



La relación entre el lugar de origen del estudiante y sus resultados de Saber 11

Gustavo Adolfo Cedeño Ocampo

Directora del proyecto:
MSc. Daniela Estrada Nates

Universidad Icesi
Facultad de Ciencias Administrativas y Económicas
Economía Y Contaduría Pública y Finanzas Internacionales
Santiago de Cali
2019

Resumen

En este trabajo analiza la relación del lugar de origen y el rendimiento académico del estudiante en las pruebas Saber 11, realizado por el Icfes, para el año 2018. Empleando una regresión logística multinivel y controlando por características personales, familiares, y escuela. Se encuentra que un estudiante de vereda tiene un efecto negativo y significativo en la probabilidad en obtener un buen desempeño. No obstante, el efecto varía por la región donde reside el estudiante, y la infraestructura de la escuela. Los resultados obtenidos se comparan por niveles socioeconómicos, en promedio, pasar de bajo a medio implica una probabilidad mayor del 10 % de obtener mejores resultados y de medio a alto del 25 %, siendo esta la brecha más amplia.

Palabras clave: Educación, veredas, unidad territorial de Colombia, regresión logística multinivel.

Abstract

This paper analyzed the relationship of the home region and the academic achievement of the student in Saber 11 tests, performed by Icfes in 2018. Using a multilevel logistic regression and controlling for personal, family, and school characteristics. It is found that a countryside student has a negative and significant effect on the probability of obtaining a good performance. However, the effect varies by the states where the student resides, and the school infrastructure. The results obtained are compared by socioeconomic levels, on average, going from low to medium implies a greater than 10 % probability of obtaining better results and from medium to high of 25 %, this being the widest gap.

Keywords: Education, countryside, territorial unit of Colombia, multilevel logit model.

Contenido

Resumen	II
1. Introducción	1
2. Revisión de la literatura	5
2.1. Investigaciones internacionales	5
2.2. Investigaciones nacionales	6
3. Aproximación empírica	9
3.1. ¿Qué es un buen desempeño académico?	9
3.2. Índice del nivel socioeconómico	10
3.3. Regresión logística de modelos lineales jerárquicos	12
3.4. Prueba de anidación	14
4. Base de datos	15
4.1. Saber 11	15
4.1.1. Índice sintético de calidad educativa	16
4.2. Información de las escuelas de las veredas colombianas	16
4.3. Educación formal	17
4.4. Estadística descriptivas	18
5. Resultados	21
5.1. Entorno del estudiante	21
5.2. Probabilidad de obtener un buen desempeño en el Saber 11	23
6. Conclusiones	26
A. Regiones de Colombia y estadísticas descriptivas	27
B. Número óptimo de cluster en INSE	29

1. Introducción

En el año 2015, Colombia se ubicó en el puesto 59 entre 70 países que participaron en las pruebas PISA¹ OCDE (2016), demostrando el bajo rendimiento académico que tienen los estudiantes del país. Esto es preocupante porque la educación es primordial para la función de producción de capital humano, debido a que, es un factor clave para el crecimiento de un país John (2010). Además, quien obtiene un bajo rendimiento escolar tienen más restricciones para participar en el mercado laboral Hanushek (2013), y está relacionado significativamente con una profundización del ciclo de las trampas de pobreza Accinelli et al. (2007).

Al mismo tiempo, en Colombia también se mide con una prueba estandarizada el desempeño académico conocida con Saber 11. Ésta no permite realizar comparaciones entre países, pero si evidencia que al interior del país hay diferencias en el desempeño académico entre Unidades Territoriales (UT)² (ver Figura 1-1). En el 2014 arrojó un resultado para los estudiantes de veredas de 232 puntos, el puntaje aumenta ligeramente por UT: 10 puntos para municipio; y 16 puntos para capital. Pasando los años esta brecha se está incrementando.

Además, hay que tener en cuenta que los estudiantes que obtienen bajos rendimientos en el Saber 11 tienen cinco veces más probabilidad de estudiar una carrera pedagógica en comparación con los que obtuvieron mejores resultados Barón y Bonilla (2011), y siguen obteniendo el mismo bajo desempeño cuando obtienen su título profesional Barón y Bonilla (2014). Esto quiere decir que, los docentes se encuentran en un círculo vicioso de bajo desempeño académico, y esto puede afectar la calidad de la educación que imparten.

De igual manera, en la Figura 1-2 se encuentra la distribución poblacional del nivel educativo de los profesores por UT y por región. Se puede observar que existen diferencias proporcionales en el nivel educativo de docentes por UT y región, donde las veredas tienen menos profesores con posgrado pedagógico. Lo anterior se debe a que esta zona enfrenta grandes dificultades para atraer y retener a los docentes capacitados debido a la alta provisionalidad de la planta Bonilla-Mejía et al. (2018). De hecho autores como Villegas et al. (2013) han encontrado que no solo que el sistema escolar en Colombia sea inequitativo y afecte la igualdad de oportunidades de los estudiantes, especialmente para la zona rural.

¹PISA es una prueba estandarizada y evalúa a estudiantes de 15 años de diferentes países del mundo para comparar el rendimiento escolar de los diferentes sistemas educativos del mundo que presentan la prueba OCDE (2016).

²Unidad Territorial hace referencia a capital distrital (departamental), municipios, y veredas de Colombia.

Figura 1-1.: Puntaje promedio global por unidad territorial.



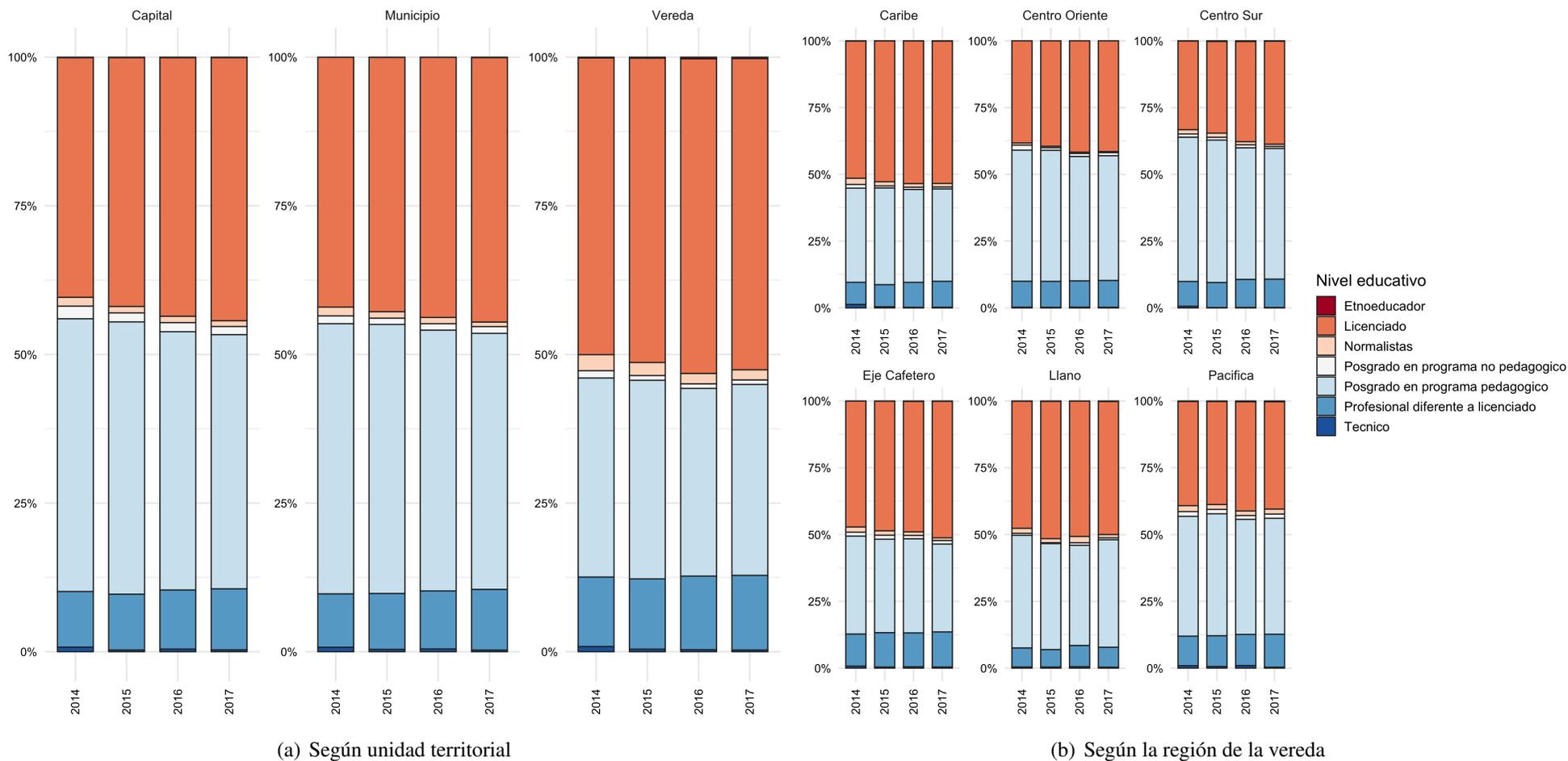
Fuente: Cálculos propios a partir del Saber 11, 2014 al 2018.

Por otro lado, en la Figura 1-3, (a) se encuentra la variación porcentual nacional del número de estudiantes que presentaron las pruebas Saber 11 por año. Este último, desde el 2008 hasta el 2015, ha tenido una reducción de crecimiento poblacional hasta el punto de tener un decrecimiento. No obstante, (b) la tasa es diferente para cada UT, pues los estudiantes de veredas han liderado el crecimiento poblacional para todos los años analizados (2008 hasta 2018), esto quiere decir, que la zona rural está creciendo en mayor proporción que la zona urbana.

