, or the fact that transfers may have been understood to be short-term in nature, resulted in these changes in behaviors and expenditure patterns.

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. Transfers provide additional resources to households, allowing them to invest more in children. Parents also change the allocation of resources in ways that appear to benefit young children. This opens up the possibility for new and innovative policies that work directly with parents and caregivers.

References

Agüero, 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*.

Armecin, G., J. Behrman, P. Duazo, S. Ghuman, S. Gultiano, E. King, N. Lee, and The Office of Population Studies, University of San Carlos ECD Team, 2006. "Early Childhood Development through Integrated Programs: Evidence from the Philippines." Cebu City, New York, Philadelphia and Washington: Universities of Pennsylvania and San Carlos (Office of Population Studies), Population Council and World Bank, 2006.

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 thorough Progresa: 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.

Behrman, J. and E. King, 2008. "Program Impact and Variation in Duration of Exposure," in S. Amin, J. Das, and M. Goldstein, eds., *Are You Being Served: New Tools for Measuring Service Delivery*, Washington, DC: World Bank, 147-172.

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. 2002. "Reconsidering the Labeling Effect for Child Benefits: Evidence from a Transition Economy." *Economics Letters* 76: 303-09.

Edmonds, E. 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.

Fan, J. and I. Gijbels 1996. *Local Polynomial Modelling and Its Applications*. New York: Chapman & Hall.

Fernald, L., P. Gertler, and L. Neufeld. 2008. "Role of Cash in Conditional Cash Transfer Programmes for Child Health, Growth, and Development: An Analysis of Mexico's *Oportunidades*." *The Lancet* 371: 828-837.

Fiszbein, Ariel, and Norbert Schady. 2008. *Conditional Cash Transfers: Reducing Present and Future Poverty*. Forthcoming, World Bank.

Fraker, T. M., A. P. Martini, J. C. Ohls. 1995. "The Effect of Food Stamp Cashout on Food Expenditures: An Assessment of the Findings from Four Demonstrations." *Journal of Human Resources* 30(4): 633-49.

Frankenberg, W.K.., and J.B. Dodds. 1996. The Denver Developmental Screening Test. Denver, CO: University of Colorado Medical Center.

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.

Ghuman, S., J. Behrman, J. B. Borja, S. Gultiano and E. King. 2005. “Family Background, Service Providers, and Early Childhood Development in the Philippines: Proxies and Interactions.” *Economic Development and Cultural Change* 54(1): 129-164.

Glewwe, P., and P. Olinto. 2004. “Evaluating the Impact of Conditional Cash Transfers on Schooling: An Experimental Analysis of Honduras. PRAF Program.” Unpublished manuscript, University of Minnesota.

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 Púb.* 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., J. A. Maluccio, J. Behrman, R. Flores and R. Martorell. 2008. “The Impact of Nutrition During Early Childhood on Income, Hours Worked, and Wages of Guatemalan Adults” *Lancet* 371: 411-416.

Islam, M. and J. Hoddinott. 2008. “Evidence of Intra-Household Flypaper Effects from a Nutrition Intervention in rural Guatemala.” *Economic Development and Cultural Change* (forthcoming).

Jacoby, H. 2002. “Is there an Intrahousehold ‘Flypaper Effect’? Evidence from a School Feeding Programme.” *Economic Journal* 112: 196-221.

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.

Martorell, R. 1999. “The Nature of Child Malnutrition and its Long-Term Implications.” *Food and Nutrition Bulletin* 19: 288-92.

Martorell, R., K.L. Khan, D. Schroeder. 1994. “Reversibility of Stunting: Epidemiological Findings in Children from Developing Countries.” *European Journal of Clinical Nutrition* 48(Suppl.): S45-S57.

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 Nutritio*n 134(9): 2336-42.

Oberhelman, R., E. Guerrero, M. L. Fernandez, M. Silio, D. Mercado, N. Comiskey, G. Ihenacho, and R. Mera. 1998. “Correlations between Intestinal Parasitosis, Physical Growth, and Psychomotor Development among Infants and Children from Rural Nicaragua.” *American Journal of Tropical Medical Hygiene* 58(4): 47—75.

Paxson, C., and N. Schady. 2007. “Cognitive Development Among Young Children in Ecuador: The Role of Health, Wealth and Parenting.” *Journal of Human Resources* 42(1): 49-84.

Paxson, C., and N. Schady. 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." *Economía* 8(2): 43-70.

Schady, N., and J. Rosero. 2008. "Are Cash Transfers Made to Women Spent like Other Sources of Income?" Forthcoming, *Economics Letters*.

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.

Strauss, J., and D. Thomas. 1998. "Health, Nutrition, and Economic Development." *Journal of Economic Literature* 36(2): 766-818.

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.

Victora, C., L. Adair, C. Fall, P. Hallal, R. Martorell, L. Richter, H. Singh Sachdev, and for the Maternal and Child Undernutrition Study Group. 2008. "Maternal and Child Undernutrition: Consequences for Adult Health and Human Capital." *Lancet* 371: 340-57.

