Are Conditional Cash Transfers Fulfilling Their Promise? Schooling, Learning, and Earnings After 10 Years

By TANIA BARHAM, KAREN MACOURS, AND JOHN A. MALUCCIO*
March 2017

Abstract:

Interventions aimed at improving the nutrition, health, and education of young children are often motivated by their potential to break the intergenerational transmission of poverty. A prominent example, conditional cash transfers (CCTs), has become the anti-poverty program of choice in many developing countries. Evidence is inconclusive as to whether the demonstrated short-term gains translate into the longer-term educational and labor market benefits needed to fully justify them. This paper uses the randomized phase-in of a 3-year CCT program in Nicaragua to estimate long-term effects. We estimate these effects using experimental variation, complemented by two alternative non-experimental identification strategies. We focus on boys aged 9–12 years at the start of the program who, due to the program's eligibility criteria and prior school dropout patterns, were more likely to have been exposed to the program in the early treatment than in the late treatment group. Previously demonstrated short-term increases in schooling are sustained after 10 years, and there are substantial gains in learning. These improvements in human capital coincide with positive labor market returns—the young men are more likely to engage in wage work, migrate temporarily for better paying jobs, and have higher earnings. In Nicaragua, schooling and learning gains hence translate into earning gains for these young men, implying important long-term returns to CCT programs.

JEL Codes: I25, I38, I28
Key words: CCT, long-term effects, education, learning, labor markets

* Barham: Department of Economics and IBS, University of Colorado at Boulder, Boulder, CO 80309-0256 (tania.barham@colorado.edu); Macours: Paris School of Economics and INRA, 48 Boulevard Jourdan, 75014 Paris, France (karen.macours@psemail.eu); Maluccio, Department of Economics, Middlebury College, Middlebury VT 05753 (maluccio@middlebury.edu). Acknowledgments: This research would not have been possible without the support of Ferdinando Regalia of the Inter-American Development Bank (IDB). We gratefully acknowledge generous financial support from IDB, the Initiative for International Impact Evaluation (3ie: OW2.216), and the National Science Foundation (SES 11239945 and 1123993). We are indebted to Veronica Aguilera, Enoe Moncada, and the survey team from CIERUNIC for excellent data collection and for their dogged persistence in tracking. We also acknowledge members of the *Red de Protección Social* program team (in particular, Leslie Castro, Carold

Herrera, and Mireille Vijil) for discussions regarding this research and Emma Sanchez Monin for facilitating the data collection process. We thank Teresa Molina Millan, Olga Larios, Jana Parsons, and Gisella Kagy for help with data preparation and Vincenzo di Maro for numerous contributions to the project. Finally, we are grateful for comments received from Norbert Schady, Guillermo Cruces and others during presentations at the IDB, Northeast Universities Development Conference 2012, Allied Social Sciences Association 2013, Colby College, Middlebury College, European Development Network Conference 2013, Population Association of America meetings, UCL, Louvain-la-Neuve, Pompeu Fabra, and at the Impact Evaluation Network conference of LACEA. This paper modifies and extends "Schooling, Learning, and Earnings: Effects of a 3-Year Conditional Cash Transfer Program After 10 Years." All remaining errors and omissions are our own.

Interventions aimed at increasing the nutrition, health, and education of children are often motivated by their potential to break the intergenerational transmission of poverty. In recent years, rigorous evidence regarding the short-term effects of such interventions has expanded tremendously, especially in low- and middle-income countries. Surprisingly little, however, is known regarding whether such interventions live up to their full promise, and increase the productive potential of the next generation.[1] The lack of a solid evidence base is due in part to the methodological challenges facing long-term assessment and to evaluations designed to measure short-term effects that are often ill suited for establishing long-term gains, or that may still be too recent.

Consideration of long-term impacts is particularly relevant for conditional cash transfer (CCT) programs. Started in 1997 in Mexico and Brazil, CCTs have spread to more than 40 countries worldwide and are now the main social program in most Latin American countries, where they reach almost one-quarter of the population, or 135 million individuals (Robles, Rubio, and Stampini 2015). The principal program components of most CCTs include regular cash transfers to women—conditional on scheduled visits to health care providers for young children and on school enrollment and regular school attendance for school-age children—and social marketing to encourage investment in nutrition, health, and education. Numerous evaluations, many based on experimental designs, consistently show positive short-term effects. These include poverty alleviation, improved nutrition and health (particularly for young

[1] Important exceptions include evidence on the long-term effects of early childhood stimulation (Gertler et al. 2014), early childhood nutrition (Hoddinott et al. 2008), deworming (Baird et al. 2016), education subsidies and HIV prevention education (Duflo, Dupas, and Kremer 2015) and school vouchers (Bettinger et al. 2016).

children), and increased school attainment (Fiszbein and Schady 2009). But evidence on their longer-term effects is sparse and inconclusive.

We begin to fill this evidence gap by analyzing the long-term effects of a randomized CCT program in Nicaragua on human capital and labor market outcomes, and address a number of methodological shortcomings affecting existing research. We demonstrate that exposure to the CCT at critical ages (when children are likely to drop out of primary school) not only led to sustained educational and learning gains of nearly 0.2 standard deviations (SD) but also to increased returns in the labor market of nearly 15 percent per month worked for young men. Young men are more likely to engage in wage work, migrate temporarily for better paying jobs, and have higher earnings.

The main analysis exploits the randomized phase-in of the program and its eligibility rules to estimate intent-to-treat (ITT) effects after 10 years.[2] Households randomly assigned to the early treatment group were eligible for transfers from 2000 to 2003, while those assigned to late treatment were eligible from 2003 to 2005. In both groups, only households with children between 7 and 13 were eligible for transfers for education (though all households were eligible for transfers for nutrition and health). We identify a cohort of boys aged 9-12 at the start of the program in 2000 for whom the randomized phase-in implied they had greater program exposure at ages critical for schooling in the early treatment group. Specifically, due to the timing of exposure, eligibility rules, and pre-program school dropout patterns, early treatment boys were exposed to the program at ages when it was more likely to prevent school dropout than boys exposed in the late treatment group. While we focus on this cohort because of the sharp differentials in exposure to education transfers and conditions between the early and late

[2] The short-term effects are similar to those found for other CCTs (Maluccio and Flores 2005).

treatment groups, the estimated program effects reflect all components of the CCT program, as is true for most CCT evaluations.

To establish the long-term program effects, we collected data in 2010 on households and individuals originally interviewed prior to the program start in 2000, including an oversample of individuals for whom the exposure differential was the largest. Special effort was made to minimize attrition, given the potential (but a priori unknown) relationship between program exposure, education, and migration for young adults, a particularly mobile age group. Consequently, migrants were tracked throughout the country, as well as to neighboring Costa Rica, the dominant destination for international migrant labor from Nicaragua. Final attrition is low (ranging from 10 to 19 percent depending on the outcome). We correct for remaining sample selection using information collected for that purpose during intensive tracking of respondents, giving more weight to individuals who were more difficult to find (Molina Millan and Macours 2017).

We timed the follow-up survey so that learning and labor market outcomes were measured approximately ten years after the start of the program, and seven years after households eligible for the program in the early treatment group stopped receiving transfers. By that time, when the young men were 19-22 years old, virtually all were in the labor force, and most had completed their schooling.[3] Therefore, the analysis offers a rare opportunity to examine the sustainability of

[3] Similar to other work examining effects of CCTs (Attanasio, Meghir, and Santiago 2012), we do not analyze effects for girls here because the educational and labor market outcomes for girls in this cohort are likely jointly determined with reproductive health decisions. In addition, girls in this context drop out at older ages and fully 30 percent in this cohort were still in school in 2010. Consequently, the same identification strategy does not provide easily interpretable results for the girls.

CCT program effects, as we assess whether individuals who might have received more schooling because of the program perform better on achievement tests and in the labor market, well after the program ended. In fact, the cost-benefit analysis suggests the program could achieve a positive net present value within two decades.

The experimental design described above yields the long-term differential program effects. These estimates, while requiring few assumptions for identification, are likely to underestimate the long-term absolute effects of the CCT. This is because boys in the 9-12 cohort in the late treatment group lived in households that were eligible for the program starting in 2003, and hence may themselves still have benefitted from the program. Therefore, we complement the experimental differential estimates with non-experimental estimates of the absolute effects using two different approaches. First, we estimate absolute effects for all outcomes after 10 years, using matching estimators and an alternative non-experimental comparison group that was added to the evaluation data in 2002 and included in our 2010 survey. Second, we use a double difference estimator to determine absolute program effects on educational outcomes in 2005, using repeated cross sections from the two most recent Nicaraguan national censuses. Results from the two non-experimental approaches, which draw on different data and require different identification assumptions, are broadly consistent. Overall, the non-experimental estimates suggest that the differential program effects are lower bounds of long-term absolute effects.

This paper goes beyond the existing literature addressing a number of the methodological challenges that weaken the ability to draw conclusive inferences from previous longer-term studies on CCTs and related programs. The CCT program for which there is the deepest evidence base, in part because it was the first to have an experimental design, is the Mexican program, *PROGRESA*. As such, our analysis most closely relates to the work of Behrman, Parker, and

Todd (2009a, 2011), who estimate effects 5.5 years after the start of the Mexican CCT. They exploit an 18-month experimental difference in exposure, complemented by non-experimental matching estimates of absolute program effects, for youth aged 15–21 at follow-up. Their results indicate that the program increased schooling but not learning, and that labor market returns are uncertain. The lack of larger significant positive results is somewhat surprising in light of the substantial and well-documented short-term effects of the program. As the authors discuss, however, there are some potential caveats for the interpretation of the findings. First, the difference in exposure of 18 months may have been too short and, because the program was ongoing, children in the late treatment group would have been eligible for up to four years of education transfers by the time the follow-up data was collected. Second, migrants outside the original communities were not tracked, leading to substantial attrition of approximately 40 percent between 1997 and 2003. As a result, labor market returns from migration are not accounted for. Moreover, as a number of pre-program characteristics were correlated with attrition, it is possible that findings are influenced by selection bias.[4] Third, the non-experimental estimates rely on data from localities that, due to the non-random program rollout, were less poor. And fourth, it may have been too early to observe labor market returns for this age group, as many of them were not yet in the labor force.

The Nicaragua program and survey design allows us to address each of these challenges. First, the experimental differential in program exposure is larger and does not diminish (relative

[4] While Behrman, Parker, and Todd (2009a, 2011) use reweighting procedures to control for such selectivity, the lack of migrants in the sample and the overall high level of attrition implies that the estimates are based on the strong assumptions that attrition is driven by selection on observables and that its effects are similar for migrants and non-migrants.

to total time in the program) over the years because the program ended in 2005. Second, migrants were intensively tracked so that attrition is much lower. Indeed, we show that migration patterns are crucial for understanding the labor market returns. Third, the non-experimental evidence uses data from neighboring localities selected for their similar poverty levels but from adjacent non-program municipalities. Fourth, as education levels in Nicaragua are much lower than in Mexico, and consequently labor market integration occurs sooner, we are able to analyze effects when all the young men in the sample are already economically active.

More generally, the literature on the long-term effects of CCTs is limited by the fact that few programs were initially set up for rigorous long-term evaluation, and often control groups are phased-in.[5] We circumvent this challenge by focusing on cohorts with differential exposure at critical ages, and by examining a program in which early and late treatment households benefitted for a similar period, allowing us to measure differential effects after the intervention ended. Existing studies relying on experimental control groups that have never been phased in are only able to present evidence two years after the end of their respective programs (Baird, McIntosh, and Özler 2016; Macours and Vakis 2016;), with the exception of the education-only CCT examined by Barrera-Osorio, Linden, and Saavedra (2015).

Some non-experimental studies are able to examine effects over longer periods before phase-in of the comparison group (García et al. 2012; Parker, Rubalcava, and Teruel 2012), but require stronger assumptions. Identification assumptions are weaker for studies using regression

[5] See Molina Millan et al. (2016) for further discussion of the literature on long-term effects of CCTs in Latin America. Related research on long-term effects of cash transfers in the U.S. is based mostly on non-experimental variation (Akee et al. 2010; Aizer et al. 2016).

discontinuity designs, but they estimate local average treatment effects (LATE) for the least poor beneficiaries, and such impacts may differ from average treatment effects (Filmer and Schady 2014; Araujo, Bosch, and Schady, 2016). Our results indicate that, at least in Nicaragua, effects were much larger for the poorest beneficiaries.

Further, as described above for Mexico, an important source of concern in many studies is sample attrition, which is often directly related to migration, and therefore likely related to labor market returns. Studies using secondary administrative data often suffer from related selection concerns when multiple data sources are combined but can only be imperfectly matched (Baez and Camacho 2011). Other such secondary data studies have only a limited set of measured outcomes (Barrera-Osorio, Linden, and Saavedra 2015). Finally, most of the existing longer-term studies have focused on beneficiaries that are still transitioning into the labor market, complicating interpretation of labor market impacts due to the potential tradeoff between additional schooling and shorter work experience.

This paper complements the body of research examining long-term effects of CCT programs on outcomes beyond learning and labor market outcomes. Gertler, Martinez, and Rubio-Codina (2012) show effects on household investment in economic activities, while several other studies analyze longer-term effects of exposure during early childhood (Behrman, Parker, and Todd 2009b; Fernald, Gertler, and Neufeld 2009; Macours, Schady, and Vakis 2012; Barham, Macours, and Maluccio 2013; and Araujo, Bosch, and Schady 2016).

In addition to those already cited, this paper relates to other recent studies analyzing long-term effects of randomized anti-poverty interventions such as the Ultra-poor transfer programs in Bangladesh (Bandiera et al. 2016) and India (Banerjee et al. 2016). It further relates to the growing literature on longer-term effects of youth training programs in Latin America (Attanasio et al. 2015; Kugler et al. 2015), one of which also draws on long-term surveys including tracking components (Ibarran et al. 2015).

Finally, our long-term evidence of the effect of CCTs on learning adds a different perspective to the existing short-term evidence. Filmer and Schady (2014) find no short-term effects of CCTs on tests of mathematics and language in Cambodia. Baird, McIntosh, and Özler (2011) show significant, but small effects on tests scores in Malawi.

Taking the learning and earning results together, this paper has important policy implications. Indeed, the lack of conclusive evidence on the long-term effects of CCTs has led some to conclude that the potential for CCTs to reduce the intergenerational transmission of poverty may be limited (Levy and Schady 2013). The evidence we present, for a program implemented by the government of Nicaragua and closely following the original Mexican CCT program design, arguably means there is room for more optimism.

II. The Nicaraguan CCT Program and Its Experimental Design[6]

A. Key program design features of the CCT

Modelled in part after *PROGRESA*, the *Red de Protección Social* was a CCT designed to address both current and future poverty through cash transfers targeted to poor households in rural Nicaragua. The program was implemented by the government of Nicaragua with technical

[6] This section and appendix C draw on the details provided in Maluccio and Flores (2005).

assistance and financial support from the Inter-American Development Bank (IDB) and benefited more than 30,000 families. On average, transfers were 18 percent of pre-program expenditures and delivered every other month. They were paid to a designated female caregiver in the beneficiary household and came with a strong social marketing message that the money was intended to be used for food and education investments.[7]

The CCT had two core components. The first component focuses on education. Households with children aged 7–13 years who had not yet completed the fourth grade of primary school were eligible for the education components of the program. They received an additional fixed bimonthly cash transfer known as the school attendance transfer, which was contingent on enrollment and regular school attendance of those children. For each eligible child, the household also received an annual cash transfer at the start of the school year, intended for school supplies, conditional on enrollment. We refer to the combined schooling attendance and school supplies transfers as the education transfer. Teachers were required to report enrollment and attendance using forms specifically designed by the program for the verification of the conditions and received a small per-student cash transfer, known as the "teacher transfer".

The second component aimed at improving nutrition and health, and all households were eligible for a transfer of fixed amount per household regardless of the household's size and composition. This nutrition and health transfer was conditional on preventive healthcare visits for any children under age five in the households, and on the household representative attending bimonthly health information workshops. More detailed information on the program is provided in the appendix C.

---

[7] A small percentage (<3%) of households were excluded ex ante as ineligible based on an asset exclusion criteria and are not included in our analyses (Maluccio and Flores 2005).

While the Nicaraguan CCT was modeled after Mexico's original CCT program, there are two differences between the programs that are important for our analyses. First, in Nicaragua beneficiary households were only eligible to receive the program for a fixed period of three years, after which it was impossible to renew eligibility. Second, the education conditionalities and transfers in the Nicaraguan CCT only applied to the first four grades of primary school, reflecting the low levels of education and high primary school dropout rates in these impoverished rural areas.

B. Experimental Design of the CCT

The randomized evaluation was built into the design of the CCT intervention in six rural municipalities in central and northern Nicaragua, chosen based on their low education and health indicators. In these six municipalities, 42 out of 59 rural *comarcas* (hereafter, localities) were selected based on a marginality index.[8] A program census of all residents in these 42 localities was collected in May 2000.

The 42 targeted localities were then randomized into one of two equally sized treatment groups, the early or late group, at a public lottery. To improve the likelihood that the selection of localities in the experimental groups would be well balanced in terms of poverty levels, the marginality index was used to classify the 42 localities into seven strata of six localities each.

---

[8] Census *comarcas* are administrative areas within municipalities based on the 1995 Nicaraguan national census that included as many as 10 small communities for a total of approximately 250 households. The marginality index was constructed from this national census and included indicators of literacy, family size, and water and sanitation conditions. See appendix C for further details.

From each stratum three localities were randomly selected as early treatment and three as late treatment.

The randomization occurred in July 2000, after the program census, and the 21 early treatment localities received their first transfers in November 2000. They were eligible to receive three years' worth of cash transfers and received the last transfer in late 2003. Households in the late treatment localities were informed that the program would start in their localities later. The 21 late treatment localities were phased-in at the beginning of 2003. They were also eligible to receive three years' worth of cash transfers. Households in the early treatment group did not receive any transfers after 2003, and therefore were not affected by any conditionalities after that date. The small teacher transfer, however, continued. At the end of 2005, all program benefits were discontinued for both groups and the program no longer operated in these municipalities.

Overall, compliance with the experimental design was high. Past analysis on the program shows that the sample was balanced at baseline and that there was little contamination of the late treatment group (Maluccio and Flores 2005). Appendix A shows balance for the main samples used in this paper. At the household level, program take-up in early and late treatment localities was approximately 85 percent. At the individual level, take-up of the education transfer for at least one year was 88 percent for 9–12-year-old boys in the early treatment group.

III. Data

We draw on a number of different data sources for the long-term analysis.

Program Census Data—The 2000 program census provides baseline data on early and late treatment localities including grades attained for all household members, household demographics, housing characteristics, and assets. We use the census data to draw an oversample of children of critical ages for the 2010 survey, and as the baseline survey for the experimental analysis.

Short-term Evaluation Surveys—The first round of the short-term evaluation surveys was conducted in September 2000, with subsequent rounds in 2002 and 2004. The sample was drawn from the program census roster and includes a random sample of early and late treatment household, 42 households in each of the 42 early and late treatment localities (Maluccio and Flores 2005). Attrition was approximately 10 percent per round. Starting in 2002, a non-experimental comparison group (that never received the program) was added from neighboring municipalities, chosen from a set of localities with marginality scores similar to the original 42 localities. The 2002 household survey targeted for interview 40 households in each of the 21 localities, and, since there is no program census data for this comparison group, provides the baseline outcomes for this sample. The household-level survey instrument is based on the 1998 Nicaraguan Living Standards Measurement Survey, with modules covering education, health, and household expenditures, among others.

2010 Evaluation Survey—Between November 2009 and November 2011, 9 to 11 years after the start of the program, we conducted a long-term follow-up evaluation survey.
For convenience, and since the bulk of data were collected in 2010, we refer to this as the 2010 survey. We expanded the short-term household evaluation survey instrument, and added a separate, individual-level instrument with measures of learning and socio-emotional outcomes. The sample frame consists of all households in the original short-term randomized evaluation survey sample including the non-experimental comparison group, as well as an additional sample of households who, according to the 2000 program census, had children of critical ages relevant

to the long-term evaluation. Specifically, we oversampled households with children born between January and June 1989 in the early and late treatment groups who were at risk of dropping out of school at the start of the program (see section IV for further details).[9] All experimental estimates account for this sampling procedure, using sampling weights, and also adjust for attrition as described in appendix F. While the detailed information from the earlier short-term evaluation surveys is not available for the oversampled households, there is full information on them from the 2000 program census. The sample frame has a total of 1,330 households from the early treatment group, 1,379 households from the late treatment group, and another 757 households from the non-experimental comparison group in 2010.[10] For the 9–12 year old boys, the sample frame included 588 observations in the early treatment group, 550 in the late treatment group, and 263 in the non-experimental comparison group.

The 2010 survey was designed to collect data on households in the target sample, as well as all split-off households containing at least one of the original household members less than 22 years old in 2010. The household instrument included questions on educational attainment and current schooling, and a detailed survey instrument module to measure participation and earnings in all economic activities each household member had engaged in over the last 12 months. It asked separately about economic activities conducted at the place of residence and activities engaged in while temporarily absent from the household. For all young adults between 15 and

[9] We also oversampled households with children born between November 2000 and mid-May 2001, for analyses on exposure to the CCT during early childhood (Barham, Macours, and Maluccio 2013).

[10] The 2002 household survey targeted 840 households in the comparison group, of which 823 were interviewed. Of those, 32 would have been ineligible for the program based on asset exclusion criteria. An additional 34 lived in border areas and ultimately received the program, so the target sample in 2010 was 757 households.

22, an additional module also collected their full labor market history, with questions on participation, location, and earnings for all non-agricultural wage jobs and self-employment. All questions in the household instrument were answered by the best-informed person available for the interview. Hence, answers were obtained from the young men themselves if they could by located at home, or from the household head or the spouse of the household head if not.