Hasta el momento no se ha realizado ningún estudio de carácter cuantitativo sobre el desempeño académico de la población que residen en las veredas, el presente trabajo espera contribuir a llenar un vacío actual en la literatura colombiana al responder: ¿Un estudiante que viva en una vereda tiene la misma probabilidad en obtener un buen desempeño en las pruebas Saber 11 que un estudiante que viva en una capital en Colombia, controlando por factores personales, de la familia y el colegio?

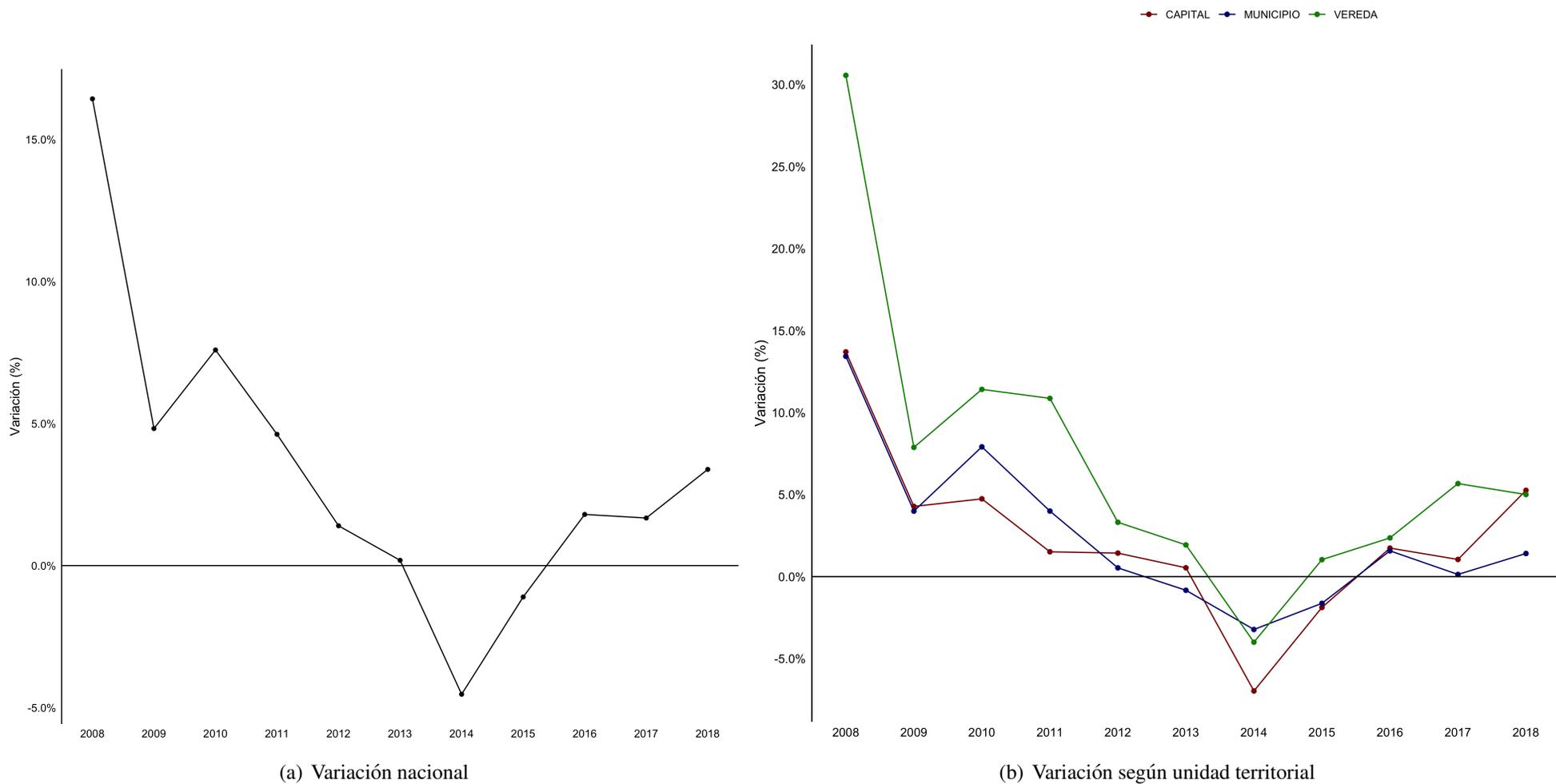
Por último, este artículo contiene además de la introducción, cinco secciones adicionales. En la siguiente a esta se describen los factores que puedan afectar al desempeño académico, y la evidencia empírica a nivel internacional y nacional que se ha encontrado al respecto. En la tercera parte se explica, ¿qué es un buen desempeño académico para la zona rural de Colombia?, y la aproximación empírica. Posteriormente, se caracterizan los datos empleados y se presentan las estadísticas descriptivas. En la quinta sección se encuentran los resultados de la regresión logística de modelos lineales jerárquicos y su respectivo análisis. En la última sección se encuentran las conclusiones.

Figura 1-2.: Evolución del nivel educativo de los docentes de educación media.



Fuente: Cálculos propios a partir de Educación Formal del DANE, 2014 al 2017.

Figura 1-3.: Variación porcentual anual del número de estudiantes.
Estudiantes que presentaron las pruebas Saber 11.



Fuente: Cálculos propios a partir del Saber 11, 2008 al 2018.

2. Revisión de la literatura

Actualmente, hay trabajos como los de Aldana et al. (1989), Restrepo y Alviar (2004), Gordillo y Ariza (2005), Iregui et al. (2006), Ramírez y Téllez (2006), Banco Mundial (2009), y Garces y Montes (2015), que estudian el bajo rendimiento académico en Colombia y dan sugerencias acerca de cómo solucionar este problema. Sin embargo, en esta literatura (y en la gran mayoría) no se tiene en cuenta el entorno del estudiante medido como la UT en la que vive. Diferentes estudios que realizan una aproximación a este factor encuentran que éste es importante para el logro educativo. La revisión bibliográfica se divide en el ámbito internacional y nacional, y exponen algunos de los factores que afectan el desempeño educativo.

2.1. Investigaciones internacionales

Inicialmente, Coleman et al. (1966) realizan las primeras investigaciones sobre los determinantes del logro educativo individual. Los autores emplean una función de producción educativa para el caso de Estados Unidos, en el que incorporaban las variables observables del individuo y de la escuela, pero ignoran el clima escolar, el nivel educativo del docente, la composición socioeconómica de las instituciones, entre otras. Los autores demuestran que los insumos o características con la que contaba una escuela no tenían un efecto importante en el logro educativo de los estudiantes, sino que el entorno familiar y las características del hogar influían más.

En el mismo sentido, autores como Hanushek (1986), Becker (1993), Benabou (1996), Patrinos et al. (2007), Baker et al. (2002), Eide y Showalter (1998), entre otros, demuestran que la calidad educativa está influida por el entorno familiar y de la escuela.

Ahora bien, es necesario mencionar estudios que incluyen variables que corresponde al entorno municipal o, en su defecto, proxies. Por ejemplo, en el caso de Argentina, Gertel et al. (2007) realizaron un estudio del rendimiento escolar por medio de un modelo jerárquico de tres niveles: hogar, salón de clase y escuela. El primer nivel contenía información del estudiante como sexo, promedio de notas, Índice de Nivel Socioeconómico (INSE) del hogar, entre otros. El segundo nivel abarcaba los factores que influían el proceso de enseñanza, tales como, número de alumnos por salón, educación y experiencia del docente, estado del aula, etc. El último contenía la gestión de las escuelas como clima escolar, metodología de enseñanza y la naturaleza de la escuela: privada u oficial. También, incluyeron variables dummy para controlar el efecto de las diferencias entre regiones del país. Los resultados encontrados son similares a los de Coleman et al.

(1966), y adicionalmente, encontraron que la población estudiantil presenta diferencias en los logros académicos según la región de residencia.

Paralelamente, Blanco (2011) crítica los estudios de eficacia escolar que suponen que las escuelas deben de ser comprendidas independientemente de las características de su entorno social. Inicialmente, el autor argumenta que los procesos educativos tiene una fuerte dependencia respecto al entorno, lo anterior es un factor clave en los resultados de esta investigación. En segundo lugar, y al igual que Manski (1993), demuestra que una vez controlada la identificación de los efectos de la interacción social, se obtienen los *efectos netos* de la escuela, tarea compleja debido a que éstos no se entienden bajo la lógica de *ceteris paribus*. Por último, el autor afirma que los indicadores de contexto social (barrio o municipio) no logran capturar la heterogeneidad socioeconómica dentro de la escuela. Por consiguiente, planteó un modelo multinivel que soporta su primer argumento con el objetivo de desagregar la varianza de cada nivel. En resumen, el artículo muestra los mismos resultados de los anteriores investigaciones, pero, el autor incluye el entorno de los individuos y la interacción entre ellos. Los niveles de escuela y entorno municipal son responsables de casi del 40 % de la variabilidad total en definir el puntaje de matemáticas y español.

Soares y Alves (2013) investigan simultáneamente el efecto de las municipalidades y las escuelas sobre el logro educativo básica individual. En el trabajo de investigación plantean un modelo de regresión lineal jerárquico de tres niveles: las características personales de los estudiantes, las dotaciones de las escuelas, y por último, el entorno municipal. Los principales resultados obtenidos van en la misma línea de la literatura. Además, los autores encontraron que las ciudades pequeñas tienden a obtener mejores desempeños educativos, lo que consideran se debe a, que tienen menos escuelas que controlar.

