

**El efecto del entorno de desarrollo municipal en la distribución de los resultados en matemáticas - Saber 11: una aproximación utilizando modelos jerárquicos lineales por cuartiles**

**Daniela Estrada Nates**

**Trabajo de Grado para optar por el título de Magíster en Economía**

**Director del trabajo de Grado:**

**Jhon James Mora Rodríguez**

**Universidad Icesi**

**Facultad de Ciencias Administrativas y Económicas**

**Cali, abril de 2017**

## **Resumen**

Este trabajo analiza el efecto del entorno de desarrollo y la tipología del municipio (categorización de las entidades territoriales propuesta por el DNP en 2014) donde vive el estudiante en el desempeño que éste tiene en el componente de matemáticas de la Prueba Saber 11, realiza por el ICFES, para el segundo semestre de 2014. Empleando modelos lineales jerárquicos en una regresión por cuantiles y controlando por características personales, familiares y del colegio, los resultados sugieren que tanto el entorno de desarrollo como la tipología municipal tiene un efecto significativo en el logro académico en matemáticas de los estudiantes en los cuartiles estudiados. Luego los hacedores de política pública deben tener en cuenta las condiciones del lugar donde viven los estudiantes a la hora de formular políticas educativas encaminadas a mejorar la calidad educativa. Especialmente, deberían enfocarse en atender las necesidades de los municipios de tipología F y G, y en estudiar las prácticas pedagógicas y administrativas que se imparten en los municipios de la tipología A y B. Los resultados obtenidos se comparan con los que se obtendrían empleando una regresión cuantílica y un modelo multinivel para el estudiante promedio, y se encuentra que hay diferencias en la magnitud y significancia de la asociación, lo que evidencian la importancia de emplear esta metodología.

Palabras claves: Educación, modelos jerárquicos lineales por cuantiles, economía, Colombia

## **Abstract**

This paper analyzes the effect that has the grade of development (categorization of territorial entities proposed by the DNP in 2014) the municipality, where the student lives, in the performance that he has in the math component of the Saber 11 test. Using linear quantile mixed model and controlling for personal, family and college characteristics, the results suggest that the grade of development a municipality has have a significant effect on achievement. Therefore, public policy makers must take into account the conditions of the place where students live when it comes to formulating educational policies aimed at improving the quality of education. In particular, they should focus on meeting the needs of F and G typology municipalities and studying the pedagogical and administrative practices that are taught in municipalities of typology A and B. The results obtained are compared with those obtained using a quantile regression and a multilevel model for the average student. Results show that there are differences in the magnitude and significance of the association, which evidences the importance of using this methodology.

Keywords: Education, linear quantile mixed model, economy, Colombia

## Tabla de contenido

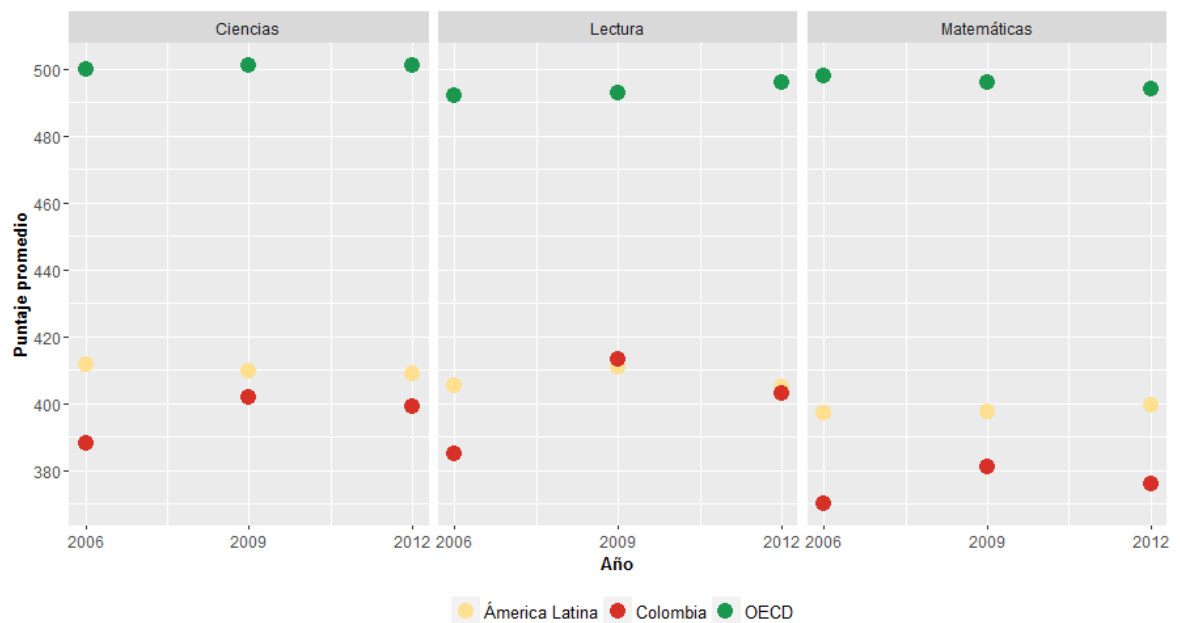
1. INTRODUCCIÓN .....	3
2. REVISIÓN DE LA LITERATURA .....	8
2.1 MODELOS .....	8
2.2 INTERNACIONAL .....	10
2.3 NACIONAL.....	10
3. APROXIMACIÓN EMPÍRICA .....	14
3.1 REGRESIÓN POR CUARTILES DE MODELOS LINEALES JERÁRQUICOS .....	15
3.2 APROXIMACIÓN A SEGUIR .....	15
4. DATOS.....	17
4.1 DNP.....	17
4.2 SABER 11.....	19
4.3 ESTADÍSTICAS DESCRIPTIVAS .....	20
5. RESULTADOS.....	22
5.1 ENTORNO DE DESARROLLO.....	23
5.2 TIPOLOGÍA MUNICIPAL .....	26
6. CONCLUSIONES .....	29
REFERENCIAS.....	29
ANEXOS .....	39

### 1. INTRODUCCIÓN

Resulta preocupante que en Colombia la mayoría de los estudiantes próximos a graduarse de secundaria presenten bajos niveles de rendimiento en la competencia de matemáticas. Como se puede observar en la Figura 1, el componente de matemáticas de las pruebas del Programa

Internacional para la Evaluación de los Estudiantes (o prueba PISA, por sus siglas en inglés)<sup>1</sup> es donde peor les va a los estudiantes colombianos. Para 2012, año en que ésta fue el área de conocimiento en que se profundizó, Colombia se ubicó en el puesto 62 entre 65 países que participaron en la prueba y 3 de cada 4 estudiantes que la tomaron fueron calificados en la categoría de bajo rendimiento (OECD, 2014). Hanushek (2013) ha encontrado que aquellos que están en la categoría de bajo rendimiento, tienen más dificultades para participar en el mercado laboral y mayor probabilidad de tener dificultades para beneficiarse de nuevas oportunidades educativas formales.

*Figura 1: Puntaje promedio para América Latina, Colombia y países que componen OECD en pruebas PISA 2006, 2009 y 2012*



Fuente: OECD PISA resultados 2006, 2009 y 2012

Nota: Los países de América Latina que toman estas pruebas son Argentina, Brasil, Uruguay, Chile y Perú; este último se tiene en cuenta a partir de 2009

<sup>1</sup> PISA es una prueba estandarizada que se realiza cada tres años y que evalúa a estudiantes de 15 años (muestras representativas, entre 4000 y 10000 estudiantes por país) de diferentes países del mundo en tres competencias: ciencias, lenguaje y matemáticas. (OCDE, 2007)

Además, una buena competencia matemática impacta otras dimensiones de la vida del ser humano, como el empleo (Smith, 2004), el ingreso (Litster, 2013), la acumulación de riqueza (Melzer et al., 2006), la prevención y tratamiento exitosa de diferentes enfermedades (Reyna y Brainerd, 2007), y el tener un peso saludable (Huizinga et al., 2008)

En el área macroeconómica, DfEE (1999) halla que una buena competencia en matemáticas tiene un efecto positivo en la productividad promedio de la fuerza laboral y que ésta explica una proporción significativa de las diferencias en el desempeño económico que hay entre los países.

Por otro lado, la estructura de la prueba PISA permite comparar el rendimiento de los alumnos a lo largo del tiempo dentro del país y con otros países, y es una medida de la calidad educativa (OCDE, 2007). Esto es importante puesto que la calidad de la educación tiene efectos a nivel microeconómico y macroeconómico: impacta las probabilidad de ingresar a la educación superior (Jimenez et al., 1991), es uno de los determinantes de los ingresos y su distribución (Hanushek y Woessmann, 2008); además que tiene una correlación significativa con el crecimiento económico y la disminución de la pobreza (Delgado et al., 2014; Hanushek y Woessmann, 2008; Cooray , 2009; van der Berg, 2008).

