DISCUSSION PAPER SERIES

IZA DP No. 5751

Assessing the Long-term Effects of Conditional Cash Transfers on Human Capital: Evidence from Colombia

Javier E. Baez
Adriana Camacho

May 2011

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

Assessing the Long-term Effects of Conditional Cash Transfers on Human Capital: Evidence from Colombia

Javier E. Baez
World Bank
and IZA

Adriana Camacho
Universidad de los Andes
and CEDE

Discussion Paper No. 5751
May 2011

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

IZA Discussion Paper No. 5751
May 2011

ABSTRACT

Assessing the Long-term Effects of Conditional Cash Transfers on Human Capital: Evidence from Colombia*

Conditional Cash Transfers (CCT) are programs under which poor families get a stipend provided they keep their children in school and take them for health checks. While there is significant evidence showing that they have positive impacts on school participation, little is known about their long-term impacts on human capital. In this paper we investigate whether cohorts of children from poor households that benefited up to nine years from *Familias en Acción*, a CCT in Colombia, attained more school and performed better in academic tests at the end of high school. Identification of program impacts is derived from two different strategies using matching techniques with household surveys, and regression discontinuity design using census of the poor and administrative records of the program. We show that, on average, participant children are 4 to 8 percentage points more likely than nonparticipant children to finish high school, particularly girls and beneficiaries in rural areas. Regarding long-term impact on tests scores, the analysis shows that program recipients who graduate from high school seem to perform at the same level as equally poor non-recipient graduates, even after correcting for possible selection bias when low-performing students enter school in the treatment group. Even though the positive impacts on high school graduation may improve the employment and earning prospects of participants, the lack of positive effects on the test scores raises the need to further explore policy actions to couple CCT's objective of increasing human capital with enhanced learning.

JEL Classification: I24, I25, I28, I38

Keywords: Conditional Cash Transfers, school completion, academic achievement, learning outcomes

Corresponding author:

Javier Baez
The World Bank
1818 H Street, NW
MSC 9-006A
Washington, DC 20433
USA
E-mail: jbaez@worldbank.org

* Special thanks for insightful comments go to Mark Sundberg, Cheryl Gray, Emmanuel Skoufias, John Hoddinott, Jennie Litvack, Ximena del Carpio, Pablo Acosta, Fabio Sanchez, Raquel Bernal, Catherine Rodriguez, Manuel F. Castro and Daniel Mejia as well as various participants of World Bank workshops (Washington, Hanoi, Jakarta y Dhaka) and seminars of the LACEA Netowrk in Inequality and Poverty (Medellin), Banco de la Republica de Colombia and CEDE (Bogota), and GRADE (Lima). Tu Chi Nguyen, Humberto Martinez, Roman A. Zarate and Roman D. Zarate provided excellent research assistance. Special recognition goes to Diego Dorado, Ana Gomez, and Jose F. Arias, from the Colombian Department of National Planning; Julian Mariño, Wilmer Martinez, Susana Ortiz, and Margarita Pena from *Icfes*; Omar Cajiao, Rita Combariza, Juanita Rodriguez, and Hernando Sanchez from *Acción Social*, and other government agencies for their support in providing data and thoughtful comments. We also thank World Bank operational staff who provided assistance, especially Aline Coudouel and Theresa Jones. We acknowledge financial support from the Norwegian Agency for Development Cooperation. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.

1. Introduction

As part of the global efforts to promote universal basic education, a number of programs have been put in place with positive effects on school enrollment and attendance. Existing evidence shows that the use of educational services responds positively to interventions such as school construction, hiring of additional teachers, regular de-worming of children, school feeding, take-home ration schemes, school vouchers, and conditional and unconditional cash transfers. In particular, Conditional Cash Transfers (CCTs), programs that transfer money to poor families contingent on specific education and health behaviors, have been on the rise in recent years. Since 1997, more than 30 countries have adopted CCTs with the goals of reducing poverty and encouraging investments in human capital. A recent review of the impact evaluation literature indicates that all 11 CCTs evaluated against school enrollment and 15 CCTs evaluated against attendance have positive effects (World Bank, 2010). However, this may not be surprising since most of these programs are conditional upon school outcomes.

CCTs' impacts on poor children's school participation are expected to lead to higher educational attainment. If students stay in school and progress, they could accumulate more human capital and enjoy higher future incomes. Additionally if conditionality of attendance to 80 percent of the classes is higher than the average attendance with no program, this might be reflected in stronger educational performance and also higher future productivity. Furthermore, the transfers of CCTs may increase household's disposable income and their spending towards activities that are beneficial for students' learning, such as foods, books and other school supplies. This additional income together with the conditions to keep children in school are also expected to reduce the pressure for eligible children to work so they can spend more time on school-related activities. Finally, the value that the program places on education could be transferred to the families, enhancing their attitude toward the importance of investing in the schooling of children. Positive peer influence that the CCT beneficiaries receive as they attend classes could also encourage them to study harder and pursue higher education.

Nevertheless, although increased children's school enrollment and time in school are important inputs for the formation of human capital, they do not automatically translate into attainment of more education and improved learning outcomes. First, if the school supply remains fixed, schools may get congested due to the rise in enrollment, increasing teacher pupil ratios and overcrowding in the classroom. Second, the marginal children who are brought into school by the transfers and conditions of the program could have lower expected returns to school compared to those already enrolled since they may be, for instance, less motivated, come from lower socio-economic background, and have less capacity or time devoted to school work. Another plausible reason for limited effects on school attainment and performance is that CCTs are often geographically targeted to poor areas where the teaching and school quality may be relatively lower. Despite the substantial amount of work devoted to assessing the educational impacts of CCTs, little is known about their long-term effects on the stock of human capital, i.e. educational attainment and academic performance in early adulthood.

This paper seeks to help fill in this knowledge gap and identify the expected but empirically uncertain link between school participation and educational achievement through an evaluation of the long-term educational impacts of a Conditional Cash Transfer program in Colombia. More specifically, we investigate whether multiple cohorts of children who are covered by *Familias en Acción* (FA) and who have different degrees of program exposure (ranging from one to nine years) complete more years of education –measured by the probability of completing high school– and perform better in a national standardized test at the end of high school. Although these are not necessarily final outcomes since they do not reflect the ultimate educational achievement, they are close determinants of human welfare and economic growth. In addition, we examine whether there is heterogeneity in program impacts by location (urban and rural) and by the gender of child. Finally, this paper explores possible indirect effects on the human capital of older children (18 years and older) who by the rules of the program are ineligible to the transfer but who reside in households with younger participant children.

Identification of program impacts is derived from two different empirical strategies that use a panel of household surveys, a census of the poor and administrative data from the information system of *Familias en Acción*. The first research design employs matching techniques to compare the school completion rates and test scores of different cohorts of children from treatment and control areas that could have finished high school during the program implementation period of 2003-2009 and were interviewed prior to the initiation of the program. The second design exploits variation in assignment to treatment arising from the sharp discontinuity that emerges at the eligibility threshold defined to participate in the program. Household are assigned a poverty index score from a census of poor people, which determines their eligibility into different social programs (including FA) with different thresholds.

We show that the program helps participant children to increase their school attainment by making them more likely to complete high school. Results from the first empirical approach (matching analysis) indicate that program effects vary between 4 and 8 percentage points. The RD approach yields estimates of program effects on school completion that are similar in direction and magnitude. Overall, focusing on the preferred specifications, we estimate that beneficiary children that belong to households near the threshold of eligibility are between 3 and 6.5 percentage points more likely to graduate from high school. There is also evidence of heterogeneity in program impacts, with effects on school completion being larger for girls and beneficiaries in rural areas. Regarding long-term impacts on tests scores, the analysis shows that program recipients who graduate from high school perform at the same level as equally poor non-recipient graduates in Mathematics, Spanish, or the overall test. This result still holds after correcting for possible selection bias when low-performing students enter school in the treatment group. Finally, there is no indication of indirect program effects on the school completion of ineligible older children residing in the same households as participant children. The results are robust to a variety of controls for observable differences between participants and non-participants, the possibility of sorting around the threshold of eligibility or manipulation of the proxy-means test used to allocate the program, misspecification bias, and differences in the accuracy of data that could lead to spurious differences in test registration between treated and control children.

This paper is structured in five chapters in addition to this introduction. The second chapter reviews the relevant literature regarding the impacts of CCTs, including previous evidence on FA, and provides an overview of the program. Chapter three describes the data sources used in the paper. The fourth chapter presents the empirical analysis, including the discussion of the two different research designs, results from the models on program impacts, the heterogeneity of program impacts and the indirect effects of the program on nonparticipating adolescents who live in the same households as participant children. Chapter five discusses robustness checks to ensure that the findings of the analysis are not subject to selection, misspecification, or data matching bias. Chapter six concludes with different interpretations of the results as well s the implications for program beneficiaries and policymakers.

2. Background

2.1 Existing Literature

There is large amount of evidence demonstrating that CCTs encourage households to increase the use of educational services. Impact evaluations of programs implemented in an array of countries including Brazil, Cambodia, Colombia, El Salvador, Honduras, Jamaica, Malawi, Mexico, Nicaragua, Pakistan, and Turkey indicate that, by and large, CCTs lead to immediate increases in school enrollment and attendance (World Bank, 2010). Even though the size of the impacts varies with the features of each program (amount of the transfers, types of conditionalities, target groups, timing of payments), the characteristics of the population (age, gender, school grade, socioeconomic status, location), and the conditions of program areas (school supply and baseline enrollment), the direction of program effects is largely consistent across programs and evaluation methods. A subset of these evaluations also tracked the school progression of participant children relative to control children, relying mostly on data that span two years of initial program participation. The large majority of these studies show positive impacts in indicators such as grade progression, grade repetition, and dropout rates. However, it is important to stress that these effects are more prominent among children in primary education, and say little about the actual accumulation of human capital in later stages of life.

The evidence is scant when it comes to the impacts of CCTs on final (or close to final) outcomes in education. Looking at school completion, existing evidence from Pakistan shows that CCT beneficiaries are more likely to complete secondary school by 4 to 6 percentage points (Alam et al., 2010). Barrera-Osorio et al. (2008) evaluated a pilot version of a CCT program implemented in Bogota (Colombia), called *Subsidios*, and found similar results.[2] As for actual

[2] The *Subsidios* program was implemented in the late 2000´s in Bogotá and targeted vulnerable population classified as *Sisben* level 1 and *Sisben* 2 based on the proxy-means test constructed with information from a census of the poor. The program's transfers and conditions are similar to *Familias en Acción*, which is a nationwide and older program. However, the *Subsidios* program offered two other treatments with a different structure of benefits. One is the "savings treatment," which reserves one third of the bimonthly payment to be given to families as a lump sum at the end of the year, just before enrollment into the subsequent grade level. The other is the "tertiary treatment," which also reserves part of the bimonthly payment, but then pays families a substantially larger amount if students graduate from high school (eleventh grade). Students who continue to tertiary education are eligible to receive that amount one year earlier than those who do not. These two additional treatments may change

school attainment, only evaluations of *Oportunidades* in Mexico (previously known as *Progresa*) has measured the impact on rural adolescents who were old enough to have plausibly completed their schooling after at least five and a half years of benefits. A first study found that children with exposure to the benefits of approximately two years or more achieve about 0.2 grades of additional schooling (Behrman et al., 2005). Subsequent studies that look at the impacts on young adults with longer periods of exposure to the benefits of *Oportunidades* (nearly 10 years) show important increases in grades of schooling achieved by program participants and their labor insertion, but no effects on the proportion of high school graduates going to college (Behrman and Parker, 2008; Freije and Rodriguez, 2008).[3] Results from evaluations of programs in Cambodia and Honduras point to similar effects, yet they are estimated through simulation analysis or on samples of younger children who were still in school (Filmer and Schady, 2009; Glewwe and Olinto, 2004).

Similarly, the evidence of program effects on learning outcomes is limited and somewhat mixed, making it difficult to draw conclusions. Improvements in cognitive development attributed to CCTs have been consistently found only for young children in pre-school and primary education. The existing literature does not find a discernible effect on learning outcomes for older children when tested during the final grades of secondary school. This is probably due partly to practical and empirical difficulties in revisiting treatment and control children long after a program has been implemented. Furthermore, an evaluation of learning could be confounded by selection problems. Evaluations that estimate program effects based on tests given to children in school may be confounded by selection problems because the beneficiary and non-beneficiary children that go to school are probably not comparable. The “marginal child” that attends school thanks to the program may be poorer and of lower ability compared to those already enrolled. Behrman et al. (2000), for example, investigate the effects of *Oportunidades* in Mexico on the academic achievement of children in school, a sample prone to suffer from nonrandom selection. To address this, the analysis reweights the data to align the age and sex distributions of the treatment and control groups and finds that there is no effect on test scores after 1.5 years of

the way families respond to the program in two ways: one is due to the reduction in the amount of cash families receive on a bimonthly basis, therefore limiting their liquidity; another is through the incentives created by linking cash transfers directly to grade progression and matriculation in tertiary education.

[3] It is worth noting that the results of these two papers are not generalizable to the entire population covered by *Oportunidades* given that the samples of analysis only contain non-migrant rural households. Migrants are expected to have higher enrollment in college and better job opportunities, including higher salaries.

exposure to the program. In contrast, evidence based on matching techniques from a scholarship program in Argentina, which operated as a CCT, shows that the program improved student performance as measured by school grades (Heinrich, 2007). Among analyses that do not condition on school enrollment to avoid contamination due to compositional changes, the results show that CCT beneficiaries in secondary school do not do better in academic tests given at home (Behrman et al., 2005 and Behrman and Parker, 2008 for Mexico; Filmer and Schady, 2009 for Cambodia).[4]

Recent evidence specific to the impacts of FA on learning among young students in primary and middle-secondary school provides mixed results. To account for the problem of selection, García and Hill (2009) focus on the impacts on school progression and academic achievement for the students who would have been enrolled in school even in the absence of the program.[5] While fifth graders in the treatment group did better in math and language tests than those in the control group, particularly in rural areas, program effects for ninth graders in both subjects are negative. Yet, the validity of the findings of this paper is limited for at least two reasons. First, if program effects do exist, they are probably difficult to identify due to the lack of enough statistical power in their analysis. Sample sizes for the nonparametric models of the paper are very low, ranging from 100 to 300 observations depending on the age groups. Second, and perhaps more troublesome, the nationally administered tests used by the authors to infer the academic performance of children (known as "*Pruebas Saber*") are only representative at the school rather than at the individual level.[6]

This paper seeks to contribute to the understanding of the effects of CCTs on school completion and learning outcomes by early adulthood. In particular, this study adds to the existing literature in three ways. First, the analysis focuses on the dynamics of program impacts in the long run as it tracks different cohorts of treatment children who have been in the program from one to nine years. Most of the few studies that measure program effects on intermediate and final outcomes in education did so with children who have been exposed to the treatment for no

[4.] Results from Behrman and Parker (2008) actually show some positive trends in math and reading tests for program participants. Results on math achievement tests are however limited by low sample sizes. The analysis on reading tests is based on single differences between program participants, which according to the authors, could underestimate the actual impacts of the program.
[5] Treated children were then matched to children in the control group based on the propensity to stay in school, derived from those in the control group that were enrolled both before and after program implementation.
[6] Schools could be comprised of children that take the test and others that do not take the test.

longer than two years. Second, we use three different data sources and samples (the baseline survey of the evaluation of the program from 2002, the program's information system, and a census of poor households) to perform two different methodological approaches which allows for a comparison of the findings across methodologies. Additionally, by using data from the program's information system for impact evaluation purposes, this paper highlights the opportunities for research that may arise from using monitoring and evaluation systems, as these are becoming increasingly popular tools to administer CCTs and other safety net programs. Finally, this paper also investigates the extent to which the final educational outcomes of older children that are not eligible to the program could be influenced by the participation of their siblings.

