

## **MÁS ALLÁ DE LA INFRAESTRUCTURA: EL IMPACTO DE LAS BIBLIOTECAS PÚBLICAS EN LA CALIDAD DE LA EDUCACIÓN**

**Paul Rodríguez-Lesmes  
José D. Trujillo  
Daniel Valderrama**

**SERIE DOCUMENTOS DE TRABAJO**

**No. 133**

**Febrero de 2013**

# Más allá de la infraestructura: el impacto de las bibliotecas públicas en la calidad de la educación.

Paul Rodríguez-Lesmes José D. Trujillo Daniel Valderrama\*

4 de diciembre de 2012

## Resumen

La literatura sobre la calidad de la educación ha prestado poca atención al papel que tienen las bibliotecas públicas dentro de los determinantes del desempeño educativo. Las bibliotecas públicas son activos externos al colegio y al hogar del estudiante, pero hacen parte del entorno social que les rodea. La puesta en marcha a finales de 2001 de tres bibliotecas de gran tamaño en Bogotá, conocidas como *megabibliotecas*, nos permite analizar el impacto de estas iniciativas sobre la calidad de la educación en los colegios aledaños. Dicho impacto se daría a través de mecanismos adicionales a la simple reducción de costos al acceso a la información: las bibliotecas renovaron el espacio público mediante la generación de espacios agradables y amigables hacia la educación, además ofrecen regularmente actividades lúdicas dirigidas a las habitantes del sector. Aprovechando la distancia del plantel educativo a la biblioteca como una aproximación al costo de acceso a la misma, utilizando para ello Diferencia en Diferencias junto a la descomposición Blinder Oaxaca. Encontramos que las mismas parecen no tener un impacto significativo sobre el desempeño académico general en los exámenes oficiales SABER 11 durante los años posteriores a su implementación. Se recomienda analizar programas específicos que aprovechen las bibliotecas para actividades escolares y otras posibles variables de impacto como actitudes hacia el estudio y aspiraciones a la educación superior.

**Palabras Clave** efecto vecindario, calidad de la educación, bibliotecas públicas.

**JEL codes** D62, I21, H52.

---

\*E-mail: p.lesmes.11@ucl.ac.uk, jdtrujillos@dane.gov.co, valderrama.daniel@ur.edu.co. Este trabajo fue desarrollado durante nuestros estudios de Maestría en la Universidad del Rosario y en la Universidad de los Andes, y bajo el programa de Jóvenes Investigadores de la Facultad de Economía de la Universidad del Rosario. El mismo fue financiado por el ICFES, bajo la Convocatoria de Investigación para Estudiantes de Posgrado 2010. Agradecemos especialmente la asesoría de Hugo Ñopo, Andrés García y los valiosos comentarios del equipo del ICFES y demás participantes de la convocatoria.

# 1. Introducción

La literatura asociada a los determinantes de la calidad de la educación suele considerar, además de variables socioeconómicas del estudiante, una serie de insumos educativos que hacen parte de las características del colegio como la infraestructura de las aulas, el acceso a computadores, libros de estudio, laboratorios y/o bibliotecas; estas últimas han sido definidas como un activo educativo al cual accede un estudiante por medio del colegio, es decir, se han considerado únicamente las bibliotecas privadas, dejando así un vacío en la literatura referente al efecto causal de las bibliotecas públicas sobre la calidad de la educación.

Tres bibliotecas mayores fueron inauguradas en el segundo semestre de 2001 en Bogotá como parte central del proyecto *BiblioRed* que buscaba mejorar el acceso de los habitantes de la ciudad a los servicios que ofrecen las bibliotecas. Además estas fueron diseñadas para proveer algo más que el acceso a la información; también buscaban ser centros culturales, artísticos y de entretenimiento. Por tal motivo se considera que uno de los objetivos de las bibliotecas es modificar hábitos de todo tipo en su entorno y que indirectamente afectan el desempeño escolar, además de suministrar, como tradicionalmente lo han hecho las bibliotecas públicas, herramientas para que los estudiantes puedan desarrollar sus competencias académicas.

En este trabajo se analiza el impacto que sobre la calidad de la educación tienen dos de las tres *megabibliotecas*, las de *El Tintal* y *El Tunal*. Estas dos se encuentran ubicadas en sectores populares de la ciudad, donde la densidad de colegios alrededor es más alta, lo que aumenta el número de colegios que reciben el posible impacto de las bibliotecas<sup>1</sup>. Aunque la biblioteca Virgilio Barco también hace parte del proyecto *BiblioRed*, se encuentra en un entorno social diferente a las dos primeras, con una menor densidad de colegios alrededor de la biblioteca que ofrecen sus servicios a hogares con características socioeconómicas diferentes. Por ejemplo, los estudiantes de alrededor de la Virgilio Barco son en mayor proporción de colegios privados y tienen mas acceso a activos educativos como libros y computadores dentro del hogar lo que implica una interacción de otra índole con

---

<sup>1</sup>La densidad alrededor de las bibliotecas es un punto crucial pues mientras más cerca esté un colegio y/o estudiante de las bibliotecas, mayor será la posibilidad de interacción con esta y así la densidad se convierte en una aproximación del número de colegios que reciben el posible impacto, un aspecto importante para hacer robusta la inferencia sobre la existencia o no de este.

la biblioteca, razón por la cual no fue considerada.

Una biblioteca pública es externa al colegio y al hogar del estudiante pero interna al vecindario de éste, esto implica que el impacto de estas sobre el desempeño educativo obedezca a un mecanismo de trasmisión distinto al que se ha considerado tradicionalmente. Estas bibliotecas presentan dos tipos de interacciones con el vecindario sobre el que se construye: la primera, es una interacción directa con el sistema educativo, donde las bibliotecas cumplen su papel de lugares que permiten el acceso a libros, salas de estudio y salas de Internet; el segundo mecanismo de interacción se refiere a los espacios de esparcimiento, la atmósfera cultural y la oferta lúdica que estas promueven, por ejemplo, las ‘vacaciones recreativas’ para niños. Si bien estos espacios no son estrictamente educativos pueden provocar cambios en los hábitos de los ciudadanos del vecindario que se pueden reflejar de manera indirecta en el desempeño de los estudiantes. Por otra parte, por medio del *efecto contagio*, quienes experimentan cambios en los hábitos por asistir a la biblioteca influyen los hábitos de las otras personas con que se relacionan, sin necesidad que estas últimas usen la biblioteca, lo cual magnifica el efecto y lo hace incondicional al uso de la biblioteca.

Desafortunadamente no tenemos información sobre el uso de las bibliotecas por parte de los colegios u otra información que nos permita identificar de forma aislada estos mecanismos. No obstante, podemos analizar el impacto neto de las bibliotecas en su vecindario aprovechando la dimensión espacial. El análisis se realiza a nivel de colegio, utilizando la cercanía de la institución como una aproximación al grado de interacción de los estudiantes con las bibliotecas en un modelo de Diferencias en Diferencias junto a la descomposición Blinder-Oaxaca. Para la medición del desempeño utilizamos la prueba de acceso a la educación superior, SABER 11, entre 2000 y 2008 cuya información fue provista por el ICFES<sup>2</sup>.

El resto de este documento está organizado de la siguiente manera: La sección 2 presenta la revisión de literatura, la sección 3 describe y explica los datos utilizados y los principales hechos estilizados, en la sección 4 se plantea la estrategia de identificación y descomposición del efecto de las bibliotecas y, finalmente, la sección 5 y 6 presentan los resultados y conclusiones.

---

<sup>2</sup>El ICFES financió este estudio a través de su programa ‘Convocatoria ICFES para estudiantes de posgrado 2010’.

## 2. Las Bibliotecas en la literatura de los determinantes del desempeño educativo

La educación es un factor relevante para el crecimiento económico y el desarrollo, y uno de los principales determinantes de la movilidad social y la lucha contra la pobreza (Hanushek y Kimko, 2000; Barro, 2001). Los países en desarrollo han mejorado sus niveles de cobertura en educación en los últimos años por lo que el gran reto de la política pública educativa actual gira en torno a la calidad de la educación. El acceso a una educación con altos niveles de calidad es inequitativo y por tanto, las políticas que buscan este objetivo deben plantear no solo una mejora transversal sobre el sistema educativo, sino también focalizar los recursos sobre los grupos más vulnerables (Gamboa, 2003). En sintonía con lo anterior, el análisis de los determinantes del desempeño escolar es de gran relevancia para definir las políticas públicas más eficaces en busca de mayores niveles de calidad en la educación. Según Vegas y Petrow (2007) los múltiples determinantes de la calidad de la educación pueden agruparse de la siguiente manera:

**Características de la demanda** Un primer grupo de estudios argumenta que el desempeño escolar depende de ciertas características relacionadas con la demanda educativa, es decir, los hogares. Dentro de este grupo se encuentran aspectos como el nivel socioeconómico del hogar, por ejemplo, para los hogares más pobres el costo de oportunidad de estudiar frente a trabajar es muy alto, lo que se refleja en un bajo desempeño educativo (Caro, 2000); otras variables, como la educación de los padres, están relacionadas con la presencia de un ambiente que promueve el estudio de forma tanto cultural (Fertig y Schmidt 2002; Abdul-Hamid 2003 y BM 2005) como tangible, por medio de los recursos educativos con los que cuenta el hogar (Murnane et al. (1981); Gamboa et al. (2010) y Blomeyer et al. (2009)).

La literatura colombiana no tiene una posición única respecto a la incidencia o no de las características de los hogares en la calidad de la educación. Gaviria y Barrientos (2001) sostienen que los padres con bajos recursos usualmente son los que tienen un nivel educativo bajo; además, dichas características están correlacionadas con la elección de la institución educativa, que finalmente sería la causa principal de las diferencias en el desempeño (Gaviria y Barrientos 2001, Sarmiento et al. 2000); así, los autores muestran

una especie de circularidad en los determinantes del desempeño educativo, que empieza por el nivel socioeconómico de los padres y desemboca en la naturaleza de la institución educativa.

**Características de la oferta** Vegas y Petrow (2007) definen este grupo de determinantes como aquel en el que se encuentran las características de la institución educativa como: el ambiente institucional y la infraestructura física y humana. La infraestructura física y de talento humano se refiere al acceso que permite una institución educativa en términos de laboratorios, aulas, computadores, bibliotecas privadas, calidad de los profesores y el acceso a otros recursos educativos; mientras que el ambiente institucional, agrupa variables como la organización y logística de la administración, el nivel de descentralización (Sánchez, 2006; Melo, 2005), la rivalidad entre las escuelas (BM, 2005) y el compromiso con el desempeño (Fusch y Woessman 2004). Hanushek y Woessman (2007) afirman que esta última clase de diferencias son mucho más importantes que las diferencias en cuanto a recursos físicos.

En la literatura colombiana se han analizado diversos determinantes de los colegios sobre la calidad de la educación, por ejemplo, Iregui et al. (2006) examinaron la estructura física, Uribe et al. (2005) la educación de los docentes, Caro (2002) analizó la asistencia de los docentes al trabajo mientras que Nuñez (2002), Gaviria y Barrientos (2001) exploraron las diferencias entre colegios públicos y privados; en todos los casos se obtuvieron conclusiones que van de la mano con la evidencia internacional, donde mejores infraestructuras, una mayor educación, menor inasistencia de los docentes al trabajo y una administración privada, implican una mejor calidad de la educación para los estudiantes de una institución.