Luego, la Figura 1 también envía una señal a los diferentes actores del sector educativo de que los jóvenes colombianos están recibiendo una educación que presenta deficiencias en la calidad<sup>2</sup> puesto que el puntaje promedio de los estudiantes colombianos en las tres

---

<sup>2</sup> La calidad de la educación es un concepto amplio; Rodríguez (1994) propone 4 categorías para interpretar la calidad: cobertura, gestión pedagógica, resultados y gestión administrativa. Este artículo se concentra en la calidad de la educación medida a través de los resultados de pruebas estandarizadas

competencias evaluadas se ha ubicado por debajo del promedio de los países que conforman la Organización para la Cooperación y el Desarrollo Económico (OECD) y los países de la región de América Latina (Argentina, Brasil, Chile, Uruguay y Perú se incluye desde 2009) que también participan de esta prueba.

El Gobierno no ha sido ajeno a esta situación y ha empezado a implementar diferentes proyectos enmarcados en el Plan Sectorial 2010-2014 y el Plan Nacional de Desarrollo 2014-2018 para solventarla. Algunos de estos proyectos son la implementación de guías y materiales didácticos para el desarrollo de competencias (Ministerio de Educación Nacional, 2010); acompañamiento y formación a docentes y directivos docentes bajo el programa Excelencia docente, Jornada Única<sup>3</sup> y Educación media para todos<sup>4</sup> (Nacional, 2015).

No obstante, uno de los problemas que existen para guiar la política pública a nivel municipal es que la categorización de las entidades territoriales<sup>5</sup> no responde a la heterogeneidad municipal ni refleja las brechas fuertes que existen entre las regiones en el país. Éstas contribuyen a la desigualdad (Cortés y Vargas, 2012), la cual ha persistido en el tiempo (Meisel Rosa y Romero, 2007) y que en términos educativos se evidencia en la inequidad espacial en la distribución de las dotaciones educativas (número de colegios y número de maestros) a nivel regional (Bonet, 2005) y la concentración regional de maestros con altos niveles de calificación (Galvis y Bonilla Mejía, 2011).

---

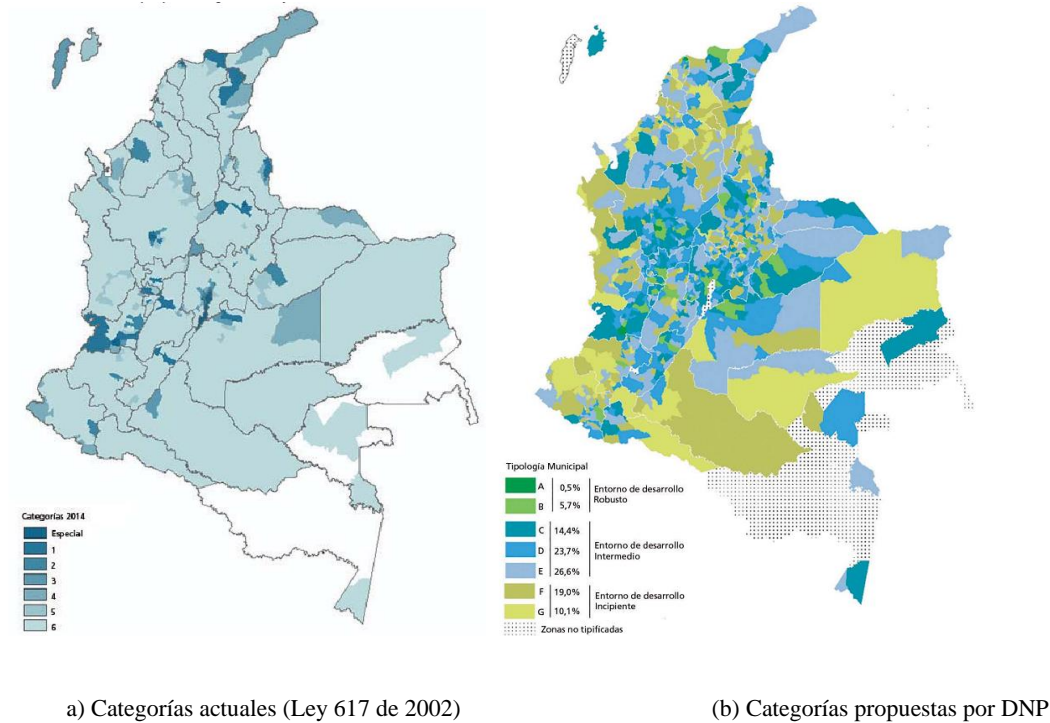
<sup>3</sup> Este programa consiste en extender el tiempo que permanecen los niños que asisten a colegios del sector oficial. Entre los objetivos que tiene está fortalecer las competencias básicas (como matemáticas) de los estudiantes al aumentar el tiempo dedicado a actividades pedagógicas. Se espera que su implementación se dé de forma gradual y finalice en 2025 en áreas urbanas y en 2030 en las zonas rurales (Nacional, 2015).

<sup>4</sup> Este programa busca ampliar la cobertura en educación media del país, la cual era del 77.31% en 2015, al volverla obligatoria. Dicho plan se está implementando de manera gradual con las entidades territoriales certificadas (Nacional, 2015).

<sup>5</sup> Actualmente se basa en dos criterios: tamaño de la población y los ingresos corrientes anuales (Ley 617 de 2002). Lo anterior implica que el 90% de los municipios se encuentran clasificados en la misma categoría, la sexta. Es decir, en municipios con bajos ingresos corrientes anuales y una población pequeña.

Como respuesta a dicha carencia, el Departamento Nacional de Planeación (DNP) propuso el diseño de una tipología municipal y departamental que permita identificar grupos más homogéneos (ver Figura 2).

*Figura 2: Categorización de las entidades territoriales*



Fuente: Tomado de Aguilar Londoño et al. (2015)

Hasta el momento dicha categorización no se ha empleado en algún estudio, luego partiendo tanto de la importancia de las matemáticas y de esta nueva categorización, el presente artículo espera contribuir a llenar un vacío actual en la literatura colombiana al responder: ¿Cómo afecta el entorno de desarrollo del municipio donde vive el estudiante a la distribución de los resultados del componente de matemáticas de la prueba Saber 11?

Este artículo contiene además de la introducción, 5 secciones adicionales. En la siguiente se describen los factores que pueden afectar el desempeño en matemáticas, y la evidencia empírica a nivel internacional y nacional que se ha encontrado al respecto. En la tercera parte se explica la aproximación empírica. Posteriormente, se caracterizan los datos empleados y se presentan las estadísticas descriptivas. En la quinta sección se encuentran los resultados de estimar el modelo jerárquico lineal por cuartiles y su respectivo análisis. En la última sección se encuentran las conclusiones.

## 2. REVISIÓN DE LA LITERATURA

Esta sección presenta los diferentes modelos que se pueden emplear para responder la pregunta planteada, y ejemplos de los hallazgos que se han encontrado en el ámbito internacional y nacional sobre los factores que pueden afectar el rendimiento en matemáticas de los estudiantes.

Generalmente para estimar cuáles factores pueden estar relacionados con los rendimientos en pruebas estandarizadas se emplean funciones de producción de educación (Hoeryner y Lemke, 2006; Hanushek ,1986). Los resultados pueden variar dependiendo del contexto que se analice, de la riqueza de la base de datos empleada para caracterizar a los colegios e individuos y si el estudio se enfoca en los resultados del estudiante promedio o en la distribución de puntajes obtenidos por los estudiantes.

### 2.1 MODELOS

Por un lado, la situación se puede modelar a nivel de estudiante, a nivel de colegio o multiniveles (es decir modelos lineales jerárquicos que incluyen la estructura jerarquizada



natural de los datos donde los estudiantes se agrupan en salones que a su vez se agrupan en colegios).

De acuerdo a Hox (2002) este último tipo de modelos es ampliamente utilizado en educación pues cubre dos de los inconvenientes de los modelos anteriores. El primero es que corrige el problema estadístico derivado de la falta de independencia entre los factores asociados al logro dentro de un mismo colegio, el cual se presenta a nivel estudiante cuyo se estima a través de mínimos cuadrados ordinarios. El segundo, evita la pérdida de información que se da en los modelos a nivel de colegio debido a la agregación, la cual puede llevar a una pérdida de poder estadístico. Por otro lado, evita un problema conceptual (falacia ecológica) que se puede dar en el momento de interpretar los resultados para un nivel de la jerarquía incorrecto. Por otra parte, cuyo se quiere analizar al estudiante, ya sea empleando un modelo a nivel de estudiante o multinivel, se puede buscar entender cómo los factores afectan al estudiante promedio (por ejemplo, empleando mínimos cuadrados ordinarios) o cómo varía el efecto a lo largo de la distribución de los estudiantes al escoger diferentes percentiles de la distribución y observar si los factores elegidos afectan de igual forma a todos los estudiantes. Este último método de estimación se llama regresión cuantílica y es un método que ha cobrado importancia puesto que diferentes estudios (Eide y Showalter, 1998; Koenker y Bassett, 1978) muestran que las variables que son significativas para el estudiante promedio no lo son necesariamente para los estudiantes a lo largo de la distribución de los resultados. Lo anterior puede resultar crucial para fines de política pública ya que al utilizar esta metodología no sólo se responde si un factor es importante, también para quién lo es.

