La educación secundaria y sus dos dimensiones. Efectos del barrio y del colegio sobre los resultados saber 11*

Recibido: 10 de noviembre de 2016 - Aceptado: 08 de marzo de 2017

Doi: dx.doi.org/10.12804/revistas.urosario.edu.co/economia/a.6148

Jacobo Rozo Alzate†

Resumen

Este trabajo estudia los resultados en matemáticas y lenguaje de 32000 estudiantes en la prueba Saber 11 de 2008, de la ciudad de Bogotá. Este análisis tiene en cuenta que los individuos se encuentran contenidos en barrios y colegios, pero no todos los individuos del mismo barrio asisten a la misma escuela y viceversa. Con el fin de modelar esta estructura de datos, se utilizan varios modelos econométricos, incluyendo una regresión jerárquica multinivel de efectos cruzados. El objetivo central es identificar en qué medida y qué condiciones del barrio y del colegio se correlacionan con los resultados educacionales de la población objetivo y cuáles características de los barrios y de los colegios están más asociadas con el resultado en las pruebas. Se usaron los datos de la prueba Saber 11, del censo de colegios C-600, del censo poblacional de 2005 y de la policía metropolitana de Bogotá.

* Agradecimientos a CEIBA (Centro de Estudios Multidisciplinarios Básicos y Avanzados) por el financiamiento; al Icfes, a la oficina de Catastro Distrital y a la Policía Nacional por proporcionar los datos. También a María José Álvarez y a Darío Maldonado por su incondicional apoyo y por sus invaluables consejos. A su vez, agradecimientos a la Facultad de Economía de la Universidad del Rosario, en particular, a Darwin Cortez y a Juan Miguel Gallego. Por otra parte, quisiera agradecer a Carlos Sepúlveda y a Harvey Vivas por sus comentarios y retroalimentación. También quisiera agradecer a Sergio Montoya, Juan Sebastián Ordóñez y Jeisson Cárdenas por su apoyo académico y personal. Por último, quisiera agradecer a mi familia, Cesar Rozo, María Fernanda Alzate, Matías Rozo y a mis abuelos, Eufrosina Rodríguez y Horacio Alzate.

† Consultor en la ONG, Gea ambiental.

Para citar este artículo: Rozo, A. J. (2017). La educación secundaria y sus dos dimensiones. Efectos del barrio y del colegio sobre los resultados Saber 11. *Revista de Economía del Rosario*, *20*(1), 33-68.
Doi: dx.doi.org/10.12804/revistas.urosario.edu.co/economia/a.6148

© *Revista de Economía del Rosario. Universidad del Rosario.*
ISSN 0123-5362 - ISSNE 2145-454x

Las estimaciones muestran que tanto el barrio como el colegio están correlacionados con los resultados en las pruebas, pero el efecto del colegio parece ser mucho más fuerte que el del barrio. Las características del colegio que están más asociadas con el resultado en las pruebas son la educación de los profesores, la jornada, el valor de la pensión y el contexto socioeconómico del colegio. Las características de los barrios más asociadas con el resultado en las pruebas son la presencia de universitarios en la UPZ, un clúster de altos niveles de educación y nivel de crimen en el barrio que se correlaciona negativamente. Los resultados anteriores fueron hallados teniendo en cuenta controles familiares y personales.

Palabras clave: educación, modelos jerárquicos, efectos vecindario.
Clasificación JEL: C31, I24, R23

Secundary Education and its Two Dimensions, Effects of the Neighborhood and the School on the Results of the "Saber 11" Test

Abstract

This paper studies the mathematics and language results of 32,000 students on the Saber 11 test for 2008 in the city of Bogota D.C. This analysis considers that individuals are contained in neighborhoods and schools, but not all individuals from the same neighborhood attend the same school, and vice versa. With the purpose of creating a proper model for this data structure various econometric models were used, including a crossed random effect multilevel hierarchical regression. The central objective is to identify the extent to which neighborhood and schooling conditions are correlated with the educational results of the objective population, and which neighborhood and school features are more strongly associated to this test's results. We used data from the Saber 11 test, the C-600 school census, the 2005 population census and the Bogota D.C. metropolitan police department. Our estimations show that both neighborhoods and schools correlate with this test's results; but the school seems to be a much stronger factor than the neighborhood. School features that have the strongest correlation with these test's results are the teacher's education, the school day's schedule, schooling expenses and the school's socioeconomic context. Neighborhoods features that are mostly associated with these test's results are the presence of university students within the UPZ, a cluster of higher educational levels as well as the crime rate within the neighborhood, which correlates negatively. Previous results were found taking in account family and personal controls.

Keywords: Education, multilevel models, neighborhood effects.
JEL Classification: C31, I24, R23

A educação secundária e suas duas dimensões, efeitos do bairro e da escola sobre os resultados "Saber 11"

Resumo

Este trabalho estuda os resultados em matemáticas e linguagem de 32000 estudantes na prova Saber 11 do ano 2008 da cidade de Bogotá. Esta análise tem em conta que os indivíduos se encontram localizados em bairros e escolas, mas não todos os indivíduos do

mesmo bairro assistem à mesma escola e vice-versa. Com o fim de modelar esta estrutura de dados utilizam-se vários modelos econométricos, incluindo uma regressão hierárquica multinível de efeitos cruzados. Nosso objetivo central é identificar em que medida e que condições do bairro e da escola se correlacionam com os resultados educacionais da população objetivo e quais características dos bairros e das escolas estão mais associadas ao resultado nas provas. Usamos dados da prova Saber 11, do censo de escolas C-600, do censo populacional do ano 2005 e da polícia metropolitana de Bogotá. Nossas estimações mostram que tanto o bairro quanto a escola estão correlacionados com os resultados das provas; mas o efeito da escola parece ser muito mais forte que a do bairro. As características da escola que estão mais associadas com o resultado nas provas são a educação dos professores, a jornada o valor das propinas e o contexto socioeconômico da escola. As características dos bairros mais associadas com o resultado nas provas são a presença de universitários na UPZ, um cluster de altos níveis de educação e nível de crime no bairro que se correlacionam negativamente. Os resultados anteriores foram achados tendo em conta controles familiares e pessoais.

Palavras-chave: Educação, modelos hierárquicos, efeitos do bairro.
Classificação JEL: C31, I24, R23

Introducción

La educación siempre ha sido una parte vital de las sociedades humanas, es la forma por la cual se trasmite toda la acumulación de conocimiento que han adquirido las generaciones anteriores; conocimiento que es en últimas la herramienta más poderosa del ser humano. Pero, en la sociedad moderna cobra aún más importancia, convirtiéndose en la forma más importante de ascensión social, es por esto que los estudios sobre educación han adquirido tanta relevancia. De igual manera, encontrar los factores que afectan el aprendizaje de un niño o un joven no es nada sencillo, estos dependen en gran medida de cualidades personales y familiares. El estudio de Patacchini y Zenou (2009) muestra que no solo importa el contexto del estudiante sino también la disposición y actitud de los padres hacia el aprendizaje de su hijo. Dado lo anterior, es importante usar las herramientas adecuadas para poder analizar los diferentes contextos a los que se ve expuesto el individuo, y así encontrar indicios sobre las desigualdades educativas, es por ello que además de controlar por condiciones familiares es válido tener en cuenta cuál es el efecto del colegio en los resultados educativos y la posibilidad de acción a ese nivel. Por otro lado, tener en cuenta el efecto del vecindario del individuo y analizar qué condiciones de este favorecen o perjudican el rendimiento educativo.

Siguiendo esta idea, se puede decir que los estudios sobre educación no pueden solo tener en cuenta los factores que la repercuten directamente, como la educación de los profesores o la educación de los padres, también

se deben tener en cuenta variables contextuales como la violencia del barrio (Hardign, 2009); pues, ya sea por efecto de los pares, por presión colectiva del grupo social o por un modelo de vida, es muy posible que condiciones del vecindario también afecten los resultados educacionales de los individuos. En esta línea, se encuentra el trabajo de Garner y Radenbush (1991), quienes encuentran que tras controlar por la habilidad y condiciones educativas del hogar y de la escuela de un individuo, las condiciones del vecindario, como mayor desempleo o presencia de pobreza, afectan negativamente los resultados educativos.

A priori, hay muchos factores en juego a la hora de predecir los resultados en educación de un estudiante. La hipótesis de esta investigación es que los factores contextuales son muy importantes, en particular, se hará foco en las condiciones contextuales del colegio y del barrio, las cuales posiblemente están asociadas con el aprendizaje en edades escolares. En términos más precisos, el ejercicio de esta investigación consiste en estudiar la contribución del barrio y del colegio sobre los resultados en matemáticas y lenguaje de una prueba estandarizada que presentan todos los estudiantes colombianos al finalizar la secundaria (prueba Saber 11 del Icfes). El ejercicio impone un reto metodológico: poder modelar los resultados educativos teniendo en cuenta una doble anidación (en barrios y colegios) cruzada, lo cual se refiere a que no todos los estudiantes de un mismo barrio asisten al mismo colegio y viceversa.

Específicamente, no existe un consenso sobre cuáles son las condiciones de los colegios que promueven o detienen el progreso educativo de un estudiante; por ejemplo, no se sabe si una composición etaria o niveles socioeconómicos más homogéneos son positivos o negativos para los resultados educativos (Hoxby & Weingarth, 2006; Owens, 2010). Sobre las condiciones del vecindario, el debate está aún más abierto, no se sabe si la segregación es positiva o negativa, o incluso si el barrio tiene algún efecto sobre la educación en lo absoluto (Goux & Maurin, 2003; Hoxby, Olmo & Weingarth, 2013; Kaztman & Retamoso, 2007). De igual manera, existen estudios como los de Manski (1993) y Lawrence y Blume et al. (2010) que abogan por la existencia de efectos par, los cuales llevan a pensar sobre la existencia de externalidades positivas en los colegios y la existencia de efecto vecindario. Por otro lado, autores como Bénabou (2002) y Hanushek y Yilmaz (2007) estudian sobre el efecto de la segregación en la educación, específicamente en estructuras de financiamiento educativo local, donde los estudiantes y sus colegios están anidados en barrios.

La estructura de datos de esta investigación exige que se tenga en cuenta simultáneamente la anidación de los datos en barrios, colegios y los cruces entre barrios y colegios, es por esto que se eligieron los modelos jerárquicos

multinivel de efectos cruzados. Para entrar en detalle, la estructura educativa en Bogotá permite medir el efecto vecindario en estudiantes que habitan y no habitan en el vecindario de su institución educativa, esto es un avance respecto a otros estudios de educación que tengan en cuenta las condiciones del barrio, pues la mayoría de estos los estudiantes habitan en barrio donde se ubica el colegio. No tener en cuenta esta estructura podría llevar a calcular erróneamente la varianza del modelo (Raudenbush & Bryck, 2002). Además, permite suponer que ciertas variables, como la educación de los padres, no tiene el mismo efecto para los distintos grupos, con lo que se halla una pendiente distinta para cada uno de ellos. El método de estimación de esta investigación modela los tres grupos (colegio, barrio y los cruces entre colegio y barrio) con un término aleatorio similar al error, se ahondará en este aspecto más adelante.