The 2010 individual instrument was conducted through direct in-person interviews with the respondents in their homes, and was designed to measure individual learning and socio-emotional outcomes of each individual born after January 1, 1988. We administered three Spanish language and two math tests to all young adults between the ages of 15 and 22 years (in 2010). The Spanish language tests included a word identification test, a spelling test, and a test of reading fluency. The math tests included a test of math fluency and a test to measure problem solving at various levels of difficulty, which we refer to as math problems and is similar to the *Peabody Individual Achievement Test* (Markwardt 1989). All tests were appropriate for different grade levels. In addition, we administered two tests that could capture both learning and cognitive development: the *Test de Vocabulario en Imágenes Peabody* (TVIP, the Spanish version of the *Peabody Picture Vocabulary Test*; Dunn et al. 1986), and a forward and backward digit span test, in which the respondent is asked to repeat a series of numbers read to him. Finally, the Raven's colored matrices (the 36-item version with sets A, AB, and B) was added to measure cognition (Raven, Court, and Raven 1984). Given that the intervention began when the cohort of boys was in late childhood, we did not necessarily expect large sustained program effects on cognition, and therefore included the Raven in the instrument to help separate more general cognitive skills from skills acquired in the classroom. An important advantage of all of the tests is that they provide observed, as opposed to self-reported, measures of learning and cognition, therein substantially reducing concerns about reporting biases. Finally, the individual survey instrument also included measures of the socio-emotional outcomes of the young men. In particular, we implemented the 20-item Center for Epidemiologic Studies of Depression (CESD) Scale and the Strengths and Difficulties Questionnaire (SDQ).[11]

Administrative Data—The household- and individual-level survey data is complemented by the CCT's administrative data from 2000 to 2005, originally obtained for registering beneficiaries, monitoring conditionalities, and determining transfers. It contains detailed information on school enrollment and attendance for both early and late treatment groups, as well as information on transfers and household eligibility for the education transfer. This data is used to examine the actual transfers received by individual boys and their consistency with the experimental design and with the program rules, which helps isolate the age groups for which we can identify clear differential program impacts on education.

Nicaraguan Population Census Data—We complement the experimental analysis for education and literacy outcomes using the two most recent Nicaraguan population censuses. The censuses provide repeated cross sections at the individual level, in 1995 before the start of the program and in 2005, the year the program ended.

IV. Methodology

A. Identification of Experimental Long-term Differential Impacts

[11] See appendix E for more information on the learning, cognitive, and socio-emotional measures.

To estimate the long-term effects of the CCT, we take advantage of the exogenous variation in treatment assignment provided by the randomized phase-in of the program and focus the analysis on a cohort of boys whose educational attainment was more likely to have been affected by the program in the early treatment group than the late treatment group. Education transfers were provided and conditionalities applied to the early treatment group from November 2000 through October 2003, and to the late treatment group from January 2003 to the end of 2005. To determine which age groups stood to benefit most from early treatment, we consider pre-program dropout patterns, as well as program rules.

First, we examine the ages at which boys commonly drop out of school in rural Nicaragua. Arguably, transfers provided at ages while most children are in school, or after most children have already left school, will be less effective at improving enrollment than transfers provided at ages when children are at higher risk of dropping out (de Janvry and Sadoulet 2006).[12] Average pre-program enrollment rates by age for boys in the program census are shown in Figure 1 (dashed line, left vertical-axis scale). Enrollment peaks around ages 10 and 11, after which it declines indicating increasing risk of dropout for boys beginning at age 11. Consequently, boys 9–12 years old in 2000 were at high risk of dropping out at some point between 2000 and 2003.[13] As 9–12 year old boys in the early treatment group were exposed to program benefits during this period, we hypothesize that they remained in school longer. On the other hand, boys in the same

[12] Children who would have been in school even in the absence of the transfer may still have been influenced in other ways, for example due to the requirement on number of days of attendance and the increase in household resources, some of which were designated for school uniforms and materials.

[13] We calculate ages on 1 November 2000, the start of the program, which we define when transfers began in the early treatment group.

age bracket from the late treatment group would have been 12–15 by the time their households became eligible for the program. By then many already would have dropped out, making it likely that the program would affect their schooling less.[14]

Second, we consider the program rules. Children between ages 7 and 13 at the start of the school year in January were eligible for the education transfer for that year, provided they had not completed fourth grade. Abstracting from the grade requirement, we calculate the potential differential in years of schooling affected by the education transfer for two boys the same age, one living in an early treatment locality and the other in a late treatment locality. If both were eleven years old at the start of the program, the early treatment boy would be eligible over (part of) four distinct school years (2000–2003).[15] The late treatment boy would be eligible for only one school year (2003) since he would be 13 in 2003.[16] Consequently, in addition to exposure at different ages, the early treatment boy was exposed during three additional school years compared to his age-mate in the late treatment group. Similar comparisons make clear that the difference in exposure to the education transfer was three school years for 11- and 12-year olds at the start of the program, two years for 10- and 13-year olds, and so on.[17]

---

[14] Enrollment patterns for girls are higher on average and decline sharply only after age 13. In part because of this distinct pattern, we do not analyze girls in this paper.

[15] The oversample in the 2010 survey is of individuals who were 11 years old at the start of the program.

[16] The announcement of the program in July 2000 means that although the transfers did not begin until late in the school year 2000, the program still had strong potential to influence schooling in that year.

[17] Exposure in this section refers specifically to child-specific exposure to the education components of the CCT program and not exposure to the other program components that are independent of the education eligibility rules.

Reincorporating the grade requirement component of eligibility into the calculation, as older children were more likely to have completed grade four, measures of potential maximum school years affected by the education transfer based on age eligibility alone do not determine the actual number of school years affected. Because some fraction of children reached the maximum grade of eligibility prior to the end of the program in their localities, the actual exposure differences are lower, on average, than the potential exposure differences described above.[18] Figure 1 (black line, right vertical-axis scale) shows the difference between the average number of school years during which boys actually received transfers in the early treatment relative to boys in the late treatment, based on program administrative data. Actual exposure differences are largest, approximately 2 school years, for 10-year-old boys and consistently above 1.5 years between ages 8 and 12.

Based on this examination of risk of dropout and exposure, we focus the analysis of the long-term effects on boys age 9–12 years. This is the age group for which we expect a large differential impact as a result of being randomly allocated into early versus late treatment and the risk of dropout being the highest.[19]

Analysis of the short-term intent-to-treat (ITT) evaluation results further supports the choice of the 9–12-year old cohort for estimating long-term differential impacts (Table 1). In 2002 (top panel)—when the early treatment group had received transfers for two years but the late

---

[18] Actual differences also may be lower than potential because take-up was less than 100 percent.

[19] Although the length of exposure for 8 year olds is also high, we do not include them because their risk of dropout is low during the program years in the early treatment group and the potential differential exposure is lower. We further do not include the 13 year olds because many of them had already dropped out of school or completed fourth grade by program start, and the maximum potential differential is only one year.

treatment group had not yet benefited—boys in the 9–12 year old age cohort had 0.36 additional grades attained, were 18 percentage points more likely to be enrolled in school, 36 percentage points more likely to attend school regularly (more than 85 percent of the time), and were 15% more likely to be literate (i.e., able to read and write according to parental report).[20] By 2004 (bottom panel), the transfers to the early treatment had ended and the late treatment group had started receiving transfers. In line with this change, the sign of the effects on enrollment and attendance in 2004 reverses, and differences between early and late treatment boys are no longer significant. Even so, the difference in grades attained between early and late treatment remains positive and significant, and at 0.49 grades is larger than in 2002.

The short-term results hence demonstrate that being exposed to the CCT during critical ages in primary school led to persistent differences between the two experimental groups by 2004, even after the early treatment group no longer received transfers. At the same time, as enrollment was higher in the late treatment group (though not significantly so), and because the late treatment group was eligible for another year of program benefits after 2004, the results also suggest that estimates of the long-term differential effects of early versus late treatment may underestimate the absolute program effects.[21] We therefore complement the differential

[20] These results build on the detailed evidence of the effect of the CCT program on education (Maluccio and Flores 2005; Maluccio, Murphy, and Regalia 2010), but focus on the specific age cohort relevant for the analyses of the long-term program effects.

[21] A second reason the experimental estimates may underestimate absolute treatment effects could be positive spillovers between the early and late treatment groups. While randomization was done at the locality level, some of the localities share borders so that spillovers are theoretically possible. Appendix G demonstrates there is little systematic evidence of such spillovers.

experimental estimates with non-experimental matching and double difference estimates of the absolute program effects as described below.

B. Empirical Specification for Long-term Differential Impacts

We estimate the following individual-level model,

$$Y_{il}^{k} = \alpha^{k} T_{l} + \boldsymbol{\beta^{k} X_{il}} + \varepsilon_{il} \; k = 1 \ldots K, \qquad (1)$$

where Y^k is one of the outcomes of interest for individual i in baseline locality l. T is an ITT indicator that takes on the value of one for boys in localities randomly assigned to early treatment and zero for those in localities randomly assigned to late treatment. All analyses are carried out on an ITT basis and using all boys from both treatment groups in the 9–12-year-old age cohort, regardless of initial levels of completed schooling or actual program participation. Given randomized assignment, our main specifications limit the set of control variables **X** to the following: age when the program started in early treatment (dummy indicators for 3-month age groups); completed grades attained prior to the program (dummy indicators for 0, 1, 2, 3, or 4+ grades); and strata and regional fixed effects. Standard errors are adjusted for clustering at the locality level and all regressions are weighted to account for sampling and attrition providing population estimates (Section IV.C). We assess robustness of the above specification with alternative sets of controls and weights, as well as by limiting the sample to only the oversampled 11-year olds, in appendix B. These alternative specifications do not substantively alter the principal findings.

C. Outcomes

CCTs can influence a wide range of behaviors and we therefore analyze a large set of outcomes, all measured in 2010 when the respondents are young men approximately 19–22 years old. To reduce concerns regarding multiple hypotheses testing we follow Kling, Liebman, and Katz (2007) and organize individual outcomes into different families capturing schooling, learning, labor market, and socio-emotional outcomes. We calculate within-sample z-scores for each outcome, using the mean and SD of the late treatment group.[22] We then obtain the average z-score for each family and estimate the ITT effect using this index, which yields the effect size in standard deviations.

The education family includes an indicator of whether the boy was enrolled in school, grades attained, and whether he had completed grade 4, after which children were no longer subject to the program's schooling-related conditionalities. To analyze learning, we classify the tests into three categories. The first measures skills most likely learned in the classroom and comprises the average impact of the five achievement tests (word identification, spelling, reading and math fluency, and math problems). The second averages tests that are likely to capture both learning and cognition (receptive vocabulary and memory test). The third has only one test to proxy for cognition, the Raven's colored matrices. We refer to the three categories as the learning family, the learning and cognition family, and the cognition family, respectively.

---

[22] The late treatment group is used for standardization because individuals in this group were less likely to have been affected by the program. Results are nearly identical when we use the mean and SD of both groups.

For the labor market we consider two families of outcomes. The labor market participation family captures labor market participation and temporary migration. The earnings family includes labor market returns or earnings for work that is off the family farm. We present two versions of the earnings family to account for outliers. One uses absolute values of earnings trimmed at the top 5 percent of values, and the other uses the rank of earnings instead of the actual values. Specific variables included in the earnings family are described in further detail in Section V.A.

For the socio-emotional outcomes, we conducted an exploratory factor analysis including all items of the CESD and the SDQ. This analysis points to four latent factors, broadly capturing optimism, positive self-evaluation, stress, and negative self-evaluation (see appendix E for details). The family outcome is measured as the average of the z-scores for the four factors, after reversing the signs that have opposite meaning (stress and negative self-perception) so that higher values always indicate more positive, or better, socio-emotional outcomes.

D. Attrition and Internal Validity

Considerable effort and resources went into minimizing attrition in the 10-year follow-up for both the individual- and household-level instruments. Individuals who could not be found in their original locations were tracked to their new locations throughout Nicaragua. Migrants to Costa Rica—the destination of 95 percent of international migrants from the sample—also were tracked. As migration in this context is often temporary, multiple return visits by the survey team to the original localities were made to interview temporary migrants after they returned. For the experimental sample, only 10 percent of the specific cohort of boys examined in this paper could not be tracked and hence has missing information for outcomes coming from the household instrument, which include the schooling, migration, and labor market outcomes.[23] Attrition is higher, 19 percent, for information obtained through the individual instrument, which required direct, in-person interviews and included all tests on learning, cognitive, and socio-emotional outcomes. Overall, these attrition levels are lower than those found in related studies with a similar time horizon and target population.

There are no significant differences in attrition levels between early and late treatment groups and attrition does not affect the balance of baseline observables (appendix F). However, attrition is correlated with baseline characteristics that are associated with migration patterns, and these correlations differ to some extent between early and late treatment groups. We account for this selection by weighting all results. Specifically, we use inverse probability weighting, allowing for differences between early and late treatment groups and incorporating information from the survey tracking process to give higher weight to individuals who were more difficult to find. The basic assumption underlying this strategy is that those not found are more similar on both observables and unobservables to those who were harder to find than to those more easily found (Molina Millan and Macours 2017). Appendix F provides further information on tracking and discusses the construction of the attrition weights, which also incorporate the sampling weights.[24]

---

[23] While 93 percent of original targeted households were surveyed, many young men no longer lived in their original locations and attrition was slightly higher for split-off households. In addition to those who could not be tracked, 15 were deceased by 2010.

[24] Results without attrition weights are shown in appendix B. Although they do not qualitatively alter the principal findings, changes in point estimates underscore the relevance of accounting for attrition.

E. Identification of Non-Experimental Long-term Absolute Effects

Matching estimators using the non-experimental comparison group—To obtain an estimate of the absolute effects on all outcomes, we take advantage of a non-experimental comparison group sample drawn in 2002 and resurveyed in 2010. Since this group never received the CCT, we use it as a counterfactual for the early treatment group and compare 2010 outcomes of 9–12 year olds between the two groups.[25] While the non-experimental comparison localities were selected based on having similar marginality indices to the treatment localities, and many share a geographical border, the ex ante match on locality-level characteristics did not completely balance household and individual characteristics across the groups. To improve the balance on observable characteristics, we estimate propensity scores at the individual-level and use the five nearest-neighbor matching estimator to estimate the absolute treatment effects. As the non-experimental estimates are based on much stronger assumptions than the experimental estimates, we present these results mainly as complementary evidence, and to facilitate interpretation of the differential effects.

The nearest neighbor estimator is bias corrected and standard errors are clustered at the locality level using the analytic asymptotic variance estimator developed by Abadie and Imbens (2008, 2011) that accounts for the fact that the propensity score is estimated. We use a typical "min-max" criteria to define the common support. Appendix H provides further details on the matching procedures, as well as a number of robustness tests, including specifications with

[25] Attrition for this non-experimental sample is 15 percent for the variables measured in the household instrument and 29 percent for variables from the individual instrument. There are 223 boys in the comparison sample with education and labor market outcomes.

alternative definitions of the common support, two nearest neighbors matching, and non-parametric kernel and local linear matching estimators.

Double difference estimation using 1995 and 2005 Nicaraguan National Censuses—Over the years 2000 to 2005, the CCT operated at some point in all rural areas of the six municipalities where the randomized evaluation was implemented, covering over 90 percent of the rural population in these municipalities. Given this high coverage, it is possible to use census data to approximate absolute program effects on schooling.

Therefore, a second non-experimental approach to estimate absolute program effects on educational outcomes uses a double difference model and the two most recent Nicaraguan censuses. Together, the censuses provide repeated cross sections at the individual level, in 1995 before the start of the program and in 2005, the year the program ended. The data include current municipality of residence (and whether urban or rural), as well as municipality of residence 5 years prior to the census administration date. We calculate double difference impacts using two cohorts of boys (those ages 9–12 in 1990 and in 2000) by comparing education outcomes in rural areas of the six program municipalities to outcomes in rural areas of the six neighboring municipalities used in the matching. The 9–12 age cohort in 2000 is, of course, the same age cohort examined in the experimental analyses. As the underlying identification assumptions for the double difference estimates are not the same as those required for the matching estimators described above, they provide independent complementary evidence. More specifically, we estimate

$$Y_{imt} = \delta_0 + \delta_1 \, T_{m,t-5} + \delta_2 \, C_t + \delta_3 \, T_{m,t-5} * C_t + \varepsilon_{imt} \qquad (2)$$

Where Y_{imt} is the educational outcome for boy *i* in municipality *m* measured in census year *t*, $T_{m,t-5}$ is an indicator for whether the boy resided in a treatment municipality 5 years prior to the census year, and C_t is one if the child is in the 2005 census. δ_3 yields the double difference estimate of the 5-year effect of the program on *Y*, which includes grades attained, enrollment, and literacy. Standard errors are robust to heteroskedasticity. We examine alternative comparison groups and assess common trends in appendix I.

V. Results

A. Long-term Differential Impacts

In this section, we present the differential experimental ITT effects using equation (1). We present results for the family of outcomes as well as the components of these families. All tables include the mean of the late treatment group (and these are what are reported in the text below) for outcomes that are not in standard deviations. Robustness tests with alternative specifications and samples described in the methodology section are shown in appendix B and do not qualitatively alter the principal findings.

Schooling and Learning—Ten years after the start of the program in the early treatment group, schooling outcomes are significantly better in early compared to late treatment (column 1). Table 2 presents the components of the education family; there is a significant differential effect on grades attained of nearly 0.3 grades (column 2), on an average of 5.5 grades attained. The significant differential effect estimated in 2004 (shown earlier in Table 1), largely persisted through 2010. Early treatment boys are also 4.5 percentage points (column 4) more likely to still be in school in 2010. Overall, a substantial minority, 18 percent in the late treatment group, is still in school, typically taking weekend classes.[26] This pattern helps explain why average grades attained for the cohort increased substantially since 2004. Despite these gains, however, the overall level of schooling remains low, with only 65 percent having completed grade 4.

The results for schooling demonstrate that boys exposed to the CCT during critical ages not only have higher grades attained but also are more likely to be continuing their studies, suggesting the grades attained differential is unlikely to diminish, and may increase further over time. We build on this result in what follows, investigating the extent to which these differences in schooling are accompanied by higher learning and better labor market outcomes. In Table 3, we see that the improvements in schooling translate into learning gains. Early treatment boys performed better than those in the late treatment on math, Spanish, and combined math and Spanish. Moreover, the difference is significant for each of the five individual tests administered (Table 4). This result is corroborated by a significant increase in self-reported literacy (column 5, Table 2). Table 3 column 1 shows that on average achievement test scores were 0.18 SD higher in the early treatment group, 10 years after the start of the program. Significant differential impacts are seen for the skills one would reasonably expect to be acquired in school: 0.16 SD for math (column 2) and 0.20 SD for Spanish reading and writing (column 3), but not for the more general cognitive outcomes. In particular, the impact on the Raven in column 5, a test less likely to capture skills learned in the classroom, is close to zero (- 0.02 SD). The mixed learning and cognition family that could capture both classroom and non-classroom skills (tests on receptive

[26] Virtually all of those studying also work; see results on labor markets below.

vocabulary and memory) show smaller and insignificant differentials in column 4 (Table 3), and this holds for both individual tests (columns 6 and 7, Table 4).[27]

Hence in Nicaragua, exposure to the CCT during critical ages in primary school led to significant learning. Moreover, the magnitude of the differential effects, which are in line with short-term absolute results from education interventions in other settings, suggest the gains are non-negligible.

Labor market outcomes—Program effects on labor market participation and outcomes are presented in Tables 5 and 6. Figure 2 shows that the off-farm monthly earnings exhibit a clear shift to the right for the early versus late treatment group. To contextualize this finding, we begin with a few key facts about the labor market experiences of the cohort of interest. First virtually all individuals (98%) in the cohort (who in 2010 are young men 19–22 years old) are working, so that labor market effects cannot come from increased participation on the extensive margin. In fact, most young men combine different types of work—work on the family farm (89 percent) with work off the family farm (83 percent), or what we more simply refer to in the tables as worked off-farm. In addition, as noted before, 18 percent are still studying. Exposure to CCTs at critical ages in primary school increases participation in off-farm work by 6 percentage points (Table 5, column 2). Given the low levels of schooling (less than 6 grades attained on average) it should not be surprising that almost all off-farm work is in non-skilled jobs, which is also where the increase is concentrated.

---

[27] We use the combined score of the forward and backward digit span. When estimating the differential ITT on each component separately, estimates of the forward digit span (measuring math memory) are in line with the TVIP and significant, while the backward digit span (often considered a measure of executive functioning) is not significant at all and similar to the Raven.

Off-farm work includes work as agricultural laborers for farms not belonging to the household or large plantations, salaried jobs in the non-farm sector (such as construction workers or security guards), and non-agricultural self-employment. Opportunities for remunerative off-farm work in such jobs are highly limited in the poor rural communities where the CCT operated. Table 5 therefore presents the differential impact on seasonal migration for work. Boys in the early treatment are 9 percentage points more likely to have migrated temporarily for work in the last 12 months (column 3), nearly one-third higher than the average 31 percent.[28] Similarly, columns 4 and 5 highlight that young men in early treatment are 8 percentage points more likely to have held a non-agricultural wage job at some point, and 7 percentage points more likely to have held an urban wage job (a 50 percent increase over the mean).[29] Taken together as a family capturing labor market participation, these different indicators point to a substantial and significant 0.24 SD shift in the type of economic activities these young men engage in and an increased degree of mobility (column 1). On the other hand, the program effect on permanent migration out of the municipality, in column 6, is close to zero.

In Table 6, we explore whether the changes in labor market participation and seasonal migration are accompanied by differential effects on earnings.[30] Earnings measures include

[28] This is not due to migration starting earlier, as there were no significant differences between early and late treatment group in work or migration patterns for this cohort in 2004.

[29] While total months worked off-farm did not change (and in fact the point estimate is negative), time worked during seasonal migration increases (from 30 to 40 days on average) while the time worked off-farm in the village of origin decreases.