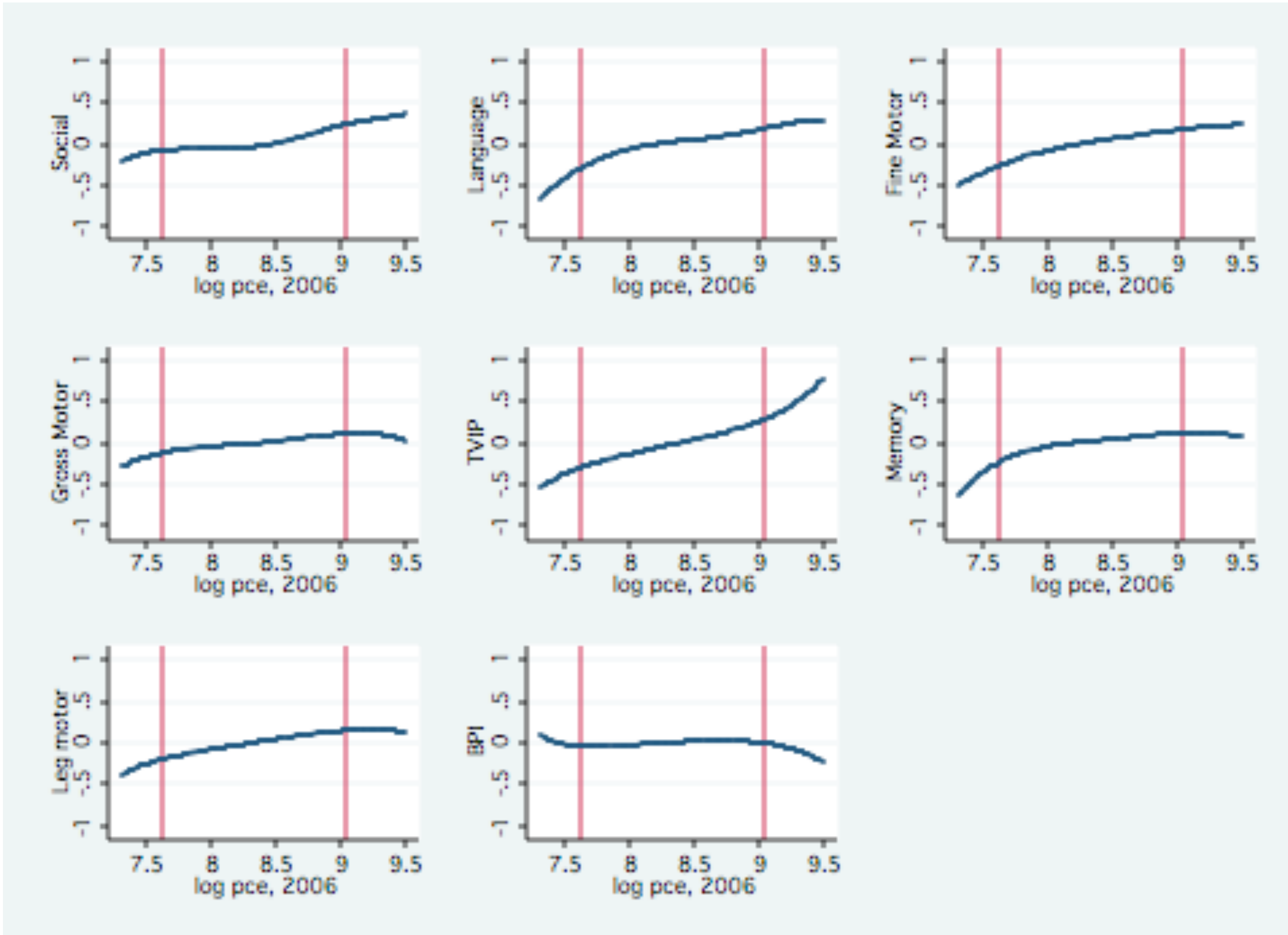
Walker, S.P., and others. 2000. "Effects of Growth Restriction in Early Childhood on Growth, IQ, and Cognition at Age 11 to 12 Years and the Benefits of Nutritional Supplementation and Psychological Stimulation." *Journal of Pediatrics* 137(1): 36-41.

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.

