

## 7. Más allá de la infraestructura: el efecto vecindario de las bibliotecas públicas en la calidad de la educación.

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

Universidad del Rosario

*Dirección electrónica:* rodriguez.paul@ur.edu.co,  
trujillos.jose@ur.edu.co,  
valderrama.daniel@ur.edu.co

### 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 internos al vecindario que las rodea, y en este sentido, el impacto de estas sobre la calidad de la educación es producto de un efecto vecindario que utiliza mecanismos de transmisión que van más allá de la simple infraestructura o activos educativos que estas bibliotecas ofrecen, estos mecanismos también se refieren al cambio que se da en los hábitos del vecindario y que son producto de la oferta cultural y la renovación urbana que en este se experimenta después de la construcción y del funcionamiento de dichas bibliotecas. En 2001, en Bogotá se construyeron tres *mega-bibliotecas* (bibliotecas públicas) con la intención de influir en las comunidades, no solo mediante un mayor acceso a la información, sino también de modificar la infraestructura urbana, sus hábitos de recreación y esparcimiento. Este documento mide el impacto de estas bibliotecas en la calidad de la educación en el colegio. Los resultados muestran que las *mega-bibliotecas* son una política progresiva en términos educativos pues se localizaron alrededor de colegios con menores capacidades educativas, el impacto de estas bibliotecas sobre la calidad de la educación a pesar de ser positivo no fue significativo en todas las estimaciones ni cumplió con los criterios de robustez; por tanto se concluye que a nivel de colegio no hay evidencia de un impacto claro de las *mega-bibliotecas* sobre la calidad de la educación.

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

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

## 7.1. Introducción

La literatura asociada a los determinantes de la calidad de la educación suele considerarse, además de variables socioeconómicas del estudiante, una serie de insumos educativos que forman 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 de la evaluación de impacto referida al efecto causal de las bibliotecas públicas sobre la calidad de la educación.

Una biblioteca pública es externa al colegio y al hogar del estudiante pero es interna a su vecindario, esto implica que el impacto de estas sobre el desempeño educativo obedezca a un mecanismo de transmisión distinto al que se ha considerado tradicionalmente. Estas bibliotecas presentan dos tipos de interacciones con el vecindario sobre el que se construyen y actúan: 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 (e.g las “vacaciones recreativas” para niños); si bien estos espacios no son estrictamente educativos pueden provocar cambios en los hábitos de los ciudadanos del vecindario como la reducción de la delincuencia y las drogas como efecto de la oferta de actividades culturales a las que los estudiantes pueden acceder por medio de las bibliotecas, dichos cambios en los hábitos 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.

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 mega bibliotecas creadas en 2001: 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. 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. 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, los cuales se componen por estudiantes cuyos hogares tienen características socioeconómicas diferentes a los de los estudiantes que rodean las Bibliotecas del Tintal y El Tunal, por ejemplo, los estudiantes de alrededor de la Virgilio Barco son de colegios privados, viven retirados de la biblioteca y tienen más acceso a activos educativos como libros y computadores dentro del hogar lo que implica una interacción de otra índole con la biblioteca, razón por la cual no fue considerada.

En síntesis, el documento busca identificar el efecto causal de las bibliotecas El Tintal y El Tunal en el desempeño escolar. 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, quien además financia este proyecto. 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. 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 las conclusiones respectivamente.

## **7.2. 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:

### 7.2.1 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.

### 7.2.2 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.

### 7.2.3 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 estas 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. Además, no se ha diferenciado claramente la interacción de las bibliotecas públicas respecto a las privadas, pues las primeras además de brindar el acceso a libros e información, promueven programas culturales y generan cambios en los hábitos del vecindario. Como se mencionó antes, en el caso de las bibliotecas públicas no solo existe el efecto directo derivado del uso de la biblioteca como fuente de acceso a la información sino también uno indirecto derivado de la transformación del vecindario en donde estas se localizan. Así, el principal mecanismo teórico en el que se fundamenta la trasmisión del impacto, que es evaluado en este documento, proviene de la literatura de las redes sociales y es conocido como “efecto vecindario”, definido de manera genérica por Durlauf (2004) como la manera en que una comunidad influye los hábitos de los ciudadanos que se encuentran alrededor por medio de un denominado “efecto contagio” que Según Jenks y Mayer (1990) es uno de los mecanismos mediante los cuales se da el efecto vecindario, dicho “efecto contagio” es definido como: la replicación por parte de un individuo de las decisiones de sus pares, Puede ser “positiva” como el cumplimiento de las normas, o “negativa” como el consumo de drogas.