## 2.2 INTERNACIONAL

Siguiendo a Restrepo y Alviar (2005) los insumos de educación se pueden agrupar en las siguientes categorías:

- Individuales: género (Demir et al., 2010), raza (Haile y Nguyen, 2008)
- Familiares: nivel socioeconómico (Bassett et al., 2002), años de educación (Chiu y Xihua, 2008), nivel educativo de los padres (Bassett et al., 2002; Haile y Nguyen, 2008), ocupación de los padres (Haile y Nguyen, 2008), número de hermanos (Chiu y Xihua, 2008), entre otras.
- Colegio: Dentro de este grupo hay una inmensa diversidad de variables. porcentaje de maestros con maestría y la razón profesor (Bassett et al., 2002; Eide y Showalter, 1998), tamaño de las clases (Hanushek, 2002), naturaleza jurídica de los colegios (si son privados o públicos) (Lubienski y Lubienski, 2006; Jimenez et al., 1991); duración de la jornada escolar (Lavy, 2015; Schiman, 2015).
- Factores del lugar de residencia: El lugar de residencia se puede capturar desde una perspectiva de país, de departamento, municipio, ciudad o incluso de barrio (Chiu y Xihua, 2008; Fryer Roly G. y Levitt, 2010).

Por otra parte, en la revisión de la literatura internacional en el campo educativo sólo se encontraron dos estudios que estiman un modelo multinivel empleando un método que tiene en cuenta la distribución de los puntajes de los estudiantes: Tian y Chen (2006) y Costanzo (2015).

## 2.3 NACIONAL

En esta sección se presentan los resultados de varios estudios realizados para el contexto colombiano por las categorías de factores que se han encontrado que están correlacionados con el rendimiento en matemáticas. La mayoría emplean los resultados de las pruebas Saber 11 ejecutadas por el Instituto Colombiano para el Fomento de la Educación Superior (ICFES).

- Factores personales y familiares

Independientemente de los datos empleados y el modelo, el género y la edad son las variables más utilizadas dentro del grupo de factores personales. Estas variables tienden a presentar el mismo signo, el cual indica que los hombres en promedio obtienen puntajes en matemáticas superiores a las mujeres, y que entre mayor es la edad del estudiante menor es su puntaje. Otras variables típicamente empleadas se pueden ver en el Cuadro 1.

Cuadro 1: Evidencia nacional de factores personales y familiares asociados al puntaje de matemáticas.

Autor	Datos	Modelo	Factores			
			Hombre	Edad	Condiciones económicas	Nivel educativo padres
Piñeros, L. y Rodríguez, A. (1998)	Saber 11 - 1997 II	Multinivel	(+)	N.I	NSE: +	N.I
Correa, J.J. (2004)	Puntaje total: Saber 11 2001- I. Cali	Multinivel	(+)	(-)	Ingresos familiares: +	Padre con el mayor nivel educativo (+)
Barrientos, J. (2008)	Saber 11 - 1999- 2004- 2006	Estudiante	(+)	(-)	Estrato: (+/- )	N.I
Bonilla, L. (2011)	Puntaje total: Saber 11 - 2009	Estudiante	(+)	(-)	Ingresos familiares: +	Madre (+)

Barón (2012)	Saber 11 2009: Barranquilla	Estudiante	NA	N.I	Ingresos familiares: +	Ambos padres (+) para mujeres y mixto para hombres
	Saber 11 2009: Bogotá	Estudiante	NA	N.I	Ingresos familiares: +	Ambos padres (+) para mujeres y hombres
Montes, I., Garcés, J., Chica, S., y Jaramillo, A. (2014)	Saber 11 - 2010	Multinivel	(+)	(-)	NSE: +	N.I

Nota: N.I Variable no se incluyó y NA: no aplica \* Sólo se incluyeron los años de la base de datos de Saber 11 utilizada y la variable dependiente en caso de que no sea el puntaje en matemáticas † Sólo se incluyeron aquellas variables que fueron significativas al menos al 95 % de confianza ‡ Regresión cuantílica.

Fuente: Elaboración propia.

- Factores del colegio

Las variables más utilizadas son la naturaleza jurídica de la institución (si es oficial o no) y la jornada escolar (ver Cuadro 2).

Cuadro 2: Evidencia nacional de factores del colegio asociados al puntaje de matemáticas.

Autor	Datos	Modelo	Factores colegio	
			Oficial	Jornada escolar
Piñeros, L. y Rodríguez, A. (1998)	Saber 11 - 1997 II	Multinivel	N.I.	Mañana (+)
				Tarde (+)
				Completa (+)
Correa, J.J. (2004)	Puntaje total: Saber 11 2001-I. Cali	Multinivel	(+)	Mañana (+)
				Tarde (-)
				Noche (-)
Barrientos, J. (2008)	Saber 11 1999- 2004-2006	Estudiante	N.I.	Tarde (+) : 0.25, 0.5 y 0.75
Bonilla, L. (2011)	Puntaje total / Saber 11 - 2009	Estudiante	(+)	Completa (+)
Barón (2012)	Saber 11 2009: Barranquilla y Bogotá	Estudiante	N.I.	Completa o mañana (+)
	Saber 11 2010	Multinivel	(-)	Mañana (+)

Montes, I., Garcés, J., Chica, S., y Jaramillo, A. (2014)				Tarde (+)
				Completa (+)
Hincapié, D. (2016)	Saber 5 y 9 - 2002, 2005, y 2009.	Estudiante	NA	Completa (+)

Nota: N.I Variable no se incluyó y NA: no aplica \* Sólo se incluyeron los años de la base de datos de Saber 11 utilizada y la variable dependiente en caso de que no sea el puntaje en matemáticas † Sólo se incluyeron aquellas variables que fueron significativas al menos al 95 % de confianza ‡ Regresión cuantílica.

Fuente: Elaboración propia.

De igual forma se pueden incluir otras variables como modalidad (técnico, académico, o ambos) (Piñeros y Rodríguez, 1998); género de la institución (masculino, femenino o mixto) (Bonilla Mejía, 2011), valor de la pensión (Barón, 2012), características de los docentes y estrategia de enseñanza (Montes et al., 2014), entre otras.

- Factores del lugar de residencia del estudiante

Hay pocos estudios que contemplen variables de este grupo: Restrepo y Alviar (2005), Bonilla Mejía (2011) y Martínez y Turriago (2015).

Cuadro 3: Evidencia nacional de factores del lugar de residencia asociados al puntaje de matemáticas.

Autor	Datos	Modelo	Municipio	Hallazgos
Restrepo, P.P., y Alviar, M. (2004)	Saber 11- 1999	Multinivel	Subregiones Antioquia	Oficial (-) Jornada mañana (-)
Bonilla, L. (2011)	Puntaje total Saber 11 - 2009	Estudiante	10 ciudades principales	% Urbano (-) % NBI (-) % analfabetismo (-) % homicidios (-) Desempeño fiscal (+) NBI (-)
Martínez, W.H., y Turriago, A. (2014)	Puntaje total: Saber 11 - 2005-2012, semestre II	NA	Todos	% Alumnos del sector privado (+) % Profesores con educación superior (+)

Nota: N.I Variable no se incluyó y NA: no aplica \* Sólo se incluyeron los años de la base de datos de Saber 11 utilizada y la variable dependiente en caso de que no sea el puntaje en matemáticas † Sólo se incluyeron aquellas variables que fueron significativas al menos al 95 % de confianza.

Fuente: Elaboración propia.

En síntesis, dentro de la revisión de la literatura colombiana se encontraron dos vacíos que este estudio pretende empezar a cubrir al emplear un modelo multinivel por cuantiles, el cual no ha sido utilizado con anterioridad; por otro lado, reconocer la heterogeneidad municipal, la cual ha sido ignorada en la mayoría de los estudios puesto que estos emplean los puntajes en matemáticas desde la perspectiva nacional o de las ciudades principales como Bogotá, Cali, Medellín y Barranquilla.

### 3. APROXIMACIÓN EMPÍRICA

Las aproximaciones empíricas típicamente utilizadas para analizar el efecto que tienen diferentes factores en el desempeño escolar son: regresiones por cuantiles (QR por sus siglas en inglés) y regresiones de modelos lineales jerárquicos. Sin embargo, la regresión por cuantil no puede lidiar con datos cuya naturaleza es jerárquica, como en el caso educativo, donde los estudiantes se agrupan en salones y estos a su vez en colegios. Parte de la solución a ese inconveniente se ha dado en años recientes con el desarrollo de modelos y algoritmos que permiten aplicar regresiones por cuantiles a datos con una estructura jerárquica (e.g. Tzavidis y Brown, 2010; Geraci y Bottai, 2007; Smith et al., 2015). En este caso se estimará el modelo empleado el método de regresión por cuantiles con efectos mixtos (LQMM) de Geraci y Bottai (2014) en el software R (2013), el cual permite estimar el cuantil condicional de la variable dependiente en función de efectos fijos y aleatorios. A continuación, se presentan las principales características del tercer método.

### 3.1 REGRESIÓN POR CUARTILES DE MODELOS LINEALES JERÁRQUICOS

La idea básica del LQMM (Geraci y Bottai,2014) es que las covariables pueden ejercer diferentes efectos dependiendo del cuantil que se analice, tal como sucede en la regresión por cuantil estándar. El grado de heterogeneidad no observada, como en los modelos lineales jerárquicos, se caracteriza con los parámetros específicos de la varianza, la cual depende del percentil estudiado y que se identifica por  $\theta$ . La ecuación básica que se estima es:

$$y_{ij} = x_{ij}^T \beta_{\theta} + z_{ij} u_j + \varepsilon_{\theta} \quad (3)$$

Donde  $y_{ij}$  es la variable dependiente,  $x$  es el vector de efectos fijos,  $z$  es en un vector de efectos aleatorios,  $\varepsilon$  es el término de error que sigue una distribución de Laplace (Sánchez y Lachos, 2013; Geraci y Bottai, 2014; Marino y Farcomeni, 2015).