Sintetizando, las investigaciones internacionales evidencian la importancia de la escuela y del municipio en la determinación del desempeño educativo. Además, muestran que las dotaciones del entorno familiar y las características del hogar son relevantes en el rendimiento académico.

2.2. Investigaciones nacionales

Esta sección presenta los limitados artículos que utilizan métodos econométricos y emplean una aproximación del entorno municipal donde vive el individuo.

Gaviria y Barrientos (2001a,b), realizaron dos investigaciones, una sobre Bogotá y otra a nivel nacional, con el objetivo de analizar la incidencia de los factores individuales sobre el rendimiento académico. Para ello, utilizaron un modelo de regresión lineal, donde la variable dependiente era el puntaje del ICFES y las variables independientes eran el sexo, la edad del estudiante, la educación de los padres, y las características de la escuelas: tipo de jornada y la naturaleza. En ambas investigaciones afirmaron que el nivel educativo de los padres es un determinante en el rendimiento académico de los estudiantes. Los autores concluyen que las diferencias individuales en los rendimientos se explican por la heterogeneidad entre planteles educativos, y porque el acceso a escuelas de buena calidad está determinado por el nivel de educación de los padres. Por último, muestran que a pesar del aumento sustancial del gasto público en educación en los últimos años del Siglo XX, el problema de la calidad de la educación pública persiste debido a la limitación de incentivos y a la ausencia de una estructura organizacional.

De la misma manera, Mina (2004) estima un modelo de datos de panel de 897 municipios durante el período 1996-1999, para revisar el impacto de las variables municipales sobre la calidad de la educación. La variable dependiente es el promedio por categoría del examen de Estado de cada municipio, y las variables independientes son las características educativas como número de maestros por alumno, porcentaje de escuelas privadas y el situado fiscal en educación; las características socioeconómicas, tales como, el coeficiente de Gini y el número de ataques de grupos insurgentes, y variables geográficas. Los resultados evidencian que hay una diferencia en el desempeño educativo entre los municipios y las capitales en favor de estas últimas. Además, Mina encuentra que en los municipios con mayor nivel de ruralidad se genera un “trade-off” entre calidad y cobertura, es decir, la calidad se ve afectada a medida que el nivel de cobertura aumenta. Por último, este autor, brinda evidencia de que la pobreza y la desigualdad del municipio tienen un impacto negativo en la calidad de la educación: a mayor pobreza y desigualdad en un municipio, menor será la calidad de la educación.

Asimismo, Villegas et al. (2013) proponen la siguiente pregunta: ¿qué tanto el sistema educativo logra atenuar las diferencias personales y sociales que afectan la igualdad de oportunidades? Para responder a esta pregunta utilizan la metodología de modelo jerárquico (o multinivel) que permite analizar la calidad de la educación, enfocándose en el promedio de los resultados de matemáticas y lenguaje del Saber 11, debido a que estas áreas son las más utilizadas para determinar la calidad de la educación. Este modelo tiene tres niveles: el entorno familiar y las condiciones socioeconómicas del estudiante, las características del colegio, y el contexto municipal.

Los resultados muestran que existe una relación positiva entre el nivel de desarrollo municipal y los resultados de las pruebas, lo que resulta ser significativo y demuestra la importancia que tiene la institucionalidad sobre las políticas de educación. Los municipios con mayor institucionalidad tendrían mejores sistemas educativos y mejores estándares de calidad. Por último, los autores concluyen que el sistema educativo colombiano ofrece niveles de calidad diferenciados, según las

clases sociales, comprometiendo la igualdad de oportunidades.

Por otra parte, Rozo Alzate (2017) realiza una investigación para identificar en qué medida y qué condiciones del barrio donde vive el estudiante se correlacionan con el rendimiento educativo. Para esto, utiliza datos de la prueba Saber 11, el censo de colegios C-600, el censo poblacional del 2005 e información de la policía metropolitana de Bogotá, y plantea el modelo de regresión lineal jerárquico de tres niveles: individuo, barrio, y colegio. El autor explica que las escuelas no necesariamente se encuentra en el mismo barrio donde habita el estudiante, por lo cual implementa el orden jerárquico mencionado. El autor afirma que tanto el barrio como la escuela están correlacionados con los resultados del estudiante, pero el efecto del colegio parece ser mucho más fuerte que el del barrio. Además, el autor encuentra que estudiantes que residen en barrios rodeados de universitarios van a tener más incentivos a estudiar y a buscar un mayor nivel educativo, lo cual implica mejores resultados. Por último, el autor afirma que el crimen muestra tener un efecto negativo sobre los rendimientos académicos, esto se puede dar por dinámicas de estrés o de sustitución, dado que, el crimen se puede presentar como una opción de vida.

Igualmente, Rodríguez y Correa (2019) evaluaron la relevancia del contexto municipal sobre el rendimiento académico individual. Para hacerlo, tomaron las variables: tasa de mortalidad infantil, violencia y conflictos de tierras del siglo XX. Los datos de cada una de estas fueron tomados del CEDE¹. Dada las variables plantearon un modelo de tres niveles que analizaba los efectos individuales, escolares y municipales. A partir de lo anterior, los autores afirman que el sistema educativo colombiano es un reproductor de desigualdades, en el cual los estudiantes más desaventajados socioeconómicamente sufren tres tipos de penalidades: la desventaja de dotación inicial, el contexto desfavorable, y la interacción entre ambas. Además, encontraron que el contexto municipal es relevante en el rendimiento académico, pero no es tan fuerte como el contexto del hogar y de la escuela. Su inferencia final es que tales consecuencias están relacionadas con una sociedad de instituciones extractivas, con políticos y una élite a la que les interesa poco el cambio y el progreso social.

En resumen, las investigaciones nacionales argumentan la importancia tanto de la escuela como del municipio para la determinación del logro educativo. Además, se evidencian que el contexto del barrio, los municipios más rurales y los bajos estratos socioeconómicos afectan negativamente el desempeño educativo del estudiante.

¹Centro de Estudio sobre Desarrollo Económico de la Universidad de los Andes

3. Aproximación empírica

En principio en esta sección, se va a definir qué se entiende por un buen desempeño académico para la zona rural colombiana. Posteriormente, se crean grupos comparables para que los resultados sean explicados por las diferencias del lugar de origen de los estudiantes. Finalmente, se describe la metodologá econométrica empleada.

3.1. ¿Qué es un buen desempeño académico?

El presente trabajo no va a medir el desempeño académico de los estudiantes como la literatura sugiere. Se va a utilizar la variable de niveles de desempeño de Saber 11, esto es debido a que marginalmente es mejor la interpretación acerca del rendimiento escolar de una variable cualitativa que cuantitativa. La ventaja de esta decisión es que proporciona una descripción de las habilidades y conocimientos que ha desarrollado el estudiante en la asignatura evaluada, mientras que la variable cuantitativa no puede ofrecer esta interpretación.

Las áreas de evaluación del Saber 11 son: la lectura crítica, matemáticas, ciencias naturales y, sociales y ciudadanas. Sus resultados son categorizados en cuatro niveles (1, 2, 3 y 4), siendo 1 el nivel más bajo y 4 el más alto ICFES (2018). Para evitar sesgos en el desempeño académico de los evaluados se va a tener en cuenta todas las áreas, esto se debe a que hay personas que tiene mejores habilidades en asignaturas cualitativas que cuantitativas, o viceversa Gardner (1995).

Un buen desempeño se define a partir del nivel 3 ICFES (2016). Aplicando a esta condición para todas las áreas evaluadas se obtiene que el 32 % de los evaluados que residen en ciudades o municipios, y el 10 % que residen en veredas, obtuvieron un buen desempeño en el Saber 11¹, esto quiere decir, que la gran mayoría de los estudiantes que viven en las zonas rurales de Colombia obtienen un mal resultado en el Saber 11, estas diferencias porcentuales pueden ser explicadas por el teorema CHAT de Vygotsky (1978).

¹Se puede explicar que cada grupo obtuvo bajos rendimientos en las áreas de ciencias naturales, y, sociales y ciudadanas.

Inicialmente esta teoría, supone un status-quo donde hay dos personas (A y B) y ambos tienen un conocimiento básico universal, es decir, que sus procesos psicológicos y de aprendizaje son homogéneos, es decir, se ignora la cultura y el contexto donde se encuentra la persona.

Sin embargo, el desarrollo humano y el contexto que habita la persona están mutuamente constituidos (no son dos fenómenos separados), esto genera diferencias en los procesos psicológicos superiores entre las dos personas mencionadas. Por ejemplo, si se supone que un estudiante A que vive en un contexto urbano y distribuye su tiempo en dos partes: ocio y estudiar; y el segundo estudiante (B) que vive en un contexto rural, y dada las condiciones socioeconómicas es necesario que se vincule al trabajo agrícola, o, a la comercialización de productos primarios, por lo cual, va a distribuir su tiempo en tres: ocio, trabajar y estudiar. Por lo tanto, es evidente que estos estudiantes tienen diferencias en sus procesos psicológicos superiores en función del contexto que habitan.

En otras palabras, y siguiendo a Gaviria y Barrientos (2001b) encontraron que los estudiantes rurales obtienen bajos niveles de desempeño en todas las áreas evaluadas. Por esta razón, en este estudio se va a modificar la condición de buen desempeño y va a tener en cuenta a los evaluados que tuvieron un bajo rendimiento en una de las cuatro áreas que evalúa el Saber 11. En la Tabla 3-1 se encuentran las condiciones para definir el desempeño en las pruebas Saber 11 de un estudiante.