El problema es computacionalmente muy exigente, lo que hace necesario una aproximación parsimoniosa. Por esto, se propone una estimación en dos etapas. En la primera etapa se descompondrá la varianza del resultado en la prueba estandarizada en el componente atribuible al barrio y al colegio, lo que permitirá controlar por condiciones familiares. En la segunda etapa, se toman estas varianzas y se buscan las variables de cada una de las dos agrupaciones (barrios o colegios) que están más correlacionadas con la misma varianza. En la base de datos se tienen variables familiares tomadas del cuestionario adjunto a la prueba Saber 11, también se tienen variables a nivel de barrio tomadas del censo del 2005 y de la policía metropolitana, y para describir los colegios se tienen variables del censo a colegios C-600.

Las preguntas centrales de esta investigación apuntan a resolver: (i) si existe o no un efecto del vecindario en los resultados educativos de los jóvenes; (ii) identificar cuál de las dos anidaciones tenidas en cuenta, dada la estructura de datos, barrios y colegios, explica una mayor porción de la varianza; y (iii) qué variables tanto de los colegios como de los barrios afectan los resultados en educación.

Los principales resultados van en la misma línea de la literatura, sobre las variables familiares se puede ver que la educación de los padres y el nivel de ingresos afectan positivamente los resultados en las pruebas; además, estudiantes en extra edad y estudiantes con hermanos que abandonaron el colegio tienen peores resultados. Por otra parte, el nivel educativo de los profesores y la educación agregada de los padres impactan positivamente los resultados del colegio.

Se encontró que efectivamente existe un efecto del vecindario en la educación, pero el efecto del colegio es mucho mayor, entrando en detalle, para lenguaje el vecindario explica casi un 8,9 % de la varianza, mientras el colegio

explica aproximadamente un 30 %; para la prueba de matemáticas se encontró que el barrio explica un 4,1 %, mientras el colegio explica un 34 % de la varianza de la variable dependiente. Lo anterior, sumado a otros resultados, lleva a pensar que la prueba de lenguaje es más susceptible a las condiciones sociales del individuo; de igual manera, el nivel individual y familiar son los más importantes, en lenguaje explican un 57 % de la varianza de esta prueba y en matemáticas explican un 56 %. Hallamos evidencia a favor de un efecto espacial del vecindario, en otras palabras, que la agrupación de vecindarios con mayores niveles de educación (clúster) genera un efecto del barrio positivo. También se puede ver que una mayor proporción de universitarios en la UPZ genera mejores condiciones para la educación a nivel de barrio, mientras que fenómenos como el crimen son perjudiciales.

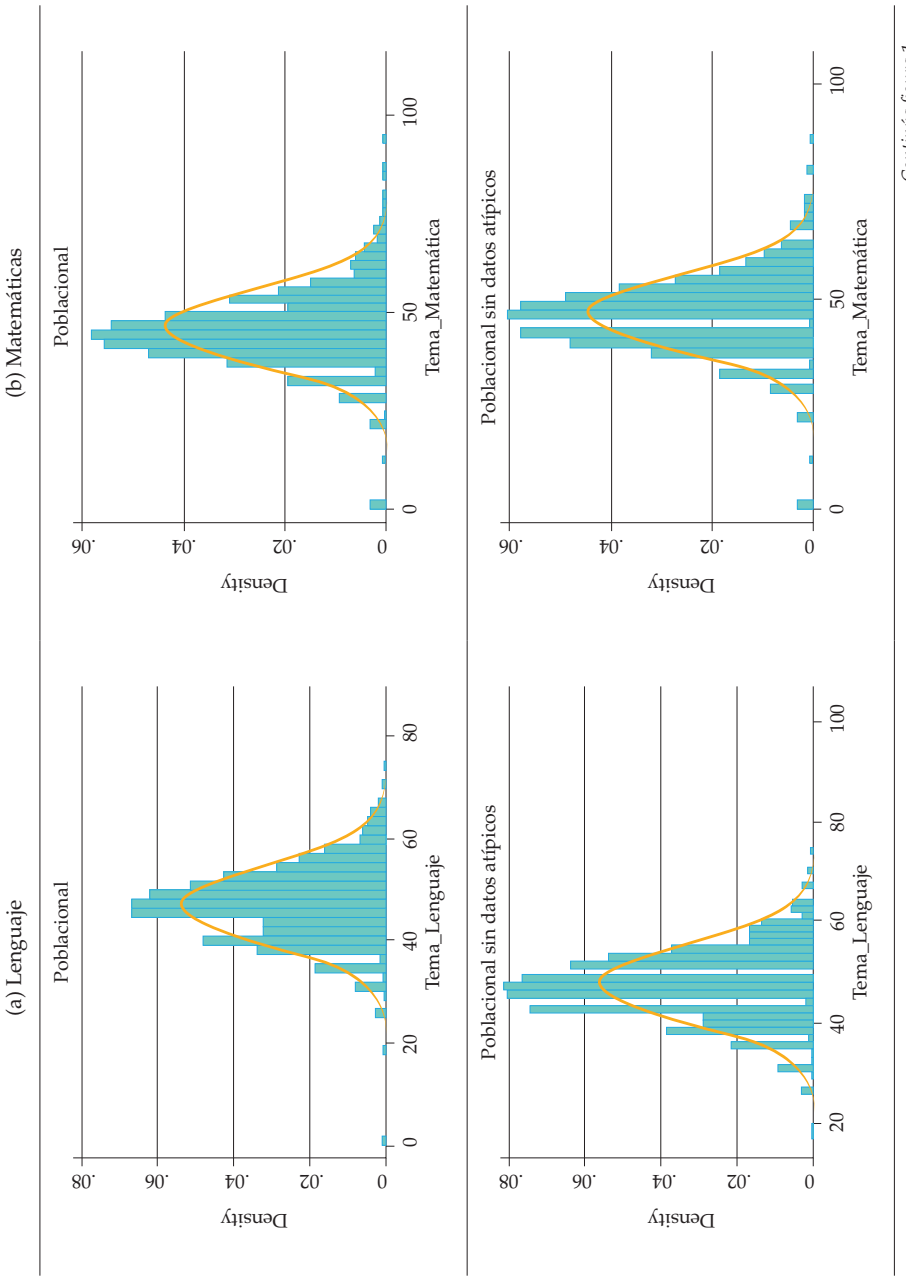
1. Estadísticas descriptivas

1.1. Descripción de la base de datos

La base de datos se compone de cuatro encuestas, la encuesta principal es el cuestionario socioeconómico adjunto a la prueba Saber 11, a esta se le agrega la información promedio del censo de colegios C-600 que se usó para describir las instituciones educativas. A esta base se le extraen los datos atípicos de la muestra, es decir, se eliminaron de la muestra los estudiantes con un resultado de cero en las pruebas de lenguaje o matemáticas, los estudiantes mayores de 26 años, los estudiantes que no registraban ningún dato en la educación de alguno de los padres y los estudiantes que asisten a jornadas nocturnas o sabatinas. Por último, para describir el vecindario de los individuos, se usó el censo poblacional del 2005 y los datos de la policía metropolitana de Bogotá. La siguiente tabla muestra el número de observaciones según se van incluyendo las bases:

Cuadro 1. Construcción de las bases de datos

Bases de datos	Número de observaciones
Sb11	95,766
C600	91,232
Sin datos atípicos	76,536
Geo-referenciados (censo)	32,692



Continúa figura 1

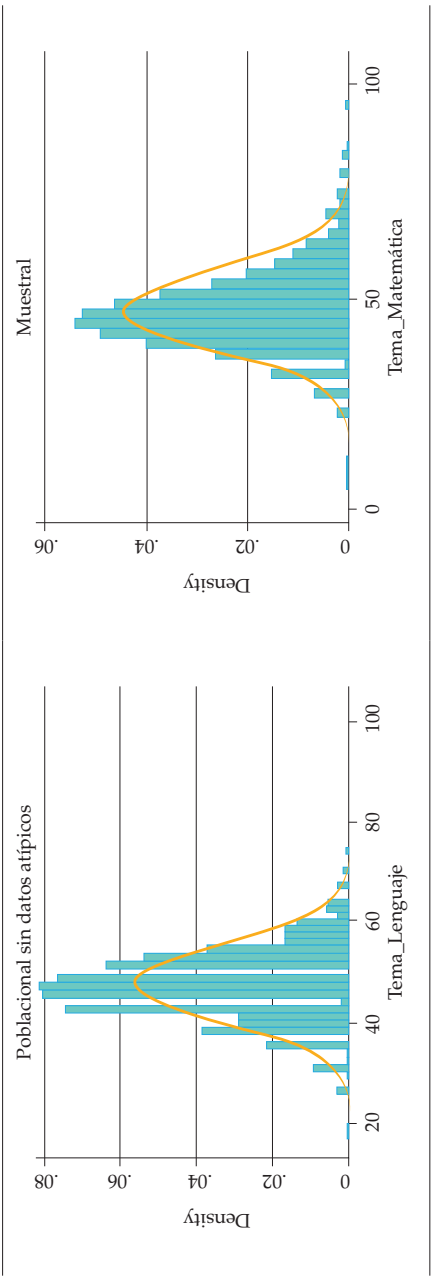


Figura 1. Histogramas de las variables dependientes

Como se mostró anteriormente, se pierde una buena cantidad de datos por la presencia de datos atípicos y por el proceso de georreferenciación. Por ello, es posible que las variables de interés tengan una distribución distinta, para esto se compararán los histogramas de las variables dependientes de la muestra poblacional (sin incluir los datos atípicos y los otros individuos excluidos de la muestra) con la muestra:

Como se puede ver en los histogramas, la diferencia entre la base poblacional y la submuestra no es radical, en especial en la prueba de lenguaje; dado esto, se puede suponer que la pérdida de datos no implica mayor sesgo. De igual manera, se realizaron varios ejercicios de regresión comparando la base poblacional, la base poblacional sin datos atípicos y la base muestral. Los resultados de esta comparación no muestran mayor diferencia, la magnitud y la significancia de los estimadores es similar; incluso la varianza explicada por la anidación de los datos en colegios es prácticamente igual.

1.2. Análisis gráfico

A continuación, se analizará la distribución espacial entre los resultados en las pruebas de lenguaje y matemáticas, para ello se utilizarán mapas con UPZ como unidad de medida (más adelante se explicarán detalladamente las dimensiones y utilidad de elegir esta unidad de medida).

En los mapas anteriores se puede ver un efecto de segregación norte-sur respecto al resultado en ambas pruebas, este resultado es similar al de Aliaga y Álvarez (2010), quienes encuentran que efectivamente existe segregación residencial a niveles grandes de agregación en Bogotá, no solo en términos educativos, sino también en términos laborales, donde el norte (parte de arriba del mapa) concentra población mejor educada y con mejores condiciones socioeconómicas. Dada esta estructura de vivienda en la ciudad, es válido pensar que se deben tener en cuenta zonas que agrupan mejores y peores condiciones, y que no solo la pertenencia a estos barrios influye, sino también la pertenencia a esta aglomeración de barrios puede tener efectos sobre el logro educativo.