**Las bibliotecas y el desempeño escolar** En la sección anterior se observa que las bibliotecas han sido consideradas como uno de los muchos factores que hacían parte de la infraestructura física de los colegios, es decir, como bibliotecas privadas, lo que muestra el poco protagonismo de éstas en la literatura. Sin embargo, cabe resaltar algunos trabajos pioneros en la literatura que se pregunta por la relación entre las bibliotecas y la calidad de la educación. Por ejemplo, Lance et al. (1994) realizan un estudio principalmente descriptivo de las mejoras sobre el desempeño escolar que están asociadas con las bibliotecas en Colorado; otros trabajos han considerado las bibliotecas como uno de los determinantes

de los resultados específicos en las evaluaciones de lectura, escritura y pensamiento crítico (Williams et al., 2001; Lance 2000; Baughman, 2000; Lance 2001; Rodney, 2003 entre otros). Según Smith (2001), quien realiza una estimación del impacto de éstas controlando por factores socio-económicos de los estudiantes, las bibliotecas mejoran en un 4% el desempeño escolar, un resultado recopilado en el trabajo de Lonsdale (2003), quien relaciona los distintos artículos donde se analiza el impacto de las bibliotecas y encuentra que el uso de estas lleva a un desarrollo de las habilidades derivadas de la lectura: comprensión de lectura, tecnología de la información, tiempo dedicado a la lectura y resultados académicos en general.

En resumen, hay un consenso sobre el impacto positivo de las bibliotecas en la calidad de la educación, no obstante, la mayoría de los estudios no tienen un sustento estadístico sólido. Más aún, la sola tenencia de una biblioteca implica de por sí un esfuerzo de la institución por mejorar la educación y probablemente esté correlacionado con otras iniciativas. Por otra parte, los estudios solo tienen en cuenta bibliotecas privadas pero no bibliotecas de carácter público, y menos instituciones al estilo de las *megabibliotecas*. Como veníamos diciendo, estas promueven programas culturales y potencialmente generan cambios en los hábitos de sus beneficiarios<sup>3</sup>.

### 3. Datos: cálculos y hechos estilizados

La información disponible está a nivel de promedio por colegio<sup>4</sup> para la ciudad de Bogotá. La base es de tipo panel, contando con información de los colegios antes y después de las bibliotecas por lo cual todas las variables no observables que se mantengan constantes en el tiempo (e.g las diferencias en las vecindades de las localidades) son controladas, identificando así con menor ruido el efecto de las bibliotecas. En esta sección vamos a

---

<sup>3</sup>El mecanismo adicional propuesto, pero que desafortunadamente no podemos contrastar empíricamente, es conocido como ‘efecto vecindario’. Definido de manera genérica por Durlauf (2004) como la manera en que una comunidad influencia los hábitos de los ciudadanos que se encuentran alrededor. Uno de los mecanismos mediante los cuales se da el efecto vecindario, es el ‘efecto contagio’: la replicación por parte de un individuo de las decisiones de sus pares, y puede ser ‘positivo’ como el cumplimiento de las normas, o ‘negativo’ como el consumo de drogas (Jenks y Mayer; 1990).

<sup>4</sup>A nivel de estudiantes se realizaron algunos ejercicios exploratorios obteniendo poca robustez en los resultados. A dicho nivel no tenemos suficiente información por parte de la demanda para controlar la variabilidad de los datos. Esperamos poder conseguir información más detallada, incluyendo información de ubicación geográfica de cada estudiante respecto a la biblioteca para una próxima versión.

detallar la información utilizada en este artículo.

### 3.1. Una aproximación a la calidad de la educación

La medición de la calidad de la educación en Colombia es un tema que siempre desata polémicas, para este caso se utilizó la prueba SABER 11. Si bien no todos los autores están de acuerdo en utilizar esta clase de pruebas como un instrumento único para medir la calidad, la prueba SABER 11 es práctica dado que su resultado influye en las posibilidades de los estudiantes en su vida académica y laboral, y sus preguntas están relacionadas con conocimientos básicos y habilidades que los estudiantes adquieren durante la educación media.

Dado que la prueba de cada año es distinta a la del año anterior, los puntajes obtenidos en una y otra prueba no son comparables pues no se sabe hasta qué punto reflejan un cambio el desempeño educativo de un estudiante o simplemente son producto de los cambios en el instrumento de medición, por ejemplo, que el formulario de preguntas diseñado por el ICFES sea ‘más fácil’ en un año que en otro. Con el fin de eliminar este error de medición, el ICFES recomienda estandarizar los resultados de las pruebas para cada año, por ende, para el análisis a nivel de institución se tomó el promedio de los resultados de cada uno de sus estudiantes en el año respectivo<sup>5</sup>. Se tienen en cuenta los resultados estandarizados<sup>6</sup> en las áreas de matemáticas, lenguaje, ciencias (un promedio entre las áreas de biología, física y química<sup>7</sup>) y un promedio entre las tres áreas que denominamos “resultado general”. Nótese que el hecho de que la variable de medición de desempeño educativo sea sometida a un proceso de estandarización implica una lectura distinta del efecto marginal de las bibliotecas. Un impacto positivo (negativo) de las *megabibliotecas* se

---

<sup>5</sup>Si bien la prueba es diferente cada semestre, cada colegio presenta sólo una de las dos versiones al año

<sup>6</sup>Respecto a media y desviación estándar de Bogotá.

$$z_{it} = \frac{x_{it} - x_t}{v(x_t)}$$

Donde  $z_{it}$  es el valor estandarizado del resultado colegio o del individuo  $i$  en el periodo  $T$ .  $x_{it}$  es el resultado del colegio o del individuo  $i$  en el periodo  $T$ .  $v(x_t)$  es la desviación estándar de los resultados en el periodo  $T$ .

<sup>7</sup>Si bien el concepto de “ciencias” puede llevar a discusiones, esta clasificación obedece a la resolución 489 de Octubre 20 de 2008 del Ministerio de educación Nacional el cual define así el área de ciencias, si bien en esta resolución se incluye el puntaje de sociales que no incluimos debido a que el cálculo de esta área cambia durante el periodo de análisis.

interpretara como una mejora (desmejora) relativa de los estudiantes o colegios expuestos a las bibliotecas respecto a los demás estudiantes o colegios. También se incluyeron como otras aproximaciones al desempeño educativo el resultado por separado en las distintas áreas (Matemáticas, Lenguaje y Ciencias).

### **3.2. Las fuentes de información para controlar el efecto causal**

Como se explico antes, los datos para observar la calidad de la educación provienen de la prueba SABER 11 aplicada por el ICFES. Además de estos datos y con el fin de identificar el efecto causal de las bibliotecas, incluimos variables que la literatura ha identificado como determinantes claves de la calidad de la educación. Estas variables de control a nivel de institución provienen de los cuestionarios C100 y C600 del Ministerio de educación Nacional, además se agregan controles del vecindario por medio de los datos del Censo General de Población del 2005 realizado por el DANE.

Por medio de la base de C-600 se han definido dos grupos de variables de control a nivel de institución: el primer grupo son las variables que hacen parte de la institucionalidad del establecimiento educativo como lo son el calendario<sup>8</sup>, la naturaleza y la jornada<sup>9</sup>. La naturaleza del colegio se refiere a si este es Oficial o No Oficial. Por último, la jornada, hace referencia al tiempo y horario que los estudiantes del colegio estudian, sólo por la mañana, sólo por la tarde, nocturno, fines de semana o en jornada completa<sup>10</sup>.

El segundo grupo de variables de control consiste en aquellas características que ofrece la institución educativa y que cambian en el corto y mediano plazo. En nuestro caso tenemos: la matrícula del colegio, que puede dar idea de las economías de escala con que cuenta el plantel educativo, para su medición se utilizo el logaritmo del número de estudiantes matriculados con el fin de incorporar la idea de rendimientos marginales decrecientes en el impacto que el tamaño de la matricula tiene sobre la calidad de la educación. El capital humano de los profesores, definido con base en el nivel de educación de estos (primaria,

---

<sup>8</sup>El calendario hace referencia a si el colegio empieza su año escolar a principio de año (calendario A) a o a mediados del año (calendario B). Alrededor de las bibliotecas sólo encontramos colegios de calendario A.

<sup>9</sup> Siguiendo a North (1996), las instituciones son normas y reglas, en este caso explícitas, que definen ciertos parámetros de funcionamiento del colegio

<sup>10</sup>Sólo tuvimos en cuenta colegios con jornadas de la mañana, tarde o completa. En la sección de datos se explica por qué.

secundaria, educación superior y posgrado). La proporción de profesoras respecto al total de docentes como idea de las diferencias en la pedagogía que tiene una profesora respecto a un profesor y que impacta finalmente sobre la enseñanza a los alumnos. El género del colegio, si bien esta variable es considerada por algunos como parte de las normas de largo plazo del colegio, se considero como parte del grupo de variables que son modificables en el corto plazo dado que en este documento el género se mide como la proporción de niñas sobre el total de alumnos, indicador que cambia año tras año para los colegios mixtos. Otras variables que se consideran y que afectan de manera más directa a los estudiantes que específicamente presentaron el examen son: el número de alumnos en el grado 11 y la proporción de niñas respecto al total de estudiantes que presentaron el examen.

Por medio de C-100 se toma un último grupo de controles que hace referencia a la infraestructura física de los colegios. Desafortunadamente esta fuente de información llega hasta 2002, por lo que los datos incluidos en la regresión para los siguientes años son los mismos que presentaban los colegios en 2002, por lo que se consideran un *stock* inicial de activos físicos. El cuestionario de C-100 permite controlar por: área construida respecto al número de estudiantes, área deportiva respecto al número de estudiantes, presencia de una biblioteca en el colegio y una medida de la calidad de los activos educativos <sup>11</sup>.

En el caso de los controles a nivel de vecindario se tomó para cada colegio las variables socioeconómicas que ofrecía el Censo a nivel de manzana, asignando a cada colegio un promedio del entorno socioeconómico de los hogares que lo rodean a 50 metros de radio: tasa de estudiantes, tasa de personas con máximo primaria, tasa de hogares con actividad económica en la vivienda, tasa de personas que llevan más de 5 años en la misma residencia, tasa de personas que vienen de otro municipio (últimos 5 años), tasa de personas que vienen de la zona rural (últimos 5 años), edad promedio, tasa de personas que ayunaron, población en edad de trabajar y población económicamente activa.

Una vez se tiene en cuenta la variable de resultado a evaluar y los controles a utilizar sólo resta definir la variable de selección de aquellos colegios que fueron afectados por las bibliotecas y aquellos que no. Para esto se utilizará la distancia euclidiana entre el colegio

---

<sup>11</sup>Se considera que los activos de un colegio son de calidad si dicho colegio cuenta simultáneamente con laboratorios de sistemas, física y química, dado que son muy pocos los que presentan estas características, el hecho de que un colegio cuente con los tres tipos de laboratorios es una aproximación de la importancia que le otorga a este tipo de inversiones

y la biblioteca más cercana como puede verse en la figura 4<sup>12</sup> (la justificación de esto se encuentra en la sección 4.1).

### 3.3. Depuraciones de los datos

Los cuadros 1 y 2 reportan para los diferentes rangos de distancia a los que se encuentra un colegio de la biblioteca más cercana (primera columna) el número de colegios<sup>13</sup> y de alumnos, en la segunda columna esta la información disponible y en la tercera columna la utilizada para las estimaciones. La diferencia entre la columna tres y dos obedece a los filtros y vacíos de información ya sea por parte de C-600 o C-100<sup>14</sup>. Los filtros aplicados a la información consisten en tomar sólo los colegios en jornada de la mañana, tarde o completa, es decir, excluir de la estimación los colegios de jornadas nocturna y fines de semana dado que obedecen a un mercado relevante diferente, donde los estudiantes por lo general trabajan, presentan rezago escolar y, donde sus incentivos y actitudes respecto a la educación son distintos. Además se aplicaron algunos filtros con el fin de eliminar observaciones atípicas e información faltante, estos fueron: no considerar los colegios con una proporción profesor-alumno mayor a 0.5 (un profesor por cada dos alumnos) o igual a 0 (ningún profesor por alumno) pues posiblemente son errores de los datos, y tampoco se incluyeron los colegios para los cuales no se cuenta con información para todos los controles usados en las estimaciones. Por otra parte trató de controlarse por otras políticas públicas que pudieran alterar el impacto como colegios en concesión, o los *megacolegios* que se han construido en la ciudad; sin embargo, según el cruce de información, los colegios que eran utilizados en las estimaciones no pertenecían a ninguna de estas políticas educativas.