2.2 The Program

In the late 1990s, Colombia was hit by its worst economic downturn in 60 years: GDP shrank by 4.5 percent in 1999 alone, and the national poverty rate increased by 7.2 percentage points, largely erasing the socio-economic gains made during the early 1990s (World Bank, 2005 and 2008). A team was put together to address the social dimensions of the crisis from the government of Colombia, the World Bank, and the Inter-American Development Bank. One of the safety net programs was *Familias en Acción* (FA), designed as an instrument to help mitigate the effects of the economic crisis on the wellbeing of poor households and protect and promote human capital formation. FA was inspired by the CCT *Oportunidades* in Mexico, and consists of subsidies to education, nutrition, and health subsidies conditional on specific behaviors associated with school participation and attendance to health checks. The program was piloted in a few municipalities in 2001, but was scaled up in the second half of 2002 in the context of great, but still insufficient, progress in improving educational coverage and attainment. [7]

[7] Net enrollment rates for children aged 7-11 (primary school) increased from 77 to 93 percent from 1992 to 2002, and from 40 to 57 percent for those in the 12-17 age group (secondary school) (UNESCO, Institute for Statistics). The average educational attainment of people 15 years and older increased from 6.4 years in 1992 to 7.6 years in 2003. This is still low, however, when compared with other countries in the region such as Argentina (8.8 years) and Panama (8.6 years) (World Bank, 2005 & 2008). In terms of school completion, the numbers had not changed significantly since the beginning of the 1990s. At the pre-program time, still only 60 percent of children who started primary school finished fifth grade, while 57 percent of those who began secondary school finished ninth grade and only 35 percent completed eleventh grade, that is, high school (García and Hill, 2009; World Bank, 2008).

The largest component of the program is educational conditional cash transfers. Households with children aged 7-18 receive a monthly grant per child, conditional on the child attending to at least 80 percent of school lessons. When the program started, the grant was $12,000 pesos (approximately $7) for each child attending primary school (grades 1-5) and $24,000 pesos ($14) for those in secondary school (grades 6-11). The level of benefits was set to compensate for the direct cost borne by low-income families to send their children to school.[8] In some urban areas, the subsidies were increased to accommodate the higher opportunity cost of secondary education in the cities, and the subsidies for primary school students were replaced by nutritional subsidies for children 7-11 years old due to almost full enrollment in urban areas (Acción Social, 2010a). Like the *Oportunidades* program, the transfers are specifically given to mothers, a mechanism designed to ensure that the money is invested in children and as an incentive for empowering women within their communities.

Within each municipality FA targets the poorest households based on a proxy-means test system constructed with information from a census of poor people (known as *Sisben*). In Colombia, all households surveyed by *Sisben* are assigned to one of six brackets of a poverty index score called *Sisben* that is used to identify the most vulnerable population. The index runs from 0 to 100, and it is a function of a set of household characteristics and variables related to the consumption of durable goods, human capital endowments, and current income to calculate a score that indicates the household economic well-being. The first level includes households that are extremely poor. While many social programs target the population with scores falling in the first two brackets, FA is only offered to households in the first bracket. Municipal governments are responsible for ensuring adequate coordination with schools and health centers working with the FA program to ensure its successful local implementation.[9] Local banks deliver the cash transfers to beneficiaries every two months. It was estimated that the annual costs of the program in 2009 were going to be equivalent to 0.27 percent of GDP in that year (World Bank, 2008).

[8] After a first expansion in 2005, the grant was increased to $14,000 pesos and $28,000 pesos, respectively. The latest change in 2007 brought the grant to corresponding $15,000 pesos and $30,000 pesos, merely keeping up with inflation.

[9] Municipalities prepare the list of families to receive the subsidies (*Sisben* level 1 families with children 18 years old or younger). The list is consolidated from municipalities every three years. The subsidies are also contingent on verification of compliance with the conditions that involves beneficiary mothers obtaining attendance certificates from schools every two months and delivering them to the municipal coordination office, who then sends that information to the regional and finally the national coordination unit. In each community, a committee of beneficiary mothers is elected to monitor program implementation.

The program FA expanded over the years to reach national coverage in 2010. Initially, the program was targeted geographically. Only municipalities that were not departmental capitals, with fewer than 100,000 inhabitants, with at least one bank branch working in the municipality, and had access to facilities that allowed for the implementation of the program were eligible (691 out of the 1,024 municipalities). Within these communities, a total of 340,000 households in 622 municipalities were registered to participate (Attanasio et al., 2006). In 2005, the program was extended to include displaced families and households in departmental capitals and municipalities which either became able to offer the required services or with services accessible in nearby towns. Most recently, during 2007, the program expanded to municipalities with more than 100,000 inhabitants to include other deprived urban areas. The program now covers nearly 2.8 million participating households in 1,093 municipalities, representing almost 65 percent of the target population (Acción Social, 2010b; Attanasio et al., 2009).

An early evaluation of FA demonstrated positive effects on short-term outcomes such as household consumption and children's school participation and nutrition status. Indeed, within the first two years of program implementation, household consumption increased by 13-15 percent, school enrollment rates increased by around 5 to7 and 2 percentage points for children in secondary and primary schools, respectively, child labor participation fell by around 10 to 12 percentage points, and health and nutrition outcomes such as morbidity, immunization and anthropometrics also improved (Attanasio et al., 2005, 2006 and 2009; Attanasio and Mesnard, 2005).

3. Data

This paper uses four sources of data (a household survey, a census of the poor, and a database with administrative records of the program) to construct two samples of participant and nonparticipant children of the program FA for the two research strategies. The first approach employs matching methods and household survey data collected for the short-term impact evaluation of FA. This survey is part of an effort to collect longitudinal data from a stratified random sample of eligible families in both treatment municipalities and matched control municipalities. The survey is a standard multi-topic household survey that includes questions on demographics, household structure, education, health, consumption, employment, anthropometry, housing characteristics, shocks, and community education and health facilities. The baseline survey was carried out between June and October 2002.[10]

The matching analysis draws only from the baseline survey which interviewed 6,722 households in 57 treatment municipalities and 4,562 households in 9 control municipalities.[11] The subsample for this analysis includes only children who were born during 1975-1994, and who may have graduated from high school between 2003 and 2009.[12] For instance, a child that had completed grade 6 at baseline (either in a treated or control area) was expected to finish high school (grade 11) by 2007 if the child progressed on schedule. In contrast, a child starting primary school (grade 1) in 2002 (baseline) will not be able to finish high school at least before 2013. Therefore, the relevant cohorts of children to estimate the average impacts of the program are those who at baseline had 4 to 10 years of schooling, and who were 18 years old or younger (called "PSM data"). The baseline survey is also used to construct most of the pre-program covariates for the matching procedures.

The samples of analysis for the RDD approach are constructed with two different administrative sources of data. The first is the monitoring and evaluation system, SIFA, created for administrative and monitoring purposes at the onset of the FA program. The system is a longitudinal census of program beneficiaries from 2001 to present. To date, there is information on approximately 2.8 million families currently participating in the program. The second source of information is the data from a census of the poor (*Sisben*) carried out between 1994 and 2003 to construct the poverty index score for the proxy-means test. Questions were asked regarding households' demographics, structure, durable goods, housing characteristics, human capital, labor force participation, income, and access to basic services. By 2003, the surveys covered

[10] Two follow-up surveys revisited the same households in 2003 and 2005.

[11] Nearly 12 percent of the households interviewed in treatment municipalities were not registered with the program (Attanasio et al., 2005).

[12] We merged the household survey data with administrative data from the standardized test score *Icfes*. Using this merged dataset, we identified that 95 percent of the population that presented the test in the 2000-2009 period were born during the period1975-1994. In order to study the indirect effects of participant children on non-eligible young adults that were still in school or could rejoin, this sample also includes individuals 19 to 23 years old that were listed as dependents, but are not eligible to the program.

over 25 million individuals.[13] Data from SIFA and *Sisben* were carefully merged using confidential information on date of birth, full name, and national identification numbers. The resulting dataset ("SIFA + *Sisben* data") contains the universe of individuals above and below the threshold of eligibility (whether or not they actually participate in the program). This information is then used to construct indicators of program participation and length of exposure. In order to focus on a period of time that is comparable to the one examined in the matching analysis, the final subsample is restricted to treatment municipalities that were covered during the first phase of expansion between 2001 and 2004. This implies that in addition to the comparison group, the samples are comprised of beneficiaries that could have been covered for up to nine years until 2009.[14]

The two resulting data sets ("PSM data" and "SIFA + *Sisben* data") are merged with the administrative records on registration and results for the *Icfes* test.[15] This exam is a nationally recognized and standardized test that is administered prior to graduation from high school and mandatory for entrance to higher education. Over four million students registered and took the test between 2003 and 2009. This database identifies test takes by date of birth, full name and national identification number. This dataset is merged with the cohorts of children identified in the "PSM data" assembled to perform the matching analysis, and with the "SIFA+SISBEN data" to implement the RDD strategy. In order to avoid problems of nonrandom mismatch, strict procedures were followed to merge the datasets including matching based on full name, birth date, national identification number, and a minimization of the phonetic *Levenshtein* distance.[16] (see Appendix A for more details about the data merging procedures). The final matching rates are around 18 percent for the matching analysis and 24 percent for the RDD approach.[17]

The long-term impacts of the program on the human capital of children are estimated on two outcome variables. The first is an indicator of high school completion that is measured through a dummy variable that identifies whether a child registered or not for the *Icfes* test during the

[13] A new *Sisben* survey was fielded between 2003 and 2007 to update the information, improve the effectiveness of the targeting scheme, and change the algorithm due to concerns regarding manipulation by local authorities (Barrera, Linden, and Urquiola, 2007; Camacho and Conover, 2011).
[14] The final sample also excludes internally displaced people who became eligible to the program much later.
[15] More recently called *Saber 11*.
[16] The *Levenshtein* distance measures the difference between two strings, in terms of edits you have to do to convert one string into the other.
[17] We considered 10 points above and below the threshold to check the matching for the RDD sample.

period 2003-2009. Although the test is given to students just prior to graduation (grade 11), registration to the test is a good proxy for high school completion since over 90 percent of the test takers end up finishing grade 11 (World Bank, 1993; Angrist et al., 2006).[18] The test is also a strong determinant of college entrance as it fulfills the qualifying requirements in several subjects. The second outcome of interest measures academic achievement. Conditional on *Icfes* registration, we measured academic learning by the actual performance of the students on the test. The exam is a standardized test that assesses the academic achievement of students in various subjects such as Mathematics, Language, Biology, Chemistry, Physics, History, Geography, and a foreign language chosen by the student. We focus on the impacts of the program on the standardized scores in the Mathematics and Language modules (35 questions each) and the overall score in the test excluding foreign language.

4. Empirical Analysis

4.1 Research Design

The FA program was not randomly assigned as eligibility requirements were based on geographic and welfare targeting. Only extremely poor households with at least one child between 7 and 18 years old and a score in the proxy-means test that falls *Sisben* 1 level are deemed eligible for the education transfer of FA. Additionally, the program was initially implemented only in certain qualified municipalities based on their supply of health, education and financial services. These eligibility criteria may be problematic for evaluation as they may induce different sources of selection bias including observable and unobservable factors that could be correlated with program eligibility and educational outcomes. For instance, small towns like those initially targeted by the program may have poorer public infrastructure, less dynamic economies, and therefore lower returns to schooling. Parents living in these places may be less willing to send their children to school. In this case, lower improvements in school attainment over time among participant children relative to nonparticipant children in other towns could be mistakenly attributed to the program as negative effects. Moreover, given that participation in the program is voluntary, those families who sign up for the program in treatment areas may be different in many aspects from those who decide not to participate. We attempt to overcome the

[18] A small fraction of individuals also take the ICFES test after they have finished high school. Also, since the test can be taken many times, only the first registration date and score is kept for the small number of students who took it more than once.

potential identification issues that may arise from non-random assignment and voluntary participation with two different quasi-experiments, each of which are explained below in more detail.

Matching Analysis

The first research strategy builds on the design of the first short-term impact evaluation of the program. This evaluation was based on a non-experimental design that compares eligible households from municipalities covered by the program with potentially eligible households (also classified as *Sisben* level 1) from selected comparable areas not targeted by the program.[19] We follow different cohorts of children in treatment and control municipalities who were interviewed as part of the baseline survey carried out in 2002 and that could have finished high school during the period 2003-2009. Hence, the location, age and grade of the children at baseline determine their treatment status and length of exposure to the program for the treated. Given that 2003 is the first year in which the program was broadly implemented in the samples for the matching analysis, the pool of treated individuals includes children with program exposure that ranges roughly from one to seven years.[20] A limitation of this research strategy is the lack of baseline measures of the outcome variables, making it difficult to test for differences between groups at the pre-program time. Given that high school graduation and test scores are in reality observed only once for the same child, program effects are therefore estimated with post-program single differences between treated and control children.[21]

Matching methods are used in the comparison of outcomes to adjust for potential biases due to nonrandom targeting and selection into the program. The standard underlying assumption for

[19] Municipalities were grouped based on the number of eligible families that reside in each of them to form 639 Primary Sampling Units (PSU). Twenty-five strata were then defined based on geographic location, level of urbanization, number of eligible families, and indexes of quality of life and availability of school and health facilities in the municipality. Fifty PSUs (two within each stratum) corresponding to 57 municipalities targeted by the program were selected and then matched to 50 "control" PSUs (equivalent to 63 municipalities) that were relatively similar to "treatment" PSUs. Matching was done within each stratum and the comparability of both types of PSUs was assessed on the basis of population size and an index of quality of life. Finally, a stratified random sample of eligible households was selected in treatment and matched control municipalities. For more details about the matching process and the stratified random methods followed in the evaluation, see Attanasio and others (2005).

