

# Transferencias Monetarias Condicionadas y desempeño académico: ¿Escogen las familias beneficiarias mejores colegios?\*

Isabel Hincapié Correa<sup>1</sup>

Fabio Sánchez Torres<sup>2</sup>

## Resumen

La literatura reciente sobre Transferencias Monetarias Condicionadas ha encontrado efectos positivos de los programas sobre variables educativas: la permanencia de los niños en los colegios, la asistencia y la progresión escolar (Attanasio, 2005; Baez & Camacho, 2011; De Brauw et al., 2015). Sin embargo, existe poca evidencia acerca de los mecanismos que pueden explicar tales resultados. El presente trabajo evalúa diferentes mecanismos. En particular, estudia si el programa de Transferencias Monetarias Condicionadas, Familias en Acción, aumentó la demanda por instituciones educativas de mayor calidad, en los niveles de educación básica, por parte de las familias beneficiarias. Para ello se hace uso de la metodología de regresión discontinua, en la cual se utiliza como variación exógena los criterios de elegibilidad del programa. Los resultados sugieren que los niños beneficiarios del programa en su primera fase asistieron a colegios de mayor calidad. Ser beneficiario de Familias en Acción aumenta entre 13 y 15 p.p. la probabilidad de asistir a un colegio de mayor calidad. Este mecanismo explica entre el 12% y el 20% del efecto sobre variables educativas en el largo plazo ya estudiadas por la literatura. De igual forma, se encuentra que el programa aumenta la probabilidad tanto de asistir a un colegio que mejora su calidad en el tiempo como de cambiarse a instituciones de mayor calidad. Estos efectos corresponden a 6.7 y 3.1 p.p. respectivamente.

Palabras Claves: Familias en Acción, Acceso educativo, Calidad educativa, Prueba Saber 11

Clasificación JEL: I25, I28, I29, I38

---

\* Los autores agradecen los comentarios realizados por Adriana Camacho y Ana María Díaz. También las sugerencias de Román David Zarate y Román Andrés Zarate.

<sup>1</sup> Universidad de los Andes, [i.hincapie2268@uniandes.edu.co](mailto:i.hincapie2268@uniandes.edu.co).

<sup>2</sup> Universidad de Los Andes y CEDE [fasanche@uniandes.edu.co](mailto:fasanche@uniandes.edu.co)

# Conditional Cash Transfers and academic performance: ¿Are the beneficiary families choosing better schools?\*

Isabel Hincapié Correa<sup>3</sup>

Fabio Sánchez Torres<sup>4</sup>

## Abstract

The recent literature on Conditional Cash Transfers has found positive effects on educational outcomes: the permanence of children in schools, attendance and school progression (Attanasio, 2005; Baez & Camacho, 2011; De Brauw et al. 2015). However, there is little evidence about the mechanisms that may explain such results. The present document assesses different mechanisms. In particular, it examines whether the beneficiary families have increased the demand for higher quality educational schools due to the Conditional Cash Transfer program *Familias in Acción*. The methodology consists of a discontinuity regression, in which the eligibility criteria are used as exogenous variation. The results suggest that children in the first phase of the program attended higher quality schools. Being a beneficiary of *Familias en Acción* increases the probability of attending a higher quality school between 13 and 15 p.p.. This mechanism explains between 12% and 20% of the effect on long term educational variables already studied by the literature. Likewise, it is found that the program increases the probability of both, attending a school that improves its quality over time and of switching to schools of higher quality. These effects correspond to 6.7 and 3.1 p.p. respectively.

Key Words: Familias en Acción, Access to Education, Quality Education, Prueba Saber 11

JEL Classification: I25, I28, I29, I3

---

\* The authors would like to acknowledge the comments made by Adriana Camacho and Ana María Díaz. The suggestions from Roman David Zarate and Román Andrés Zarate are also very appreciated.

<sup>3</sup> Universidad de los Andes, [i.hincapie2268@uniandes.edu.co](mailto:i.hincapie2268@uniandes.edu.co).

<sup>4</sup> Universidad de Los Andes y CEDE [fasanche@uniandes.edu.co](mailto:fasanche@uniandes.edu.co)

## 1. Introducción

Los impactos de los programas de Transferencias Monetarias Condicionadas sobre distintas variables educativas han sido ampliamente estudiados. Entre estos se encuentran la permanencia de los niños en los colegios, la asistencia y la progresión escolar, etc (Attanasio, 2005; Baez y Camacho, 2011). Sin embargo, existe poca evidencia sobre los mecanismos que explican dichos efectos. En general, no se conocen evaluaciones que analicen el impacto de Familias en Acción (FeA) sobre potenciales mecanismos de calidad educativa. El presente estudio tiene como objetivo estudiar un potencial mecanismo que permita explicar los efectos de programas de Transferencias Monetarias Condicionadas sobre variables educativas. En particular, se busca determinar si los beneficiarios de FeA asistieron a mejores colegios oficiales explotando la variación cuasi experimental en el punto de corte de elegibilidad del programa, el cual se determina a través del puntaje SISBEN. Adicionalmente, se evalúa el efecto del programa sobre otros potenciales mecanismos a nivel individual, tales como la tenencia de computador o acceso a internet por familia.

De igual forma, para ser consistente con evaluaciones pasadas de esta literatura, se mide el efecto del programa sobre variables educativas de resultado en el largo plazo a nivel individual. Específicamente, sobre la probabilidad de graduación del colegio y el puntaje en la prueba Saber 11. Posteriormente, se procede a estimar, a través de modelos de mediación, qué porcentaje del efecto sobre estas variables se explica a través de la asistencia a mejores colegios oficiales.

En términos generales, existen diferentes potenciales mecanismos de calidad educativa que podrían explicar los resultados de los niños beneficiarios. Por ejemplo: 1) una mejoría en la calidad por parte de los colegios a los que asisten los niños; 2) los “efectos de pares”, los cuales hacen referencia a que los niños estarían interactuando con niños de mejores habilidades y 3) los beneficiarios de FeA estarían asistiendo a colegios que brindan mayor calidad. Este trabajo busca identificar si los niños beneficiarios de FeA en su primera fase asistieron a colegios de mayor calidad.

El análisis parte desde la perspectiva de la demanda, ya que analiza la forma en que reaccionaron los padres de familia a raíz de ser elegibles para el programa, bajo el argumento de que escogieron mejores colegios para sus hijos. En términos económicos esto implica que la

población beneficiaria de FeA generó un cambio en la valoración que se le da a la calidad de los colegios. De esta forma, dichos programas podrían incidir en la percepción de la población sobre el valor de la educación.

Existen estudios que han analizado potenciales mecanismos de calidad educativa desde el punto de vista de la oferta. En especial, la forma en que la provisión de recursos ofrecida por los colegios se ve afectada a raíz del programa (Benson, 2012). El presente estudio contribuye a la literatura a partir de una perspectiva hasta ahora ignorada; su principal objetivo es explicar potenciales mecanismos de los programas de Transferencias Monetarias Condicionadas desde la demanda. Para esta literatura es relevante el análisis tanto desde la oferta como desde la demanda para determinar qué factores explican los resultados encontrados sobre variables educativas. Es así, que el presente documento es un paso más para explicar los distintos mecanismos que podrían generar impactos gracias a dichos programas.

Para evaluar los mecanismos de calidad educativa mencionados anteriormente y las variables de desempeño escolar, este estudio utiliza diferentes fuentes de información de datos administrativos. La primera es la encuesta del SISBEN (Sistema de Identificación de posibles Beneficiarios de Programas Sociales). La segunda son los datos administrativos del sistema de información de Familias en Acción, que permite identificar los beneficiarios del programa. La tercera es el SIMAT (Sistema Integrado de Matrículas Estudiantil), cuyos datos permiten establecer a qué colegio asiste los niños. La última base de datos son los resultados de la prueba Saber 11, los cuales permiten construir las variables de resultado del presente estudio.

Con esta información se busca evaluar el impacto de FeA a través de la metodología de regresión discontinua. Se decide utilizar esta metodología, ya que a los hogares se les asigna una puntuación del índice de pobreza según el SISBEN y es éste el que determina su elegibilidad en el programa. Esto permite explotar la variación cuasi experimental del umbral establecido con el fin de estimar los efectos que este tuvo sobre las variables de interés.

Los resultados sugieren efectos positivos de FeA en el umbral de elegibilidad. La probabilidad de asistir a un colegio de mayor calidad para los beneficiarios de Familias en Acción aumenta entre 13 y 15 p.p. Así, a raíz del programa los niños beneficiarios si están asistiendo a mejores colegios. Sin embargo, si bien estos colegios son de mayor calidad, no están ubicados en la parte derecha de la distribución de Saber 11, ya que no se encuentran

efectos positivos cuando se evalúa el mecanismo de asistencia a estos colegios. Adicionalmente, se encuentra que el programa aumenta la probabilidad tanto de asistir a un colegio que mejora su calidad en el tiempo como de cambiarse a instituciones de mayor calidad. Estos efectos corresponden a 6.7 y 3.1 p.p. respectivamente. Los resultados son robustos a diferentes anchos de banda y municipios focalizados<sup>5</sup>. A su vez, respecto a mecanismos a nivel individual, como lo es la tenencia de activos fijos en educación, se encuentra un efecto positivo más no significativo. Finalmente, al llevar a cabo el ejercicio para cuantificar el efecto de los mecanismos sobre las variables de desempeño académico se encuentra que el efecto del programa sobre la asistencia a mejores instituciones explica entre 12% y 18% del efecto sobre graduación y entre un 20% y 23% sobre los resultados de la prueba Saber 11.

El resto del documento se estructura de la siguiente manera. La sección 2 presenta una revisión de la literatura relacionada. La sección 3 describe el contexto institucional colombiano en el que se diseñó FeA, su relación con el SISBEN y la prueba Saber 11. La sección 4 presenta las principales fuentes de datos para este trabajo. En la sección 5 se explica la metodología que permite identificar el impacto de FeA. La sección 6 presenta los principales resultados y la sección 7 corrige el posible sesgo de selección y presenta los resultados de las pruebas de robustez. Por último, la sección 8 concluye.

## **2. Literatura relacionada**

El presente estudio se relaciona con tres ramas de la literatura: 1) impactos de programas de Transferencias Monetarias Condicionadas en variables educativas; 2) efectos de programas de Transferencias Monetarias Condicionadas en la oferta educativa en Colombia y 3) efectos de “agrupamiento” en los resultados educativos de los niños y en la calidad de los colegios.

La primera rama de la literatura con la que se relaciona este estudio es aquella que ha evaluado el impacto de Transferencias Monetarias Condicionadas sobre los resultados educativos de los niños. En general, existen diversos estudios que han encontrado efectos positivos de los programas de Transferencias Monetarias Condicionadas en variables educativas. Por ejemplo, para el caso colombiano, Attanasio (2005) encuentra efectos positivos

---

<sup>5</sup> Los efectos fueron estimados sobre dos muestras diferentes: la primera contiene todos los municipios del SISBEN y la segunda únicamente los que fueron focalizados en la primera fase del programa.

sobre la asistencia escolar y la probabilidad de matricularse por parte de los beneficiarios; y Camacho y Baez (2011) encuentran efectos positivos del programa Familias en Acción (FeA) sobre variables de desempeño en el largo plazo. En particular, la probabilidad de presentar la prueba Saber 11 aumenta entre 3.3 y 4.5 p.p. en el punto de corte de elegibilidad para los beneficiarios.<sup>6</sup> Similarmente, para el contexto internacional, también existe evidencia empírica positiva. Por ejemplo, De Brauw *et al.* (2015) encuentran, a través de una metodología de emparejamiento, efectos positivos sobre participación escolar y progresión escolar del programa de Bolsa Familia en Brasil. Galassi (2015) también encuentra, a través de una metodología de diferencias en diferencias, efectos positivos de Bolsa Familia sobre las pruebas Pisa en Brasil; mediante una comparación entre las regiones con una mayor proporción de beneficiarios antes y después del programa. Por último, un estudio del Banco Mundial (2010) resume los resultados de diferentes programas de Transferencias Monetarias Condicionadas en el mundo: el estudio concluye que dichos programas generan un efecto positivo en la asistencia escolar de los niños.