Por último, un filtro que se realizó con el fin de mejorar la estrategia de identificación fue incluir en la estimación únicamente los colegios de Bogotá que se encuentran hasta 3500 mil metros de distancia de las bibliotecas del Tintal y del Tunal, la razón para acotar la distancia es que si se incluyen colegios muy distantes de las bibliotecas se perderá la homogeneidad del entorno que había entre los cercanos y los no cercanos a estas. Al perderse esta homogeneidad (no observable) se entraría en sesgos de estimación debido a que no se

---

<sup>12</sup>La información sobre la localización espacial de cada colegio proviene de la Secretaría de educación Distrital

<sup>13</sup>En el caso de instituciones públicas un colegio es considerado como la combinación sede-jornada

<sup>14</sup>En los casos que fue posible se imputaron los datos utilizando el método de hot deck (véase Daza y Escobar, 2010) que se basa en la metodología de *matching* de Ñopo (2008)

tendría un contrafactual adecuado, esto se explicara con mayor detalle más adelante en la sección 4.1.

### **3.4. Un mirada desde los datos a los determinantes de la calidad de la educación**

En los cuadros 3 y 4 se toman solo el universo de colegios que se usa en la estimación, como se explico antes son los colegios de Bogotá que se encuentran hasta 3.5 kilómetros alrededor de las bibliotecas según muestra en la figura 4. En estas tablas se muestra el valor promedio del puntaje (estandarizado) de los colegios de acuerdo con las características de estos.

Al mirar el cuadro 3 se tiene que los colegios de mayor tamaño (medido por medio de la matricula del colegio) suelen tener mejores resultados, esto puede obedecer que aquellas instituciones que tienen más alumnos pueden adquirir herramientas de estudio más costosas cuya lógica obedece a economías de escala (laboratorios modernos, profesores con altos niveles de educación, bibliotecas privadas, entre otros) mientras que aquellos colegios de pocos estudiantes no pueden hacerlo. Por otra parte, la razón de profesores por alumno muestra que las instituciones que reportan una mayor valor presentan mejores puntajes, esto puede explicarse por el hecho de que mientras más personalizado sea el proceso de enseñanza mayor es la capacidad de atención del profesor a falencias individuales o de grupos pequeños. No obstante, tenemos que tener cuidado al interpretar estos datos: por ejemplo, cuando la proporción de profesores con nivel de educación de secundaria es menor la calidad parece aumentar, mientras que para el caso de la proporción de profesores con los más altos niveles de educación (posgrado), y de una forma aparentemente contraintuitiva, los mejores resultados se obtienen cuando esta proporción de profesores es baja. Lo anterior se explica porque los profesores con mayores niveles de educación se encuentran principalmente en colegios públicos, que generalmente tienen un desempeño inferior (Núñez, Stainer, Cadena y Pardo, 2002)<sup>15</sup>. Por último, si se observa el porcentaje de niñas en el colegio se tiene que las instituciones en las que esta proporción es muy alta o muy baja se dan mejores puntajes que en las que esta proporción tiene un puntaje intermedio. En promedio, los colegios que tienen una proporción de niñas entre 0.48 y 0.85 son los que tienen peor desempeño.

---

<sup>15</sup>En este caso, especialmente por selección de estudiantes.

En el primer panel de cuadro 4 se muestra una relación favorable para los estudiantes con jornada completa respecto a las demás jornadas, siguiéndole la jornada de mañana y por último la de la tarde. En el segundo panel de este mismo cuadro se presenta la diferencia en los puntajes entre los colegios públicos y privados, según se observa para todos los años las instituciones privadas tiene mejor desempeño. Por otra parte, si se fragmentan espacialmente los puntajes, se tiene que existe una trayectoria similar año a año en los resultados por localidades y los mejores puntajes se obtuvieron en la localidad de Fontibón, Antonio Nariño y Puente Aranda<sup>16</sup>, mientras los peores se dieron en Ciudad Bolívar, Usme y Bosa<sup>17</sup>.

El cuadro 5 presenta los resultados promedio (estandarizados) de los colegios que se encuentran a tres diferentes rangos de distancia: los colegios que están a menos de 1km, colegios entre 1km y 2.5km, y colegios que está a más de 2.5km de la biblioteca. También se tienen 3 grupos de años, el primero es de 2000 a 2002 (antes de las bibliotecas), el segundo entre 2003 y 2005, y el último entre 2006 y 2008. Para el primer grupo de años (2000-2002) los colegios que están más lejos de la biblioteca son los que tienen mejores resultados, mientras que para los dos siguientes grupos de años 2003-2005 y 2006-2008 este hecho tiende a cambiar pues son los colegios que están más cerca los que se desempeñan mejor. La figura 3 nos da otra perspectiva de esta relación entre distancia y puntaje esta vez comparando el año 2002 con el 2008, donde no hay un claro patrón ni antes ni después de la implementación de las bibliotecas.

Después de observar en los datos las relaciones entre algunas características del plantel educativo y la calidad de la educación, y considerando el impacto causal que la literatura atribuye a estas características, es prudente identificar si la localización de las bibliotecas presenta alguna discrecionalidad o es totalmente aleatoria en términos del tipo de colegios que rodean el lugar donde estas se construyeron. El cuadro 6 intenta responder a esta pregunta calculando el tipo de colegio promedio que se encuentra a diferentes rangos de distancia respecto a la biblioteca más cercana, como se observa en este, los colegios más cercanos tienen una mayor probabilidad de ser oficiales mientras que a medida que nos

---

<sup>16</sup>debido a que los datos están filtrados por aquellos que se encuentren hasta 3.5 kilómetros alrededor de las bibliotecas, lo anterior no quiere decir que las localidades mencionadas arriba sean las mejores de Bogotá

<sup>17</sup>En las tablas anteriores no se mostraron diferencias de resultados por calendario del colegio ya que todos los colegios en el rango utilizado alrededor de las *megabibliotecas* tienen calendario A

alejamos de las *megabibliotecas* la proporción de este tipo de colegios va disminuyendo, esto implica que si no se controlara por tipo de colegio el efecto de las bibliotecas estaría siendo subestimado debido a que el desempeño promedio de los colegios públicos es menor (Nuñez, 2002; Gaviria y Barrientos, 2001). Otra variable importante del colegio es su género (mixto, masculino o femenino), si se reagrupa esta variable en dos categorías: los que tienen un sólo género (masculino o femenino), y los que tienen ambos; se observa como todos los colegios que se encuentran a menos de 1 Km de las bibliotecas pertenecen a colegios mixtos. Adicionalmente, este cuadro presenta las variables asociadas con la cantidad y la calidad del cuerpo docente por medio de la relación profesores-alumnos y la proporción de profesores con nivel de educación de posgrado respectivamente. Según se aprecia, la probabilidad de que los colegios cuenten con mayor proporción de profesores con alta educación aumenta mientras más cerca estén de las bibliotecas (resultado influenciado por la concentración de colegios públicos alrededor de las bibliotecas, hacia los cuales migra la mano de obra más calificada), mientras que sucede lo contrario con la relación de profesores por alumno. En términos de la jornada y el tamaño de los colegios la distribución es similar en todas las distancias. En síntesis, según el cuadro 6 existe evidencia de una posible selección implícita del programa de las *megabibliotecas*, donde los colegios cercanos a estas presentan características más desfavorables en términos de factores que impactan la calidad de la educación.

### 3.5. Exploración inicial

Una primera aproximación al impacto de la implementación de las bibliotecas y el puntaje es explorar la relación puntaje-distancia tras descontar la variación de determinantes comunes del puntaje. Para ello recurrimos a un modelo semiparamétrico clásico, una regresión lineal parcial que nos permite ver la relación de una forma no lineal<sup>18</sup>

$$Y = m(d) + X\beta + u \quad (1)$$

Donde  $Y$  es el puntaje,  $X$  son los controles,  $u$  es un error tal que  $E[u|d, X] = 0$ . La figura 7 nos muestra el valor ajustado para  $m(\hat{d}) + u$ , que nos da la relación entre el

---

<sup>18</sup>La estimación se realizó siguiendo el algoritmo de diferencias de Yatchew (1997), implementado por Loskshin (2006).

puntaje y la distancia descontando la variación por controles usuales. Podemos encontrar una relación similar a una U, donde el mínimo está centrado cerca a los 1500 metros. Debido a ello nuestro análisis será especialmente cuidadoso entre los 750 y 2000 metros, que podría indicarnos que el impacto de las bibliotecas llega hasta esa zona. No obstante, estas gráficas son solo para explorar la relación dado que incluyen determinantes no observados  $u$ ; de hecho el patrón tipo U se encuentra tanto antes y después de 2002<sup>19</sup>.

Para estimar el efecto tendremos que asumir que las variables no observables pueden variar a través de la distancia, pero las variaciones en el tiempo de dichas variables no deben variar con la distancia. Esta restricción nos permitirá identificar el impacto promedio sobre los colegios ‘ceranos’ respecto a los ‘lejanos’ y es la motivación para utilizar la estrategia de Diferencia en Diferencias, como veremos en la siguiente sección.

## 4. Estrategia de identificación

La estrategia de identificación involucra tres etapas de decisión: la primera se refiere a cuáles son los colegios y/o estudiantes que efectivamente se ven afectados por la creación y funcionamiento de las *megabibliotecas*; la segunda hace referencia a la estrategia de identificación para medir la magnitud y significancia del impacto que las *megabibliotecas* tienen sobre estos colegios y/o estudiantes, y la última etapa presenta una manera de descomponer el impacto estimado en la segunda etapa.

### 4.1. El papel de la distancia en la incidencia del efecto vecindario de las *megabibliotecas*

El objetivo de este estudio es identificar el efecto causal de las bibliotecas públicas en la calidad de la educación. Dado que no es posible contar con información sobre cuáles estudiantes de los que presentan la prueba SABER 11 hacen uso de las bibliotecas, se propone utilizar como variable alternativa la distancia. Ello bajo el supuesto de que el uso de las bibliotecas es mayor por quienes viven mas cerca a estas, dado que quienes viven lejos enfrentan mayores costos de desplazamiento que reducen sus incentivos a visitarlas de

---

<sup>19</sup>Lo ‘deseable’ sería que este estimador diera una línea horizontal antes de la implementación, y luego una función decreciente con la distancia. Aún si  $m(d)$  fuese así, la no posibilidad de observar  $u$  no permitiría que viéramos dicho patrón.

manera frecuente. Por otra parte, como se explico antes, las bibliotecas pretenden generar un impacto en los hábitos del vecindario que las rodea por lo cual la cercanía a estas también se vuelve clave para identificar en su totalidad la interacción entre los estudiantes y las bibliotecas públicas. Según el cuadro 12, el 77% de los estudiantes de Bogotá viven a menos de 20 minutos de los establecimientos educativos a los que asisten, por lo que la distancia del colegio a la biblioteca aproxima la distancia de la biblioteca al hogar y por ende a la mayor probabilidad de vivir en un entorno afectado por las bibliotecas.

Con el fin de realizar una inferencia causal necesitamos un grupo de comparación o ‘contrafactual’, es decir, colegios que no reciben ningún impacto por la creación de las bibliotecas pero que sean lo más parecido posible a los que si reciben el impacto de estas, en este punto la distancia juega nuevamente un aspecto clave, se considerara como ‘tratados’ a los colegios que están ‘cerca’ a una de las dos bibliotecas y como ‘controles’ a los que están lo suficientemente ‘lejos’ de estas, de tal forma que sea muy poco probable que experimenten impacto alguno, pero lo suficientemente ‘cerca’ de los vecindarios de los considerados como ‘tratados’, de forma que sus vecindarios tengan características socio-económicas y urbanísticas similares a los de estos. Por esta razón después de repetidos análisis se tomo la decisión de acotar hasta 3.5 Km el universo de colegios que se consideran en la estimación.