[20] It would be important to measure how the impacts of the program vary with different years of exposure. However, due to empirical limitations, this marginal impact cannot be properly identified. Despite the fact that the presence of cohorts would induce variation of treatment dosage, the evaluation design used in this paper cannot disentangle the effects of age on the outcomes from those of length of exposure. Conditional on being enrolled in school, older beneficiaries (when they joined the program) have fewer years ahead of them in school, and perhaps, are more likely to be observed finishing high school. If this is the case, shorter length of exposure may be wrongly attributed to higher school completion rates.

[21] Only a very small fraction of children were found to have taken the test more than once. However, although the test can be taken several times, it is administered only to students who have achieved grade 11.

this approach is that matching on the propensity score (i.e. the estimated probability of participation in the program) eliminates any bias generated by pre-treatment differences between the two groups as long as there are no differences in unobservable variables that jointly influence program participation and the outcomes under analysis. The availability in the baseline survey of relevant pretreatment information to model program targeting and participation, in principle, makes the application of the matching methodology suitable for the evaluation of FA.

Although careful procedures were followed in the early evaluation to select comparable control areas, a comparison of baseline characteristics between treated and control children in our sample using standardized t-tests and normalized differences, presented in Table 1, reveal a number of differences that are statistical significant. In order to balance the distribution of covariates between the two groups and assess the sensitivity of the results, we matched children on the basis of three different model specifications to predict the probability of treatment. Table 2 presents the group of variables included in the three models used to predict participation into the program. The first model (Model 1) includes standard individual and household pre-treatment socio-economic and demographic characteristics such as age and order of the child, dummy variables for married and participating in the labor force head of the household, age, education and gender of household head, urban location, the number of children in the household ages 7 to 11 and 12 to 18, and monthly expenditures. The second model (Model 2) extends Model 1 by adding a number of municipality-level covariates that proxy for measures of educational supply and demand at baseline (for instance, pupil/teacher ratio and access to schools). In addition, the third model (Model 3) includes all but the health variables used by Attanasio et al. (2005) in their participation models estimated for the evaluation of the short-term impacts of the FA program. This augmented specification includes additional household- and municipality-level variables mostly associated with the attributes of the dwellings, family structure, access to basic services, population, relevant public infrastructure, population, and geographic characteristics.[22] Various kernel techniques and bandwidths are used to match children and define common supports that exclude poor matches between treatment and control groups. The estimated propensity scores for each of the three different model specifications

[22] See tables 1, 2 and 3 for a complete list of all variables used in each of the models of participation.

suggest that there is a strong overlap between the treatment and comparison groups. We then estimate the treatment effect in a standard way as follows:

$$\Delta ATE = \frac{1}{n}\sum[E(Y_1|D=1,X) - E(Y_0|D=0,X)] \qquad (1)$$

Balancing checks were conducted to assess the comparability of treatment and control children after conditioning for observable characteristics that explain participation into the program. The results of these balancing tests are particularly robust for Model 1.[23] Furthermore, we do not find major statistically significant differences in the conditioning variables between treated and control within strata having similar probabilities of program participation. Even though there is not a formal way to rule out the existence of unobserved factors that could determine participation and the final outcomes, these results provide confidence in the ability of matching with the estimated propensity scores to indentify program impacts.

An additional concern for identification remains when program impacts on learning are estimated with academic tests given to children enrolled and present in school. By raising school enrollment, CCTs –including FA, make program participants more likely to take tests given their higher school participation. The "marginal" children that are brought into school and promoted through grades due to FA are probably different (for instance, poorer or less motivated) from those who would have been enrolled or attended school regardless of the subsidy. There may also be heterogeneity in the expected returns to education between those previously enrolled in school and the new enrollees. If this type of selection exists, the test score distributions of treated and control children tested in school are not comparable. In order to address this issue, we follow Lee (2002) and Angrist et al. (2006) to construct nonparametric upper bounds of program effects (for the matching analysis) on learning by symmetrically truncating the two distributions at some specific quantile. In contrast, unadjusted (selection-contaminated) comparisons of test scores – conditional on positive scores– provide lower bound estimates of program impacts (see Appendix B for details on the methodology for nonparametric bounds of program impacts).

[23] Additional regression analysis also shows that, after controlling for the estimated probability to participate, no additional conditioning variables help predict the receipt of treatment.

Regression Discontinuity Design

The second research design employed in this paper exploits variation in assignment to treatment arising from the discontinuous rule that determines eligibility to the program. As noted before, a proxy-means test designed with the goal of identifying the most vulnerable population that qualifies for various social programs was used to define the target population of FA. The index (*Sisben*) varies between 0 and 100 and is the result of an algorithm that weights households' variables associated with their socio-economic wellbeing. Households placed in the first bracket of *Sisben* (level 1), namely those with scores below 18 and 36 in rural and urban areas, respectively, were considered eligible for the first phase of the program. In principle, the expected discrete change in participation produced by the rule offers an opportunity to estimate the causal effects of the program on education outcomes with a Regression Discontinuity Design (RDD). The intuition behind this strategy is that households that lie just below and just above the threshold are statistically comparable except for their participation in the program. As a result, any discontinuity in the conditional distribution of high school graduation rates and tests scores at the cutoff could be interpreted as the effect of FA.

We used data from SIFA, an information management system that compiles information on a number of operational aspects of the program including historical records of all beneficiaries. When merging it with the *Sisben* scores, these databases allow identification of the universe of individuals in households that lie below or above (in a smaller proportion) the cutoff of eligibility and participate in the program between one and nine years. Information on the rest of the population relevant for the analysis – those eligible but not participating and those with *Sisben* scores above the threshold of eligibility— is obtained from a census of the poor that was collected between 1994 and 2003 used to compute the *Sisben* proxy-means test. Both datasets are merged together to create the sample of analysis for the RDD.[24]

In practice, the *Sisben* poverty score predicts substantial but not perfect changes in the probability of receiving the treatment. In fact, the data show a significant discontinuity in the probability of assignment to treatment at the threshold of around 66 to 72 percentage points as

[24] Given that the information system of the program does not reports SISBEN scores for participant families, uniform scores for the whole sample are calculated by using the proxy means testing algorithm and the information from the Census of the Poor 1994-2003.

presented in Figure 1 and Table 3.[25] Therefore, we perform a *fuzzy* instead of a *sharp* RDD. The average causal effect of this design is given by the ratio of the jump in the outcome variable at the threshold to the jump in the probability of participation in FA also at the threshold. This is equivalent to an instrumental variable setting in which the average effect of the treatment is obtained from compliers -- individuals whose participation is affected by the cutoff. To check the sensitivity of the results to different specifications, estimates of program impacts based on the RDD are computed using different parametric functional forms and nonparametric procedures. Besides, to restrict the sample close to the cutoff, optimal bandwidths for the nonparametric analysis were also estimated using Imbens and Kalyanaraman (2010) methods.

Our first stage regression formally tests if the probability of treatment D_{imt} for individual *i*, in municipality *m*, and in year *t* changes discontinuously at the cutoff point. We estimate different polynomial specifications of the model allowing the regression function to differ on both sides of the threshold as follows:

$$D_{imt} = \alpha_0 + \delta T_{imt} + \alpha_1 f(s_{imt}|s_{imt} \leq s^*) + \alpha_2 f(s_{imt}\,|s_{imt} > s*) + \gamma_m + \vartheta_{imt} \quad (2)$$

Where $T_{imt} = 1[s_i \leq s^*]$ is an index function that indicates whether the *Sisben* score of the individual *i* is below the eligibility threshold s^*, γ_mcorrespond to municipality fixed effects included in the regression. Results from the first stage of the fuzzy design, summarized in Table 3, show that there is a large and significant jump in the treatment probability δ at the cutoff of the assignment variable. In fact, and confirming the graphical analysis presented in Figure 1, all point estimates of δ given by the models vary from 0.69 to 0.73 significant at the 1 percent level, regardless of the flexibility specified in the functional form around the threshold..

Our reduced form equation is described by the following equation:

$$Y_{imt} = \tau_o + \beta D_{imt} + \tau_1 f(s_{imt}|s_{imt} \leq s^*) + \tau_2 f(s_{imt}\,|s_{imt} > s*) + \gamma_m + \tau_3 X_{imt} + \varepsilon_{imt} \quad (3)$$

where Y_{imt} corresponds to our outcome of interest (i.e. high school graduation). The *fuzzy* Regression Discontinuity analysis instruments the treatment dummy, D_{imt}, with T_{imt} to identify

[25] This may happen for at least two reasons. First, take up rates were in the order of 65 percent so participation is not universal among the eligible population. Second, households above the threshold may lobby with local authorities to gain access to the program. Concerns of nonrandom sorting that could arise from this or other gaming behaviors to influence the *Sisben* poverty index score are discussed in the robustness section.

the “intent-to-treat” effect. Our coefficient of interest corresponds to the ratio of the coefficient of the treatment effect from the reduced form, β, and the coefficient of that identifying individual eligibility, δ. Results from this approach and the matching analysis are presented next. However, before discussing the findings, it is important to mention that the effect estimated with the RDD framework is only applicable to the group of individuals around the threshold of eligibility. For this reason, although program impacts derived from matching and RDD analyses seek to estimate the same parameter, they are not strictly comparable.

4.2 Program Impacts

High School Completion

In theory, the net effect of CCTs on children’s school completion is ambiguous. On the one hand, there are two obvious positive direct effects. One is an income effect arising from the cash transfer which increases the budget of the family so that they could afford keeping the children in school. The second comes from the program’s condition on regular school attendance, which introduces a substitution effect for children not in school or not attending regularly as it reduces the relative price of education. Together these effects are expected to increase the demand for education among the target population, a prediction widely confirmed in the literature in the form of positive effects on school enrollment, attendance and progression. If the additional investments in educational services are continued over time, one might also expect an increase in high school completion. On the other hand, there are other possible mechanisms that could reverse these positive effects. For instance, classrooms may be overcrowded by the additional enrollment affecting the academic performance and progress of children. Additionally, perverse incentives could encourage families to delay children’s graduation just to prolong their participation in the program if the child still meets the age criteria for eligibility, i.e. below 18 years of age. In addition to the standard income and substitution effects, child-specific conditional transfers like the ones offered by FA could also have negative impacts on the school outcomes of ineligible siblings due to a displacement effect (Ferreira et al., 2009).

Findings from the first empirical approach (matching analysis) on the net effects on high school completion are summarized in Table 4. We report 3 sets of regression for each matching model. For comparison purposes the first column in each pair corresponds to the OLS specification including a linear form all the variables from the matching model, whereas the second column presents the matching model. Overall, OLS results suggests that on average treated children are between 3 and5 percentage points more likely to finish high school. Results from the three different specifications of the model of participation also show positive and statistically significant effects of the program on high school completion. In this case impact estimates vary between 4 and 8.4 percentage points depending on the specification of the propensity score. Looking at the existing evidence produced for other programs, these effects appear to be comparable to those estimated for similar CCT and education fee waiver programs in Pakistan and Colombia where, as discussed previously in the background section, participants are more likely to complete secondary by 4 to 7 percentage points.[26]

Table 4 also includes impacts based on the gender and location (rural or urban) to test for heterogeneous effects of the program. The analysis by gender uncovers clear differences in the magnitude of impacts for participant boys and girls. On one hand, results for the sample of girls based on OLS and matching estimates always yield positive and statistically significant effects on high school graduation rates. The magnitude of the effects ranges from 4.6-6.5 percentage points in the OLS framework to 5.2-8.9 percentage points in the matching analysis. For all these cases, results are strongly significant in statistical terms at the 1 percent level. On the other hand, the pattern in impact estimates for boys is less obvious. While OLS point estimates indicate that there are no program effects, two of the three point estimates obtained from matching models (Model 2 and Model 3) appear to be just marginally significant at the 10 percent level. Results from the sample used in the RD analysis, presented in Table 5, also seem to suggest that program effects on high school completion are a little larger among girls.

The distribution of impacts on high school completion varies with the location of program beneficiaries as well. However, there is a clear difference between the results from the matching and RDD approaches. Evidence from the PSM models suggests that the effects on high school graduation rates accrue mostly to participant children whose families resided in rural areas at the

[26] A similar CCT program implemented in Bogota (Colombia), *Subsidios*, increased the probability of completing high school by 4 percentage points. Beneficiaries of Colombian PACES, which offered vouchers to attend private secondary schools to students from poor urban neighborhoods, are also found to be 5-7 percentage points more likely to graduate from high school. In Pakistan, the Female School Stipend program, a CCT targeting girls, appears to improve the chance of completing grade 9 in high school for girls aged 15-16 years old by 5 percentage points (Angrist and others, 2004; Barrera-Osorio and others, 2008; Alam and others, 2010).

baseline (most likely girls based on the previous results). Whereas the effects on the samples of children in rural areas are positive and strongly significant in both economic and statistical sense in all models, analogous point estimates obtained for beneficiaries in urban areas are mostly insignificant – except for a marginally significant impact from the second matching model. Even if the discussion is limited to the parameter estimates of this particular model, the findings show that the size of the effect in urban centers is still half of that for children in rural municipalities - about 5 percentage points compared to 10.4 percentage points. Even though there are no data to empirically identify the channels that may explain these differences in terms of location and gender, a possible explanation is that the marginal effects of the transfer and the conditionality are larger for girls in rural settings for whom the opportunity costs of education are relatively lower (for instance, if households are less dependent on the labor of the girls for farming activities).

We then turn to program effects on high school completion estimated on a different sample and using the RDD approach. To inform the discussion, the results of the RDD analysis are first shown in a graphical way. Figure 2 show the average and estimated high school completion rates of children with respect to their ranking in the *Sisben* poverty index score relative to the threshold.[27] Means of high school completion rates for each value of the normalized poverty index score also provide suggestive evidence that FA had a positive effect on high school graduation. The "jump" at the threshold indicates that the discontinuous change in eligibility increases the probability of finishing high school.