En términos generales, la estimación se basa en la relación existente entre la minimización de las desviaciones estándar de la regresión cuantílica y la maximización de una distribución Laplace. La estimación de parámetros se lleva a cabo mediante el uso de la función de distribución asimétrica de Laplace que se ha introducido en el contexto de regresiones por cuantiles por Yu et al. (2003) y Koenker y Machado (1999). Para un detalle del algoritmo de estimación ver Geraci y Bottai (2007) y Geraci y Bottai (2014).

### 3.2 APROXIMACIÓN A SEGUIR

1. Estimar el modelo vacío o de efectos fijo (Ecuación 4) y calcular el coeficiente de correlación interclase (Ecuación 5):

$$y_{ij} = 1 * \theta^\tau + 1 * u_j + \varepsilon_{ij}^\tau \quad (4)$$

$$ICC = p^\tau = \frac{\Psi_{u_\tau}^2}{\Psi_{u_\tau}^2 + \Psi_{\varepsilon_\tau}^2} \quad (5)$$

Donde  $\Psi_{\varepsilon_\tau}^2$  representa la varianza al interior del colegio (entre los alumnos que se encuentran en el cuartil  $\tau$ ), y  $\Psi_{u_\tau}^2$  representa la varianza entre colegios en el cuartil  $\tau$ . Luego  $\Psi_{u_\tau}^2 + \Psi_{\varepsilon_\tau}^2$  indica la varianza total.

2. Estimar el modelo incluyendo la variable independiente del entorno de desarrollo:

$$y_{ij} = x_{ij}^T \theta^\tau + z_{ij}^T u_j + \varepsilon_{ij}^\tau \quad (6)$$

Donde

- $\tau$ : son los percentiles 0.25, 0.5 y 0.75
- $y_{ij}$ : es el puntaje en matemáticas del estudiante  $i$  que asiste al colegio  $j$
- $x_{ij}^T = (1, \text{munintermedia}_{ij}, \text{munincipiente}_{ij}, \text{edad}_{ij}, \text{mujer}_{ij}, \text{msecundaria}_{ij}, \text{mterciaria}_{ij}, \text{oficial}_{ij}, \text{jmanana}_{ij}, \text{jtarde}_{ij})$  (Ver Anexos para una descripción de cada una de las variables.)
- $\theta^\tau = (\theta_0^\tau, \theta_1^\tau, \theta_2^\tau, \theta_3^\tau, \theta_4^\tau, \theta_5^\tau, \theta_6^\tau, \theta_7^\tau, \theta_8^\tau, \theta_9^\tau)^T$
- $z_{ij}$ : 1
- $u_j$ : son los efectos aleatorios en el intercepto  $u \perp \varepsilon_{ij}$  y  $u \sim N(0, \Psi^\tau)$  por lo tanto las estimaciones se calcularán mediante integración numérica utilizando el método de cuadratura Gauss- Hermite (Geraci y Bottai, 2014).
- $\varepsilon$ : término de error  $\varepsilon \sim \text{Al}(0, \sigma, \tau)$ .

3. Estimar el modelo incluyendo la variable independiente de las tipologías municipales que conforman los entornos de desarrollo:



$$y_{ij} = x_{ij}^T \delta^\tau + z_{ij}^T u_j + \varepsilon_{ij}^\tau \quad (6)$$

Donde

- $x_{ij}^T = (1, B_{ij}, C_{ij}, D_{ij}, E_{ij}, F_{ij}, G_{ij}, \text{mujer}_{ij}, \text{msecundaria}_{ij}, \text{mterciaria}_{ij}, \text{edad}_{ij}, \text{oficial}_{ij}, \text{jmanana}_{ij}, \text{jtarde}_{ij})$
- $\delta^\tau = (\delta_0^\tau, \delta_1^\tau, \delta_2^\tau, \delta_3^\tau, \delta_4^\tau, \delta_5^\tau, \delta_6^\tau, \delta_6^\tau, \delta_7^\tau, \delta_8^\tau, \delta_9^\tau, \delta_{10}^\tau, \delta_{11}^\tau, \delta_{12}^\tau, \delta_{13}^\tau)^T$

4. Estimar el modelo de la ecuación 6 y 7 empleando una regresión cuantílica y un modelo lineal jerárquico.

Los errores estándar para determinar la significancia individual de los coeficientes estimados se obtienen por el método de bootstrapping (se emplean 50 repeticiones).

#### 4. DATOS

Los datos que se emplean provienen de dos fuentes. La información sobre el entorno de desarrollo municipal y tipología municipal del Departamento Nacional de Planeación(DNP). Esta información está disponible sólo para el 2014. Los datos sobre los puntajes en matemáticas y los factores individuales, familiares y de los colegios provienen de Saber 11 del segundo semestre del 2014.

##### 4.1 DNP

Las tipologías y clasificaciones de entorno de desarrollo municipal<sup>6</sup> provienen de la propuesta elaborada por el DNP. La metodología empleada para construir las tipologías se basó en el análisis de componentes principales (ver Aguilar Londoño et al. (2015) para un detalle de la

---

<sup>6</sup> Bogotá D.C no se incluyó en la Tipología Municipal sino en la Tipología Departamental, por lo que de ésta se extrajo la información correspondiente.

metodología y las variables utilizadas en la construcción del índice). De lo anterior resultan tres entornos de desarrollo: i) Entorno de Desarrollo Robusto que corresponde a los municipios pertenecientes a la tipología municipal A y B ii) Entorno de Desarrollo Intermedio corresponde a los municipios que pertenecen a la tipología C, D y E; iii) Entorno Incipiente, son los municipios pertenecientes a la tipología F y G. Como se observa en el Cuadro 4, si bien la tipología A agrupa la menor cantidad de municipios del país (0.5%), contiene casi un tercio de los habitantes del país para el 2014. Las otras tipologías contienen un menor número de habitantes, siendo la tipología G la que contiene la menor cantidad (3.6% del total de habitantes). Por otro lado, entre todas las tipologías, es la E la que agrupa la mayor cantidad de municipios, seguida de la D, F y C.

Cuadro 4: Número de municipios y habitantes por Tipología municipal

Entorno de Desarrollo	Tipología	Número de municipios	Participación (%)	Población	% Población	Ejemplo
Robusto	A	6	0,5%	15 293 275	32,17%	Cali
	B	63	5,7%	9 179 477	19,31%	Manizales
Intermedio	C	158	14,4%	7 250 124	15,25%	Tunja
	D	261	23,7%	5 812 304	12,23%	Florencia
	E	293	26,6%	5 126 952	10,79%	El Cocuy
Incipiente	F	209	19,0%	3 173 444	6,68%	Moñitos
	G	111	10,1%	1 700 315	3,58%	Currillo
Total general		1101	100%	47535891	100%	

Fuente: Adaptado de Carmona, Supelano y Osejo (2015).

Las tipologías reflejan las características propias de cada municipio, en relación con seis dimensiones: Urbano regional, condiciones sociales, económica, ambiental, institucional y de seguridad. Cada una de éstas se mide por medio de un índice que va de 0 a 1, donde entre más cerca esté el índice de la entidad territorial a 1 mejor se encuentra en esa dimensión.

## 4.2 SABER 11

La prueba Saber 11 es efectuada por el Instituto Colombiano para la Evaluación de la Educación (ICFES). Esta prueba es importante porque todos los estudiantes que se deseen graduar del bachillerato deben tomarla (es decir, es de carácter censal y obligatoria), además las instituciones de educación superior emplean los resultados como un mecanismo para seleccionar a los estudiantes y otorgar becas educativas<sup>7</sup> (Ministerio de Educación Nacional, 2015).

De acuerdo a ICFES (2015) la prueba Saber 11 evalúa 5 competencias: matemáticas, lectura crítica, sociales y competencias ciudadanas, ciencias naturales e inglés. Los puntajes de estos componentes se miden en una escala de 0 a 100, con media 50 y desviación estándar 10. La distribución de estos puntajes se debe a la aplicación de la metodología de respuesta al ítem a los puntajes iniciales que obtienen los estudiantes (Abadía y Bernal, 2016). Por lo tanto, el puntaje no indica el número de respuestas correctas, y por ahora no tiene asociado una caracterización de los niveles de desempeño para la competencia, excepto para inglés.

Se escogió la base del segundo semestre de 2014 por dos razones. Primero, el indicador de tipología municipal está construido sólo para el 2014. Segundo, en ese semestre se introdujeron cambios a la forma y estructura de la prueba por lo que los resultados del primer semestre no son estrictamente comparables con los del segundo semestre (Ministerio de Educación Nacional, 2014).

---

<sup>7</sup> Uno de los principales criterios utilizado por el Gobierno en el programa de Ser Pilo Paga, que consiste en otorgar las becas 100 % condonables para acceder a las instituciones de educación superior acreditadas en Alta Calidad, es el puntaje total que los estudiantes obtienen en Saber 11.