1.3. Variables

Las variables incluidas en el modelo se dividen en tres grupos, primero están las variables a nivel familiar e individual, que son tomadas de la encuesta socioeconómica adjunta a la prueba Saber 11. En este grupo de variables están la educación de los padres medida en años de educación, que según la literatura se muestra como uno de los principales predictores de la educación

del individuo (Kaztman & Retamoso, 2007; Raudenbush, 1993). También está una *dummy* que indica si el estudiante tiene hermanos y si tienen hermanos menores de 15 años que no están estudiando, la presencia de hermanos en el hogar puede correlacionarse con mejores resultados educacionales (Bordalejo & Calero, 2009) y el nivel educativo de estos es una buena *proxy* del ambiente educativo en el hogar. Entre estas variables también hay variables dicotómicas que indican la ocupación de los jefes de hogar y el ingreso mensual del hogar, un mayor ingreso y ocupaciones de alto estatus suelen tener una relación positiva en el logro educativo (Garner & Raudenbush, 1991; Gaviria & Barrientos, 2001).

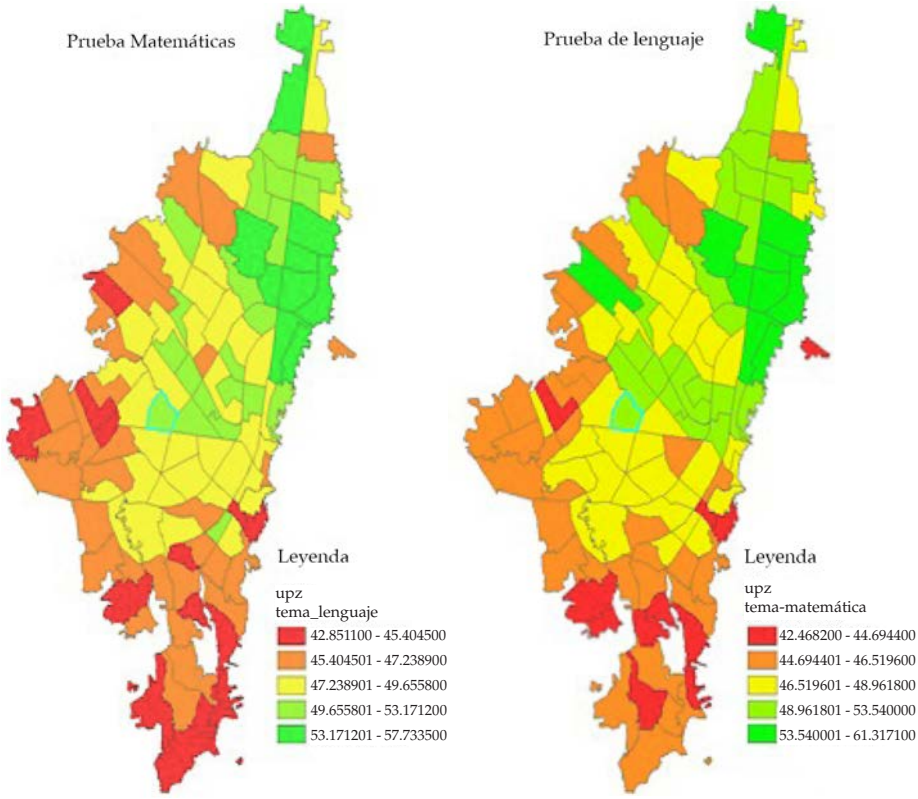


Figura 2. Resultados saber11 en Bogotá

Teniendo en cuenta el método de esta investigación, para evitar posibles sesgos de variables omitidas en la primera etapa, como la educación de los padres agregada a nivel de colegio, es importante hacer que las variables no dicotómicas tengan media cero, ya que esto permite que la correlación entre las variables familiares y las variables agregadas sea igual a cero y de esta

manera se calcula con mayor precisión la porción de la varianza explicada por barrios y colegios.

A la hora de centrar variables con una estructura de datos anidados existen dos opciones, centrar las variables respecto a la media general de los datos o centrar las variables respecto a la media del grupo o nivel de anidación. Para este trabajo, se prefirió centrar las variables respecto a la media general de los datos (respecto a la media general de la muestra, si se habla de la base muestral; y respecto a la media general de la base poblacional sin datos atípicos, si se habla de esta base de datos), esto se debe principalmente a dos motivos: (i) al tener dos niveles de anidación la decisión de centrar los datos respecto a la media de alguno de estos dos niveles no es clara y puede sesgar los resultados al usar el modelo completo que tiene en cuenta la agrupación en barrios y colegios, y (ii) según Hofman y Mark (1998), el centrar respecto a la media general de la muestra permite hacer un mejor cálculo de la varianza y, por lo tanto, de los efectos aleatorios, esto es sumamente importante si se tiene en cuenta que el método de esta investigación de estimación funciona en dos etapas y se necesita un cálculo lo más preciso posible de los efectos aleatorios.

La literatura dice que hay características de los colegios que impactan el aprendizaje, estas variables a nivel de colegio son tomadas del censo educativo C-600, entre estas se tiene la educación de los profesores, que está especificada como la proporción de profesores a nivel de colegio que alcanzaron cada uno de los niveles educativos diferentes, por ejemplo, la proporción de profesores que tiene posgrado en el colegio (García, Maldonado, Perry, Rodríguez & Saavedra, 2014). También se construyó un índice que tiene en cuenta si la jornada del colegio es ordinaria, la naturaleza del colegio (público o privado), y una medida de la pensión del colegio que es válida según la prueba de Alpha de Cobranch; este índicese realizó por problemas de multicolinealidad en la segunda etapa. Por otra parte, se tienen distintas especificaciones donde se incluyen la educación de los padres agregada, el resultado en la prueba agregada (también el rezago de las pruebas en matemáticas y lenguaje del año 2006) y el cambio de variaciones en el resultado de la prueba (matemáticas o lenguaje) y de la edad de los estudiantes a nivel de colegio (Hoxby & Weingarth, 2006).

Por último, se tienen variables a nivel de barrio tomadas del censo poblacional del 2005, entre las cuales se encuentran la cantidad de universitarios a nivel de UPZ. Por su parte, Goux y Maurin (2003) muestran que la presencia de personas con diploma universitario es un factor determinante del efecto vecindario sobre la educación. Además de las variables anteriores, se tienen datos de homicidios agregados por UPZ, tomados de los datos anuales de la policía nacional (Formisano, 2002; Harding, 2009).

También se incluye un índice que muestra la presencia local de clúster a nivel de UPZ de altos resultados, de bajos resultados y de UPZ de altos resultados rodeadas de UPZ de bajos resultados y viceversa, estas variables son dicotómicas y se construyeron usando el Local Morans Index. Lo anterior se hace con el ánimo de identificar si existe un efecto espacial sobre la educación, lo cual cobra validez si, como se mencionó, se tiene en cuenta que existe segregación a gran escala entre el sur y el norte de la ciudad, pues es importante identificar si el individuo se encuentra en una zona que grupa buenas o malas características barriales. Aliaga y Álvarez (2010) también encuentran que a una escala más pequeña Bogotá presenta otra estructura de segregación con un patrón menos claro, este fenómeno también se está controlando al incluir el Local Morans Index, pues este índice tiene en cuenta cada barrio o UPZ en comparación con sus vecinos.

1.4. Estructura de los datos

Como se verá más adelante, los ejercicios empíricos tendrán en cuenta la agrupación de individuos en barrios y colegios. La base de datos de este estudio se compone de 32692 estudiantes, los cuales se encuentran contenidos en 110 UPZ (barrios) de las 117 UPZ en Bogotá y 1181 colegios (esto es teniendo en cuenta cada sede por separado), no sobra aclarar que en la mayoría de estudios sobre educación que tienen en cuenta las condiciones del barrio, los estudiantes viven en el mismo barrio del colegio, por otro lado este estudio tiene tanto estudiantes que viven en el mismo barrio del colegio como estudiantes que no; idealmente el tener esta estructura de datos permite medir con mayor rigurosidad el efecto del vecindario.

En los ejercicios no solo va a importar la pertenencia a un barrio o a un colegio, sino también el traslape entre estos dos grupos. En otras palabras, dado que no todos los estudiantes de un colegio viven en la misma UPZ y viceversa, se generan múltiples grupos de estudiantes que asisten a la misma escuela y viven en el mismo barrio, es decir, los modelos generan un nuevo nivel de anidación para los estudiantes que asisten al mismo colegio y viven en el mismo barrio, donde no necesariamente el colegio queda en el mismo barrio donde habitan. Para ser más específicos, en los datos existen 9571 cruces entre barrios y colegios (que de ahora en adelante denominaremos celdas), las cuales tienen en promedio 15 estudiantes (mínimo tienen un estudiante y máximo 117); de igual manera, la desviación típica es de 18,6 estudiantes.

El proceso para elegir la medida espacial que se va a tomar como barrio, se basó en la convergencia del modelo de efectos cruzados, en otras palabras, si se elige una medida espacial más pequeña de UPZ, el número de grupos se

aumenta significativamente y por lo tanto el número de celdas. Esto implica que el modelo va a tener que calcular un mayor número de efectos aleatorios con grupos más pequeños y una muestra más desbalanceada, lo que hace más difícil la convergencia del modelo. Se puede pensar que una medida espacial más pequeña podría ser más adecuada para desentrañar el efecto de las interacciones sociales a nivel de barrio, pero es difícil definir el tamaño del "hábitat" social de un individuo; además de lo anterior, el nivel de UPZ es más adecuado para controlar por las dotaciones físicas del barrio, esto se debe a que la asignación de recursos públicos, políticas educativas y otros elementos del ordenamiento territorial se organizan a nivel de UPZ[1].

Respecto a lo anterior, uno de los canales de transmisión del efecto vecindario en la educación más estudiado por la literatura, se da por medio del financiamiento a nivel de barrio de las escuelas; esta estructura de financiamiento de la educación ha sido estudiada por varios autores, entre ellos Bénabou (2002), quien encuentra un efecto negativo de la segregación espacial a la hora de financiar la educación, pues los hogares más educados de mayores ingresos tienden a agruparse, lo que deja los barrios de los hogares peor educados con peores fuentes de financiación, este efecto se puede contrarrestar utilizando un mecanismo como subsidios e impuestos. De igual manera, Hanushek y Yilmaz (2007) analizan este problema de hogares heterogéneos en la valoración de la educación y advierten sobre el efecto que puedan tener algunas de estas políticas redistributivas. Afortunadamente, en la base de datos las escuelas no se financian de esta manera, lo que permite un mejor estudio del efecto vecindario, pues el efecto del barrio y del colegio se pueden separar. A pesar de esto, al anidar los datos a nivel de UPZ se está controlando por las dotaciones físicas del barrio, como bibliotecas, centros comerciales, centros culturales, parques, bares, entre otros posibles aspectos del barrio que se determinan por el plan de ordenamiento territorial y que, dado el caso, pudieran afectar el logro educativo de un estudiante.