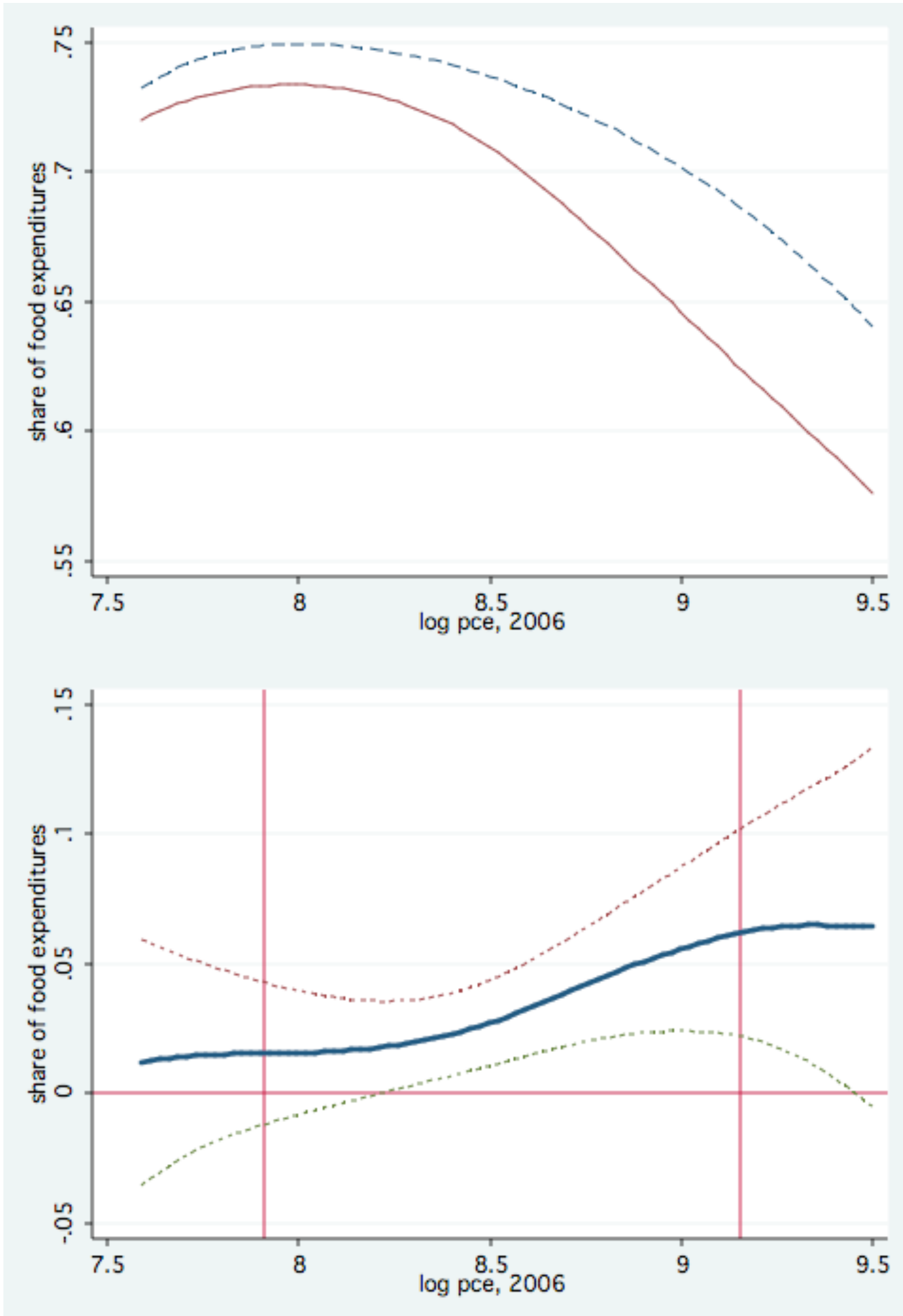
Ward-Batts, J. 2008. "Out of the Wallet and Into the Purse: Using Micro Data to Test Income Pooling." *Journal of Human Resources* 43(2): 325-51.

Figure 1: Socioeconomic gradients in cognitive development



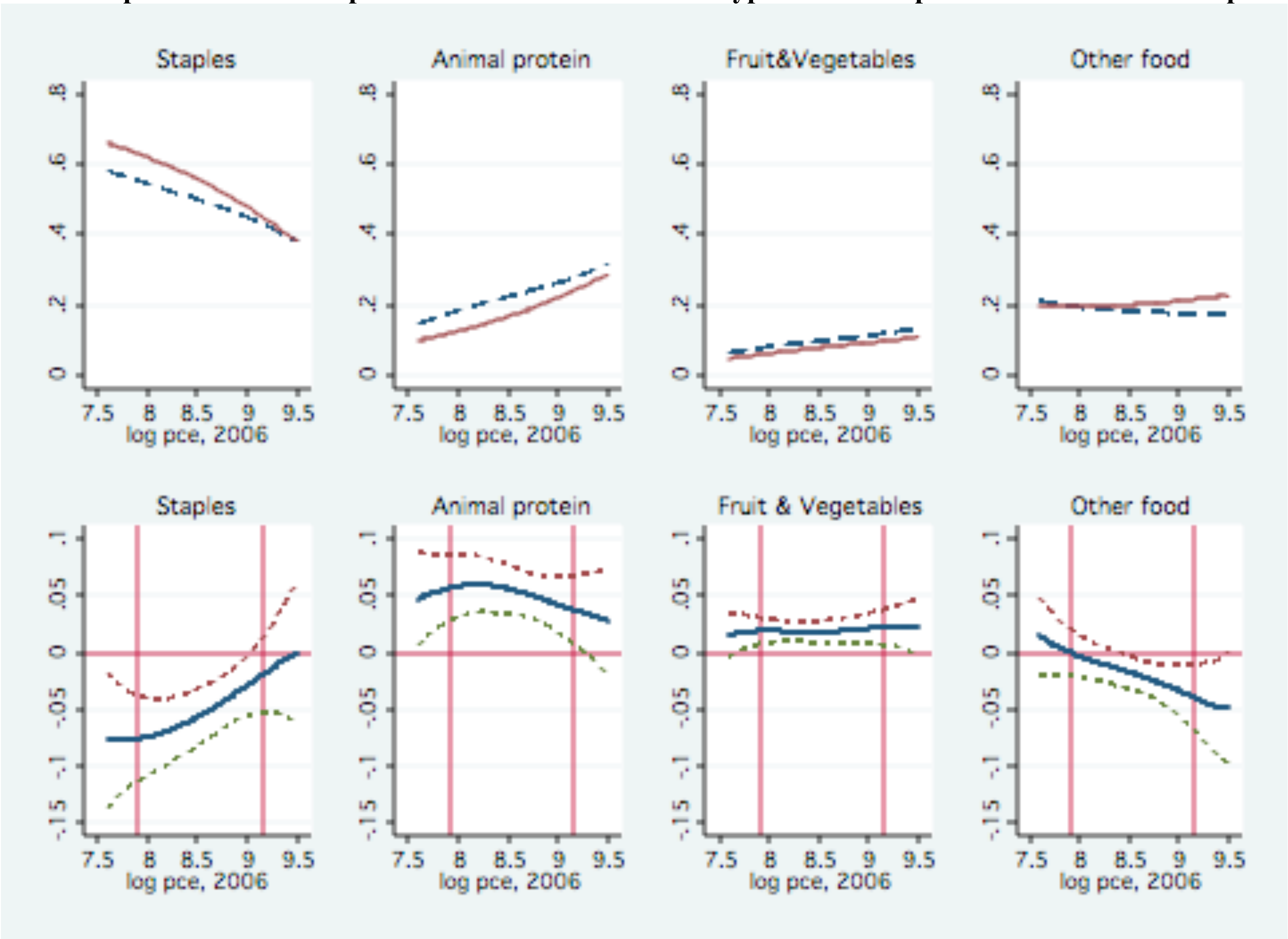
Note: Outcomes for the control group only. Age trends have been removed from all outcomes. All outcomes are then standardized by subtracting the mean and dividing by the standard deviation of the control group. For the Denver (personal, language, fine motor, gross motor), the sample includes children 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. Vertical lines are included at 10th and 90th percentiles of log per capita expenditures in control communities. Fan regressions with bandwidth of 1. 2.5% highest and lowest outliers of log(pce) trimmed from graph.