Tabla 3-1.: Condiciones de un buen desempeño académico.

Grupo	Criterio para pertenecer al grupo
Bueno	El estudiante alcanza un nivel de desempeño de 3 o 4 en todas las asignaturas. El estudiante alcanza un nivel de desempeño de 1 o 2 en sólo una de sus asignaturas.
Malo	El estudiante alcanza un nivel de desempeño de 1 o 2 en más de una de sus asignaturas.

3.2. Índice del nivel socioeconómico

El objeto de estudio son los estudiantes que residen en veredas, y así como realizar una comparación entre ellas, y los estudiantes residentes en capitales y municipios de Colombia. Esta investigación debe poder analizar estudiantes comparables, es decir, los niveles socioeconómicos de los estudiantes de veredas deben de ser similares que los residentes en ciudades o municipios. Para que los resultados de esta investigación puedan ser explicados por las diferencias de su lugar de origen, y no, por sus características socioeconómicas.

La variable proxy para crear los grupos comparables es el Índice del Nivel Socioeconómico (INSE). Este mide las dotaciones de los bienes y servicios que posee el núcleo familiar, la posición social y los ingresos del hogar del estudiante. Además, esta inspirado en los cuatro componentes de la

Encuesta de Calidad de Vida: vivienda, capital humano, servicios, y hacinamiento. El Icfes se encarga de construir este indicador para todos los estudiantes que presentan el examen de Estado. El INSE ha tenido como objetivo mejorar los análisis del desempeño educativo, y evaluar el impacto de las políticas públicas educativas para los más desaventajados ICFES (2010).

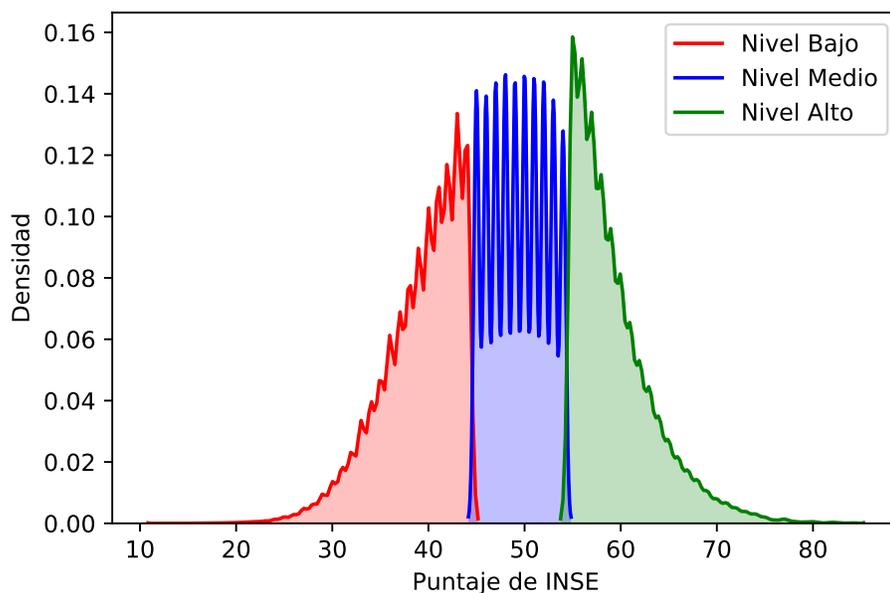
Para la creación de los grupos de estudiantes comparables se va a utilizar el método de K-Means. Ésta es una técnica de aprendizaje no supervisado, que es utilizado en datos que no se encuentran etiquetados, es decir, sin categorías o grupos definidos. Este método particiona un conjunto de datos heterogéneos en grupos por agrupamiento de proximidad, es decir, crea grupos de estudiantes que sean similares (o comparables) socioeconómicamente entre sí, pero que en conjunto se diferencian de los otros grupos Hastie et al. (2001). En la figura **B-1** del Anexo A: (a) se encuentra el test de WSS y se observa que particionar en tres grupos es óptimo, por otro lado, (b) se realizó el test de silhouette e indica que el número de cluster escogido es razonablemente robusto, y los grupos encontrados se van a denominar: bajo, medio, y alto socioeconómico.

En la Tabla **3-2** y en la Figura **3-1** se encuentra los intervalos de puntaje para los grupos de estudiantes comparables, y la densidad del INSE.

Tabla 3-2.: Grupos de INSE

Nivel Socioeconómico	Intervalo de Puntaje
Bajo	0 - 44
Medio	45 - 54
Alto	55 - 100

Figura 3-1.: Gráfico de densidad del INSE



3.3. Regresión logística de modelos lineales jerárquicos

El objetivo de este modelo es igual al que plantea un modelo de regresión logística: estimar la probabilidad de que ocurra un evento. Pero, este modelo se diferencia por tres características importantes: *a)* ser una función de densidad probabilística que relaciona las variables independientes con la variable latente (VL), *b)* ser un modelo lineal mixto (efectos fijos y aleatorio) en una VL , y *c)* estructura las varianzas que describen las (co)varianzas esperadas entre los efectos fijos y aleatorios.

La importancia de los efectos aleatorios en este modelo, permite medir el nivel de correlación de obtener un buen desempeño académico con el lugar de origen de los estudiantes. Hay muchas razones por las cuales esto podría suceder. Por ejemplo, las capitales tienen mejor infraestructura educativa, o más profesores capacitados que los municipios o veredas, de modo que dentro una UT, los estudiantes pueden ser más homogéneos que entre UT. Esta decisión se puede explicar por dos razones: la primera, el entorno donde habita el estudiante tiene un efecto en su desarrollo humano, y la segunda, la educación rural tiene dificultad de acceso para contratar profesores calificados, una gran carencia en recursos, ausencia de salones adecuados, falta de materiales didácticos, laboratorios e implementos deportivos Arias Gaviria (2017). Por lo cual, se plantea la siguiente ecuación a estimar para una VL , y los supuestos que permiten su estimación:

$$\log\left(\frac{P(y=1)}{1-P(y=1)}\right) = \beta X + Z\theta + e \quad (3-1)$$

donde X es un vector de efectos fijos, y Z es un vector de efectos aleatorios. Esta estimación tiene parámetros (β y θ), y un vector de residuos (e), que siguen una distribución normal multivariable Hadfield (2010). Por otra parte, la literatura plantea que la evaluación educativa requiere un modelo multinivel para tener en cuenta la estructura jerárquica de los datos; además, para estudiar el efecto municipal en el desempeño educativo es necesario crear un modelo de tres niveles: estudiante, escuela y municipio.

Sin embargo, no es posible implementar el modelo con los niveles establecidos por la literatura, esto es debido, que el 90,24 % de las veredas tienen una sola IE, por ende, esto va a generar problemas en la variación en el nivel de las escuelas para las veredas. Por lo cual, se va a formular tres modelos de regresión logística multinivel (n)² de dos niveles: el primer nivel corresponde a (i) estudiantes que presentaron el examen de Estado, y el segundo nivel (j) es la UT donde se encuentra la escuela. Se reestructura la ecuación (3-1) para su estimación:

$$\log\left(\frac{P(y_{ij}^n = 1)}{1-P(y_{ij}^n = 1)}\right) = \theta_j^n + \beta_j^n X_{ij}^n + \varepsilon_{0j}^n \quad (3-2)$$

Donde:

$$y_{ij}^n = \begin{cases} 1 & \text{obtuvo un buen desempeño en las pruebas Saber 11.} \\ 0 & \text{o.w.} \end{cases}$$

el cual relaciona al vector X (**entorno familiar**: edad, educación madre, género; **entorno escolar**: jornada escolar, calidad de la escuela, profesores capacitados; y **entorno municipal**: UT, región a la que pertenece la UT (ver Anexo A, Tabla A-2), naturaleza de SE, y una perturbación. Por otra parte, el intercepto y la pendiente del modelo son específicos para cada grupo; esto es uno de las características del modelo que permite que los coeficientes varíen entre unidades de agregación, lo cual se refleja en las siguientes ecuaciones Rodríguez y Correa (2019):

$$\theta_j^n = \theta_0^n + \Lambda_j^n \quad (3-3)$$

$$\beta_j^n = \beta_0^n + \Lambda_{1j}^n \quad (3-4)$$

²Los tres modelos econométricos se refieren a cada grupo de estudiantes comparables establecidos en la sección anterior.

Partiendo de la ecuación (3-3) y (3-4) y reemplazándolo en el (3-2), se obtendría el modelo combinado y completo de los dos niveles, donde ya tiene en cuenta la jerarquía de los datos, y se expresa de la siguiente forma:

$$\log \left(\frac{P(y_{ij}^n = 1)}{1 - P(y_{ij}^n = 1)} \right) = \theta^n_0 + \beta^n_0 X^n_{ij} + (\Lambda^n_j + \Lambda^n_{1j} X^n_{ij} + \varepsilon^n_{0j}) \quad (3-5)$$

$$\Lambda^n_j \sim N(0, \sigma_{\Lambda^n}^2) \quad \Lambda^n_{1j} \sim N(0, \sigma_{\Lambda^n_1}^2) \quad \varepsilon^n_{0j} \sim N(0, \sigma_{\varepsilon^n_0}^2) \quad (3-6)$$

Este modelo incluye tanto efectos fijos (los que están fuera del paréntesis), como aleatorios (los que están dentro del paréntesis); permitiendo que el intercepto y la pendiente sean específicos para cada grupo. Donde Λ^n_j representa la desviación de la unidad j respecto al promedio general, y Λ^n_{1j} representa el efecto diferencial de la variable X^n_{ij} en el grupo j . Además, uno de los supuestos del modelo es que considera la independencia entre los términos de error del primer y segundo nivel. Para una mejor descripción del algoritmo de estimación ver Hadfield (2010). Por último, la significancia individual de los coeficientes estimados se obtienen por el método de bootstrapping (se emplean 1000 repeticiones).