2. Modelos con un nivel de anidación

2.1. Apuntes sobre los modelos jerárquicos multi-nivel

Antes de continuar con nuestros análisis preliminares, es importante revisar un poco la teoría de los modelos multinivel para tener claridad sobre cómo se

1 Lo anterior está legislado en el Decreto 190 del 2004, más específicamente en Artículo 46 de ese documento.

debe hacer la comparación con los modelos panel que son nuestro referente. Los modelos multinivel se estiman por medio de una máxima o cuasimáxima verosimilitud, para este estudio se prefiere la estimación por cuasimáxima verosimilitud. A pesar de que ambos métodos de estimación no son muy distintos a la hora de encontrar los parámetros fijos (por ejemplo, el estimador de la educación de los padres), son distintos a la hora de estimar los efectos aleatorios; pues mientras el estimador por cuasimáxima verosimilitud tiene en cuenta los grados de libertad perdidos por los parámetros a estimar (variables explicativas), el método de máxima verosimilitud no. Esto hace que el método de máxima verosimilitud tenga un sesgo hacia el límite inferior al estimar los efectos aleatorios (Snijders & Bosker, 1999).

El objetivo de estos modelos es tener en cuenta la anidación de los datos en grupos para descomponer la varianza de la variable dependiente en los distintos niveles de agregación, esto permite saber qué porción de la varianza no explicada por las covariantes se encuentra explicada por cada uno de los niveles de agregación. Para ilustrar lo anterior, se usará de ejemplo la estructura del modelo objetivo. Los modelos jerárquicos de efectos cruzados, con variables explicativas al primer nivel tienen la siguiente forma:

Dónde:

$$Yijk = \gamma 0jk + \gamma 1jk\Omega ijk + eijk \quad (1)$$

$$\gamma 0jk = \alpha 0 + b00j + c00k + v0jk \quad (2)$$

$$\gamma 1jk = \alpha 1 + b10j + c10k + v1jk \quad (3)$$

Reemplazando:

$$Yijk = \alpha 0 + (\alpha 1 + b10j + c10k + v1jk)\Omega ijk + b00j + c00k + v0jk + eijk \quad (4)$$

La primera ecuación muestra cómo sería la estimación por OLS. La segunda ecuación muestra la descomposición de la varianza del intercepto de la primera ecuación, en los efectos de barrio y colegio, el cruce específico de barrio y colegio en el que se encuentra el individuo (el efecto de la celda anteriormente explicado). La última ecuación muestra el modelo a estimar. Si se piensa en un modelo de panel, la ecuación 1 sería el modelo "within" y la ecuación 2 sería el modelo "between". La gran diferencia con un modelo de panel sería que en los modelos panel solo se puede tener en cuenta un único

nivel de agregación para estimar el modelo "between" (Leeden, Busing & Meijer, 1997). La ecuación 3 muestra la descomposición de la varianza del estimador asociado con las variables familiares en la primera ecuación, según la anidación de los datos en barrios y colegios y la celda a la que pertenece el individuo; lo anterior implica que el efecto de las variables familiares se puede desviar del efecto promedio según el barrio, el colegio o el cruce específico barrio-colegio del individuo, esta es otra de las diferencias de estos modelos con los modelos panel.

Para este modelo, $Yijk$ (la variable dependiente) representa el resultado en la prueba Saber 11 en matemáticas o lenguaje, en donde los subíndices i se refieren al individuo, j al barrio y k al colegio; $\alpha 0$ es el intercepto, $c00k$ es el efecto aleatorio asociado con el nivel de colegio, $b00j$ es el efecto aleatorio a nivel de barrio, y $v0jk$ es la interacción del efecto aleatorio asociado con asistir al colegio j y vivir en el barrio k (la variación intracelda). Estos modelos sufren de problemas de convergencia a medida que se tiene un mayor número de grupos no balanceados, lo anterior se refleja particularmente en el hecho de que $v0jk$ va a ser un conjunto de efectos aleatorios por cada cruce de barrio y de colegio presente en la muestra o celda como se definió anteriormente. Por otra parte, $eijk$ es el término de error, y $\gamma 0jk$ sería la constante asociada con un modelo OLS (*naive*).

El término Ωijk es un vector que incluye todas las variables familiares. Por otra parte, el término $\alpha 1$ es el efecto promedio (o estimador promedio) de las variables a nivel de familia. Respecto a la parte aleatoria del modelo, el término $b10j$ indica las perturbaciones del efecto promedio de las variables familiares a nivel de barrio, el término $c10k$ indica las perturbaciones del efecto promedio de las variables familiares a nivel de colegio, y el término $v1jk$ indica las perturbaciones del efecto promedio de las variables familiares a nivel de celda. si cualquiera de los tres términos anteriores es distinto a cero, implica que este modelo tiene pendientes aleatorias referentes a las condiciones familiares.

El método de este análisis permitirá obtener los efectos aleatorios asociados con el nivel de barrio y de colegio al eliminar la parte de la varianza de la variable dependiente explicada por las condiciones familiares (observables). Para los estimadores de las variables de familia se permitirá la presencia de efectos aleatorios a nivel de colegio que hasta el momento es el nivel de anidación que más explica la varianza de la variable dependiente, el no tener en cuenta estos cambios en las pendientes podría llevar a no estimar correctamente la varianza del modelo.

No sobra aclarar que en esta especificación el intercepto y la pendiente son aleatorias; la interpretación de lo anterior, como lo muestran Albright y

Marinova (2015), es básicamente que cada institución educativa o cada UPZ puede tener un efecto fijo o pendiente distinta. De igual manera, en este tipo de regresiones existe una constante general para toda la muestra. Lo anterior tiene mucho sentido si se tiene en cuenta que los distintos grupos no son réplicas independientes de la misma estructura, en otras palabras, estos modelos permiten suponer que tanto los barrios como los colegios pueden tener un efecto distinto sobre la educación y una pendiente distinta para el efecto de las variables familiares (esto se llama heterogeneidad de las regresiones).

2.2. Modelos vacíos de un nivel

A continuación, se compararán los modelos panel de efectos fijos y aleatorios sin variables explicativas con los modelos jerárquicos de un nivel, lo importante es comparar las distintas estimaciones de los efectos entre grupos y al interior de los grupos. En las siguientes tablas se muestran los resultados en matemáticas y lenguaje para barrios y colegios de la muestra.

Cuadro 2. Comparación de modelos, barrio y colegio (Panel de efectos fijos, aleatorios y jerárquicos de un nivel)

	Efectos Aleatorios	Efectos Fijos	Multinivel	Cuasi Máx Vero	Efectos Aleatorios	Efectos Fijos	Multinivel	Cuasi Máx Vero
	A nivel de colegio Modelo con controles de sexo y edad							
	Leng.	Mate.	Leng.	Mate.	Leng.	Mate.	Leng.	Mate.
between	3,842	4,766	4,445	5,563	3,774	4,967	3,774	4,967
within	6,135	7,661	6,135	7,661	6,136	7,659	6,136	7,659
	Modelo vacío							
between	3,852	4,752	4,406	5,47	3,699	4,853	3,699	4,853
within	6,172	7,818	6,172	7,818	6,174	7,815	6,174	7,815
	A nivel de barrio Modelo con controles de sexo y edad							
	Leng.	Mate.	Leng.	Mate.	Leng.	Mate.	Leng.	Mate.
between	2,935	3,848	3,13	4,102	2,97	3,86	2,97	3,86
within	6,689	8,737	6,689	8,737	6,689	8,737	6,689	8,737
	Modelo vacío							
between	2,468	3,134	3,245	4,218	3,097	3,981	3,097	3,981
with in	6,656	8,596	6,656	8,596	6,656	8,596	6,656	8,596

En las tablas anteriores se puede notar que las diferencias entre los modelos panel con efectos aleatorios y los modelos jerárquicos de un nivel no son muy grandes, pero las diferencias entre modelos jerárquicos de un nivel y los modelos con efectos fijos sí lo son. Según Snijders y Bosker (1999), estas diferencias entre efectos aleatorios pueden generarse por lo desbalanceado de los distintos grupos de colegios y barrios (en otras palabras, existen grupos con muy pocos individuos y grupos con más de 100 individuos). A la hora de tratar este problema, los modelos con efectos aleatorios son más adecuados que los modelos con efectos fijos, puesto que los modelos de efectos fijos estiman cada uno de los errores aparte, entonces los coeficientes serán sobreestimados en los casos de los grupos pequeños, por falta de información, esto hace que los parámetros tengan grandes errores *standard*. Por otra parte, los efectos aleatorios tienen en cuenta el supuesto de que los efectos de grupo son independientes e idénticamente distribuidos, en otras palabras, para el cálculo tienen en cuenta la distribución de los datos; esto permite contrarrestar la escasez de casos, lo que hace la inferencia más precisa.

2.3. Prueba de anidación

Antes de continuar con la estimación de los modelos a un nivel, se debe poner a prueba que el pertenecer a un grupo sí tiene un efecto sobre los resultados en las pruebas Saber 11, esto se realiza al comparar un modelo vacío con un modelo vacío anidado por medio de un test de tasa de verosimilitud (Likehood-ratio test), más específicamente se está comparando el logaritmo de la función de verosimilitud maximizada de un modelo con anidación y uno sin anidación (Rabe-Hesketh Skrondal, 2008).

Cuadro 3. Pruebas de anidación

Matemática - Barrios	LR = 4318,68 Valor p = 0,00
Lenguaje - Barrios	LR = 4445,35 Valor p = 0,00
Matemática - Colegios	LR = 9461,1 Valor p = 0,00
Lenguaje - Colegios	LR = 7618,05 Valor p = 0,00

(Supuesto: modelo anidado en barrios barrio no contiene el modelo sin anidación)

A partir de la tabla anterior, se puede decir que, tanto para matemáticas como para lenguaje, la anidación en barrios y colegios sí explica parte de la

varianza de la variable dependiente. Se podría pensar que estos modelos jerárquicos de un nivel contienen a los modelos sin anidación. El paso a seguir es estimar modelos jerárquicos de un solo nivel con variables explicativas a nivel familiar y personal. Sobre la especificación de esta primera etapa se hicieron las pruebas correspondientes y no se encontró evidencia de efectos no lineales, en este aspecto Cook et al. (2002) encontraron que tanto pares, colegios, barrios y familia eran significativos, pero que los efectos de estas variables eran aditivos.

2.4. Regresiones con variables explicativas y efectos aleatorios, modelos de un nivel

En las regresiones de la tabla 4 se puede ver que, para casi todas las variables, tanto el signo como la significancia de estas es igual si se usa la muestra poblacional sin datos atípicos que si se usa la muestra final. Entrando más en detalle, se puede ver que la regresión replica algunas de las evidencias empíricas halladas por otros autores (Mediavilla & Martínez, 2009; Ready, 2010), más detalladamente se puede ver que el ser hombre tiene una relación positiva con las pruebas de matemáticas, al comparar el resultado de ser hombre en la prueba de lenguaje respecto a una regresión *naive*, se puede ver que deja de ser significativo, es posible que esto se deba a una mejor especificación de la varianza del modelo; también se puede ver que los estudiantes en extra edad y/o que tienen hermanos desertores (menores de 15 años) tienden a tener peores resultados en ambas pruebas. Por último, se puede ver que a mayor educación e ingreso de los padres mejores resultados en las pruebas Saber 11.