In addition to examining the possibility of impacts and understanding the functional form through graphical analysis, program effects are estimated econometrically with different parametric regressions that include different polynomial functions and non-parametric regressions following Imbens and Kalyanaraman (2010). In specifications 1 and 2 of Table 5, we include quadratic forms of the control function $f(s_{imt}|s_{imt} \leq s^{*}), f(s_{imt}\ |s_{imt} > s\ *)$, where s_i correspond to the value of the proxy-means test (*Sisben*) and s^* denotes the threshold of eligibility, and let the regression function differ on both sides of the cutoff point. In columns 3 to

[27] The cut-off of eligibility takes different values for urban and rural municipalities. Therefore, rather than presenting the outcomes as a function of the *Sisben* poverty index score, they are normalized as the distance of each child's score to the area-specific cut-off that is used to classify households as level 1 in the proxy-means test system, and determine eligibility for this specific program. For instance, a child with a value -5 is in reality 5 points below the cutoff and is therefore eligible to *FA*. In contrast, positive values of the normalized score represent children that belong to ineligible households, given the eligibility rule.

6, we also run other flexible specifications of the regression function including cubic and quartic functional forms of the control function. Non-parametric estimations of the models are presented in column 7. Throughout all functional form specifications with or without including controls, we identify a positive and significant impact on the high school completion of children that participate in the program. Overall, focusing on our preferred specifications that are the quartic functional form with controls and nonparametric models (columns 6), we estimate that beneficiary children are between 3.3 and 4.5 percentage points more likely to finish high school, including municipality fixed effects. These results are robust to inclusion of controls for school quality (average score in the ICFES test and class size by school and year).

Finally, we looked at the possible indirect effects of FA on the school completion of nonparticipating young adults who reside with participant group children. This indirect effect could be caused by a substitution of resources and time allocated to work between eligible and non-eligible children within the household that, although positive for participant children, is expected to run in the opposite direction for ineligible children.[28] The treatment group was comprised of ineligible young adults who were more than 18 years old and had not finished high school when at least one of the children of the same household joined the program. The high school graduation rate of this group is compared with those of young adults of similar characteristics who reside either in eligible families in control areas (matching analysis) or in families that were ineligible for the program, but that are otherwise similar to participant households (RDD).[29] Overall, the analysis (results not shown) does not reveal consistent evidence of either positive or negative indirect effects of the program on the school completion of nonparticipating young adults.[30]

Test Scores

Analogous to the existent ambiguity in the link between CCTs and high school completion, the net effect of interventions that increase school participation on learning outcomes is difficult

[28] A clear example is the reallocation of labor or child caring away from participant children—so that they are able to comply with the conditionality—towards ineligible children in the household. Existing empirical studies have provided evidence of either negative effects or no effects of CCTs or similar programs on the school enrollment of ineligible siblings in Cambodia, Colombia and Pakistan (Ferreira and others 2009; Barrera-Osorio and others 2008; Alam and others 2009).
[29] The final sample includes young adults that meet the criteria outlined in the text and that were identified in the survey as household members but not as the main breadwinner.
[30] Results available from the authors upon request.

to establish theoretically. In principle, students who attend school more regularly and spend more time doing school work due to FA are supposed to have higher academic achievement than non-recipient children who are out of school or attend less regularly[31]. In fact, cross-country evidence shows that school attainment correlates strongly with test scores (Filmer and others, 2006). In addition, CCTs like FA could enhance learning at least through two additional channels. First, cash and nutritional transfers have been found to encourage positive behaviors towards investments in cognitive enhancing inputs such as books, more nutritious food, and parental time, and less child work. Second, the conditionality on attendance required by the program could lead to more learning for enrolled children who do not attend school regularly. Nonetheless, there may be other effects running in the opposite direction. For instance, the extra influx of students may increase class size and put additional pressure on existing educational resources. In addition to congestion, increased enrollment may also affect class composition and trigger negative peer effects in learning. Moreover, CCTs are often targeted to the neediest areas where school quality may be low. Finally, with just a few exceptions -not in the Colombian case, CCTs have not been designed with explicit objectives and incentives to raise academic performance.

However, as noted above, selection bias can confound the analysis due to different observable and unobservable characteristics of the marginal students who would not have joined school, progressed until grade 11 and taken the exam in absence of the program. If it exists, this sort of selection will most probably introduce low-scorers into the distribution of the treatment group. To address the probability of sample selection for children that took the test, we performed bounding procedures to symmetrically truncate the distributions of the treatment and control group at a quantile φ_0 where non-selected control children begin having positive scores (i.e. start taking the exam), that is, for each score Y, the following should hold: $Y\left(q_{\varphi 0}\right) > 0$ (Appendix B provides more details on this methodology). This is equivalent to estimating the impacts for the students who would have taken the exam in the absence of the program ("always takers"). Given that the procedure rests on the assumption that the selection bias is negative, the unadjusted conditional-on-positive comparison of test scores provides a lower bound of the impact of the program. Table 6 presents Model 1, 2 and 3 in the following way: the OLS

[31] If regular attendance on average is lower than 80 percent of the classes.

estimates, the unadjusted nonparametric lower bound estimates, the quantile φ_0 where non-selected control children begin having positive scores, and the corresponding adjusted nonparametric upper bounds estimates. Overall, these findings do not provide an indication of program effects on test scores either. For the most part, program effects estimated at φ_0 are insignificant. There are, however, some positive and significant effect of 0.098 and 0.069 of a standard deviation on Spanish and the overall test score, respectively. However, these results are very unstable and hold only for the specifications of the first model of participation in the matching analysis.

In general, findings from the second research design (RDD) on the "SIFA + *Sisben*" sample, in Table 7, also indicate that program recipients who graduate from high school perform at the same level as equally poor non-recipient graduates. The graphical analysis[32] does not provide visual evidence of a jump in the regression function at the threshold. Overall, econometric results based on functional forms with second, third and fourth-order polynomials in general show that participant children do as good as the children in the control group in their math test scores. There is a partial negative effect of the program on Spanish (significant at the 5 percent level) but not consistent throughout all the functional form specifications. Furthermore, the findinga indicate that there is no systematic evidence of differential performance between participant and non-participant children based on the scores in the overall test. The non-parametric estimations (estimated on an optimal bandwidth close to the cutoff following Imbens and Kalyanaraman (2010)) consistently show that test scores are lower for treated children. Nevertheless we should take into that account that we were not able to correct for the negative selection bias, given that the controls appear to be over-represented in this data. Additionally, results coming from the non-parametric estimation use a local sample very close to the threshold assigned by the optimal bandwidth. When compared against the little evidence available, the absence of impacts of FA on learning outcomes reported in this paper mirrors the findings available from previous evaluations of comparable interventions in Mexico and Cambodia (Behrman and others, 2005; Filmer and Schady, 2009).

[32] Not presented, but available upon request to the Authors.

5. Robustness Checks

This section discusses the robustness of the findings regarding a number of identification issues that may affect the internal validity of the analysis. We start by discussing in more detail the quality of the treatment and control groups in terms of their comparability. In the case of matching, the underlying assumption states that there are no unobservables that could create nonrandom selection into the program after matching treated and control children on the estimated probability of assignment to treatment. Although there is no definitive test to formally rule out selection on unobservables, a series of checks on observable variables do not provide a serious indication of this type of bias in our matching analysis. We performed balancing checks to assess the comparability of treatment and control children after conditioning for a large set of observable pre-program characteristics at the individual, household, and community levels that explain participation into the program. Regression analysis shows that, after controlling for the probability of participation, no additional conditioning variables help predict the receipt of treatment. Furthermore, there are no major statistically significant differences in the conditioning variables between the treatment and control within the same strata of similar probabilities of program participation. These balancing tests are particularly robust for the first specification of the model of participation (Model 1). In addition to the common support restriction, we also dropped observations with probability of participation below 0.1 or above 0.9.

As for the RDD approach, the analogous condition for identification is that the groups of people right below and above the threshold of eligibility need to be statistically equivalent and that the only difference between them is the treatment itself. To account for possible differences in addition to participation in the program, we run RDD models based on specifications including municipality fixed effects. Overall, the main findings of the paper regarding the impacts on high school completion appear to be stable across these different model specifications, signaling robustness to a variety of covariates. Unfortunately, it was not possible to control directly for school fixed effects in the econometric analysis. [33] This, however, seems to

[33] It was not possible to control for school fixed effects in the econometric analysis for at least three reasons. First, high school completion is observed only for individuals in the relevant ages that registered for the ICFES test and therefore it is impossible to know the school of children that were not matched to the ICFES database. Second, information with the name and code of schools exists only for a – probably not random- subsample of the students who took the ICFES test. Third, in many cases – particularly in small municipalities where the school supply is fixed – the inclusion of municipality fixed effects is expected to remove any existing school fixed effects.

be less problematic in the RDD since treatment and control groups are in some cases comprised of children that attend the same schools, particularly in small municipalities. Additionally, for the RDD analysis, we constructed a more balanced (trimmed) sample using the distribution of the p-score and dropping units with a propensity score below 0.1 or above 0.9 to make estimates more precise and less sensitive to changes in specification (Crump et al. 2009).

To test the identifying assumption of the RDD approach formally, we carried out a number of continuity checks on baseline characteristics at the individual and household level –all of them available from the census of the poor and used for the *Sisben* proxy-means test – that could be associated with the outcomes of interest. Figure 3 and Table 9 show graphical and regression analysis on the relationship between the *Sisben* poverty index score and these variables. Because of the large size of the sample used in the analysis, these differences are very precisely estimated. In general, there are not remarkable statistically significant differences in pre-program characteristics between the two groups on each side of the cutoff that could be argued to drive the results. In cases where the differences are statistically significant, the magnitude of the discontinuity is either relatively small in economic sense to drive the results (on average 0.046 more children in households right above the threshold of eligibility) or the direction of the difference (smaller household size or larger home ownership) is unlikely to introduce a bias that could affect the interpretation of our results.

Another potential concern in the RDD is the possibility that individuals could manipulate the assignment variable (the *Sisben* poverty index score) and generate nonrandom sorting around the threshold. For example, the RDD approach may be invalid if more motivated and education-driven people seek to influence the value of their scores by taking actions such as hiding assets, 'borrowing' children from other families to increase the size of the household, or bribing local authorities and program administrators. This may undermine the comparability of people on each side of the cutoff. While it is impossible to fully rule out this type of behavior in the context of FA, there are several reasons to believe that direct manipulation of the assignment mechanism is not a major concern for identification. First, an examination of the density of the *Sisben* score, presented in Figure 4, itself shows that there are no jumps in the distribution at the eligibility threshold of this program[34] for the whole sample or when broken down by gender and location. Second, and related to the continuity checks discussed above, there is no evidence of discrete changes in the distribution of other observable dimensions beyond the probability of participation in the program that could indicate some degree of manipulation of the scoring and ranking system.

Another issue is the possibility of nonrandom migration and crossover. In theory, families in control areas could try to migrate to treatment areas to become eligible for the program. This may affect the internal validity of the analysis if, for instance, families that are poor – and perhaps systematically different from non-migrants in other dimensions – are more likely to engage in this type of migration. Similarly, parents that live in control areas may prefer to send their children to schools that serve program beneficiaries if they are perceived to be of higher quality and are within their geographical reach. There are, however, a number of reasons to think that migration and crossover effects of this sort are unlikely to invalidate the findings of this paper. First, migration from control to treatment municipalities is close to zero among the baseline households examined in the PSM approach and that were resurveyed in the two subsequent rounds of longitudinal data collected in 2003 and 2005. Second, as noted above, part of the control group used for the RD analysis is comprised of children and adolescents that are above the cutoff of eligibility but reside in the same municipalities than participating children. Third, the economic incentives induced by the amount of the transfer (around $7-14 per eligible child) are probably not enough to compensate for the direct and indirect costs of migration. Fourth, *Acción Social*, the national agency that administers the program opens calls for inscription and enrollment into the *FA* program only on specific dates for each municipality, making it impossible for people to join the program at other times.

Problems of misspecification of the underlying regression models may introduce bias in treatment effects, particularly in the RD design. In order to avoid such problems, we checked the sensitivity of program impacts using various alternative specifications. Overall, the existence or lack of program effects on school completion and test scores, respectively, from the first approach is fairly stable across parametric (OLS) and nonparametric (matching) models.

[34] Camacho and Conover (2011) find evidence of manipulation for the *Sisben* proxy-means test constructed with the same census of the poor. However, this problem occurs only for people around the level 2 threshold which determines eligibility for other social programs such the Subsidized Health Insurance (*Régimen Subsidiado*).

Similarly, and given that the consequences of model misspecification are more serious in the RDD approach, the empirical models in this design were run for a number of low-order polynomial functions as well as other more flexible functional forms including third- and fourth-order polynomials and non-parametric models. The direction – and even in some cases the magnitude – of the RDD estimates to test for school completion are robust to the inclusion of different polynomial terms in the control function.

Finally, another concern is that the higher rates of matching survey data and test records observed for the treatment group may be driven by differences in the merging procedures and quality of information rather than by the effect of the program. In particular, considering that part of the information used to construct the treatment and control groups come from different datasets, it could be that the individual-level variables used to merge them (name, date of birth and national identification number) may be more accurate for people in the treatment group. For instance, surveyors that collected the data for the short-term impact evaluation could have been more careful to correctly keep the information to identify individuals in the treatment group for the subsequent rounds of data collection. Similarly, as part of the regular updates of the data entered in the information systems of FA (SIFA), program administrators may be more likely to correct mistakes in names, birth dates, and national identification numbers of program beneficiaries. As a result, individuals in the treatment group may be more likely to be matched with records in the *Icfes* database, not because they are more likely to take the test, but because of better information quality. In practice, however, measures computed from the matching algorithms such as the *Levenshtein* distance do not indicate systematic differences in the accuracy of data between the treatment and control groups. Therefore, it seems unlikely that this could be a reason to find children covered by the FA program to be more likely to be matched to *Icfes* registration records.

6. Conclusions

Despite growing efforts directed to assessing the impacts of CCTs, the most popular type of safety net applied in developing countries, on education, gaps in knowledge exist as to whether the largely documented positive effects on enrollment and attendance are sustained over time and result in higher school attainment. The evidence on the relationship between higher utilization of school inputs due to CCTs and learning outcomes is equally scant. This paper seeks to help fill these gaps by empirically investigating the schooling trajectories and academic performance of various cohorts of participant children who have different levels of program exposure ranging from one to nine years. We find robust evidence that the FA program increases school attainment by helping participant children to finish high school. Indeed, treated children are on average between 4 and 8 percentage points – equivalent to an increase of 8-16 percent – more likely to graduate from high school relative to those in the control group. If at present the program supports nearly 3.5 million poor children and only about 36 percent of the children in Colombia who start primary school are expected to graduate from high school[35], a conservative extrapolation of program impacts to this population would be equivalent to around 100,000-200,000 additional high school graduates.[36]. Moreover, the size of these program impacts is in the same range of magnitude as the effects found in similar CCT and education fee waiver programs in Colombia and Pakistan.