### 7.3. Datos: cálculos y hechos estilizados

Los datos usados se encuentran disponibles a dos niveles de desagregación: a nivel de colegio y a nivel de estudiantes. Los datos a nivel de colegio presentan una ventaja crucial sobre la información a nivel de estudiante, esta consiste en la ventana de tiempo que cada nivel permite observar; mientras el nivel mas agregado (los colegios) cuenta con información para antes y después de las bibliotecas el nivel mas desagregado (los estudiantes) únicamente cuenta con información completa para después de la creación de estas (específicamente a partir de 2008). A pesar del aumento ostensible de controles que la especificación a nivel de estudiantes permite, no contar con información para antes de la creación de la bibliotecas impide lograr una adecuada identificación del impacto, este sesgo se refuerza si se considera que las vecindades sobre las que las bibliotecas tienen influencia son distintas de aquellas en las que la influencia no existe o es menor (vecindades distantes a las bibliotecas), dichas diferencias son características no observables específicas a la localidad (e.g criminalidad, vandalismo, zonas de recreación, entre otros) y por tanto una estimación de corte transversal, como la que habría que realizar si se utiliza la información a nivel de estudiantes, implicaría un sesgo no controlado; contrario a una estimación a nivel de colegios, la cual puede incluir información sobre la situación de los colegios antes y después de las bibliotecas y así corregir por el sesgo de las variables no observables.

De acuerdo a lo anterior, en el documento se utilizará la información a nivel de colegio<sup>1</sup>, para la ciudad de Bogotá, este nivel de agregación permite tener un una base 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.

#### 7.3.1. 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.

---

1 A nivel de estudiantes se realizaron múltiples estimaciones obteniendo poca robustez en los resultados.

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 (i.e el formulario de preguntas diseñado por el ICFES). 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>2</sup>. Se tienen en cuenta los resultados estandarizados<sup>3</sup> en las áreas de matemáticas, lenguaje, ciencias (un promedio entre las áreas de biología, física y química<sup>4</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, ahora un impacto positivo (negativo) de las mega bibliotecas se 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).

### 7.3.2. Las fuentes de información para controlar el efecto causal

Como se explicó 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, la estimación debe controlar por todo sesgo de selección que pueda existir en los colegios y estudiantes que reciben el impacto de las bibliotecas, para esto se utilizan las 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 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, esta institucionalidad se define con base en el concepto de instituciones de North (1996) como las normas y reglas, en este caso explícitas, que definen

2 Si bien la prueba es diferente cada semestre, cada colegio presenta sólo una de las dos versiones al año.

3 Respecto a media y desviación estándar de Bogotá.

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

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  $t$ .  $v(xt)$  es la desviación estándar de los resultados en el periodo  $t$ .

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.

4 Si bien el concepto de “ciencias” puede llevar a discusiones, esta clasificación obedece a la resolución.

ciertos parámetros de funcionamiento del colegio. Dichas normas son relativamente estables en el tiempo como: el calendario, la naturaleza y la jornada. 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). 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>5</sup>.

El segundo grupo de variables de control consiste en aquellas características que ofrece la institución educativa y que cambian en el corto plazo como: El tamaño del colegio, este puede dar idea de las economías de escala con que cuenta el plantel educativo, para su medición se utilizó 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 matrícula tiene sobre la calidad de la educación. El capital humano de los profesores, definido con base en el nivel de educación de estos (primaria, secundaria, educación superior y posgrado). La *proporción* de profesoras femeninas 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, asumiendo así que estas variables no cambian de manera importante en el tiempo. 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>6</sup>.

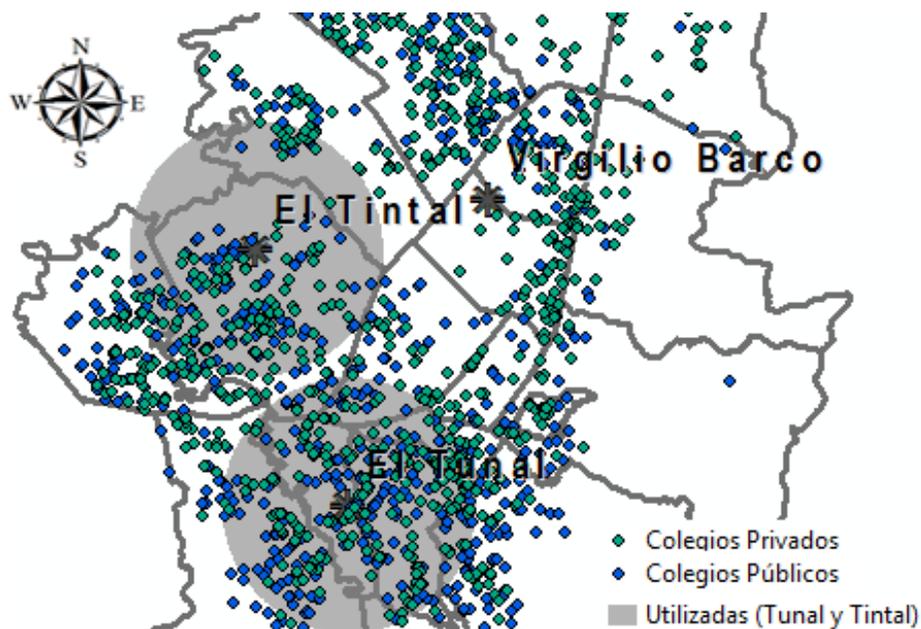
5 Si bien otro control importante es la localidad en la que se ubica el colegio, no se consideró debido a los altos niveles de colinealidad que esta tiene con el impacto de la mega biblioteca, es decir, por lo general una biblioteca tiene incidencia en un alto porcentaje a una única localidad, por lo cual el efecto de la localidad podría reducir el efecto de la biblioteca.