Cuadro 4. Modelo a un nivel – colegios

	Muestral		Poblacional	
	lenguaje	matemática	lenguaje	matemática
Sexo	-0,0341 (0,0756)	2,804*** (0,0949)	-0,117** (0,0476)	2,683*** (0,0584)
Edad centrada del estudiante	-0,518*** (0,0331)	-0,715*** (0,0416)	-0,524*** (0,0205)	-0,697*** (0,0252)
Años de educación de la madre centrados	0,103*** (0,0130)	0,119*** (0,0163)	0,0845*** (0,00804)	0,0929*** (0,00987)
Años de educación del padre centrados	0,145*** (0,0122)	0,110*** (0,0153)	0,117*** (0,00756)	0,0907*** (0,00928)
Hermano desertor menor de 15 años	-0,772*** (0,179)	-0,613*** (0,224)	-0,731*** (0,105)	-0,586*** (0,129)

Continúa

	Muestral		Poblacional	
	lenguaje	matemática	lenguaje	matemática
Sin hermanos	-0,0318 (0,0750)	-0,201** (0,0939)	-0,0280 (0,0485)	-0,218*** (0,0594)
Ingreso familiar centrado	0,507*** (0,0344)	0,595*** (0,0435)	0,420*** (0,0234)	0,473*** (0,0290)
Constante	48,24*** (0,161)	46,03*** (0,212)	47,98*** (0,110)	45,84*** (0,148)
varianza entre colegios (between)	2,661	3,984	2,868	4,146
varianza al interior (within)	6,110	7,637	6,124	7,505
Controles laborales	si	si	si	si
Observations	32,692	32,692	79,231	79,231
Numero de Sedes	1,181	1,181	1,441	1,441

Se han excluido datos atípicos para las estimaciones. Errores estándar en paréntesis.

*** p<0,01, ** p<0,05, * p<0,1

Ahora, se analizarán los resultados teniendo en cuenta la anidación en barrios: los resultados de las regresiones de la tabla 5, que controlan por variaciones barriales, son similares a los encontrados en los modelos anteriores, en otras palabras, la educación de los padres y su ingreso tienen un efecto positivo; y el estar en extra edad y tener hermanos desertores tiene un efecto negativo. Por otra parte, el efecto del género es el mismo, positivo para hombres en la prueba de matemáticas. Sobre el género, en la prueba de lenguaje, al comparar el modelo jerárquico de un nivel anidado en colegios con el modelo jerárquico de un nivel anidado en barrios y con el modelo OLS (*naive*) se puede ver que desaparece el efecto positivo de ser mujer.

Cuadro 5. Modelo a un nivel – barrios

	Muestral	
	lenguaje	matemática
Sexo	-0,222*** (0,0718)	2,660*** (0,0931)
Edad centrada del estudiante	-0,462*** (0,0328)	-0,585*** (0,0425)
Años de educación de la madre centrados	0,154*** (0,0133)	0,194*** (0,0173)
Años de educación del padre centrados	0,184*** (0,0125)	0,170*** (0,0163)

Continúa

	Muestral	
	lenguaje	matemática
Hermano desertor menor de 15 años	-0,944*** (0,185)	-0,808*** (0,240)
Sin hermanos	0,0292 (0,0772)	-0,189* (0,100)
Ingreso familiar centrado	0,780*** (0,0334)	1,077*** (0,0433)
Constante	48,67*** (0,203)	46,59*** (0,277)
varianza entre barrios (between)	1,445	2,043
varianza al interior del barrio (within)	6,428	8,320
Controles laborales	si	si
Observaciones	32,692	32,692
Numero de UPZ	110	110

Errores estándar en paréntesis.

*** p<0,01, ** p<0,05, * p<0,1

En las pruebas de lenguaje, es muy probable que esto se deba a no tener en cuenta el efecto de la anidación de los datos en barrios o colegios, pues según Snijders y Bosker (1999), el omitir este tipo de estructuras aumenta el error de tipo 1.

3. Regresiones con variables explicativas y efectos aleatorios cruzados, modelos multi nivel

3.1. Prueba de doble anidación

Estos modelos son el objetivo central de este trabajo, el cual es lograr descomponer la varianza no explicada del modelo en los dos grupos de anidación, barrios y colegios. Como ya se mostró anteriormente, es válido usar este tipo de modelos con efectos aleatorios para tratar este problema. Por otra parte, las pruebas anteriores eran de suma importancia, pues la prueba de tasa de verosimilitud (Likehood-ratio test) de la que se habló anteriormente no se puede aplicar con toda libertad a modelos con más de un parámetro aleatorio. Lo anterior se debe a que se desconoce con exactitud el número de efectos aleatorios que deben ser restringidos a cero, en especial si se tiene en cuenta la existencia de covarianzas; de igual manera, si se rechaza la hipótesis nula

(que dice que los efectos aleatorios son iguales a cero), se está encontrando un límite superior, por lo tanto, si se rechaza la hipótesis nula con los grados de libertad propuestos implica que se rechaza esta hipótesis con los grados de libertad que en realidad tiene el modelo. Dado lo anterior, para estar totalmente seguros de la doble anidación que se propone como modelo ideal, se puede comparar un modelo con un único nivel de anidación, con un modelo que posee doble anidación, es decir, se estaría comparando si un modelo que solo tiene en cuenta la anidación en colegios estaría contenido por un modelo anidado en colegios y barrios; de esta manera, se reduce el número de parámetros a ser comparados, de todas maneras el estimado continua siendo un límite superior pero de esta forma no estará tan lejano:

Cuadro 6. Pruebas de anidación

Matemática - Colegios	LR = 18,01 Valor p = 0,000
Lenguaje - Colegios	LR = 71,68 Valor p = 0,000
Matemática - Barrios	LR = 5177,88 Valor p = 0,000
Lenguaje - Barrios	LR = 3249,34 Valor p = 0,000

Hipótesis: modelo vacío multi nivel no contiene modelo anidado en solo colegios.

Las pruebas anteriores muestran que la hipótesis de doble anidación es válida en todos los casos. Este procedimiento también puede ser usado para saber si la celda (los grupos generados por el cruce de barrios y colegios) explica efectivamente algo de la varianza de la variable dependiente, en particular este último nivel de anidación es sumamente interesante, pues si existe un sesgo de selección que relacione la pertenencia a estos dos grupos estaría contenido en la varianza explicada por la celda, pero en sí mismo controla por la combinación de las condiciones del barrio y el colegio. A continuación, se muestran los resultados de esta prueba:

Cuadro 7. Pruebas de anidación con celda

Matemática	LR = 7,63 Valor p = 0,006
Lenguaje	LR = 3,74 Valor p = 0,053

Hipótesis: modelo vacío multi nivel no contiene modelo anidado en solo colegios.

En este caso también se puede decir que la celda es válida tanto para matemáticas como para lenguaje. De igual manera, en el caso de la regresión de lenguaje esta prueba se encuentra justo en el límite. Dados los resultados anteriores, se puede continuar con la estimación del modelo ideal.

3.2. Resultados de los modelos con doble anidación

En los siguientes modelos se tienen dos grupos en lo que respecta a la estimación de efectos aleatorios. Modelos que solo suponen un componente aleatorio en el intercepto, en barrios, colegios y el cruce de estos dos grupos (las celdas), y modelos que además de tener estos componentes aleatorios suponen que el efecto de la educación de los padres puede variar entre colegios, en resumidas cuentas, hay un componente asociado con el estimador de la educación del padre y de la madre, que varía según colegios. Esto se supone porque en algunos estudios se ha mostrado que el efecto de las condiciones contextuales del individuo depende de la actitud de los padres respecto al aprendizaje del hijo (Patacchini & Zenou, 2009; Ready 2010), dado que no se tiene esta variable, se puede modelar si pertenecer a un colegio dado potencia o reduce el efecto de la educación de los padres sobre los resultados educacionales de su hijo.

Es posible observar en la tabla 8 que el hecho de ser hombre no tiene efecto sobre las pruebas de lenguaje, esto puede implicar que el género deja de ser importante en el aspecto del lenguaje; por otra parte, no solo este estudio ha encontrado que ser mujer se correlaciona con peores resultados educativos, entre otros está el estudio de Gaviria y Barrientos (2001), quienes muestran que ser hombre se correlaciona con mejores resultados tanto en el puntaje general como en la prueba de matemáticas y lenguaje. De igual manera, en varios de los modelos estimados anteriormente el ser mujer tenía efectos positivos sobre la prueba de lenguaje, es posible que este cambio en la significancia se deba a una correcta estimación de la varianza del modelo. Además, este resultado indica que persisten ciertas estructuras patriarcales en Bogotá. En este aspecto, Domínguez (2004) argumenta que no solo existen diferencias reales en las instituciones educativas (como normas o cursos especializados para cada género), sino que además persisten diferencias en las expectativas educativas por género.

Ahondando en las variables familiares, la edad del estudiante es una *proxy* de su desempeño académico, pues es muy posible que estudiantes en extra edad hayan reprobado algún curso, y como es de esperarse su signo es negativo. El hecho de tener un hermano desertor muestra que, por un lado, la familia puede enfrentar dificultades económicas, las cuales se relacionan

con peores resultados educativos (Brook-Gun et al., 2007) y, por otro lado, la familia valora menos la educación, condición que a su vez se relaciona con un peor desempeño (Patacchini & Zenou, 2009). Por otra parte, padres más educados y con mayores ingresos se correlacionan con mejores resultados en las pruebas. De igual manera, hay que tener cuidado con los canales de trasmisión, pues el efecto de la educación puede ser o un efecto directo de aprendizaje o un efecto de la valoración de la familia por la educación. Sobre el ingreso, el canal de transmisión tampoco es claro, según Gaviria y Barrientos (2001) puede ser un efecto de elección del colegio, pero el efecto de las condiciones materiales no debe ser despreciable.

Cuadro 8. Modelo con doble anidación, prueba de lenguaje

	Vacío		No vacío	
	Interc. Al.	Pend. Al.	Interc. Al.	Pend. Al.
Sexo			-0,0464 (0,0755)	-0,0399 (0,0755)
Edad centrada del estudiante			-0,526*** (0,0331)	-0,525*** (0,0331)
Años de educ. madre centrados			0,108*** (0,0138)	0,100*** (0,0130)
Años de educ. padre centrados			0,152*** (0,0130)	0,141*** (0,0122)
Hermano desertor < 15 años			-0,754*** (0,179)	-0,767*** (0,179)
Sin hermanos			-0,0274 (0,0751)	-0,0333 (0,0751)
Ingreso familiar centrado			0,492*** (0,0347)	0,489*** (0,0347)
Constante	48,08*** (0,146)	47,99*** (0,146)	48,38*** (0,169)	48,43*** (0,171)
Varianza UPZ	0,967	0,885	0,575	0,560
Varianza Colegio	3,212	2,814	2,511	2,345
Varianza Celda	0,558	0,503	0,411	0,411
Varianza Error	6,159	6,121	6,104	6,090
Varianza colegio – educ. padre		0,221		0,120
Varianza colegio – educ. madre		0,194		0,116
Controles laborales			Si	Si
Observaciones	32,692	32,692	32,692	32,693

Errores estándar en paréntesis. *** p<0,01, ** p<0,05, * p<0,1

Sobre los efectos aleatorios, como era de esperarse, en todas las especificaciones, el colegio explica una porción de la varianza mucho mayor que el barrio, más específicamente para el modelo completo, mientras el barrio explica casi un 8,95 % el colegio aproximadamente un 30 % y la celda cerca de un 5,1 % de la varianza de la variable dependiente. Además, se puede ver cómo la porción de la varianza explicada por colegios y barrios baja al incluir las variables familiares. También es muy interesante que al incluir un efecto aleatorio de la educación de los padres a nivel de colegio, el efecto del colegio se reduce en una medida no despreciable, esto indica que no solo importan las condiciones del colegio, sino que además el efecto del colegio se puede ver potenciado o reducido según la educación de los padres, esto puede ser producto de la valoración de la educación a nivel familiar (como el efecto de estudiar con o ser educado en parte por los padres) y de que ciertos colegios tengan programas de padres de familia.