Figure 2: Food Engel curves



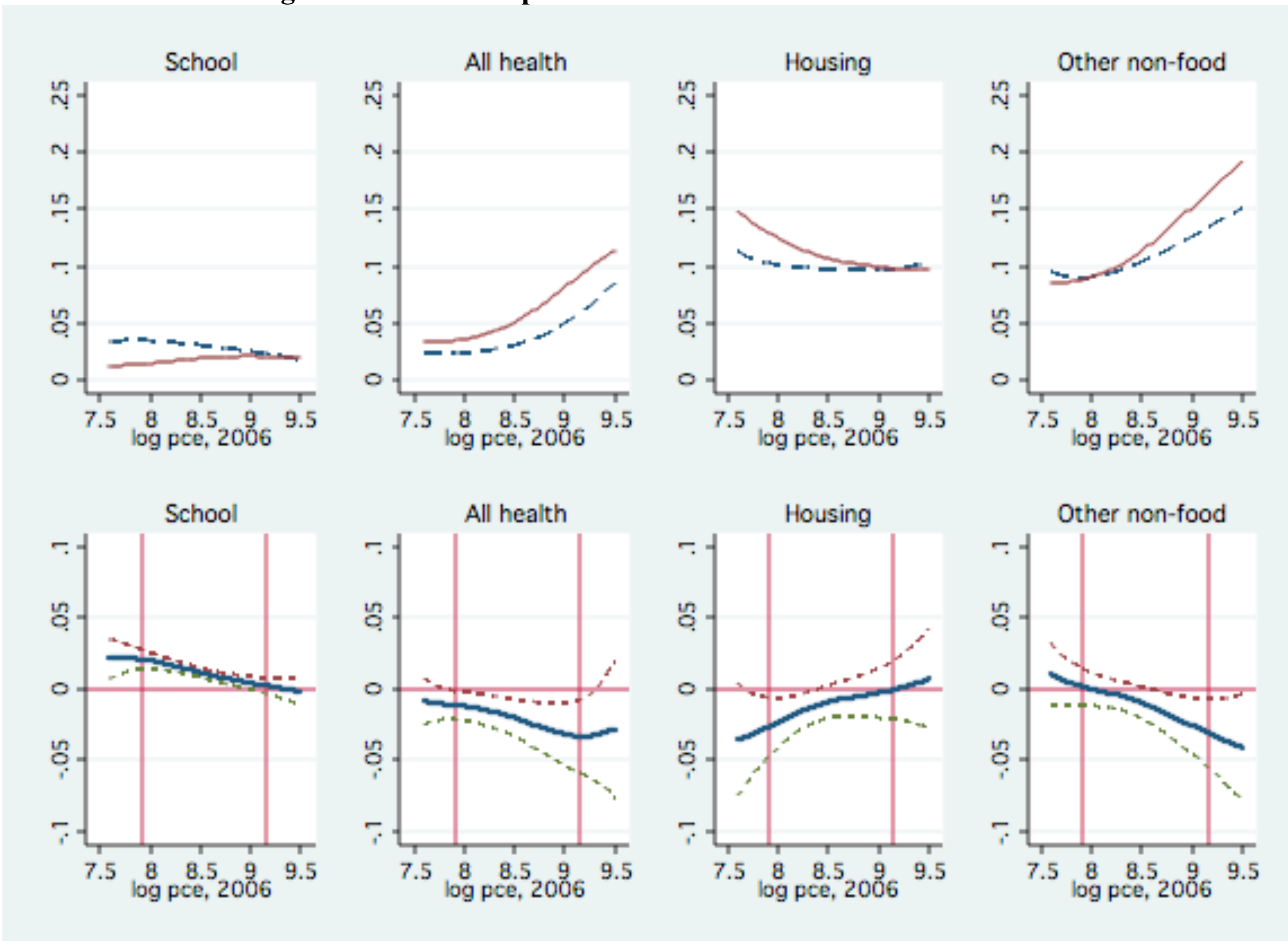
Note: In upper panel the dashed line correspond to treatment, the solid line to control. In lower panel, the bold line corresponds to the difference between treatment and control, and the bounds of the 95% confidence interval are indicated in short dashes. Vertical lines are included at 10^{th} and 90^{th} percentiles of log per capita expenditures (eligible in treatment and control). Fan regressions with bandwidth of 1. 2.5% highest and lowest outliers of log(pce) trimmed from graph.