La base de datos de 2014 se caracteriza en términos generales por incluir:

- Estudiantes que tienen entre 14 y 28 años.
- Estudiantes que asisten a colegios públicos y privados, en jornada diurna (mañana, tarde, y completa).
- Viven en cualquier municipio de Colombia excepto en San Yrés Islas, Belén de Bajira, El Encanto, La Chorrera, La Pedrera, La Victoria, Mirití - Parana, Puerto Santyer, Tarapaca, Barranco Minas, Maripipana. Lo anterior debido a que para estos municipios no existe una tipología definida por el DNP.
- Pertenecen a alguna institución educativa.
- Toman la prueba por primera vez.
- Libres, es decir no están privados de su libertad.
- Reportaron información para todas las variables de interés.
- Pertenecen a instituciones educativas que presentaron como mínimo 10 estudiantes.

Dado lo anterior, se obtienen 410.608 observaciones de 533.133, es decir el 77,0% de las observaciones originales.

#### 4.3 ESTADÍSTICAS DESCRIPTIVAS

El Cuadro 5 muestra las estadísticas descriptivas de todas las variables usadas en la estimación a nivel nacional y por entorno de desarrollo.

*Cuadro 5: Estadísticas descriptivas a nivel nacional y por entorno de desarrollo*

Variable	Entorno de desarrollo			
	Nacional	Incipiente	Intermedio	Robusto

No. De estudiantes	410,709	35,111	150,928	224,670
<b>Personales</b>				
Promedio puntaje Matemáticas	50.88	46.58	49.32	52.60
Desviación Estándar puntaje Matemáticas	9.68	8.29	9.04	9.90
Primer cuartil puntaje Matemáticas	45	41	44	46
Segundo cuartil puntaje Matemáticas	50	46	49	52
Tercer cuartil puntaje Matemáticas	56	52	55	58
Mujer (%)	54.89	53.30	55.61	54.65
Edad (promedio años)	16.82	17.14	16.87	16.74
<b>Familiares (%)</b>				
Nivel educativo madre: ninguno o primaria	33.21	58.14	43.11	22.67
Nivel educativo madre: secundaria	45.18	33.34	41.45	49.53
Nivel educativo madre: terciaria	21.61	8.52	15.44	27.80
<b>Colegio (%)</b>				
Oficial	77.47	97.82	90.51	65.54
Jornada Mañana	59.63	77.23	64.86	53.37
Jornada Tarde	18.36	8.40	14.45	22.55
Jornada Completa	21.98	14.36	20.64	24.07

Fuente: Cálculos propios a partir de Saber 11, 2014-2

En cuanto a las características personales, se puede observar que el puntaje promedio en matemáticas de los estudiantes que viven en los entornos de desarrollo incipiente e intermedio está por debajo del puntaje promedio nacional; mientras que, el promedio en el entorno robusto está por encima. Este comportamiento se mantiene si se comparan los puntajes en matemáticas obtenidos por cuartil, tanto por entornos de desarrollo con en el ámbito nacional.

De igual forma hay diferencias al interior de los entornos, por ejemplo, se observa que la diferencia entre la cola inferior (primer cuartil) y la cola superior (tercer cuartil) de la distribución del puntaje aumenta ligeramente por entorno de desarrollo: 9 puntos para entorno de desarrollo incipiente; 11 puntos para entorno de desarrollo intermedio; y 12 para entorno de desarrollo robusto.

Asimismo, hay diferencias marcadas entre los entornos de desarrollo para un mismo cuartil lo que indica que el lugar de residencia puede ser un factor relevante. Por ejemplo, el puntaje del primer cuartil para el estudiante que vive en un municipio incipiente es de 41 puntos, y para un estudiante que vive en un entorno de desarrollo robusto es de 46; mientras que, para el tercer cuartil el puntaje es de 52 y 58, respectivamente.

Por otra parte, el porcentaje de mujeres es relativamente similar y superior al 50 % para todos los casos. La edad promedio del estudiante es de 16.82 años a nivel nacional, para el municipio incipiente e intermedio es ligeramente mayor y para los estudiantes de municipios robustos es menor.

En cuanto a las características familiares, en el ámbito nacional el 45.18 % de los estudiantes reportaron que el máximo nivel educativo alcanzado por su madre había sido secundaria (incompleta o completa), en este mismo nivel educativo se concentra la mayoría de respuestas de los estudiantes que viven en un entorno de desarrollo robusto. Por el contrario, tanto en el entorno incipiente como en el intermedio el mayor porcentaje de respuesta se concentró en ninguno o primaria (incompleta o completa).

Por último, se encuentra más del 90 % de los estudiantes que viven en un entorno de municipio incipiente o intermedio asisten a colegios oficiales, mientras que en los municipios robustos ese porcentaje es del 65.5 %. En los tres entornos, más del 50 % de los estudiantes asiste a jornada mañana, seguido de jornada completa y por último jornada de la tarde.

## 5. RESULTADOS

En esta sección se presentan los resultados de la aplicación de la aproximación empírica tanto a nivel de entorno de desarrollo como de tipología municipal. Previo a eso se estima el modelo vacío de la ecuación 4.

Los resultados de iniciales indican que el colegio explica el 64% de la variación en los resultados del puntaje de matemáticas en 2014 II de los estudiantes que se encuentran en la cola inferior de la distribución (ver Cuadro 6). Mientras que, ese porcentaje aumenta en 12 puntos porcentuales si el estudiante se encuentra en la mediana de la distribución, y en 4 puntos porcentuales para el estudiante del tercer cuartil. Lo anterior apoya la hipótesis de que es adecuado emplear modelos lineales jerárquicos.

Cuadro 6: Coeficiente de correlación interclase del modelo vacío

<b>Cuartil</b>	<b>CCI (%)</b>
0.25	64
0.50	76
0.75	68

Fuente: Cálculos propios a partir de Saber 11, 2014-2

### 5.1 ENTORNO DE DESARROLLO

En el Cuadro 7 se encuentran los resultados de la estimación del modelo jerárquico lineal por cuartiles (LQMM), del modelo lineal jerárquico (HML) y de la regresión cuantílica (RQ). El intercepto, el cual es significativo y positivo para todos los cuartiles, se puede interpretar como el puntaje en matemáticas promedio respectivo a cada cuartil que obtendría un estudiante hombre con 16.82 años en promedio, cuya madre no ha alcanzado un nivel educativo o sólo alcanzó la primaria (completa o incompleta), que vive en un entorno de desarrollo robusto, asiste a un colegio no oficial en la jornada de completa, respectivamente

para cada cuartil. Como era de esperarse la magnitud del coeficiente aumenta a medida que se pasa de cuartil de estudio. El intercepto del modelo HML si bien es significativo, por su misma naturaleza, oculta las diferencias en el desempeño que existen a lo largo de la distribución.

Por otro lado, vivir en un municipio catalogado con un entorno de desarrollo incipiente o intermedio tiene un efecto negativo en el desempeño en matemáticas comparado con vivir en un municipio de entorno de desarrollo robusto. Lo anterior también se evidencia con la aplicación del modelo HML y RQ. Vivir en un municipio intermedio para un estudiante de bajo desempeño (primer cuartil) no tiene un efecto significativo sobre su desempeño; en cambio para un estudiante que se encuentra en la mediana o en la cola superior de la distribución el efecto es significativo con el 99 % de confianza. Por otra parte, un estudiante de la cola inferior que viva en un entorno de desarrollo incipiente obtiene en promedio dos puntos menos que aquel que vive en un entorno de desarrollo robusto, esa diferencia es mayor para aquellos que se encuentran en el segundo o tercer cuartil.

Por otro lado, los coeficientes que acompañan los controles utilizados también son relevantes y evidencian los signos esperados de acuerdo a la revisión de la literatura realizada. El coeficiente de regresión para la edad y el sexo son aproximadamente constantes a lo largo de la distribución y negativos para todos los cuartiles. Aquellos individuos con más edad tienen un desempeño en matemáticas menor a sus pares más jóvenes. Las mujeres tienen un desempeño menor al de los hombres. De hecho, se puede observar que el ser mujer afecta más a los estudiantes de alto desempeño (tercer cuartil). El nivel educativo alcanzado por la madre también es positivo y estadísticamente significativo. Un estudiante cuya madre haya



alcanzado un nivel terciario de educación obtiene en promedio 3 puntos adicionales en promedio si se compara con un estudiante cuya madre no alcanzó algún nivel educativo o sólo alcanzó primaria.