Evidentemente el ‘cerca’ o ‘lejos’ es arbitrario y requiere de una norma de selección que es parte de la pregunta de investigación. Se consideraron una alternativa discreta y otra continua para definir la exposición al tratamiento utilizando la distancia de cada colegio a la *megabiblioteca* más cercana,  $d$ .

**Alternativa Discreta** La alternativa discreta es una definición dicótoma que separa a los colegios entre tratados  $T = 1$  y no tratados o controles  $T = 0$  por la biblioteca. Dos versiones de la alternativa discreta son presentadas a continuación<sup>20</sup>. Esta es una aproximación no estructural a la estimación del efecto causal.

1. Asignar  $T = 1$  a los colegios tales que  $d \in [0, R1]$ , y  $T = 0$  a aquellos que  $d \in (R1, R2)$ .

Ver la especificación I de la figura 5.

---

<sup>20</sup>Otra alternativa evaluada fue similar a la primera aquí presentada pero definiendo una ‘zona gris’ entre colegios tratados y no tratados. Los colegios de dicha zona (500 metros) no eran incluidos. Los resultados no varían considerablemente frente a los aquí presentados.

2. Asignar  $T = 1$  a los colegios tales que  $d \in [0, R1]$ , y  $T = 0$  a aquellos que  $d \in (R3, R2)$ , donde  $R3 \in [R1, R2]$ . Ver la especificación II de la figura 5.

Si bien las dos estrategias siguen el mismo principio, en la primera tanto el grupo de tratamiento como el de control se modifican al variar  $R1$  mientras que en la segunda siempre comparamos contra un mismo grupo de control, so pena de excluir algunos colegios de la estimación. Al presentar ambos resultados queremos mostrar que los mismos no están dados por la particularidad de la técnica. El procedimiento consistió en realizar múltiples estimaciones del estimador teniendo en cuenta diferentes valores para  $R1, R2, R3$ . Debido a la cantidad de posibles combinaciones, los resultados que se presentarán definen a  $R2 = 3500$  y a  $R3 = 2000$  mientras que  $R1 \in 750, 1000, 1250, 1500, 1750, 2000$ . Estos resultados son un buen resumen de los hallazgos de la estimación y se centran en la zona de mayor variación que encontramos con la regresión lineal parcial descrita anteriormente. Por otra parte, radios  $R1 < 750$  determinan una muestra muy pequeña de colegios tratados.

**Alternativa Continua** Una alternativa más sencilla es especificar una forma funcional al impacto que tienen las bibliotecas a través de la distancia.

1. La version más sencilla define la variable de intensidad del tratamiento  $T$  que experimenta un colegio como  $T = 1 - \frac{d}{R1}$  si  $d \leq R1$  y  $T = 0$  si  $d \geq R1$ , incluyendo los colegios hasta  $R2 \geq R1$ . Los resultados se presentan para  $R1 \in [1500, 3500]$  con  $R2 = 3500$  como en el caso discreto. Ver la especificación I de la figura 6.
2. Otra alternativa, siguiendo el patrón visto en la regresión lineal parcial, es asumir que el efecto decae de forma inversa con la distancia. Para ello definimos  $T = \frac{R1}{d} - 1$  si  $d \leq R1$  y  $T = 0$  si  $d \geq R1$ , incluyendo los colegios hasta  $R2 \geq R1$ . Ver la especificación II de la figura 6.

## 4.2. Estimación del impacto de las bibliotecas

Una vez se ha definido la forma de identificación del grupo de tratamiento o conjunto de colegios y estudiantes sobre los que las bibliotecas tienen incidencia, se procede a definir la técnica que permita realizar una estimación adecuada del impacto. Se propone estimar dos especificaciones diferentes para analizar la forma en la que las bibliotecas impactan el

desempeño escolar. La primera es la especificación tradicional del modelo que se denominará DDT, esta evalúa únicamente el efecto nivel de las bibliotecas, es decir, asume que las *megabibliotecas* generan un impacto sobre los colegios independiente de las características de los mismos (efecto nivel). La segunda especificación calcula un impacto agregado que es la suma del efecto nivel y la interacción de las bibliotecas con las características de los colegios que se denominará DD-BO.

#### 4.2.1. DDT: Diferencias a nivel

Considerando que es posible que los vecindarios donde las bibliotecas operan sean distintos a los vecindarios control, la técnica de estimación debe capturar estas diferencias iniciales que no obedecen a la construcción de las bibliotecas. Para esto, la técnica mas apropiada es la estimación por medio de DD. En ella se estima la diferencia entre ‘tratados’ y ‘controles’ después de implementar el programa de las *megabibliotecas* y a esta se le descuenta la diferencia entre estos dos grupos que había antes de la creación de las bibliotecas y que obedece en parte a las diferencias en variables no observables del vecindario. Esta estimación asume que ambos tipos de colegios hubiesen tenido una evolución similar si el programa nunca hubiese sido implementado. Para un colegio  $i$  en el año  $t$ , la variable dependiente (el puntaje estandarizado de SABER 11) es  $Y_{it}$ ,  $A_{it}$  es la variable indicadora del periodo posterior a la implementación (toma el valor de 1 para el periodo 2003-2008 y el valor de 0 para antes de 2003),  $T_{it}$  es el estado del tratamiento (como vimos, puede ser discreto o continuo),  $\gamma_i$  y  $\gamma_t$  son efectos fijos por colegio y año, y  $X_{it}$  es la matriz de los controles (jornada, naturaleza del colegio, matricula, etc.). El efecto causal sobre los tratados  $\delta_{ATT}$  puede estimarse por medio de la siguiente regresión de mínimos cuadrados ordinarios<sup>21</sup>.

$$Y_{it} = \delta_{ATT}T_{it} \cdot A_{it} + \beta_1T_{it} + \beta_2A_{it} + \beta X_{it} + \gamma_i + \gamma_t + e_{it} \quad (2)$$

---

<sup>21</sup>Como vimos anteriormente, hay diferencias entre los colegios tratados y los no tratados, por lo que no es de esperar que el efecto sea homogéneo. Por ello lo más seguro es que el efecto sea específico a los tratados.

#### 4.2.2. DD-BO: Descomposición del efecto de las *megabibliotecas*

Al implementar el modelo de DDT obtenemos un efecto nivel de la biblioteca sobre los colegios cercanos, sin embargo es válido pensar que las bibliotecas además de generar un efecto sobre el puntaje, independiente de las características de los colegios, pueden también generar un efecto a través otros factores educativos. Por ejemplo, se puede pensar que los impactos que generan una alta razón de docentes por alumno se potencian con las bibliotecas dado que los profesores encuentran en estas un medio para complementar la educación que brindan a los estudiantes. Otro efecto interactivo que se puede dar es que la cercanía de los colegios a las bibliotecas reduzca la brecha entre colegios públicos y privado debido a que los colegios públicos pueden usar mas intensivamente estas dado su carencia en activos educativos como los libros y sistemas de información.

La identificación de estos efectos interactivos sumada al efecto nivel es planteada por la descomposición de Blinder-Oaxaca (BO), (Blinder y Oaxaca, 1973). Esta permite descomponer la diferencia que existe entre dos grupos de colegios (aquellos que se encuentran cerca a la biblioteca y quienes se encuentran lejos) en dos partes: el primer componente  $-\Delta_X$ - define la brecha como respuesta a distintas dotaciones de los factores educativos entre ambos grupos, por ejemplo, las características socio-económicas de los estudiantes, proporción de colegios privados, principal tipo de jornada en que estudian, entre otros. El segundo componente,  $-\Delta_0$ -, explica la brecha como producto de variables no observables.

Para llevar el modelo BO al contexto dinámico de DD, realizamos la descomposición en ambos momentos de tiempo. De esta forma podemos aislar en  $\Delta_X$ , la variación del  $Y$  debido a la diferencia en las diferencias de los controles antes y después de las bibliotecas; y en  $\Delta_0$ , que sería el impacto de las *megabibliotecas* sobre  $Y$  no explicado por variaciones en los controles pero sí por posibles variaciones en la relación de ellos con el puntaje debido a las bibliotecas. Para su estimación utilizamos la siguiente regresión lineal<sup>22</sup>.

$$Y_{it} = \alpha_0 + \alpha_1 T_{it} + \alpha_2 A_{it} + \alpha_3 X + \alpha_4 X_{it} \cdot T_{it} + \alpha_5 X_{it} \cdot A_{it} + \alpha_6 T_{it} \cdot A + \alpha_7 X \cdot T_{it} \cdot A_{it} + u$$

<sup>22</sup>Por simplicidad de la notación se omiten los subíndices. Por otra parte, sólo se incluyen efectos fijos por año en esta especificación dado que el gran número de interacciones resultantes al utilizar efectos fijos por colegio llega a limitar las posibilidades de estimación.

De ahí, tenemos que el impacto de las bibliotecas  $\delta$  sobre el desempeño estaría dado por:

$$\begin{aligned}\delta &= (E[y|T = 1, A = 1] - E[y|T = 0, A = 1]) - (E[y|T = 1, A = 0] - E[y|T = 0, A = 0]) \\ \delta &= \Delta_0 + \Delta_x \\ \delta &= \alpha_6 + (\alpha_4 + \alpha_5 + \alpha_7)E[X|T = 1, A = 1] - \alpha_5E[X|T = 0, A = 1] - \alpha_4E[X|T = 1, A = 0] \\ &\quad + \alpha_3[(E[X|T = 1, A = 1] - E[X|T = 0, A = 1]) - (E[X|T = 1, A = 0] - E[X|T = 0, A = 0])]\end{aligned}$$

Los errores son calculados mediante Bootstrapping. Para más detalles sobre el estimador, favor referirse al anexo.

## 5. Resultados

Las tablas 7 y 8 presentan los resultados para la estimación discreta DDT en sus dos especificaciones. Se utiliza la información de 2000 a 2002 y el año especificado en cada columna. Encontramos que el estimador tiene signo positivo en casi todos los casos pero no es significativo en la mayor parte de ellos. Tal vez en el único caso en el que se encuentra una diferencia significativa para diferentes variantes de los radios es el año 2008, donde en ambas especificaciones encontramos un efecto que decae con la distancia, cercano a 0.1 desviaciones estándar (DE) a los 1250 metros.

La tabla 9 muestra los resultados para la descomposición BO-DD bajo la especificación I de DD. Encontramos que generalmente la variación en el puntaje proveniente de la variación en los controles representada por  $\Delta_X$  tiende a ser negativa, lo que nos dice que el impacto de las bibliotecas públicas, que hace parte de  $\Delta_0$ , puede ser cercano a 0.2 DE. No obstante, si bien los signos son positivos, tampoco hay resultados significativos.

Las tablas 10 y 11 presentan los resultados de las especificaciones estructurales del efecto. Los resultados son similares a los discretos: efectos positivos pero no significativos.

Los resultados son en esencia un posible impacto de signo positivo, con un patrón decreciente en la distancia, pero no significativo. Por lo tanto, existe evidencia de una posible mejoría en el desempeño escolar debido a las bibliotecas, pero no es posible descartar con un alto nivel de certeza que no haya efecto alguno.