3.4. Prueba de anidación

Antes de continuar con las estimaciones de los modelos jerárquicos, se debe poner a prueba que pertenecer a una organización municipal tiene un efecto sobre el desempeño académico en el Saber 11. Inicialmente, se deben comparar dos regresiones vacías (es decir, sin ninguna variable explicativa): logit y logit jerárquica, por medio de un test de tasa de verosimilitud (Likelihood-ratio test), donde la hipótesis nula se refiere que la varianza entre organizaciones municipales es cero: $\sigma_{\varepsilon^n_0}^2 = 0$.

Tabla 3-3.: Pruebas de anidación.

Nivel Bajo	LR = 4879,612 Valor p = 0,00
Nivel Medio	LR = 6459,822 Valor p = 0,00
Nivel Alto	LR = 2957,41 Valor p = 0,00

En la tabla 3-3 se encuentran los resultados de test estadísticos para cada nivel socioeconómico, y se evidencia que la hipótesis nula es rechazada para los tres modelos econométricos. Esto quiere decir que el lugar de origen de los estudiantes sí explica parte de la varianza de su desempeño académico.

4. Base de datos

4.1. Saber 11

El Saber 11 es un examen de Estado con carácter obligatorio para los estudiantes que están finalizando el grado undécimo, y es un requisito si desean continuar sus estudios en la educación superior. Los resultados de estas evaluaciones permiten analizar los conocimientos, habilidades y valores desarrollados por todos los estudiantes colombianos Ministerio de Educación Nacional de Colombia (2018). Para evaluar la incidencia que en su formación tiene las Instituciones Educativas (IE), las Secretarías de Educación (SE), el Ministerio de Educación Nacional (MEN) y la sociedad en general. Además, algunas universidades ofrecen becas tanto parciales como completas a los estudiantes que obtienen buenos desempeños en la prueba Saber 11. A partir del análisis de estos resultados se pueden definir planes de mejoramiento en el Sistema Educativo teniendo en cuenta la procedencia, condiciones sociales, económicas y culturales de los estudiantes.

Además de ser una herramienta retroalimentadora al MEN, esta prueba tiene por objetivos, según lo dispuesto por el Decreto 869 de 2010:

- 1. Proporcionar a las IE información pertinente sobre las competencias de los aspirantes a ingresar a programas de educación superior, así como sobre las de quienes son admitidos, que sirva como base para el diseño de programas de nivelación académica y prevención de la deserción en este nivel.*
- 2. Proporcionar información para el establecimiento de indicadores de valor agregado, tanto de la educación media como de la educación superior.*
- 3. Proporcionar información a los establecimientos educativos que ofrecen educación media para el ejercicio de la autoevaluación y para que realicen la consolidación o reorientación de sus prácticas pedagógicas.*

La base de datos de Saber 11 de 2018 utilizada en este estudio se caracteriza por estudiantes que:

- Tienen entre 14 y 18 años.¹
- Asisten a colegios públicos², en jornada diurna (mañana, tarde, única, y completa).

¹Colbert (1999) afirma que la baja calidad e ineficiencia del sistema escolar en la educación rural y urbano-marginal de América Latina y El Caribe es debido a la alta heterogeneidad de edades y extraedad. Por lo cual, se filtró este tipo de población para evitar sesgos en las estimaciones obtenidas.

²Se hace este filtro, debido a que, las veredas no tienen escuelas privadas.

- Toman la prueba por primera vez.
- Reportaron información para todas las variables de interés.
- Pertenecen a IE que presentaron como mínimo 10 estudiantes.

Dado lo anterior, la muestra objeto de esta investigación quedó conformada por 343.107 de 549.934, es decir el 62 % de las observaciones originales.

4.1.1. Índice sintético de calidad educativa

EL ISCE es un índice, calculado por el ICFES, que apoya en el seguimiento del progreso de cada IE. Ésta tiene cuatro aspectos de medición de la calidad educativa del país: progreso, desempeño, eficiencia y ambiente escolar (ver ICFES (2016) para un detalle de la metodología y las variables utilizadas en la construcción del índice). Este indicador permite saber cómo se encuentra la calidad educativa y cuánto falta para mejorar. El ISCE es una variable continua y tiene una escala del 1 al 10, siendo 10 el valor más alto que se puede obtener. Este índice va a estar centralizado (desviaciones) por el nivel socioeconómico, para evitar problemas de convergencia en los modelos econométricos.

4.2. Información de las escuelas de las veredas colombianas

Actualmente, el MEN no maneja una base de datos oficial que especifique las instituciones nacionales que se encuentran en las veredas con su respectivo código del DANE³. Por este motivo se empleó la información del **Sistema Interactivo de Consulta de Colegios (SICOLE)** del DANE. Esta plataforma del Sistema de Información Geográfica (SIG) muestra a nivel espacial todas las IE oficiales y no oficiales de Colombia. Esta base de datos cuenta con 57.135 IE de educación preescolar, básica y media. No obstante, el DANE aclara que al SICOLE le hace falta información, debido a ausencia de ciertas encuestas.

El 62 % de las IE que se encuentran en SICOLE están en las veredas de Colombia. Dada la motivación de este trabajo se tomará el registro de las IE de educación media y oficiales, lo que equivale al 9,8 % de la información. Por otro lado, se excluyeron los modelos educativos: etnoeducación y tradicional. La etnoeducación es definida por el artículo 55 de La Ley General de Educación como:

una educación para grupos étnicos la que se ofrece a grupos o comunidades que integran la nacionalidad y que poseen una cultura, una lengua, unas tradiciones y unos fueros propios autóctonos. Esta educación debe estar ligada al ambiente, al proceso

³Pero, sí existe esta información para las escuelas que se encuentran en municipios y ciudades.

productivo, al proceso social y cultural, con el debido respeto de sus creencias y tradiciones. Congreso de la República de Colombia (1994)

Por otra parte, el artículo 27 de la misma Ley explica que: *la educación media tiene como fin la comprensión de las ideas y los valores universales y la preparación para el ingreso del educando a la educación superior y al trabajo*. Dada las definiciones de los dos modelos pedagógicos que existen en las IE de las veredas colombianas, no es posible establecer una comparación entre estas, por lo tanto, se depurará el 7,5 % de la información que respecta a este tipo de escuelas. En conclusión, se obtuvieron 2.524 IE de 57.135, es decir el 4,4 % de las observaciones originales.

4.3. Educación formal

Educación Formal (EF) es una base de datos del DANE que es recolectada por medio del Formulario Electrónico Único Censal - C600. El objetivo principal de EF es:

generar y difundir información estadística básica sobre la EF en Colombia, en los niveles de preescolar, básica primaria, básica secundaria y educación media, base para la formulación y seguimiento de la política pública sectorial, y el ejercicio de planeación y administración del servicio público de educación según las categorías político-administrativas del país DANE (2019).

EF maneja información de matrículas, capital humano como número de docentes por IE y su nivel educativo, infraestructura de las TIC por colegio, y la situación académica de las y los estudiantes al terminar el año lectivo. En esta base de datos se va a utilizar el nivel educativo del docente de educación media para determinar si la IE se encuentra con profesores capacitados para la enseñanza del año 2017. Profesores capacitados se hace referencia a los docentes graduados de un posgrado pedagógico. Se va a construir una variable dummy por cada colegio donde el criterio de decisión es partir la información por cuartiles teniendo en cuenta su respectiva UT y región. Si el porcentaje de profesores capacitados por IE se encuentra por encima del último cuartil (ver Tabla 4-1), se considera que la escuela tiene docentes capacitados para la enseñanza:

$$Escuela = \begin{cases} 1 & \text{Escuela con profesores capacitados.} \\ 0 & \text{o.w.} \end{cases}$$

Tabla 4-1. Último cuartil de porcentaje de profesores con posgrado en programa pedagógico según región y unidad territorial.

Und. Territorial \ Regiones	Caribe	Centro Oriente	Centro Sur	Eje Cafetero	Llano	Pacífico
Capital	61,9 %	79 %	85,5 %	57,3 %	74,2 %	67,5 %
Municipio	62 %	73,6 %	76,7 %	55,4 %	55,7 %	68,1 %
Vereda	44,5 %	62,5 %	57,2 %	36,4 %	43,3 %	66,7 %

4.4. Estadística descriptivas

Las Tablas 4-3 y A-2 del Anexo A, muestran las estadísticas descriptivas de todas las variables usadas en la estimación a nivel nacional, por UT y niveles socioeconómicos.

Inicialmente, se puede observar que el porcentaje de los estudiantes de bajos niveles socioeconómicos, que obtienen un nivel 1 o 2 en todas las áreas que evalúa el Saber 11, está por encima del porcentaje poblacional nacional; mientras que, el porcentaje en los niveles medios y altos está por debajo. De igual manera, para los niveles 3 y 4, el análisis sería de manera contraria. Este comportamiento se mantiene si se comparan por UT.