La tabla 9 muestra los resultados para los puntajes de matemáticas. Como muestra la literatura, a los hombres les va estadísticamente mejor en matemáticas, incluso al controlar por variables familiares y por la anidación de los individuos en barrios y colegios. La edad continúa siendo una *proxy* de rendimiento académico. Por otra parte, el resto de variables familiares, como la educación de los padres o el ingreso, tienen el mismo signo que en las regresiones anteriores y magnitudes similares.

Cuadro 9. Modelo con doble anidación, prueba de matemáticas

	Vacío		No vacío	
	Interc. Al.	Pend. Al.	Interc. Al.	Pend. Al.
Sexo			2,802***	2,658***
			(0,0948)	(0,0944)
Edad centrada del estudiante			-0,717***	-0,358***
			(0,0416)	(0,0423)
Años de educ. madre centrados			0,118***	0,232***
			(0,0163)	(0,0174)
Años de educ. padre centrados			0,109***	0,209***
			(0,0153)	(0,0162)
Hermano desertor < 15 años			-0,612***	-0,845***
			(0,224)	(0,243)
Sin hermanos			-0,205**	-0,156
			(0,0939)	(0,102)
Ingreso familiar centrado			0,592***	1,491***
			(0,0436)	(0,0408)

Continúa

	Vacío		No vacío	
	Interc. Al.	Pend. Al.	Interc. Al.	Pend. Al.
Constante	47,27*** (0,164)	47,18*** (0,164)	46,13*** (0,214)	46,41*** (0,184)
Varianza upz	0,586	0,522	0,257	0,259
Varianza Colegio	4,672	4,325	3,982	3,915
Varianza Celda	0,864	0,720	0,677	0,671
Varianza Error	7,781	7,742	7,619	7,613
Varianza colegio – educ. padre		0,201		0,099
Varianza colegio – educ. madre		0,216		0,057
Controles laborales			Si	Si
Observaciones	32,692	32,692	32,692	32,692

Errores estándar en paréntesis. *** p<0,01, ** p<0,05, * p<0,1

Al igual que en la prueba de lenguaje, para todas las especificaciones, la varianza explicada por colegio es mayor, pero en este caso el efecto del barrio es radicalmente menor, más específicamente, para el modelo vacío, el barrio explica aproximadamente un 4,3 % de la varianza, mientras que el colegio explica cerca de un 34 % y la celda un 6,2 %; esto puede indicar que la transmisión del aprendizaje del lenguaje se da más por interacciones sociales (de igual manera la escuela sigue siendo lo más importante), y el efecto sobre matemáticas sea más una cuestión de calidad educativa. Dado que en la estructura educacional en Colombia el barrio de los estudiantes y el de los colegios no se superpone perfectamente, es importante comparar el efecto celda con el efecto del barrio; en el caso de la prueba de lenguaje el efecto aleatorio de la UPZ es mayor que el efecto de la celda, en cambio en la prueba de matemáticas sucede lo contrario. Esto se evidencia a favor de que el aprendizaje del lenguaje pueda depender más del contexto social del individuo, mientras que en matemáticas, a pesar de que el barrio importa, es mucho más importante la escuela, y la combinación barrio-colegio; mientras en lenguaje es más importante el barrio que la combinación barrio-colegio. Este resultado ha sido mencionado por algunos autores (Steele, Vignoles & Jenkins 2007).

Otro resultado interesante es que el efecto de la educación de los padres (el estimador) cambia según la pertenencia al colegio, más específicamente, el impacto que tiene la educación de los padres sobre los resultados en las pruebas Saber 11 cambia entre los distintos colegios (distintas pendientes).

De igual manera, la educación de la madre tiende a ser más estable entre grupos que la educación del padre, al controlar por variables familiares, lo anterior podría indicar que las madres tienden a estar más involucradas en la educación de sus hijos en promedio, en cambio en los padres hay más heterogeneidad en el efecto. Para complementar lo anterior, al extraer de la base poblacional aquellos estudiantes que no conocen la educación de sus padres, solo 770 no conocen la educación de su madre y en cambio 7778 no conocen la educación de su padre. Esto puede indicar que el abandono familiar es más común por parte del padre que de la madre, pero no solo esto, sino que podría llevar a pensar que es posible que la tarea de educar y estar pendiente de los hijos sea una tarea más femenina, lo que lleva a pensar en una estructura familiar patriarcal.

3.3. Sobre la especificación de la primera etapa y el "sorting" familiar

Uno de los principales desafíos en la estimación de efectos par y efectos vecindarios es el problema de reflexión, identificado por Manski (1993). Un modelo 'lineal en medias', en el que el efecto vecindario se mide mediante un promedio de la variable de interés grupal, no está plenamente identificado en términos de efectos endógenos (efecto del promedio del grupo de referencia en cuanto a la variable de interés) y efectos exógenos (efecto del promedio del grupo de referencia en cuanto a variables de contexto). Sin embargo, Manski (1993) y Lawrence y Blume et al. (2010) plantean que estos dos tipos de efectos no están identificados separadamente, pero sí de forma conjunta. En la primera etapa solo se estima la forma reducida del modelo, en la cual el resultado en la prueba Saber 11 para ambos temas se estima únicamente usando las variables observables de cada alumno, y no se intenta modelar el efecto endógeno, lo que implicaría agregar los resultados de la prueba Saber 11 tanto a nivel de barrio como a nivel de colegio. Por tanto, no hay problemas de identificación por cuenta de reflexión.

Otro problema al que se puede enfrentar la especificación de la primera etapa es la multicolinealidad, pues es probable que los padres a la hora de elegir su pareja busquen personas con niveles de educación similares y que, a la vez, este nivel de educación tenga una correlación directa con el ingreso familiar. En otras palabras, la educación del padre, de la madre y el ingreso del hogar pueden estar correlacionados.

Respecto a lo anterior, Fernández (2001) afirma que para Estados Unidos la correlación entre la educación de los padres es de 0,6 y para el Reino Unido es de 0,5. Para esta muestra, la correlación es de 0,688, mayor en Estados Unidos, pero al no ser tan alta no se genera multicolinealidad perfecta en

los modelos. Por otra parte, la correlación entre el ingreso familiar y la educación de la madre es de 0,546; y la correlación entre el ingreso familiar y la educación del padre es de 0,5683. Como se puede notar, es mayor la correlación entre la educación de los padres que la correlación entre la educación y el ingreso, por esto se hará foco en analizar si incluir la educación de ambos padres aporta información al modelo o si es mejor incluir un índice por los problemas ya mencionados.

Una de las posibilidades para revisar el efecto de la multicolinealidad es hacer un análisis sobre la inflación de la varianza de los estimadores, que se aumenta con esta correlación entre variables. Según Belsley, Kuh, y Welsch (1980), el valor donde un alto índice de inflación de varianza puede generar problemas de estimación es de 10; de igual manera, otros autores recomiendan para la correcta estimación de la varianza utilizar un índice de inflación de la varianza menor a 4. Para las variables usadas en los análisis anteriormente expuestos, el mayor índice de inflación fue de 3,46. Como lo recomiendan Robinson y Schumaecker (2009), este análisis se realizó utilizando las variables centradas, el resultado anterior indica que la correlación entre las variables del modelo no genera problemas de estimación de la varianza.

Continuando con este análisis, para el modelo de primer nivel es válido incluir la educación de ambos padres, solo si ambos niveles de educación tienen efectos ligeramente distintos o si la diferencia en niveles de educación afecta a los resultados educativos de los hijos. Para evaluar lo anterior, se ha decidido reemplazar la educación del padre por la resta de la educación de ambos padres, si esta variable es significativa implica que, de haber diferencias en educación en los padres, estas pueden afectar al rendimiento académico de los estudiantes.

Para la mayoría de las especificaciones, los errores típicos y el valor de los estimadores es similar, también el error del modelo y los efectos aleatorios asociados con el barrio o el colegio; esto es de esperar, pues los problemas de multicolinealidad tienden a inflar el valor de la varianza de los estimadores (Robinson & Schumaecker, 2009). Este patrón se mantiene para modelos de un nivel o de dos niveles, excepto para una variable, la educación de la madre, cuyo valor se reduce a la mitad para todas las especificaciones, pero su varianza se mantiene prácticamente igual. La variable que indica la diferencia entre la educación de los padres fue significativa para todas las especificaciones.

3.4. Apuntes sobre la segunda etapa

A partir de los efectos aleatorios anteriormente estimados, se tendrán las variables dependientes del segundo paso de estimación. Este paso consiste

en encontrar las variables más correlacionadas con el barrio y el colegio. Específicamente, se estimará el efecto aleatorio asociado con el vecindario, con las variables del vecindario, y el efecto aleatorio del colegio con las variables de colegio (modelo de intercepto aleatorio). Con estas estimaciones se podrá saber cuáles son las variables significativas y a qué nivel, pero lo más importante, se podrá comparar la significancia de los estimadores de las regresiones que usan los efectos aleatorios del modelo vacío contra los modelos que usan los efectos aleatorios del modelo con variables familiares. En este punto los modelos a estimar serían los siguientes:

$$b0\hat{}0j = A00j + A01j\ Xj + Ej$$
$$c0\hat{}0k = A00k + A01kZk + Ek$$

En este caso, la primera y segunda ecuación son los modelos a estimar por OLS, que buscan explicar los efectos aleatorios a nivel de barrio y de colegio, respectivamente; estos son los efectos aleatorios calculados para el intercepto del modelo con variables familiares.

En estas ecuaciones, *Xj* y *Zk* son vectores de variables a nivel barrio y colegio, respectivamente. Los términos *A*00*j* y *A*00*k* indican el intercepto de las regresiones; los términos *A*01*j* y *A*01*k* indican la pendiente de las regresiones (el efecto de las condiciones de barrio y colegio, respectivamente, sobre los efectos aleatorios); y, por último, *Ej* indica el error asociado con esta estimación.

Las estimaciones a continuación utilizan errores estándar robustos clúster (por estimador de White). Las estimaciones de colegio tienen en cuenta el clúster de errores a nivel de barrio y las estimaciones de barrio tienen en cuenta el clúster a nivel de colegios.

3.5. Resultados de la segunda etapa: vecindarios

La primera columna de regresión se corre contra la variable dependiente (ya sea matemáticas o lenguaje), esto se hace con el ánimo de mostrar que el efecto es robusto no solo para explicar los efectos aleatorios, sino también para explicar como tal a la variable dependiente de la primera etapa.