La no significancia de los resultados puede tener varias razones, entre ellas dos a destacar. La primera es sencillamente el número de observaciones: a nivel de colegio son pocos los colegios tratados por lo que los estimadores son altamente ineficientes. La segunda es la posible mezcla de efectos heterogéneos: algunos colegios pueden tener mecanismos que utilicen las bibliotecas directamente mientras otros no, por lo que al promediar el resultado no es lo suficientemente grande como para ser significativo. Desafortunadamente hasta el momento no ha sido posible obtener información sobre esta clase de prácticas para cada colegio y biblioteca. Explorar el impacto de esta clase de prácticas, más allá del análisis de datos agregados, puede enriquecer enormemente la literatura relacionada con intervenciones que puedan mejorar la calidad de la educación al futuro.

Una alternativa a nuestra estimación actual sería utilizar la información de residencia de los estudiantes para tratar de asociar el efecto a la distancia de la vivienda. Ello nos permitiría obtener un mayor número de observaciones y conseguir mayor varianza para identificar el efecto, aunque con el costo de incluir una mayor heterogeneidad (la demanda) para la cuál quizás no tengamos los suficientes controles para explicar. Este proyecto hace parte de nuestra agenda de investigación.

Un tema final a discutir es la posible existencia de patrones que no cumplan el supuesto de tendencias paralelas en las no observables que soportan el DD. La construcción de las bibliotecas pudo mejorar las condiciones de vida de la zona, además de estar en zonas aledañas al sistema de transporte Transmilenio que se fue desarrollando durante los años analizados. Este fenómeno pudo atraer a ciudadanos de mayor ingreso a las cercanías de las bibliotecas, que a su vez están más preocupados por la calidad de la educación de sus hijos. Si bien, con la información actual no es posible aislar esta posible explicación alterna, los resultados encontrados a una distancias relativamente corta (menos de dos kilómetros) nos dicen que posiblemente esta explicación alterna no sea lo suficientemente creíble como para minar el signo positivo encontrado.

Por otra parte, es posible pensar en que las bibliotecas no afectan fuertemente resultados asociados a exámenes específicos sino que impactan aptitudes y actitudes hacia el estudio que pueden ser útiles también. Por ejemplo, puede impactar las aspiraciones de los estudiantes sobre nivel educativo y aumentar la precision de sus creencias sobre salarios y retornos esperados a la educación superior. Sobre ello la literatura aún guarda silencio.

Un último tema a tener en cuenta es la calidad de los datos de control. Se encontraron varios detalles en las bases C-600, y la no continuidad de la encuesta C-100 dificultan la tarea de la investigación relacionada con educación en Colombia.

## 6. Conclusiones

La distribución de las dotaciones de factores educativos alrededor de las bibliotecas deja entrever un criterio de selección implícita en la localización de la *megabibliotecas*, es decir, si bien la política pública jamás definió un perfil específico de colegios para construir las *megabibliotecas*, el escoger lugares periféricos de la ciudad llevo a que se seleccionara indirectamente cierto tipo de instituciones educativas. Estas son instituciones que tienen bajas dotaciones de factores educativos que la literatura ha identificado como determinantes que generan un mejor desempeño. Por tanto, de existir un efecto positivo de las *megabibliotecas* en el desempeño educativo, estas serían políticas progresivas dado que permitirían reducir la desigualdad de oportunidades en los activos educativos que se presenta en Bogotá.

Las estimaciones planteadas a nivel de colegio indican que no hay efectos notorios de las bibliotecas públicas sobre el desempeño de los colegios alrededor de estas, ya sea utilizando una estimación discreta o continua, el efecto es positivo pero no significativo. Por otra parte, aún cuando no hay resultados significativos es importante destacar que el efecto para un mismo año decrece con la distancia, lo cual muestra que quizás sí haya un impacto pero no sobre todos los estudiantes.

## 7. Referencias

- Abdul Hamid, H. (2003). What Jordan needs to do to prepare for the knowledge economy: Lessons learned from TIMSS-R. University of Maryland (processed).
- Attanasio, O., E. Fitzsimons, A. Gomez, D. Lopez, C. Meghir, A. Mesnard, and C. for Economic Policy Research (Great Britain) (2006). Child education and work choices in the presence of a conditional cash transfer programme in rural Colombia. Centre for Economic Policy Research.
- Attanasio, O. and M. Vera-Hernández (2004). Medium-and long run effects of nutrition and child care: evaluation of a community nursery programme in rural Colombia.
- Barro, R. (2001). Human capital and growth. The American Economic Review 91(2), 12–17.
- Becker, S., C. CES, and I. Siebern-Thomas (2007). Schooling infrastructure, educational attainment and earnings.
- Blinder, A. (1973). Wage discrimination: reduced form and structural estimates. Journal of Human Resources, 436–455.
- Blomeyer, D., K. Coneus, M. Laucht, and F. Pfeiffer (2009). Initial risk matrix, home resources, ability development, and children’s achievement. Journal of the European Economic Association 7(2-3), 638–648.
- Blundell, R. and M. Dias (2009). Alternative approaches to evaluation in empirical microeconomics. Journal of Human Resources 44(3), 565–640.
- Caro, B. (2000). Factores asociados al logro académico de los alumnos de 3 y 5 de primaria en Bogotá. Coyuntura Social, 65–80.
- Casas, A., L. Gamboa, and L. Piñeros (2002). El valor que agrega la escuela: una aproximación a la calidad de la educación en Colombia. Borradores de investigación.
- CVN (2008, Jun). Comienzan vacaciones robóticas en biblioteca el tinal. <http://www.mineducacion.gov.co/cvn/1665/article-164284.html>.
- Duflo, E. (2000). Schooling and labor market consequences of school construction in Indonesia: Evidence from an unusual policy experiment.
- Durlauf, S. (2004). Neighborhood effects. Handbook of regional and urban economics 4, 2173–2242.
- El Tiempo (2009, Dic). Las megabibliotecas públicas de la ciudad, todo un plan para estas vacaciones.

- Fertig, M., C. Schmidt, and C. for Economic Policy Research (2002). The role of background factors for reading literacy: Straight national scores in the PISA 2000 study. Centre for Economic Policy Research.
- Fuchs, T. and L. Woessmann (2007). What accounts for international differences in student performance? a re-examination using pisa data. Empirical Economics 32(2), 433–464.
- Gamboa, L., M. Rodríguez-Acosta, and A. García-Suaza (2010). Academic achievement in sciences: the role of preferences and educative assets. DOCUMENTOS DE TRABAJO.
- Giraldo Giraldo, Y., G. Román Betancur, and R. Quiroz Posada (2009). The public library as educational environment for creating citizenship: a study in the "comuna 1" in medellín. Rev. Interam. Bibliot 32(1), 47–84.
- Hanushek, E. and D. Kimko (2000). Schooling, labor-force quality, and the growth of nations. The American Economic Review 90(5), 1184–1208.
- Hanushek, E. and L. Woessmann (2007). The role of education quality in economic growth. World.
- Heckman, J. J., H. Ichimura, and P. E. Todd (1997, October). Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme. Review of Economic Studies 64(4), 605–54.
- Iregui, A., B. Ligia Melo, and J. Ramos (2006). Evaluación y análisis de eficiencia de la educación en colombia. Borradores de Economía.
- Iregui, A., B. Ligia Melo, and J. Ramos (2010). La educación en colombia: análisis del marco normativo y de los indicadores sectoriales. Revista de Economía del Rosario (2).
- Jencks, C. and S. Mayer (1990). The social consequences of growing up in a poor neighborhood. Inner-city poverty in the United States, 111.
- Lance, K. (1994). The impact of school library media centers on academic achievement. School Library Media Quarterly 22(3), 167–70.
- Lance, K., M. Rodney, and C. Hamilton-Pennell (2000). Measuring up to standards: The impact of school library programs & information literacy in pennsylvania schools.
- Leventhal, T. and J. Brooks-Gunn (2000). The neighborhoods they live in: the effects of neighborhood residence on child and adolescent outcomes. Psychological Bulletin 126(2), 309.
- Lindauer, B. (1998). Defining and measuring the library's impact on campuswide outcomes. College & Research Libraries 59(6), 546.

- Lokshin, M. (2006). Semi-parametric difference-based estimation of partial linear regression models. Stata Journal 6(3), 377–383.
- Lonsdale, M. (2003). Impact of School Libraries on Student Achievement: A Review of the Research. For full text: <http://www.asla.org.au/research/>.
- Melo, L. (2005). Impacto de la descentralización fiscal sobre la educación pública colombiana. Borradores de economía.
- Murnane, R., R. Maynard, and J. Ohls (1981). Home resources and children's achievement. The Review of Economics and Statistics 63(3), 369–377.
- Núñez, J., R. Steiner, X. Cadena, and R. Pardo (2002). cuáles colegios ofrecen mejor educación en colombia? Archivos de Economía 193.
- Oaxaca, R. (1973). Male-female wage differentials in urban labor markets. International economic review 14(3), 693–709.
- Pal, S. (2010). Public infrastructure, location of private schools and primary school attainment in an emerging economy. Economics of Education Review 29(5), 783–794.
- Rodney, M., K. Lance, C. Hamilton-Pennell, and M. Center (2002). Make the connection: Quality school library media programs impact academic achievement in Iowa. Mississippi Bend Area Education Agency.
- Rubin, D. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. Journal of educational Psychology 66(5), 688.
- Sánchez, F. (2006). Descentralización y progreso en el acceso a los servicios sociales de educación, salud y agua y alcantarillado. Documento cede 15.
- Sarmiento, A., L. Becerra, and J. González (2000). La incidencia del plantel en el logro educativo del alumno y su relación con el nivel socioeconómico. Coyuntura Social 22, 53–64.
- Smith, E. (2001). Texas school libraries: Standards, resources, services, and students' performance. EGS Research & Consulting.
- Uribe, C., R. Murnane, J. Willett, and M. Somers (2005). Expanding school enrollment by subsidizing private schools: lessons from bogotá.
- Vegas, E. and J. Petrow (2008). Raising student learning in Latin America: The challenge for the 21st century. World Bank Publications.
- Williams, D., C. Wavell, and L. Coles (2001). Impact of school library services on achievement and learning. Aberdeen: Robert Gordon University.

World Bank (2005). Mexico: Determinants of learning policy note.

Yatchew, A. (1997). An elementary estimator of the partial linear model. Economics Letters 57(2), 135–143.