La segunda rama de literatura estudia la relación de FeA y programas sociales con la oferta educativa. Benson (2012) analiza por medio de una metodología de diferencias en diferencias los efectos de FeA sobre la oferta escolar. El estudio encuentra que la oferta educativa se ha deteriorado por falta de inversión en los colegios. De igual forma, la literatura de colegios también estudia efectos de equilibrio general inducidos por programas sociales, ya que la competencia entre colegios podría aumentar a raíz del programa y como resultado generar cambios en la calidad de las instituciones. En este sentido, Neilson (2013) encuentra efectos positivos sobre las pruebas estandarizadas de Chile por el aumento de una transferencia a los niños de bajos recursos en 2008. El autor establece que 2/3 del efecto se explica a través de una mayor competencia entre colegios a raíz del aumento en la transferencia.

La última rama de literatura está relacionada con los efectos de “agrupamiento” de los niños. El término “agrupamiento” se refiere a la posibilidad que tienen los padres de familia de elegir el colegio de sus hijos, por lo que niños con ciertas características deciden agruparse en ciertos colegios.<sup>7</sup> Respecto a esto, diferentes estudios han evaluado los efectos de

---

<sup>6</sup> Esta variable captura la probabilidad de graduarse del colegio, ya que la prueba Saber 11 sólo la presentan aquellos estudiantes con aspiraciones a terminar la educación básica.

<sup>7</sup> En la literatura anglosajona esto se conoce como “*sorting*”.

“agrupamiento” sobre resultados escolares. Por ejemplo, Kim *et al.* (2008) encuentra que cuando los niños pueden elegir a que colegio asistir aumentan sus puntajes escolares. Similarmente, Hoxby (2002) sugiere que un conjunto más grande de alternativas incrementa la competencia entre colegios generando efectos positivos en pruebas estandarizadas de los niños. Por último, Urquiola y Verhoogen (2009) encuentran efectos positivos de “agrupamiento” a través de un modelo de elección de colegios. En este sentido, este estudio busca determinar si la forma en que los beneficiarios de Familias en Acción se agrupan en diferentes colegios cambió a raíz del programa de Transferencias Monetarias Condicionadas.

### **3. Contexto institucional**

Esta sección describe el contexto institucional en el cual se diseñó tanto la encuesta SISBEN, como el programa Familias en Acción y la relación entre ambos para determinar los diferentes criterios de elegibilidad del programa de Transferencias Monetarias Condicionadas.

#### **SISBEN**

La constitución Nacional de 1991 de Colombia, dentro del ámbito del estado social de derecho, establece la necesidad de focalizar o dirigir el gasto social a la población más pobre y vulnerable por parte del gobierno nacional y de los gobiernos departamentales y locales. Por esta razón, la administración pública debe contar con mecanismos técnicos que permitan la identificación y selección de posibles beneficiarios para los programas sociales (Gobierno de Cundinamarca, 2008). Es así, que durante la década de los 90 nace la primera versión del SISBEN<sup>8</sup>, que estuvo vigente entre 1994 y 2003, y buscaba focalizar el gasto público para garantizar que el gasto social sea asignado a los grupos de población más pobres y vulnerables.

Esta encuesta es utilizada por el Gobierno para establecer los criterios de elegibilidad de diferentes programas sociales a través de un índice que se construye a partir de diferentes características socioeconómicas de los hogares. El SISBEN permite establecer el grado de pobreza de los hogares por medio de este índice, el cual es un puntaje entre 0 y 100. Una familia es más pobre cuanto más se acerca a un puntaje de 0 y más rica cuanto más se

---

La literatura de elección de colegios y como los niños se agrupan en diferentes colegios ha estado muy activa a partir de la discusión planteada por Friedman (1962). En su libro, Friedman señala que, en términos de política pública, es más eficiente brindarles transferencias a los padres para elegir el colegio de sus hijos que proveer los servicios educativos a través de colegios públicos.

<sup>8</sup> Sistema de Identificación de Potenciales Beneficiarios de Programas Sociales.

aproxima a 100 (Secretaría Distrital de Salud, 2015). En otras palabras, un aumento en el valor del indicador representa una mejora en dichas condiciones, por lo que hogares más vulnerables obtienen los puntajes más bajos. Los gobiernos locales y central definen diferentes umbrales de este puntaje para la focalización de los programas sociales, entre ellos, Familias en Acción.

## **Familias en Acción**

El programa Familias en Acción es un programa de subsidios condicionados establecido por el Gobierno nacional en junio de 2001. Este surge como resultado de la disminución del bienestar de los colombianos menos favorecido tras la crisis económica de final de siglo (particularmente 1999). Los objetivos establecidos fueron: (a) reducir la inasistencia y deserción de los alumnos de educación primaria y secundaria; (b) complementar el ingreso de las familias con niños menores de siete años en extrema pobreza para incrementar el gasto en alimentación; (c) aumentar la atención de salud de los niños menores de siete años; (d) mejorar prácticas de cuidado de niños en aspectos tales como salud, nutrición, estimulación temprana y prevención de la violencia intrafamiliar (Banco Mundial, 2003).

Los subsidios se crearon para cumplir con estándares mínimos de salud y nutrición para niños menores de 7 años, y a la asistencia a la escuela por parte de niños entre los 7 y los 18 años. A continuación, el documento explica la primera fase del programa, cuyos beneficiarios corresponden a la población de análisis del presente estudio.

### **Familias en acción: Fase 1**

Para esta primera fase se estableció una meta de 340.000 familias beneficiarias. Para ello, se preseleccionaron 686 municipios aplicando ciertos criterios de elegibilidad. Entre estos, mayor población SISBEN 1 en relación con la población inscrita en SISBEN y mayor agilidad en el proceso de conformación de la unidad coordinadora Regional, entidad encargada de manejar el programa a nivel local. Finalmente se obtuvieron 300 municipios beneficiarios en los que se esperaban cumplir con el número de familias inicialmente establecidas. Sin embargo, por diferentes motivos entre estos la falta de conocimiento y poca credibilidad en el programa, esta cuota no se alcanzó. Es por esta razón, que el Gobierno reajustó el número de municipios elegibles a 627 en su primera parte. Los primeros giros fueron recibidos en 2001 por un valor de \$14.000 para estudiantes en primaria y \$28.000 para secundaria. Por concepto de salud se recibía \$46.500 sin importar el número de hijos (DNP, 2010).



En 2003, dado que Familias en Acción había logrado grandes avances en consumo, salud y educación, se decidió extender el programa hasta 2006. Para poder aumentar la cobertura se decidió incluir nuevos municipios. Entre estos beneficiados por otros programas nacionales, capitales de departamentos con población mayor a 100.000 habitantes y permitieron que el municipio con contaran con entidades bancarias. Finalmente, en la fase 1 de Familias en Acción se logró la vinculación de 848 municipios de los cuales salieron 778.000 familias inscritas y 666.000 familias beneficiadas. De éstas, 83% fueron familias de SISBEN 1 y 17% familias desplazadas. (DNP, 2010)

Familias en Acción se convirtió en el principal instrumento del Gobierno para la superación de la pobreza en zonas rurales. Por esta razón, en 2007, se realiza la segunda fase del programa. Así mismo, por medio de la Ley 1532 del 2012, se establece la permanencia del programa Familias en Acción indefinidamente, convirtiéndose en una política de Estado.

#### **4. Datos**

En este trabajo se utilizan cuatro fuentes de información de datos administrativos: 1) SISBEN, 2) Familias en Acción, 3) SIMAT (Anexo 6A de la Reforma 166 del Ministerio de Educación) y 4) Saber 11. De esta forma, la base final consiste en el cruce de estas fuentes.<sup>9</sup> La primera fuente es la población del SISBEN 1, la cual consiste en todos los potenciales beneficiarios de programas sociales. Estos datos proveen información de diferentes características socioeconómicas tanto a nivel de hogar, como a nivel individual, para el año en el que se les realizó la encuesta. Algunas de las características observadas son: edad, género, tamaño del hogar, posesión de diferentes activos e ingresos. Como se mencionó en la sección de contexto institucional, estos datos cuentan también con información de un índice de pobreza que el Gobierno colombiano construye con base en las características socioeconómicas. Este puntaje es un mecanismo de asignación de los diferentes programas sociales.<sup>10</sup>

La segunda fuente de información son los datos administrativos de Familias en Acción, el programa de Transferencias Monetarias Condicionadas en Colombia, en su primera fase.

---

<sup>9</sup> Las estadísticas descriptivas se presentan para la muestra final, es decir, para la muestra sobre la cual se realizan las estimaciones.

<sup>10</sup> Para el SISBEN 1 el algoritmo se puede encontrar en Castañeda (2005). Para las siguientes encuestas de SISBEN el algoritmo es confidencial del Departamento Nacional de Planeación (DNP).

Estos datos contienen todos los beneficiarios del programa entre 2001 y 2006 y también proveen información de diferentes características socioeconómicas tanto a nivel del hogar como del individuo, en el momento de la encuesta. Una vez cruzadas estas primeras dos fuentes de información, se logra identificar los grupos de tratamiento y de control.

La tercera fuente de información consiste en los datos administrativos del SIMAT. El cruce con esta tercera fuente de información se realiza para identificar el colegio al que el niño asiste. Los datos sólo incluyen colegios oficiales y se cuenta con información para los años entre 2005 y 2012<sup>11</sup>. De esta forma, teniendo identificados los beneficiarios y no beneficiarios del programa, así como los colegios en los que se encuentran matriculados, se realiza el cruce con la última fuente de información. Esta provee la información de las variables de resultado del estudio.

La cuarta fuente de información son los resultados de la prueba Saber 11 entre 2000 y 2012.<sup>12</sup> Esta es una prueba que presentan los estudiantes en grado once y constituye un requisito necesario para poder obtener el título de bachiller. Según el Ministerio de Educación Nacional, esta prueba proporciona información a la comunidad educativa sobre las competencias básicas que debe desarrollar un estudiante durante el paso por la vida escolar (MEN, 2015). El resultado del Saber 11 es utilizado por diferentes investigaciones para identificar la calidad de la educación que recibe un estudiante (Banco Mundial, 2008; Gaviria & Barrientos, 2001). Es así, que los resultados de esta prueba se utilizan agregados a nivel de institución para determinar la calidad del colegio, y a nivel individual para evaluar el desempeño del niño en el largo plazo. Además, para el periodo 2008-2012, los datos de la prueba Saber 11 incluyen preguntas sobre la posesión de activos fijos educativos en el hogar, específicamente el acceso a Internet y la tenencia de computador. La construcción de las variables dependientes se realiza a partir de esta base de información.

## **Variables de resultado**

En total se construyen siete variables de resultado. Las primeras cinco variables corresponden a los potenciales mecanismos. La primera es una variable binaria de calidad a nivel de institución que toma el valor de 1 si el resultado promedio en la prueba de los últimos

---

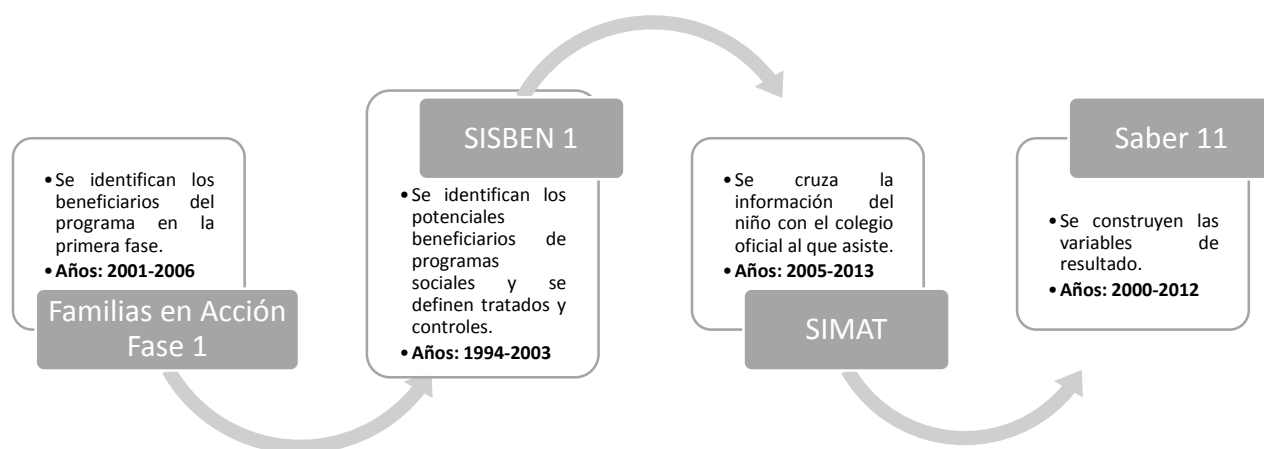
<sup>11</sup> Los datos del SIMAT hacen un seguimiento académico al niño. Este estudio toma únicamente la primera observación del niño en la base de datos y utiliza la información de ese periodo.