De igual forma, hay diferencias en el porcentaje poblacional sobre el desempeño académico al interior de las UT, por ejemplo, se observa que el 22.58 % de los estudiantes rurales obtuvieron un buen rendimiento académico en el Saber 11, y este porcentaje aumenta por UT: 10.4 puntos porcentuales para municipios; y 16.57 puntos porcentuales para capitales. Esto quiere decir que los estudiantes rurales y más pobres de Colombia tienen un bajo desempeño en las pruebas de Estado, siguiendo a los autores de la literatura nacional mencionada, se puede confirmar que si el estudiante tiene mejores condiciones socioeconómicas, el estudiante va a obtener un mejor desempeño académico en el Saber 11.

Tabla 4-2.: Estadísticas descriptivas a nivel nacional y por unidades territoriales

Variables	Nacional	Unidades Territoriales			Variables	Nacional	Unidades Territoriales		
		Capital	Municipio	Vereda			Capital	Municipio	Vereda
Nº de estudiantes	242.850	75.842	114.722	52.286	Nivel Educativo Madre (%)				
Niveles de Desempeño (%)					Ninguno	7.1 %	6.91 %	6.45 %	7.75 %
Lectura Crítica: 1	4.5 %	3.04 %	4.15 %	6.85 %	Educación Primaria	27.2 %	17.28 %	26.38 %	42.21 %
Lectura Crítica: 2	43 %	35.50 %	42.41 %	52.93 %	Educación Secundaria	43 %	45.87 %	43.94 %	37.50 %
Lectura Crítica: 3	45 %	50.32 %	45.73 %	36.57 %	Educación Técnica	12.8 %	17.14 %	13.02 %	7.01 %
Lectura Crítica: 4	7.5 %	11.14 %	7.71 %	3.65 %	Educación Superior	9.9 %	12.80 %	10.21 %	5.53 %
Matemáticas: 1	12.2 %	9.30 %	11.67 %	16.93 %	Colegio (%)				
Matemáticas: 2	43 %	38.96 %	42.42 %	49.19 %	Jornada Mañana	64.3 %	62.86 %	64.66 %	64.95 %
Matemáticas: 3	42 %	47.72 %	43.03 %	32.62 %	Jornada Tarde	13.8 %	21.40 %	12.15 %	7.24 %
Matemáticas: 4	2.8 %	4.02 %	2.88 %	1.26 %	Jornada Completa	10 %	2.19 %	11.22 %	18.43 %
Ciencias Naturales: 1	23.6 %	18.98 %	23.02 %	29.98 %	Jornada Única	11.9 %	13.55 %	11.97 %	9.38 %
Ciencias Naturales: 2	52.1 %	50.36 %	52.20 %	54.16 %	ISCE (Promedio)	5.60	5.95	5.57	5.13
Ciencias Naturales: 3	22.9 %	28.41 %	23.43 %	15.29 %	Escuela: 1	30.3 %	31.75 %	29.96 %	29.63 %
Ciencias Naturales: 4	1.4 %	2.25 %	1.35 %	0.57 %	Municipio (%)				
Sociales y Ciudadanas: 1	32.9 %	26.79 %	32.07 %	41.95 %	Región Caribe	29 %	33.30 %	27.65 %	25.59 %
Sociales y Ciudadanas: 2	43.9 %	43.68 %	44.30 %	43.26 %	Región Centro Oriente	20.4 %	9.16 %	27.29 %	21.58 %
Sociales y Ciudadanas: 3	21.3 %	26.65 %	21.74 %	13.98 %	Región Centro Sur	9.2 %	9.33 %	6.05 %	16.03 %
Sociales y Ciudadanas: 4	1.9 %	2.88 %	1.89 %	0.81 %	Región Eje Cafetero	21.6 %	25.62 %	21.76 %	15.45 %
Desempeño Académico (%)					Región Llano	5.4 %	5.1 %	4.52 %	4.52 %
Bueno	32 %	39.15 %	33.02 %	22.58 %	Región Pacífico	14.4 %	17.49 %	12.73 %	17.13 %
Malo	68 %	60.85 %	66.98 %	77.42 %	Nivel Bajo	32.91 %	19.69 %	35.72 %	59.89 %
Personales					Nivel Medio	43.39 %	43.82 %	38.85 %	26.90 %
Edad (promedio años)	17.9	17.47	17.51	17.97	Nivel Alto	23.70 %	36.49 %	25.43 %	13.21 %
Mujer (%)	43.14 %	43.07 %	42.87 %	43.85 %	No Certificado	50.1 %	0 %	68.94 %	80.74 %

Fuente: Cálculos propios a partir de Saber 11 (2018), Educación Formal (2018), y SICOLE (2019)

Por otra parte, el porcentaje de mujeres es relativamente similar y superior al 40 % para todos los casos. La edad promedio del estudiante es de 17.9 años a nivel nacional, para todas las UT es ligeramente menor. En cuanto a las características familiares, en el ámbito nacional, el 43 % 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 mayor proporción de los estudiantes de niveles medios socioeconómicamente. Por el contrario, en los bajos y altos niveles, el mayor porcentaje de respuesta se concentró en primaria (incompleta o completa) y superior, respectivamente.

Por otro lado, a nivel nacional, el puntaje promedio del ISCE es 5,6, para todas las UT es ligeramente menor. Asimismo, se mantiene el comportamiento con las escuelas con profesores capacitados. De igual manera, en el ámbito nacional, el 43.39 % en los estudiantes pertenecen a niveles socioeconómicos medios, en este mismo nivel social se concentra la mayor proporción de los estudiantes de capital y municipio. Por el contrario, para los estudiantes de las veredas el mayor porcentaje de respuesta se concentró en niveles bajos. En otras palabras, la gran mayoría de los estudiantes rurales son pobres, y por el contrario, la mayor proporción de los que residen en capitales y municipios son de niveles medios socioeconómicos.

Finalmente, a nivel nacional, el 29 % de los estudiantes reportaron que residen en la región caribe, en este mismo lugar de origen se concentra la mayor proporción de los estudiantes de niveles bajos y medios socioeconómicamente. Por el contrario, los estudiantes de altos niveles se concentran en la región del eje cafetero. Por último, se encuentra que el 80.74 % de los estudiantes rurales estudiaron en una escuela administrada por una secretaría de educación departamental, y se reduce 11.8 puntos porcentuales para los municipios. Esto quiere decir que la gran mayoría de la población rural y que residen en municipios estudió en una escuela administrada por una secretaría de educación departamental.

5. Resultados

En esta sección se presentan los resultados de la aplicación de la aproximación empírica de la pregunta de investigación. Antes de continuar con los resultados de las estimaciones, se interpreta los modelos vacíos de las tres regresiones.

Los resultados iniciales indican que la UT y la escuela explica el 20 % de la variación en el desempeño académico de los estudiantes que se encuentran en los niveles bajos socioeconómicos (ver Tabla 5-1). Mientras que, ese porcentaje disminuye en 3 puntos porcentuales si el estudiante es de niveles medios, y en 7 puntos porcentuales para el estudiante de niveles altos.

Tabla 5-1.: Coeficiente de correlación interclase del modelo vacío.

Nivel Socioeconómico	CCI
Bajo	20 %
Medio	17 %
Alto	13 %

Fuente: Cálculos propios a partir de Saber 11 (2018).

5.1. Entorno del estudiante

En la tabla 5-2 se encuentran los resultados de las estimaciones de las regresiones logísticas de modelos lineales jerárquicos. El intercepto, el cual es significativo para los niveles bajos (22,4 %) y medios (30,8 %), se puede interpretar como la probabilidad en obtener un buen desempeño en las pruebas Saber 11 en relación a cada nivel socioeconómico que obtendría un estudiante hombre con 18 años en promedio que vive en una capital de la región caribe y su madre no tiene ningún nivel educativo, y estudia en una escuela que tiene el puntaje en el ISCE de 5.17 en promedio para los niveles bajos socioeconómicos (y 5.58 en promedio para niveles medios), el colegio es de jornada completa y no cuenta con profesores capacitados. Como era de esperarse la probabilidad aumenta a medida que mejora el nivel socioeconómico.

Por otro lado, un estudiante que viva en una vereda o municipio tiene menos probabilidad en obtener un buen desempeño comparado con vivir en una capital. Además, estos resultados sigue a

Tabla 5-2.: Estimación regresión logística de modelos lineales jerárquicos: ¿Obtuvo un buen desempeño en las pruebas Saber 11?