6 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.

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 y la biblioteca más cercana (la justificación de esto se encuentra en la sección 4.1). Como puede verse en la figura 3, solo se encuentran dos bibliotecas en esta pues como se mencionó antes sólo se incluyen las bibliotecas El Tunal y El Tintal, pues la biblioteca Virgilio Barco se encuentra en un entorno social diferente a las dos primeras y con una concentración más baja<sup>7</sup>.

Figura 3: Bibliotecas y area de influencia



7 La información sobre la localización espacial de cada colegio proviene de la Secretaría de educación Distrital.

### 7.3.3. 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>8</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>9</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 (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 redireccionar el impacto como colegios en concesión, o los mega colegios 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.

Cuadro 1: Colegios por distancias

Distancia desde la biblioteca	Todos los colegios	Usados en la estimación
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
<b>Total</b>	<b>275</b>	<b>232</b>

Fuente: Cálculos propios

8 En el caso de instituciones públicas un colegio es considerado como la combinación sede-jornada.

9 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 Nopo (2008).

Cuadro 2: Estudiantes por distancias

Distancia desde la biblioteca	Todos los colegios	Usados en la estimación
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
<b>Total</b>	<b>41738</b>	<b>38943</b>

Fuente: Cálculos propios

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

#### 7.3.4. 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 3. En estas tablas se muestra el valor promedio del puntaje (estandarizado) de los colegios de acuerdo con las características de estos. Así, en los cuadros se puede observar ciertas relaciones estadísticas sobre las que la literatura previa ha evaluado la existencia de un efecto causal, se enfatiza en que en estos cuadros se presentan simplemente relaciones estadísticas mas no de causalidad, bajo el argumento que la causalidad requiere de un examen minucioso que implica analizar un aspecto a la vez (Rubin 1974), dado que el objetivo de este documento se concentra únicamente en observar el impacto de las bibliotecas públicas en la calidad de la educación, se deja de lado todo tema referente a la causalidad de otras variables sobre esta.

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								
	2000 prom	2001 prom	2003 prom	2004 prom	2005 prom	2006 prom	2007 prom	2008 prom	Total 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>	-0.082	-0.234	-0.174	-0.217	-0.197	-0.235	-0.245	-0.234	-0.204
<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>	-0.082	-0.234	-0.174	-0.217	-0.197	-0.235	-0.245	-0.234	-0.204
<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>	-0.082	-0.234	-0.174	-0.217	-0.197	-0.235	-0.245	-0.234	-0.204

Al mirar el cuadro 3 se tiene que los colegios de mayor tamaño (medido por medio de la matrícula 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, si se continúa observando las variables que involucran los docentes se tiene que las instituciones que, según los últimos 2 paneles de este cuadro, 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, si bien esto parece extraño, la literatura lo ha explicado por medio del esquema de incentivos de los profesores que rige el sistema educativo del país, donde los profesores con mayores niveles de educación se encuentran principalmente en colegios públicos por tanto se combina tanto el efecto de los profesores como el efecto de la naturaleza del plantel. (Núñez, Stainer, Cadena y Pardo, 2002).

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.

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>10</sup>, mientras los peores se dieron en Ciudad Bolívar, Usme y Bosa<sup>11</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. Este resultado es un indicio de un impacto positivo de las bibliotecas, sin embargo, aun falta validar si este es o no un efecto causal.

10 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.

11 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 mega bibliotecas tienen calendario A.

Cuadro 5: Puntaje total promedio segun 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>	-0.161	-0.196	-0.238	-0.204

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 alejamos de las mega bibliotecas 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)).

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>Genero del colegio</b>			
Un solo genero	0	11.79	11.67
Mixto	100	88.21	88.33
<b>Total</b>	<b>100</b>	<b>100</b>	<b>100</b>

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 mega bibliotecas, donde los colegios cercanos a estas presentan características más desfavorables en términos de factores que impactan la calidad de la educación.