<sup>12</sup> Para información sobre esta prueba leer la sección Contexto Institucional.

cinco años está por encima de la media municipal del mismo periodo y 0 de lo contrario. La segunda variable es igualmente binaria; le asigna el valor de 1 si el promedio del colegio de los últimos 5 años se encuentra por encima del percentil 75 del municipio en este mismo periodo. La tercera variable consiste en la tenencia de activos escolares en el hogar; esta es una variable dicótoma que toma el valor de 1 si el niño posee en el hogar computador o si tiene acceso a internet. Esta sólo se puede construir para el periodo 2008-2012<sup>13</sup>. La cuarta variable corresponde a una binaria que toma el valor de 1 si la posición (ranking) del colegio mejora entre  $t=1$  y  $t=5$ . Por último, el quinto mecanismo hace referencia a la migración o cambio de colegio de los beneficiarios. Nuevamente es una variable binaria que toma el valor de 1 si el niño se cambia a un mejor colegio según la prueba Saber 11. Es importante resaltar que para esta variable es necesario observar tanto a tratados como controles en dos colegios diferentes. Es por este motivo que el número de observaciones disminuye significativamente respecto a las demás variables. Las últimas dos variables corresponden a las variables de desempeño académico en largo plazo: la graduación de bachillerato utilizando como proxy si el niño presenta o no la prueba Saber 11 y el puntaje estandarizado a nivel nacional obtenido en la prueba.

La Figura 1 resume la construcción de la base final. Aquí se muestra en qué orden se realizaron los cruces y la información que se espera obtener de cada fuente de información.<sup>14</sup>

*Figura 1. Descripción construcción de la base de datos final*



<sup>13</sup> La encuesta incluye preguntas relacionadas a la tenencia de internet y computador a partir del año 2008.

<sup>14</sup> Para el mecanismo de cambio a un mejor colegio la información se presenta como un panel con datos para dos momentos del tiempo. Los datos para las demás variables de resultado se presentan como un corte transversal repetido con información para cada individuo en un momento del tiempo. Para mayor información remitirse al apéndice A1.

La tabla 1 presenta estadísticas descriptivas de las principales variables de análisis, así como de algunas características socioeconómicas de la muestra final para los grupos de tratamiento y control.

Tabla 1: Estadísticas descriptivas en el ancho de banda

	Familias en Acción =1			Familias en Acción =0		
	Media	Desv. Est.	Obs	Media	Desv. Est.	Obs
<b>Demográficas (SISBEN)</b>						
Edad jefe de hogar	39.49	11.14	178,445	38.55	11.8	1,142,698
Proporción de padres jefe de hogar	0.8	0.4	178,445	0.72	0.45	1,142,707
Edad del niño	11.19	10	178,644	11.37	11.15	1,143,912
Proporción de niños hombres	0.46	0.49	178,665	0.47	0.49	1,144,172
Número de miembros del hogar	6.58	2.78	178,665	5.26	2.35	1,144,172
Proporción de menores de edad en el hogar	0.57	0.16	178,665	0.55	0.17	1,144,172
<b>Educación (SISBEN)</b>						
Años de educación jefe de hogar	7.91	2.37	177,894	8.41	2.41	1,138,867
<b>Características de la Vivienda (SISBEN)</b>						
Proporción de hogares con paredes	0.97	0.18	178,665	0.98	0.15	1,144,172
Proporción de hogares con acueducto	0.44	0.49	178,665	0.69	0.47	1,900,986
Proporción de hogares con pisos diferente a arena	0.53	0.5	178,665	0.78	0.41	1,144,172
Proporción de hogares con luz eléctrica	0.74	0.44	178,665	0.88	0.28	1,144,172
<b>Variables de resultado (Saber 11)</b>						
Colegio mejor que la media del municipio.	0.49	0.49	118,412	0.36	0.48	883,554
Colegio mejor que el percentil 75 del municipio.	0.013	0.11	118,412	0.021	0.15	883,554
Tenencia de activos en educación (Periodo 2008-1012)	0.2	0.39	67,942	0.33	0.47	592,188
Mejora en el ranking del colegio en su municipio	0.19	0.39	117,226	0.21	0.41	877,035
Cambio a un mejor colegio	0.5	0.5	21,420	0.52	0.5	199,262
Graduación (Nacidos entre 1986 y 1994)	0.5	0.5	108,977	0.56	0.5	593,297
Puntaje estandarizado Prueba Saber 11	-0.34	0.8	79,765	-0.23	0.84	499,782

Nota: La tabla presenta las estadísticas descriptivas para la muestra que contiene las 4 fuentes de información dividido entre beneficiarios de FeA y no. Las variables relacionadas a las características del hogar fueron creadas a partir de Oxford Poverty Index, que corresponde al índice multidimensional de pobreza. Género hace referencia a 1 si es hombre y 0 si es mujer.

Respecto a las características demográficas, tomadas de la encuesta del SISBEN, se encuentra que las medias entre grupos de tratamiento y control son bastante similares. Por ejemplo, en promedio, el jefe de hogar tiene 39.49 años para los beneficiarios y 38.55 años para los no beneficiarios de Familias en Acción. De igual forma, las edades de los niños difieren en magnitudes pequeñas entre ambos grupos, siendo los tratados en promedio 0.18 años menores que los controles. En relación a la proporción de padres jefe de hogar se encuentra que para el grupo de tratamiento corresponde a 0.8, mientras que para el grupo de control a 0.72. El número de niños hombres en la muestra final es 46% para beneficiarios y 47% para no beneficiarios. Además, la proporción de menores de edad en el hogar es de 0.57 para tratados y 0.55 para los controles del programa. El número de miembros promedio en el hogar para los

beneficiarios de Familias en Acción es de 6.58 miembros, mientras que para los no beneficiarios es de sólo 5,26. Por último, en términos de los años de educación del jefe del hogar, la media es de 7.91 para beneficiarios de Familias en Acción y 8.41 para no beneficiarios.

En relación con las características del hogar se encuentra que, en promedio, los no beneficiarios del programa tienen mejores viviendas. Por ejemplo, la proporción de hogares con acueducto corresponde a 0.44 para los tratados del programa, mientras que para sus respectivos controles corresponde a 0.69. Igualmente, la proporción de viviendas con pisos diferentes a arena es mayor para beneficiarios que para no beneficiarios, correspondiendo a 0.53 y 0.78 respectivamente. Por último, la proporción de hogares con luz eléctrica es de 0.74 para el grupo de tratamiento y 0.88 para el grupo de control, mientras que la proporción de hogares con paredes es del 97% para beneficiarios y 98% para no beneficiarios del programa.

En términos de las variables de resultado se presentan las estadísticas descriptivas tanto para los posibles mecanismos como las variables de desempeño académico en el largo plazo. Respecto a los mecanismos se encuentra que la proporción de niños que asiste a un colegio con resultado en la prueba Saber 11 por encima de la media municipal corresponde a 0.49 para los beneficiarios y 0.36 para los no beneficiarios del Familias en Acción. De igual forma, la proporción de estudiantes que asiste a un colegio con resultado por encima del percentil 75 del municipio es de 0.013 para tratados y de 0.021 para controles. En relación al ranking del colegio en el municipio, el 19% del grupo de tratamiento y el 21% del grupo de control asiste a colegios que mejora su ranking. Además, la proporción de niños que se cambia a un mejor colegio es de 0.5 para los beneficiarios y de 0.53 para los no beneficiarios. Por otro lado, respecto a las variables de desempeño de largo plazo, se encuentra que la proporción de graduados para beneficiarios es de 0.5 y 0.56 para no beneficiarios. Finalmente, en promedio, los tratados obtienen 0.43 desviaciones estándar por debajo de la media nacional en la prueba estandarizada, mientras que los controles solo 0.23 inferiores a la media.

## **5. Metodología**

La metodología busca solucionar el sesgo de selección de los hogares a través de un diseño de regresión discontinua. Se utiliza como umbral el puntaje SISBEN definido por el Gobierno para determinar la elegibilidad de una persona en el programa FeA. Esta

metodología consiste en explotar la variación cuasi experimental del umbral para estimar los efectos del programa sobre las variables de interés. Dado que existen familias que pueden ser elegibles, pero deciden no entrar al programa, se debe realizar una estimación de regresión discontinua borrosa. Para esto se instrumenta la participación en el programa con la elegibilidad en el mismo.<sup>15</sup> Este método ha sido utilizado para el caso colombiano en diferentes investigaciones (Báez y Camacho (2011); Báez *et al.* (2012); Melguizo, Sánchez y Velasco (2015)).

Luego, para determinar qué porcentaje del efecto del programa sobre la variable de graduación y puntaje en la prueba Saber 11 es explicado por la asistencia a colegios de calidad, se estima un modelo de mediación<sup>16</sup>.

### 5.1 Modelo por regresión discontinua borrosa

Con este método empírico se estima el efecto sobre las cinco variables de resultado mencionadas. La ecuación (1) presenta la especificación del modelo:

$$Y_i = \beta_0 + \tau D_i + f(Z_i) + X_i + u_i \quad (1)$$

Donde  $Y_i$  es alguna de las cinco variables de resultado,  $D_i$  es un indicador de tratamiento del programa o de la elegibilidad,  $f(Z_i)$  es una función de control que tiene en cuenta la distancia entre el puntaje del individuo y el umbral de elegibilidad, y  $X_i$  es un vector de características socioeconómicas. Entre ellas se incluyen: edad, género, urbano, años de educación del jefe del hogar, tamaño del hogar y proporción de menores de edad en el hogar. Por último,  $u_i$  es el error individual de la regresión. El parámetro de interés que mide el efecto del programa es  $\tau$ .

La regresión discontinua borrosa utiliza los criterios de elegibilidad del programa para instrumentar la participación en el mismo. Con esto, se estima la probabilidad de que la persona elegible sea efectivamente beneficiario. La ecuación a estimar es:

$$D_i = \beta_0 + \beta_1 ITT_i + \epsilon_i \quad (2)$$

---

<sup>15</sup> En la literatura de evaluación de impacto anglosajona, estos individuos se denominan “never takers”.

<sup>16</sup> Modelo para cuantificar el efecto de los mecanismos sobre las variables de desempeño académico.

Donde ambas variables son binarias siendo la dependiente el tratamiento y la independiente la elegibilidad. En otras palabras,  $D_i$  explica si el individuo  $i$  recibe efectivamente el programa e  $ITT_i$  hace referencia a si el individuo era elegible dado su puntaje SISBEN.<sup>17</sup> Luego, en la segunda etapa, se estima la ecuación (3) incluyendo esta probabilidad predicha en la primera etapa, que, como se explicó, corrige el sesgo por autoselección.

$$Y_i = \beta_0 + \tau \widehat{D}_i + f(Z_i) + X_i + u_i \quad (3)$$

Es importante mencionar que, a pesar de que la regresión discontinua es fuerte en términos de validez de las estimaciones, los resultados son efectos locales. Es decir, los resultados sólo son válidos para niños que tienen puntajes cercanos al punto de corte de elegibilidad y no se puede extrapolar el efecto sobre los estudiantes que están muy por encima o por debajo del mismo. Además, el presente estudio se basa en el ancho de banda óptimo propuesto por Imbens & Kalyanaraman (2012), que establece cuál es la distancia óptima al puntaje de corte del SISBEN<sup>18</sup> para que cada una de las estimaciones sean válidas.