[30] We exclude earnings on the family farm, as person-specific individual returns from work on the family farm cannot be quantified in this setting. Only 11 percent of the cohort are head of their own household and nearly all of

average and maximum monthly earnings, as well as total earnings in the last 12 months (columns 2–4). The final column shows the monthly salary in the highest paid salary job ever held, including prior to the previous 12 months.[31] To avoid selectivity, all values are unconditional with zero earnings when an individual did not have off-farm (or salary) earnings during the reference period.[32] As the earnings data are highly skewed, we trim the 5 percent highest outliers in panel A. Alternatively, we follow Athey and Imbens (2016) and estimate the effect on the ranks of each of the indicators in panel B. Results are broadly consistent across the different indicators and methods, with the earnings family outcome in column 1 showing an overall increase of about 0.2 SD for both the absolute value as well as the rank. The point estimate for average monthly earnings in column 2 indicates a differential increase of 14 percent, with estimates of the other indicators showing differential increases between 7 and 62 percent. Finally, Table 7 shows the results of a quantile regression for the earnings family outcome. The results suggest positive gains across the distribution, with point estimates that tend to increase for higher deciles. Estimates are mostly significant between the 40th and the 80th percentiles, and gradually increase from 0.15 SD to 0.3 SD in this range.

---

the others do not have an independent farm but rather work on the farm of an older household member, typically their father or father-in-law. Finally, we do not have information on agricultural inputs, preventing us from calculating agricultural profits for these household farms.

[31] This information comes from the separate section capturing the full history of labor market experiences, excluding agricultural wage work, and is not limited to the previous 12-month reference period.

[32] As relatively few men have ever had a salaried job, the unconditional mean of the monthly salary is much lower than for the off-farm earnings in the last 12 months.

Differential exposure to the CCT during primary school hence led to non-negligible increased returns in the labor market for young men in Nicaragua. Moreover, the results are likely underestimates of the long-term annual returns for three reasons: 1) they capture only the differential effects; 2) more of the young men in the early treatment group are still in school; and 3) these returns are measured very early in the working life cycle.

Socio-emotional outcomes—Does exposure to CCTs during critical ages in primary school affect socio-emotional outcomes? The causal pathway for such outcomes is potentially quite complex, as cash transfers during childhood could affect them both directly and indirectly. A positive shock during a critical period in childhood when personality traits may be forming could permanently affect socio-emotional outcomes. Additionally, changes in schooling, learning, or labor market outcomes also could affect these measures. Table 8, column 1, shows that on average, there is a small (0.05 SD) but insignificant differential improvement in socio-emotional outcomes.

Considering each of the four factors separately, however, reveals that this average finding masks offsetting effects on the different latent traits. The differential effects on positive self-evaluation and optimism are positive, significant, and relatively large at around 0.25 (columns 2 and 3), possibly reflecting higher learning and earnings for the early treatment group. At the same time, however, the early treatment group is also more stressed, and more likely to agree with statements referring to negative self-evaluation, even if the point estimates in columns 4 and 5 for these outcomes are smaller, around 0.16 SD. A plausible explanation for this latter finding is the higher stress is related to temporary migration.

B. Interpretation

The findings provide relatively positive evidence regarding the long-term returns to CCTs and their potential to reduce the intergenerational transmission of poverty. This stands somewhat in contrast with other findings in the literature on CCTs, which have suggested limited learning and labor market returns. Setting aside the methodological differences with other studies discussed above, we explore why, in this setting, the program effects could differ.

Most notably, the population studied in this paper is poorer and starts from much lower levels of education than many of the other previously studied CCTs in rural populations in Latin America. At such low levels of education, it is possible that spending additional time in the classroom as a result of the CCT more effectively leads to learning gains than in other settings where children already acquired basic language and math skills. If this interpretation is correct, one could expect impacts to be larger for the poorest segments of the target population. We use the original stratification of the randomized experiment at the locality level to investigate treatment heterogeneity, distinguishing between children from the three lowest (or poorest) strata and the four highest (or better off) strata. We estimate the effects by interacting the ITT estimator with a dummy that identifies the four highest strata. Results in Table 9 show that the ITT effects are indeed larger for boys in the poorest strata. Impacts on grades attained, learning, and socio-emotional outcomes are all significantly higher for the poorest children. Specifically, differential learning gains in column 3 are 0.36 SD for the three lowest strata, but are only 0.07 SD and not significantly different from zero for the higher strata, and results are also significantly different between strata for grades attained and socio-emotional outcomes. Overall, these results suggest that the differential learning results are driven primarily by the poorest half of the population who had lower average levels of education at baseline.[33] Hence encouraging children to enroll and stay in school may have particularly high returns in learning for the poorest, possibly because the marginal gains are higher for them than for children in the higher strata.[34]

Given the overall low levels of education, it should not be surprising that improvements in these basic skills did not automatically lead to skilled wage jobs. Findings from qualitative interviews in 2009 suggest, however, that such skills may have led to higher labor market returns in seasonal migration (CIERUNIC 2009). While temporary migration comes with costs, there are important differences in wages across regions in Nicaragua and wages are substantially higher in Costa Rica. Better math skills may have helped the young men adequately assess the cost-benefit trade-offs of seasonal migration, and better reading and writing skills may have enabled them to complete any relevant paper work, particularly for international migration.

That said, Table 9 does not show significant differences in labor market returns between the lower and the higher strata. This may indicate that earning gains are not necessarily the direct result of the additional grades attained alone. Indeed, more generally we do not interpret the results as the effects of the education transfers and schooling conditions alone, but rather as the result of being exposed to the entire package offered by the CCT program. Households were also eligible to receive the nutrition and health transfer, independent of whether their children were

[33] At baseline, educational attainment of boys in the 9–12 cohort from the three poorest strata already lagged behind those from higher strata by half a year (1 versus 1.5 grades attained). Moreover, their mothers also had less education (1.4 versus 2 grades attained).

[34] The cross-sectional relationship between learning and grades attained also shows decreasing returns to additional years attained, with the slope coefficient being the steepest for children with less than 3 years of grades attained.

eligible for the education transfer. By relaxing the household liquidity constraint, the nutrition and health transfer might have enabled children to stay in school even beyond the ages and grade levels the education transfers were targeting. In addition, children who did not meet the eligibility requirements of the education transfers may live in households with younger children who were eligible for education transfers, further relaxing household liquidity constraints.

The gains in off-farm earnings and the role of temporary migration in improving earnings could also capture other direct or indirect effects of the CCT program. One possible alternative mechanism for the migration and earnings results could be that the CCT program increased the strength of local networks (e.g., by children spending more time in school together or beneficiaries attending regular meetings), and that this in turn led to lower costs or improved information on migration and earnings opportunities. Variation in pre-program family networks and geographical variation in the density of the early treatment localities suggest some of these additional mechanisms may indeed be at play, but lack statistical power for us to make conclusive inferences (see Appendix G). In general, the design of the experiment, with the same relatively complex benefit package offered to all, is not well suited to disentangling the various potential specific mechanisms underlying the long-term gains.

C. Non-experimental Absolute Effects

Matching—In Table 10, we present non-experimental matching estimates of the absolute program effects, using the five nearest neighbors (NN5) bias-adjusted estimator to compare 2010 outcomes between the early treatment group and the non-experimental comparison group (see appendix H for robustness analysis). The findings confirm the overall direction and interpretation of the experimental results. The estimates suggest that exposure to the CCT for three years during critical ages in primary school yields 1.4 more grades attained and increases learning by 0.4 SD after 10 years. These results are consistent with the pattern in the differential results, which not only showed a significant differential in grades attained after 10 years, but also in enrollment in 2010. Results for the other outcome families overall point in the same direction as the experimental differential results, even if they are less precise. We interpret the difference in the absolute and differential program effects as evidence that exposure to the CCT increased school attainment and learning both for the early and late treatment groups, not only because it kept children in school during the program, but also because households (or the children themselves) continued to invest in education after the program ended. While these estimates are based on a much stronger set of assumptions than the experimental results, they suggest that the differential estimates, if anything, underestimate absolute effects of the CCT for the early treatment group, consistent with the likelihood that boys in the late treatment also benefited from their families receiving the program.

Double Difference—To provide additional evidence of absolute program impacts with a different identification strategy, Table 11 (panel A) presents double difference results that compare treatment municipalities to a comparison group using the Nicaraguan national census. The double difference results show clear improvements for all four educational indicators in 2005, all of which are the same or larger than the 2010 experimental differential effects. By 2005, boys in the 9–12 cohort (who would have been 14–17 in 2005) in treatment municipalities had attained 0.6 more grades than similarly aged boys in the comparison area and were 9 percentage points more likely to be literate, while they were still nearly 4 percentage points more likely to be enrolled.

These double difference absolute effects may be smaller than the 10-year absolute results because: 1) they are measured after a shorter period (2005 compared to 2010); 2) the program was implemented at different times in different areas and was still operating in the late treatment area in 2005; and 3) there may have been continued investment in children's education even after the end of the program. Panel B presents the absolute effects for these education variables for 2010, using the same matching estimator and sample as Table 10. The absolute effects measured in 2010 are consistently larger than the 2005 absolute estimates using the double difference approach. The two sets of results, at different time points and using different identification strategies, are consistent with continued improvements in education after the end of the program.

D. Cost-Benefit Analysis

Increased earnings hint at the potential for recouping costs of the CCT. A full cost-benefit analysis would require monetary assessment of all of the many potential benefits of the CCT, not only for these young men (e.g., for their learning gains) but also for all other members of the beneficiary households. More simply, we describe the potential for the program to have recovered its direct costs as follows. First, to estimate costs relevant for the education component of the program, we applied half of the total per household annual program costs in the first three years of the program to the boys in the sample. Second, to estimate benefits, we use the additional earnings generated by the program starting in 2009 and assume they persist. All flows are deflated to year 2000 real U.S. dollars.[35] Using a discount rate of 10 percent, the net present

[35] Following the approach outlined in Dhaliwal et al. (2012), all values are first translated into U.S. dollars using market exchange rates after which they are deflated to 2000, the base year, using U.S. CPI. Costs included are direct

value (NPV) turns positive in 2027, approximately 20 years after the program ended; at 5 percent, the NPV turns positive in 2016. While admittedly uncertain, under relatively conservative assumptions regarding cost and benefit flows for a subset of the beneficiaries, the program achieves positive NPV within a few decades.[36]

VI. Conclusions

CCTs and related interventions—combining short-term poverty reduction with enhanced investment in human capital to strengthen the productive capacity of future generations—have broad policy appeal. Numerous rigorous empirical studies have established that such programs have been effective at reducing contemporaneous poverty and increasing nutrition, health, and children's school attainment in the short term. Establishing and understanding the longer-term

program costs for running the program (administrative/management, targeting, monitoring and conditionality) but as recommended for cost-benefit analysis do not include transfers or evaluation costs (Caldés and Maluccio 2005). We allocate 50 percent of the average per household program cost as an estimate of costs related to the education components of the program, in nominal terms $60 a year for three years. Benefits are calculated as the estimated average increase monthly earnings ($10) starting in 2009, multiplied by average number of months worked (3.5) and by the average number of men in the cohort in each household (1.37). All these dollar figures are first deflated to year 2000 constant U.S. dollars and then discounted at 5 or 10 percent discount rates.

[36] We suggest the estimates are a lower bound on NPV for several reasons. First, we only consider benefits of one type, increased earnings. Second, we only consider benefits for one category of beneficiaries of the education components of the program (boys in the 9–12 cohort) within the household. Third, we only consider benefits starting in 2009. Last, we ignore that there would have been program costs associated with boys in the late treatment group further offsetting the costs of program in early treatment.

impacts of these programs is necessary for assessing the extent to which the investments in human capital fulfill their promise of improving the welfare of the next generation.

Building on the short-term randomized evaluation of a CCT program, we designed a 10-year follow-up study addressing the most important pitfalls for long-term analysis of human capital and labor market outcomes. In particular, we tracked migrants domestically and internationally to reduce attrition, and implemented the survey at a point when entry into the labor market for the cohort of interest was complete. Using the experimental variation in exposure, we estimate differential program effects between early and late treatment groups.

The experimental results indicate the CCT led to long-term gains in schooling, learning, and labor market returns. Moreover, the cohort of young men was more likely to still be enrolled at follow-up, suggesting the gains may increase further. The increased labor market returns reflect in particular gains in employment through temporary migration. One plausible interpretation is that with more education and learning, the young men developed core competencies that made them better at finding higher paying jobs that are often further away from home. Finally, as the impacts are estimated at the start of these young men's working lives, they potentially set them on a higher earnings trajectory for the future.

The sustained learning improvements are important in their own right, as they relate to the global sustainable development goal of quality education. In addition, we find that impacts on learning are larger for the poorest households. This has implications for targeting of CCTs, which currently do not always reach the poorest populations, even in countries with widespread coverage. Equally important, while the gains are substantial, they were experienced from low initial levels and in a population facing many other constraints. It is unsurprising, therefore, that the program did not completely transform the lives of these young men and their families to a fundamentally higher level of welfare.

Overall, our results show that exposure to a CCT during critical ages in primary school translated into gains in both learning and earnings, and point to important and sustained benefits several years after the end of the program. Based on the earnings gains, and under relatively conservative assumptions regarding cost and benefit flows for this subset of beneficiaries, the program achieves positive NPV within two decades. Despite skepticism in some policy circles, these widespread interventions may well have a role to play in reducing the intergenerational transmission of poverty.

VII. References

Abadie, A., and G. Imbens. 2008. "On the Failure of the Bootstrap for Matching Estimators." *Econometrica* 76(6): 1537–1557.

Abadie, A., and G. Imbens. 2011. "Bias-Corrected Matching Estimators for Average Treatment Effects." *Journal of Business & Economics Statistics* 29(1): 1–11.

Centro de Investigación y Estudios Rurales y Urbanos de Nicaragua (CIERUNIC). 2009. "Qualitative Findings for the Evaluation of the Long-term Impact of the *Red de Protección Social* in Nicaragua." Unpublished.

Aizer, A., S. Eli, J. Ferrie, and A. Lleras-Muney. 2016. "The Long-Run Impact of Cash Transfers to Poor Families." *American Economic Review* 106(4): 935–971.

Akee, R., W. Copeland, G. Keeler, A. Angold, and E.J. Costello. 2010. "Parent's Incomes and Children's Outcomes: A Quasi-Experiment with Casinos on American Indian Reservations." *American Economics Journal: Applied Economics* 2(1): 86–115.

Araujo, M. C., M. Bosch, and N. Schady. 2016. "Can Cash Transfers Help Households Escape an Intergenerational Poverty Trap?" National Bureau of Economic Research Working Paper No. 22670.

Athey, S., and G.W. Imbens. 2016. "The Econometrics of Randomized Experiments." Banerjee, A. and E. Duflo, (eds.), *Handbook of Economic Field Experiments*. Volume 1. Elsevier.

Attanasio, O., A. Guarín, C. Medina, C. Meghir. 2015. "Long Term Impacts of Vouchers for Vocational Training: Experimental Evidence for Colombia." NBER Working Paper 21390.

Attanasio, O., C. Meghir, and A. Santiago. 2012. "Education Choices in Mexico: Using a Structural Model and a Randomized Experiment to Evaluate PROGRESA." *Review of Economic Studies* 79(1): 37–66.

Baez, J. E., and A. Camacho. 2011. "Assessing the Long-Term Effects of Conditional Cash Transfers on Human Capital: Evidence from Colombia." World Bank Policy Research Working Paper No. 5681. Washington, DC, United States: World Bank.

Baird, S., C. McIntosh, and B. Özler. 2011. "Cash or Condition? Evidence from a Cash Transfer Experiment." *The Quarterly Journal of Economics* 126(4): 1709–1753.

Baird, S., C. McIntosh, and B. Özler. 2016. "When the Money Runs Out: Do Cash Transfers have Sustained Effects?" World Bank Policy Research Working Paper No. 7901. Washington, DC, United States: World Bank.

Baird, S., J.H. Hicks, M. Kremer, and E. Miguel. 2016. "Worms at Work: Long-run Impacts of a Child Health Investment." *Quarterly Journal of Economics* 131 (4): 1637–1680.

Bandiera, O. R. Burgess, S. Gulesci, N.Das, I. Rasul, M. Sulaiman. 2016. "Labor Markets and Poverty in Village Economies." *Quarterly Journal of Economics*, forthcoming.

Banerjee, A., E. Duflo, R. Chattopadhyay, and J. Shapiro. 2016. "The Long Term Impacts of a "Graduation" Program: Evidence from West Bengal." Unpublished.

Barham, T., K. Macours, and J.A. Maluccio. 2013. "Boys' Cognitive Skill Formation and Physical Growth: Long-Term Experimental Evidence on Critical Ages for Early Childhood Interventions," *American Economic Review: Papers and Proceedings* 103(3): 467–471.

Barrera-Osorio, F., Linden, L.L., and Saavedra, J.E., 2015. "Medium Term Educational Consequences of Alternative Conditional Cash Transfer Designs: Experimental Evidence from Colombia." CESR-Schaeffer Working Paper No. 2015-026

Behrman, J. R., S.W. Parker and P.E. Todd. 2009a. "Medium-Term Impacts of the *Oportunidades* Conditional Cash Transfer Program on Rural Youth in Mexico." In: S. Klasen and F. Nowak-Lehmann, eds. *Poverty, Inequality, and Policy in Latin America*, 219–270. Cambridge, United States: MIT Press.

Behrman, J. R., S.W. Parker and P E. Todd. 2009b. "Schooling Impacts of Conditional Cash Transfers on Young Children: Evidence from Mexico." *Economic Development and Cultural Change* 57(3): 439–477.

Behrman, J. R., S.W. Parker and P.E. Todd. 2011. "Do Conditional Cash Transfers for Schooling Generate Lasting Benefits? A Five-Year Followup of PROGRESA/Oportunidades." *Journal of Human Resources* 46(1): 93–122.

Bettinger, E., M. Kremer, M. Kugler, C. Medina, C. Posso, and J.E. Saavedra. 2016. "Can Educational Voucher Programs Pay for Themselves?" Unpublished.

Caldés, N., and J.A. Maluccio. 2005. "The Cost of Conditional Cash Transfers." *Journal of International Development* 17(2): 151–168.

de Janvry, A., and E. Sadoulet. 2006. "Making Conditional Cash Transfer Programs More Efficient: Designing for Maximum Effect of the Conditionality." *World Bank Economic Review* 20(1): 1–29.

Dhaliwal, I., E. Duflo, R. Glennerster, and C. Tulloch. 2013. "Comparing Cost-Effectiveness Analysis to Inform Policy in Developing Countries: A General Framework with Applications for Education." in P. Glewwe, Ed. *Education Policy in Developing Countries*, Oxford UK: Oxford University Press.

Duflo, E., P. Dupas, and M. Kremer. 2015. "Education, HIV, and Early Fertility: Experimental Evidence from Kenya." *American Economic Review* 105(9): 2757–2797.

Dunn, Lloyd M., D.E. Lugo, E.R. Padilla, and Leota M. Dunn. 1986. *Test de Vocabulario en Imágenes Peabody*. Circle Pines, Minnesota, United States: American Guidance Service, Inc.

Fernald, L., P. Gertler and L. Neufeld. 2009. "10-Year Effect of Oportunidades, Mexico's Conditional Cash Transfer Programme, on Child Growth, Cognition, Language, and Behaviour: a Longitudinal Follow-up Study." *Lancet* 371:828–837.

Filmer, D., and N. Schady. 2014. "The Medium-Term Effects of Scholarships in a Low-Income Country." *Journal of Human Resources* 49(3): 663–694.

Fiszbein, A., and N. Schady. 2009. "Conditional Cash Transfers: Reducing Present and Future Poverty." *World Bank Policy Research Report*. Washington, DC, United States: World Bank.

García, A., Romero, O.L., Attanasio, O. and Pellerano, L. 2012. "Impactos de Largo Plazo del Programa Familias en Acción en Municipios de Menos de 100 mil Habitantes en los Aspectos Claves del Desarrollo del Capital Humano." Technical report, Union Temporal Econometria S.A. SEI. con la asesoria del IFS.

Gertler, P. J. Heckman, R. Pinto, A. Zanolini, C. Vermeersch, S. Walker, S.M. Chang, and S. Grantham-McGregor. 2014. "Labor Market Returns to an Early Childhood Stimulation Intervention in Jamaica." *Science* 344(6187) 997–1001.

Gertler, P., S. Martinez and M. Rubio-Codina. 2012. "Investing Cash Transfers to Raise Long Term Living Standards." *American Economic Journal: Applied Economics* 4(1):164–192.

Hoddinott, J., J.A. Maluccio, J.R. Behrman, R. Flores, and R. Martorell. 2008. "Effect of a Nutrition Intervention During Early Childhood on Economic Productivity in Guatemalan Adults." *Lancet* 371(9610, 2): 411–416.

Ibarran, P., J. Kluve, L. Ripani and D. Rosas-Shady. 2015. "Experimental Evidence on the Long Term Impacts of a Youth Training Program." IZA discussion paper 9136.

Kling, J., J. Liebman and L. Katz. 2007. "Experimental Analysis of Neighborhood Effects." *Econometrica* 75(1): 83–119.

Kugler, A., M. Kugler, J. Saavedra, L.Omar and H. Prada. 2015. "Long-term Direct and Spillover Effects of Job Training: Experimental Evidence from Colombia." NBER working paper 21607.

Levy, S., and N. Schady. 2013. "Latin America's Social Policy Challenge: Education, Social Insurance, Redistribution." *Journal of Economic Perspectives* 27(2): 193–218.

Macours, K., N. Schady and R. Vakis. 2012. "Cash Transfers, Behavioral Changes, and Cognitive Development in Early Childhood: Evidence from a Randomized Experiment." *American Economic Journal: Applied Economics* (4)2: 247–273.

Macours, K. and R. Vakis. 2016. "Sustaining Impacts When Transfers End: Women Leaders, Aspirations, and Investment in Children." NBER Working Paper No. 22871.

Maluccio, J. A., and R. Flores. 2005. "Impact Evaluation of a Conditional Cash Transfer Program: The Nicaraguan *Red de Protección Social*." *Research Report 141*. Washington, DC, United States: International Food Policy Research Institute.

Maluccio, J. A., A. Murphy and F. Regalia. 2010. "Does Supply Matter? Initial Schooling Conditions and the Effectiveness of Conditional Cash Transfers for Grade Progression in Nicaragua." *The Journal of Development Effectiveness* 2(1): 87–116.