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



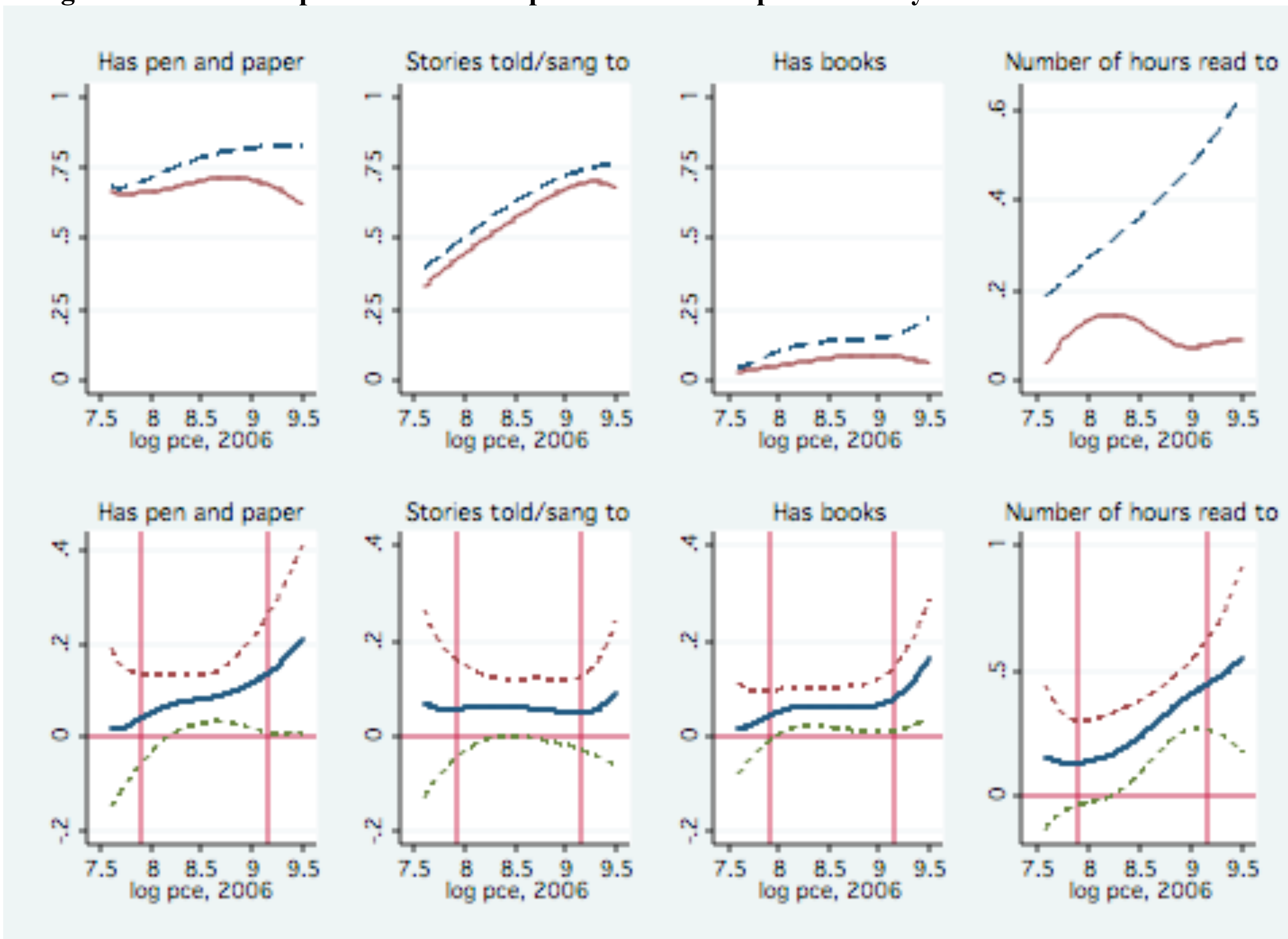
Note: In upper panel dashed lines correspond to treatment, solid lines to control. In lower panel, the bold lines corresponds to the difference between treatment and control, and the bounds of the 95% confidence intervals are indicated in short dashes, reflecting bootstrapped standard errors, clustered by community. Vertical lines are included at 10^{th} and 90^{th} percentiles of log per capita expenditures (eligible in treatment and control). Fan regressions with bandwidth of 1. 2.5% highest and lowest outliers of log(pce) trimmed from graph.