## Validez del RD

Es necesario revisar las condiciones necesarias para que la metodología de regresión discontinua borrosa identifique estimadores insesgados y consistentes. De acuerdo con Bernal y Peña (2011), las condiciones necesarias para esta metodología son: (1) discontinuidad en el punto de corte, (2) no manipulación de las variables que determinan la participación y (3) continuidad de las variables observables y no observables alrededor del puntaje de corte.<sup>19</sup> A continuación, se evalúan estas condiciones para el diseño de RD de esta investigación:

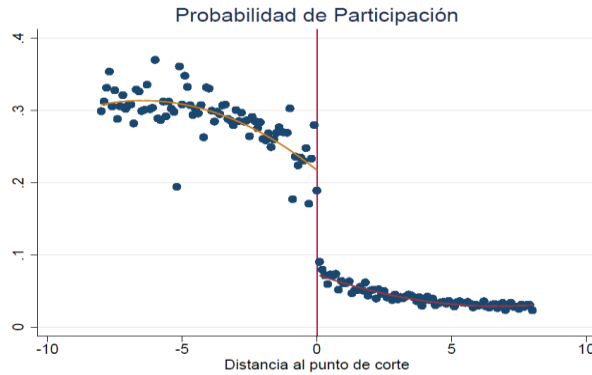
---

<sup>17</sup> Se realiza la primera etapa para cada una de las cinco variables de resultados, dado que esta puede variar por: 1) los diferentes anchos de banda óptimos y 2) el número de observaciones para cada estimación.

<sup>18</sup> La primera fase del programa los puntajes de corte establecidos fueron 36 para el área urbana y 18 para rural. Estos puntajes se alinean para ambos grupos con el fin de tener el punto de elegibilidad alineado en 0 y poder realizar inferencias sobre ambos grupos.

<sup>19</sup> Dado que las variables no observables no se pueden observar, la tercera condición solo se puede probar para las variables observables.

Figura 2. Probabilidad de Participación en el Familias en Acción<sup>20</sup>

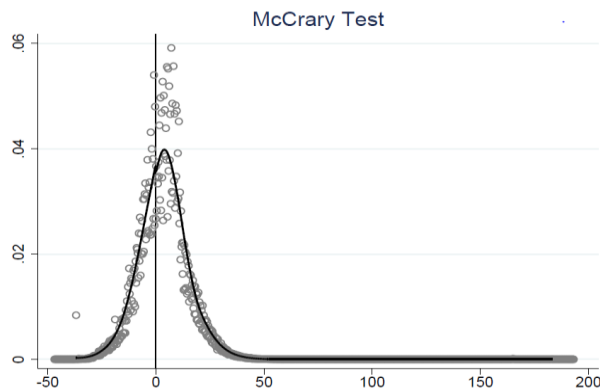


Nota: El gráfico muestra el resultado de la primera etapa del RD en el que se instrumenta la participación con la elegibilidad. Los datos utilizados son el cruce entre las cuatro fuentes de datos administrativos.

1) La probabilidad de entrar al programa es discontinua en el punto de corte. La Figura 2 muestra que la probabilidad de participación salta justo en el puntaje 0 por lo que se cumple este supuesto. En particular, en la figura se puede observar que la probabilidad de ser beneficiario del programa en el lado izquierdo es cercana a 27%, mientras que en el lado derecho la probabilidad es sólo del 9%.

2) No manipulación. En la Figura 3 se presenta el gráfico correspondiente al Test propuesto por McCrary en (2008) en el que se chequea posible manipulación. También se presenta el estadístico t-student donde la hipótesis nula es la no existencia de manipulación.

Figura 3. Test de manipulación en la variable de asignación



Nota: El gráfico muestra el Test de McCrary de potencial manipulación para la base final que contiene las cuatro fuentes de información cruzadas.

<sup>20</sup> Esta discontinuidad corresponde a la primera etapa del diseño de RD borrosa. En este aspecto se hará más énfasis en la sección de resultados, donde se presentará la magnitud del salto para cada una de las variables de resultado.

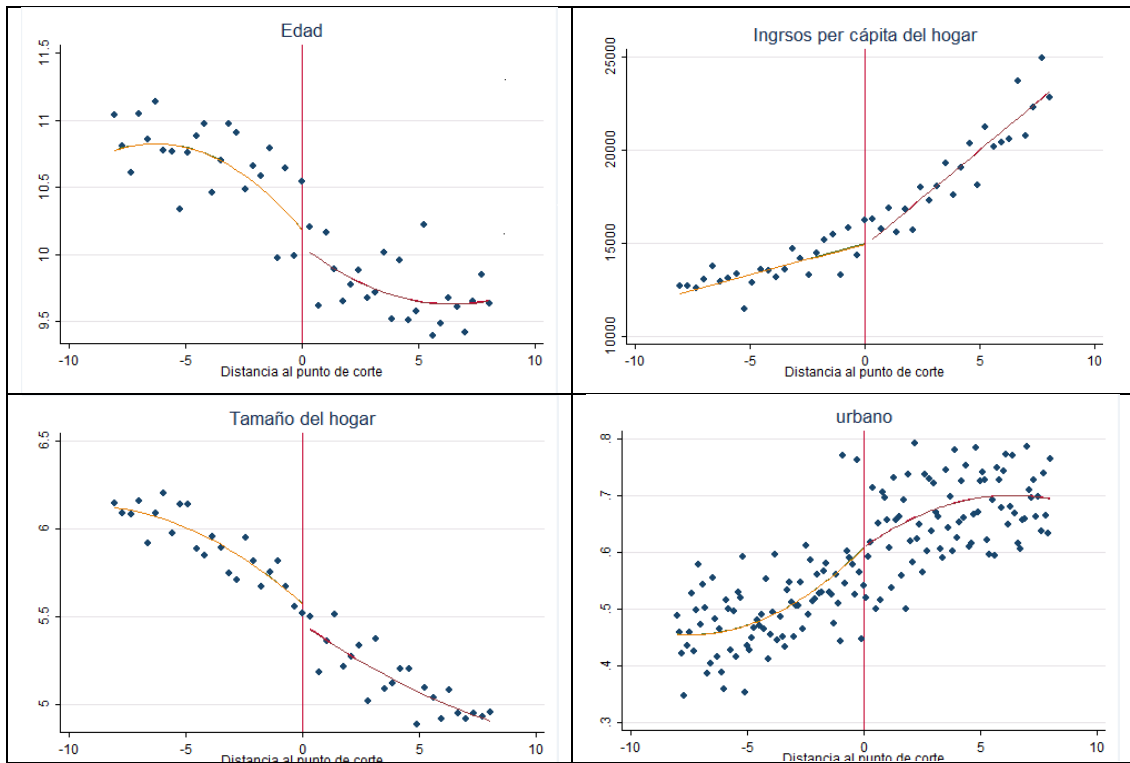


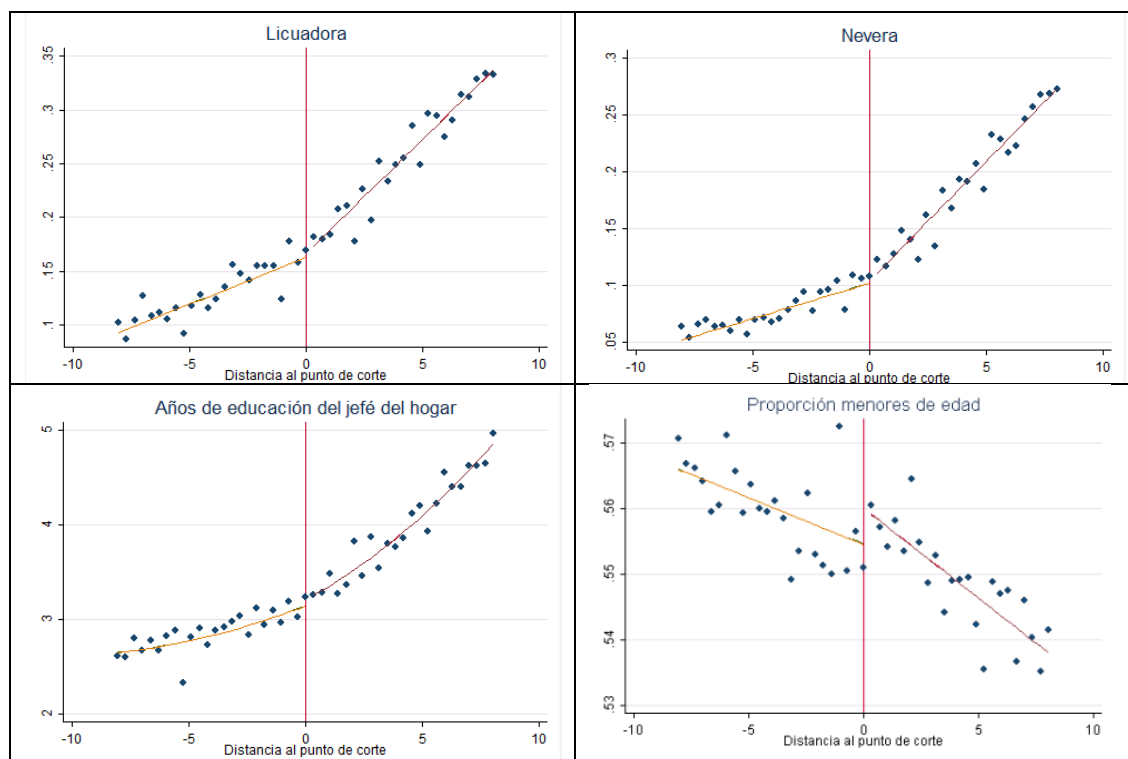
$$t = \frac{-0.044}{0.0026} = -17.2$$

El estadístico t muestra que existe cierta manipulación en el punto de corte. Sin embargo, la literatura reciente ha utilizado una estrategia empírica complementaria para lidiar con la manipulación. Este procedimiento se conoce como “donut-hole” de regresión discontinua borrosa (Barreca *et al.* (2011); Grau *et al.* (2016)). El método determina que, si al eliminar la manipulación, el efecto del RD sigue siendo robusto, se concluye estadísticamente que la manipulación no afectó los estimadores. En la sección 8 se realizan algunas pruebas de robustez en las que se incluye este procedimiento.

3) Las variables observables no presentan saltos en el punto de corte. En la figura 4 se puede observar los gráficos correspondientes a las siguientes covariables: edad, ingresos per cápita tamaño del hogar, zona urbana, licuadora, nevera, años de educación del jefe del hogar y proporción de menores de edad en el hogar. La Tabla 2 muestra la magnitud del salto, su significancia y el grado del polinomio incluido en la regresión.

Figura 4. Validez del RD: Gráficos de las covariables.





Nota: En la figura se muestran las gráficas correspondientes a algunas de las variables con las que se estima el puntaje del SISBEN para verificar la validez del RD. Las variables son: edad, ingresos per cápita tamaño del hogar, zona urbana, licuadora, nevera, años de educación del jefe del hogar y proporción de menores de edad en el hogar.

Tabla 2. Validez del RD: Resultados de las covariables .

	Edad	Genero	Urbano	Nevera	Televisión	Lavadora
ITT	-0.039 (0.074)	0.002 (0.003)	-0.03 (0.003)	-0.001 (0.001)	-0.002 (0.003)	0.0003 (0.000)
Forma funcional	Cuadrática	Cuadrática	Cuadrática	Lineal	Cuadrática	Lineal

	Licuadora	Años de educ del jefe	Propor. menores de edad	Ingreso per cápita	Tamaño del hogar
ITT	0.002 (0.002)	-0.034** (0.015)	-0.002** (0.001)	105.2 (102.6)	0.060*** -0.015
Forma funcional	Lineal	Cuadrático	Lineal	Lineal	Cuadrático

Nota: La tabla presenta los resultados correspondientes al salto en el puntaje de corte para algunas de las variables con las que se estima el puntaje del SISBEN. Errores estándar en paréntesis. \*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

De esta forma, es posible hacer una verificación tanto visual como estadística de la continuidad de las variables en el umbral de elegibilidad. En la Figura 4 se observa la continuidad de las variables en el punto de corte, sin embargo, en la tabla 2 se encuentra que a pesar de que los saltos son muy pequeños en magnitud para las variables de años de educación del jefe, proporción de menores de edad y tamaño del hogar, su efecto es significativo. Es por este motivo que se controla por estas variables en las estimaciones principales.