Markwardt, F. C. 1989. *Peabody Individual Achievement Test-Revised Manual*. Circle Pines, Minnesota, United States: American Guidance Service.

Molina Millan, T., T. Barham, K. Macours, J.A. Maluccio, and M. Stampini. 2016. "Long-term Impacts of Conditional Cash Transfers in Latin America: Review of the Evidence." IDB Working Paper Series No. IDB-WP-732.

Molina Millan, T. and K. Macours. 2017. "Attrition in Randomized Control Trials: Using Tracking Information to Correct Bias." Unpublished.

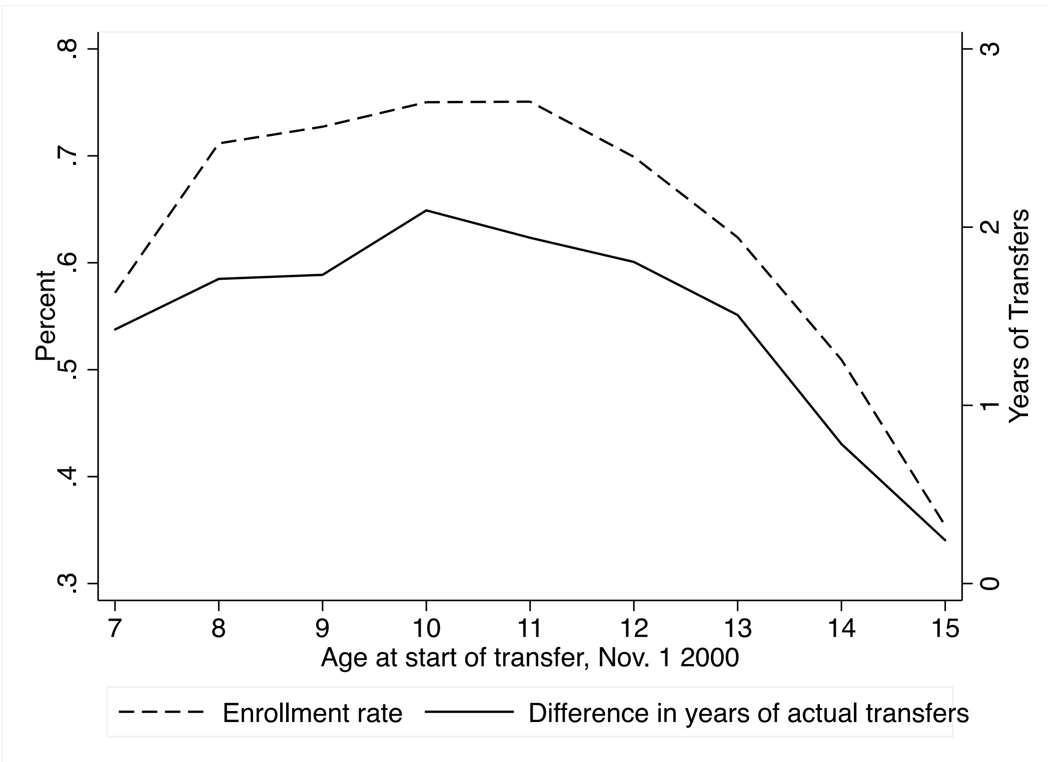
Parker, S., Rubalcava, L., Teruel, G. 2012. "Do Conditional Cash Transfer Programs Improve Work and Earnings among its Youth Beneficiaries? Evidence after a Decade of a Mexican Cash Transfer Program." Unpublished.

Raven, J.C., Court, J.H. and Raven, J. 1984. *Manual for Raven's Progressive Matrices and Vocabulary Scales. Section 2: Coloured Progressive Matrices*. London: H. K. Lewis.

Robles, M., M.G. Rubio, and M. Stampini. 2015. "Have Cash Transfers Succeeded in Reaching the Poor in Latin America and the Caribbean?" Policy Brief No. IDB-PB-246, Inter-American Development Bank, Washington DC.

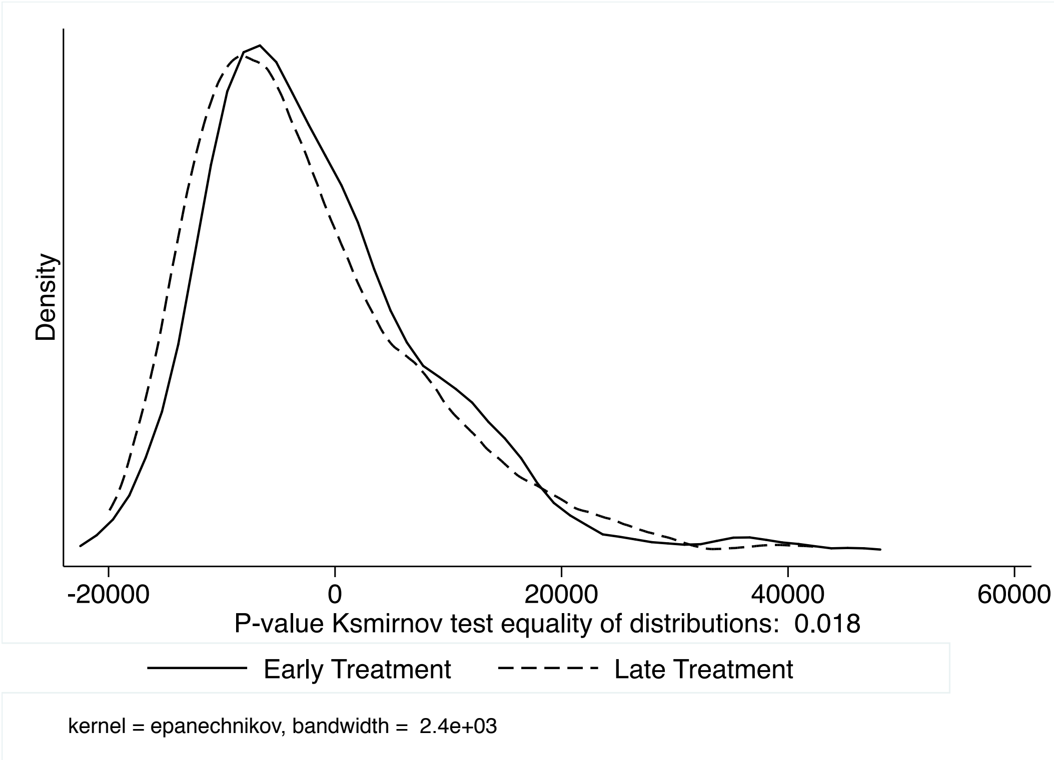
Figures and Tables

FIGURE 1: DIFFERENCE IN MEAN YEARS OF TRANSFERS RECEIVED BETWEEN EARLY AND LATE TREATMENT GROUPS AND MEAN ENROLLMENT RATE FOR BOYS



Notes: The difference in mean years of transfers refers to the mean difference in the total number of school years that children in the early and late treatment groups received transfers (left vertical-axis scale). The data on transfers is taken from the program administrative records and the enrollment rate (right vertical-axis scale) is from the 2000 census data. The means of the actual transfers are for all boys in the early and late treatment localities, not just those in the sample. Source: Author's calculations.

FIGURE 2: MONTHLY EARNINGS OFF-FARM, BOYS 9-12 IN 2000



Notes: Results for boys 9–12 in 2000. Earnings per month worked are demeaned using three-month age group indicators and indicator variables for grades attained at baseline, stratification, and region and trimmed at the top five percent of values.

Source: Author's calculations.

TABLE 1: 2002 AND 2004 EXPERIMENTAL IMPACTS ON EDUCATION

	Grades Attained	Completed Grade 4 =1	Enrolled =1	Attended School More Than 85% of Time =1	Read and Write =1
	(1)	(2)	(3)	(4)	(5)
Panel A: 2002 — Absolute Effects					
ITT	0.361***	0.015	0.182***	0.360***	0.150***
	(0.094)	(0.022)	(0.042)	(0.055)	(0.034)
N	475	475	475	475	475
R^2	0.828	0.743	0.191	0.271	0.324
Mean late treatment	2.396	0.182	0.733	0.544	0.735
Panel B: 2004 — Differential Effects					
ITT	0.487***	0.039	-0.049	-0.100	0.124***
	(0.155)	(0.027)	(0.063)	(0.066)	(0.029)
N	458	458	458	458	458
R^2	0.598	0.595	0.262	0.239	0.241
Mean late treatment	3.585	0.360	0.626	0.564	0.815

*Notes:**** p<0.01, ** p<0.05, * p<0.10. Standard errors are clustered at the locality level and given in parentheses. Compares ITT effects of early versus late treatment groups. The late treatment group started to receive the program in 2003, so 2002 represents absolute effects and 2004 differential effects. Results for boys that were 9–12 in 2000. Controls include three-month age group indicators and indicator variables for grades attained at baseline, stratification, and region. Attended school for more than 85% of the time is zero for those who were not enrolled in school at the time.

TABLE 2: 2010 DIFFERENTIAL EXPERIMENTAL IMPACTS FOR EDUCATION FAMILY AND LITERACY

	Education Family Z-Score	Education Family Components			Read and Write =1
		Grades Attained	Completed Grade 4 =1	Enrolled =1	
	(1)	(2)	(3)	(4)	(5)
ITT	0.093**	0.288*	0.031	0.045**	0.052**
	(0.042)	(0.167)	(0.025)	(0.021)	(0.021)
N	1,007	1,006	1,006	1,005	1007
R^2	0.346	0.425	0.356	0.096	0.175
Mean late treatment	-0.026	5.498	0.647	0.181	0.874

Notes: *** p<0.01, ** p<0.05, * p<0.10. Standard errors are clustered at the locality level and given in parentheses. Regressions are weighted to account for sampling and attrition providing population estimates. Differential ITT results compare early to late treatment groups in 2010. The family z-score averages the z-scores for the individual components and is calculated if any component is available. Z-scores are calculated using the mean and standard deviation of the late treatment group. Controls include three-month age group indicators and indicator variables for grades attained at baseline, stratification, and region.

TABLE 3: 2010 DIFFERENTIAL EXPERIMENTAL IMPACTS FOR LEARNING AND COGNITION FAMILIES (Z-SCORES)

	Learning			Mixed Cognition and Learning	Cognition (Raven)
	Math and Spanish	Math	Spanish		
	(1)	(2)	(3)	(4)	(5)
ITT	0.183**	0.160**	0.204**	0.113	-0.016
	(0.070)	(0.069)	(0.081)	(0.082)	(0.095)
N	907	905	907	906	906
R^2	0.448	0.395	0.437	0.380	0.215

Notes: *** p<0.01, ** p<0.05, * p<0.10. Standard errors are clustered at the locality level and given in parentheses. Regressions are weighted to account for sampling and attrition providing population estimates. Differential ITT results compare early to late treatment groups in 2010. Z-scores are calculated using the mean and standard deviation of the late treatment group. Controls include three-month age group indicators and indicator variables for grades attained at baseline, stratification, and region.

TABLE 4: 2010 DIFFERENTIAL EXPERIMENTAL IMPACTS FOR LEARNING AND COGNITION BY TEST (Z-SCORES)

	Learning Family Components					Mixed Cognition and Learning Family Components	
	Math Fluency	Math Problems	Reading Fluency	Spelling	Word Identification	Receptive Vocabulary	Memory Math
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
ITT	0.179**	0.137**	0.252***	0.206**	0.147*	0.128	0.094
	(0.082)	(0.067)	(0.078)	(0.087)	(0.075)	(0.102)	(0.074)
N	904	904	898	905	899	906	902
R^2	0.349	0.346	0.425	0.358	0.408	0.353	0.291

Notes: *** p<0.01, ** p<0.05, * p<0.10. Standard errors are clustered at the locality level and given in parentheses. Regressions are weighted to account for sampling and attrition providing population estimates. Differential ITT results compare early to late treatment groups in 2010. The family z-score is calculated by averaging the z-score for the individual components where the average is determined even if one component is missing. Z-scores are calculated using the mean and standard deviation of the late treatment group. Controls include three-month age group indicators and indicator variables for grades attained at baseline, stratification, and region.

TABLE 5: 2010 DIFFERENTIAL EXPERIMENTAL IMPACTS FOR LABOR MARKET PARTICIPATION AND PERMANENT MIGRATION

	Labor Market Participation Family Z-Score	Labor Market Participation Family Components				Permanent Migration Out of Municipality =1
		Worked Off-Farm =1 (last 12 months)	Migrated for Work =1 (last 12 months)	Ever Had a Salaried Non-Agricultural Job =1	Ever Worked in Urban Area =1	
	(1)	(2)	(3)	(4)	(5)	(6)
ITT	0.236***	0.062***	0.093***	0.084**	0.065*	-0.019
	(0.065)	(0.022)	(0.032)	(0.036)	(0.034)	(0.028)
N	1,006	1,006	1,006	998	998	1,007
R^2	0.146	0.074	0.169	0.132	0.115	0.102
Mean late treatment	-0.003	0.828	0.312	0.226	0.127	0.150

Notes: *** p<0.01, ** p<0.05, * p<0.10. Standard errors are clustered at the locality level and given in parentheses. Regressions are weighted to account for sampling and attrition providing population estimates. Differential ITT results compare early to late treatment groups in 2010. The family z-score is calculated by averaging the z-score for the individual components where the average is determined even if one component is missing. Z-scores are calculated using the mean and standard deviation of the late treatment group. Controls include three-month age group indicators and indicator variables for grades attained at baseline, stratification, and region.

TABLE 6: 2010 DIFFERENTIAL EXPERIMENTAL IMPACTS FOR EARNINGS FAMILY AND COMPONENTS

	Family Z-Score	Earnings Family Components (C$)			
		Earnings Per Month Worked (last 12 months)	Annual Earning (last 12 months)	Maximum Earnings (last 12 months)	Maximum Salary Ever
	(1)	(2)	(3)	(4)	(5)
Panel A: Earnings — Five Percent Trim					
ITT	0.192***	201.152***	595.013	211.421***	142.260*
	(0.067)	(63.624)	(619.322)	(69.318)	(71.919)
N	997	956	956	956	955
R^2	0.097	0.085	0.084	0.107	0.071
Mean late treatment	0.012	1436	8222	1619	228
Panel B: Rank of Earnings					
ITT	0.194***	41.780**	43.899**	49.313**	25.568
	(0.057)	(19.471)	(19.290)	(19.684)	(18.493)
N	1,006	1,006	998	1,006	1,006
R^2	0.097	0.082	0.103	0.094	0.095
Mean late treatment	0.003	497	487	498	503

Notes: *** p<0.01, ** p<0.05, * p<0.10. Standard errors are clustered at the locality level and given in parentheses. Regressions are weighted to account for sampling and attrition providing population estimates. Differential ITT results compare early to late treatment groups in 2010. Earnings include wage work off the family farm. Earnings in panel A are trimmed at the top five percent of values. Earnings are in Nicaragua Cordobas (C$) and the exchange rate is approximately 20. The family z-score is calculated by averaging the z-score for the individual components where the average is determined even if one component is missing. Z-scores are calculated using the mean and standard deviation of the late treatment group. Controls include three-month age group indicators and indicator variables for grades attained at baseline, stratification, and region.

TABLE 7: 2010 DIFFERENTIAL QUANTILE REGRESSIONS ON EARNINGS FAMILY (Z-SCORE)

	Percentile of Earnings Family								
	10	20	30	40	50	60	70	80	90
ITT	0.219	0.130	0.133	0.150*	0.172*	0.230*	0.223	0.300**	0.154
	(0.153)	(0.134)	(0.087)	(0.082)	(0.095)	(0.132)	(0.141)	(0.137)	(0.216)
N	997	997	997	997	997	997	997	997	997

Notes: *** p<0.01, ** p<0.05, * p<0.10. Standard errors are clustered at the locality level and given in parentheses. Regressions are weighted to account for sampling and attrition providing population estimates. Differential ITT results compare early to late treatment groups in 2010. The family z-score is calculated by averaging the z-score for the individual components where the average is determined even if one component is missing. Z-scores are calculated using the mean and standard deviation of the late treatment group. Controls include three-month age group indicators and indicator variables for grades attained at baseline, stratification, and region.

TABLE 8: 2010 DIFFERENTIAL EXPERIMENTAL IMPACTS FOR SOCIOEMOTIONAL FAMILY OUTCOMES

	Family Z-Score	Socioemotional Family Components			
		Positive Self Evaluation	Optimism	Stress	Negative Self Evaluation
	(1)	(2)	(3)	(4)	(5)
ITT	0.053	0.249**	0.287***	0.170**	0.155*
	(0.039)	(0.093)	(0.078)	(0.071)	(0.086)
N	900	900	900	900	900
R^2	0.152	0.194	0.180	0.097	0.106

Notes: *** p<0.01, ** p<0.05, * p<0.10. Standard errors are clustered at the locality level and given in parentheses. Regressions are weighted to account for sampling and attrition providing population estimates. Differential ITT results compare early to late treatment groups in 2010. Socioemotional components are the first four factors resulting from exploratory factor analysis of all socioemotional questions. The family z-score is calculated by averaging the z-score for the individual components. Z-scores are calculated using the mean and standard deviation of the late treatment group. Controls include three-month age group indicators and indicator variables for grades attained at baseline, stratification, and region.

TABLE 9: 2010 DIFFERENTIAL EXPERIMENTAL IMPACTS FOR ALL FAMILIES BY STRATA

	Education		Learning Family Z-Score	Labor Market Participation Family Z-Score	Earnings Family Z-Score		Socio-emotional Family Z-Score
	Grades Attained	Family Z-Score			Absolute (5 % Trim)	Rank	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
ITT	0.667**	0.155**	0.360***	0.174**	0.081	0.223**	0.185***
	(0.250)	(0.075)	(0.127)	(0.076)	(0.094)	(0.085)	(0.055)
Four Highest Strata (=1) * ITT	-0.629*	-0.102	-0.294**	0.103	0.185	-0.049	-0.220***
	(0.336)	(0.094)	(0.145)	(0.120)	(0.119)	(0.106)	(0.079)
Test: ITT + Highest Strata * ITT = 0							
P-value	0.861	0.316	0.342	0.005	0.002	0.018	0.504
N	1,006	1,007	907	1,006	997	1,006	900
R^2	0.427	0.347	0.453	0.146	0.099	0.097	0.158

Notes: *** p<0.01, ** p<0.05, * p<0.10. Standard errors are clustered at the locality level and given in parentheses. The strata are ordered from 1 to 7 with one being the poorest. Highest strata includes strata 4–7. Regressions are weighted to account for sampling and attrition providing population estimates. Differential ITT results compare early to late treatment groups in 2010. Z-scores are calculated using the mean and standard deviation of the late treatment group. Controls include three-month age group indicators and indicator variables for grades attained at baseline, stratification, and region.

TABLE 10: 2010 MATCHING ABSOLUTE IMPACTS FOR ALL FAMILIES

	Education		Learning Family Z-Score	Economic Participation Family Z-Score	Earnings Family Z-Score		Socio-emotional Family Z-Score
	Grades Attained	Family Z-Score			Absolute (5% Trim)	Rank	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
ATT	1.379***	0.380***	0.387***	0.160	0.106	0.066	0.102
	(0.337)	(0.085)	(0.125)	(0.120)	(0.107)	(0.098)	(0.078)
N	690	690	616	690	687	690	613

Notes: *** p<0.01, ** p<0.05, * p<0.10. Standard errors are clustered at the locality level and given in parentheses. Absolute effects compare early treatment to comparison group in 2010. Mean of grades attained in the comparison group is 5.3 in 2010. ATT biased adjusted estimator (Abadie and Imbens 2011) using five nearest neighbors. Z-scores are calculated using the mean and standard deviation of the late treatment group.

TABLE 11: 2005 AND 2010 ABSOLUTE IMPACTS ON EDUCATION

	Grades Attained	Completed Grade 4 =1	Enrolled =1	Read and Write =1
	(1)	(2)	(3)	(4)
Panel A: 2005 Absolute Effects — Double-Difference Estimates				
Treatment municipality * 2005 (δ3)	0.597***	0.124***	0.037***	0.091***
	(0.078)	(0.014)	(0.014)	(0.013)
N	18,399	18,399	18,421	18,403
R^2	0.107	0.088	0.037	0.061
Mean untreated municipalities in 2005	3.922	0.559	0.456	0.779
Panel B: 2010 Absolute Effects — NN5 Matching Estimates				
ATT	1.379***	0.218***	0.096***	0.121***
	(0.344)	(0.057)	(0.027)	-0.027
N	690	690	690	690
Mean comparison group in 2010	5.304	0.583	0.157	0.839

Notes: *** p<0.01, ** p<0.05, * p<0.10. 2005 absolute effects use national census data to compare rural areas of program municipalities to rural areas of comparison group municipalities. Heteroskedasticity-robust standard errors are given in parentheses. 2010 absolute effects use evaluation survey data to compare early treatment to the comparison group. Estimates are from the ATT biased adjusted estimator (Abadie and Imbens 2011) using five nearest neighbors. Standard errors are clustered at the locality level and given in parentheses.