Las regresiones OLS anteriores muestran un efecto positivo del porcentaje de personas con grado universitario en la UPZ para todas las especificaciones, tanto en la prueba de lenguaje como de matemáticas; esto puede ser evidencia a favor de un efecto barrio por medio de un modelo (Jencks & Mayer, 1990; Wilson, 1987), lo anterior significa que si los estudiantes se encuentran rodeados por más universitarios tienen más incentivos a estudiar más y a buscar un mayor nivel educativo, lo cual implica mejores resultados.

Por otra parte, el crimen muestra tener un efecto negativo sobre los resultados, además explica parte del efecto negativo asociado con los barrios, esto se puede dar por dinámicas de estrés o de sustitución, pues el crimen se puede presentar como una opción de vida. Otros estudios también han encontrado que el crimen del barrio se traduce en condiciones desfavorables para la educación; por ejemplo, Harding (2009) encuentra que el efecto del barrio que se correlaciona con el abandono educativo se explica, para los hombres, en un 65 % por la violencia del barrio y para las mujeres en un 100 %. Por otra parte, es válido tener esta variable en cuenta, pues muchos de los homicidios ocurren en la cercanía del hogar del victimario y, por lo tanto, es un buen indicador de las condiciones del barrio (Formisano, 2002).

Cuadro 10. Segundas etapas, prueba de matemáticas – barrios

	(1)	(2)	(3)	(4)
% universitarios por UPZ	16,23***	1,717***	0,325***	0,336***
	(1,209)	(0,0498)	(0,0210)	(0,0217)
Homicidios totales en la UPZ	-0,00198*	-0,00096***	-5,53e-05**	-6,50e-05***
	(0,00106)	(8,16e-05)	(2,20e-05)	(2,27e-05)
Clúster de UPZ: Alto-Alto	3,721***	0,0986***	0,0561***	0,0577***
	(0,741)	(0,0164)	(0,00844)	(0,00867)
Clúster de UPZ: Bajo-Bajo	-0,236	-0,0648***	0,0111**	0,0103**
	(0,300)	(0,0166)	(0,00506)	(0,00518)
Clúster de UPZ: Alto-Bajo	5,403	-0,485***	-0,0789***	-0,0829***
	(3,823)	(0,0129)	(0,00546)	(0,00565)
Clúster de UPZ: Bajo-Alto	-0,659	0,0173	0,0240***	0,0240***
	(0,631)	(0,0113)	(0,00340)	(0,00350)
Constante	44,94***	-0,268***	-0,0689***	-0,0703***
	(0,220)	(0,0121)	(0,00468)	(0,00482)
Observaciones	32,616	32,616	32,616	32,616
R-squared	0,120	0,606	0,267	0,270
Variables primera etapa	Vacío	Vacío	No vacío	No vacío
Modelo primera etapa	Interc. Al.	Pend. Al.	Interc. Al.	Pend. Al.

Errores estándar en paréntesis. *** p<0.01, ** p<0.05, * p<0.1

Por último, la medida de Local Morans Index muestra que posiblemente exista un efecto espacial en lo referente al efecto del vecindario. En otras palabras, clústeres de UPZ con mayores resultados en promedio se relacionan con efecto vecindario positivo sobre los resultados en las pruebas; en cambio, clústeres de barrios con bajos resultados en las pruebas se correlacionan con

un peor efecto vecindario. El efecto de vivir en un barrio malo rodeado de barrios buenos no es robusto en las distintas regresiones. De igual manera, el hecho de que exista un efecto significativo de la existencia de clústeres puede ser indicio de que no solo importe el barrio del individuo, sino que además importan las condiciones de los barrios circundantes.

Otro resultado que da soporte a una de las hipótesis, en la que se afirma que el resultado en lenguaje depende más del efecto de las interacciones sociales que de resultados en matemáticas, es que el $R2$ de los modelos que explican el efecto vecindario es mucho mayor para la prueba de lenguaje que para la de matemáticas. Esto cobra real importancia si se tiene en cuenta que la variable de porcentaje de universitarios en la UPZ tiene una magnitud mucho mayor en la prueba de lenguaje que en la de matemáticas. En contra de esto, el efecto de los homicidios deja de ser significativo en las últimas dos especificaciones del efecto aleatorio de la prueba de lenguaje, pero uno de los canales de transmisión del efecto del crimen es el que se muestra como un sustituto a la educación, lo que hace que este resultado no invalide la hipótesis de que la prueba de lenguaje depende más del contexto social del individuo.

Cuadro 11. Segundas etapas, prueba de lenguaje – barrios

	(1)	(2)	(3)	(4)
% de universitarios por UPZ	13,80***	3,822***	1,675***	1,671***
	(0,760)	(0,0949)	(0,0578)	(0,0577)
Homicidios totales en la UP	-0,00241***	-0,000825***	2,80e-05	2,56e-05
	(0,000814)	(0,000192)	(0,000132)	(0,000130)
Clúster de UPZ: Alto-Alto	2,017***	0,311***	0,378***	0,351***
	(0,431)	(0,0296)	(0,0205)	(0,0202)
Clúster de UPZ: Bajo-Bajo	-0,794***	-0,675***	-0,308***	-0,311***
	(0,195)	(0,0295)	(0,0192)	(0,0185)
Clúster de UPZ: Alto-Bajo	2,914	-0,792***	-0,296***	-0,313***
	(1,932)	(0,0242)	(0,0155)	(0,0154)
Clúster de UPZ: Bajo-Alto	-2,592***	0,0753**	0,0566***	0,0594***
	(0,946)	(0,0322)	(0,0146)	(0,0139)
Constante	46,14***	-0,634***	-0,342***	-0,339***
	(0,158)	(0,0228)	(0,0147)	(0,0145)
Observaciones	32,616	32,616	32,616	32,616
R-squared	0,128	0,804	0,675	0,676
Variables primera etapa	Vacío	Vacío	No vacío	No vacío
Modelo primera etapa	Interc. Al.	Pend. Al.	Interc. Al.	Pend. Al.

Errores estándar en paréntesis. *** p<0.01, ** p<0.05, * p<0.1

3.6. Resultados de la segunda etapa: colegios

Los resultados a nivel de colegio (ver tablas 12 y 13) muestran que existe un efecto positivo del índice que se compone de la pensión del colegio, si tiene jornada ordinaria y si es privado (este índice pasa la prueba de Alpha de Cobranch). Este resultado es acorde con la literatura: por un lado, varios autores han corroborado que estudiar en colegios privados se traduce en mejores resultados en las pruebas Saber 11 (Gaviria & Barrientos, 2001; Ireguri, Melo & Ramos, 2006); por otro, es posible que un colegio más costoso se relacione con mejores condiciones físicas, lo cual algunos autores argumentan que repercute positivamente en los resultados educativos (Mediavilla & Calero, 2009). Por último, se puede suponer que una mayor exposición al colegio debería generar mejores resultados académicos, por esto es posible que la jornada completa resulte en un mejor puntaje en la prueba Saber 11 (García, Maldonado, Perry, Rodríguez y Saavedra, 2014).

Por otra parte, las *dummys* de proporción de educación de los profesores (estructura tomada de García, Maldonado, Perry, Rodríguez y Saavedra, 2014) también muestran que colegios con profesores más calificados tienen un efecto colegio relacionado con mejores resultados, y a pesar de que este efecto no es tan fuerte en la prueba de lenguaje, se puede ver que una institución con mayor proporción de profesores que normalistas tiene un efecto negativo sobre los resultados educacionales.

Un resultado interesante es que el efecto del promedio de la educación de los padres a nivel de colegio es positivo (Ann Owens, 2010; Otto, 1977), lo cual se podría traducir en un efecto par positivo. Por otra parte, un curso con más estudiantes en extra edad o con mayor variación en la edad de los estudiantes tiene un efecto negativo a nivel de colegio. Este efecto puede tener dos interpretaciones: (i) la existencia de un modelo negativo, donde la presencia de estudiantes con atraso académico empeore los resultados de otros estudiantes (Wilson, 1987); (ii) se podría estar en presencia de un efecto "focus", donde el profesor al enfrentarse a un curso más homogéneo puede diseñar un programa enfocado a la habilidad media de los estudiantes, lo que generaría mejores resultados educativos (Hoxby & Weingarth, 2006). De igual manera, es posible que los dos resultados anteriores sufran de problemas de identificación, por lo tanto, se deben tomar como un indicio.

Lo más interesante que se puede concluir de estas regresiones es que el efecto del cambio de variaciones de la prueba Saber 11 es radicalmente distinto en lenguaje comparado con matemáticas, se vuelve no significativo en matemáticas, pero en lenguaje sí lo es y, además, es negativo. Esto puede ser evidencia a favor de que el aprendizaje en lenguaje tiene muchos más

canales de transmisión sociales, mientras que los canales de trasmisión de matemáticas son más académicos.

Cuadro 12. Segundas etapas, prueba de matemática – colegios

	(1)	(2)	(3)	(4)
Coef. var. edad por colegio	-85,76***	-63,15***	-56,47***	-0,0205***
	(4,750)	(4,739)	(4,773)	(0,00604)
Coef. var. prueba de matemáticas	4,862*	3,349	2,735	0,0323***
	(2,887)	(2,454)	(2,507)	(0,00467)
índice de colegios privados, valor de la pensión y jornada.	0,179***	0,228***	0,191***	-4,26e-05
	(0,0503)	(0,0301)	(0,0348)	(0,000145)
Promedio de educación de los padres	0,587***	0,564***	0,299***	0,000754***
	(0,0401)	(0,0370)	(0,0385)	(9,65e-05)
Porcentaje de profesores con bachillerato pedagógico.	0,173	0,474	0,120	-0,00631
	(0,994)	(0,637)	(0,692)	(0,00431)
Porcentaje de profesores normalista superiores.	0,886	1,156*	1,443*	0,00965***
	(0,845)	(0,680)	(0,750)	(0,00285)
Porcentaje de profesores con estudios técnicos.	2,325**	2,376***	2,595***	0,00616*
	(1,055)	(0,879)	(0,941)	(0,00349)
Porcentaje de profesores con estudios técnicos en pedagogía.	1,524	1,895**	1,710*	0,00129
	(0,996)	(0,777)	(0,907)	(0,00270)
Porcentaje de profesores con estudios profesionales.	2,179***	2,090***	2,232***	0,00233
	(0,807)	(0,537)	(0,620)	(0,00330)
Porcentaje de profesores con estudios profesionales en pedagogía.	2,419*	2,061**	2,499**	0,00631
	(1,294)	(0,944)	(1,159)	(0,00412)
Porcentaje de profesores con postgrado.	5,094***	5,057***	4,881***	0,0124***
	(1,197)	(1,166)	(1,219)	(0,00307)
Porcentaje de profesores con postgrado pedagógico.	3,761***	3,712***	4,498***	0,00421
	(0,969)	(0,820)	(0,876)	(0,00328)
Constante	69,84***	14,94***	15,13***	-0,00965***
	(1,290)	(1,136)	(1,147)	(0,00269)
Observaciones	32,643	32,643	32,643	32,643
R-squared	0,271	0,701	0,582	0,120
Variables primera etapa	Vacío	Vacío	No vacío	No vacío
Modelo primera etapa	Interc. Al.	Pend. Al.	Interc. Al.	Pend. Al.