## 5.2 Modelo de mediación para cuantificar el efecto de los mecanismos sobre las variables de desempeño académico

Con este modelo se busca determinar qué fracción del efecto sobre las variables de largo plazo es explicado a través de mecanismo de asistencia a colegios de calidad. Esta metodología es la implementada por Baron and Kenny (1986) y Preacher and Hayes (2008). Para ello, se estima conjuntamente las siguientes dos ecuaciones:

$$Y_i = \beta_0 + \alpha M_i + \delta D_i + f(Z_i) + X_i + e_i \quad (4)$$

$$M_i = \gamma_0 + \theta D_i + f(Z_i) + X_i + \varepsilon_i \quad (5)$$

Donde  $Y_i$  es una de las variables de resultado de largo plazo, y  $M_i$  corresponde a el mecanismo que toma el valor de 1 si el niño asiste a un colegio con resultados en la prueba por encima de la media del municipio. La variable  $D_i$  es un indicador de participación o elegibilidad en el programa FeA y la variable  $f(Z_i)$  sigue siendo la función de control del modelo anterior.  $X_i$  es un vector de características socioeconómicas en el que se incluyen las mismas variables del modelo anterior: edad, género, urbano, años de educación del jefe del hogar, tamaño del hogar y proporción de menores de edad en el hogar. Por último,  $e_i$  y  $\varepsilon_i$  son los errores de las dos ecuaciones, dado que se estiman simultáneamente, el modelo permite que exista correlación entre estos dos errores. En otras palabras, se estiman dos regresiones discontinuas borrosas de manera simultánea con el fin de descomponer el efecto de FeA sobre las variables de largo plazo en dos partes. La primera parte del efecto total es el efecto directo por del programa; la segunda parte del efecto corresponde indirecto generado por el mecanismo de calidad educativa propuesto en el presente trabajo.

Para cada variable de desempeño de largo plazo se realizan dos estimaciones del modelo de mediación. En la primera, el indicador  $D_i$  corresponde a la participación en el programa, mientras que en la segunda  $D_i$  representa la elegibilidad en el programa. Dado que la participación no cumple el supuesto de exogeneidad, en una primera etapa, se instrumenta con la elegibilidad. De esta forma, la variable incluida es el valor predicho de la participación, que se obtiene de esta primera estimación.

Reemplazando la ecuación (5) en la (4) se obtiene:

$$Y_i = \beta_0 + \alpha (\gamma_0 + \theta D_i + f(Z_i) + X_i + \varepsilon_i) + \delta D_i + f(Z_i) + X_i + e_i \quad (6)$$

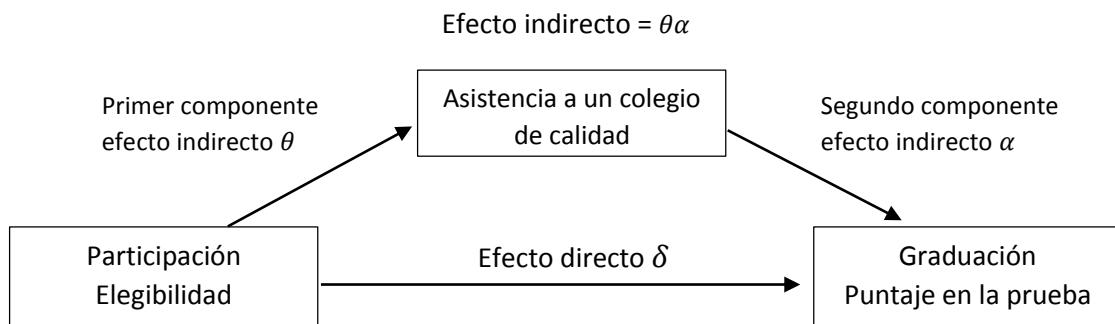
$$Y_i = \beta_0 + \alpha \gamma_0 + \theta \alpha D_i + \delta D_i + (1 + \alpha)f(Z_i) + (1 + \alpha)X_i + e_i + \alpha \varepsilon_i \quad (7)$$

Los coeficientes de interés son  $\alpha$ ,  $\delta$  y  $\theta$ . El parámetro  $\delta$  captura el efecto directo del programa, mientras que la interacción  $\alpha\theta$  estima el efecto indirecto a través del mecanismo de calidad educativa. De esta manera, con base en los modelos de mediación, el porcentaje que explicado a través del mecanismo de calidad educativa:

$$\% \text{ Explicado} = \frac{\alpha\theta}{\alpha\theta + \delta} \quad (6)$$

La figura 5 muestra una representación gráfica de los modelos de mediación. Se muestra la forma de obtener el efecto directo y cada uno de los componentes del efecto indirecto.

*Figura 5. Modelo de mediación para variables de desempeño de largo plazo*



Por último, se realiza un proceso de bootstrapping para determinar la significancia estadística de los estimadores que miden el efecto directo y el indirecto.

## 6. Resultados

En esta sección se presentan los resultados de las principales estimaciones para dos muestras de municipios diferentes: 1) todos los municipios de Colombia; 2) los municipios focalizados del programa FeA entre los años 2001-2006, que corresponden a los años de la primera fase.

En la primera parte se presentan los resultados para los diferentes mecanismos estudiados tanto de asistencia a colegios de calidad como de posesión de activos fijos<sup>21</sup>; la segunda sección describe los resultados de las variables de desempeño académico de largo plazo. En la tercera parte se presentan los resultados para el modelo en el cual se cuantifican los efectos de los mecanismos sobre las variables de desempeño académico<sup>22</sup>.

## Mecanismos

La tabla 3 presenta los resultados correspondientes a los primeros tres potenciales mecanismos que pueden explicar el efecto de FeA sobre variables educativas. Las columnas 1 y 2 centran su análisis sobre la variable de asistencia escolar a colegios con resultados en la prueba Saber 11 por encima de la media municipal. Para la primera etapa se encuentra que la elegibilidad cumple con las dos condiciones de un instrumento para explicar la participación en el programa, además se obtiene un coeficiente asociado entre 15.7 y 17.5 puntos porcentuales. A su vez, los resultados sobre asistencia escolar a instituciones de calidad sugieren que la participación en FeA aumenta la probabilidad de asistir a un colegio de mayor calidad.

El efecto encontrado para este primer mecanismo de calidad educativa se encuentra en un rango entre 13.8 y 15.4 puntos porcentuales. Este efecto corresponde al LATE (Angrisy & Imbens, 1994), es decir, al efecto promedio del tratamiento sobre aquellas personas que cambiaron su participación dado que eran elegibles en el programa<sup>23</sup>. Esta población está compuesta por aquellas personas que son beneficiarias de familias en acción y elegibles y que si no fueran elegibles no participarían en el mismo. Este efecto corresponde a 0.25 desviaciones estándar respecto al grupo de control, por lo que se puede inferir que FeA genera un impacto significativo en la demanda por calidad de colegios. De igual forma, la Figura 6 permite ver de forma gráfica el efecto del programa sobre la asistencia a colegios por encima de la media municipal. Sin embargo, cuando se analiza el efecto sobre colegios con resultados en la prueba Saber 11 por encima del percentil 75 municipal (columnas 3 y 4 de la tabla 4) se evidencia que estos colegios no hacen parte de la cola derecha de la distribución. Esto porque el efecto LATE para este segundo mecanismo de calidad educativa es no significativo y cercano a 0. Es

---

<sup>21</sup> Los anchos de bandas óptimos corresponden a 8.02 y 11.00 para asistir a un colegio por encima de la media del municipio o del percentil 75 del municipio respectivamente, 7.34 para puntaje estandarizado, 4.67 para graduación y 10.72 para la tenencia de activos en educación.

<sup>22</sup> Para la estimación se utilizan sólo formas lineales o cuadráticas siguiendo a Gelman y Imbens (2014).

<sup>23</sup> En la literatura de evaluación de impacto anglosajona, estos individuos se denominan “compliers”.

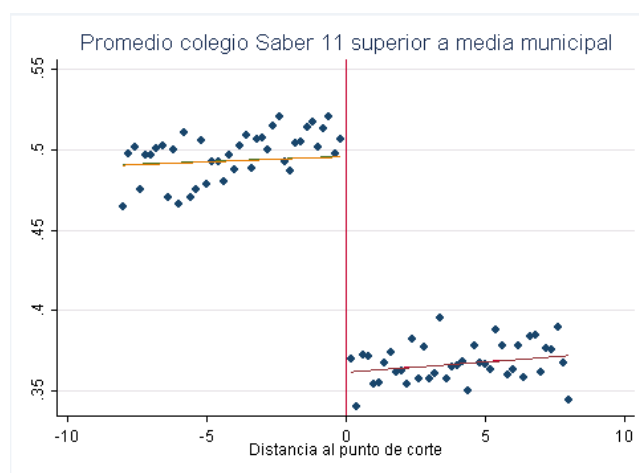
importante señalar que el error estándar asociado al coeficiente es muy pequeño por lo que los resultados son muy precisos.

Finalmente, las columnas 5 y 6 de la tabla 4 presentan los resultados para el último mecanismo que se analiza: la posesión de activos fijos en computador y acceso a Internet en el hogar. Los resultados sugieren que el impacto del programa es estadísticamente igual a 0, ya que se encuentran efectos positivos más no significativos.

Tabla 3: Mecanismos						
	Asistencia a colegios por encima de la media municipal		Asistencia a colegios por encima del per75 municipal		Activos fijos en educación	
<i>LATE</i>	(1)	(2)	(3)	(4)	(5)	(6)
Beneficiario FeA	0.154*** (0.012)	0.138*** (0.011)	-0.003 (0.003)	-0.002 (0.003)	0.003 (0.013)	0.013 (0.012)
<i>Primera etapa</i>						
Elegibilidad	0.157*** (0.001)	0.175*** (0.002)	0.167*** (0.001)	0.184*** (0.001)	0.209*** (0.002)	0.234*** (0.002)
F -Test instrumento	13,278	13,876	9,647	10,483	8,100	8,552
Observaciones	1,010,735	971,764	1,337,838	1,213,048	420,542	414,404
Grado del polinomio	Lineal	Lineal	Lineal	Lineal	Lineal	Lineal
Controles	Si	Si	Si	Si	Si	Si
Muestra	Todos	2001-2006	Todos	2001-2006	Todos	2001-2006

Nota: La tabla presenta los resultados correspondientes a la primera y segunda etapa de los mecanismos tanto para todos los municipios del SISBEN, como los focalizados por el programa entre 2001 y 2006. Errores estándar en paréntesis. \*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

Figura 6. Resultado gráfico para la asistencia a un colegio con mejor resultado que la media municipal<sup>24</sup>



Nota: En la figura se muestran las gráficas correspondientes al efecto LATE de la variable de asistencia a un colegio con mejor resultado que la media municipal.

<sup>24</sup> Los demás resultados gráficos se presentan en el apéndice A3

La tabla 4 presenta en las columnas (1) y (2) el mecanismo de asistencia a un colegio que mejoran en el tiempo su ranking dentro del municipio. Se encuentra un coeficiente asociado a la primera etapa de entre 15 y 16 p.p., es decir, ser elegible aumenta la probabilidad de participar en dicha magnitud. Adicionalmente, los beneficiarios de FeA aumentan la probabilidad de asistir a uno de estos colegios en entre 6.77 p.p. y 7.85 p.p., frente a los no beneficiarios. Respecto a la migración de colegios se encuentra que los tratados del programa en municipios focalizados aumentan la probabilidad de cambiarse a un mejor colegio en 3.1 p.p. respecto a sus controles.

Tabla 4: Mecanismos				
	Asistencia colegios que mejora en el ranking del colegio en el municipio		Cambian a un mejor colegio	
<i>LATE</i>	(1)	(2)	(3)	(4)
Beneficiario FeA	0.0785*** (0.012)	0.0677*** (0.011)	0.026 (0.019)	0.031* (0.018)
<i>Primera etapa</i>				
Elegibilidad	0.150*** (0.001)	0.166*** (0.002)	0.162*** (0.002)	0.176*** (0.002)
F -Test				
instrumento	10,000	10,500	6,561	6,828
Observaciones	812,406	1,113,851	386,478	352,216
Grado del polinomio	Lineal	Lineal	Lineal	Lineal
Controles	Si	Si	Si	Si
Muestra	Todos	2001-2006	Todos	2001-2006

Nota: La tabla presenta los resultados correspondientes a la primera y segunda etapa de los dos últimos mecanismos estudiados, tanto para todos los municipios del SISBEN, como los focalizados por el programa entre 2001 y 2006. Errores estándar en paréntesis. \* p<0.1, \*\* p<0.05, \*\*\*p<0.01.