Online Appendices for “Are Conditional Cash Transfers Fulfilling Their Promise? Schooling, Learning, and Earnings After 10 Years”

By TANIA BARHAM, KAREN MACOURS, AND JOHN A. MALUCCIO

March 2017

APPENDIX A: BASELINE BALANCE

TABLE A1: BASELINE BALANCE FOR 9–12 COHORT SAMPLE

	Early Treatment			Late Treatment			Diff. in Means		Mean/ SD
	Mean	SD	N	Mean	SD	N	Diff.	P-value	
Individual Characteristics									
Age at start of transfers in months	11.0	1.12	516	11.0	0.43	490	-0.02	0.67	-0.06
No grades attained (=1)	0.46	1.31	516	0.43	1.18	490	0.03	0.75	0.02
Highest grade attained	1.21	3.39	516	1.19	2.86	490	0.02	0.88	0.01
Worked last week (=1)	0.17	0.61	516	0.21	0.48	490	-0.04	0.16	-0.09
Mother not living in same household	0.08	0.28	516	0.07	0.33	490	0.01	0.43	0.03
Father not living in same household	0.22	0.66	516	0.18	0.40	490	0.04	0.16	0.11
Child of household head	0.86	0.41	516	0.88	0.35	490	-0.02	0.39	-0.06
Mother no grades attained (=1)	0.45	0.81	516	0.49	0.77	490	-0.04	0.32	-0.05
Mother 3 plus grades attained (=1)	0.37	0.85	516	0.32	0.75	490	0.04	0.33	0.06
Household Head Characteristics									
Age	44.8	14.6	516	44.4	13.2	490	0.40	0.58	0.03
No grades attained (=1)	0.53	0.80	516	0.50	0.55	490	0.04	0.33	0.06
3 plus grades attained (=1)	0.29	0.57	516	0.28	0.47	490	0.02	0.54	0.04
Worked last week (=1)	0.85	0.68	516	0.90	0.39	490	-0.05	0.06	-0.13
Household Characteristics									
Log predicted expenditures (pc)	7.71	0.73	516	7.74	0.68	490	-0.03	0.38	-0.04
Number of household members	8.26	4.57	516	8.22	5.20	490	0.04	0.91	0.01
Number of children aged 0-8	2.10	2.31	516	2.08	2.57	490	0.02	0.92	0.01
Number children aged 9-12	1.76	0.89	516	1.80	1.38	490	-0.04	0.53	-0.03
Log of size of landholdings	7.81	7.38	516	8.10	8.87	490	-0.29	0.57	-0.03
Family network size (individuals)	92.2	201.3	516	68.36	162	490	23.83	0.03	0.15
Own house (=1)	0.81	0.89	516	0.88	0.62	490	-0.08	0.11	-0.12
Some in household work in ag	0.82	0.78	516	0.85	0.86	490	-0.03	0.61	-0.03
Wealth index - housing characteristics	0.10	3.81	516	-0.02	2.90	490	0.13	0.57	0.04
Wealth index - productive assets	-0.13	1.53	516	0.14	2.08	490	-0.27	0.02	-0.13
Wealth index - other assets	-0.04	3.18	516	0.00	3.45	490	-0.03	0.78	-0.01
Number of rooms in house	1.59	1.65	516	1.58	1.34	490	0.01	0.89	0.01
Cement block walls (=1)	0.17	0.90	516	0.15	0.64	490	0.02	0.68	0.04
Zinc roof (=1)	0.54	1.60	516	0.49	1.39	490	0.05	0.62	0.04
Dirt floor (=1)	0.85	0.67	516	0.87	0.58	490	-0.02	0.64	-0.04
Latrine (=1)	0.65	1.25	516	0.60	1.14	490	0.04	0.57	0.04
Electric light (=1)	0.26	1.18	516	0.20	1.01	490	0.06	0.24	0.06
Radio (=1)	0.22	0.59	516	0.22	0.81	490	-0.01	0.86	-0.01
Work animals (=1)	0.11	0.35	516	0.18	0.73	490	-0.07	0.04	-0.09
Fumigation sprayer (=1)	0.31	0.81	516	0.38	1.15	490	-0.07	0.13	-0.06
Distance to nearest school (minutes)	24.9	80.5	516	23.9	66.8	490	1.09	0.84	0.02
Live in Tuma region (=1)	0.52	2.74	516	0.29	2.40	490	0.23	0.13	0.09
Live in Madriz region (=1)	0.19	2.07	516	0.18	2.19	490	0.01	0.89	0.00
Characteristics of Nearest School									
Highest grade school offers	4.78	3.34	512	4.85	6.12	484	-0.07	0.80	-0.01
Student-teacher ratio	36.7	28.2	507	36.8	40.7	419	-0.16	0.83	0.00
School under local governance	0.30	2.22	512	0.27	2.09	484	0.03	0.76	0.02

Notes: Standard errors and deviations are clustered at the locality level following the stratified randomization design. Means are weighted to account for sampling and attrition providing population estimates. The sample includes boys age 9–12 at program start. To match the 2010 differential analysis, boys without years of grades attained in 2010 are dropped. The p-value for the difference in means includes controls for strata following the program design (difference in means do not include strata controls). The mean divided by the standard deviation uses the standard deviation of the late treatment group. The predicted per capita expenditures are from the program census data and uses the proxy means method developed by the government of Nicaragua for the purpose of household targeting (based on the 1998 Nicaraguan Living Standards Measurement Survey, Maluccio [2009]). The asset indices are constructed using principal component analysis for the assets list (appendix D). School characteristics are constructed from program administrative data collected for monitoring conditionalities. School under local governance refers to schools that participated in Nicaraguan's school autonomy reform, which provided schools and parents a certain level of autonomy over their own management and operations.

APPENDIX B: ROBUSTNESS—ALTERNATIVE SPECIFICATIONS AND SAMPLES

In Appendix Table B1, we examine the robustness of the 2010 differential results to a number of alternative specifications and samples. Results from the main specification are reproduced in panel A for comparison.

First, we limit the sample to 11-year olds at the start of the program, the group we oversampled because of their large differential exposure. We expect program effects to be larger for this group than for the overall sample, but possibly less precise as the sample size is substantially smaller (353 versus 1,007). Results in panel B indeed show higher point estimates for educational gains (0.7 grades attained), learning (0.3 SD), and earnings (approximately 0.2 SD).

Second, we estimate effects controlling only for strata to follow the stratified randomized design of the program. The results in panel C are broadly consistent with the main results.

Third, in panel D we include additional controls. These are estimated per capita expenditures, number of children 9–12 in the household, distance to school, the productive assets component of the wealth index, an indicator for whether the child worked at baseline, and family network size. These are important correlates of the outcomes, and the last three were not balanced at baseline. Results are similar to the main results.

Fourth, to explore how the weights affect the analysis, results are presented with sampling as opposed to attrition-corrected weights, and without weights in panels E and F. Again, results are consistent with the main results.

Finally, in panel G we re-estimate effects for the variables derived from the household instrument for the sample of boys for whom there is information from the individual instrument, and use the weights designed to account for attrition and sampling in the individual survey.

Results are qualitatively similar even if the earnings estimates are slightly lower and no longer significant, while the estimate on the rank of earnings is slightly higher and significant.

TABLE B1: 2010 DIFFERENTIAL EXPERIMENTAL IMPACTS, ALTERNATIVE SPECIFICATIONS

	Education		Learning Family Z-Score	Labor Market Participation Family Z-Score	Earnings Family Z-Score		Socio-emotional Family Z-Score
	Grades Attained	Family Z-Score			Absolute (5% Trim)	Rank	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Age 9–12 Main results							
ITT	0.288*	0.093**	0.183**	0.236***	0.192***	0.194***	0.053
	(0.167)	(0.042)	(0.070)	(0.065)	(0.067)	(0.057)	(0.039)
N	1,006	1,007	907	1,006	997	1,006	900
Panel B: Age 11 Only							
ITT	0.681*	0.177*	0.283***	0.223	0.183	0.245**	-0.046
	(0.362)	(0.099)	(0.079)	(0.135)	(0.129)	(0.110)	(0.070)
N	352	353	321	353	349	353	318
Panel C: Age 9–12 Strata Controls Only							
ITT	0.361	0.101	0.326**	0.165**	0.133**	0.143	0.128*
	(0.324)	(0.078)	(0.141)	(0.063)	(0.059)	(0.085)	(0.070)
N	1,006	1,007	907	1,006	997	1,006	900
Panel D: Age 9–12 Extended Controls							
ITT	0.289*	0.106**	0.165**	0.182***	0.179***	0.207***	0.045
	(0.167)	(0.046)	(0.067)	(0.064)	(0.057)	(0.060)	(0.039)
N	1,006	1,007	907	1,006	997	1,006	900
Panel E: Age 9–12 Sample Population Weights							
ITT	0.351**	0.096**	0.175***	0.188***	0.195***	0.230***	0.033
	(0.153)	(0.042)	(0.063)	(0.069)	(0.057)	(0.061)	(0.041)
N	1,006	1,007	907	1,006	997	1,006	900
Panel F: Age 9–12 No Weights							
ITT	0.267	0.082*	0.170***	0.111	0.103*	0.169***	0.021
	(0.165)	(0.043)	(0.060)	(0.072)	(0.059)	(0.060)	(0.039)
N	1,006	1,007	907	1,006	997	1,006	900
Panel G: Restrict Household Variables to Test Sample and Weights							
ITT	0.351*	0.128***		0.121	0.100	0.212***	
	(0.176)	(0.046)	-	(0.077)	(0.071)	(0.069)	-
N	906	907		907	898	907	

Notes: *** p<0.01, ** p<0.05, * p<0.10. Standard errors are clustered at the locality level and in parentheses. Regressions are weighted to account for sampling and attrition providing population estimates. Differential ITT results compare early to late treatment groups in 2010. The family z-score averages the z-scores for the individual components and is calculated if any component is available. Z-scores are calculated using the mean and standard deviation of the late treatment group. Controls include three-month age group indicators and indicator variables for grades attained at baseline, stratification, and region.

Appendix C: Details of the CCT's Program Design

The Nicaraguan CCT, the *Red de Protección Social* (RPS) was designed to address both current and future poverty through cash transfers targeted to poor households in rural Nicaragua. The program was implemented by the government of Nicaragua with technical assistance and financial support from the Inter-American Development Bank (IDB). The randomized evaluation was built into the design of the program and the International Food Policy Research Institute carried out the original short-term evaluation. The program started in 2000 and comprised two budgetary phases over six years. Phase I lasted three years with a budget of $11 million. In late 2002, based in part on the positive findings of various evaluations, the government of Nicaragua and the IDB agreed to a continuation and expansion of the program until 2006 with a budget of $22 million.

Program Targeting—The CCT first targeted rural areas in six municipalities of the central region of Nicaragua, on the basis of poverty as well as on their capacity to implement the program. The focus on rural areas reflected the distribution of poverty in Nicaragua—of the 48 percent of Nicaraguans identified as poor in 1998, 75 percent resided in rural areas (World Bank 2001). While the targeted municipalities were not the poorest in the country, nor in the central region for that matter, the proportion of impoverished people living in these areas was still well above the national average (World Bank 2003). In addition, these areas had easy physical access and communication (for example, less than a day's drive to Managua, where the central office was located), relatively strong institutional capacity and local coordination, and good access to primary schools.

In the next stage of pre-program targeting, a marginality index was constructed for all 59 rural census localities in the selected municipalities. The index was the weighted average of a set

of locality-level indicators (including average family size, lack of access to potable water, access to latrines, and illiteracy rates, all calculated from the 1995 *National Population and Housing Census*) in which higher index scores are considered to be more impoverished areas.
The 42 localities with the highest calculated scores were selected for inclusion in the program. Finally, while all households in the 42 targeted localities were to be eligible in the initial program design, the government excluded about 3 percent of households deemed to have substantial resources (in particular those who owned a vehicle or substantial agricultural land) ex ante from the program (Maluccio 2009) and these are excluded from the analyses in the paper.

While not statistically representative of rural Nicaragua, the 42 localities are similar to other rural areas in the central region and elsewhere in the country. Three-quarters of the approximately 1,000 rural localities in the country as a whole have marginality index scores in the same range as the program areas.

Program Components and Conditionalities—The transfers were conditional on a household's health and education behaviors, and the conditionalities were monitored by teachers and healthcare providers. The specific stated objectives of the program included: i) supplementing household income for up to three years to increase expenditures on food; ii) reducing dropout rates during the first four years of primary school; and iii) increasing the healthcare and nutritional status of children under the age of five years. Only the designated household representative was allowed to collect the transfers and, where possible, the CCT appointed the mother or another female caregiver to this role. As a result, more than 95 percent of the household representatives were women. The CCT also worked with local volunteer coordinators (beneficiary women chosen by the community) to implement the program.

The coordinators were charged with keeping beneficiary household representatives informed about upcoming healthcare appointments for their children, upcoming transfers, and any failures in fulfilling conditions. The conditions were monitored for compliance, and the specific transfers were withheld if conditions for that transfer were not met. Conditionalities and benefits were explained to eligible families in the early treatment group during registration assemblies that took place in September and October 2000 and transfers began in November.

The CCT had two core components. The first core component focused on education. Each eligible household received a bimonthly cash transfer known as the "school attendance transfer," contingent on enrollment and regular school attendance of children aged 7–13 years who had not yet completed the fourth grade of primary school. Additionally, for each eligible child, the household received an annual cash transfer at the start of the school year intended for school supplies (including uniforms and shoes) known as the "school-supplies transfer," which was contingent on enrollment. Unlike the school attendance transfer, which was a fixed amount per household regardless of the number of children in school, the school supplies transfer was per child.

To provide incentives to the teachers and to increase resources available to the schools, there was also a small cash transfer, known as the "teacher transfer".[37] This was given for each beneficiary child, and caregivers in turn delivered it to the teacher. The teacher was meant to keep one-half, while the other half was earmarked for the school. Although the delivery of the funds to the teacher was a condition of the program that was monitored, their ultimate use by the teacher and the school was not monitored.

[37]In rural Nicaragua, school's parents' associations often request small monthly contributions from parents to support the teacher and the school; the teacher transfer was, in part, intended to substitute for this type of fee.

The second component focused on food security, nutrition, and health. Each eligible household received a bimonthly (every two months) cash transfer we call the "nutrition and health transfer" that was a fixed amount per household, regardless of household size and regardless of whether a household had children affected by the associated conditionalities. It was contingent upon the designated household representative attending bimonthly health information workshops and bringing children under age five for scheduled preventive healthcare appointments with specially trained and contracted providers. The workshops were held within the communities and covered household sanitation and hygiene, nutrition, and related topics. For preventive health care visits, children under age two were seen monthly and those age two to five bimonthly. Health services at the scheduled visits included growth monitoring, vaccination, iron supplementation, and provision of anti-parasite medicine.

The supply of healthcare was increased to ensure the program could meet increased demand for healthcare without reducing quality. In particular, the CCT contracted and trained private health providers to provide services free of charge (Regalia and Castro 2007), and beneficiaries were required to use those providers for fulfillment of the conditions. Providers visited program areas on scheduled dates and delivered services in existing health facilities, community centers, or private homes. In 2003, additional services (and corresponding conditions) were added, including vaccination for school-age children, family planning services for women of childbearing age, prenatal care consultations, and a health educational workshop for adolescents and household representatives. These services were scheduled to be continued in the early treatment group after 2003, although in practice this was only partially implemented.

The initial U.S. dollar annual amounts and their Nicaraguan Córdoba (C$) equivalents of the transfers (using the September 2000 average exchange rate of C$ 12.85 to US$ 1.00) were as follows: the nutrition and health transfer was $224 a year, the school attendance transfer $112, the school supply transfer $21, and the teacher transfer $5. On its own, the nutrition and health transfer represented about 13 percent of total annual household expenditures in beneficiary households before the program. A household with one child benefiting from the education component would have received additional transfers of about 8 percent, yielding an average total potential transfer of 21 percent of total annual household expenditures. The nominal value of the transfers remained constant, with the consequence that the real value of the transfers declined by about 8 percent due to inflation during budgetary Phase I. In budgetary Phase II, which began in 2003 and incorporated the late treatment group, the size of the transfers was reduced. The nutrition and health transfer started at $168 for the first year of program participation and then declined to $145 and $126 in the second and third years. The school attendance transfer also declined slightly, to $90 per year, but the teacher transfer rose to $25 per student. All figures represent potential transfers, i.e. transfers received when fully complying with all conditions.

To enforce compliance with program requirements, beneficiaries did not receive the nutrition and health, or separately education, transfers, in a given transfer period when they failed to carry out all of the relevant conditions described above. Repeated violation led to households losing their overall eligibility.

APPENDIX D: WEALTH INDEX

The baseline program census data contain a number of variables to proxy for household wealth, including variables capturing characteristics of the housing structure and household assets. Following Filmer and Pritchett (2001), we aggregate these characteristics using principal components analysis. The principal components are estimated using the baseline target sample of all 9–12 year olds regardless of whether they were interviewed in 2010. We retain the first three principal components as they each have an eigenvalue of more than one; jointly, they account for 53 percent of the variation of the nine underlying variables included. The first principal component mostly captures characteristics of the house, the second productive assets (ownership of work animals and a fumigation sprayer), and the third specific household amenities (roof materials and latrines).

TABLE D1. PRINCIPAL COMPONENT SCORING COEFFICIENTS

Variable	PC 1	PC 2	PC 3
Household Characteristics			
Work animals (=1)	0.13	0.62	-0.03
Fumigation sprayer (=1)	0.12	0.59	0.36
Number of rooms in the house	0.35	0.27	0.01
Radio (=1)	0.39	0.07	-0.21
Cement block walls (=1)	0.44	-0.10	0.11
Zinc roof (=1)	0.21	-0.16	0.69
Dirt floor (=1)	-0.44	0.26	-0.18
Latrine (=1)	0.28	0.14	-0.49
Electricity light (=1)	0.43	-0.27	-0.23

Notes: PC refers to principal component

APPENDIX E: LEARNING, COGNITIVE, AND SOCIO-EMOTIONAL OUTCOMES

All standardized tests included in the 2010 individual survey instrument were piloted extensively and minor adjustments made for the local context as necessary, such as rephrasing questions for maximum understanding. Similar tests have been applied in other populations in Latin America, including in the evaluations of CCT programs in Ecuador and Mexico, and a different CCT program in Nicaragua (Behrman, Parker, and Todd 2009a; Fernald, Gertler, and Neufeld 2009; Paxson and Schady 2010; Macours, Schady, and Vakis 2012).

Tests were conducted in the young adult respondents' homes by specially trained female test administrators. Therefore, the results were obtained independent of whether the respondent was in school, avoiding potential selection concerns.

Test administrators were selected for their background (trained as psychologists, social workers, or similar fields) and for their ability to quickly establish a strong rapport with children and young adults. They were trained to motivate the respondents to participate in the tests, keeping final non-response to a minimum. Tests were administered inside the home (or in the compound) and the privacy of the test-taker and the confidentiality of the results were assured throughout the process. During the test administrators' training, emphasis was placed both on gaining the confidence of the respondents before starting the tests and on the standardized application of each of the tests. The quality and standardized application of the tests was monitored closely in the field, and given the long survey period, several re-standardization trainings were carried out.

Data collection and test administration was organized in such a way that the test administrators would maintain a balance between the number of children visited in early and late treatment localities. Visits to early and late treatment localities were also balanced over time to avoid possible seasonal differences in measurement between the experimental groups. Consistent with these field protocols, the main results are robust to controls for the identity of the test administrator (not shown).

Exploratory Factor Analysis for Socio-Emotional Outcomes—Two standardized instruments to measure socio-emotional outcomes were applied in the individual instrument. The first was the Strength and Difficulties test (SDQ), a self-reported behavioral screening test consisting of 25 questions aimed at measuring a set of positive and negative behaviors. In addition, we implemented the Center for Epidemiologic Studies Depression Scale or CESD (Radloff 1977), a commonly used mental health scale, developed as a screening test for depression and depressive disorder and consisting of 20 questions asking for the frequency of both positive and negative self-perceptions. Both tests are available in Spanish.

We first analyze the internal consistency for the sample of boys studied (9–12 year olds at baseline) through exploratory factor analysis. The overall Cronbach alpha of the 25 items of the SDQ together (0.70) indicates that the scale as a whole is internally consistent. But the alphas are much lower when considering the five usual subdomains (emotional symptoms, conduct problems, hyperactivity, peer relationships, and pro-social behavior), and vary from 0.26 to 0.51, hence much lower than the usual threshold for statistical validity. Exploratory factor analysis on the 25 items suggests there are only two factors that can be meaningfully retained (i.e., two factors have eigenvalues above one and the scree plot leads to a similar conclusion). Moreover, when imposing the 5-factor structure, the items do not group along the original five subscales, with the first factor having high factor loads on items from three of the five subscales.[38] When we consider the CESD, the Cronbach alpha for internal consistency of the 20 items is high (0.83) but the factor analysis only points to one or two factors, and does not allow further differentiation.

As the data suggests that we should not pool questions together based on regular subcategories of the SDQ, we construct new indices capturing the relevant latent traits, based on all items in the SDQ and CESD. We pool together all questions from the SDQ and the CESD scales and identify the latent socio-emotional traits in our sample, following, among others, Cunha, Heckman, and Schennach (2010) and Attanasio et al. (2015). Based on both the eigenvalue and the scree plot, and using an oblique quartimin rotation to allow the different factors to be correlated with one another, we retain four factors, two factors with high loads on items from the SDQ scale, and two factors with high loads on items from the CESD scale. Notably, questions referring to positive, respectively negative, attitudes or behavior are pooled in each of the scales. Hence considering the factor loadings of the different items points to a plausible interpretation of these factors as capturing stress, positive self-evaluation, negative self-evaluation, and optimism.[39] Results are shown in Appendix Table E1.

[38] Similar findings have been obtained on other measures of socio-emotional outcomes when scales originally designed for developed country settings are used in developing country settings (Laajaj and Macours 2017).

[39] All results are qualitatively similar with or without inclusion of the non-experimental comparison group.