By encouraging participant children to finish high school, the FA program is expected to have other positive effects on further human capital gains (increasing the probability of entering into higher education), future employability, and income growth. As is the case in most developing countries, finishing high school is a critical achievement for low-income individuals and may lead to significant positive externalities in various dimensions. The first obvious channel is eligibility to apply for college or formal technical training, which may increase their qualifications and economic prospects.[37] Likewise, having a high school diploma already has a fairly high value in the labor market in the form of improved access to more and better jobs and higher wages. Moving to other plausible dynamic effects, empirical findings in the Colombian context suggest that more educated individuals tend to not only have fewer children but also

[35] Reference about work from Sanchez

[36] If households classified in SISBEN 1 have on average 1.37 children between 7 and 18 years and the number of families covered by the program is about 2.8 million, *FA* supports nearly 3.8 million children. The effect of the program in terms of the number of students is proxied by the increment in the high school completion rates induced by the program (from 36 percent to 40-44 percent) with respect to the 3.8 million that are currently participating in FA. Estimates based on the high school completion rates observed for the control group from the PSM analysis (children between grades 4 and 10 at baseline) suggests that approximately 50 percent of these children would graduate from high school. Applying the effects of the program to this completion rate implies that the aggregate effect would be in the order of 140,000-280,000 additional high school graduates.

[37] Indeed, recent evidence for Colombia shows that higher education (college and technical training) provides positive returns on wages that range from 7.4 to 12.8 percent (García-Suaza et al., 2009).

invest more in their human capital, consistent with the prediction of theories about the trade-offs in the quantity and quality of children.[38]

The impact analysis on learning outcomes shows mostly no statistical differences in test scores between the treatment and equally disadvantaged control group children. Furthermore, there is no clear pattern in the direction, size, and significance of impact estimates across different model specifications after adjusting for the probability of sample selection that restricts the treated sample to children that would have finished school even without the program. The fact that children covered by FA do not perform better than non participant children despite the monetary transfer and the conditionality is in line with existing evidence on similar programs in Mexico and Cambodia. CCT programs could have various conflicting effects on the learning outcomes of participant children. On one hand, beneficiary children may be expected to perform better as they stay more in school and their parents invest more time and money in their nutrition, health and education. On the other hand, these interventions are often targeted to disadvantaged areas where the quality and supply of education are probably low. Besides, the increase in the demand for schooling could also cause classroom congestion and induce negative peer effects.

Although interventions like CCTs are designed to improve school participation of poor children, not to directly raise learning, there is growing concern regarding the level of skills and quality of education with which program participants seek admission to higher education or enter the labor force after exiting the program. Therefore, assessing the potential of CCTs and/or supplementary interventions for increasing learning is critical for policy making. Innovations in project design (for instance, changing the timing of transfers or tying them to performance rather than attendance) as well as supplementary supply-side interventions aimed at improving school quality and increasing resources for low-performing students are possible options. Pilot tests, together with careful evaluations, would surely yield valuable knowledge about the efficacy of these policies in linking the objectives of increasing human capital with improving learning outcomes.

[38] Conditional correlations show, for instance, that having a secondary education degree reduces the expected number of children by 27 percent, almost twice the effect calculated for primary education; in addition, children's enrollment status and educational attainment are shown to be largely determined by their parents' education (N. Forero and L. Gamboa, 2009; Nunez and Sanchez, 2003).

References

Acción Social. 2010a. Familias en Acción: Corbetura Geografica. Accion Social: http://www.accionsocial.gov.co/contenido/contenido.aspx?catID=204&conID=157&pagID=275

Acción Social. 2010b. Familias en Acción: Subsidios. Accion Social: http://www.accionsocial.gov.co /contenido /contenido.aspx?catID=204&conID=157&pagID=282

Alam, A., J. Baez, and X. del Carpio. 2011. "Does Cash for School Influence Young Women's Behavior in the Longer Term? Evidence from Pakistan." IZA Discussion Papers N. 5703, Bonn, Germany.

Angrist, Joshua, Eric Bettinger, Erik Bloom, Elizabeth King and Michael Kremer. 2002. "Vouchers for Private Schooling in Colombia: Evidence from a Randomized Natural Experiment" *American Economic Review* 92(5): 1535-58.

Angrist, Joshua, Eric Bettinger, and Michael Kremer. 2006. "Long-Term Educational Consequences of Secondary School Vouchers: Evidence from Administrative Records in Colombia." *American Economic Review* 96(3): 847-72.

Angrist, J., E., Bettinger, and M. Kremer. 2004. "Long-Term Consequences of Secondary School Vouchers: Evidence from Administrative Records in Colombia," NBER Working Papers 10713.

Attanasio, O., E. Battistin, and A. Mesnard. 2009. "Food and Cash Transfers: Evidence from Colombia." Discussion Paper No. 7326. London, UK: Center for Economic Policy Research.

Attanasio, O., E. Fitzsimmons, A. Gómez, D. López, C. Meghir, and A. Mesnard. 2006. "Child Education and Work Choices in the Presence of a Conditional Cash Transfer Programme in Rural Colombia." Working Paper W06/01. London, UK: IFS.

Attanasio, O., E. Fitzsimmons, and A. Gómez. 2005. "The Impact of a Conditional Education Subsidy on School Enrollment in Colombia." Report Summary: Familias 01. London, UK: IFS.

Attanasio, O. and A. Mesnard. 2005. "The Impact of a Conditional Cash Transfer Programme on Consumption in Colombia." Report Summary: Familias 02. London, UK: Institute for Fiscal Studies.

Attanasio, O., L. C. Gómez, P. Heredia, and M. Vera-Hernández. 2005. "The Short-Term Impact of a Conditional Cash Subsidy on Child Health and Nutrition in Colombia." Report Summary: Familias 03. London, UK: Institute for Fiscal Studies.

Barrera-Osorio, F., M. L. Bertrand, L. Linden, and F. Perez-Calle. 2008. "Conditional Cash Transfers in Education: Design Features, Peer and Sibling Effects Evidence from a Randomized Experiment in Colombia." Policy Research Working Paper 4580. Washington, DC: The World Bank.

Barrera-Osario, F., L. L. Linden, and M. Urquiola. 2007. "The Effects of User Fee Reductions on Enrollment: Evidence form a Quasi-Experiment," Columbia University, Department of Economics, Mimeo.

Camacho, A. and E. Conover. 2011. "Manipulation of a Targeting System". *American Economic Journal: Economic Policy* (forthcoming).

Crump, R., V. J. Hotz, G. Imbens and O. Mitnik, "Dealing with Limited Overlap in Estimation of Average Treatment Effects", *Biometrika*, Vol. 96, Number 1, 187-199, March 2009

Behrman, Jere R., S. W. Parker, and P. E. Todd. 2005. "Long-Term Impacts of the Oportunidades Conditional Cash Transfer Program on Rural Youth in Mexico." Discussion Paper 122. Göttingen, Germany: Ibero-America Institute for Economic Research.

Behrman, J. R., P. Sengupta, and P. Todd. 2000. "The Impact of PROGRESA on Achievement Test Scores in the First Year." Final Report. Washington, DC: IFPRI.

Behrman, Jere R., P. Sengupta, and P. E. Todd. 2005. "Progressing through PROGRESA: An Impact Assessment of a School Subsidy Experiment in Rural Mexico." *Economic Development and Cultural Change* 54 (1): 237–75.

Bhargava, A. 2007. The Impact of Conditional Cash Transfers on the Educational Performance of Children: Evidence from Colombia's Familias en Acción Program.University of British Colombia.

Dubois, P. and M. Rubio-Codina. 2009. "Child Care Provision: Semiparametric Evidence from a Randomized Experiment in Mexico." CEPR Discussion Paper No. DP7203.

Fan, Jiangqing and Irene Gijbels. 1996. Local Polynomial Modelling and Its Applications. New York: Chapman and Hall.

Ferreira, F. H. G., D. Filmer, and S. Norbert. 2009. "Own and Sibling Effects of Conditional Cash Transfer Programs: Theory and Evidence from Cambodia," Policy Research Working Paper Series 5001. Washington, DC: World Bank.

Filmer, D. and N. Schady. 2009. "School Enrollment, Selection and Test Scores." Policy Research Working Paper Series 4998. Washington, DC: The World Bank.

Filmer, D., L. Pritchett, and A. Hasan. 2006. "A Millennium Learning Goal: Measuring Real Progress in Education." Working Paper No. 97. Center for Global Development.

Fiszbein, Ariel, and Norbert Schady. 2009. *Conditional Cash Transfers: Reducing Present and Future Poverty*. Washington, D.C., The World Bank.

Fernando Gamboa, L. and N. Forero Ramírez. 2008. "Fertility and Schooling: How this Relation Changed between 1995 and 2005 in Colombia." Working Paper 004711, Universidad del Rosario – Department of Economics.

García, S. and J. Hill. 2009. "The Impact of Conditional Cash Transfers on Children's School Achievement: Evidence from Colombia." Universidad de los Andes, CEDE Working Paper Series. Available at SSRN: http://ssrn.com/abstract=1485841

García-Suaza, A.F.; J.C. Guataquí; J.A. Guerra; and D. Maldonado. 2009. "Beyond the Mincer Equation: The Internal Rate of Return to Higher Education in Colombia." Working Paper Series No. 68. Department of Economics, University del Rosario.

Glewwe, P. and P. Olinto. 2004. "Evaluating the Impact of Conditional Cash Transfers on Schooling: An Experimental Analysis of Honduras." Mimeo. Minneapolis, MN: University of Minnesota.

Heinrich, C. J. 2007. "Demand and Supply-Side Determinants of Conditional Cash Transfer Program Effectiveness." *World Development*, Elsevier, 35(1): 121-43.

Imbens, Guido W., and Karthik Kalyanaraman. 2010. “Optimal Bandwidth Choice for the Regression Discontinuity Estimator.” CeMMAP working papers CWP05/10, Centre for Microdata Methods and Practice, Institute for Fiscal Studies.

Imbens, Guido W., and Thomas Lemieux. 2008. “Regression Discontinuity: A Guide to Practice.” *Journal of Econometrics* 142(2): 615-35.

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

Institute for Fiscal Studies, Econometría, and Sistemas Especializados de Información. 2006. “Evaluación del Impacto del Programa Familias en Acción–Subsidios Condicionados de la Red de Apoyo Social.” Bogotá, Colombia: National Planning Department.

Lee, David S., and Thomas Lemieux. 2009. “Regression Discontinuity Designs in Economics.” NBER Working Paper 14723, Cambridge, MA.

Prada, C.F. 2006. “Is the Decision of Study in Colombia Profitable?” Essays on Political Economy No. 51. Prepared for Degree in Economics, Javeriana University.

Rawlings, L. 2002. “Response to a Crisis and Beyond.” In Spectrum 23878: Protecting People with Social Safety Nets. Washington, DC: The World Bank.

Sánchez Torres, F.; Núñez Méndez, J. 2003 “A Dynamic Analysis of Human Capital, Female Work-force Participation, Returns to Education and Changes in Household Structure in Urban Colombia, 1976-1998.” *Colombian Economic Journal*, 1(1): 109-49.

World Bank. 2010. “Evidence and Lessons Learned from Impact Evaluations in Social Safety Nets.” Manuscript. Washington, DC: The World Bank.

World Bank. 2008. Project Appraisal Document on a Proposed Loan in the Amount of $636.5 million to the Republic of Colombia for a Support for the Second Phase of the Expansion of the Program of Conditional Cash Transfers-Familias en Acción Project. Washington, DC: The World Bank.

World Bank. 2005. Project Appraisal Document on a Proposed Loan in the Amount of $86.4 million to the Republic of Colombia for a Social Safety Net Project. Washington, DC: The World Bank.

Table 1. Summary Statistics, t-tests and Normalized Differences by Treatment Status
(Sample for the Matching Analysis)

Variable	Treated		Control		Difference	t-stat	Significance	ND
	Mean	N	Mean	N				
Demographic								
Age (household head)	45.330	2,415	44.940	1,766	0.390	1.076		0.034
Age (spouse)	41.388	2,415	41.002	1,766	0.386	1.115		0.035
Age (child)	12.359	2,420	12.481	1,766	-0.123	-1.746	*	-0.055
Gender (household head)	0.772	2,420	0.840	1,766	-0.068	-5.576	***	-0.173
Gender (child)	0.436	2,420	0.428	1,766	0.008	0.517		0.016
Household structure								
Is household single headed?	0.021	2,420	0.020	1,766	0.002	0.376		0.012
Number of household members	6.072	2,420	6.165	1,766	-0.093	-1.327		-0.042
Number of children	1.337	2,420	1.419	1,766	-0.082	-2.484	**	-0.078
Consumption and assets								
Monthly household consumption	200,000	2,362	212,000	1,697	-12,600	-3.80	***	-0.12
Does the family own the house?	0.667	2,420	0.649	1,766	0.018	1.184		0.037
Education, health and work								
Does household head read?	0.826	2,276	0.831	1,673	-0.004	-0.362		-0.012
Household head completed secondary or more?	0.057	2,267	0.078	1,673	-0.02	-2.493	**	-0.081
Years of schooling (household head)	3.715	2,129	4.037	1,576	-0.322	-2.969	***	-0.099
Did children suffer from diarrhea?	0.112	1,086	0.112	702	0	0.013		0.001
Does household head work?	0.887	2,316	0.885	1,720	0.002	0.182		0.006
Dwelling characteristics								
Located in an urban area?	0.562	2,420	0.695	1,766	-0.133	-8.943	***	-0.278
No walls?	0.007	2,414	0.011	1,765	-0.004	-1.41		-0.045
Connected to piped water?	0.685	2,409	0.762	1,766	-0.078	-5.599	***	-0.174
Connected to gas?	0.12	2,390	0.14	1,759	-0.021	-1.955	*	-0.062
Connected to sewage?	0.334	2,417	0.324	1,766	0.01	0.708		0.022
Community								
Altitude	658.161	2,415	567.069	1,766	91.092	3.846	***	0.121
Students per teacher	22.485	2,415	22.678	1,766	-0.192	-1.139		-0.036
Square metres of classroom per student	2.949	2,415	2.502	1,766	0.448	6.969	***	0.215
Number of banks	1.694	2,369	0.909	1,766	0.784	14.215	***	0.446
Number of health centers	1.134	2,369	0.844	1,766	0.29	6.875	***	0.215
Region = East?	0.214	2,415	0.245	1,766	-0.03	-2.281	**	-0.072
Region = Central?	0.289	2,415	0.162	1,766	0.127	9.953	***	0.307
Region = Pacific?	0.114	2,415	0.127	1,766	-0.013	-1.227		-0.039
Affected by violent attacks?	0.025	2,415	0.031	1,766	-0.006	-1.107		-0.035

Notes: Significant at 90(*), 95(**), 99(***) percent confidence. Summary statistics calculated for households with at least one child (enrolled or not in school) who were 18 or below when she joined the program and that, based on their school attainment at the pre-program time, could have achieved grade 11 between 2003 and 2009.