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



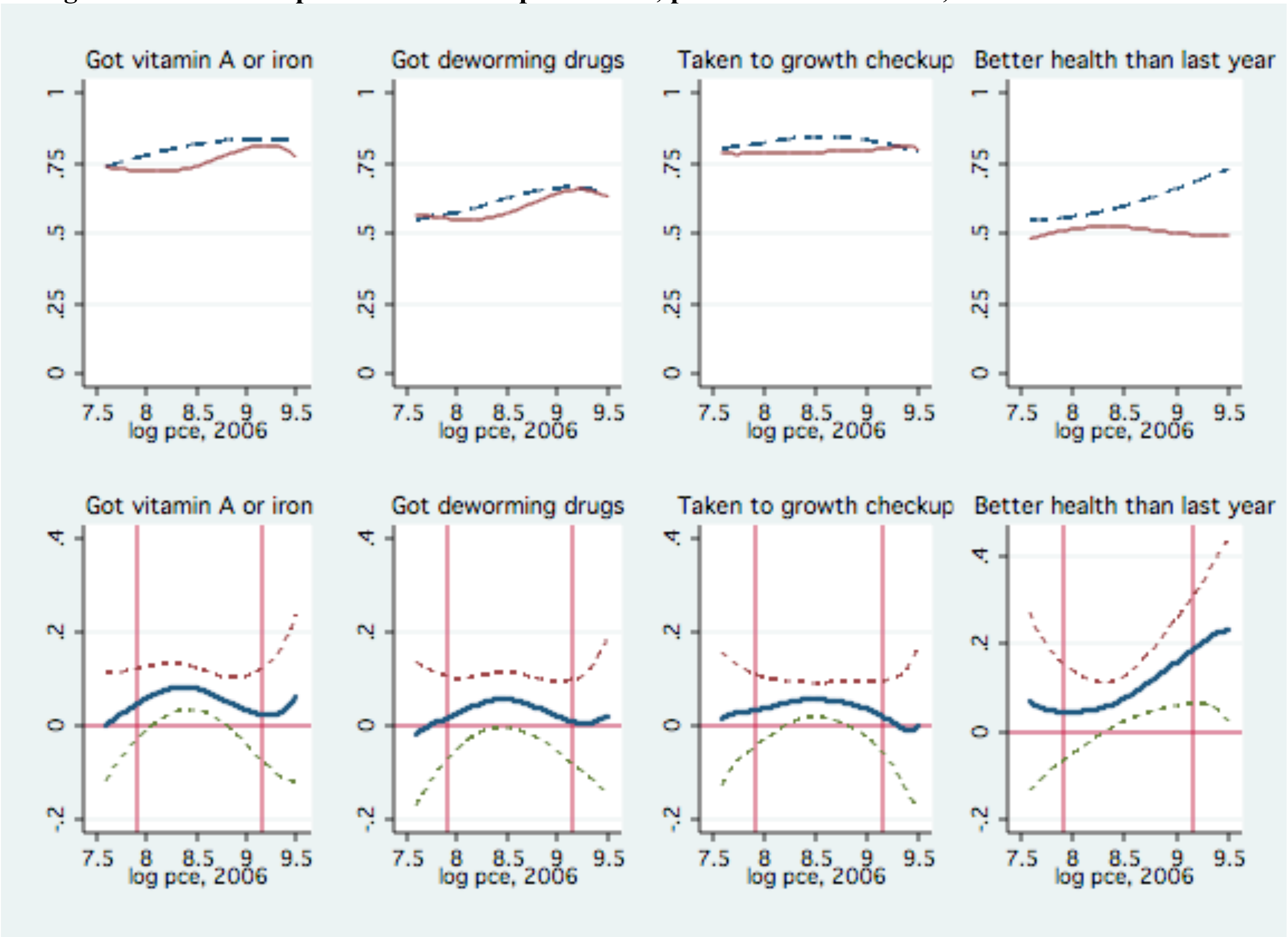
Note: In upper panel dashed lines correspond to treatment, solid lines to control. In lower panel, the bold lines correspond to the difference between treatment and control, and the bounds of the 95% confidence intervals are indicated in short dashes, reflecting bootstrapped standard errors, clustered by community. Vertical lines are included at 10th and 90th percentiles of log per capita expenditures (eligible in treatment and control). Fan regressions with bandwidth of 1. 2.5% highest and lowest outliers of log(pce) trimmed from graph.

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 10^{th} and 90^{th} percentiles of log per capita expenditures (eligible in treatment and control). Fan regressions with bandwidth of 1. 2.5% highest and lowest outliers of log(pce) trimmed from graph.

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



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 10^{th} and 90^{th} percentiles of log per capita expenditures (eligible in treatment and control). Fan regressions with bandwidth of 1. 2.5% highest and lowest outliers of log(pce) trimmed from graph.

Table 1: Randomization results

	N	Mean control	Mean treatment	P-value difference
Program info				
Take-up rate (contamination in control)		0.1	95.3	
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 vitamines 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
Acces 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 delay 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				TVIP	McCarthy		BPI
	Social	Language	Fine Motor	Gross Motor	Receptive Language	Memory	Leg Motor	Behavior Problems
				Separate Treatments				
Basic	.136*	.190***	0.099	0.018	.209**	.153*	-0.026	0.101
	(.070)	(.067)	(.070)	(.074)	(.086)	(.091)	(.117)	(.080)
Training	.155**	.176***	0.07	0.017	.186**	0.053	-0.004	0.081
	(.063)	(.066)	(.073)	(.066)	(.082)	(.085)	(.114)	(.073)
Grant	.112*	.135**	0.130*	-0.012	.278***	0.074	0.038	0.017
	(.067)	(.065)	(.076)	(.076)	(.105)	(.076)	(.112)	(.080)
F-test	0.654	(.462)	0.384	0.819	0.535	0.429	0.597	0.489
				Joint Treatment				
Treated	.134**	.166***	0.099	0.007	.223***	0.092	0.001	0.067
	(.059)	(.060)	(.068)	(.066)	(.078)	(.072)	(.109)	(.065)
N	3454	3432	3420	3405	1971	1985	1993	2080

Note: The table reports coefficients and standard errors corrected for clustering at the community level (in parentheses).Age trends have been removed from all outcomes. All outcomes are 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. The F-test reports the p-value of a test of equality of coefficients of the basic, training, and grant packages.