Cuadro 7: Estimación modelo jerárquico lineal, modelo jerárquico lineal por cuantiles y regresión cuantílica por entornos de desarrollo: Puntaje Matemáticas

Variable	HML	Puntaje en matemáticas					
		LQMM			Regresión cuantílica		
		Cuartil 1	Cuartil 2	Cuartil 3	Cuartil 1	Cuartil 2	Cuartil 3
<b>Efectos fijos</b>							
Intercepto	54.91*** (0.11)	49.19*** (0.58)	54.25*** (0.46)	60.15*** (0.37)	47.51*** (0.09)	53.48*** (0.08)	59.67*** (0.19)
Entorno de desarrollo	-1.85*** (0.09)	-1.00 (0.69)	-2.00*** (0.44)	-2.02*** (0.33)	-1.95*** (0.05)	-2.00*** (0.04)	-2.06*** (0.05)
Entorno de desarrollo Intermedio	-3.12*** (0.14)	-2.00* (1.04)	-3.00*** (0.52)	-3.69*** (0.43)	-3.48*** (0.09)	-3.66*** (0.06)	-3.98*** (0.06)
Entorno de desarrollo Incipiente	<b>Efectos fijos: Controles</b>						
Mujer	-2.74*** (0.03)	-2.00*** (0.01)	-2.06*** (0.42)	-3.04*** (0.04)	-1.76*** (0.05)	-2.45*** (0.05)	-2.96*** (0.05)
Edad (centrada)	-1.19*** (0.01)	-1.00*** (0.01)	-1.02*** (0.05)	-1.18*** (0.04)	-1.28*** (0.02)	-1.39*** (0.02)	-1.47*** (0.03)
Nivel educativo madre: secundaria	0.87*** (0.03)	0.99*** (0.01)	0.98*** (0.06)	0.92*** (0.07)	1.00*** (0.05)	1.22*** (0.05)	1.51*** (0.09)
Nivel educativo madre: terciaria	3.04*** (0.04)	2.99*** (0.03)	2.98*** (0.03)	3.35*** (0.11)	4.45*** (0.07)	5.22*** (0.05)	6.53*** (0.10)
Oficial	-1.67*** (0.12)	-1.00 (0.72)	-1.01 (0.70)	-1.60*** (0.25)	-0.29*** (0.05)	-0.63*** (0.05)	-1.14*** (0.06)
Jornada mañana	-2.38*** (0.12)	-3.00*** (0.66)	-3.01*** (0.49)	-2.40*** (0.38)	-2.01*** (0.05)	-2.21*** (0.06)	-2.31*** (0.08)
Jornada tarde	-2.81*** (0.15)	-3.00*** (0.99)	-4.01*** (0.72)	-2.83*** (0.39)	-2.50*** (0.08)	-2.76*** (0.08)	-3.25*** (0.10)
<b>Efectos aleatorios</b>							
Colegio (sd)	3.94	3.23	3.22	3.69			
CCI	20.00 %	50.62 %	54.34 %	56.18 %			

Error estándar entre paréntesis

\*\*\* Nivel de confianza 99 % \*\* Nivel de confianza 95 % \* Nivel de confianza 90 %  
Fuente: Cálculos propios a partir de Saber 11, 2014-2

Respecto a las variables del colegio, se encuentra que asistir a un colegio oficial está correlacionado negativa y significativamente sólo con aquellos estudiantes que se encuentran en la cola superior, lo anterior no es capturado por el modelo HML o RQ. Se ha encontrado para el caso colombiano que esto puede obedecer no solamente a unas mejores condiciones socioeconómicas de los estudiantes que asisten a estos colegios, también a la forma en que son administrados y estructura de incentivos que se maneja al interior de estos (Gaviria y Barrientos, 2001).

De igual forma se encuentra que los estudiantes que asisten a la jornada de mañana o de la tarde tienen un menor desempeño que aquellos que asisten a la jornada completa. Ese efecto negativo es mayor para aquellos que asisten a la jornada de la tarde que para aquellos que asisten a la jornada de la mañana. Por último, se puede observar que los coeficientes de correlación interclase para el modelo LQMM disminuyeron respecto a los obtenidos con el modelo vacío. Estos indican que aproximadamente el 50 % de la variabilidad en los resultados se debe a factores del colegio, una vez se controla por ciertas variables de la dimensión personal, familiar y del lugar de residencia del individuo.

## 5.2 TIPOLOGÍA MUNICIPAL

Los resultados encontrados por tipología son similares que los que derivaron de emplear la variable de entorno de desarrollo. Lo anterior, era de esperarse puesto que la tipología municipal es una subcategorización de los entornos de desarrollo. Al igual que en la sección

anterior el intercepto para bajo las tres metodologías es positivo y significativo (Ver Cuadro 8).

Al analizar por las tipologías municipales que componen cada entorno se encuentra que para aquellos estudiantes (de la parte inferior o superior de la distribución del puntaje) que viven en un municipio de tipología B hay una asociación significativa entre obtener puntajes levemente mayores que los estudiantes que viven en municipios de tipología A; es decir en los municipios que de acuerdo a la clasificación del DNP son los que tienen las mejores condiciones de desarrollo.

Al comparar con los resultados de la regresión cuantílica, el efecto sólo se captura para el estudiante en el tercer cuartil, y el modelo HML evidencia que hay un efecto positivo y significativo, pero no logra capturar los cambios que hay a lo largo de la distribución.

Por otro lado, al revisar las tipologías municipales del entorno de desarrollo intermedio e incipiente se evidencia que vivir en cualquiera de estos municipios está asociado con tener un menor resultado que aquellos que viven en un municipio de tipología A, independientemente del cuartil en el que se encuentren y de la metodología aplicada. La única excepción a esa generalización se encuentra en el estudiante de bajo y alto desempeño que vive en un municipio de tipología C, ya que para éste no se encontró una relación significativa por la metodología LQMM.

Además, se evidencian algunas diferencias en la magnitud y significancia de la asociación encontrada entre tipología municipal y puntaje en matemáticas. Por ejemplo, hay una diferencia entre pertenecer a un municipio de tipología D y E de casi un punto si se es un

estudiante de alto desempeño, esta diferencia se tendría que confirmar si es estadísticamente significativa. Otro ejemplo que se puede observar es el del estudiante que se encuentra en el segundo cuartil y vive en un municipio de tipología C. Para éste se halla que obtiene 1 punto menos en la prueba de matemáticas que el que vive en una tipología A; mientras que, aquellos que viven en un municipio de tipología D o E obtienen dos puntos menos, esto con un 99% de confianza.

Para las tipologías que conforman el entorno de desarrollo incipiente (F y G) no se observa que haya diferencias a lo largo de la distribución, es decir resulta casi lo mismo vivir en un municipio de tipología F que vivir en uno de tipología G. Lo anterior puesto que en ambos casos se obtiene en promedio tres puntos menos en la prueba de matemáticas para el estudiante del primer y segundo cuartil respecto al estudiante que vive en la tipología A, y de aproximadamente 4 puntos menos para el estudiante de alto desempeño.

Por último, los coeficientes que acompañan las variables de control (individuales, familiares y del colegio) son en su mayoría significativas y tienen los signos esperados

Cuadro 8: Estimación modelo jerárquico lineal, modelo jerárquico lineal por cuantiles y regresión cuantílica por tipología municipal: Puntaje Matemáticas

Variable	HML	Puntaje en matemáticas					
		LQMM			Regresión cuantílica		
		Cuartil 1	Cuartil 2	Cuartil 3	Cuartil 1	Cuartil 2	Cuartil 3
<b>Efectos fijos</b>							
Intercepto	54.81*** (0.11)	49.20*** (0.70)	54.23*** (0.43)	60.13*** (0.30)	47.58*** (0.08)	53.52*** (0.07)	59.39*** (0.15)
<b>Entorno Robusto</b>							
Tipología B	0.22** (0.09)	0.99** (0.50)	0.00 (0.33)	0.49** (0.24)	-0.08 (0.06)	0.00 (0.04)	0.25*** (0.08)
Tipología C	-1.08***	-1.00	-1.01**	-0.86	-1.25***	-1.19***	-0.92***

	(0.12)	(0.74)	(0.38)	(0.39)	(0.08)	(0.05)	(0.08)
Entorno Intermedio							
Tipología D	-1.92***	-2.01**	-2.01***	-1.95***	-2.08***	-2.08***	-2.01***
	(0.13)	(0.80)	(0.45)	(0.34)	(0.06)	(0.06)	(0.06)
Tipología E	-2.59***	-2.01**	-2.03***	-2.93***	-3.08***	-3.24***	-3.34***
	(0.14)	(0.82)	(0.44)	(0.18)	(0.08)	(0.07)	(0.09)
Entorno Incipiente							
Tipología F	-3.07***	-3.01***	-3.01***	-3.61***	-3.43***	-3.58***	-3.62***
	(0.16)	(0.99)	(0.31)	(0.29)	(0.11)	(0.06)	(0.10)
Tipología G	-3.19***	-3.00**	-3.02***	-3.88***	-3.96***	-4.08***	-4.12***
	(0.21)	(1.36)	(0.33)	(0.36)	(0.10)	(0.08)	(0.11)
<b>Efectos fijos:</b>							
<b>Controles</b>	Sí		Sí			Sí	
<b>Efectos aleatorios</b>	3.91	3.21	3.22	3.69			
<b>Colegio (sd)</b>	20.00 %	50.44 %	54.41 %	56.15%			

Error estándar entre paréntesis

\*\*\* Nivel de confianza 99 % \*\* Nivel de confianza 95 % \* Nivel de confianza 90 %

Fuente: Cálculos propios a partir de Saber 11, 2014-2

## 6. CONCLUSIONES

Lo anterior cobra aún más importancia si se tiene en cuenta los hallazgos del estudio de Cendales et al. (2015). En éste los autores establecen que el sistema de educación media en la zona industrial de Colombia (Medellín, Cali y Bogotá, las tres hacen parte del entorno de desarrollo robusto y la tipología A) incentiva la reproducción de una estructura social inequitativa. Lo anterior puesto que los estudiantes de bajos recursos tienen menores oportunidades de acceder a la educación superior, por ende tienen a repetir la baja productividad que caracteriza a sus padres; lo mismo sucede para aquellos con mejores condiciones socioeconómicas. Los autores concluyen que si se tiene en cuenta que en estas ciudades la mayor cantidad de la mano obra industrial (91% en 2012 y 2013) corresponde a estudiantes de bajo recursos entonces estos resultados tienen implicaciones negativas en el desarrollo de la industria moderna del país.