Table 2. Variables included in Models 1, 2 and 3 as determinants of participation in FA program

Variables	Model 1	Model 2	Model 3
Age	X	X	
Order of the child	X	X	
Household head is married	X	X	X
Household head works	X	X	
Male household head	X	X	
Age household head	X	X	X
Urban	X	X	
Household head years of schooling	X	X	
Number of children ages 7 to 11	X	X	
Number of children ages 12 to 17	X	X	
Monthly expenditures	X	X	
Teacher-pupil ratio in municipality		X	
Classroom space[a]		X	
Resides in most dense part of municipality		X	
Resides in least dense part of municipality		X	
Number of urban schools registered in the municipality		X	
Number of rural Schools registered in the municipality		X	
Live in a rural disperse area			X
Live in a rural populated area			X
Age spouse of household head			
Education of household head: incomplete primary			X
Education of household head: complete primary			X
Education of household head: incomplete secondary			X
Education of household head: complete secondary			X
Education spouse: incomplete primary			X
Education spouse: complete primary			X
Education spouse: incomplete secondary			X
Education spouse: complete secondary			X
Family lives in a house or room			X
Wall materials: Tapia, Abobe or Bahareque			X
Wall materials: wood			X
Wall materials: bad quality wood			X
Wall materials: cardboard or no Walls			X
House has is connected to natural gas			X
House has water pipe			X

Notes: [a] Measured in square meters per student in the municipality.

Table 3. Probability of Participation in FA Program
(First Stage Estimates – 'Fuzzy' RDD)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Same functional form above and below threshold							
Eligibility	0.732***	0.726***	0.731***	0.726***	0.705***	0.714***	
	[0.002]	[0.001]	[0.001]	[0.001]	[0.002]	[0.002]	
Observations	624,028	624,028	624,028	624,028	624,028	624,028	
R2	0.688	0.741	0.688	0.741	0.689	0.742	
Panel B: Different functional form above and below threshold							
Eligibility	0.702***	0.711***	0.706***	0.708***	0.696***	0.702***	0.703***
	[0.002]	[0.002]	[0.002]	[0.002]	[0.003]	[0.002]	[0.008]
Observations	624,028	624,028	624,028	624,028	624,028	624,028	14,647
R2	0.689	0.742	0.689	0.742	0.690	0.742	
Quadratic	Yes	Yes					
Cubic			Yes	Yes			
Quartic					Yes	Yes	
Municipality fixed effects		Yes		Yes		Yes	
Imbens Optimal Bandwidth							0.7573

Notes: Heteroskedasticity-robust standard errors reported in square bracket. Significant at 90(*), 95(**), 99(***) percent confidence. The units of observation are children (enrolled or not in school) who were 18 years old or younger when they joined the program and that, based on their school attainment at the preprogram time, could have achieved grade 11 between 2003 and 2009 and the number of years needed to complete high school was lower than the number of years of treatment. Models include quadratic $[(Si - S^*)^2]$, cubic $[(Si - S^*)^3]$, and quartic $[(Si - S^*)^4]$ specifications of the control function below and above the cutoff of eligibility S*. Optimal bandwidths for non-parametric models were computed following a cross-validation method suggested by Imbens and Kalyamaran (2010)

Table 4 - OLS and Matching Estimates of the Impacts of FA Program on High School Completion

Dependent Variable:School Completion	(1)		(2)		(3)	
	OLS	Matching	OLS	Matching	OLS	Matching
All sample	0.030*	0.0401**	0.050***	0.0840**	0.049***	0.0696 **
	[0.017]	[0.0187]	[0.018]	[0.0220]	[0.017]	[0.0214]
Observations	3,452	3,476	3,452	3,476	3,861	3,888
Boys	0.011	0.0206	0.036	0.0661*	0.041	0.0587*
	[0.026]	[0.0301]	[0.027]	[0.0363]	[0.026]	[0.0348]
Observations	1,478	1,490	1,478	1,490	1,676	1,687
Girls	0.046**	0.0523**	0.065***	0.0856***	0.062***	0.0899***
	[0.022]	[0.0245]	[0.023]	[0.0290]	[0.023]	[0.0290]
Observations	1,974	1,986	1,974	1,986	2,185	2,198
Urban	-0.008	-0.0052	0.015	0.0492*	0.019	0.0391
	[0.021]	[0.0229]	[0.022]	[0.0274]	[0.022]	[0.0254]
Observations	2,102	2,120	2,102	2,120	2,352	2,372
Rural	0.091***	0.0868***	0.115***	0.1044***	0.108***	0.1176***
	[0.028]	[0.0314]	[0.029]	[0.0402]	[0.029]	[0.0398]
Observations	1,350	1,356	1,350	1,356	1,509	1,514

Significant at 90(*), 95(**), 99(***) percent confidence. Bootstrapped standard errors reported in square brackets are obtained from 200 replications.The units of observation are children (enrolled or not in school at baseline) who were 18 or below at baseline (2002) and that, based on their school attainment at the preprogram time, could have achieved grade 11 between 2003 and 2009 and the number of years needed to complete high school was lower than the number of years of treatment. The mean high school completion rate of the control group for the period 2003-2009 is 0.501. Units of analysis are matched on the propensity score from three different specifications of a logistic regression on participation in the program. Preprogram covariates of each specification of the logit models of participation are listed in Table 2.

Table 5. RDD- 2SLS Estimates of the Impacts of *FA* on High School Completion

Dependent Variable: High School Completion 2SLS. Effect of Predicted FEA							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
All sample	0.011***	0.018***	0.026***	0.027***	0.052***	0.039***	0.024**
	[0.003]	[0.003]	[0.004]	[0.004]	[0.005]	[0.005]	[0.011]
Observations	624,028	624,028	624,028	624,028	624,028	624,028	25,249
Imbens Optimal Bandwidth							1.3101
Boys	0.016***	0.023***	0.023***	0.025***	0.045***	0.033***	0.023
	[0.004]	[0.004]	[0.005]	[0.005]	[0.007]	[0.006]	[0.015]
Observations	308,345	308,345	308,345	308,345	308,345	308,345	11,374
Imbens Optimal Bandwidth							1.204
Girls	0.007	0.014***	0.029***	0.029***	0.059***	0.045***	0.018
	[0.005]	[0.005]	[0.006]	[0.006]	[0.007]	[0.007]	[0.017]
Observations	315,544	315,544	315,544	315,544	315,544	315,544	11,191
Imbens Optimal Bandwidth							1.143
Urban	0.000	-0.000	0.018**	0.022***	0.060***	0.044***	0.042**
	[0.006]	[0.006]	[0.007]	[0.007]	[0.009]	[0.009]	[0.019]
Observations	257,689	257,689	257,689	257,689	257,689	257,689	10,202
Imbens Optimal Bandwidth							1.414
Rural	0.026***	0.029***	0.021***	0.022***	0.051***	0.038***	0.021*
	[0.004]	[0.004]	[0.005]	[0.005]	[0.006]	[0.006]	[0.013]
Observations	359,952	359,952	359,952	359,952	359,952	359,952	16,078
Imbens Optimal Bandwidth							1.342
Quadratic	Yes	Yes					
Cubic			Yes	Yes			
Cuartic					Yes	Yes	
Municipality fixed effects		Yes		Yes		Yes	

Notes: Heteroskedasticity-robust standard errors reported in square bracket. Significant at 90(*), 95(**), 99(***) percent confidence. The units of observation are children (enrolled or not in school) who were 18 or below when they joined the program and that, based on their school attainment at the pre-program time, could have achieved grade 11 between 2002 and 2009 and the number of years needed to complete high school was lower than the number of years of treatment. Models include quadratic [(Si - S*)2], cubic [(Si - S*)3], and quartic [(Si - S*)4] specifications of the control function below and above the cutoff of eligibility S*. Optimal bandwidths for non-parametric models were computed following a cross-validation method suggested by Imbens and Kalyamaran (2010).

Table 6. OLS and lower and upper bound Matching Estimates of Program Impacts on Mathematics, Spanish, and Overall Test Scores

Outcome: Test Score	(1)				(2)				(3)			
	OLS	Lower Bound	Control Φ	Upper Bound	OLS	Lower Bound	Control Φ	Upper Bound	OLS	Lower Bound	Control Φ	Upper Bound
Mathematics	-0.401	-0.0189	48	0.033	-0.593	-0.0141	50	0.0374	-0.265	-0.0439	60	0.0179
	[0.392]	[0.0474]		[0.0419]	[0.413]	[0.0545]		[0.0495]	[0.410]	[0.0558]		[0.0464]
Observations	1,867	1,867		1,850	1,867	1,867		1,850	2,047	2,044		2,023
Spanish	0.398	0.0502	48	0.0982**	0.079	-0.0113	50	0.0421	0.246	0.0171	60	0.0689
	[0.350]	[0.0465]		[0.0438]	[0.368]	[0.0902]		[0.0525]	[0.370]	[0.0579]		[0.0567]
Observations	1,867	1,867		1,847	1,867	1,867		1,847	2,047	2,044		2,021
Overall test score	0.179	0.0292	48	0.0698**	-0.086	0.0040	60	0.0413	0.184	0.0171	60	0.0498
	[0.226]	[0.0327]		[0.0328]	[0.237]	[0.0340]		[0.0419]	[0.239]	[0.0361]		[0.0358]
Observations	1,867	1,867		1,850	1,867	1,867		1,850	2,047	2,044		2,022

Notes: Test scores are normalized by the mean and the standard deviation in each subject by semester. The definition of the overall test score excludes results of the foreign language test chosen by the student. Bootstrapped standard errors reported in square brackets are obtained from 200 replications. Significant at 90(*), 95(**), 99(***) percent confidence. The units of observation are children (enrolled or not in school) who were 18 or below at baseline (2002) and that based on their school attainment at the preprogram time could have achieved grade 11 between 2003 and 2009, the number of years needed to complete high school was lower than the number of years of treatment, and registered for the ICFES test. Units of analysis are matched on the propensity score from three different specifications of a logistic regression on participation in the program. Preprogram covariates of each specification of the logit models of participation are listed in Table 2.

Table 7. RDD - 2SLS Estimates of the Impacts of FA Program on Test Scores.

Outcome: Test scores	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Mathematics	-0.024	-0.022	-0.034	-0.034	-0.002	-0.015	-0.048***
	[0.018]	[0.018]	[0.023]	[0.022]	[0.027]	[0.027]	[0.015]
Observations	131,744	131,744	131,744	131,744	131,744	131,744	11,689
Imbens Optimal Bandwidth							0.60
Spanish	-0.044**	-0.039**	-0.037	-0.037	-0.039	-0.048*	-0.100***
	[0.018]	[0.018]	[0.023]	[0.023]	[0.028]	[0.027]	[0.010]
Observations	131,744	131,744	131,744	131,744	131,744	131,744	24,724
Imbens Optimal Bandwidth							1.283
Overall test score	-0.020	-0.020	-0.025	-0.028*	-0.009	-0.025	-0.057***
	[0.013]	[0.013]	[0.016]	[0.016]	[0.020]	[0.019]	[0.009]
Observations	131,744	131,744	131,744	131,744	131,744	131,744	17,031
Imbens Optimal Bandwidth							0.886
Quadratic	Yes	Yes					
Cubic			Yes	Yes			
Quartic					Yes	Yes	
School fixed effects		Yes		Yes		Yes	

Notes: Test scores are normalized by the mean and the standard deviation in each subject by year. The definition of the overall test score excludes results of the foreign language test chosen by the student. Heteroskedasticity-robust standard reported in square bracket. Significant at 90(*), 95(**), 99(***) percent confidence. The units of observation are children (enrolled or not in school) who were 18 or below when they joined the program and that, based on their school attainment at the pre-program time, could have achieved grade 11 between 2003 and 2009, and the number of years needed to complete high school was lower than the number of years of treatment. Models include linear [(Si - S*)], quadratic [$(Si - S^*)^2$], cubic [$(Si - S^*)^3$], and quartic [$(Si - S^*)^4$] specifications of the control function below and above the cutoff of eligibility S*. Optimal bandwidths for non-parametric models were computed following a cross-validation method suggested by Imbens and Kalyamaran (2010)

Table 8. RDD- 2SLS Estimates of the Impacts of FA on Mathematics, Spanish, and Overall Test Scores (by gender and area)