TABLE E1. FACTOR LOADINGS OF SOCIO-EMOTIONAL QUESTIONS

	Factor 1 Stress	Factor 2 Positive Self-Evaluation	Factor 3 Negative Self-Evaluation	Factor 4 Optimism
CESD				
During the last 7 days, how many days …				
… were you bothered by things that usually don't bother you?	0.45	0.08	-0.10	0.12
… did you not feel like eating? (your appetite was poor)	0.53	-0.06	0.09	-0.07
… did you feel that you could not shake off the blues even with help from your family and friends?	0.66	-0.06	0.06	-0.01
… did you feel that you were just as good as other people?	0.60	0.05	0.00	-0.05
… did you have trouble keeping your mind on what you were doing?	0.56	0.01	-0.03	0.14
… did you feel depressed?	0.58	0.04	-0.06	0.08
… did you feel that everything you did was an effort?	0.38	-0.02	-0.04	0.23
… were you hopeful about the future?	0.16	-0.07	0.01	0.54
… did you think your life had been a failure?	0.56	0.08	0.01	-0.07
… did you feel fearful?	0.53	-0.03	-0.02	0.00
… was your sleep restless?	0.48	-0.02	0.05	0.09
… were you happy?	-0.11	0.00	0.01	0.47
… did you talk less than usual?	0.39	-0.04	-0.02	0.10
… did you feel lonely?	0.56	-0.01	-0.02	-0.06
… people were unfriendly?	0.54	0.10	-0.06	0.06
… did you enjoy life?	-0.29	0.01	0.06	0.41
… did you have crying spells?	0.45	-0.01	0.03	-0.13
… did you feel sad?	0.62	-0.07	-0.04	-0.16
… did you feel that people disliked you?	0.48	0.05	-0.14	0.06
... could you not get 'going'?	0.49	-0.03	-0.04	0.01
… did you feel you were moving ahead in life?	-0.05	-0.03	0.02	0.61
… where you thinking abou the way to move ahead in life?	0.04	-0.09	0.01	0.59
SDQ				
I try to be nice to other people. I care about other people's feelings	0.01	0.19	0.10	-0.21
I am restless, I cannot stay still for long	0.02	0.15	0.07	-0.15
I get a lot of headaches, stomach-aches or sickness	-0.04	0.14	0.25	0.11
I usually share with others, for example food, pencils/	0.01	0.23	0.17	-0.08
I get very angry and often lose my temper	-0.04	0.00	0.43	0.04
I would rather be alone than with other people	-0.05	0.16	0.18	0.07
I usually do as I am told	-0.01	0.41	-0.09	0.01
I worry a lot*	-0.02	0.39	0.08	-0.03
I am helpful if someone is hurt, upset or feeling ill	0.00	0.48	0.01	-0.02
I am constantly fidgeting or squirming*	0.13	0.32	0.13	-0.13
I have one good friend or more	0.06	0.34	0.02	0.04
I fight a lot. I can make other people do what I want	-0.02	-0.15	0.35	-0.05
I am often unhappy, depressed or tearful	-0.14	0.13	0.41	0.01

Other people my age generally like me	0.05	0.42	-0.03	0.00
I am easily distracted, I find it difficult to concentrate	0.05	0.30	0.16	0.11
I am nervous in new situations. I easily lose confidence	-0.05	0.11	0.35	0.17
I am kind to younger children	0.02	0.36	0.08	-0.11
I am often accused of lying or cheating	0.04	-0.07	0.47	-0.06
Other young people pick on me or bully me	-0.07	-0.03	0.45	0.06
I often offer to help others (parents, teachers, children)	-0.07	0.53	-0.02	0.00
I think before I do thing	-0.03	0.46	-0.01	-0.07
I take things that are not mine from home, school or elsewhere	0.04	-0.16	0.32	-0.03
I get along better with adults than with people my own age	0.03	0.39	0.03	-0.01
I have many fears, I am easily scared	0.00	0.19	0.38	0.06
I finish the work I'm doing. My attention is good	0.05	0.46	-0.06	-0.13

Notes: * denotes items that are meant to capture negative traits (difficulties) in English but in the Spanish translation may have been interpreted as positive by the respondents.

APPENDIX F: TRACKING AND ATTRITION CORRECTION

In the 2010 survey, we placed special emphasis on tracking all migrants and other difficult-to-find individuals. In the first phase of the survey, lasting about 6 months, we interviewed individuals and households located in or nearby their original localities. We refer to this period as the "regular tracking" phase. This was followed by an "intensive tracking" phase, lasting approximately 1.5 years, during which we made exhaustive efforts to find all individuals not found during regular tracking, through repeat visits to original locations and tracking to any location in Nicaragua or Costa Rica.

We distinguish between two sets of outcomes, those collected in the household instrument and those from the individual instrument. Outcomes collected in the household instrument include educational attainment and all labor market outcomes, self-reported by the individual or—in cases when he was temporarily absent—the household head or spouse of the household head. Outcomes collected in the individual instrument include all achievement and cognitive tests, and socio-emotional outcomes. Attrition rates are different for the two sets of outcomes as the latter required direct contact with the individual. Only 10 percent of the specific cohort of boys examined in this paper could not be tracked for the household instrument, and 19 percent for the individual instrument.[40] We calculate two sets of attrition-correction weights, one for each survey instrument.

[40] A small number of individuals (15) in the target age group were deceased by 2010. They are not used to predict the probability of attrition as selection for them is clearly driven by other factors. Including the deceased individuals, final attrition rates are 12 and 20 percent respectively for the household and individual instruments.

This appendix briefly describes the methods used to construct the attrition-correction weights. See Molina Millan and Macours (2017) for a more detailed explanation and rationale for the selection correction, as well as for further insights into the nature of the selective attrition. Several baseline characteristics are correlated with the probability of being found during regular tracking, and these correlations differ for the early versus late treatment groups. Appendix Table F1 shows that prior to the intensive tracking phase, the number of baseline characteristics that are not balanced is higher (columns 3 and 7) than what is to be expected by chance but baseline balance was restored through the intensive tracking (columns 5 and 9).[41] This holds for both household- and individual-level data. Intensive tracking proved important for internal validity. Those found during the intensive tracking phase differ in observed baseline characteristics from those found during the regular tracking phase. Moreover, individuals that were difficult to find, but were ultimately found, are more similar on baseline observable characteristics to those that were ultimately not found. Among other differences, those that were difficult to find had lower compliance rates with the program, suggesting that attrition correction is important to account for heterogeneity in the ITT effects.

To correct for the remaining attrition, we use a modified version of the standard inverse probability weighting adjustment, putting more weight on individuals who were more difficult to find, because the baseline data show they more closely resemble those we could not interview. The basic assumption underlying the estimation strategy is that the probability of being found during the intensive tracking phase is explained by observable characteristics. Overweighting individuals whose observed characteristics predict they were difficult to find corrects for the

[41] The few variables that are not balanced in the final 2010 sample (column 5 and 9) are similarly not balanced on the full baseline target sample (column 1).

sample selection. We assign a weight of 1 to individuals found during the regular tracking phase, and estimate the weight only for those found during the intensive tracking phase. To determine the weights, we estimate the probability of being found among those tracked during the intensive phase. Observable characteristics are much better predictors for this subsample than for the full sample, confirming that selection on observables for this subsample is a more plausible assumption.

A wide set of socio-economic variables observed in the program census was considered as potential predictors of attrition, informed by the nature of migration in the region. This includes all of the baseline variables shown in the balance table for the sample in Appendix Table A1 capturing individual characteristics, household characteristics (parental education, demographics, economic activities, and assets), and locality characteristics. As connectivity could be a good predictor of tracking success, we also include two variables to capture the social network of the individual (village size and family network size), some more detailed household structure variables, and a set of proxy variables meant to capture the possible temporary nature of some households' residence in the village at baseline.[42] We similarly add locality characteristics that could be pull or push factors for migration: remoteness (measured using distance to night light and altitude), location in coffee producing region and having been affected by hurricane Mitch. Finally, as a proxy for locations with more temporary workers, we also use two variables capturing the level of attrition between the census and the first baseline survey (i.e., between May and August of 2000) in the locality of origin of the individual, the share of individuals

[42] In particular, we include a set of indicators as proxies for whether the household comprised temporary workers on one of the large coffee plantations (haciendas). These households were captured in the program census but likely were not permanent residents.

attrited, and whether anybody attrited.[43] Individuals from such locations not only were more likely to attrit, but also could be harder to trace, as contacts with the community of origin could be limited.

As there are a wide range of observed characteristics to consider, and because there are relatively few individuals not found after intensive tracking, we follow Doyle et al. (2016) to select a reduced set of predictors. Separate estimations are carried out to model attrition for the household and individual instruments. We first estimate bivariate regressions in which each potential predictor was examined to determine whether a significant difference existed between the means for those found and not found during intensive tracking. All estimates use the survey sample weights and standard errors are clustered at the locality level. The testing was conducted separately for the early and late treatment groups. Results are shown in first four columns of Appendix Table F2 for attrition in the household instrument, and in final four columns for the individual instrument. The correlates of being found during intensive tracking are not the same for both treatment groups and also differ between the individual and the household instruments.

Any indicator found to be statistically significantly different for the early or late treatment group was retained as a potential predictor. We then estimate the probability of being found on this set of baseline predictor variables for each experimental group on the sample of those not found during regular tracking. To account for collinearity between measures, the baseline predictor set was restricted further by conducting stepwise selection of variables with backward elimination and using the adjusted R^2 as the information criterion. The strata and regional fixed effects, as well as 6-monthly age dummies were included as fixed predictors in all models.

In the final step, we estimate the probability of being found during the intensive tracking

[43] The baseline survey was conducted right after the public lottery and before the start of the transfers.

phase for both early and late treatment group together, keeping only the predictors as indicated by the stepwise procedure, as well as the strata and regional fixed effects and the 6-monthly age dummies, all interacted with the treatment variable. In addition, we add a set of field supervisor fixed effects, to capture potential differences between survey teams in effectiveness in tracking. The resulting regression has good predictive power (the linear probability model has an R^2 of 51 percent for household level attrition, and 37 percent for individual level attrition). Appendix Table F3 shows the linear probability model estimates.

The probability of being found during intensive tracking (conditional on not having been found during regular tracking) using a probit estimation for this last specification is then used to determine the weights for the attrition selection. All observations found during regular tracking retain a weight of 1, while those found during the intensive tracking obtain a weight 1/prob (found during intensive tracking). Finally, these weights are then multiplied with the sample weights. Final attrition-correction weights vary between 1 and 35 for the household instrument, and 1 and 86 for the individual instrument.

TABLE F1. COMPARISON OF BASELINE BALANCE BY TRACKING STATUS AND 2010 SURVEY INSTRUMENT

	Baseline Sample: Cohort 9-12 Boys		Household Instrument Outcomes				Individual Instrument Outcomes			
			Found During Regular Tracking		Final Sample		Found During Regular Tracking		Final Sample	
	N=1138		N=826		N=1006		N=527		N=907	
	Coef.	SE	Coef.	SE	Coef.	SE	Coef.	SE	Coef.	SE
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Individual Characteristics										
Age at start of transfer in months	-0.046*	(0.026)	-0.086*	(0.047)	-0.068**	(0.032)	-0.129*	(0.068)	-0.052	(0.041)
No grades attained (=1)	-0.008	(0.064)	-0.025	(0.065)	-0.014	(0.063)	-0.018	(0.073)	-0.016	(0.065)
Highest grade attained	0.071	(0.16)	0.136	(0.16)	0.094	(0.15)	0.173	(0.18)	0.111	(0.16)
Worked in last week (=1)	-0.054*	(0.027)	-0.064**	(0.029)	-0.064**	(0.030)	-0.039	(0.038)	-0.060**	(0.029)
Participated in some economic activity (=1)	-0.009	(0.035)	-0.016	(0.037)	-0.014	(0.035)	0.009	(0.042)	-0.008	(0.037)
Household Characteristics: Education										
Distance to nearest school (minutes)	0.744	(4.76)	2.547	(4.56)	0.594	(4.54)	4.308	(5.91)	1.153	(4.31)
Household head no grades attained (=1)	0.005	(0.031)	0.005	(0.040)	0.020	(0.035)	0.024	(0.034)	0.036	(0.036)
Household head 3 plus grades attained (=1)	0.025	(0.029)	0.046	(0.031)	0.032	(0.027)	0.060*	(0.033)	0.033	(0.026)
Mother no grades attained (=1)	-0.046	(0.037)	-0.097**	(0.042)	-0.053	(0.043)	-0.100*	(0.053)	-0.040	(0.045)
Mother 3 plus grades attained (=1)	0.072	(0.044)	0.111***	(0.038)	0.068	(0.041)	0.171***	(0.052)	0.057	(0.042)
Household Characteristics: Demographics										
Father not living in same household (=1)	0.017	(0.031)	0.041	(0.029)	0.013	(0.027)	0.092**	(0.040)	0.011	(0.025)
Mother not living in same household (=1)	0.011	(0.017)	0.027	(0.017)	0.013	(0.017)	0.022	(0.018)	0.015	(0.019)
Child of household head (=1)	-0.016	(0.024)	-0.061**	(0.024)	-0.020	(0.023)	-0.074**	(0.031)	-0.023	(0.024)
Number of children of household head	-0.241	(0.22)	-0.511**	(0.23)	-0.326	(0.23)	-0.487*	(0.29)	-0.311	(0.23)
Female household head (=1)	0.030	(0.020)	0.035*	(0.018)	0.018	(0.018)	0.053*	(0.026)	0.019	(0.018)
Age of household head	0.339	(0.87)	0.556	(0.87)	0.473	(0.79)	0.290	(1.11)	0.843	(0.85)
Number of household members	-0.047	(0.19)	-0.205	(0.24)	-0.159	(0.22)	0.028	(0.27)	-0.089	(0.22)
Nuclear household (=1)	-0.017	(0.040)	-0.023	(0.044)	-0.019	(0.043)	-0.071	(0.052)	-0.038	(0.044)
Multigenerational household (=1)	-0.037	(0.035)	-0.005	(0.034)	-0.020	(0.037)	0.023	(0.039)	-0.009	(0.038)

Other household structure (=1)	0.054*	(0.028)	0.028	(0.031)	0.038	(0.028)	0.047	(0.038)	0.047	(0.032)
Number of children aged 0-8	0.021	(0.11)	-0.067	(0.12)	-0.097	(0.11)	0.095	(0.14)	-0.087	(0.12)
Number of children age 9 to 12	-0.036	(0.058)	-0.078	(0.070)	-0.072	(0.062)	-0.087	(0.076)	-0.066	(0.068)
Household Characteristics: Economic Activities & Assets										
Household head main occupation is agric. (=1)	-0.021	(0.038)	-0.003	(0.046)	-0.018	(0.041)	0.004	(0.049)	-0.018	(0.042)
Size of landholdings (1000 square meters)	-1.987	(1.85)	-2.822	(2.02)	-2.500	(1.80)	-3.730	(2.28)	-1.656	(1.82)
Log of size of landholdings	-0.270	(0.41)	-0.049	(0.43)	-0.087	(0.42)	0.144	(0.60)	0.013	(0.44)
Number of parcels of land	-0.036	(0.073)	-0.018	(0.076)	-0.008	(0.074)	0.010	(0.094)	0.017	(0.073)
Log predicted expenditures (per capita)	-0.009	(0.022)	0.029	(0.026)	0.003	(0.025)	0.009	(0.034)	0.001	(0.025)
Wealth index - housing characteristics	0.200	(0.16)	0.328**	(0.16)	0.223	(0.15)	0.226	(0.17)	0.159	(0.15)
Wealth index - productive assets	-0.258**	(0.11)	-0.209*	(0.12)	-0.263**	(0.11)	-0.188	(0.15)	-0.201*	(0.11)
Wealth index - other assets	-0.040	(0.18)	-0.127	(0.19)	-0.059	(0.18)	-0.123	(0.21)	-0.067	(0.18)
Number of rooms in house	0.030	(0.074)	0.067	(0.086)	0.051	(0.074)	0.022	(0.089)	0.021	(0.075)
Cement block walls (=1)	0.044	(0.043)	0.065	(0.044)	0.036	(0.043)	0.043	(0.058)	0.007	(0.042)
Zinc roof (=1)	0.050	(0.085)	0.020	(0.082)	0.052	(0.083)	0.020	(0.090)	0.057	(0.085)
Tile roof (=1)	-0.093	(0.097)	-0.047	(0.097)	-0.094	(0.095)	-0.042	(0.10)	-0.098	(0.096)
Dirt floor (=1)	-0.030	(0.036)	-0.049	(0.038)	-0.039	(0.036)	-0.017	(0.045)	-0.016	(0.038)
Latrine or toilet (=1)	0.035	(0.063)	0.083	(0.067)	0.062	(0.065)	0.103	(0.071)	0.073	(0.065)
Electric light (=1)	0.074	(0.050)	0.116**	(0.057)	0.084	(0.052)	0.120**	(0.059)	0.078	(0.054)
Radio (=1)	0.020	(0.034)	0.010	(0.040)	0.005	(0.040)	-0.027	(0.042)	0.004	(0.041)
Work animals (=1)	-0.0541*	(0.029)	-0.026	(0.030)	-0.0573*	(0.031)	-0.046	(0.043)	-0.048	(0.031)
Fumigation sprayer (=1)	-0.073	(0.047)	-0.062	(0.058)	-0.073	(0.050)	-0.027	(0.056)	-0.053	(0.050)
Village Characteristics										
Village affected by hurricane Mitch (=1)	-0.062	(0.050)	-0.062	(0.051)	-0.059	(0.053)	-0.016	(0.057)	-0.039	(0.049)
Altitude of village	-21.10	(34.6)	-25.69	(35.9)	-22.51	(35.9)	-26.79	(37.0)	-24.49	(36.9)
Village in coffee producing region (=1)	-0.018	(0.065)	-0.021	(0.066)	-0.010	(0.066)	-0.008	(0.065)	-0.008	(0.066)
Distance to night light (meters)	2276	(2642)	1291	(2619)	1880.0	(2607)	2730	(2758)	1980	(2645)
Live in Tuma region (=1)	0.193	(0.15)	0.155	(0.14)	0.194	(0.14)	0.162	(0.15)	0.187	(0.14)
Live in Madriz region (=1)	0.017	(0.12)	0.044	(0.12)	0.027	(0.13)	0.021	(0.12)	0.017	(0.13)
Social Capital										
Family network size (individuals)	23.52**	(10.3)	27.65**	(12.1)	25.94**	(10.9)	19.20	(14.1)	26.17**	(11.7)
Population size village	257.6***	(88.6)	249.0**	(94.6)	257.2***	(88.7)	250.1**	(101)	242.1***	(86.9)

Proxy's of Permanent Residence in Village										
Own house (=1)	-0.052	(0.043)	-0.021	(0.046)	-0.064	(0.043)	-0.008	(0.041)	-0.048	(0.041)
House is obtained in exchange for service/labor (=1)	0.019	(0.039)	-0.016	(0.032)	0.023	(0.036)	-0.014	(0.033)	0.011	(0.037)
Address in hacienda (=1)	-0.022	(0.058)	-0.023	(0.058)	-0.004	(0.056)	-0.028	(0.059)	-0.007	(0.054)
Address in hacienda & house rented (=1)	0.032	(0.033)	0.018	(0.031)	0.047	(0.031)	0.017	(0.022)	0.038	(0.028)
Variables Indicating Very Early Attrition										
Probability of attrition prior to program start in locality	-0.081**	(0.033)	-0.083**	(0.034)	-0.079**	(0.034)	-0.070**	(0.035)	-0.075**	(0.033)
Nobody of target sample attrited before program start	0.100	(0.15)	0.121	(0.15)	0.095	(0.15)	0.101	(0.14)	0.104	(0.15)

Notes: *** p<0.01, ** p<0.05, * p<0.10. Standard errors are clustered at the locality level and given in parentheses. All estimates control for strata fixed effects. Regressions are weighted to account for sampling providing population estimates. Distance to night light (meters) is linear distance from household to an area with stable night light detected by a satellite (DMSP-OLS Nighttime Lights). House received in exchange for services is an indicator variable for households who received the house in exchange for labor services. Address in hacienda is an indicator for households whose address refers to a location on a large plantation (hacienda).