Con base en lo anterior y en caso de existir un efecto causal de las bibliotecas en la calidad de la educación este debe ser interpretado como un *efecto promedio sobre los tratados* (ATT, por sus siglas en inglés) en lugar de un *efecto promedio del tratamiento* (ATE, por sus siglas en inglés) dado que los colegios que se encuentran alrededor de las bibliotecas tienen características específicas y por tanto el impacto, en caso de existir, sería concluyente para solo este tipo de colegios.

## 7.4. 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 *mega-bibliotecas*; la segunda hace referencia a la estrategia de identificación para medir la magnitud y significancia del impacto que las *mega-bibliotecas* tienen sobre estos colegios y/o estudiantes, y la última etapa presenta una manera de descomponer el impacto estimado en la segunda etapa.

### 7.4.1. El papel de la distancia en la incidencia del efecto vecindario de las mega-bibliotecas

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 bajo el supuesto de que el uso de las bibliotecas es mayor por quienes viven más cerca a estas, dado que quienes viven lejos enfrentan mayores costos de desplazamiento que reducen sus incentivos a visitarlas de manera frecuente.

Por otra parte, como se explicó 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. En otras palabras, si bien no se discute que las bibliotecas públicas pueden y son visitadas por personas que viven tanto cerca como lejos, son las personas que viven cerca las que presentan una mayor probabilidad de interactuar de manera continua con las bibliotecas y sus programas, y además son quienes reciben el efecto vecindario el cual es incondicional a usar o no las bibliotecas.

Según lo anterior, se puede concluir que los estudiantes que viven y/o estudian cerca de las *mega-bibliotecas* son quienes experimentan con mayor intensidad el posible impacto de estas. Así, la variable idónea para identificar la interacción de las bibliotecas con estos sería contar las observaciones georeferenciadas a nivel de estudiantes, y así poder observar si los puntajes obtenidos en SABER 11 por los estudiantes que viven y/o estudian alrededor de las bibliotecas son significativamente mayores que los de aquellos que viven lejos. Sin embargo, aun no están disponibles dichos datos y, según se explico antes, las estimaciones a este nivel de desagregación (estudiantes) implican importantes sesgos por la carencia de una ventana de tiempo amplia que considere tanto el antes como el después de la construcción de las mega bibliotecas.

De otro lado, según el Cuadro 11 el 77 % de los estudiantes del país viven a menos de 20 minutos de los establecimientos educativos a los que asisten. Esto es especialmente cierto para el caso de los niveles socioeconómicos bajo y medio, en los cuales la mayoría de hogares llevan a sus hijos a colegios públicos en los que la inscripción depende de la distancia entre el colegio y el hogar pues las familias de estos niveles socioeconómicos no cuentan usualmente con los recursos para pagar diariamente un transporte al colegio. Por tal motivo la posición espacial del colegio es una buena aproximación de la localización de los estudiantes; así si se consideran los colegios que se encuentran alrededor de las bibliotecas como aquellos que reciben el impacto directo e indirecto de estas se obtiene al mismo tiempo una buena aproximación de los estudiantes que reciben este mismo impacto, dado que existe una probabilidad del 77 % que el vecindario de estudiantes circunscrito al radio de acción de las bibliotecas asista a un colegio que también se encuentra cerca a estas.

Cuadro 11: Tiempo de desplazamiento al colegio

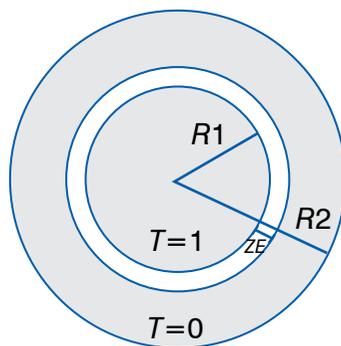
Tiempo	Frec.	Frec. Acum.
Menos de 10 min.	51%	51%
Entre 10 y 20 min.	26%	77%
Entre 20 y 30 min.	23%	100%

Fuente: DANE Censo nacional de población 2005.

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 socioeconómicas 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, este es el maximo del segundo radio ( $R2$ ) en la Figura 4.

Figura 4: Ejercicio Discreto



Evidentemente el “cerca” o “lejos” es arbitrario y requiere de una norma de selección que es parte de la pregunta de investigación. Se consideraron dos alternativas que utilizan la distancia, una discreta y otra continua, para seleccionar el nivel de exposición de un colegio a las *mega-bibliotecas*. La alternativa discreta es una definición dicótoma que separa a los colegios entre tratados o no tratados (controles) por la biblioteca, esta alternativa se explica por medio de la Figura 4, en esta se escoge un primer radio  $R1$  cuya distancia define los colegios que son considerados como tratados, aquellos que se encuentren a una distancia de la biblioteca menor a  $R1$ , los colegios que se encuentren por fuera del radio  $R1$  pero al interior de un segundo radio  $R2$  son definidos como controles. Se realizaron múltiples ejercicios teniendo en cuenta un amplio rango de opciones para estos radios lo que llevo a que finalmente a tomar la decisión de dejar el radio  $R2$  estático en 3.5 Km mientras que el radio  $R1$  varia continuamente con el fin de evaluar para cuales rangos se puede encontrar un impacto significativo de las bibliotecas<sup>12</sup>.