Outcome: Test scores	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Boys							
Mathematics	-0.008	-0.005	-0.007	-0.007	0.039	0.025	0.036**
	[0.029]	[0.029]	[0.035]	[0.035]	[0.042]	[0.041]	[0.015]
Observations	57,637	57,637	57,637	57,637	57,637	57,637	9,319
Spanish	-0.072**	-0.061**	-0.075**	-0.066*	-0.064	-0.057	-0.111***
	[0.028]	[0.028]	[0.035]	[0.034]	[0.042]	[0.041]	[0.012]
Observations	57,637	57,637	57,637	57,637	57,637	57,637	12,828
Overall test score	-0.018	-0.014	-0.029	-0.028	-0.005	-0.015	-0.064***
	[0.021]	[0.021]	[0.026]	[0.025]	[0.031]	[0.030]	[0.010]
Observations	57,637	57,637	57,637	57,637	57,637	57,637	11,619
Girls							
Mathematics	-0.041*	-0.036	-0.051*	-0.049*	-0.035	-0.045	-0.132***
	[0.024]	[0.024]	[0.029]	[0.029]	[0.035]	[0.035]	[0.021]
Observations	74,087	74,087	74,087	74,087	74,087	74,087	6,992
Spanish	-0.024	-0.022	-0.011	-0.012	-0.021	-0.034	-0.073***
	[0.024]	[0.024]	[0.030]	[0.030]	[0.037]	[0.036]	[0.019]
Observations	74,087	74,087	74,087	74,087	74,087	74,087	8,119
Overall test score	-0.024	-0.026	-0.021	-0.025	-0.011	-0.028	-0.068***
	[0.017]	[0.017]	[0.021]	[0.021]	[0.026]	[0.025]	[0.011]
Observations	74,087	74,087	74,087	74,087	74,087	74,087	12,881
Urban							
Mathematics	0.019	0.026	0.020	0.019	0.006	-0.011	-0.063***
	[0.028]	[0.028]	[0.034]	[0.034]	[0.042]	[0.042]	[0.019]
Observations	64,036	64,036	64,036	64,036	64,036	64,036	9,076
Spanish	0.003	0.020	0.014	0.022	0.024	0.024	0.018
	[0.027]	[0.027]	[0.033]	[0.033]	[0.042]	[0.041]	[0.025]
Observations	64,036	64,036	64,036	64,036	64,036	64,036	4,443
Overall test score	0.020	0.034*	0.032	0.037	0.030	0.018	-0.017
	[0.020]	[0.020]	[0.024]	[0.024]	[0.031]	[0.030]	[0.014]
Observations	64,036	64,036	64,036	64,036	64,036	64,036	8,413
Rural							
Mathematics	-0.057**	-0.060**	-0.063**	-0.051	-0.048	-0.064*	-0.017
	[0.028]	[0.027]	[0.032]	[0.032]	[0.039]	[0.039]	[0.011]
Observations	65,523	65,523	65,523	65,523	65,523	65,523	8,925
Spanish	-0.072***	-0.077***	-0.092***	-0.075**	-0.090**	-0.102***	-0.154***
	[0.028]	[0.027]	[0.032]	[0.033]	[0.040]	[0.039]	[0.012]
Observations	65,523	65,523	65,523	65,523	65,523	65,523	14,116
Overall test score	-0.056***	-0.064***	-0.087***	-0.074***	-0.066**	-0.084***	-0.065***
	[0.020]	[0.020]	[0.023]	[0.023]	[0.029]	[0.028]	[0.011]
Observations	65,523	65,523	65,523	65,523	65,523	65,523	9,223
Quadratic	Yes	Yes					
Cubic			Yes	Yes			
Quartic					Yes	Yes	
Municpality fixed effects		Yes		Yes		Yes	

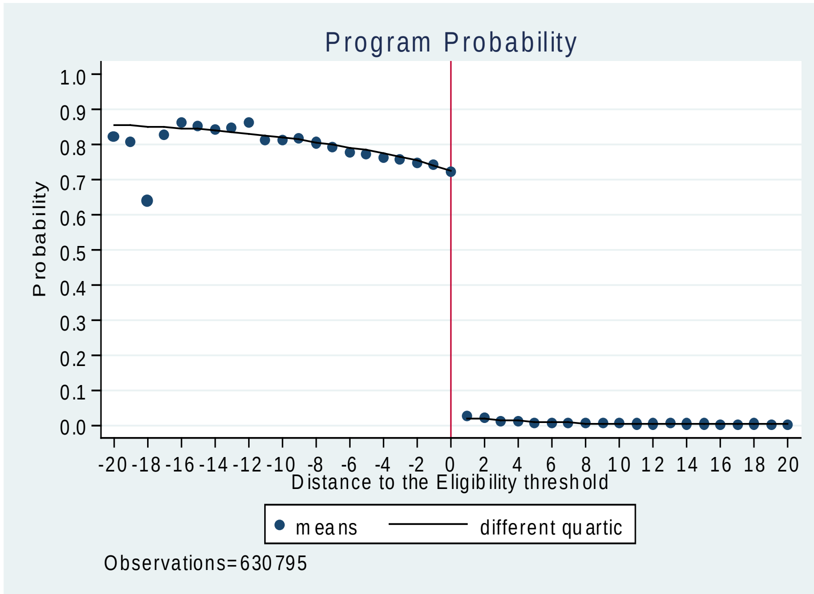
Notes: Scores are normalized by the mean and the standard deviation in each subject by year. The definition of the overall test score excludes results of the foreign language test chosen by the student. Heteroskedasticity-robust standard reported in square bracket. Significant at 90(*), 95(**), 99(***) percent confidence. The units of observation are children (enrolled or not in school) who were 18 or below at baseline (2002) and that based on their school attainment at the preprogram time could have achieved grade 11 between 2003 and 2009, the number of years needed to complete high school was lower than the number of years of treatment and registered for the ICFES test. Models include linear [(Si - S*)], quadratic [$(Si - S^*)^2$], cubic [$(Si - S^*)^3$], and quartic [$(Si - S^*)^4$] specifications of the control function below and above the cutoff of eligibility S*. Optimal bandwidths for non-parametric models were computed following a cross-validation method suggested by Imbens and Kalyamaran (2010).

Table 9. Continuity Checks for Preprogram Household and Individual Level Variables

Household variables			Individual variables			Characteristics of the house		
Variable	Imbens	0.25 points	Variable	Imbens	0.25 points	Variable	Imbens	0.25 points
Household size	-0.145***	0.470	Social Security	0.012	-0.039	Fridge	-0.01	0.092*
	[0.054]	[0.322]		[0.013]	[0.066]		[0.009]	[0.052]
Observations	26,970	9,907		498,138	9,907		23,608	9,897
Kids	0.046*	0.290	Farmer	0.001	0.014	Dining	-0.007	-0.029
	[0.028]	[0.183]		[0.002]	[0.009]		[0.010]	[0.043]
Observations	34,368	9,907		15,660	9,907		16,077	9,907
Household head age	-0.168	-0.657	Male	-0.002	0.030	Bed	0.015	0.033
	[0.193]	[1.560]		[0.014]	[0.084]		[0.028]	[0.134]
Observations	55,962	9,907		24,477	9,907		18,764	9,907
Household head male	0.016	0.017	Age	-0.007	-0.062	Owner	0.039**	-0.056
	[0.012]	[0.064]		[0.080]	[0.476]		[0.017]	[0.077]
Observations	20,076	9,907		34,045	9,907		16,608	9,907
Household head education	-0.033	0.120	Regular Activity	0.087	0.117			
	[0.029]	[0.187]		[0.060]	[0.379]			
Observations	30,456	9,886		35,097	9,907			
Partner education	-0.039	0.060						
	[0.025]	[0.206]						
Observations	32,181	7,749						
Married household	-0.011	-0.033						
	[0.014]	[0.070]						
Observations	23,186	9,907						

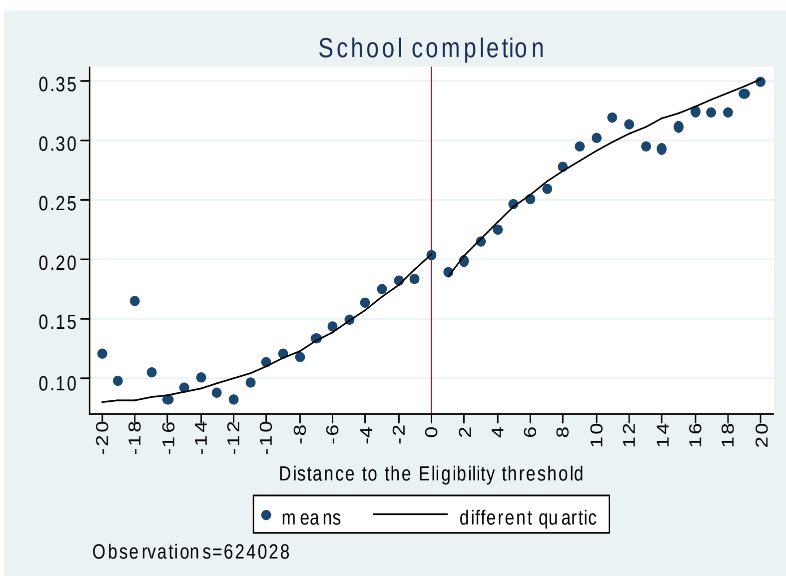
Notes: Heteroskedasticity-robust standard errors reported in square bracket. Significant at 90(*), 95(**), 99(***) percent confidence. Regressions include quartic specifications of the control function below and above the cutoff of eligibility S*. Optimal bandwidths were computed following a cross-validation method suggested by Imbens and Kalyamaran (2010).

Figure 1. Effects of the SISBEN Score on Participation in the Program



Notes: The X axis presents the normalized distance of each child's proxy-means score to the cutoff that is used to classify households as SISBEN 1 and determines eligibility to the program. The Y axis presents the program participation probability.

Figure 2. Impacts of FA on High School Completion (RD Analysis)



Notes: The X axis presents the normalized distance of each child's proxy-means score to the cutoff that is used to classify households as SISBEN 1 and determines eligibility to the program. The Y axis presents the probability of the child completing high school.

Figure 3. Continuity Checks for Household- and Individual-level Variables

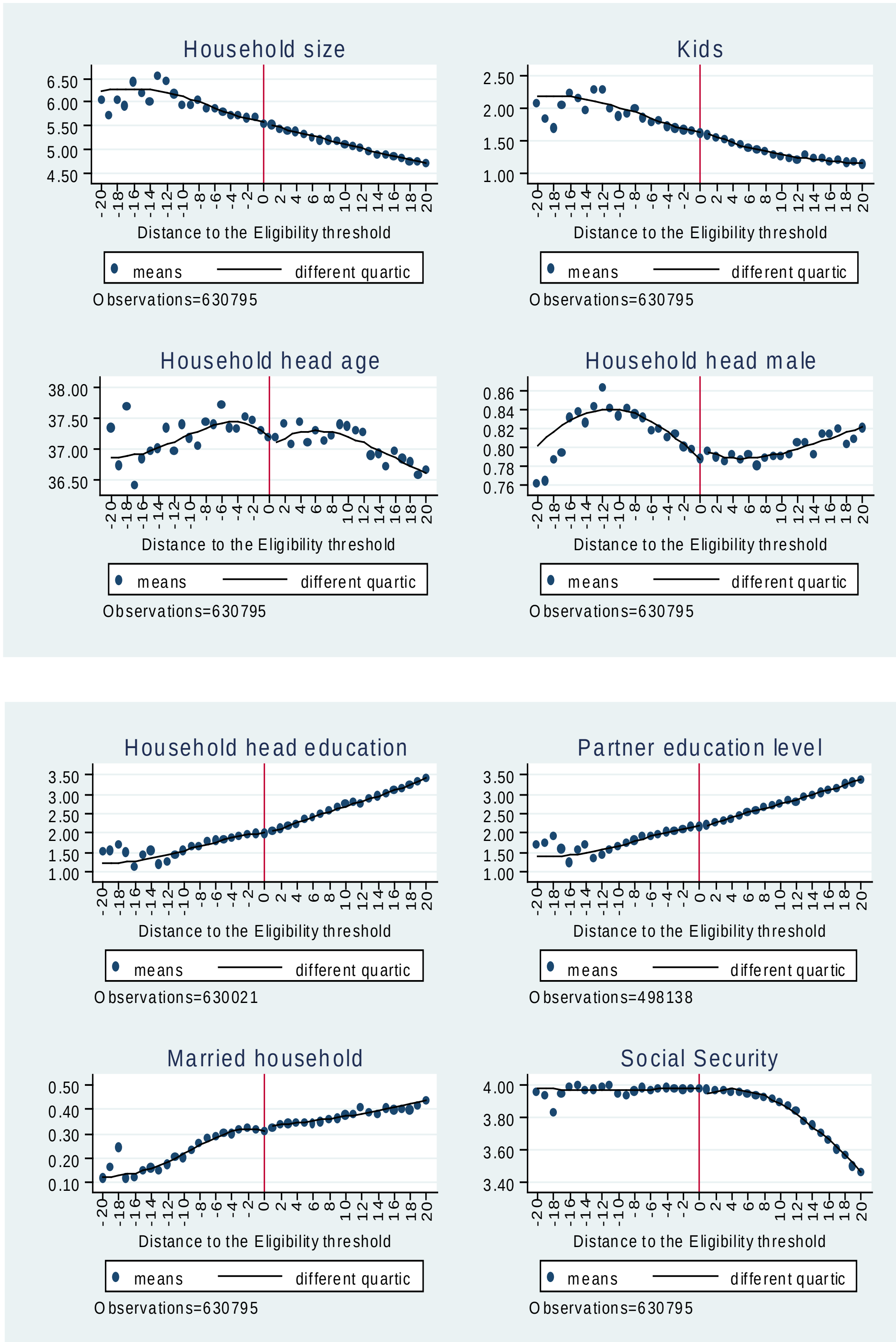


Figure 3. Continuity Checks for Household- and Individual-level Variables (*continued*)