## A. Cuadros

### A.1. Número de Colegios y estudiantes según distancia

Cuadro 1: Colegios por distancias

<b>Distancia desde la biblioteca</b>	<b>Todos los colegios</b>	<b>Usados en la estimación</b>
0-500m	5	4
500m-1000m	15	11
1000m-1500m	28	27
1500m-2000m	30	24
2000m-2500m	48	40
2500m-3000m	45	39
3000m-3500m	45	38
3500m-4000m	59	49
Total	275	232

Fuente: Calculos propios

Cuadro 2: Estudiantes por distancias

<b>Distancia desde la biblioteca</b>	<b>Todos los colegios</b>	<b>Usados en la estimación</b>
0-500m	237	115
500m-1000m	2996	2888
1000m-1500m	5372	5178
1500m-2000m	5229	4820
2000m-2500m	6322	5629
2500m-3000m	7634	7032
3000m-3500m	6263	6086
3500m-4000m	7685	7195
Total	41738	38943

Fuente: Calculos propios

## A.2. Estadísticas Descriptivas

Cuadro 3: Puntaje total promedio según características de infraestructura y planta docente

	Año								
	2000 Prom	2001 Prom	2003 Prom	2004 Prom	2005 Prom	2006 Prom	2007 Prom	2008 Prom	Total Prom
<b>Estudiantes</b>									
Menor a 300	-0.26	-0.52	-0.44	-0.42	-0.41	-0.33	-0.41	-0.29	-0.38
Entre 300-600	-0.22	-0.36	-0.15	-0.21	-0.13	-0.19	-0.26	-0.16	-0.21
Entre 600-1000	-0.02	-0.15	-0.03	-0.14	-0.19	-0.17	-0.05	-0.14	-0.12
Mayor a 1000	0.13	-0.01	-0.12	-0.09	-0.10	-0.16	-0.16	-0.18	-0.10
<b>Total</b>	-0.06	-0.21	-0.15	-0.18	-0.19	-0.20	-0.20	-0.19	-0.17
<b>Propor. prof. alumnos</b>									
Menor a .03	-0.08	-0.57	-0.32	-0.29	-0.31	-0.34	-0.25	-0.28	-0.31
Entre .03-.04	-0.06	-0.16	0.01	-0.19	-0.21	-0.25	-0.20	-0.27	-0.18
Entre .04-.05	-0.11	-0.21	-0.00	-0.18	-0.00	-0.04	-0.17	0.01	-0.11
Entre .05-.06	0.01	-0.10	-0.46	-0.03	-0.12	-0.19	-0.19	-0.27	-0.13
Más que .06	-0.21	-0.45	-0.37	-0.35	-0.30	-0.34	-0.40	-0.36	-0.36
<b>Total</b>	-0.08	-0.23	-0.17	-0.22	-0.20	-0.23	-0.25	-0.23	-0.20
<b>Propor. niñas alumnos</b>									
Menor a 0.15	0.15	0.41	0.29	0.11	0.13	0.11	0.03	-0.03	0.15
Entre 0.15-0.43	0.10	-0.06	-0.08	-0.11	-0.11	-0.15	-0.28	-0.15	-0.11
Entre 0.43-0.48	-0.11	-0.29	-0.19	-0.19	-0.13	-0.18	-0.06	-0.20	-0.17
Entre 0.48-0.52	-0.20	-0.40	-0.31	-0.30	-0.36	-0.37	-0.40	-0.33	-0.34
Entre 0.52-0.85	-0.22	-0.36	-0.18	-0.39	-0.23	-0.28	-0.34	-0.42	-0.30
Mayor a 0.85	0.24	0.26	0.33	0.13	0.09	0.01	0.03	0.16	0.16
<b>Total</b>	-0.08	-0.23	-0.17	-0.22	-0.20	-0.23	-0.25	-0.23	-0.20
<b>Prof. nivel básico</b>									
Menor a .25	-0.03	-0.20	-0.11	-0.16	-0.19	-0.22	-0.20	-0.21	-0.17
Entre .25-.5	-0.26	-0.35	-0.43	-0.47	-0.13	-0.38	-0.36	-0.21	-0.33
Entre .5-.75	0.21	-0.03		-0.22	-0.36				-0.05
Mayor a .75	-0.31	-0.83	-0.66	-0.52	-0.76	-0.59		-0.90	-0.61
<b>Total</b>	-0.06	-0.22	-0.15	-0.20	-0.19	-0.23	-0.21	-0.21	-0.19
<b>Prof. nivel superior</b>									
Menor a .25	-0.08	-0.25	-0.19	-0.24	-0.14	-0.18	-0.17	-0.17	-0.18
Entre .25-.5	-0.16	-0.37	-0.15	-0.55	-0.17	-0.18	-0.81	-0.26	-0.30
Entre .5-.75	-0.12	-0.28	-0.19	-0.27	-0.33	-0.32	-0.41	-0.34	-0.29
Mayor a .75	-0.04	-0.20	-0.08	-0.18	-0.11	-0.30	-0.24	-0.32	-0.19
<b>Total</b>	-0.09	-0.26	-0.18	-0.24	-0.18	-0.21	-0.23	-0.22	-0.20

Cuadro 4: Puntaje total promedio según características institucionales y de entorno

	Año								Total prom
	2000 prom	2001 prom	2003 prom	2004 prom	2005 prom	2006 prom	2007 prom	2008 prom	
<b>Jornada</b>									
Completa	0.040	-0.075	-0.017	-0.046	0.020	0.016	0.030	0.051	0.002
Mañana	-0.065	-0.209	-0.193	-0.288	-0.251	-0.313	-0.394	-0.365	-0.263
Tarde	-0.252	-0.451	-0.356	-0.362	-0.420	-0.474	-0.460	-0.475	-0.408
<b>Total</b>	<b>-0.082</b>	<b>-0.234</b>	<b>-0.174</b>	<b>-0.217</b>	<b>-0.197</b>	<b>-0.235</b>	<b>-0.245</b>	<b>-0.234</b>	<b>-0.204</b>
<b>Sector</b>									
Público	-0.139	-0.312	-0.256	-0.289	-0.339	-0.410	-0.432	-0.414	-0.328
Privado	-0.034	-0.162	-0.097	-0.147	-0.062	-0.068	-0.073	-0.060	-0.088
<b>Total</b>	<b>-0.082</b>	<b>-0.234</b>	<b>-0.174</b>	<b>-0.217</b>	<b>-0.197</b>	<b>-0.235</b>	<b>-0.245</b>	<b>-0.234</b>	<b>-0.204</b>
<b>Localidad</b>									
4	0.049	-0.178	-0.268	-0.350	-0.354	-0.475	-0.470	-0.375	-0.323
5	-0.682	-0.592	-0.574	-0.690	-0.534	-0.346	-0.931	-0.521	-0.581
6	-0.061	-0.204	-0.183	-0.204	-0.216	-0.243	-0.264	-0.257	-0.206
7	-0.574	-0.673	-0.248	-0.428	-0.249	-0.474	-0.474	-0.361	-0.435
8	-0.117	-0.280	-0.237	-0.285	-0.212	-0.297	-0.304	-0.244	-0.248
9	0.436	0.441	0.568	0.451	0.495	0.354	0.540	0.429	0.465
15	0.106	-0.083	0.141	0.134	0.028	0.065	-0.032	0.049	0.052
16	-0.056	-0.144	0.019	-0.088	-0.048	-0.017	-0.043	-0.088	-0.058
18	-0.075	-0.149	-0.071	-0.137	-0.139	-0.117	-0.142	-0.158	-0.124
19	-0.332	-0.675	-0.644	-0.601	-0.598	-0.626	-0.633	-0.691	-0.604
<b>Total</b>	<b>-0.082</b>	<b>-0.234</b>	<b>-0.174</b>	<b>-0.217</b>	<b>-0.197</b>	<b>-0.235</b>	<b>-0.245</b>	<b>-0.234</b>	<b>-0.204</b>

Cuadro 5: Puntaje total promedio según distancias

Distancia a la biblioteca	Años			Total Prom
	2000-2002 Prom	2003-2005 Prom	2006-2008 Prom	
Menor a 1000	-0.141	-0.048	-0.150	-0.110
Entre 1000-2500	-0.239	-0.306	-0.346	-0.306
Mayor a 2500	-0.112	-0.140	-0.177	-0.147
<b>Total</b>	<b>-0.161</b>	<b>-0.196</b>	<b>-0.238</b>	<b>-0.204</b>

Cuadro 6: Distribución de las características de los colegios en la distancia

	Distancia		
	Entre 0 y 1 Km %	Entre 1 y 2 Km %	Entre 2 y 4 Km %
<b>Administración de Colegios</b>			
Oficial	59.84	58.96	50.80
No Oficial	40.16	41.04	49.20
<b>Total</b>	<b>100</b>	<b>100</b>	<b>100</b>
<b>Proporción de profesores con posgrado</b>			
Menor a 30 %	51.18	61.32	63.19
Entre 30 % y 60 %	25.20	16.98	19.93
Más de 70 %	23.62	21.70	16.88
<b>Total</b>	<b>100</b>	<b>100</b>	<b>100</b>
<b>Jornada</b>			
Completa	31.50	37.26	42.90
Mañana	35.43	28.07	24.64
Tarde	33.07	34.67	32.46
<b>Total</b>	<b>100</b>	<b>100</b>	<b>100</b>
<b>Estudiantes por profesor</b>			
Menos de 20 alumnos por profesor	20.47	21.70	23.84
Entre 20 y 30 alumnos por profesor	59.84	58.96	54.06
Más de 30 alumnos por profesor	19.69	19.34	22.10
<b>Total</b>	<b>100</b>	<b>100</b>	<b>100</b>
<b>Tamaño del colegio</b>			
Más de 1000 estudiantes	39.37	50.47	27.90
Entre 500 y 1000 estudiantes	37.01	25.47	38.84
Menos 500 estudiantes	23.62	24.06	33.26
<b>Total</b>	<b>100</b>	<b>100</b>	<b>100</b>
<b>Género del colegio</b>			
Un sólo género	0	11.79	11.67
Mixto	100	88.21	88.33
<b>Total</b>	<b>100</b>	<b>100</b>	<b>100</b>

### A.2.1. Forma discreta

Cuadro 7: DD Discreto Especificación I

	Trat./Contr.	2003	2004	2005	2006	2007	2008
<b>Resultado Total</b>							
R1=750	10/182	0,2109 (0,2731)	0,0992 (0,2631)	0,2882 (0,4086)	0,1846 (0,4654)	0,1136 (0,2358)	0,1624 (0,2379)
R1=1000	19/173	0,1613 (0,1383)	0,0703 (0,1303)	0,0012 (0,0830)	0,0572 (0,1916)	-0,0145 (0,1199)	0,0398 (0,0718)
R1=1250	27/165	0,1315 (0,1027)	0,0941 (0,1088)	0,0045 (0,0353)	0,0613 (0,1256)	0,0108 (0,1087)	0,0651 (0,0732)
R1=1500	44/148	0,0515 (0,0525)	0,0655 (0,0602)	-0,0160 (0,0413)	0,0620 (0,0770)	0,0071 (0,0348)	0,0646 (0,0683)
R1=1750	52/140	0,0365 (0,0469)	0,0169 (0,0369)	-0,0455 (0,0334)	0,0100 (0,0500)	-0,0169 (0,0257)	0,0495 (0,0583)
R1=2000	70/122	0,0347 (0,0377)	-0,0077 (0,0161)	-0,0147 (0,0702)	-0,0424* (0,0210)	-0,0339 (0,0392)	0,0279 (0,0525)

R2=3500. Errores estándar en cluster por Localidad.

Significancia: \* 10 %, \*\* 5 %, \*\*\* 1 %.

Cuadro 8: DD Discreto Especificación II

	Trat./Contr.	2003	2004	2005	2006	2007	2008
<b>Resultado Total</b>							
R1=750	10/122	0,2193 (0,2860)	0,0911 (0,2591)	0,2456 (0,3799)	0,1447 (0,4486)	0,0995 (0,2105)	0,1581 (0,2191)
R1=1000	19/122	0,1588 (0,1375)	0,0527 (0,1235)	-0,0205 (0,0800)	0,0186 (0,1690)	-0,0299 (0,0964)	0,0328 (0,0730)
R1=1250	27/122	0,1284 (0,0976)	0,0791 (0,0990)	-0,0177 (0,0472)	0,0227 (0,1140)	-0,0018 (0,0823)	0,0491 (0,0715)
R1=1500	44/122	0,0530 (0,0547)	0,0499 (0,0527)	-0,0242 (0,0592)	0,0274 (0,0719)	-0,0060 (0,0322)	0,0598 (0,0761)
R1=1750	52/122	0,0381 (0,0518)	0,0112 (0,0337)	-0,0434 (0,0501)	-0,0099 (0,0437)	-0,0239 (0,0219)	0,0467 (0,0629)
R1=2000	70/122	0,0347 (0,0377)	-0,0077 (0,0161)	-0,0147 (0,0702)	-0,0424* (0,0210)	-0,0339 (0,0392)	0,0279 (0,0525)

R2=3500, R3=2000. Errores estándar en cluster por Localidad.

Significancia: \* 10 %, \*\* 5 %, \*\*\* 1 %.