TABLE F2. CORRELATES OF THE PROBABILITY OF BEING FOUND DURING THE INTENSIVE TRACKING PHASE

	Household Instrument				Individual Instrument			
	Treatment N=160		Control N=137		Treatment N=311		Control N=300	
	Coef. (1)	SE (2)	Coef. (3)	SE (4)	Coef. (5)	SE (6)	Coef. (7)	SE (8)
Individual Characteristics								
Age at start of transfer in months	-0.446**	(0.18)	-0.382	(0.23)	-0.375***	(0.12)	-0.421***	(0.091)
No grades attained (=1)	-0.122	(0.16)	-0.071	(0.12)	-0.209**	(0.084)	-0.147**	(0.055)
Highest grade attained	0.042	(0.44)	-0.074	(0.37)	0.220	(0.29)	0.054	(0.18)
Worked in last week (=1)	-0.038	(0.066)	0.043	(0.084)	-0.029	(0.052)	0.047	(0.047)
Participated in some economic activity (=1)	-0.025	(0.076)	0.003	(0.069)	-0.009	(0.058)	0.029	(0.048)
Household Characteristics: Education								
Distance to nearest school (minutes)	-4.960**	(2.33)	1.910	(7.59)	-11.18**	(4.85)	-8.546	(5.25)
Household head no grades attained (=1)	0.067	(0.13)	-0.114	(0.11)	0.030	(0.11)	-0.128**	(0.061)
Household head 3 plus grades attained (=1)	0.001	(0.068)	-0.011	(0.088)	0.049	(0.054)	0.054	(0.068)
Mother no grades attained (=1)	0.323**	(0.14)	0.220	(0.14)	0.143	(0.086)	0.027	(0.073)
Mother 3 plus grades attained (=1)	-0.177	(0.15)	0.074	(0.099)	-0.115	(0.091)	0.112*	(0.056)
Household Characteristics: Demographics								
Father not living in same household (=1)	-0.190**	(0.068)	-0.008	(0.12)	-0.179***	(0.052)	-0.034	(0.057)
Mother not living in same household (=1)	-0.049	(0.084)	-0.001	(0.085)	0.003	(0.052)	-0.022	(0.029)
Child of household head (=1)	0.123	(0.082)	-0.068	(0.092)	0.058	(0.053)	0.014	(0.032)
Number of children of household head	0.658	(0.56)	0.711	(0.55)	0.134	(0.21)	0.302	(0.28)
Female household head (=1)	-0.133**	(0.051)	0.025	(0.063)	-0.083**	(0.035)	0.017	(0.040)
Age of household head	4.009	(2.69)	4.487*	(2.56)	2.987*	(1.51)	-0.579	(1.86)
Number of household members	-0.134	(0.73)	1.054	(0.67)	-0.472	(0.33)	0.078	(0.39)
Nuclear household (=1)	0.002	(0.11)	-0.022	(0.11)	0.062	(0.070)	0.102	(0.090)
Multigenerational household (=1)	0.169*	(0.085)	0.065	(0.091)	0.016	(0.084)	-0.069	(0.078)
Other household structure (=1)	-0.171*	(0.088)	-0.043	(0.052)	-0.077	(0.063)	-0.033	(0.048)
Number of children aged 0-8	-0.748	(0.59)	0.649*	(0.33)	-0.804***	(0.23)	0.107	(0.16)
Number of children age 9 to 12	-0.183	(0.20)	0.167	(0.12)	-0.110	(0.13)	0.023	(0.087)

Household Characteristics: Economic Activities & Assets								
Household head main occupation is ag. (=1)	0.018	(0.13)	0.080	(0.063)	0.017	(0.065)	0.0493	(0.059)
Size of landholdings (1000 square meters)	-5.661	(10.7)	-2.630	(2.46)	-2.683	(5.69)	-5.810**	(2.29)
Log of size of landholdings	1.342	(1.68)	-0.055	(0.55)	1.394	(0.96)	0.392	(0.60)
Number of parcels of land	0.288	(0.19)	0.025	(0.080)	0.199	(0.13)	-0.047	(0.084)
Log predicted expenditures (per capita)	-0.070	(0.079)	-0.116	(0.10)	0.007	(0.051)	-0.040	(0.056)
Wealth index - housing characteristics	-0.402	(0.42)	-0.327	(0.39)	-0.631	(0.41)	-0.338	(0.22)
Wealth index - productive assets	0.180	(0.21)	0.439**	(0.18)	0.372*	(0.19)	0.117	(0.12)
Wealth index - other assets	-0.164	(0.17)	-0.277	(0.20)	-0.484***	(0.17)	-0.287*	(0.14)
Village Characteristics								
Village affected by hurricane Mitch (=1)	-0.022	(0.065)	-0.106*	(0.054)	0.0339	(0.073)	-0.051*	(0.026)
Altitude of village	-16.20	(44.1)	-16.76	(52.2)	-0.0909	(0.081)	-0.049	(0.054)
Village in coffee producing region (=1)	0.025	(0.094)	-0.071	(0.081)	-27.87	(38.4)	-5.22	(22.7)
Distance to night light (meters)	-3017**	(1288)	-1407	(1707)	-5302***	(1486)	-2108	(1512)
Live in Tuma region (=1)	-0.161**	(0.068)	-0.339***	(0.098)	-0.387***	(0.076)	-0.335***	(0.098)
Live in Madriz region (=1)	0.106	(0.061)	0.087	(0.063)	0.160**	(0.074)	0.086	(0.056)
Social Capital								
Family network size (individuals)	31.65**	(11.5)	17.98**	(6.86)	41.72***	(11.3)	15.41**	(7.01)
Population size village	-3.007	(80.2)	-40.26	(52.0)	-142.6**	(68.0)	-68.24	(43.1)
Proxy's of Permanent Residence in Village								
Own house (=1)	-0.046	(0.13)	0.230***	(0.078)	0.089	(0.083)	0.110	(0.074)
House is obtained in exchange for service/labor (=1)	0.062	(0.12)	-0.114**	(0.052)	-0.094	(0.085)	-0.073	(0.055)
Address in hacienda (=1)	0.081	(0.11)	-0.128	(0.10)	-0.020	(0.067)	-0.096	(0.085)
Address in hacienda & house rented (=1)	0.065	(0.12)	-0.172	(0.10)	-0.042	(0.069)	-0.094	(0.060)
Variables Indicating Very Early Attrition								
Probability of attrition prior to program start in comarca	-0.005	(0.012)	-0.034	(0.028)	-0.032**	(0.013)	-0.047*	(0.025)
Nobody of target sample attrited before program start	0.077	(0.078)	0.221**	(0.090)	0.283***	(0.085)	0.188**	(0.082)

Notes: *** p<0.01, ** p<0.05, * p<0.10. Standard errors are clustered at the locality level and given in parentheses.

TABLE F3. LINEAR PROBABILITY ESTIMATES FOR PROBABILITY OF BEING FOUND DURING INTENSIVE TRACKING PHASE

	Household Instrument		Individual Instrument	
	Coef. (SE)	Coef. ET Interaction (SE)	Coef (SE)	Coef. ET Interaction (SE)
Early treatment (ET)=1		-0.581		-0.347
		(0.38)		(0.30)
No grades attained (=1)			-0.097	-0.014
			(0.073)	(0.12)
Distance to nearest school (minutes)			0.000	-0.001
			(0.001)	(0.001)
Mother no grades attained (=1)	0.159*	0.022		
	(0.094)	(0.13)		
Mother 3 plus grades attained (=1)			-0.032	-0.0402
			(0.061)	(0.078)
Household head no grades attained (=1)			-0.103*	0.140*
			(0.054)	(0.082)
Father not living in same household (=1)	-0.144	0.134		
	(0.17)	(0.19)		
Child of household head (=1)	0.0501	0.209		
	(0.18)	(0.22)		
Female household head (=1)	0.374***	-0.506***	0.039	-0.196*
	(0.12)	(0.18)	(0.088)	(0.10)
Other household structure (=1)	0.426***	-0.434***		
	(0.12)	(0.15)		
Number of children aged 0-8	0.072***	-0.131***	0.008	-0.058**
	(0.022)	(0.036)	(0.017)	(0.022)
Number of parcels of land	-0.168**	0.303***		
	(0.066)	(0.092)		
Size of landholdings ('000 sq meters)			-0.002	0.003
			(0.002)	(0.002)
Wealth index - housing characteristics			-0.035	0.007
			(0.023)	(0.033)
Wealth index - productive assets	0.058	-0.033	-0.005	0.046
	(0.035)	(0.060)	(0.033)	(0.056)
Village affected by hurricane Mitch (=1)	-0.278**	0.163	-0.104*	0.208*
	(0.12)	(0.15)	(0.053)	(0.11)
Distance to night light (km)			0.007*	-0.009

			(0.004)	(0.007)
Live in Tuma region (=1)	-0.283***	0.379***	-0.276***	0.246
	(0.082)	(0.11)	(0.079)	(0.15)
Live in Madriz region (=1)	0.179	-0.080	0.162**	-0.257**
	(0.13)	(0.19)	(0.070)	(0.10)
Population size village	0.001	0.001	0.000	0.001*
	(0.001)	(0.001)	(0.000)	(0.001)
Own house (=1)	0.371**	-0.331*		
	(0.15)	(0.18)		
House is obtained in exchange for service/labor (=1)	0.407***	-0.356*		
	(0.12)	(0.21)		
Address in hacienda & house rented (=1)	-0.231	0.427	-0.204	0.306*
	(0.23)	(0.27)	(0.16)	(0.18)
Probability of attrition prior to program start in comarca			-0.133	-0.475
			(0.31)	(0.97)
Nobody of target sample attrited before program start	0.120	-0.075	0.043	0.006
	(0.11)	(0.15)	(0.11)	(0.16)
Age fixed effects	YES	YES	YES	YES
Strata fixed effects	YES	YES	YES	YES
Supervisor fixed effects	YES		YES	
Observations	297		611	
R-squared	0.51		0.37	

Notes: *** p<0.01, ** p<0.05, * p<0.10. Standard errors are clustered at the locality level and in parentheses. First column shows coefficient of variable alone, second column coefficient of the variable interacted with the early treatment dummy.

APPENDIX G: SPILLOVERS AND FAMILY NETWORKS

Spillovers—One reason the experimental estimates may underestimate absolute treatment effects is that there could be positive spillovers between the early and late treatment groups. As detailed in the description of the evaluation (Section II.B), a set of 42 rural localities in six municipalities were randomized into early or late treatment. Many of these shared geographic borders (with no buffer zone), including some with the opposite treatment group, opening up the possibility of spillover effects. In this appendix, we describe the methodology we use to test for spillover effects exploiting GPS information on initial household location.

In contrast to a more typical treatment and control design, in which one might expect spillovers in one direction only (positive or negative), the phase-in design of early and late treatment implies that the direction of spillovers may vary over time. For example, if there were positive spillovers from treatment to non-treatment areas during program operation, then there could have been spillovers from early to late treatment areas in the first years of the program when only early treatment households received transfers, followed by spillovers from late to early treatment areas in later years when only late treatment households received transfers. Even if all spillover effects were positive, the net effect of spillovers on early treatment (measured after 10 years), and therefore on the differential between early and late treatment, would be ambiguous.

To explore the potential influence of spillover effects on the main differential findings, we exploit GPS location data for sample households and analyze whether differential program effects differ depending on whether households were living in close proximity to many other households receiving the program during early treatment.[44] To do so, we use the variation in the density of households with early treatment in a 3-kilometer radius around each household, which results from the locality-level randomization, and the fact that the geographical borders of the locality (based on census segments) do not necessarily correspond to village borders. The density itself is likely to capture other unobserved characteristics of the environment, but the interaction of the density with early treatment provides an estimate that accounts for both spillovers, as well as for any potential multiplier effects among early treatment households. The possibility of the latter, in areas where treatment density was high, is consistent with evidence from contemporaneous qualitative work that demonstrated strong united enthusiasm for improved schooling during the CCT, evidenced by community mobilization to ensure additional teachers if needed (Adato and Roopnaraine 2004).

Results in Appendix Table G1 suggest that treatment effects appear to be larger when early treatment density is higher but only one of the interactions, for the socio-emotional family, is significant at 10 percent. If anything, the signs of the interaction effects do not suggest differential effect estimates are underestimated due to spillovers, and they are instead consistent with the possibility of a positive multiplier effect within the early treatment group.

Family Kinship Networks—Family networks have been shown to be important for anti-poverty programs in many contexts. Moreover, heterogeneous treatment effects related to such networks can shed light on the mechanisms underlying impacts. Building on the literature

[44] Initial baseline GPS locations are available only for households targeted for 2010 follow-up, and not for all households in the 2000 program census. We use the observed program density in the sample as a proxy for actual program density in the population.

examining the importance of family and broader kinship networks for impacts of CCTs in the short run (Angelucci and Di Giorgio 2009; Angelucci et al. 2010), we analyze to what extent program effects differ depending on the strength of pre-existing family networks.

The program census carried out in May 2000 included all households in both early and late treatment localities, and administered individual rosters for each household. In addition to the demographic and schooling information for individuals, for the purpose of program administration the census also recorded and digitized first and last names of all household members. In many parts of Latin America including Nicaragua, individuals have two surnames and naming conventions follow the pattern in which a child typically takes the first surname of his father as his first surname, and the first surname of his mother as his second surname. For example, if the paternal surname is Godoy Sandino (F1, f1) and the maternal surname is Darío Martí (F2, f2), then their child's surname would be Godoy Darío (F1, F2).

In the spirit of the approach taken for constructing family networks in Mexico in analyses of PROGRESA (Angelucci and De Giorgi 2009; Angelucci et al. 2010; Angelucci, De Giorgi, and Rasul 2016), we use this naming convention to construct different measures of the potential family network in each village based on surname matches. Specifically, we use the program census data to determine, for each boy, the total number of other individuals living in the same village[45] who have at least one surname in common with that boy, regardless of its position (i.e., first or second surname). Alternatives to this definition, for example requiring the same position (first or second) or using the name of the household head instead, lead to quantitatively smaller

[45] We use village, as defined in the 1995 National Census, as the relevant geographic area rather than the larger locality area used for randomization and clustering in the evaluation, as the former constitutes a more concentrated settlement and is arguably a more natural geographic area in which to consider network effects.

networks that are highly correlated with our broader definition. While we cannot verify that all individuals within the family network as defined here would have a direct familial relationship, many likely do (including siblings and half-siblings, cousins, grandfathers, and aunts/uncles). Median family network size in these rural communities is nearly 60 persons (average 83) and the median family network as a share of the total number of individuals in the community is nearly 20 percent (average 26 percent). We transform the family network size into a binary variable indicating above and below median network size, in part to mitigate concerns about measurement error for those with the same name who do not have direct family relationships (Angelucci et al. 2010).

We use these additional variables to analyze heterogeneous treatment effects. Results for grades attained and the families of outcomes presented in the paper are shown in Appendix Table G2. With the additional variables, the interpretation on the treatment variable is now the ITT differential impact for boys with below median family network size at baseline and the interaction captures differences in that impact for those with initially high family network size. There is little evidence of heterogeneous treatment effects, though for the education and socio-emotional families the point estimates hint at somewhat smaller effects on those outcomes for boys with initially larger family networks. Assuming the lack of differential effects is not due only to low power,[46] we conclude that the long-term program effects were not concentrated among those with initially stronger or weaker family networks.

[46] For comparison, Angelucci et al. (2010) have a male sample four times as large.

TABLE G1: 2010 DIFFERENTIAL EXPERIMENTAL IMPACTS, BY EARLY TREATMENT DENSITY

	Education		Learning Family Z-Score	Economic Participation Family Z-Score	Earnings Family Z-Score		Socio-emotional Family Z-Score
	Grades Attained	Family Z-Score			Absolute (5% Trim)	Rank	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
ITT	0.237	0.077	0.039	0.019	-0.305	-0.247	-0.071
	(0.494)	(0.117)	(0.197)	(0.209)	(0.187)	(0.161)	(0.135)
Early treatment density	-0.789	-0.344**	-0.223	0.081	0.233	0.085	-0.318**
	(0.526)	(0.153)	(0.203)	(0.176)	(0.214)	(0.227)	(0.155)
ITT * Early treatment density	0.673	0.286	0.350	0.206	0.434	0.481	0.399*
	(0.801)	(0.203)	(0.333)	(0.283)	(0.290)	(0.289)	(0.215)
Observations	1,006	1,007	907	1,006	997	1,006	900
R-squared	0.426	0.349	0.450	0.148	0.107	0.104	0.156

Notes: *** p<0.01, ** p<0.05, * p<0.10. Standard errors are clustered at the locality level and are given in parentheses. Early treatment density is defined as the fraction of early treatment households within a 3 km radius. Regressions are weighted to account for sampling and attrition providing population estimates. Differential ITT results compare early to late treatment groups in 2010. Z-scores are calculated as the mean and divided by the standard deviation of the late-treatment group. Controls include three-month age group indicators and indicator variables for grades attained at baseline, stratification and region.

TABLE G2: 2010 DIFFERENTIAL EXPERIMENTAL IMPACTS, BY FAMILY NETWORK SIZE

	Education		Learning Family Z-Score	Economic Participation Family Z-Score	Earnings Family Z-Score		Socio-emotional Family Z-Score
	Grades Attained	Family Z-Score			Absolute (5% Trim)	Rank	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
ITT	0.295	0.146**	0.147	0.225**	0.177**	0.185*	0.096
	(0.226)	(0.068)	(0.089)	(0.086)	(0.085)	(0.094)	(0.063)
Family network (>median =1)	-0.023	-0.025	0.110	0.119	-0.033	0.001	0.120*
	(0.270)	(0.065)	(0.081)	(0.091)	(0.127)	(0.143)	(0.066)
ITT * Family network (>median =1)	-0.005	-0.083	0.025	-0.022	0.038	0.015	-0.116
	(0.344)	(0.090)	(0.105)	(0.115)	(0.164)	(0.168)	(0.095)
Observations	1,006	1,007	907	1,006	997	1,006	900
R-squared	0.425	0.348	0.452	0.149	0.097	0.097	0.155

Notes: *** p<0.01, ** p<0.05, * p<0.10. Standard errors are clustered at the locality level and are given in parentheses. The late treatment mean for grades attained is 5.5. Family network size is the number of individuals with common surnames in the village. Regressions are weighted to account for sampling and attrition providing population estimates. Differential ITT results compare early to late treatment groups in 2010. Z-scores are calculated as the mean and divided by the standard deviation of the late-treatment group. Controls include three-month age group indicators and indicator variables for grades attained at baseline, stratification and region.

APPENDIX H: NON-EXPERIMENTAL ABSOLUTE PROGRAM EFFECTS —MATCHING

To estimate the absolute effects of the program, we compare 2010 outcomes for boys ages 9–12 at baseline in the early treatment group to the 2010 outcomes for the same cohort of boys in a non-experimental comparison group. In 2002, prior to the phase-in of the late treatment group, 21 non-experimental comparison localities from neighboring rural municipalities were added to enhance the potential for an evaluation of the longer-term effects of the program. The principal criteria for selection included: 1) the same marginality index score cut-offs from the Nicaraguan national census used in the selection of the original 42 localities; 2) minimal ongoing or planned development interventions related to the CCT's objectives; and 3) coverage of the geographic regions of the original municipalities.[47] The comparison group was surveyed in 2002, 2004, and 2010 using the same survey instruments as the experimental groups.

As the ex-ante match of the comparison area to the program areas on locality-level characteristics may not be sufficient to balance household and individual characteristics, we estimate the absolute effects using five nearest neighbors matching (NN5), two nearest neighbors

[47] More specifically, the comparison sample was drawn from rural municipalities adjacent to or neighboring the six original municipalities. Six comparison municipalities without any major planned development initiatives but with similar levels of poverty and density of schools and health clinics were selected to capture the geographic diversity of the original municipalities. After excluding a small number of localities, the same marginality index used to select the original 42 localities (Arcia 1999; Maluccio 2009), and based on the 1995 Nicaraguan National Census, was calculated for each remaining rural locality. From this exercise, 22 localities with marginality scores in the range targeted by the CCT were identified; one locality that was further way, and thus less likely to be similar, was dropped, leaving 21 comparison localities. A random sample of households was drawn in each. For additional details, see IFPRI (2005).

matching (NN2), and non-parametric kernel and local linear matching. We draw on the relatively rich data available and include household and individual characteristics in the propensity score. The matching aims to balance these observable characteristics. To satisfy the unconfoundedness assumption we must assume balance of the unobservables (Rosenbaum and Rubin 1983), as required with any non-experimental method.

The nearest neighbor matching estimators are bias adjusted (Abadie and Imbens 2006; Imbens 2015) and standard errors are clustered at the locality level using the analytic asymptotic variance estimator developed by Abadie and Imbens (2008, 2011) that accounts for the fact that the propensity score is estimated. For kernel and local linear estimates, the standard errors are bootstrapped and the bandwidth is set to be small (0.06) to limit the bias (Todd 2007). We estimate average treatment effects on the treated (ATT) which matches all boys who are 9–12 at the start of the program in the early treatment group with same aged boys in the comparison group. For more direct comparison with the ITT experimental differential results, the early treatment group includes all children who were eligible for treatment, regardless of whether they had taken up the treatment (i.e., it includes the non-compliers).

Estimation of Propensity Score—To estimate the propensity score, we combine data from the 2000 program census, with data from the 2002 household survey for the non-experimental comparison group. There are two important caveats. First, the baseline data used to determine the propensity score for the early treatment and comparison groups are from different years: 2000 for the early treatment group and 2002 for the comparison group. We do not use 2002 data for the early treatment group because this group had already received two years of CCT benefits by 2002. We argue the difference in the timing of the surveys is not likely to be a major source of bias as the value of the variables used in the propensity score are unlikely to have changed much between 2000 and 2002 (e.g., mother's age, mother's and household head's years of education, head's gender, distance to the municipality center).[48]

Second, the data come from different types of survey instruments; census and household surveys.[49] We use the 2000 program census data, rather than the 2000 baseline household survey, in order to include the oversample group in the estimate of the propensity score. The inclusion of the oversample is important for comparability with the differential experimental estimates and also increases the precision of the propensity score estimate. The 2000 program census has a more limited set of variables though all questions in the census and survey instrument are similar for the variables included in the propensity score.

The logit model used to estimate the propensity score is presented in Appendix Table H1. We estimate the propensity score using data on boys who are 9–12 at baseline for both early and late treatment groups and the comparison group. While the estimation of the absolute effects does not include the late treatment group, we include it in the construction of the propensity score to increase the sample size and hence the precision of the propensity score estimates.[50] We use all available variables that are similar between the 2000 program census and 2002 household survey and important predictors of either treatment status or the outcomes of interest. We exclude

---

[48] When appropriate, comparison group data from 2002 was adjusted in order to be consistent with the data for the early treatment group coming from 2000. For example, age of mother and the child were calculated for the same year, 2000.

[49] An important exception is the locality level marginality index, which is based on the 1995 national census for both the early treatment and the comparison group.

[50] Due to the randomization, the distributions of variables at baseline are similar between early and late treatment groups and using both groups in the estimation of the propensity score does not introduce bias.

variables whose values are likely to have changed between 2000 and 2002, binary variables which did not have sufficient variation, and information about fathers that was incomplete (e.g., father's age at baseline was missing for more than 20 percent of the sample because it was only asked if the father was a resident of the same household).[51,52] Because of the two-year gap in measurement, it is not possible to consistently measure baseline education variables at the same point in time. As a result, we do not include grades attained or enrollment in the main propensity score, but do include year of birth fixed effects, which are correlated with the education variables.

Propensity Score Balance—We follow Dehejia and Wahba (1999) to determine if the propensity score is balanced across the non-experimental groups and use initial estimates as guides to include interactions or polynomials of variables in the propensity score. We divide the common support into four blocks and test that the propensity score, and each of the variables in the propensity score, are balanced within each block using a t-test.

Appendix Figure H1 presents the distributions of the estimated individual-level propensity score model. Observations above the x-axis are from the early treatment group, and those below from the comparison group. In contrast to what we might see had the groups been randomly allocated, the overlap, while substantial, is imperfect with the treatment group skewed to the

[51] We use mother's education from the 2010 survey because it is more complete. In the 2010 survey mother's education was collected for all household members, while in the earlier surveys it was collected for all individuals on the household roster, so mother's education was only available if the mother was living in the same household.