Modelo	(1: Bajo)	(2: Medio)	(3: Alto)	Modelo	(1: Bajo)	(2: Medio)	(3: Alto)
Efectos Fijos				Efectos Fijos: Controles			
Intercepto	-1.24*** (0.36)	-0.81** (0.32)	-0.12 (0.30)	Edad (Centrada)	-0.38*** (0.01)	-0.35*** (0.01)	-0.34*** (0.01)
Vereda	-0.38*** (0.12)	-0.34*** (0.11)	-0.27** (0.11)	Genero: Mujer	-0.54*** (0.01)	-0.51*** (0.01)	-0.44*** (0.02)
Municipio	-0.05 (0.13)	-0.02 (0.11)	-0.02 (0.11)	Educación Primaria: Madre	0.11*** (0.03)	0.09*** (0.03)	-0.06 (0.06)
Centro Oriente	0.76*** (0.05)	0.78*** (0.05)	0.62*** (0.06)	Educación Secundaria: Madre	0.34*** (0.03)	0.24*** (0.03)	0.12*** (0.04)
Centro Sur	0.55*** (0.06)	0.57*** (0.06)	0.40*** (0.08)	Educación Técnica: Madre	0.80*** (0.05)	0.68*** (0.03)	0.51*** (0.04)
Eje Cafetero	0.35*** (0.05)	0.47*** (0.05)	0.41*** (0.06)	Educación Superior: Madre	0.63*** (0.09)	0.63*** (0.04)	0.61*** (0.04)
Llano	0.68*** (0.08)	0.60*** (0.08)	0.38*** (0.10)	Jornada: Mañana	-0.13*** (0.04)	-0.14*** (0.03)	-0.28*** (0.04)
Pacífica	0.74*** (0.05)	0.67*** (0.05)	0.48*** (0.06)	Jornada: Tarde	-0.22*** (0.05)	-0.21*** (0.04)	-0.34*** (0.05)
Escuela	0.24*** (0.02)	0.12*** (0.02)	0.12*** (0.02)	Jornada: Única	0.00 (0.05)	-0.10** (0.04)	-0.19*** (0.05)
ISCE (Centrada)	0.25*** (0.00)	0.27*** (0.00)	0.29*** (0.00)	Efecto Aleatorio			
No Certificado	-0.27*** (0.05)	-0.27*** (0.05)	-0.20*** (0.06)	Varianza entre municipios	0.480	0.269	0.204
				CCI	12.73 %	7.57 %	5.83 %
				N	79,944	105,382	57,524
				UT	2,683	2,601	1,869

* $P < 0,1$ ** $P < 0,05$ *** $P < 0,01$ (Entre paréntesis es el error estándar)

la literatura mencionada donde existe una fuerte correlación negativa del nivel de desempeño con los niveles socioeconómicos del evaluado. Es decir, un estudiante de bajos niveles que viva en una vereda tiene en promedio 31.61 % menos de probabilidad que un residente en una capital, el efecto es significativo con el 99 % de confianza. Este resultado disminuye por cada nivel social, y sigue siendo significativo.

Se evidencia el mismo resultado para los estudiantes que viven en municipio pero en menor proporción. No obstante, esto no tiene un efecto significativo sobre la probabilidad. También, al analizar por las regiones de Colombia se evidencia que residir en la zona caribe tiene menos probabilidad en obtener un buen desempeño comparado con las demás regiones.

Por otra parte, sobre las variables de la escuela se encuentra que asistir a un colegio con profesores capacitados tiene una correlación positiva y significativamente con la probabilidad de tener un buen desempeño, para los estudiantes de bajos niveles el efecto es mucho mayor. Asimismo, estudiar en una escuela con mayor puntaje en el ISCE tiene un efecto positivo y significativo en la probabilidad que los de menor puntaje. En cambio, si la IE es administrada por la SE departamental tiene un efecto negativo y significativo en la probabilidad, esto se podría analizar como un abandono del Estado o políticas educativas no adecuadas como trade-off de cobertura y calidad educativa Duraisamy et al. (1998).

Los estimadores de los controles utilizados son relevantes y tienen los efectos esperados de acuerdo a la revisión de la literatura. Los coeficientes de edad y de la mujer son aproximadamente iguales, negativos y significativos, para todos los niveles socioeconómicos. Entre más edad del estudiante menor la probabilidad de obtener un buen desempeño en el Saber 11 que los más jóvenes. Las mujeres tienen menos posibilidad en lograr un adecuado rendimiento académico que los hombres. Sin embargo, se puede observar que ser una mujer de bajos recursos tiene un mayor efecto que los otros niveles.

Además, el nivel educativo alcanzado por la madre es positivo y estadísticamente significativo. Por ejemplo, un estudiante que su madre tenga una carrera profesional (educación superior) tiene en promedio 87.8 % más de probabilidad en obtener un buen desempeño que un estudiante cuya madre no alcanzó algún nivel educativo. En cambio, los evaluados que asisten en una escuela de jornada de la mañana, tarde y única tienen menos posibilidad en lograr un buen rendimiento que los estudiantes que estudian en jornada completa. La jornada de la tarde tiene un mayor efecto negativo que la jornada de la mañana y única. Como revelan las correlaciones intraclase, el nivel de la escuela y el entorno del estudiante (medido como el lugar de origen) explica entre el 12.73 % y 5.83 % de las variaciones del buen desempeño en las pruebas Saber 11.

5.2. Probabilidad de obtener un buen desempeño en el Saber 11

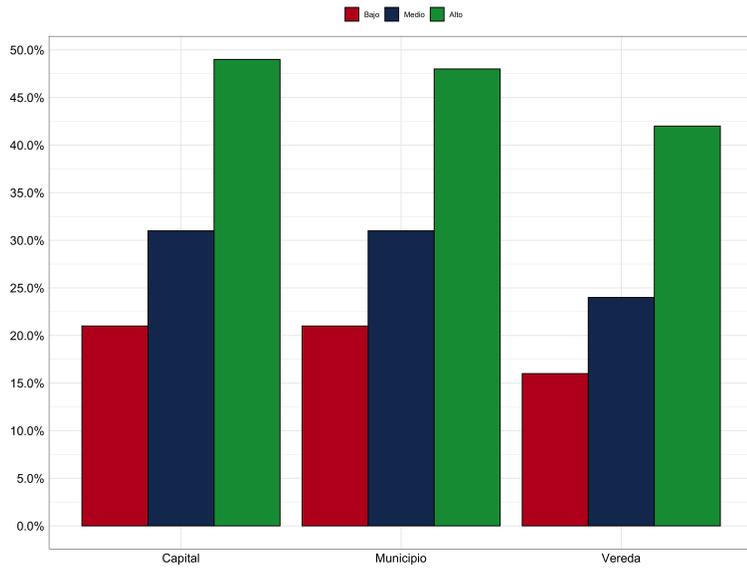
A partir de los resultados arrojados por los modelos, es posible extraer la Figura 5-1 que permite comprender que los estudiantes con nivel socioeconómico más alto tienen una probabilidad mayor de obtener mejores desempeños en las pruebas Saber 11; esto es así para todas las unidades territoriales (a). Sin embargo, a nivel de la capital esta es un poco más alta que la probabilidad que obtienen los estudiantes de los municipios y, la de estos últimos, mayor que la de las veredas. No obstante, cabe aclarar que los estudiantes de veredas cuyo nivel socioeconómico alto, tienen mayor probabilidad de obtener mejores resultados en las pruebas Saber, que los estudiantes de niveles socioeconómicos bajo y medio que residen en las capitales y los municipios.

El segundo gráfico (b) da cuenta de la probabilidad que los estudiantes obtienen para un buen desempeño en las pruebas Saber 11 diferenciándolos por la región donde esta ubicada la capital, el municipio o la vereda en la que habitan. Se constató que los estudiantes del centro-oriente del país, tanto los que viven en capitales, como en los municipios y las veredas, tienen una mayor probabilidad de obtener mejores resultados en las pruebas Saber que los estudiantes que residen en las demás regiones de Colombia. Incluso, los estudiantes de niveles altos socioeconómicamente a nivel de las veredas, unidad territorial en la que los resultados de la región Centro Oriente son los más deficientes, son superados por muy pocos puntos porcentuales por los estudiantes de altos niveles sociales de las capitales de tan solo tres de las restantes regiones del país: Eje Cafetero, Pacífico y Centro-Sur. Por último, la gráfica permite evidenciar que la región Caribe es en la que los estudiantes tienen una menor probabilidad de obtener buenos resultados.

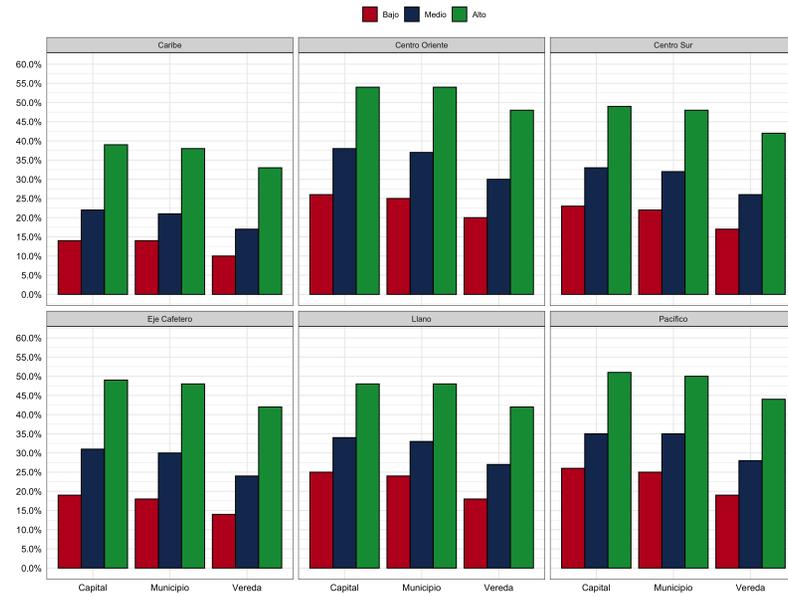
Por otro lado, (c) se pone en evidencia que pese a que hay colegios que obtienen la misma puntuación según el ISCE esto no garantiza que la probabilidad del desempeño de sus estudiantes en las pruebas Saber 11 sea la misma, fenómeno que debería esperarse. Por el contrario, colegios con la misma puntuación pero cuyo nivel socioeconómico de los estudiantes es diferente, obtienen resultados asimismo diferentes. La probabilidad de que los estudiantes obtengan mejores resultados es mayor para quienes tienen un alto nivel socioeconómico, independientemente del puntaje ISCE que se le haya otorgado a la IE del que estos hacen parte. Un ejemplo claro que se puede constatar en la gráfica es el caso de un colegio cuya puntuación es alta, pero sus estudiantes tienen un nivel socioeconómico bajo, estos terminan obteniendo una probabilidad menor de obtener un buen desempeño, hasta 30 puntos porcentuales por debajo del que obtendrían estudiantes de un mismo colegio pero con un nivel socioeconómico más alto.