Table 4: Robustness checks

	Denver Test			
	Social	Language	Fine Motor	Gross Motor
Basic specification	.134** (.059)	.166*** (.060)	.099 (.068)	.007 (.066)
Without removing age effects	.122** (.058)	.152*** (.056)	.085 (.067)	.010 (.063)
With age and gender controls, without removing age effects first	.111** (.053)	.136*** (.051)	.082 (.068)	.004 (.059)
Including additional controls	.120** (.057)	.144*** (.052)	.090 (.070)	.015 (.067)
Without dropping outliers	.144** (.056)	.162*** (.058)	.076 (.066)	-.034 (.068)
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)	0.067 (0.065)
Without removing age effects	.166*** (.070)	.052 (.060)	.001 (.093)	.032 (.065)
With age and gender controls, without removing age effects first	.228*** (.085)	.070 (.059)	.018 (.088)	.037 (.065)
Including additional controls	.189*** (.071)	.071 (.066)	.013 (.111)	.071 (.065)
Without dropping outliers	.241*** (.078)	.087 (.069)	.006 (.107)	.074 (.056)
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: See table 3. The table reports coefficients and standard errors corrected for clustering (in parentheses).

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

	Denver Test				TVIP	McCarthy		BPI
	Social	Language	Fine Motor	Gross Motor	Receptive Language	Memory	Leg Motor	Behavior Problems
Heterogeneity by child gender								
T*Boy	.188**	.206**	.125	-.023	.233**	.068	-.073	.058
	(.077)	(.092)	(.083)	(.073)	(.095)	(.095)	(.112)	(.084)
T*Girl	.077	.121**	.072	.042	.219**	.114	.081	.076
	(.068)	(.056)	(.075)	(.088)	(.096)	(.082)	(.134)	(.075)
Boy	-.141*	-.170*	-.090	.082	.060	.005	.111	.001
	(.074)	(.088)	(.072)	(.081)	(.090)	(.089)	(.102)	(.071)
F-test	0.193	0.374	0.520	0.489	0.901	0.659	0.184	0.849
Heterogeneity by child age								
T*AGE1	.079	.056	.028	-.016				
	(.076)	(.060)	(.076)	(.093)				
T*AGE2	.234***	.174	.144	.005	.054	.111	-.032	.172**
	(.088)	(.110)	(.103)	(.109)	(.052)	(.079)	(.131)	(.082)
T*AGE3	.120	.300***	.152	.040	.362***	.077	.031	-.023
	(.105)	(.088)	(.113)	(.066)	(.118)	(.093)	(.107)	(.080)
AGE2	.064	.075	.008	.042				
	(.098)	.095	(.076)	(.105)				
AGE3	.002	.038	.023	.013	.013	.044	.015	.033
	(.109)	(.075)	(.114)	(.077)	(.088)	(.077)	(.079)	(.078)
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	0.049
F-test 3	0.743	0.011	0.340	0.565				
N	3454	3432	3420	3405	1971	1985	1993	2080

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 report the p-values on the following tests: F-test 1 tests equality of the coefficients on (T*AGE1) and (T*AGE2), F-test 2 tests equality of the coefficients on (T*AGE2) and (T*AGE3), and F-test 3 tests equality of the coefficients on (T*AGE1) and (T*AGE3).

Table 6: Treatment effects on intermediate inputs, full sample

Variable	Mean control	Coefficient	s.e.
Hh-level food consumption per capita (logs)			
Total food	8.028	0.310***	(0.04)
Staples	7.214	0.195***	(0.05)
Animal protein	5.488	1.071***	(0.12)
Fruit and vegetables	4.580	1.005***	(0.12)
Child food intake (nr days in last week)			
Tortilla	5.950	0.043	(0.10)
Milk	1.580	1.141***	(0.23)
Meat	0.564	0.764***	(0.08)
Eggs	1.594	1.258***	(0.14)
Fruit	2.552	0.452**	(0.20)
Vegetables	1.468	0.713***	(0.21)
Stimulus			
Got toy in last 6 months	0.271	0.068**	(0.03)
Has pen and paper to draw	0.690	0.101***	(0.03)
Has books	0.073	0.067***	(0.02)
Somebody tells stories	0.520	0.125***	(0.04)
Somebody reads to child	0.080	0.081***	(0.02)
Nr of hours read to	0.134	0.257***	(0.06)
Enrolled in stimulus program or pre-school	0.424	0.065	(0.04)
Early enrollment in primary school	0.096	0.025*	(0.01)
Preventive health care and health status			
Improved health status since last year	0.510	0.102***	(0.03)
Probability of being in bed for illness	0.099	-0.035**	(0.02)
Number of days in bed for illnes	0.610	-0.330**	(0.13)
Consulted doctor if ill	0.730	0.057**	(0.03)
Taken to growth check-up	0.782	0.056***	(0.02)
Weighed in last 6 months	0.705	0.063***	(0.02)
Received vitamin A or iron in last 6 months	0.734	0.086***	(0.02)
Received deworming drugs in last 6 months	0.566	0.066***	(0.03)
Birth weight in kg (0-5 months old)	2.987	0.161	(0.13)
Waz-score	-0.958	-0.052	(0.10)
Whz-score	-0.070	-0.025	(0.06)
Parenting and maternal health care			
HOME scale	3.826	0.035	(0.27)
CESD depression scale	14.786	-1.400*	(0.80)

Note: Standard errors corrected for clustering at the community level (in parentheses). Regressions for household per capita food consumption and parenting and mental health use household as unit of observation (*N*=2270); all other regressions use individual children as unit of observation (*N*=3506). In child-specific regressions, sample includes children between 0 and 83 months, except for birthweight, where sample is restricted to children younger than 6 months (*N*=170). All regressions estimated by OLS.