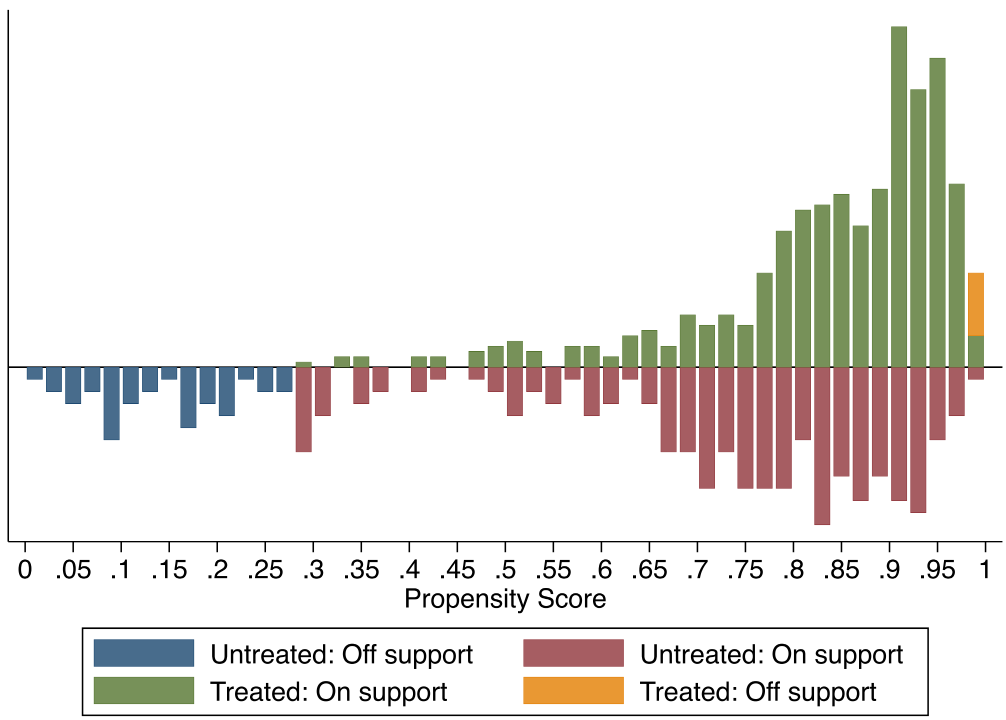
[52] We further do not include the household asset index because not all of the variables are available, however we include the variables that are the same between the two surveys and likely to be correlated with the outcomes.

right and the comparison group to the left. Matching estimators address and correct for this difference in the distributions. To improve the overlap in the covariate distribution, we restrict the analysis to the common support between the experimental group and comparison group. In the main specification reported in the text, we define the common support by trimming all observations that have a propensity score lower than the minimum of the early treatment group distribution, as well as, observations whose propensity score value is greater than the maximum value of the comparison group (i.e. a typical "min-max" common support).

In Appendix Table H2 we test the balance of the matching using the "min-max" trim. Columns 1–4 present the p-value on the t-test of the difference in the propensity score and each of the variables used to make the propensity score between the early treatment and comparison group by block. They indicate that the p-score is balanced between the early treatment and comparison group in all blocks, and that only one variable is statistically different within any block at the 5 percent level or lower (less than 1 percent of the block-variable combinations). We further test whether the early treatment and comparison groups are balanced using the NN5 matching estimator on the variables in the propensity score. Appendix Table H2 columns 5 and 6 show that the baseline variables used in the propensity score are balanced between the early treatment and comparison groups; the estimates are close to zero and none are statistically significant at 10 percent level or below.

Absolute Program Effects—The absolute effects using the matching estimators are reported in Appendix Table H3. We reproduce the NN5 results from Table 10 in panel A, and then a number of robustness checks to examine the stability of these estimates. First, given the sparse distribution of treatment observations (see Appendix Figure H1), we estimate results with the common support defined by trimming all observations that have a propensity score lower than the first and second percentiles (propensity score equal to 0.423 and 0.499, respectively) of the treatment group distribution in panels B and C, rather than the minimum value. Second, we use NN2, non-parametric kernel and local linear matching (panels D, E, and F) to test robustness to different matching estimators. Results are very similar in magnitude and significance regardless of which trim or matching estimator is used.

FIGURE H1—INDIVIDUAL-LEVEL PROPENSITY SCORE DISTRIBUTION,
BOYS 9–12 IN 2000



Notes: The treated include the early treatment groups and the untreated the comparison group.

TABLE H1— LOGIT RESULTS FOR PROPENSITY SCORE MATCHING AT INDIVIDUAL LEVEL, BOYS 9–12 IN 2000

	Logit	OLS
	(1)	(2)
Age 10 in 2000 (=1)	-0.243	-0.028
	(0.237)	(0.029)
Age 11 in 2000 (=1)	0.549**	0.063**
	(0.243)	(0.027)
Age 12 in 2000 (=1)	-0.171	-0.016
	(0.247)	(0.031)
1995 marginality index	521.454***	81.659***
	(66.590)	(8.046)
1995 marginality index squared	-57.673***	-9.024***
	(7.486)	(0.906)
Mom no education (=1)	0.459**	0.054**
	(0.225)	(0.027)
Mom less than 3 years of education (=1)	15.533*	2.538**
	(8.321)	(1.057)
Mothers age in 2000	-0.017	-0.002
	(0.012)	(0.001)
Household head male (=1)	0.108	0.010
	(0.250)	(0.032)
Household head years of education	-0.186***	-0.025***
	(0.056)	(0.007)
Household head has no education (=1)	-1.081***	-0.130***
	(0.280)	(0.032)
Family size	-0.208	0.003
	(0.180)	(0.014)
Family size squared	0.020*	0.001
	(0.011)	(0.001)
Share of household members age 0-13	1.891***	0.240***
	(0.616)	(0.075)
Has electric light (=1)	-0.744***	-0.096***
	(0.202)	(0.025)
Has work animals (=1)	-0.194	-0.022
	(0.227)	(0.027)
Log (km to the municipality capital)	-0.461***	-0.042**
	(0.168)	(0.019)
Mother no ed. * 1995 marginality index	-3.504*	-0.571**
	(1.873)	(0.237)
R squared		0.203
Observations	1,230	1,230

Notes: *** p<0.01, ** p<0.05, * p<0.10. The dependent variable is the treatment variable for matching, which is 1 if in the early treatment and 0 if in the comparison group.

TABLE H2: BASELINE BALANCE — BY BLOCK AND NN5 MATCHING

	P-value of Difference in Means				NN5	
	Block 1	Block 2	Block 3	Block 4	Diff. in mean	P-value
	(1)	(2)	(3)	(4)	(5)	(6)
Propensity score	0.465	0.537	0.274	0.282		
Age 9 (=1)	0.692	0.636	0.967	0.980	0.042	(0.272)
Age 10 (=1)	0.782	0.455	0.400	0.416	-0.057	(0.138)
Age 11 (=1)	0.638	0.296	0.389	0.931	0.039	(0.350)
Age 12 (=1)	0.506	0.724	0.924	0.614	-0.023	(0.612)
Marginality Index	0.838	0.692	0.142	0.059*	-1.584	(0.453)
Mother no grades attained (=1)	0.717	0.282	0.243	0.871	-0.038	(0.454)
Mother 3 plus grades attained (=1)	0.578	0.492	0.268	0.835	-0.051	(0.415)
Mother's age	0.279	0.078*	0.844	0.013**	0.144	(0.917)
Household head male (=1)	0.540	0.491	0.293	0.905	0.003	(0.924)
Household head years of education	0.293	0.282	0.566	0.298	0.110	(0.665)
Household head no grades attained (=1)	0.433	0.369	0.183	0.352	-0.037	(0.598)
Family size	0.911	0.288	0.452	0.286	-0.054	(0.925)
Share of household members age 0-13	0.506	0.747	0.330	0.537	-0.011	(0.413)
Household has electric light (=1)	0.376	0.180	0.069*	0.412	0.075	(0.341)
Household had work animals (=1)	0.933	0.093*	0.713	0.826	-0.055	(0.102)
Log distance to the municipality capital (km)	0.759	0.567	0.306	0.317	-0.117	(0.578)

Notes: *** p<0.01, ** p<0.05, * p<0.10. Compares baseline values from early treatment group (2000) with comparison group (2002). ATT biased adjusted estimator (Abadie and Imbens 2011) using five nearest neighbors. Standard errors on the differences are clustered at the locality level.

TABLE H3: 2010 ABSOLUTE IMPACT —ALTERNATIVE SPECIFICATION

	Education		Learning Family Z-Score	Economic Participation Family Z-Score	Earnings Family Z-Score		Socio-emotional Family Z-Score
	Grades Attained	Family Z-Score			Absolute (5% Trim)	Rank	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: NN5 — Min-Max Trim							
ATT	1.379***	0.380***	0.387***	0.160	0.106	0.066	0.102
	(0.337)	(0.085)	(0.125)	(0.120)	(0.107)	(0.098)	(0.078)
N	690	690	616	690	687	690	613
Panel B: NN5 — 1 Percent Trim							
ATT	1.409***	0.388***	0.388***	0.160	0.108	0.069	0.096
	(0.339)	(0.086)	(0.126)	(0.116)	(0.105)	(0.094)	(0.079)
N	665	665	594	665	662	665	591
Panel C: NN5 —2 Percent Trim							
ATT	1.467***	0.399***	0.400***	0.182	0.107	0.069	0.094
	(0.320)	(0.085)	(0.125)	(0.117)	(0.104)	(0.094)	(0.080)
N	652	652	581	652	649	652	578
Panel D: NN2 — Min-Max Trim							
ATT	1.560***	0.431***	0.455***	0.135	0.112	0.228**	0.127*
	-0.337	-0.078	-0.098	-0.116	-0.098	-0.113	-0.066
N	690	690	616	687	690	690	613
Panel E: Kernel Matching —Min-Max Trim							
ATT	1.447***	0.397***	0.414***	0.165	0.105	0.069	0.121*
	(0.429)	(0.105)	(0.108)	(0.130)	(0.114)	(0.113)	(0.064)
N	690	690	616	690	687	690	613
Panel F: Local Linear Regression Matching — Min-Max Trim							
ATT	1.280***	0.374***	0.377***	0.158	0.110	0.065	0.085
	(0.403)	(0.106)	(0.106)	(0.129)	(0.109)	(0.101)	(0.067)
N	690	690	616	690	687	690	613

Notes: *** p<0.01, ** p<0.05, * p<0.10. Standard errors are clustered at the locality level and are given in parentheses. Absolute effects compare early treatment to comparison group in 2010. Mean of grades attained in the comparison group is 5.3 in 2010 for panel A. Z-scores are calculated as the mean and divided by the standard deviation of the late treatment group.

APPENDIX I: NON-EXPERIMENTAL ABSOLUTE PROGRAM EFFECTS —DOUBLE DIFFERENCE

Over the years 2000 to 2005, the CCT operated at some point in all rural areas of the six municipalities where the randomized evaluation took place, covering over 90 percent of the rural population in these municipalities. This includes of course the 42 rural localities randomized into early and late treatment in the six municipalities. In addition to those 42 localities, the six municipalities include an additional 17 rural localities that were less poor in 1995 according to the marginality index used for stratification. In these other 17 localities, the CCT was offered to 80 percent of the population based on a household-level proxy means targeting model, also for three years, and beginning in late 2001. Consequently, by 2005 the program had been implemented (to modestly different degrees) in all (59) rural localities in the six municipalities. Given this high coverage, it is possible to use national census data to provide additional evidence of absolute program effects with a different identification strategy and over a different period. We present results for education outcomes only, because none of the other outcomes used in the differential analysis are available in the Nicaraguan national census. For this reason, we also include literacy (if someone can read and write) to provide some information about learning.

More specifically, our second non-experimental approach to estimate absolute program effects on educational outcomes uses a double difference model and the two most recent Nicaraguan censuses. Together, the censuses provide repeated cross sections at the individual level, in 1995 before the start of the program and in 2005, the year the program ended. The data include current municipality of residence (and whether urban or rural), as well as municipality of residence 5 years prior to the census administration date, and at birth.[53] For those who moved, however, whether the place of residence is urban or rural is known for current residence, but not past. We calculate double difference impacts using two cohorts of boys (those ages 9–12 in 1990 and in 2000, calculating ages on November 1 as done for the main analyses) by comparing education outcomes in rural areas of the six program municipalities to outcomes in rural areas of the six neighboring municipalities used in the matching. The 9–12 age cohort in 2000 is, of course, the same age cohort examined in the experimental analyses. The underlying identification assumptions for this approach differ from those required for the matching estimators described in Appendix H, and therefore provide complementary evidence. More specifically, we first estimate

$$Y_{imt} = \delta_0 + \delta_1 \; T_{m,t-5} + \delta_2 \; C_t + \delta_3 \; T_{m,t-5} * C_t + \varepsilon_{imt} \qquad (2)$$

Where Y_{imt} is the educational outcome for boy i in municipality m measured in census year t, $T_{m,t-5}$ is an indicator for whether the boy resided in a treatment municipality 5 years prior to the census year, and C_t is one if the child is in the 2005 census. δ_3 yields the double difference estimate of the 5-year effect of the program on Y, which includes grades attained, enrollment, and literacy. Standard errors are robust to heteroskedasticity.

All estimates limit the sample to individuals living in rural areas (the CCT did not operate in urban areas, and educational outcomes there are on average quite different). The main double

[53] While the data also include more detailed location information, changes in the definition and boundaries of census areas between 1995 and 2005 make it impossible to match them across time. Municipality boundaries, however, remained identical.

difference estimation equation takes a first difference between outcomes measured in 2005 for those living in program and non-program comparison municipalities in 2000, and a second difference between outcomes measured in 1995 for those living in program and non-program municipalities in 1990, as indicated by municipality of residence five years prior. These results are presented in Table 11, and reproduced in panel A of Appendix Table I1.

We explore the sensitivity of the main double difference findings to different definitions of comparison municipalities and treatment status. First, we expand the comparison municipalities to include rural areas in all non-program municipalities in the Central Region of Nicaragua (where the program was located) in panel B. Second, in panels C and D we examine whether results differ when we instead use current residence or, separately, birth residence, to determine if the person lived in a program municipality. While there are some differences in point estimates (particularly for enrollment which ranges from 0.027 to 0.075), the different approaches all suggest significant absolute impacts of the program on educational outcomes after five years.

Finally, to provide support for the identifying assumption, we explore common trends and show that the same specification does not suggest any effects on characteristics of the household heads (for households with a boy in the 9–12 cohort using one observation per household) in panel E. We analyze common trends using household heads rather than an older age cohort of children, as older children may still have been influenced by the program, including through migration patterns that are difficult to disentangle using the national census data. As expected, the point estimates are all close to zero.

TABLE I1: 2005 ABSOLUTE IMPACTS ON EDUCATION AND HOUSEHOLD HEAD CHARACTERISTICS

	Grades Attained	Completed Grade 4 =1	Enrolled =1	Read and Write =1	Female Head =1	Age in Years
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Treatment Municipality (5 Years Prior) vs 6 Municipality Comparison Group						
Treatment municipality * 2005 (d3)	0.597***	0.124***	0.037***	0.091***	-	-
	(0.078)	(0.014)	(0.014)	(0.013)		
N	18,399	18,399	18,421	18,403		
Mean comparison group	3.922	0.559	0.456	0.779		
Panel B: Treatment Municipality (5 Years Prior) vs Central Region Comparison Group						
Treatment municipality * 2005 (d3)	0.574***	0.114***	0.075***	0.080***	-	-
	(0.051)	(0.010)	(0.010)	(0.009)		
N	93,712	93,712	93,867	93,709		
Mean comparison group	3.685	0.523	0.398	0.749		
Panel C: Treatment Municipality (Current) vs 6 Municipality Comparison Group						
Treatment municipality * 2005 (d3)	0.646***	0.129***	0.038***	0.091***	-	-
	(0.078)	(0.014)	(0.014)	(0.013)		
N	18,324	18,324	18,348	18,332		
Mean comparison group	3.896	0.556	0.453	0.779		
Panel D: Treatment Municipality (of Birth) vs 6 Municipality Comparison Group						
Treatment municipality * 2005 (d3)	0.552***	0.113***	0.027*	0.084***	-	-
	(0.079)	(0.014)	(0.014)	(0.013)		
N	18,206	18,206	18,232	18,210		
Mean comparison group	3.860	0.547	0.445	0.771		
Panel E: Household Head for 9–12 Boys — Panel A Groups						
Treatment municipality * 2005 (d3)	-0.020	0.014	-	0.005	0.007	0.342
	(0.079)	(0.013)		(0.016)	(0.013)	(0.417)
N	15,192	15,192		15,286	15,292	15,292
Mean comparison group	1.673	0.226		0.507	0.221	46.901

Notes: *** p<0.01, ** p<0.05, * p<0.10. 2005 absolute effects use census data to compare rural areas of program municipalities to rural areas of other municipalities. The mean of the comparison group is for the 6 comparison municipalities in 2005. Heteroskedasticity-robust standard errors are given in parentheses.

Appendix References

Abadie, A., and G. Imbens. 2006. "Large Sample Properties of Matching Estimators for Average Treatment Effects," *Econometrica* 74(1): 235–267.

Abadie, A. and G. Imbens. 2008. "On the Failure of the Bootstrap for Matching Estimators." *Econometrica* 76(6): 1537–1557.

Abadie, A., and G. Imbens. 2011. "Bias-Corrected Matching Estimators for Average Treatment Effects," *Journal of Business & Economic Statistics* 29(1): 1–11.

Adato, M. and T. Roopnaraine. 2004. "A social analysis of the Red de Protección Social. Report submitted to the *Red de Protección Social*." International Food Policy Research Institute. Washington, DC.

Angelucci, M. and G. De Giorgi. 2009. "Indirect Effects of an Aid Program: How Do Cash Transfers Affect Ineligibles' Consumption?" *American Economic Review* 99(1): 486–508.

Angelucci, M., G. De Giorgi, M.A. Rangel, and I. Rasul. 2010. "Family Networks and School Enrolment: Evidence from a Randomized Social Experiment." *Journal of Public Economics* 94: 197–221.

Angelucci, M., G. De Giorgi, and I. Rasul. 2016. "Consumption and Investment in Resource Pooling Family Networks." Unpublished.

Arcia, G. 1999. "*Proyecto de Red de Protección Social: Focalización de la fase piloto*" Report to the Inter-American Development Bank, Washington DC.

Attanasio, O., S. Cattan, E. Fitzsimon, C. Meghir, and M. Codina. 2015. "Estimating the Production Function for Human Capital: Results from a Randomized Controlled Trial in Colombia." *IFS Working Paper* 15/06.

Behrman, J. R., S. W. Parker and P. E. Todd. 2009a. "Medium-Term Impacts of the Oportunidades Conditional Cash Transfer Program on Rural Youth in Mexico." In: S. Klasen and F. Nowak-Lehmann, editors. Poverty, Inequality, and Policy in Latin America, 219–270. Cambridge, United States: MIT Press.

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

Dehejia, R., and S. Wahba. 1999. "Causal effects in nonexperimental studies: Reevaluating the evaluation of training programs". *Journal of the American Statistical Association* 94 (448): 1053–1062.

Doyle, O., C. Harmon, J.J. Heckman, C. Logue, and S.H. Moon. 2016. "Early skill formation and the efficiency of parental investment: A randomized controlled trial of home visiting." *Labour Economics*, forthcoming.

Fernald, L., P. Gertler and L. Neufeld. 2009. "10-Year Effect of Oportunidades, Mexico's Conditional Cash Transfer Programme, on Child Growth, Cognition, Language, and Behaviour: a Longitudinal Follow-up Study." Lancet 374:1997–2005.

Filmer D, and L. Pritchett. 2001. "Estimating Wealth Effects Without Expenditure Data – or Tears: An Application to Educational Enrollments in States of India." *Demography* 38(1): 115–132.

IFPRI. 2005. *Sistema de Evaluación de la Red de Protección Social (RPS) – Mi Familia, Nicaragua: Evaluación de Impacto 2000–04*, Report submitted to the *Red de Protección Social*. International Food Policy Research Institute, Washington, DC. Photocopy.

Imbens, G. 2015. "Matching Methods in Practice: Three Examples," *Journal of Human Resources* 50(2): 373–419.

Imbens, G. W. and Jeffrey M. Wooldridge. 2009. "Recent Developments in the Econometrics of Program Evaluation." *Journal of Economic Literature* 47(1): 5–86.

Laajaj, R. and K. Macours. 2017. "Measuring Skills in Developing Countries." *World Bank Policy Research Paper* No. 8000. Washington, DC, United States: World Bank

Macours, K., N. Schady and R. Vakis. 2012. "Cash Transfers, Behavioral Changes, and Cognitive Development in Early Childhood: Evidence from a Randomized Experiment." *American Economic Journal: Applied Economics* (4)2: 247–273.

Maluccio, J. A. 2009. "Household Targeting in Practice: The Nicaraguan Red de Protección Social." *Journal of International Development* 21(1): 1–23.

Molina Millan, T. and K. Macours. 2017. "Attrition in Randomized Control Trials: Using Tracking Information to Correct Bias." Unpublished.

Paxson, C., and N. Schady. 2010. "Does Money Matter? The Effects of Cash Transfers on Child Health and Development in Rural Ecuador." *Economic Development and Cultural Change* 59(1): 187–230.

Radloff, Lenore S. 1977. "The CES-D Scale: A Self-Report Depression Scale for Research in the General Population." *Applied Psychological Measurement* 1 (3): 385–401.

Regalia, F., and L. Castro. 2007. "Performance-Based Incentives for Health: Demand and Supply-Side Incentives in the Nicaraguan *Red de Protección Social.*" *Center for Global Development Working Paper* No. 119. Washington, DC, United States: Center for Global Development.

Rosenbaum, P. and Rubin, D. 1983. "The Central Role of the Propensity Score in Observational Studies for Causal Effects." *Biometrika* 70: 41–50.

Todd, P. 2007. "Evaluating Social Programs with Endogenous Program Placement and Selection of the Treated." In *The Handbook of Development Economics* edited by. T. Paul Schultz and John A. Strauss, Vol. 4: 3848–3891.

World Bank. 2001. "Nicaragua Poverty Assessment: Challenges and Opportunities for Poverty Reduction," Report No. 20488-NI, The World Bank, Washington, D.C.

World Bank. 2003. "Nicaragua Poverty assessment: Raising Welfare and Reducing Vulnerability," Report No. 26128-NI. Washington DC: The World Bank.