### **Variables de desempeño académico en largo plazo**

La tabla 5 presenta los resultados para las variables de desempeño de largo plazo. Las columnas 1 y 2 muestran las estimaciones para graduación, mientras que las columnas 3 y 4 se enfocan en el puntaje obtenido en la prueba Saber 11.

Para la variable graduación se puede observar que, para la primera etapa, se presenta una discontinuidad en el puntaje de corte, en particular, ser elegible aumenta entre 17.3 y 19.2 puntos porcentuales la probabilidad de participar en el programa alrededor del umbral de elegibilidad. Dado que este umbral es aleatorio, el instrumento cumple con las dos condiciones de un buen instrumento, por una parte, es relevante, ya que explica la participación en FeA, y por otra, es exógeno en el umbral dado el supuesto principal de esta metodología. El estadístico F asociado a la primera etapa es mayor a 9.

Respecto al efecto LATE, se encuentra que la participación de FeA aumenta entre 4.9 y 5.1 puntos porcentuales la probabilidad de presentar la prueba Saber 11, la cual es una proxy para la variable graduación. Se puede concluir que el programa genera efectos positivos en el largo plazo, ya que se encuentran un efecto positivo y significativo sobre graduación. Es importante resaltar que a diferencia de Báez y Camacho (2011) estas estimaciones sólo se realizan para estudiantes que asisten a colegios oficiales.

Por otra parte, las estimaciones para el puntaje estandarizado en la prueba Saber 11 se presentan en las columnas 3 y 4 de la tabla 5. Los resultados sugieren que la elegibilidad es un buen instrumento de la participación. Se encuentra un efecto entre 17.9 y 19.0 puntos porcentuales en la primera etapa, por lo que ser elegible aumenta la participación en FeA de manera positiva y significativa. De igual manera se encuentran efectos positivos sobre la prueba estandarizada. En particular, para los niños beneficiarios de FeA se obtiene un efecto LATE sobre el puntaje de la prueba entre 0.042 y 0.074 desviaciones estándar en el punto de corte.

Tabla 5: Variables de desempeño en el largo plazo				
	Graduación		Puntaje Saber 11	
<i>LATE</i>	(1)	(2)	(3)	(4)
Beneficiario FeA	0.051*** (0.018)	0.049 *** (0.016)	0.042* (0.024)	0.074* (0.039)
<i>Primera etapa</i>				
Elegibilidad	0.173*** (0.001)	0.192*** (0.002)	0.190*** (0.002)	0.179*** (0.002)
F -Test instrumento	7,357	7,482	8,922	9,452
Observaciones	536,105	519,301	536,105	519,301
Grado del polinomio	Lineal	Lineal	Lineal	Cuadrático
Controles	Si	Si	Si	Si
Muestra	Todos	2001-2006	Todos	2001-2006

Nota: La tabla presenta los resultados correspondientes a la primera y segunda etapa de las variables de desempeño en el largo plazo, tanto para todos los municipios del SISBEN, como los focalizados por el programa entre 2001 y 2006. Errores estándar en paréntesis. \* p<0.1, \*\* p<0.05, \*\*\*p<0.01.

La siguiente sección analiza los potenciales mecanismos que pueden explicar los impactos positivos de FeA sobre variables educativas.



## Modelo de mediación para cuantificar el efecto de los mecanismos sobre las variables de desempeño académico de largo plazo

En esta sección se busca descomponer el efecto de las variables de desempeño académico en el porcentaje del efecto total que es explicado por el mecanismo de asistencia a un colegio con resultados en la prueba Saber 11 por encima de la media del municipio, a través de dos modelos de mediación. Se escoge este mecanismo, ya que es el que muestra efectos mayores del programa Familias en Acción.

Como se explicó en la sección de metodología, el valor agregado de esta investigación es evaluar algunos de los potenciales mecanismos que explican el efecto de FeA sobre variables educativas en el largo plazo. En este sentido, a través de modelos de mediación se puede descomponer el efecto en dos partes: 1) el efecto directo que es explicado por la participación o la elegibilidad en el programa a y 2) el efecto indirecto que es explicado a través del mecanismo. Los modelos de mediación se estiman tanto para la participación<sup>25</sup> en el programa como para la elegibilidad en el mismo, los errores estándar se construyen con bootstrapping.

Tabla 6. Modelos de mediación				
Mecanismo de calidad sobre graduación				
	Participación		Elegibilidad	
	Coef.	Err. Estándar	Coef.	Err. Estándar
Efecto directo	0.053	0.019	0.018	0.002
Efecto indirecto	0.014	0.002	0.002	0.000
Proporción	0.209	-	0.121	-
Mecanismo de calidad sobre prueba saber 11				
	Participación		Elegibilidad	
	Coef.	Err. Estándar	Coef.	Err. Estándar
Efecto directo	0.084	0.032	0.012	0.006
Efecto indirecto	0.022	0.003	0.004	0.001
Proporción	0.207	-	0.234	-

Nota: La tabla presenta los resultados para el modelo de mediación. La columna 1 presenta los estimadores del efecto directo de FeA, y el efecto indirecto a través de un aumento por colegios de calidad. Por otra parte, la columna 3 presenta los estimadores para el efecto directo de elegibilidad y el efecto indirecto de un aumento en la calidad de colegios. Los errores estándar se presentan en las columnas 2 y 4.

La tabla 6 presenta los resultados del modelo de mediación y los errores estándar del proceso de bootstrapping con 50 repeticiones. Para la variable graduación, el efecto directo de participar en FeA corresponde a 5.3 puntos porcentuales, mientras que el efecto indirecto que

<sup>25</sup> Para cumplir con el supuesto de exogeneidad, la participación se instrumentó con la elegibilidad en una primera etapa.

se explica a través de la asistencia a colegios de mejor calidad es de 1.4 puntos porcentuales. Es decir, este mecanismo explica cerca de un 20% del efecto total. Cuando se estima el modelo tomando como variable de interés la elegibilidad, el efecto directo corresponde a 1.8 puntos porcentuales, mientras que el indirecto es de 0.25 puntos porcentuales, es decir, el mecanismo explica un 12% del efecto total.

Para el efecto sobre puntaje estandarizado de la prueba Saber 11,<sup>26</sup> se encuentra que el efecto directo de participar en el programa corresponde a 0.084 desviaciones estándar, mientras que el efecto indirecto, explicado a través del mecanismo, es de 0.022 desviaciones estándar. Es decir, el principal mecanismo que se quiere evaluar explica cerca de un 20% el efecto total del programa. De igual forma, en el caso en el que se analiza la elegibilidad, el efecto directo corresponde a 0.012 desviaciones estándar, mientras que el indirecto corresponde a 0.004 desviaciones estándar. Entonces, el mecanismo de demanda por colegios de calidad explica en 23% el efecto sobre la prueba Saber 11 de la población elegible en el programa.

Se puede concluir que la asistencia a colegios a colegios de calidad, tomando como brecha la media municipal, explica entre un 12% y 20% los efectos sobre graduación y entre un 20% y 23% los efectos sobre el puntaje en la prueba Saber 11.

## **8. Posible sesgo de selección y pruebas de robustez**

En esta sección se estudia en detalle el posible sesgo de selección en la muestra sobre la que se estiman los efectos del programa para los resultados de la prueba Saber 11 y la tenencia de activos fijos en educación. Adicionalmente, se presentan algunas pruebas de robustez donde se corrige la manipulación de puntaje SISBEN.

### **Posible sesgo de selección positiva y negativa**

Es importante resaltar que para los resultados de la prueba Saber 11 y la tenencia de activos en educación, los efectos causales se estiman en una muestra seleccionada, ya que muchos de los estudiantes que presentan la prueba, lo hacen a raíz de la intervención del

---

<sup>26</sup> Se estima sobre el mejor escenario, es decir, en el que se elimina la proporción de elegibles de la cola izquierda de la distribución del puntaje estandarizado en la prueba Saber 11. Estos escenarios se desarrollan en el capítulo 8.

programa.<sup>27</sup> En otras palabras, una fracción de los estudiantes tratados que presentan la prueba Saber 11, no la hubieran presentado en ausencia del programa. Para resolver este problema de selección no aleatoria se sigue el método desarrollado por Lee (2009), el cual consiste en acotar el efecto del programa entre dos valores. El primer valor corresponde al límite inferior del efecto, y se obtiene eliminando una proporción de la distribución superior de la variable de resultado de los individuos elegibles. El segundo valor corresponde al límite superior, donde se elimina la proporción de la distribución inferior.<sup>28</sup> El porcentaje de la muestra que es eliminado corresponde al efecto de graduación, ya que este es el porcentaje que no hubiera presentado la prueba en ausencia del programa. El límite inferior corresponde al escenario más negativo para el programa, mientras que el límite superior corresponde al escenario más optimista. De esta manera, se acota el efecto del programa sobre estas variables entre estos dos límites.

Tabla 7: Lee Bounds				
Panel A: Prueba Saber 11				
	Límite inferior		Límite superior	
Beneficiario FeA	-0.075*** (0.025)	-0.0514 (0.039)	0.118*** (0.024)	0.148*** (0.039)
Observaciones	532,203	514,634	532,275	514,697
Grado del polinomio	Lineal	Cuadrático	Lineal	Cuadrático
Controles	Si	Si	Si	Si
Muestra	Todos	2001-2006	Todos	2001-2006
Panel B: Activos fijos en educación				
	Límite inferior		Límite superior	
Beneficiario FeA	-0.008 (0.015)	-0.005 (0.014)	0.005 (0.018)	0.003 (0.017)
Observaciones	292,195	281,310	291,381	282,058
Grado del polinomio	Lineal	Lineal	Lineal	Lineal
Controles	Si	Si	Si	Si
Muestra	Todos	2001-2006	Todos	2001-2006

Nota: La tabla presenta los resultados correspondientes a la segunda etapa del resultado en la prueba Saber 11 y el mecanismo de activos fijos, corrigiendo por sesgo de selección positivo y negativo. Se realizan estimaciones incluyendo todos los municipios del SISBEN y los focalizados por el programa entre 2001 y 2006. Errores estándar en paréntesis. \* p<0.1, \*\* p<0.05, \*\*\*p<0.01.

La tabla 7 presentan las estimaciones de los Lee-Bounds, en el panel A, la variable de resultado corresponde al puntaje estandarizado en la prueba Saber 11 y el panel B analiza la posesión de activos fijos en educación. Los resultados para prueba Saber 11 muestran que en el peor de los escenarios el efecto del programa decrece el puntaje, mientras que en el mejor lo aumenta entre 0.118 y 0.148 desviaciones estándar. Se podría inferir que FeA si genera un

<sup>27</sup> Para el caso de los activos en educación, la selección se genera debido a que la variable se observa condicional a que el niño presente el examen estandarizado, es decir, esta variable se construye a partir de los datos administrativos de la Prueba Saber 11.

<sup>28</sup> En el apéndice se explica detalladamente en que consiste el método de manera teórica.

efecto positivo sobre los resultados en la prueba Saber 11, ya que los niños de menor habilidad (aquellos que en ausencia del programa no presentarían la prueba) corresponde a la cola izquierda de la distribución, por lo que el efecto correspondería a la cota superior. Para el mecanismo de activos fijos en educación no se encuentran efectos significativos. Se puede observar que en el escenario más optimista el efecto esta entre 0.03 y 0.5 p.p., pero no es estadísticamente diferente de cero.

### Corrección manipulación

Como se mencionó en la sección de validez del RD, el puntaje SISBEN presenta manipulación en el punto de corte. Es por este motivo que es necesario realizar las estimaciones sin incluir los puntajes del SISBEN que presentan manipulación y corroborar que los resultados son consistentes a los encontrados en la sección anterior. Para ello, se realiza la prueba “*donut-hole*” (Barreca *et al.* (2011); Grau *et al.* (2016)) en la que se elimina los puntajes del SISBEN que presentan dicha manipulación, es decir los puntajes entre -0.3 y 0.3. Las tablas 8 y 9 presentan las estimaciones para cinco variables de resultado cuando se lleva a cabo este procedimiento.