<sup>12</sup> Dado que la selección del umbral es arbitraria y con el fin de identificar de manera más clara la diferencia entre el grupo que se consideraba recibe el impacto de la biblioteca y el grupo que no, se definió una zona gris, tal y como se aprecia en la Figura 4, que es un área intermedia entre el área de los colegios que son definidos como tratados por la biblioteca y el área de los que son definidos como controles, después de realizar distintas especificaciones de radios y de área de esta zona gris se encontró que el uso o no de la zona gris no cambiaba de manera importante los resultados (debido tal vez a la poca densidad de colegios) por lo que en función de hacer mas armoniosa la estimación se tomo la decisión de no aplicarse dicha zona gris, quien este interesado en dichos resultados le serán enviados bajo solicitud.

Otra alternativa para definir la interacción de los colegios con las bibliotecas es definir una especificación continua del tratamiento, en esta se asume que cada colegio experimenta una intensidad diferente de exposición al efecto de la biblioteca y que esta tiene una relación inversa con la distancia a la que el colegio se encuentra de la biblioteca, bajo esta especificación una de las *mega-bibliotecas* tiene mayor incidencia sobre un colegio que se encuentra a 500 metros que sobre uno que se encuentra a 600 metros, la especificación que reproduce esta idea y que fue usada en las estimaciones define la variable de intensidad del tratamiento  $T$  que experimenta un colegio  $i$  como  $T_i = 1 - [(dmáx - d_i) / dmáx]$ , donde  $dmáx$  es el radio de 3.5 Km, el umbral máximo del segundo radio definido en la figura 4 para delimitar los colegios que sirven de contrafactual, mientras  $d_i$  es la distancia entre el colegio  $i$  y la *mega-biblioteca* más cercana. Dado que bajo la especificación continua no existe ningún colegio que tenga en su variable de exposición al tratamiento el valor de cero, se propone censurar el valor de los colegios que se encuentran a más de 2 Km de la biblioteca a este cero<sup>13</sup>. Esta censura es necesaria para realizar la descomposición continua de Blinder Oaxaca desarrollada por Ñopo (2008) y que se mencionara en la Sección 4.3.<sup>14</sup>

#### 7.4.2. El modelo de Diferencias en Diferencias como técnica de estimación del impacto

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. Considerando que los vecindarios donde las bibliotecas operan son distintos a los que no tienen incidencia,<sup>15</sup> la técnica de estimación debe capturar estas diferencias iniciales que no obedecen a la construcción de las bibliotecas, para esto la técnica más apropiada es la estimación por medio de diferencias en diferencias (*DD*). En ella se estima la diferencia entre “tratados” y “controles” después de implementar el programa de las *mega-bibliotecas* 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 (tratados y controles) hubiesen tenido una evolución similar si el programa nunca hubiese sido implementado. Si la variable dependiente (el puntaje estandarizado de SABER 11) es  $Y$ ,  $A$  es la variable indicadora

13 La decisión de la distancia a la cual se realizó la censura (2 Km) se tomó con base en las distintas estimaciones hechas, en estas se observó que bajo ninguna especificación se encontró un impacto de la biblioteca para colegios localizados a más de 2 Km de distancia.

14 Además se consideró una especificación en la que la distancia reducía exponencialmente la incidencia de la biblioteca en el colegio. Para esta especificación exponencial la intensidad de tratamiento  $T_e i$  que experimenta un colegio  $i$  es  $T_e i = 1/d_i$ , dado que esta especificación depende mucho de contar con observaciones cerca a la biblioteca sobre las que se recoge los valores más altos del impacto, lo cual no es el caso de los datos con que se cuenta. Se obtuvieron estimaciones poco robustas, por tanto no se colocaron en el documento y se pueden proveer bajo solicitud.