Cuadro 9: BO-DD discreto

	Trat./Contr.	2003	2004	2005	2006	2007	2008
<b>Resultado Total</b>							
R1=750	10/182	0,2008 (0,3981)	0,2252 (0,3644)	0,4293 (0,4220)	0,3343 (0,5306)	0,1444 (0,4816)	0,1726 (0,4339)
	$\Delta_0$	0,0310 (0,4573)	0,2346 (0,3555)	0,5356 (0,4093)	0,4517 (0,4972)	0,3101 (0,4399)	0,3223 (0,4144)
	$\Delta_X$	0,1698 (0,1710)	-0,0094 (0,0985)	-0,1063 (0,1055)	-0,1174 (0,1137)	-0,1657 (0,1078)	-0,1497 (0,1005)
R1=1000	19/173	0,1626 (0,2353)	0,1460 (0,2138)	0,0851 (0,2474)	0,1524 (0,2810)	-0,0009 (0,2593)	0,0465 (0,2562)
	$\Delta_0$	0,1253 (0,2616)	0,1787 (0,2075)	0,2083 (0,2410)	0,2776 (0,2625)	0,1597 (0,2414)	0,2296 (0,2469)
	$\Delta_X$	0,0373 (0,1308)	-0,0327 (0,1149)	-0,1232 (0,1120)	-0,1252 (0,1241)	-0,1605 (0,1073)	-0,1831* (0,1080)
R1=1250	27/165	0,1308 (0,1595)	0,1056 (0,1489)	0,0286 (0,1789)	0,0639 (0,1891)	-0,0229 (0,1787)	0,0463 (0,1774)
	$\Delta_0$	0,1360 (0,1895)	0,1890 (0,1644)	0,1740 (0,1890)	0,2745 (0,2016)	0,2032 (0,1754)	0,2613 (0,1850)
	$\Delta_X$	-0,0051 (0,1091)	-0,0833 (0,1158)	-0,1455 (0,1079)	-0,2107 (0,1331)	-0,2262** (0,1118)	-0,2149* (0,1173)
R1=1500	44/148	0,0560 (0,1194)	0,0840 (0,1145)	0,0115 (0,1239)	0,0790 (0,1245)	-0,0202 (0,1430)	0,0637 (0,1270)
	$\Delta_0$	0,0535 (0,1390)	0,1459 (0,1278)	0,1349 (0,1330)	0,2530 (0,1586)	0,1201 (0,1466)	0,1831 (0,1568)
	$\Delta_X$	0,0026 (0,0990)	-0,0619 (0,1230)	-0,1235 (0,1052)	-0,1740 (0,1341)	-0,1403 (0,1155)	-0,1193 (0,1303)
R1=1750	52/140	0,0437 (0,1081)	0,0320 (0,1109)	-0,0172 (0,1106)	0,0379 (0,1213)	-0,0237 (0,1337)	0,0574 (0,1144)
	$\Delta_0$	0,0515 (0,1211)	0,0611 (0,1157)	0,0967 (0,1204)	0,2114 (0,1469)	0,1084 (0,1370)	0,1779 (0,1436)
	$\Delta_X$	-0,0078 (0,0906)	-0,0291 (0,1131)	-0,1139 (0,0958)	-0,1735 (0,1302)	-0,1321 (0,1204)	-0,1206 (0,1273)
R1=2000	70/122	0,0208 (0,1074)	-0,0132 (0,1079)	-0,0178 (0,1071)	-0,0446 (0,1175)	-0,0722 (0,1178)	0,0125 (0,1107)
	$\Delta_0$	0,0230 (0,1206)	0,0657 (0,1033)	0,0530 (0,1103)	0,1115 (0,1387)	0,0862 (0,1334)	0,1253 (0,1350)
	$\Delta_X$	-0,0021 (0,0853)	-0,0789 (0,1138)	-0,0708 (0,0952)	-0,1561 (0,1292)	-0,1584 (0,1195)	-0,1128 (0,1343)

R2=3500. Errores estándar con Bootstrapping con 200 repeticiones.

Significancia: \* 10 %, \*\* 5 %, \*\*\* 1 %.

## A.2.2. Forma Continua

Cuadro 10: DD Continuo Lineal

	2003	2004	2005	2006	2007	2008
Resultado Total						
R1=1500	0,2842 (0,2805)	0,1222 (0,2980)	0,1259 (0,1847)	0,1750 (0,4046)	0,0881 (0,2807)	0,1882 (0,2443)
R1=2000	0,1855 (0,1558)	0,0868 (0,1724)	0,0371 (0,0792)	0,0948 (0,2268)	0,0303 (0,1370)	0,1584 (0,1687)
R1=2500	0,1238 (0,1194)	0,0681 (0,1173)	0,0217 (0,0938)	0,0236 (0,1340)	-0,0076 (0,0590)	0,1171 (0,1287)
R1=3000	0,0972 (0,1373)	0,0640 (0,1000)	0,0068 (0,1006)	-0,0022 (0,1017)	-0,0303 (0,0379)	0,0975 (0,1001)
R1=3500	0,0871 (0,1599)	0,0692 (0,1109)	-0,0156 (0,1042)	-0,0157 (0,0953)	-0,0508 (0,0424)	0,0994 (0,0934)

R2=3500. Errores estándar en cluster por Localidad  
Significancia: \* 10 %, \*\* 5 %, \*\*\* 1 %.

Cuadro 11: DD Continuo Exponencial

	2003	2004	2005	2006	2007	2008
Resultado Total						
R1=1500	0,0732 (0,1057)	0,0031 (0,0932)	0,0718 (0,0929)	0,0592 (0,1381)	0,0500 (0,1168)	0,0718 (0,1086)
R1=2000	0,0495 (0,0662)	0,0051 (0,0620)	0,0421 (0,0518)	0,0380 (0,0878)	0,0297 (0,0703)	0,0501 (0,0707)
R1=2500	0,0361 (0,0470)	0,0045 (0,0455)	0,0301 (0,0350)	0,0251 (0,0611)	0,0196 (0,0475)	0,0372 (0,0516)
R1=3000	0,0286 (0,0378)	0,0041 (0,0363)	0,0234 (0,0267)	0,0189 (0,0474)	0,0145 (0,0358)	0,0297 (0,0406)
R1=3500	0,0239 (0,0320)	0,0036 (0,0309)	0,0191 (0,0221)	0,0156 (0,0396)	0,0116 (0,0298)	0,0252 (0,0341)

R2=3500. Errores estándar en cluster por Localidad  
Significancia: \* 10 %, \*\* 5 %, \*\*\* 1 %.

## A.3. Otras tablas

Cuadro 12: Tiempo de desplazamiento al colegio

Tiempo	Frec.	Frec. Acum.
Menos de 10 min.	51 %	51 %
Entre 10 y 20 min.	26 %	77 %
Entre 20 y 30 min.	23 %	100 %

(Fuente): DANE Censo Nacional de población 2005

## B. Figuras

Figura 1: Características educativas de los colegios según la distancia (1)

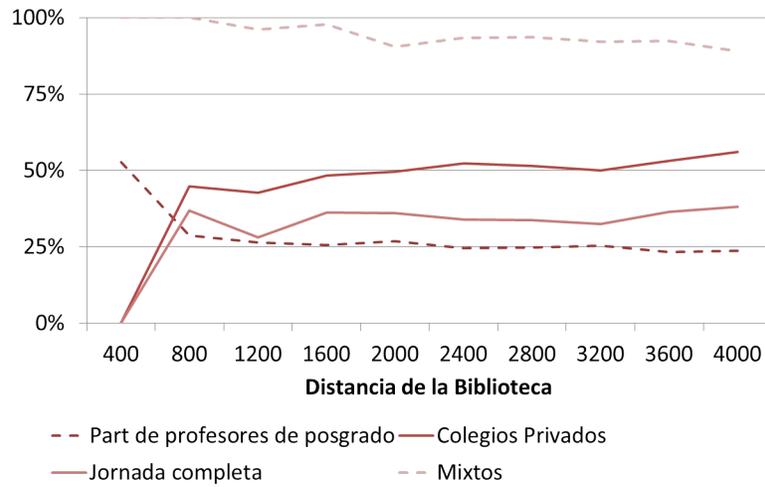


Figura 2: Características educativas de los colegios según la distancia (2)