Tabla 8: Robustez variables de desempeño en el largo plazo				
	Graduación		Puntaje Saber 11	
<i>LATE</i>	(1)	(2)	(3)	(4)
Beneficiario FeA	0.057*** (0.018)	0.043*** (0.016)	0.023 (0.024)	0.0773* (0.042)
<i>Primera etapa</i>				
Elegibilidad	0.182*** (0.002)	0.213*** (0.002)	0.203*** (0.002)	0.166*** (0.003)
F -Test instrumento	8,281	8,576	8,463	9,040
Observaciones	417,314	400,748	550,141	541,636
Grado del polinomio	Lineal	Lineal	Lineal	Cuadrático
Controles	Si	Si	Si	Si
Muestra	Todos	2001-2006	Todos	2001-2006

Nota: La tabla presenta las estimaciones “*donut-hole*” para las variables de desempeño en el largo plazo, tanto para todos los municipios del SISBEN como los focalizados por el programa entre 2001 y 2006. Errores estándar en paréntesis. \*p<0.1, \*\*p<0.05, \*\*\*p<0.01.

En la tabla 8 se puede observar que el efecto LATE sobre graduación es de magnitud similar cuando se eliminan estos puntajes. En particular, el efecto de FeA sobre la variable graduación se estima entre 5.4 y 6.12 puntos porcentuales. Similarmente, para los puntajes en la prueba Saber 11 los resultados siguen siendo positivos. Sin embargo, para el caso en el que se incluyen todos los municipios el efecto es de sólo 0.023 desviaciones estándar y no

significativo, aunque para la muestra de municipios focalizados entre 2001 y 2006 por el programa se encuentra un efecto LATE positivo y significativo de 0.077 desviaciones estándar.

Por otra parte, los mecanismos se analizan en la tabla 9. Se puede observar, que para el caso de la variable dicótoma de asistencia a colegios con un resultado promedio en la prueba Saber 11 mayor a la media municipal, el efecto LATE sigue siendo positivo y de similar magnitud. El efecto varía entre 14.6 y 14.8 puntos porcentuales. Por otra parte, para el caso en el que el criterio es la asistencia a colegios con resultados por encima del percentil 75 municipal se encuentran efectos LATE estadísticamente igual a 0 para las familias beneficiarias. Por último, los resultados siguen siendo robustos para la posesión de activos fijos, acceso a Internet y computador, ya que cuando se eliminan los puntajes de manipulación, el efecto LATE para los beneficiarios de FeA es estadísticamente igual a 0.

Tabla 9: Robustez Mecanismos						
	Asistencia a colegios por encima de la media municipal		Asistencia a colegios por encima del per75 municipal		Activos fijos en educación	
<i>LATE</i>	(1)	(2)	(3)	(4)	(5)	(6)
Beneficiario FeA	0.148*** (0.012)	0.146*** (0.012)	0.000 (0.004)	0.002 (0.004)	-0.002 (0.014)	0.002 (0.013)
<i>Primera etapa</i>						
Elegibilidad	0.168*** (0.001)	0.162*** (0.001)	0.172*** (0.001)	0.189*** (0.001)	0.221*** (0.002)	0.215*** (0.002)
F -Test instrumento	12,612	13,216	13,870	14,614	7,313	7,768
Observaciones	1,023,483	1,023,951	1,195,529	1,067,606	405,884	404,422
Grado del polinomio	Lineal	Lineal	Lineal	Lineal	Lineal	Lineal
Controles	Si	Si	Si	Si	Si	Si
Muestra	Todos	2001-2006	Todos	2001-2006	Todos	2001-2006

Nota: La tabla presenta las estimaciones “donut-hole” para los mecanismos tanto para todos los municipios del SISBEN, como los focalizados por el programa entre 2001 y 2006. Errores estándar en paréntesis. \*p<0.1, \*\*p<0.05, \*\*\*p<0.01.

Se puede concluir que los resultados tanto de las variables educativas como de potenciales mecanismos son robustos a eliminar los puntajes que presentan manipulación

## 8. Conclusiones

El presente trabajo busca estudiar diferentes mecanismos que expliquen los efectos positivos de los Programas de Transferencias Monetarias Condicionadas sobre variables educativas en el largo plazo estudiadas por la literatura. Específicamente, la probabilidad de graduarse de educación básica y el puntaje de la prueba Saber 11. En particular, se estudia si existe un impacto del programa FeA sobre la demanda por colegios de calidad. En otras

palabras, si los niños beneficiarios del programa asisten a mejores colegios, a raíz del ingreso extra recibido por la participación en el programa.

Para ello, se explota la variación cuasi experimental generada por los criterios de elegibilidad de Familias en Acción a través de una metodología de Regresión Discontinua en dos etapas. En la primera etapa se estima la probabilidad de tratamiento con base en los criterios de elegibilidad, mientras que en la segunda se estima el efecto LATE sobre la población que cambia su decisión de participación en el programa por la elegibilidad. Esta metodología permite comparar niños similares tanto en características observables como en no observables alrededor del puntaje de corte establecido por el gobierno. De igual manera, se estudiaron otros posibles mecanismos como la posesión de activos fijos educativos en el hogar como computador y acceso a Internet.

El estudio construye una medida de calidad de los colegios basada en los resultados de la prueba estandarizada Saber 11. Se define a un colegio de alta calidad como aquel que obtiene un resultado en la prueba Saber 11 por encima de la media del municipio, durante los últimos 5 años. Además, se construye otra variable similar, donde se establece si el promedio del colegio en la prueba está por encima o por debajo del percentil 75 del municipio. También se construye una variable que captura si el niño se cambia a un colegio que mejora el resultado en la prueba Saber 11 en el tiempo y otra que estudia si los estudiantes se cambian a colegios de mejor calidad.

Los resultados sugieren que los niños beneficiarios del programa aumentan su probabilidad de asistir a un mejor colegio entre 13 p.p. y 15 p.p. respecto a sus controles. Este efecto corresponde a 0.25 desviaciones estándar de dicha variable. Estos resultados son robustos a diferentes especificaciones: 1) estimaciones incluyendo todos los municipios del SISBEN, 2) estimaciones incluyendo únicamente los municipios focalizados por la fase 1 del programa y 3) estimaciones corrigiendo la manipulación del puntaje SISBEN.

De igual forma, se analizó el segundo mecanismo de calidad educativa que hace referencia a la asistencia de los niños a colegios con resultados promedio en la prueba Saber 11 por encima del percentil 75. Se encuentra que no es significativa, lo que sugiere que los niños beneficiarios si están asistiendo a mejores colegios alrededor del punto de corte, pero no están inscritos en los mejores colegios de sus municipios. En otras palabras, asisten a mejores

colegios, pero no a los colegios en la cola derecha de la distribución de la prueba Saber 11. Adicionalmente, para el caso de activos fijos educativos en el hogar no se encuentran efectos LATE significativos.

Adicionalmente, se logró establecer que ser beneficiario de Familias en Acción aumenta la probabilidad de asistir a un colegio que mejora su calidad en el tiempo. En particular, se encuentra un efecto LATE de entre 6.77 p.p. y 7.85 p.p. También se encuentran resultados positivos para el mecanismo de cambio a un mejor colegio; ser tratado aumenta la probabilidad de cambiarse a un colegio con mejor resultado en la prueba saber 11 en 3.1 p.p.

A través de modelos de mediación se procede a descomponer el efecto directo e indirecto (explicado a través del mecanismo de calidad educativa) del programa. Se encuentra que la asistencia a un colegio con resultados promedio en la prueba Saber 11 por encima de la media municipal explica entre 12% y 20% del efecto total sobre graduación. Para el caso en que la selección aleatoria afecta la cola izquierda de la distribución, es decir, aquellos estudiantes que en ausencia del programa no presentarían el examen, la asistencia a un mejor colegio explica entre un 20% y 23% los resultados sobre la prueba estandarizada.

Es importante señalar que estos efectos son locales, por dos razones: 1) se estiman a través de variables instrumentales, por lo que los efectos se estiman sobre la población que participa en el programa cuando es elegible (Angrist y Imbens, 1994); 2) el efecto sólo se estima en la vecindad del punto de elegibilidad.

Por último, la pregunta de investigación de este estudio es relevante ya que evalúa uno de los posibles mecanismos de calidad educativa, desde el lado de la demanda, que podría explicar los resultados obtenidos en evaluaciones pasadas sobre variables de desempeño académico. Es así, que el presente documento es un paso más para explicar los distintos mecanismos que podrían generar impactos gracias a los programas de Transferencias Monetarias Condicionadas. Este estudio intenta generar una contribución a esta literatura.

## Bibliografía

Attanasio, O., Fitzsimons, E., & Gomez, A. (2005). "The impact of a conditional education subsidy on school enrolment in Colombia." Londres: *The institute for Fiscal Studies*.

Baez, J. E., Camacho, A. (2011). "Assessing the Long-term Effects of Conditional Cash Transfers on Human Capital: Evidence from Colombia." Bonn: *Forschungsinstitut zur Zukunft der Arbeit*.

Baez, J.E., Camacho, A., Conover, E., Zárate, R.A. (2012). "Conditional Cash Transfers, Political Participation, and Voting Behavior," IZA Discussion Papers 6870, Institute for the Study of Labor (IZA).

Banco Mundial. (2010). "Evidence and Lessons Learned from Impact Evaluations in Social Safety Nets." Manuscript. Washington, DC: *The World Bank*.

— (2008). "Calidad de la educación en Colombia: un análisis y algunas opciones para un programa de política." Washington, DC: *Banco Mundial*.

— (2003). "Programa Familias en Acción" Obtenido de <http://siteresources.worldbank.org/SAFETYNETSANDTRANSFERS/Resources/281945-1131468287118/1876750-1140107387177/ColombiaCCT.pdf>

Baron, R. M., Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality & Social Psychology*, 51, 1173–1182.

Barreca, A. I., M. Guldi, J. M. Lindo, and G. R. Waddell (2011): "Saving Babies? Revisiting the effect of very low birth weight classification," *The Quarterly Journal of Economics*, 126(4), 2117–2123.

Bernal, R., Peña, X. (2011). *Guía práctica para la evaluación de impacto: Guía práctica para la evaluación de impacto*. Universidad de Los Andes, Colombia. Retrieved from <http://www.jstor.org/stable/10.7440/j.ctt1b3t82z>

Castañeda, T. (2005) "Targeting social spending to the poor with proxy - means testing: Colombia's SISBEN system". *Banco Mundial*.

De Brauw, A., Gilligan, D. O., Hoddinott, J., & Roy, S. (2015). "The Impact of Bolsa Familia on Schooling." *World Development*.

Díaz, J., Grau, N., Reyes, T., & Rivera, J. The Impact of Grade Retention on Juvenile Crime. *Serie documentos de trabajo Universidad de Chile departamento de Economía*. No. 429, pp. 1 - 38, Septiembre, 2016

DNP. (25 de Julio de 2014). Red Unidos para Superación de la Pobreza Extrema. Obtenido de <https://www.dnp.gov.co/programas/desarrollo-social/pol%C3%ADticas-sociales-transversales/Paginas/red-unidos-para-superacion-de-la-pobreza-extrema.aspx>

DNP. (2010). "El camino recorrido: Diez años de Familias en Acción." Bogotá.

Friedman, M (1962), Capitalism and Freedom. Chicago: *University of Chicago Press*



Galassi, (2013). „The effects of Conditional Cash Transfers (CCTs) on cognitive skills in Latin America: The case of the Brazilian "Bolsa Família"“. *Research Gate*

Gaviria, A., & Barrientos, J. H. (2001). “Calidad de la educación y rendimiento académico en Bogotá.” Bogotá.