15 Tanto en términos de factores educativos como de variables socioeconómicas y de vandalismo.

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),  $T$  es el estado del tratamiento (es mayor para los colegios que reciben el impacto de las bibliotecas) y  $X$  es la matriz de los controles (jornada, naturaleza del colegio, matrícula, etc), el efecto causal  $\delta_{ATE}$  está dado por:

$$\begin{aligned} \delta_{ATE} = E[E[Y|T=1] - E[Y|T=0]|X] = \\ E[(E[Y|A=1, T=1] - E[Y|A=1, T=0]) - \\ (E[Y|A=0, T=1] - E[Y|A=0, T=0])|X] \end{aligned} \quad (1)$$

Esto puede estimarse por medio de la siguiente regresión de mínimos cuadrados ordinarios.

$$Y = \delta_{ATE} T A + \beta_1 T + \beta_2 A + \beta X + e \quad (2)$$

Bajo la aproximación discreta, la variable  $T$  de la ecuación anterior es una dicótoma que indica si el colegio es parte del grupo de tratamiento y que se genera como se explico antes para varios tipos de radios, mientras que bajo la aproximación continua se utiliza el inverso de la distancia, bajo dos tipos de especificaciones una lineal y otra cuadrática con el fin de identificar si el efecto causal además de ser continuo es lineal o no lineal.

### 7.4.3. Descomposición del efecto de las mega bibliotecas

Al implementar el modelo de  $DD$  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 que interactúa con los 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 socioeconómicas de los estudiantes, proporción de colegios privados, principal tipo de jornada en que estudian, entre otros. Dicho componente según se observe en el análisis del Cuadro 6 revelará como los controles tienen mejores dotaciones que los tratados. El segundo componente,  $-\Delta_0$ , explica la brecha como producto de la “discriminación” o variables no observables, en este caso las *mega-bibliotecas*.

La descomposición se ha realizado usualmente bajo escenarios de corte transversal y considerando únicamente dos grupos de comparación, en este caso se aplicó para un escenario dinámico como el planteado por el modelo *DD* y además utilizando una extensión del *BO* a un tratamiento continuo propuesta por Ñopo (2008), donde el inverso de la distancia del colegio a la mega biblioteca explica la intensidad del tratamiento<sup>16</sup>.

## 7.5. Resultados

Según se explicó en la sección anterior, si bien se utilizó el modelo de Diferencias en Diferencias en todas las estimaciones del impacto de las bibliotecas, se propone estimar dos especificaciones diferentes que implican dos preguntas de investigación distintas, 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 *mega-bibliotecas* generan un impacto sobre los colegios independiente de las características de los mismos. 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*, por tanto el modelo asume que hay un impacto diferenciado para cada colegio según sus características, esta última está inspirada en la descomposición de Blinder Oaxaca (1973).

Como pruebas de robustez se plantea realizar estimaciones por separado para cada uno de los años posteriores a la creación de las bibliotecas (Desde 2003 hasta 2008), esto permitirá observar la evolución y consistencia del efecto. Además, como se mencionó antes se plantearon dos definiciones de la interacción de las bibliotecas con los colegios, una discreta (para la cual se estiman diferentes radios) y otra continua. Esto lleva a que se tengan cuatro paneles de estimación, dos que utilizan la especificación *DDT* bajo la definición continua y discreta que son el Cuadro 7 y el Cuadro 9 respectivamente y otros dos paneles que son la especificación *DD-BO* bajo la definición continua, en el Cuadro 8, y bajo la definición discreta, en el Cuadro 10.

<sup>16</sup> Dado que el *DD* es un modelo dinámico, se realizaron dos descomposiciones estáticas de *BO* (en el caso de la distancia continua se utiliza la especificación propuesta en Ñopo (2008)), una para antes de las mega bibliotecas y otra para después de estas. En cada una se calcula el componente que mide el nivel de discriminación o diferencias en no observables, si se asume que las bibliotecas tienen algún impacto entonces se esperaría que este coeficiente aumente tras la construcción y funcionamiento de la bibliotecas y por tanto que la diferencia entre ambos coeficientes debería ser positiva (Las regresiones se realizan anidadas con el fin de poder hacer inferencia estadística sobre la diferencia entre ambos componentes)

En el Cuadro 7 se observa los resultados de la especificación *DDT* en la que se utiliza el tratamiento como una variable continua, según este existe un efecto positivo pero no significativo para ninguno de los años, este efecto oscila entre 0.05 y 0.12 desviaciones estándar (*DE*) para los colegios que se encuentran alrededor de la biblioteca. Se debe recordar que esta especificación estima un impacto que decrece a medida que la distancia entre el colegio y la biblioteca aumenta, lo que significa que el valor del coeficiente es el impacto cuando un colegio se encuentra muy cerca<sup>17</sup>, mientras mas lejos se encuentre el colegio el impacto esperado será menor que el valor del coeficiente.