El problema de tener una estructura social inequitativa que se regenera a sí misma se vuelve mayúsculo cuando se tiene en cuenta que en los 1117 municipios restantes, donde vive el 73.3% de la población de estudiantes de educación media según las pruebas Saber 11 2014-2, el 96.2% de los estudiantes son de bajos recursos. Cuantificar la magnitud de ese potencial obstáculo en el desarrollo de la industria moderna del país requiere de su estudio respectivo, pero estas cifras ponen de manifiesto la importancia de incluir en estos análisis a los municipios de los entornos de desarrollo intermedio e incipiente.

Asimismo, otro aspecto meritorio de investigación, y que se relaciona con lo anterior, es la diferencia que se observa entre los estudiantes que asisten a colegios oficiales de aquellos que asisten a colegios privados, en beneficio de este segundo grupo. Lo anterior resulta relevante porque el 77.5% de los estudiantes a nivel nacional asisten a colegios oficiales, esa cifra es de más del 90% en los entornos de desarrollo menos avanzados. Por ejemplo, Celis et al. (2012) encuentran que existen diferencias entre el sector oficial y no oficial, de hecho en las instituciones oficiales la varianza asociada a la sede educativa alcanza el 11%; mientras que en el sector no oficial puede llegar hasta el 22% en algunas regiones. Es decir, los estudiantes que ingresan a un plantel privado de alta calidad pueden recibir mucho de parte del éste que contribuya a su logro educativo en relación a los que asisten a colegios oficiales, y aún más a los que asisten a instituciones no oficiales de baja calidad. Por lo tanto, a los hacedores de política pública educativa debería interesarles si hay una relación sistemática entre las condiciones del entorno y la calidad de la educación que ofrecen las instituciones educativas (oficiales y no oficiales) para poder incluir este aspecto dentro de la formulación de la política pública. Una forma de contribuir a la solución de esta incógnita sería incluir

interacciones entre la naturaleza del colegio y el entorno de desarrollo en el modelo estimado presentado en este estudio.

Por otro lado, entre las limitaciones del estudio se encuentra que hay varias variables a nivel de colegio y de la familia que no se tuvieron en cuenta por restricciones de los datos, por ejemplo, los años de experiencia de los profesores, tamaño de la clase, salarios de los profesores o la mensualidad del colegio, el estrato socioeconómico o el ingreso de la familia. Una segunda limitación consiste en que no diferenció a los colegios privados que prestan su servicio al sector educativo oficial, de los que no. Lo anterior podría ser importante puesto que se ha encontrado que hay una diferencia entre los estudiantes que asisten a los colegios por concesión de los que asisten a colegios oficiales (Godoy, 2013).

Una tercera limitación consiste en que, aunque aquí se incluye una conceptualización del entorno de desarrollo, se desconoce cuál o cuáles dimensiones tienen una asociación más fuerte con el desempeño y por lo tanto en qué condición o característica del lugar donde reside el estudiante se deberían focalizar la inversión pública. De igual forma el índice construido por el DNP carece de una dimensión importante que es la educativa. En esta se podría incluir elementos como si la secretaría de educación municipal está certificada, nivel educativo de los profesores, características de la infraestructura educativa, entre otras. Luego, un próximo estudio podría construir un índice de desempeño educativo que permita analizar cómo esos factores educativos, que se derivan de decisiones institucionales, afectan el desempeño de los estudiantes en matemáticas, y otras competencias básicas, como lectura crítica e inglés.

Asimismo, en investigaciones posteriores se debería tener en cuenta el tiempo. Por ahora el entorno de desarrollo sólo se encuentra para el 2014; si el DNP sigue construyendo la serie se podría cuantificar el efecto de cambios en el entorno de desarrollo en los resultados de matemáticas, por ejemplo, incluyendo la variable dependiente rezagada. Finalmente, esperamos que una futura investigación incluya las interacciones entre los niveles del modelo jerárquico que logren evidenciar si hay una relación entre el lugar donde reside el estudiante y las características del colegio al que asiste.

## REFERENCIAS

Abadía, L. K y Bernal, G. (2016). Brechas de género en el rendimiento escolar a lo largo de la distribución de puntajes: evidencia pruebas Saber 11.

Aguilar Londoño, C. M., Osejo Villamil, I., Carmona Sanchez, C. O., Supelano González, D., Gaitán Alvarez, J., & Taborda, J. C. (2015). Tipologías municipales de Colombia: herramienta para la consolidación de regiones desarrolladas

Barón, J. D. (2012). Diferencias en las características de los estudiantes y la brecha de rendimiento académico entre Barranquilla y Bogotá: Una descomposición semiparamétrica. *Ensayos sobre Política Económica*, 30(68), 164-215.

Barrientos, J. H., & Rios, P. (2007). Evaluación de la gestión privada del servicio público educativo en Medellín. *Lecturas de Economía*, 66, 147.



- Barrientos, M. (2012). Diferencias regionales en la distribución del ingreso en Colombia. *Sociedad y Economía*, 108, 43–68.
- Bassett Jr, G. W., Tam, M. Y. S., & Knight, K. (2002). Quantile models and estimators for data analysis. *Metrika*, 55(1-2), 17-26.
- Bonet, J., & Bonett, J. (2005). *Inequidad espacial en la dotación educativa regional en Colombia*. Banco de la República.
- Castaño, E. (2010). El efecto colegio sobre la variabilidad del rendimiento en matemáticas. *Lecturas de economía*, 49(49), 47-57.
- Cendales, A., Mora, J.J., y Campo, J. (2015). Educación media y mano de obra industrial en Colombia: Inequidad y pobreza, En VI Seminario Internacional de Investigación sobre la Calidad Educativa Bogotá.
- Chiu, M. M., & Xihua, Z. (2008). Family and motivation effects on mathematics achievement: Analyses of students in 41 countries. *Learning and Instruction*, 18(4), 321-336.
- Cooray, A. (2009), The role of education in economic growth, en 38 Australian Conference of Economists University of Wollongong, 1–27.
- Cortés, D. y Vargas J.F (2012). Inequidad Regional en Colombia, *Serie Documentos Cede*, 34.
- Costanzo, A. (2015). The Effect of M@ tabel on Italian Students' Performances: A Quantile Regression Approach. *Procedia-Social and Behavioral Sciences*, 197, 236-244.

Educación Nacional, Ministerio, Colombia, la mejor educada en el 2025, Technical Report, Ministerio Nacional de Educación, Bogotá 2015.

Delgado, M. S., Henderson, D. J., & Parmeter, C. F. (2014). Does education matter for economic growth?. *Oxford Bulletin of Economics and Statistics*, 76(3), 334-359.

Demir, İ., Kılıç, S., & Ünal, H. (2010). Effects of students' and schools' characteristics on mathematics achievement: findings from PISA 2006. *Procedia-Social and Behavioral Sciences*, 2(2), 3099-3103.

DfEE (1999). A, Moser Report.

Eide, E., & Showalter, M. H. (1998). The effect of school quality on student performance: A quantile regression approach. *Economics letters*, 58(3), 345-350.

Foy, P. (2004). Intraclass correlation y variance components as population attributes y measures of sampling efficiency in PIRLS 2001, in The 1st IEA International Research Conference, 1–13.

Fryer, R. G., & Levitt, S. D. (2010). An empirical analysis of the gender gap in mathematics. *American Economic Journal: Applied Economics*, 2(2), 210-240.

Galvis, L. A., & Bonilla, L. (2012). Desigualdades regionales en el nivel educativo de los profesores en Colombia. *Revista de Economía Institucional*, 14(26), 223-240.

Gaviria, A., & Barrientos, J. (2001). Características del plantel y calidad de la educación en Bogotá. *Coyuntura Social*, 25, 81-98.

Geraci, M., & Bottai, M. (2007). Quantile regression for longitudinal data using the asymmetric Laplace distribution. *Biostatistics*, 8(1), 140-154.

- Geraci, M., & Bottai, M. (2014). Linear quantile mixed models. *Statistics and Computing*, 24(3), 461-479.
- Godoy, S. (2013), Análisis del programa de ampliación de cobertura en Cali con respecto a la calidad de la educación, medido a través de las pruebas de Estado.
- Haile, G. A., & Nguyen, A. N. (2008). Determinants of academic attainment in the United States: A quantile regression analysis of test scores. *Education Economics*, 16(1), 29-57.
- Hanushek, E. A. (1986). The economics of schooling: Production and efficiency in public schools. *Journal of economic literature*, 24(3), 1141-1177.
- Hanushek, E. A. (2002). Evidence, politics, y the class size debate, in Lawrence Mishel y Richard Rothstein, eds., *The Class Size Debate*, Washington, DC: Economic Policy Institute, chapter 2, 37–66.
- Hanushek, E. A. (2013). Economic growth in developing countries: The role of human capital. *Economics of Education Review*, 37, 204-212.
- Hanushek, E. A., & Woessmann, L. (2008). The role of cognitive skills in economic development. *Journal of economic literature*, 46(3), 607-668.
- Hoerandner, C. M., & Lemke, R. J. (2006). Can No Child Left Behind Close the Gaps in Pass Rates on Standardized Tests?. *Contemporary Economic Policy*, 24(1), 1-17.
- Hox, J. (2002). *Multilevel Analysis: Techniques y Applications.*, New Jersey: Lawrence Erlbaum Associates.
- Huizinga, M. M., Beech, B. M., Cavanaugh, K. L., Elasy, T. A., & Rothman, R. L. (2008). Low numeracy skills are associated with higher BMI. *Obesity*, 16(8), 1966-1968.