Gelman, A., Imbens, G., (2014). "Why High-order Polynomials Should not be Used in Regression Discontinuity Designs," *NBER Working Papers 20405, National Bureau of Economic Research, Inc.*

Gobierno de Cundimarca. (2008). “Generalidades del Sisben” Obtenido de [http://soacha-cundinamarca.gov.co/apc-aa-files/65336436636635633532306433346433/Normas\\_Sisb\\_n.pdf](http://soacha-cundinamarca.gov.co/apc-aa-files/65336436636635633532306433346433/Normas_Sisb_n.pdf)

Hoxby, C. M. (2002). School Choice and School Productivity (or Could School Choice be a Tide that Lifts All Boats?). *NBER Working Papers 8873, National Bureau of Economic Research, Inc.*

Imbens, G., & Kalyanaraman, K. (2012). Optimal bandwidth choice for the regression discontinuity estimator. *The Review of Economic Studies*, 79(3), 933-959.

Imbens, G., & Angrist, J. (1994). Identification and estimation of local average treatment effects. *Econometrica*, 62(2), 467-476.

Kim, T., Lee, J.-H., & Lee, Y. (2008). “Mixing versus sorting in schooling: Evidence from the equalization policy in South Korea.” Seoul. *Economics of Education Review*.

Lee Benson , A. (2012). “Los efectos de Familias en Acción sobre la calidad de la oferta educativa.” Bogotá: *Documentos CEDE*.

Lee, D. S. (2009). “Training, wages, and sample selection: Estimating sharp bounds on treatment effects”. *The Review of Economic Studies*, 76(3), 1071-1102

McCrary, J., (2008). “Manipulation of the running variable in the regression discontinuity design: A density test”, *Journal of Econometrics*, 142, issue 2, p. 698-714.

Melguizo, T., Sanchez, F., Velasco, T., (2015) "Credit for Low-Income Students and Access to and Academic Performance in Higher Education in Colombia: A Regression Discontinuity Approach," *DOCUMENTOS CEDE 012854*, Universidad de los Andes.

MEN. (17 de Diciembre de 2015). “Pruebas Saber.” Obtenido de <http://www.mineducacion.gov.co/1759/w3-article-244735.html>

Neilson, C. (2013). Targeted vouchers, competition among schools, and the academic achievement of poor students. Technical report, mimeo.

Preacher, K.J., Hayes, A.F. (2008). *Behavior Research Methods*. 40: 879. doi:10.3758/BRM.40.3.879

Urquiola, Miguel, and Eric Verhoogen. (2009). "Class-Size Caps, Sorting, and the Regression Discontinuity Design." *American Economic Review*, 99(1): 179-215.

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

## **Apéndices**

### *A1. Cruce base de datos*

Como se mencionó en la sección de datos este estudio contiene información de cuatro fuentes de datos administrativos. En esta sección se explican en detalle las estadísticas del cruce realizado, en particular, se muestra como se obtiene el conjunto de individuos de la muestra final.

En primer lugar, se realiza el cruce fundamental del estudio, que consiste en las fuentes de información de Familias en Acción Fase 1 y SISBEN 99. El fin de éste es identificar tanto los beneficiarios como los no beneficiarios del programa con sus respectivos puntajes SISBEN. Por un lado, la muestra de FeA Fase 1 contiene 2,858,040 observaciones, mientras que la de SISBEN99 contiene 24,302,461 individuos. En total cruzan 2,148,367 personas, es decir, el 75.17% del total de tratados en esta fase se encontraron en los datos administrativos del SISBEN 99.

Al saber que familias recibieron el programa y que familias no, se procede a cruzar con los registros de los datos administrativos del SIMAT. En otras palabras, se busca la información académica para los niños que asisten a colegios oficiales entre los años 2005 y 2013. Esta última base de datos tiene información para 21,313,849 estudiantes y como resultado del cruce de datos se identifican 4,021,915 individuos en ambas bases. De esta forma se consolida una base de datos con estas 3 fuentes de información que cuenta con 485,602 individuos tratados y 3,536,313 no tratados. Es así que se obtiene el universo de individuos con el cual se realizan las estimaciones.

Finalmente, se realizan cruces a nivel individual o de colegio con los datos administrativos del ICFES. Estos últimos se realizan para crear las variables de resultado del estudio. Es importante resaltar que en caso de que no se logre construir una variable de resultado para un individuo se deja en vacío y de esta forma puede variar las observaciones finales de cada una de las variables. En la Tabla 16 se muestran las variables de resultado diferenciando el número de observaciones para participantes y no participantes en el programa.

Por último, se resalta que los individuos pertenecientes a un departamento focalizado entre 2001 y 2006 son 3,642,757.

Tabla 10: Estadísticas de cruce base final

	Familias en Acción =1	Familias en Acción =0	Total
Colegio mejor que la media del munic.	183,505	1,800,969	1,984,474
Colegio mejor que el percentil 75 del munic.	183,505	1,800,969	1,984,474
Graduación	166,284	1,221,788	1,388,072
Puntaje estandarizado Prueba Saber 11	117,756	1,121,456	1,239,212
Tenencia de activos en educación	67,942	592,188	660,130

Nota: La tabla presenta el total de observaciones de la base final para cada una de las variables de resultado distinguiendo beneficiarios de no beneficiarios.

Para el cruce de las distintas fuentes de información se realizaron varias metodologías. La primera fue utilizando como llave el documento de identidad del individuo. La segunda es similar y se creó una llave con el documento de identidad del jefe del hogar y la fecha de nacimiento debido a que, en muchas de estas encuestas, se asigna el documento del padre al niño. Por último, el principal método de cruce consistió en un algoritmo que reconoce distancias fonéticas entre palabras, en el cual se utilizó como variable de cruce el nombre completo. Además, se incluyó dentro del algoritmo la fecha de nacimiento para evitar el error tipo I, es decir, que no crucen falsos positivos. Métodos similares de cruce entre distintas fuentes de información han sido empleados por otras investigaciones como Báez y Camacho (2011), y Melguizo, Sánchez y Velasco (2014) que utilizan algunas de las mismas fuentes de datos de este documento.

#### *A2. Lee Bounds*

Lee (2008) propone un método de identificación parcial para estimar el efecto de un programa bajo la existencia de selección no aleatoria. El autor motiva esta aplicación con un programa de entrenamiento que genera un aumento en los ingresos de las personas, sin embargo, este efecto se encuentra condicionado a estar trabajando. Para los efectos de este estudio, los ingresos corresponderían al puntaje de la prueba Saber 11, mientras que la probabilidad de trabajar corresponde a la probabilidad de graduarse.

Sea  $Y_i$  el puntaje de la prueba Saber 11, sea  $FEA_i$  el tratamiento del programa;<sup>29</sup> y por último, sea  $S_i$  una variable binaria que toma el valor de 1 si la persona se gradúa de los niveles de educación básica ó 0 dlc. Considere el siguiente modelo de variable aleatoria latente:

$$Y_i = \alpha_1 + \beta FEA_i + \epsilon_i$$

$$S_i^* = \alpha_2 + \gamma FEA_i + v_i$$

$$S_i = I[S_i^* > 0]$$

De esta manera el tratamiento de FeA afecta por una parte el resultado en la prueba y a su vez también afecta la probabilidad de graduación. Siendo consistentes con los resultados sobre la variable graduación asumamos que  $\gamma > 0$ . En ausencia de selección el efecto del programa estaría dado por la diferencia de medias entre el grupo de tratamiento y el grupo de control. Sin embargo, la media para cada uno de los grupos está dada por:

$$E[Y_i|D_i = 1, S_i = 1] = \alpha_1 + \beta + E[\epsilon_i|D_i = 1, S_i = 1]$$

$$= \alpha_1 + \beta + E[\epsilon_i|v_i > -\alpha_2 - \gamma]$$

$$E[Y_i|D_i = 0, S_i = 1] = \alpha_1 + \beta + E[\epsilon_i|D_i = 0, S_i = 1]$$

$$= \alpha_1 + \beta + E[\epsilon_i|v_i > -\alpha_2]$$

Si se toma la diferencia de medias se obtiene que:

$$E[Y_i|D_i = 1, S_i = 1] - E[Y_i|D_i = 0, S_i = 1]$$

$$= \beta + E[\epsilon_i|v_i > -\alpha_2 - \gamma] - E[\epsilon_i|v_i > -\alpha_2]$$

De esta manera el efecto del programa sobre la prueba estaría sesgado, ya que en el modelo estructural el efecto sobre la prueba está dado por el parámetro  $\beta$ . Es así, que el sesgo contiene la diferencia de los dos términos del error. Note que el primer término del error está compuesto por dos grupos, aquellas personas que debido al programa se graduaron del colegio, y aquellas personas que incluso en ausencia del programa se hubieran graduado del programa. Al primer grupo lo llamaremos “Graduados por FeA” y al segundo “Graduados

---

<sup>29</sup> Para los propósitos de esta sección asumamos que  $FEA_i$  es aleatorio.

siempre<sup>30</sup>). Si en este grupo no estuvieran los “Graduados por FeA” el valor esperado de los dos errores se cancelaría. De esta manera por la ley de probabilidad total se puede descomponer el primer término del error en la siguiente forma:

$$\begin{aligned}
 E[\epsilon_i | v_i > -\alpha_2 - \gamma] &= E[\epsilon_i | -\alpha_2 > v_i > -\alpha_2 - \gamma] \Pr[v_i < -\alpha_2 | v_i > -\alpha_2 - \gamma] \\
 &\quad + E[\epsilon_i | v_i > -\alpha_2] \Pr[v_i > -\alpha_2 | v_i > -\alpha_2 - \gamma] \\
 &= E[\epsilon_i | -\alpha_2 > v_i > -\alpha_2 - \gamma] p \\
 &\quad + E[\epsilon_i | v_i > -\alpha_2] (1 - p)
 \end{aligned}$$

El primer término corresponde al valor esperado del error de los graduados por la participación en el programa, mientras que el segundo es la selección de los “Graduados siempre”. De esta manera, el error de selección consiste en un promedio ponderado de aquellas personas que en ausencia del programa se hubieran graduado y de los que dado el efecto del programa pudieron graduarse.

La contribución de Lee (2008) consiste en encontrar cuál es la proporción de aquellas personas que se hubieran graduado en ausencia del programa. Si bien no se puede identificar quiénes son aquellas personas en los datos, sí se puede identificar la proporción de este grupo de personas en la población, bajo el supuesto de aleatoriedad del tratamiento. En particular, la proporción de “graduados por FeA” corresponde a:

$$E[S_i | FeA_i = 1] - E[S_i | FeA_i = 0] = \Pr[-\alpha_2 > v_i > -\alpha_2 - \gamma]$$

De esta manera, esta proporción  $p$  consiste en:

$$p = \frac{E[S_i | FeA_i = 1] - E[S_i | FeA_i = 0]}{E[S_i | FeA_i = 1]}$$

Esta ecuación es muy intuitiva. Por una parte,  $E[S_i | FeA_i = 1]$  son todas las personas que con el programa se gradúan, este grupo se compone de los graduados siempre y los graduados por FeA. Por otra parte, la proporción de graduados siempre en ausencia del programa corresponde a  $E[S_i | FeA_i = 0]$ , por lo que la diferencia entre los dos grupos correspondería a los “graduados por FeA”. Este se divide sobre el denominador para que la

---

<sup>30</sup> Haciendo referencia al término anglosajón “Always takers”

proporciones sumen 1. En otras palabras, la proporción de graduados siempre es la misma en grupo de tratamiento y control, de esta forma, si se toma la diferencia entre el tratamiento y control se obtiene la proporción de graduados por FeA. Dado que se asumimos que  $FEA$  es aleatorio, podemos identificar la proporción  $p$ .

Por último, para construir el límite inferior de los Lee Bounds se supone que los graduados siempre corresponde a aquellas personas por debajo del percentil  $1 - p$ , y para el límite superior aquellas por encima del percentil  $p$ . Definamos los dos siguientes parámetros:

$$\underline{\mu} = E[Y_i | FEA = 1, S_i = 1, Y_i < y_{1-p}]$$

$$\bar{\mu} = E[Y_i | FEA = 1, S_i = 1, Y_i > y_p]$$

Por lo que el límite inferior y superior del efecto del programa sobre el puntaje de la prueba es:

$$\underline{\beta} = \underline{\mu} - E[Y_i | D_i = 0, S_i = 1]$$

$$\bar{\beta} = \bar{\mu} - E[Y_i | D_i = 0, S_i = 1]$$

### A3. Variables de resultado

Las siguientes 5 figuras muestran los resultados gráficos para cada una de las variables de largo plazo y los mecanismos evaluados en el presente estudio.