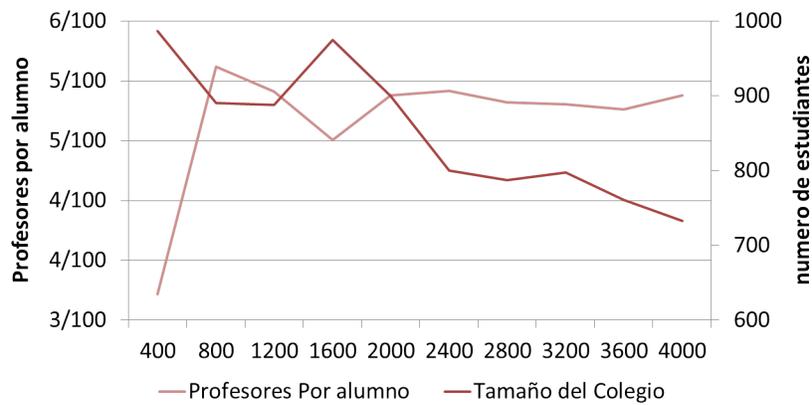


Figura 3: Distancia y puntaje: dispersión

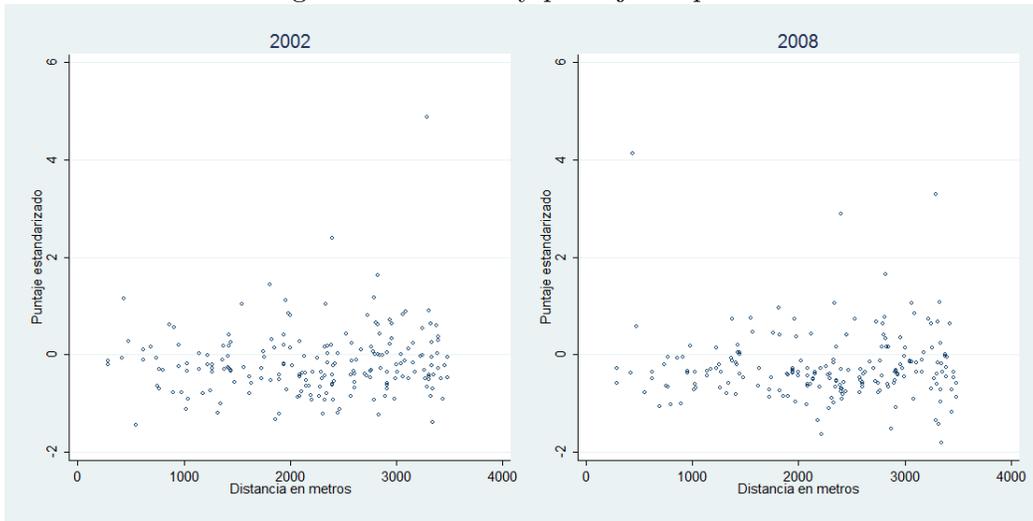


Figura 4: Bibliotecas y área de influencia

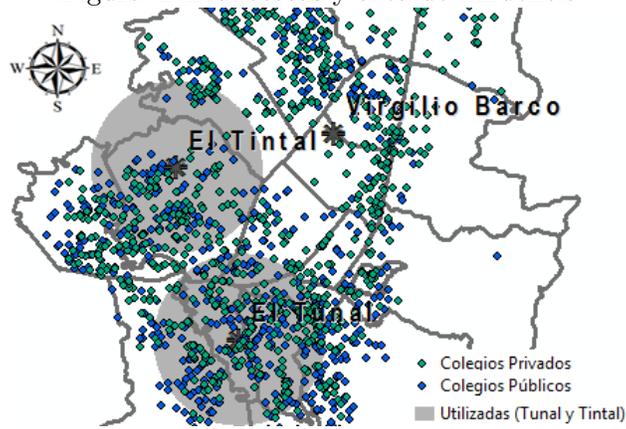


Figura 5: Ejercicio Discreto

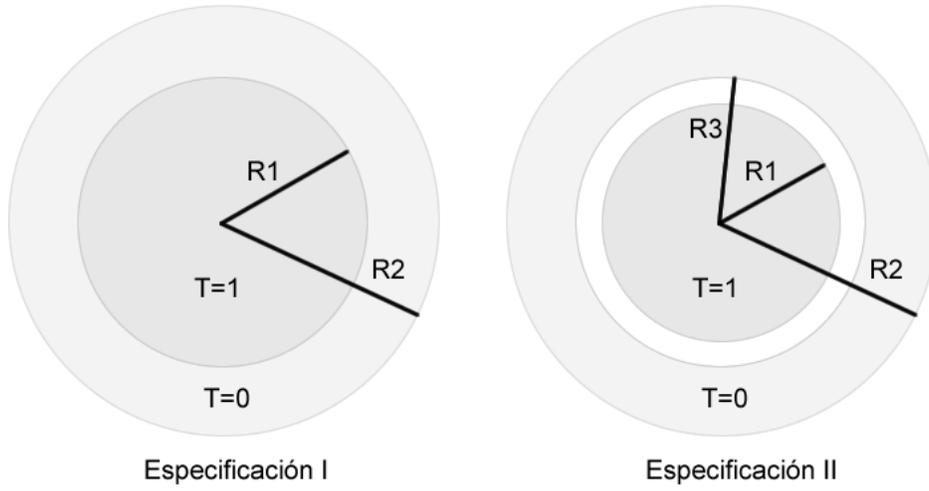


Figura 6: Ejercicio Continuo

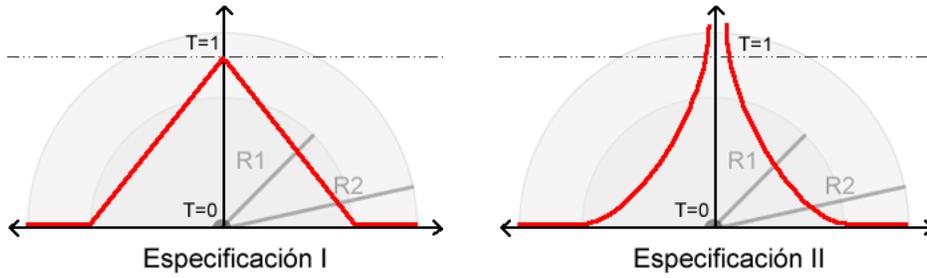
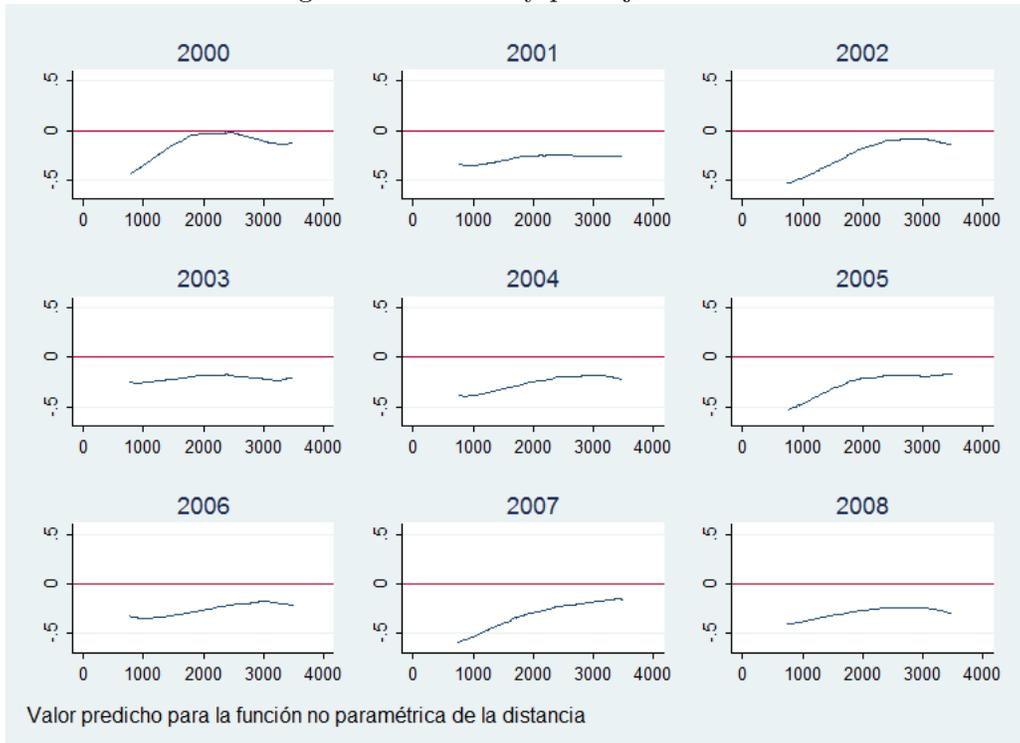


Figura 7: Distancia y puntaje: relación



## C. Anexo: Blinder-Oaxaca en Diferencia en Diferencias

La descomposición de Blinder y Oaxaca (1973) permite analizar la diferencia para una variable  $y$  entre dos grupos  $T = 1$  y  $T = 0$ , a la que nos referimos como  $\delta = E[y|T = 1] - E[y|T = 0]$ . La relación entre  $y$  y  $x$  está dada por la especificación paramétrica  $y_i = \alpha_0 + \alpha_1 x_i + u_i$ , y es el caso que la pertenencia a cada grupo implica tanto una diferencia en el intercepto  $\alpha_0$  como en las relaciones marginales  $\alpha_1$ . Por tal motivo,  $\delta$  puede dividirse entre una parte generada por la diferencia entre observables  $\Delta_x$ , y por otra parte basada en la diferencia entre no observables  $\Delta_0$ . Para estimarle, utilizamos una regresión lineal simple:

$$y = \beta_0 + \beta_1 x + \beta_2 t + \beta_3 t \cdot x + e_2$$

En tal caso, bajo la restricción  $E[e_2|T = 1] = E[e_2|T = 0]$ , podemos identificar puntualmente la diferencia  $\delta$ :

$$\begin{aligned} \delta &= E[y|T = 1] - E[y|T = 0] \\ &= [\beta_0 + \beta_2 + (\beta_1 + \beta_3)E[x|T = 1]] - [\beta_0 + \beta_1 E[x|T = 0]] \\ &= \beta_2 + (\beta_1 + \beta_3)E[x|T = 1] - \beta_1 E[x|T = 0] \\ &= \beta_2 + \beta_3 E[x|T = 1] + \beta_1 (E[x|T = 1] - E[x|T = 0]) \\ &= \Delta_0 + \Delta_x \end{aligned}$$

Nótese que si no hubiese un impacto sobre  $x$  por parte de  $T$ ,  $\beta_3 = \beta_1 = 0$ , y estaríamos frente al caso típico de un experimento bajo aleatorización del tratamiento  $T$ . Si ese no es el caso, podemos definir  $\Delta_x = \beta_1 (E[x|T = 1] - E[x|T = 0])$ , que refleja la diferencia de pertenecer a  $T = 1$  y no a  $T = 0$  sobre  $x$ . La otra parte,  $\Delta_0 = \beta_2 + \beta_3 E[x|T = 1]$ , sería la diferencia en  $y$  no explicada por la diferencia entre las  $x$ . En economía laboral, se suele explicar como un posible fuente a la ‘discriminación’ por pertenecer a  $T = 1$  y no a  $T = 0$ . Además de lo ya mencionado, es claro que estamos empleando restricciones sobre la linealidad de las relaciones.

Si pensamos en el contexto de evaluación de impacto, es útil considerar que una variable  $y$  sea impactada por un programa, pero el mecanismo afecte también a variables  $x$  que también son relevantes para explicar la variabilidad de  $y$ . Generalmente queremos aislar

el impacto neto del programa sobre  $y$ , pero también quisiéramos poder analizar en más detalle qué mecanismos hacen parte de dicho impacto. En tal contexto esta descomposición puede ser bastante útil. El problema es que nuestro supuesto básico para la identificación es equivalente a la aleatorización, y usualmente queremos algo menos restrictivo.

### **Blinder-Oaxaca en Diferencia en Diferencias**

La extensión más sencilla es hacia Diferencia en Diferencias. Supongamos que podemos observar dos periodos,  $A \in \{0, 1\}$ . Con ello vamos a poder definir un estimador de DD:

$$\begin{aligned} \delta &= (E[y|T = 1, A = 1] - E[y|T = 0, A = 1]) \\ &\quad - (E[y|T = 1, A = 0] - E[y|T = 0, A = 0]) \end{aligned}$$

Que puede estimarse con la siguiente ecuación lineal:

$$y = \alpha_0 + \alpha_1 T + \alpha_2 A + \alpha_3 x + \alpha_4 x \cdot T + \alpha_5 x \cdot A + \alpha_6 T \cdot A + \alpha_7 x \cdot T \cdot A + u$$

Nuestro estimador se puede identificar bajo unas condiciones mucho más laxas. Siguiendo la notación de Blundell y Costa-Dias (2009), lo no observable puede estar compuesto por un choque agregado  $m_A$  y un efecto fijo idiosincrático  $n_i$ :  $E[u_{i,A}|t, A] = E[n_i|t_i] + m_A$

Con ello podemos calcular los términos del estimador

$$\begin{aligned} E[y|T = 0, A = 0] &= \alpha_0 + \alpha_3 E[x|T = 0, A = 0] \\ E[y|T = 1, A = 0] &= \alpha_0 + \alpha_1 + (\alpha_3 + \alpha_4) E[x|T = 1, A = 0] \\ E[y|T = 0, A = 1] &= \alpha_0 + \alpha_2 + (\alpha_3 + \alpha_5) E[x|T = 0, A = 1] \\ E[y|T = 1, A = 1] &= \alpha_0 + \alpha_1 + \alpha_2 + \alpha_6 + (\alpha_3 + \alpha_4 + \alpha_5 + \alpha_7) E[x|T = 1, A = 1] \end{aligned}$$

Como resultado:

$$\begin{aligned}
\delta &= ((\alpha_0 + \alpha_1 + \alpha_2 + \alpha_6 + (\alpha_3 + \alpha_4 + \alpha_5 + \alpha_7)E[x|T = 1, A = 1]) \\
&\quad - (\alpha_0 + \alpha_2 + (\alpha_3 + \alpha_5)E[x|T = 1, A = 1])) \\
&\quad - ((\alpha_0 + \alpha_1 + (\alpha_3 + \alpha_4)E[x|T = 1, A = 0]) - (\alpha_0 + \alpha_3E[x|T = 0, A = 0])) \\
\delta &= \alpha_6 + (\alpha_4 + \alpha_5 + \alpha_7)E[x|T = 1, A = 1] - \alpha_5E[x|T = 0, A = 1] - \alpha_4E[x|T = 1, A = 0] \\
&\quad + \alpha_3[(E[x|T = 1, A = 1] - E[x|T = 0, A = 1]) - (E[x|T = 1, A = 0] - E[x|T = 0, A = 0])] \\
&= \Delta_0 + \Delta_x
\end{aligned}$$

De nuevo, si  $x$  no es impactado por  $T$ , obtenemos un clásico DD:  $\delta = \alpha_6$ . Si ese no es el caso, podemos definir  $\Delta_x = \alpha_3[(E[x|T = 1, A = 1] - E[x|T = 0, A = 1]) - (E[x|T = 1, A = 0] - E[x|T = 0, A = 0])]$ . Y adicionalmente,  $\Delta_0 = \alpha_6 + (\alpha_4 + \alpha_5 + \alpha_7)E[x|T = 1, A = 1] - \alpha_5E[x|T = 0, A = 1] - \alpha_4E[x|T = 1, A = 0]$ .