Errores estándar en paréntesis. *** p<0.01, ** p<0.05, * p<0.1

Cuadro 13. Segundas etapas, prueba lenguaje colegios – colegios

	(1)	(2)	(3)	(4)
Coef. var. edad por colegio.	-43,36***	-26,54***	-22,08***	-20,45***
	(1,902)	(1,704)	(1,719)	(1,480)
Coef. var. prueba de lenguaje.	-26,36***	-24,96***	-22,52***	-21,00***
	(2,240)	(1,590)	(1,511)	(1,396)
índice de colegios privados, valor de la pensión y jornada.	0,120***	0,129***	0,102***	0,101***
	(0,0369)	(0,0230)	(0,0212)	(0,0198)
Promedio de educación de los padres.	0,621***	0,474***	0,237***	0,201***
	(0,0292)	(0,0222)	(0,0208)	(0,0178)
Porcentaje de profesores normalista superiores.	-0,693	-0,140	-0,305	-0,149
	(0,785)	(0,608)	(0,575)	(0,538)
Porcentaje de profesores con estudios técnicos.	-1,494**	-0,825	-0,882*	-0,969**
	(0,694)	(0,529)	(0,513)	(0,465)
Porcentaje de profesores con estudios técnicos en pedagogía.	-0,100	0,497	0,275	0,178
	(0,933)	(0,658)	(0,624)	(0,561)
Porcentaje de profesores con estudios profesionales.	-0,223	0,264	0,103	-0,0104
	(0,645)	(0,459)	(0,457)	(0,418)
Porcentaje de profesores con estudios profesionales en pedagogía.	-0,405	-0,258	-0,276	-0,470
	(0,559)	(0,397)	(0,407)	(0,380)
Porcentaje de profesores con postgrado.	-0,249	-0,230	0,175	0,0506
	(0,931)	(0,672)	(0,710)	(0,655)
Porcentaje de profesores con postgrado pedagógico.	0,652	0,671	1,048*	0,761
	(0,711)	(0,489)	(0,467)	(0,419)
Porcentaje de profesores normalista superiores.	(0,790)	(0,622)	(0,631)	(0,583)
	0,954	0,597	0,877*	0,594
Porcentaje de profesores con estudios técnicos.	(0,711)	(0,489)	(0,467)	(0,419)
Constante	60,79***	8,001***	8,403***	8,091***
	(0,905)	(0,699)	(0,707)	(0,622)
Observaciones	32,643	32,643	32,643	32,643
R-squared	0,246	0,726	0,582	0,576
Variables primera etapa	Vacío	Vacío	No vacío	No vacío
Modelo primera etapa	Interc. Al.	Pend. Al.	Interc. Al.	Pend. Al.

Errores estándar en paréntesis. *** p<0.01, ** p<0.05, * p<0.1

4. Conclusiones

De este estudio se puede concluir que el barrio tiene efectos sobre la educación de los individuos, esto se comprueba con varios métodos de estimación

y distintas especificaciones. El elemento a resaltar en este aspecto es que la prueba de lenguaje tiende a ser más sensible a las condiciones contextuales del estudiante que la prueba de matemáticas, que depende principalmente de las condiciones del colegio. Este resultado cobra mayor fuerza si se tiene en cuenta que Bogotá cuenta con una estructura donde barrio y colegio no se superponen exactamente, el vecindario es importante de todas formas. Esta estructura separa mejor el efecto del barrio del efecto del colegio.

Además de lo anterior, se puede ver cómo el tener en cuenta ambos niveles de anidación permite conocer la significancia real de los efectos fijos, en particular, este efecto es muy claro a la hora de tener en cuenta el género en la prueba de lenguaje. Además, este resultado deja rastros de problemas de desigualdad de género en la educación. En general, sobre la primera etapa del modelo, se puede concluir que se mantienen muchos de los resultados encontrados por la literatura, entre ellos está que el nivel que explica una mayor porción de la varianza es el de las condiciones familiares y personales, seguido por el colegio, y, por último, se encuentra el barrio.

Por otra parte, se puede ver que la metodología en dos etapas funciona bastante bien, en especial si se tiene una muestra grande pero desbalanceada, pues es muy funcional a la hora de superar los problemas computacionales que traen consigo los modelos jerárquicos multinivel. Este método permite a su vez, incluir una gran cantidad de variables explicativas para ambos grupos de anidación, lo cual da pie para reforzar la hipótesis de que el lenguaje depende más de condiciones contextuales y permite tener indicios sobre otras hipótesis planteadas por la literatura en educación.

Sobre la segunda etapa del modelo, se muestran efectos muy interesantes; por un lado, se encuentra evidencia a favor de que la dispersión en resultados de la prueba y en edad de los estudiantes genere efectos negativos a nivel de institución. Por otra parte, se encuentra evidencia a favor de un efecto espacial y también se puede observar que el crimen afecta negativamente los resultados y que los universitarios en la UPZ afectan positivamente el efecto del barrio.

Referencias

Albright, J. J., & Marinova, D. M. (2015). *Estimating multilevel models using SPSS, Stata, SAS and R*. Ciudad: editorial.

Aliaga-Linares, L., & Álvarez-Rivadulla, M. J. (2010). *Residential Segregation in Bogotá across Time and Scales*. Ciudad: Lincoln Institute of Land Policy.

Barajas, R. G., Philipsen, N., & Brooks-Gunn, J. (2007). *Cognitive and emotional outcomes for children in poverty. Handbook of families and poverty*. Thousand Oaks, CA: Sage.

Belsley, D. A., Kuh, E., & Welsch, R. E. (1980). Detecting and assessing collinearity. Regression diagnostics: Identifying influential data and sources of collinearity, 85-191.

Benabou, R. (2002). Tax and Education Policy in a Heterogeneous-Agent Economy: What Levels of Redistribution Maximize Growth and Efficiency? *Econometrica*, *70*(2), 481-517.

Blanco, M. E. D. (2004). Equidad de género y diversidad en la educación colombiana. *Revista Electrónica de Educación y psicología*, 1(2), x-x.

Blume, L. E., Brock, W. A., Durlauf, S. N., & Ioannides, Y. M. (2010). Identification of social interactions.

Cook, T. D., Herman, M. R., Phillips, M., & Settersten Jr, R. A. (2002). Some ways in which neighborhoods, nuclear families, friendship groups, and schools jointly affect changes in early adolescent development. *Child development*, *73*(4), 1283-1309.

Fernández, R. (2001). Education, segregation and marital sorting: Theory and an application to UK data. *National Bureau of Economic Research*, w8377, x-x.

Fielding, A. (1999). Why use arbitrary points scores?: ordered categories in models of educational progress. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *162*(3), 303-328.

Formisano, M. (2002). Econometría espacial: características de la violencia homicida en Bogotá, 10. Ciudad: CEDE.

Garner, C. L., & Raudenbush, S. W. (1991). Neighborhood effects on educational attainment: A multilevel analysis. *Sociology of education*, x(x), 251-262.

Gaviria, A., & Barrientos, J. H. (2001). Determinantes de la calidad de la educación en Colombia. Documento 159. *Departamento Nacional de Planeación, Dirección de Estudios Económicos*.

Gibbons, S., Silva, O., & Weinhardt, F. (2013). Everybody needs good neighbours? Evidence from students' outcomes in England. *The Economic Journal*, *123*(571), 831-874.

Goux, D., & Maurin, E. (2003). *Neighborhood Efects on Performances at School*. Conference on Changing Condition in Education (Uppsala, 2003).

Hanushek, E., & Yilmaz, K. (2007). The complementarity of Tiebout and Alonso. *Journal of Housing Economics*, *16*(2), 243-261.

Harding, D. J. (2009). Collateral consequences of violence in disadvantaged neighborhoods. *Social forces; a scientific medium of social study and interpretation*, *88*(2), 757.

Hofmann, D. A., & Gavin, M. B. (1998). Centering decisions in hierarchical linear models: Implications for research in organizations. *Journal of Management*, *24*(5), 623-641.

Hoxby, C. M., & Weingarth, G. (2005). Taking race out of the equation: School reassignment and the structure of peer effects. *Working paper* 7867.

Iregui, A. M., Melo, L., & Ramos, J. (2006). La educación en Colombia: análisis del marco normativo y de los indicadores sectoriales. *Revista de economía del Rosario*, *9*(2), 175-238.

Jaramillo, S. G., Carrizosa, D. M., Rubio, G. P., Orgales, C. R., Calvo, J. E. S., Ospina, I. S., Montaña, L. B. et al. (2014). Tras la excelencia docente. Cómo mejorar la Cali.

Kaztman, R., & Retamoso, A. (2007). Efectos de la segregación urbana sobre la educación. *Revista de la CEPAL*, 91, 134.

Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *The review of economic studies*, *60*(3), 531-542.

Mayer, S. E., & Jencks, C. (1989). Growing up in poor neighborhoods: How much does it matter? *Science*, 243(4897), 1441-1446.

Mediavilla, M., & Calero, J. (2009). Determinantes internos y externos en el proceso de aprendizaje. Una aproximación al caso español a partir de la ECV-05. *Revista Iberoamericana de Educación* (OEI), 50(6), x-x.

O'brien, R. M. (2007). A caution regarding rules of thumb for variance inflation factors. *Quality & Quantity*, *41*(5), 673-690.

Owens, A. (2010). Neighborhoods and schools as competing and reinforcing contexts for educational attainment. *Sociology of Education*, *83*(4), 287-311.

Patacchini, E., & Zenou, Y. (2011). Neighborhood effects and parental involvement in the intergenerational transmission of education. *Journal of Regional Science*, *51*(5), 987-1013.

Rabe-Hesketh, S., & Skrondal, A. (2008). Multilevel and longitudinal modeling using Stata. Ciudad: STATA press.

Raudenbush, S. W. (1993). A crossed random effects model for unbalanced data with applications in cross-sectional and longitudinal research. *Journal of Educational Statistics*, *18*(4), 321-349.

Raudenbush, S. W., & Bryk, A. S. (2002). Hierarchical linear models: Applications and data analysis methods.

Ready, D. D. (2010). Socioeconomic disadvantage, school attendance, and early cognitive development the differential effects of school exposure. *Sociology of Education*, *83*(4), 271-286.

Robinson, C., & Schumacker, R. E. (2009). Interaction effects: centering, variance inflation factor, and interpretation issues. *Multiple Linear Regression Viewpoints*, *35*(1), 6-11.

Small, M. L., & Newman, K. (2001). Urban poverty after the truly disadvantaged: The rediscovery of the family, the neighborhood, and culture. *Annual Review of sociology*, *27*(1), 23-45.

Snijders, T. A. B., & Bosker, R. J. (1999). Multilevel analysis. An introduction to basic and advanced multilevel modeling. Ciudad: Sage.

Steele, F., Vignoles, A., & Jenkins, A. (2007). The effect of school resources on pupil attainment: a multilevel simultaneous equation modelling approach. *Journal of the Royal Statistical Society*: *Series A (Statistics in Society)*, *170*(3), 801-824.

Van der Leeden, R., Busing, F. M. T. A., & Meijer, E. (April, 1997). Applications of bootstrap methods for two-level models. In Multilevel Conference, Amsterdam.

Wilson, W. J. (1987). *The truly disadvantaged: The inner city, poverty and the underclass*. Chicago: Univ. Chicago Press.