Cuadro 7: *DD* continuo

Tratados	2003	2004	2005	2006	2007	2008
<b>Resultado Total</b>						
<b>70</b>	-0.044 ( 0.0619)	0.1189 ( 0.0760)	0.0628 ( 0.0832)	0.1034 ( 0.1368)	0.0511 ( 0.1033)	0.1276 ( 0.1633)

En el Cuadro 9 se tiene la misma especificación *DDT* pero con múltiples definiciones discretas para seleccionar a los tratados y no tratados por la biblioteca. Según se observa en esta estimación el impacto sigue siendo en general positivo y no significativo; a pesar de esto es importante resaltar que para cada uno de los años, a medida que es mayor la distancia del radio *R1* (se avanza verticalmente en cada columna del Cuadro 9) es menor el impacto de las bibliotecas, lo cual es consistente con la idea que de existir algún impacto, este se reduce con la distancia. Según se observa en el Cuadro 9, los colegios que se encuentran a 500 metros alrededor de la biblioteca más cercana tienen entre 0.53 y 0.67 *DE* más en la Prueba SABER-11 que sus homólogos que están lejos de las bibliotecas, si bien este impacto es positivo y significativo a partir de 2005, no es robusto pues en dicho radio se encuentran únicamente 5 colegios definidos como tratados, lo cual son muy pocas observaciones para definir un efecto causal<sup>18</sup>, por otra parte, si se define al grupo de control como los colegios que se encuentran hasta 750, 1000, 1250 o 1500 metros alrededor de las bibliotecas las conclusiones sobre el impacto cambian drásticamente, ahora no es significativo el efecto de las bibliotecas y además la magnitud del impacto alcanza un máximo de 0.42 *DE*.

17 Por ejemplo a menos de 100 metros, pues si se recuerda la fórmula para calcular la intensidad del tratamiento continua era  $T_i = 1 - [(d_{max} - d_i) / d_{max}]$ , por ende a menos de 100 metros el valor de  $T_i \cong 1$

18 Efectivamente al observar los datos se encuentra que el resultado encontrado en 500 metros es causado principalmente por uno de los 5 colegios, lo cual demuestra que con tan pocas observaciones no puede obtenerse una conclusión robusta.

Cuadro 9: *DD* discreto

Trat./Contr.		2003	2004	2005	2006	2007	2008
<b>Resultado Total</b>							
R1=500	5/187	0,3195 (0,2892)	0,1119 (0,2693)	0,5894 (0,2804)	0,5344 (0,2747)	0,6234 (0,2858)	0,6763 (0,2735)
R1=750	10/182	0,1726 (0,2087)	0,2089 (0,1997)	0,4254 (0,2095)	0,2994 (0,2054)	0,1448 (0,2016)	0,1848 (0,1930)
R1=1000	19/173	0,1723 (0,1526)	0,1566 (0,1455)	0,1074 (0,1532)	0,1525 (0,1496)	0,0269 (0,1512)	0,0938 (0,1449)
R1=1250	27/165	0,1358 (0,1280)	0,1323 (0,1212)	0,0562 (0,1279)	0,0760 (0,1250)	0,0144 (0,1278)	0,0746 (0,1226)
R1=1500	44/148	0,0406 (0,1057)	0,0943 (0,0998)	0,0318 (0,1050)	0,0865 (0,1031)	0,0045 (0,1053)	0,0664 (0,1014)4

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

Luego si se consideran los resultados de la especificación *DD-BO* la pregunta de investigación se amplía a no solo si existe un efecto nivel sobre los colegios alrededor de las bibliotecas si no si además de este hay un efecto de interacción de las bibliotecas con colegios de ciertas características (e.g público, jornada de la mañana, pequeños, entre otros). En el Cuadro 8 se muestran los resultados de la estimación de *DD-BO* bajo la especificación continua desarrollada por (Ñopo, 2008) el impacto de las mega-bibliotecas sigue siendo positivo con un valor bastante volátil que oscila entre 0.126 *DE* (en el 2005) y 1.42 *DE* (en el 2003) sin embargo ningún resultado es significativo.

 Cuadro 8: *DD* con Blinder Oxaca continuo

Tratados	2003	2004	2005	2006	2007	2008
<b>Resultado Total</b>						
<b>70</b>	1,4263 (0,8790)	0,4479 (0,8384)	0,1267 (0,8647)	0,1512 (0,8672)	0,0573 (0,8563)	0,5119 (0,8267)

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

Por último, en el Cuadro 10 se realiza la especificación de *DD-BO* para múltiples rangos de distancia, en estos se valida nuevamente un impacto positivo pero no significativo y mucho más volátil que en el caso en que se observaba únicamente el efecto nivel bajo la especificación de *DD* del Cuadro 9.