Table 7: Treatment effects on intermediate inputs, by child age

	T*AGE1	T*AGE2	T*AGE3	F-test 1	F-test 2	F-test 3
Child food intake						
Tortilla	-0.155	0.144	0.202*	0.153	0.667	0.081
	(0.194)	(0.100)	(0.107)			
Milk	0.661**	1.321***	1.578***	0.010	0.279	0.001
	(0.304)	(0.278)	(0.212)			
Meat	0.533***	0.895***	0.939***	0.000	0.647	0.000
	(0.087)	(0.095)	(0.096)			
Eggs	0.878***	1.506***	1.522***	0.006	0.944	0.004
	(0.158)	(0.205)	(0.208)			
Fruit	0.421**	0.364	0.561*	0.834	0.508	0.616
	(0.198)	(0.289)	(0.286)			
Vegetable	0.606***	0.789***	0.781***	0.459	0.976	0.481
	(0.193)	(0.292)	(0.292)			
Stimulation						
Received toy	0.095**	0.048	0.051	0.266	0.926	0.314
	(0.039)	(0.043)	(0.036)			
Told stories	0.091*	0.135***	0.158***	0.308	0.590	0.149
	(0.050)	(0.040)	(0.043)			
Read to	0.019	0.118***	0.128***	0.001	0.739	0.000
	(0.019)	(0.031)	(0.031)			
Hours read to	0.192***	0.353***	0.259***	0.051	0.399	0.482
	(0.072)	(0.086)	(0.090)			
Stimulation program or preschool	0.061	0.068	0.069	0.898	0.987	0.884
	(0.049)	(0.062)	(0.052)			
Primary school	0.000	0.006**	0.079**	0.040	0.061	0.043
	(0.000)	(0.003)	(0.038)			
Preventive health care and health outcomes						
Improved health status	0.050	0.155***	0.123***	0.035	0.474	0.144
	(0.037)	(0.040)	(0.042)			
Being ill	-0.027	-0.033	-0.047*	0.825	0.612	0.482
	(0.018)	(0.022)	(0.026)			
Days ill	-0.186*	-0.413	-0.437*	0.452	0.941	0.310
	(0.111)	(0.273)	(0.237)			
Consulted doctor if ill	0.026	0.046	0.105**	0.642	0.223	0.078
	(0.027)	(0.044)	(0.044)			
Taken to growth check-up	0.001	0.018	0.153***	0.420	0.007	0.001
	(0.010)	(0.022)	(0.044)			
Received vitamin A or iron	0.048*	0.042	0.167***	0.878	0.008	0.012
	(0.028)	(0.028)	(0.038)			
Received deworming	0.025	0.040	0.136***	0.751	0.086	0.012
	(0.033)	(0.040)	(0.037)			
Waz-score	-0.059	-0.147	0.046	0.495	0.094	0.380
	(0.133)	(0.124)	(0.112)			
Whz-score	-0.104	0.000	0.070	0.386	0.484	0.126
	(0.094)	(0.081)	(0.086)			

Note: N=3506. See table 6. All regressions include AGE1, AGE2, and AGE3 as independent variables. The F-tests report the p-value on the following tests: F-test 1 tests equality of the coefficients on (T*AGE1) and (T*AGE2), F-test 2 tests equality of the coefficients on (T*AGE2) and (T*AGE3), and F-test 3 tests equality of the coefficients on (T*AGE1) and (T*AGE3).

Appendix 1: Attrition

Attrition can potentially introduce serious biases into the estimation of program effects. In this study, attrition between the baseline and follow-up surveys was minimal. Only 1.3 percent 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 in our paper. 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.

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
Owns 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
N	4021	3969

Note: Sample includes eligible households with children between the ages of 0 and 83 months at follow-up.

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

	All		Did all tests	
	N	mean	*N*	mean
Child-specific characteristics				
All Children				
Male	3505	0.50	3325	0.50
Age at follow-up	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 diarea	1599	0.75	1494	0.75
Weighed in last 6 months	2488	0.91	2329	0.91
Received vitamin A 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
Owns toilet/latrine	2270	0.72	2197	0.72
Water in house	2270	0.12	2197	0.12
Acces 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: Based on sample of children between the ages of 0 and 83 months at follow-up.

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

	Did all tests		Did not do all tests		P-value difference
Outcome	*N*	Mean	*N*	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: P-value of the difference calculated based on clustering at the community level. Age trends have been removed from all outcomes. All outcomes are 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.

Appendix 2: Details on early childhood development tests

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 that 75 percent of children of their age in the reference population can perform, the test falls back to easier tasks, up to the point where tasks are reached that the child can perform.[25] 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.[26]

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.[27] 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

[25] 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.
[26] 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).
[27] 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).

expected to make if she were randomly guessing.[28] 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 on one foot.[29] 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".[30] 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.[31]

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 in the paper). 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.[32] 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.[33]

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

[28] Before the test starts, the enumerator explains the test with the help of a few example slides. She proceeds to the actual test slides only once the child has demonstrated understanding of the test.
[29] See Stoltzfus et al. (2001), Gertler and Fernald (2004), Fernald, Gertler, and Neufeld (2008), and Cogill et al. (1986) for other applications.
[30] Recent applications of the BPI in Latin America include Fernald, Gertler, and Neufeld (2008); Paxson and Schady (2008).
[31] 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.
[32] 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.
[33] As discussed above, this implies that households with children 7 years or older received an extra amount of cash for school supplies.

was further verified during pre-testing.[34] In the case of the Denver test of language, there also does not appear to be an obvious concern with cultural appropriateness—the test measures whether infants can utter various sounds and, for older children, whether they can identify and name simple concepts, such as body parts. Additional evidence that the Denver is appropriate for our study population is provided by the fact that the national early childhood stimulation program in Nicaragua uses a slightly modified version of this test for child monitoring. For all these reasons, it seems unlikely that the observed delays are due to possible cultural inappropriateness of the tests for children we study.

[34] 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.