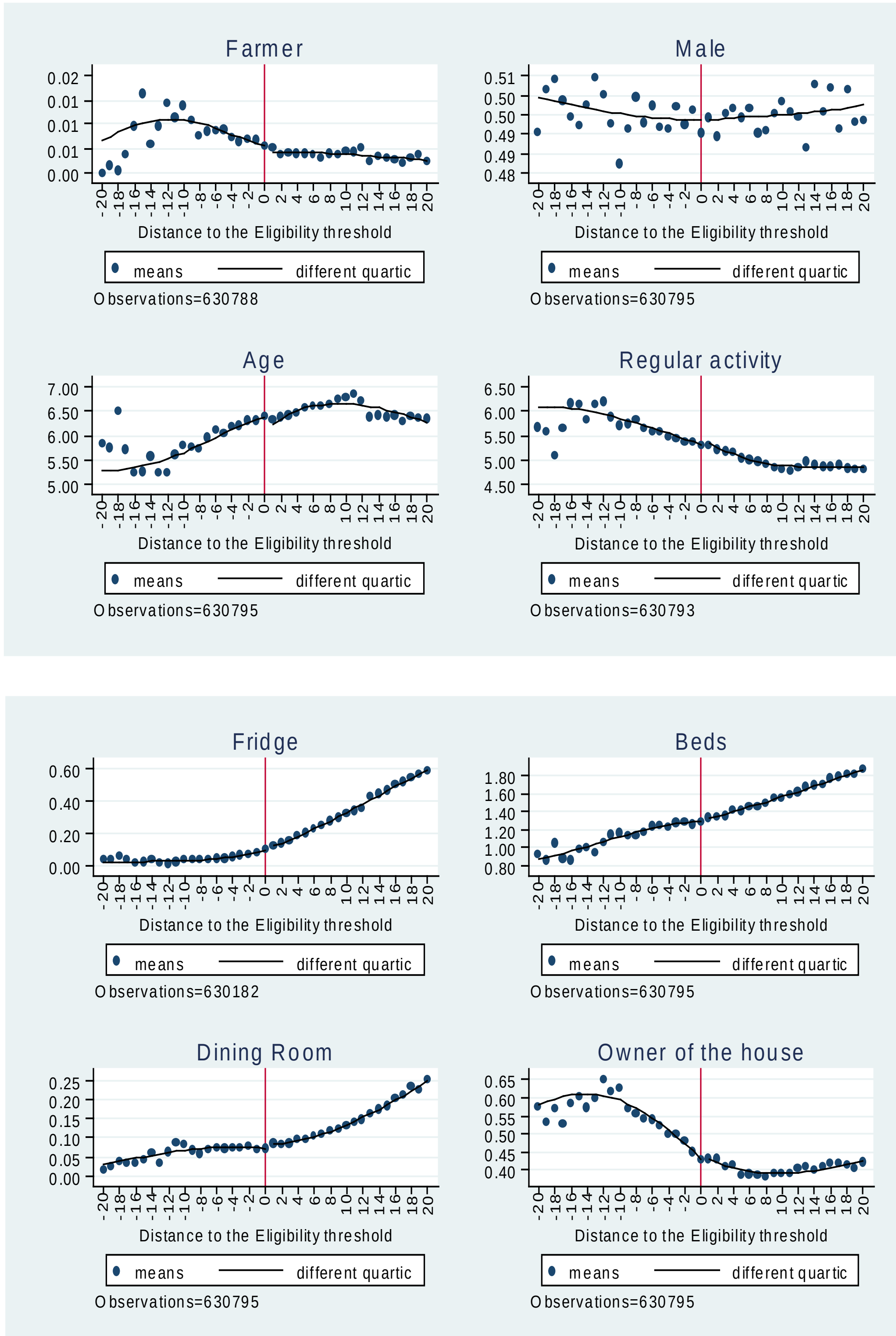
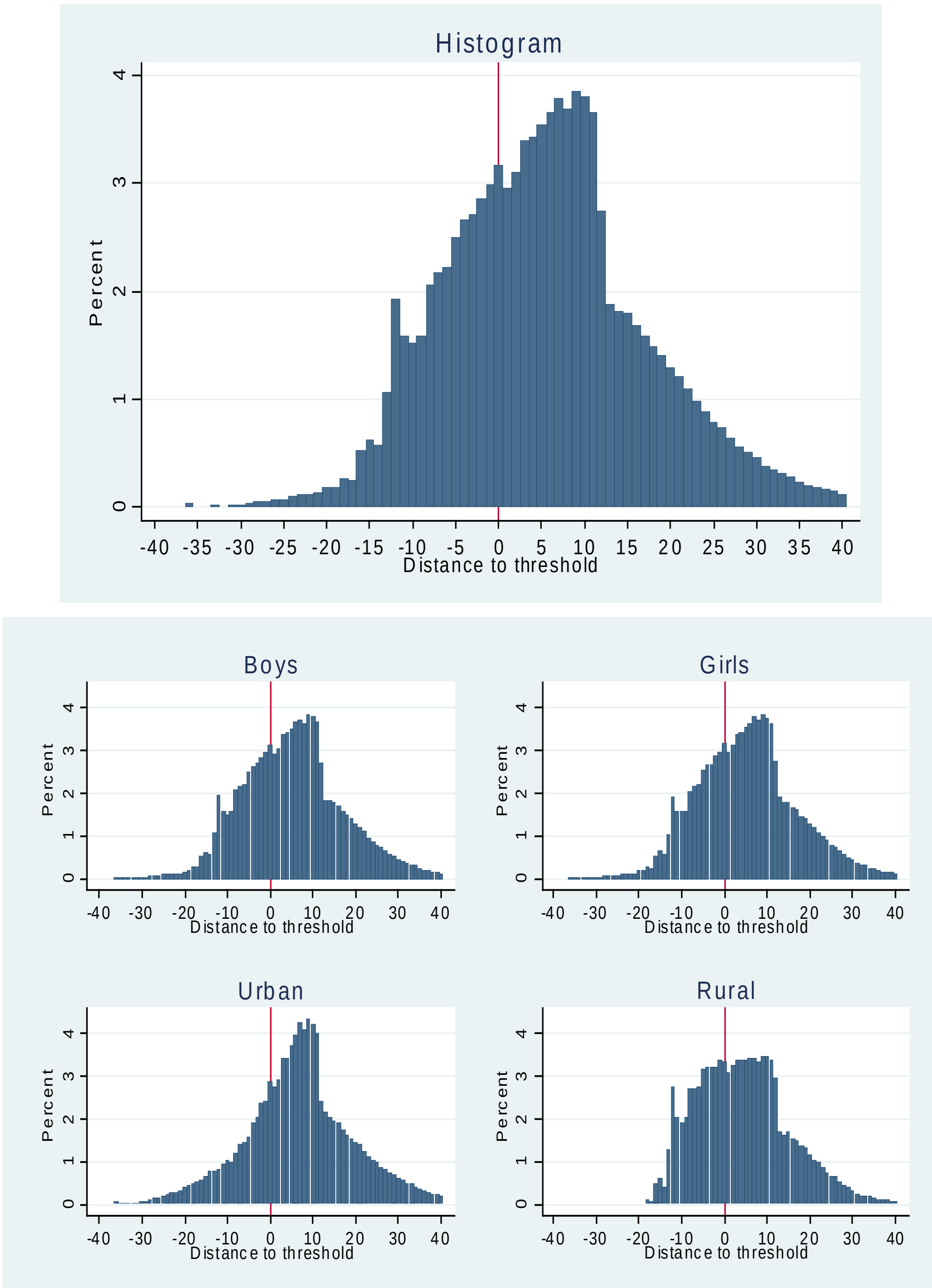


Figure 4. Distribution of the SISBEN Score
(total and by gender and area)



Appendix A -- Data Merging Procedures

Propensity Score Matching. The propensity score matching exercise builds on the short- and middle-term evaluation and utilizes data from the survey collected in 2002 (There are more than 10,000 households,and51,056 individuals who were seven years or older at the baseline. This sample is merged with the ICFES database of around 6 million registrants between 2000 and 2009 to obtain their test scores and estimate whether they have completed high school. The ICFES tests are administered in Colombia twice a year - October and May – and are used as a pre-requisite for enrollment into tertiary education. The majority of students who took the ICFES tests (90 percent) have finished their 11th grade and around 90 percent of high school graduates take the test so the test registration is a good estimator for secondary school graduation (World Bank, 1993; Angrist et al., 2006). Although the ICFES exam takers in 2000 and 2001 could not have participated in the FA, which started in 2002, they are kept in the dataset because they may include older siblings of FA participants for the analysis of the program indirect effects.

Since one of the relevant outcomes is high school completion, different merging strategies are used to enhance the probability of finding the surveyed individuals in the ICFES database, making sure that relevant people are not excluded. If an individual is not matched to any ICFES registration, it is assumed that the child did not take the test, either because they did not finish 11th grade or because they chose not to. The latter is more unlikely because of the high proportion of secondary school graduates that take the ICFES exam (Angrist et al., 2006). There are many difficulties in matching the sampled children with the ICFES registration. However, there is no reason to believe that the matching errors, often due to name and ID mismatches, would be systematically different between the treatment and control groups.

Four different merging methods are employed:

- The first method uses only the ID numbers reported in the surveys as the matching criterion. This is a unique ID number assigned to all citizens of Colombia when they turn 11-12 years old. However, the ID numbers change when individuals turn 18, which may result in the failure to match individuals who took the ICFES test at or after 18 years of age. Furthermore, since an ID includes 11 digits, there are, expectedly, many occasions of IDs being misreported in the evaluation surveys. This merge, consequentially, gives only 4,048 matched observations.
- The second method uses only full name as the matching criterion with the probability of orthographical mistakes (for instance, Catherine vs. Katherine). While this may resolve the issue of ID change or misreporting as observed in the first method, it has some potential mismatches due to a number of common last names in Colombia. It is therefore important to be cautious of the likelihood of matching different individuals with the same name. This merge results in 6,563 matches.
- The third method uses both the two last names and date of birth for merging. Again, due to many common last names in Colombia, this method does not guarantee unique matches. The merge provides 46,360 observations.
- The final method uses the two last names and the first 7 digits of the ID number for the merge.[39] For children under 18 years of age, the first 6 digits correspond to the date of birth so this strategy is potentially more accurate than the third method of using last names and date of birth. With shorter IDs, it is also expected that there are fewer misreporting cases than in the first method. However, since the short ID numbers are not unique, this strategy has similar issue with the full name merge. It provides 5,927matches.

The results from all four matches are used to minimize the probability of exclusion. In order to ensure that the matches are correct, the four merges are appended and subject to three cleaning processes. First, the

[39] The IDs distributed before 2004 have 9 digits. We used 7 digits in order to maximize the number of merges we could get.

records with the exact name and similar date of birth are kept (either the same date and month of birth and within four years of birth, or the same year of birth and within two months and two days of birth). Second, those that do not fulfill the first check are tested whether they have the exact date of birth and similar name (again, to account for orthographical mistakes). Finally, among those that fail both tests, the observations with similar name and similar date of birth are kept to account for mistakes in recording.
After each cleaning, duplicates are checked within both the evaluation survey sample and the ICFES dataset. Only the duplicates across both datasets are deleted to avoid eliminating individuals that took the exam multiple times. The result is 5,022 observations. When multiple test scores are found for one individual, the first test result is kept, which produces 4,820 records corresponding to unique individuals. This final data set makes up the sample used in the analysis.

The accuracy of the merges is tested using the information on students enrolled in schools in the evaluation surveys. This test follows the cohorts of students who in principle could have completed the 11th grade between 2000 and 2009, assuming no grade repetition, and obtains 5,395 records. Again, this includes students who could have finished high school in 2000 and 2001 since they serve as the analysis group for another outcome. Incorporating the average dropout rates of students from grades 7 to 11, there should be 4,073 individuals who completed the 11th grade within this time period. The final sample of 4,820 individuals obtained from the merging and cleaning process comes quite close to this estimate.
Among the 4,820 matched individuals, 3,002 have exactly the same full name and date of birth. Among the 5,022 matched records, 67 percent are 18 years old or younger and 88 percent are under 20 at the time they took the test. Most of the observations correspond to tests taken in the second semester of the year when public schools administer the exam, which is consistent with the fact that most of FA beneficiaries attend public schools.

Regression Discontinuity Design. As for the RD analysis, the merge uses three set of administrative data: (1) the System of Information of Beneficiaries of the program (SIFA) provided by Acción Social; (2) the Poverty Index Score Survey collected between 1994 and 2003 (SISBEN); and (3) records on registration for the national ICFES test. The following steps describe the use of these data:

- SIFA is used to construct the treatment groups whereas SISBEN provides the control group – individuals under and above the threshold of eligibility, respectively.
- After running the same merging procedures followed in the PSM analysis, the merge is able to identify that 95 percent of the matching distribution was born between the years of 1975 and 1994. This information is used to restrict the sample to those individual that are mostly likely to be merged between SIFA+SISBEN and ICFES.
- The analysis focuses on information from the SISBEN survey that was collected between 1994 and 2003, since *FA* targeted the beneficiaries during the first phase of the program with the scores from the first version of the poverty score index. For consistency of result, the sample is restricted in this way to evaluate comparable children who joined the program during the first expansion that took place between 2002 and 2003; these children come from the records of SIFA in 2001-06.
- The same four merging strategies employed in the PSM analysis were followed to guarantee the comparability of the outcome variable that measures secondary school completion.
- Individuals with score zero were excluded from the analysis for two reasons: First, it is not possible to establish whether a score zero is the result of a problem in the algorithm -- the probability of getting 0 is very low. Second, the probability of getting *FA* is much lower for this group when compared with the probability of receiving the treatment for people with scores equal to one.

Appendix B – Methodology for Nonparametric Bounds of Program Impacts

This paper uses the results of ICFES, a mandatory academic test given to students in school, to infer the impacts of FA on learning outcome. Such approach is problematic from a methodological standpoint because the program increases school enrollment and attendance among beneficiaries and with it the probability that they take the exam. Participant children are therefore more likely than nonparticipants to register for the ICFES test, which makes the scores distributions of beneficiaries and nonbeneficiaries not comparable. The intuition behind this is a selection bias created by the 'marginal' child who is brought into school due to the incentives of the program and may be different in many dimensions (socio-economic background, inner ability, motivation, expected returns to schooling, etc.) than those already enrolled in school. For these reasons, a simple comparison of participant and nonparticipant children may be deceptive and – given that the bias is expected to be negative – would probably underestimate the actual effect of the program on learning.

In order to address this identification issue, this evaluation uses bounding procedures on specific quantile average treatment effects estimated with matching techniques to correct for the selection bias brought about by the likely introduction of low-scorers into the group of program beneficiaries who ended up taking the test (Lee, 2002; Angrist et al., 2006). The two key assumptions of the procedure are: (1) independence of the treatment status and the errors in the outcome and selection equations (expected to be addressed by the PSM strategy); (2) a monotonicity condition in the sense that assignment to treatment affects the outcome in only one direction, namely that the program does not reduce the test scores of program participants:

$$S_{1i} \geq S_{0i} \text{ for all } i$$

where S, 1 and i denote test scores, treatment status and individuals, respectively. Following Angrist and others (2006), students are assumed to choose to take the test if their expected scores are above a certain threshold so that the quantiles of test-takers are identified from the quantiles of non-takers ($S_{tk} = 0$ where $t = 0,1$ indexes treated and control children and k equals 1 if the student takes the test) for $\varphi \geq \varphi_k$, where $q_k(\varphi_k) = 0$. The φ-quantile of the distributions of non-participants and participants are denoted by $q_0(\varphi_0)$ and $q_1(\varphi_1)$, respectively.

The main idea of the procedure is to find the quantile φ_0 for the control group such that $q_0(\varphi_0) = 0$ and restrict the distribution of S_1 to the percentiles above φ_0. This defines the upper bound based on the subsample of individuals who would have taken the test regardless of the program $E[S_1 - S_0|k = 1]$ as follows:

$$E[S|D = 1, S > q_0(\varphi)] - E[S|D = 0, S > q_0(\varphi)]$$

Unadjusted comparisons between treated and control children – conditional on positive test scores for each group – provide lower bounds on the actual effects of the program:

$$E[S|D = 1, S > q_1(\varphi)] - E[S|D = 0, S > q_0(\varphi)]$$

Since the problem of selection is expected to be more prevalent at the bottom of the distribution, upper bounds should be tighter at upper parts of the distribution. Angrist (1997) shows that under the assumptions (1) and (2), the symmetric truncation of the score distributions is expected to eliminate the sample selection bias. As noted above, this equivalent to restricting the sample to the individuals thought to be the 'always takers'