Cuadro 10: DD con Blinder Oaxaca Discreto

Trat./Contr.		2003	2004	2005	2006	2007	2008
<b>Resultado Total</b>							
R=750	10/182	0;1064 (0;2282)	0;2107 (0;2177)	0;5276 (0;2213)	0;2876 (0;2125)	0;1480 (0;2062)	0;1940 (0;2003)
R=1000	19/173	0;1469 (0;1733)	0;1753 (0;1651)	0;1762 (0;1706)	0;1804 (0;1614)	0;0528 (0;1629)	0;1535 (0;1579)
R=1250	27/165	0;1111 (0;1580)	0;1465 (0;1494)	0;0746 (0;1543)	0;1324 (0;1478)	0;0078 (0;1526)	0;1292 (0;1467)
R=1500	44/148	0;0309 (0;1457)	0;0893 (0;1370)	0;0040 (0;1404)	0;0922 (0;1380)	0;0604 (0;1404)	0;0938 (0;1375)

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

Existen tres hipótesis que explican la poca significancia de las estimaciones: la primera se refiere a la no viabilidad de hacer la estimación a nivel de estudiantes, por ende la ausencia de contar con un mayor número de observaciones y además diferentes controles pudieron haber contribuido a explicar la varianza tan alta que presentan los resultados a nivel de colegios. La segunda se refiere al poco número de observaciones lo cual se deriva de contar con solo 2 bibliotecas y además de contar con una base de datos que no alcanza a pegar todos los colegios en una estimación de carácter tan local que los pocos datos que no pegan hacen crucial la capacidad de inferencia, se reitera que este problema es mas significativo en este documento que en cualquier otro pues la pregunta de investigación se desarrolla a un nivel local que no supero los 3500 metros alrededor de una biblioteca. La tercera hipótesis se refiere a que una vez se cuente con las dos anteriores sugerencias, el siguiente paso es incluir variables que permitan medir la interacción efectiva que estudiantes y colegios tienen con las bibliotecas como una medida más precisa de impacto.

## 7.6. 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 *mega-bibliotecas*, es decir, si bien la política pública jamás definió un perfil específico de colegios para construir las *mega bibliotecas*, 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 *mega-bibliotecas*

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 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, aun cuando no hay resultados significativos es importante destacar que el efecto para un mismo año decrece con la distancia, lo cual muestra cómo la biblioteca tiene efectivamente un mayor impacto mientras mas cerca esté de los colegios y hogares de los estudiantes.

Como recomendación de líneas de investigación futuras en este nivel se propone que dado que puede hacerse un seguimiento de antes y después de la política pública, sería prudente contar con variables como las relaciones que tienen los colegios con las bibliotecas y la distancia a la que se encuentra el hogar de cada estudiante de la biblioteca. Por otra parte la base de C-600 presenta muchas inconsistencias en términos de la matrícula y valores no reportados los cuales requieren otras fuentes de información o una metodología para ser corregidas y tener así mas observaciones en el análisis. Estas ausencias de información estan siendo adelantadas por el ICFES quien ha apoyado y manifestado su intención de continuar apoyando esta investigación en términos de brindar la información que este a su alcance.

## Bibliografía

- **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.
- **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 tinta!. <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.
- **Giralda Giralda, 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. W**  
"oBmann (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., L. Melo, and J. Ramos** (2010). La educación en colombia: análisis del marco normativo y de los indicadores sectoriales. Revista de Economía del Rosario 9(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.
- **Leuven, E. and B. Sianesi** (2003, April). Psmatch2: Stata module to perform full mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing.
- **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.
- **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.
- **Ñopo, H.** (2008). An extension of the blinder-oaxaca decomposition to a continuum of comparison groups. *Economics Letters* 100(2), 292-296.
- **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.

- **Rosenbaum, P. and D. Rubin** (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika* 70(1), 41.
- **Rubin, D.** (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of educational Psychology* ()(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.

Otras figuras

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

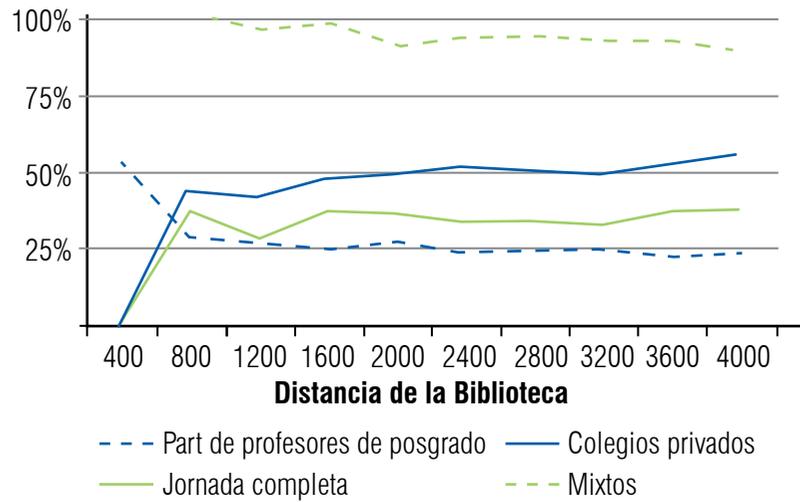


Figura 2: Características educativas de los colegios segun la distancia (2)