- ICFES (2015), Guía de Interpretación y Uso de Resultados de las pruebas SABER 11, 57.
- Jimenez, E., Lockheed, M. E., & Paqueo, V. (1991). The relative efficiency of private and public schools in developing countries. *The World Bank Research Observer*, 6(2), 205-218.
- Koenker, R., & Machado, J. A. (1999). Goodness of fit and related inference processes for quantile regression. *Journal of the american statistical association*, 94(448), 1296-1310.
- Koenker, R. (2005). *Quantile regression* (No. 38). Cambridge university press.
- Koenker, R., & Bassett Jr, G. (1978). Regression quantiles. *Econometrica: journal of the Econometric Society*, 33-50.
- Lavy, V. (2015). Do differences in schools' instruction time explain international achievement gaps? Evidence from developed and developing countries. *The Economic Journal*, 125(588), F397-F424.
- Litster, J. (2013). The impact of poor numeracy skills on adults, Technical Report , University of London, London.
- Lubienski, C., & Lubienski, S. T. (2006). *Charter, private, public schools and academic achievement: New evidence from NAEP mathematics data* (Vol. 16). New York: National Center for the Study of Privatization in Education, Teachers College, Columbia University.
- Marín, J. B. (2008). Calidad de la educación pública y logro académico en Medellín 2004-2006. Una aproximación por regresión intercuartil. *Lecturas de Economía*, 68(68), 121-144.
- Marino, M. F., & Farcomeni, A. (2015). Linear quantile regression models for longitudinal experiments: an overview. *Metron*, 73(2), 229-247.

Mateus, W. A. M. (2016). Análisis de distribución geográfica y espacial de los resultados de las Pruebas Saber 11 del Instituto Colombiano para el Fomento de la Educación Superior-ICFES-. 2005-2012. Colombia. *Cuadernos Latinoamericanos de Administración*, 11(21), 39-50.

Meisel, R. y Romero, J. (2007). Igualdad de oportunidades para todas las regiones, en Fernández, M., Guerra W., y Meisel A., eds., Políticas para reducir las desigualdades regionales en Colombia, primera ed., Cartagena: Banco de la República, 1–415.

Melzer, David, E Gardener, Iain Lang, B McWilliams, y JM Guralnik, Retirement , health y relationships of the older population in Engly: The 2004 English Longitudinal Study of Ageing number July 2006.

Ministerio de Educación Nacional (2010). Plan sectorial de educación 2010 - 2014, Technical Report.

Ministerio de Educación Nacional (2014). El ICFES cambia la clasificación de resultados, jul 2014.

Ministerio de Educación Nacional (2015), Información de la prueba Saber 11.

Montes, I., Garcés, J. D., Chica, S. M., & Jaramillo, A. (2014). Rendimiento académico:¿ qué papel juegan los factores institucionales?. Documentos de trabajo Economía y Finanzas, (14), 1–39.

OCDE (2007). El programa PISA de la OCDE Qué es y para qué sirve, Technical Report, Organización para la cooperación y el desarrollo económico, París.

OECD (2014). PISA 2012 Results in Focus, Technical Report.

Jiménez, L. J. P., & Pinzón, A. R. (1998). Los insumos escolares en la educación secundaria y su efecto sobre el rendimiento académico de los estudiantes: un estudio en Colombia. *Washington, DC: The World Bank/Latin America and the Caribbean Region/Department of Human Development.*

Raudenbush, S., & Bryk, A. S. (1986). A hierarchical model for studying school effects. *Sociology of education*, 1-17.

Restrepo, P. P., & Alviar, M. (2009). El logro académico y el efecto colegio en las pruebas ICFES en Antioquia. *Lecturas de Economía*, 60(60), 67-95.

Reyna, V. F., & Brainerd, C. J. (2007). The importance of mathematics in health and human judgment: Numeracy, risk communication, and medical decision making. *Learning and Individual Differences*, 17(2), 147-159.

Rivkin, Steven G y Jeffrey C Schiman, Instruction time, classroom quality, y academic achievement, *The Economic Journal*, 2015, 125 (588), F425–F448.

Sánchez, C., Supelano, D., y Osejo, I. (2014). Tipologías Departamentales y Municipales: una propuesta para comprender las entidades territoriales colombianas, Technical Report, Departamento Nacional de Planeación - Dirección de Desarrollo Territorial Sostenible.

Sánchez, B. L., Lachos, H. V., & Labra, V. F. (2013). Likelihood based inference for quantile regression using the asymmetric Laplace distribution. *Journal of Statistical Computation and Simulation*, 81, 1565-1578.

Smith, A. (2004). Making mathematics count, Report Commissioned by the UK Government, 2004.

Smith, L. B., Fuentes, M., Gordon-Larsen, P., & Reich, B. J. (2015). Quantile regression for mixed models with an application to examine blood pressure trends in China. *The Annals of Applied Statistics*, 9(3), 1226-1246.

Tian, M., & Chen, G. (2006). Hierarchical linear regression models for conditional quantiles. *Science in China Series A: Mathematics*, 49(12), 1800-1815.

Tzavidis, N., & Brown, J. (2010). Using M-quantile models as an alternative to random effects to model the contextual value-added of schools in London.

Yu, K., Lu, Z., & Stander, J. (2003). Quantile regression: applications and current research areas. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 52(3), 331-3

## ANEXOS

Cuadro 12: Diccionario de Variables

Variable	Descripción	Categoría Base
munintermedia <sub>ij</sub>	Variable dummy, toma el valor de 1 si el estudiante <i>i</i> que asiste al colegio <i>j</i> vive en un municipio de entorno de desarrollo intermedio, 0 en caso contrario.	
munincipiente <sub>ij</sub> ,	Variable dummy, toma el valor de 1 si el estudiante <i>i</i> que asiste al colegio <i>j</i> vive en un municipio de entorno de desarrollo incipiente, 0 en caso contrario.	munrobusto <sub>ij</sub> ,
edad <sub>ij</sub>	La edad del individuo se encuentra centrada en 16.81971 años, por lo que esta variable indica el número de años por debajo o por encima de esa cifra que el individuo <i>i</i> que asiste al colegio <i>j</i> tiene.	
mujer <sub>ij</sub>	Variable dummy, toma el valor de 1 si el estudiante <i>i</i> que asiste al colegio <i>j</i> es mujer, 0 en caso contrario.	hombre <sub>ij</sub>
msecundaria <sub>ij</sub>	Variable dummy, toma el valor de 1 si nivel educativo más alto alcanzado por la madre del estudiante <i>i</i> que asiste al colegio <i>j</i> es de secundaria incompleta o completa, 0 en caso contrario.	Ninguno o primaria

	Variable dummy, toma el valor de 1 si nivel educativo más alto alcanzado por la madre del estudiante $i$ que asiste al colegio $j$ es de técnica, tecnológica, universitaria o de postgrado.	
$mterciaria_{ij}$		
$oficial_{ij}$	Variable dummy, toma el valor de 1 si el colegio $j$ al que asiste el estudiante $i$ es oficial, 0 en caso contrario.	$nooficial_{ij}$
$jmanana_{ij}$	Variable dummy, toma el valor de 1 si la jornada a la que asiste el estudiante $i$ del colegio $j$ es la de la mañana, 0 en caso contrario.	
$jtarde_{ij}$	Variable dummy, toma el valor de 1 si la jornada a la que asiste el estudiante $i$ del colegio $j$ es la de la tarde, 0 en caso contrario.	$jcompleta_{ij}$
$B_{ij}$	Variable dummy, toma el valor de 1 si el estudiante $i$ del colegio $j$ vive en un municipio de tipología B, 0 en caso contrario.	
$C_{ij}$	Variable dummy, toma el valor de 1 si el estudiante $i$ del colegio $j$ vive en un municipio de tipología C, 0 en caso contrario.	
$D_{ij}$	Variable dummy, toma el valor de 1 si el estudiante $i$ del colegio $j$ vive en un municipio de tipología D, 0 en caso contrario.	
$E_{ij}$	Variable dummy, toma el valor de 1 si el estudiante $i$ del colegio $j$ vive en un municipio de tipología E, 0 en caso contrario.	
$F_{ij}$	Variable dummy, toma el valor de 1 si el estudiante $i$ del colegio $j$ vive en un municipio de tipología F, 0 en caso contrario.	$A_{ij}$
$G_{ij}$	Variable dummy, toma el valor de 1 si el estudiante $i$ del colegio $j$ vive en un municipio de tipología G, 0 en caso contrario.	

---