Al mismo tiempo, (d) permite constatar que tanto en las capitales, como en los municipios y las veredas, si los colegios que estas unidades territoriales albergan poseen profesores capacitados, la probabilidad de que los estudiantes obtengan un buen desempeño en las pruebas Saber 11, independientemente del nivel socioeconómico, va a ser mayor comparada con escuelas que no tienen profesores capacitados en los mismos niveles territoriales. Por último, las brechas entre los niveles socioeconómicos son muy significativas, es decir, un cambio de nivel socioeconómico afecta en gran medida el desempeño en las pruebas de Estado. En promedio, pasar de bajo a medio implica una probabilidad mayor del 10 % de obtener mejores resultados y de medio a alto del 25 %, siendo esta la brecha más amplia.

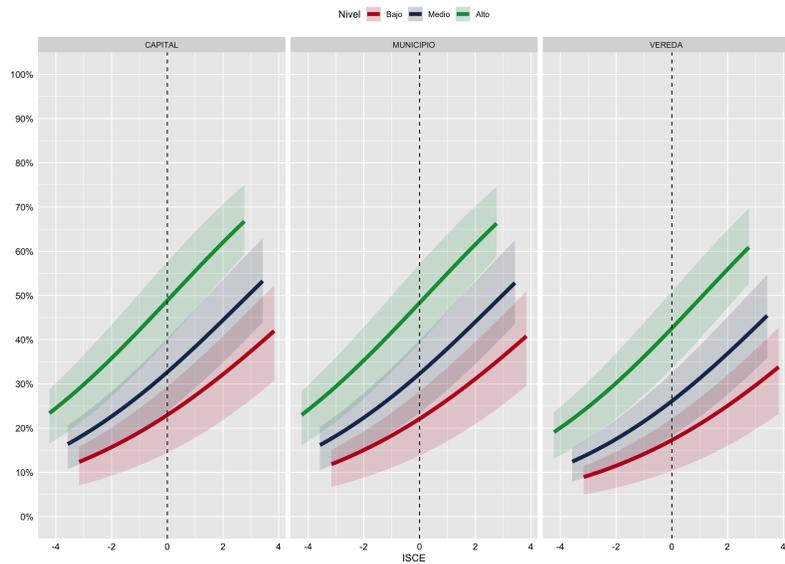
Figura 5-1.: Probabilidad de obtener un buen desempeño en las pruebas Saber 11 (2018)



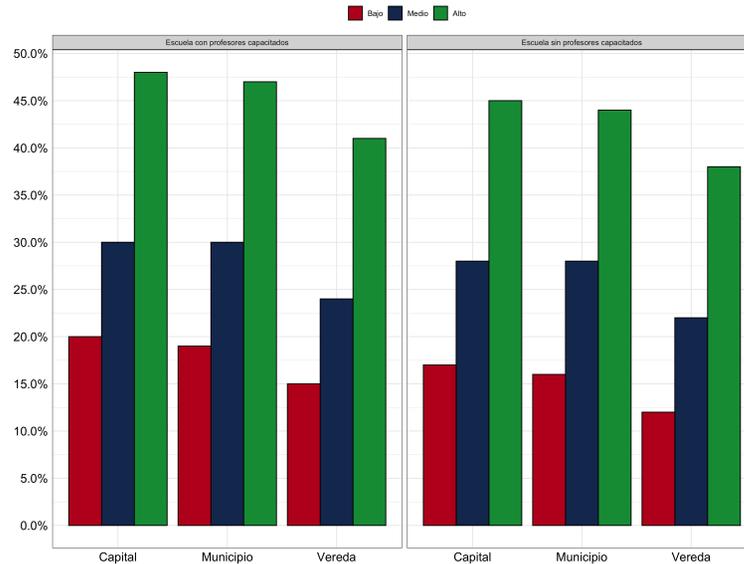
(a) Según unidad territorial



(b) Según la región



(c) Según el puntaje ISCE



(d) Según con profesores capacitados

6. Conclusiones

La gran mayoría de los estudiantes de veredas, en promedio, obtienen un bajo rendimiento en las pruebas de Estado. Esto se debe a que la zona rural debe superar una barrera para garantizar la educación de calidad. No obstante, analizando la misma UT se encuentra que el rendimiento académico al interior de esta varía, esto se debe a la heterogeneidad de las políticas educativas implementadas en el país. Por otro lado, es alarmante que el hecho de que una escuela sea administrada por la SE departamental implique un efecto negativo en la probabilidad de obtener un buen desempeño en las pruebas Saber 11. Futuros estudios podrían estudiar cómo se relaciona esto con la descentralización del sistema educativo en Colombia.

Las conclusiones a las que llega Coleman et al. (1966) se ven reafirmadas en esta investigación. Siguiendo los argumentos de dicho autor, mejorar las IE o tener docentes más capacitados no va a generar un efecto importante en el logro educativo de los estudiantes. Por el contrario, el nivel socioeconómico es un determinante más diciente del desempeño en las pruebas Saber 11.

Por otro lado, entre las limitaciones del estudio se encuentra que hay varias variables a nivel del colegio y UT que no se tuvieron en cuenta por restricciones en los datos, por ejemplo, los años de experiencia de los maestros, el tamaño de la clase, los salarios de los profesores, la distancia en KM entre la vereda (o municipios no certificados) y la SE. Una segunda limitación consistió en que no se diferenció los modelos educativos y, además, no se tuvo en cuenta a la población indígena. Finalmente, en este trabajo queda demostrada la importancia de incluir, en futuras investigaciones sobre el rendimiento académico, una variable alusiva a la unidad territorial en la que los estudiantes residen.

A. Regiones de Colombia y estadísticas descriptivas

Tabla A-1.: Regiones de Colombia (DNP)

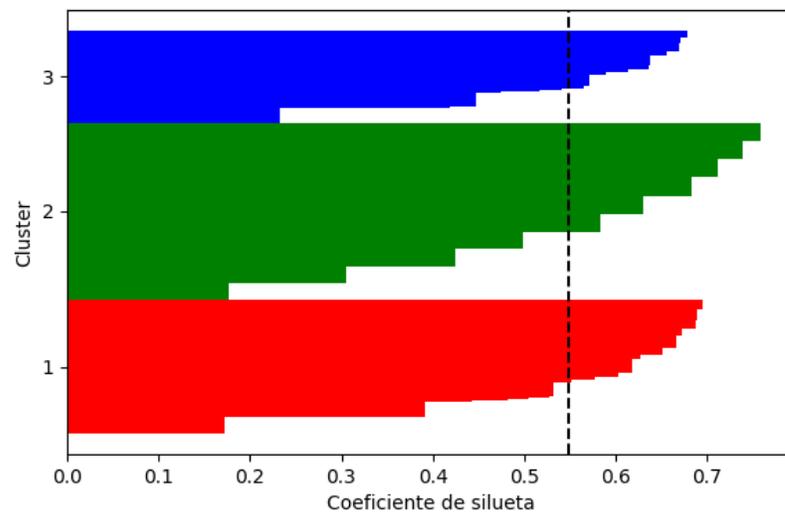
Región	Departamento	Región	Departamento
Pacífico	Cauca	Centro Oriente	Bogotá D.C.
	Chocó		Boyacá
	Nariño		Cundinamarca
	Valle del Cauca		Norte de Santander
Caribe	Atlántico	Llano	Santander
	Bolívar		Arauca
	Cesar		Casanare
	Córdoba		Guainía
	La Guajira		Guaviare
	Magdalena		Meta
	San Andrés y Providencia		Vuapés
	Sucre		Vichada
Eje Cafetero	Antioquia	Centro Sur	Amazonas
	Caldas		Caquetá
	Quindío		Huila
	Risaralda		Putumayo
			Tolima

Tabla A-2.: Estadísticas descriptivas a nivel nacional y por niveles socioeconómicos

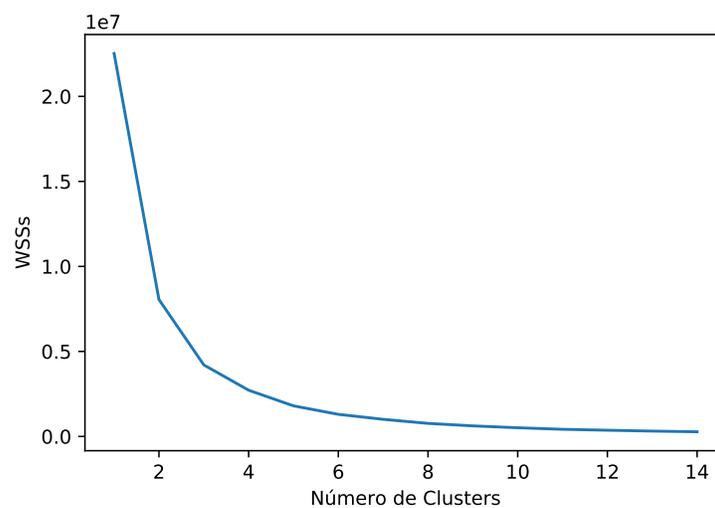
Variables	Nacional	Niveles Socioeconómicos			Variables	Nacional	Niveles Socioeconómicos		
		Bajo	Medio	Alto			Bajo	Medio	Alto
Nº de estudiantes	242.850	79.944	105.382	57.524	Nivel Educativo Madre (%)				
Niveles de Desempeño (%)					Ninguno	7,1 %	9,6 %	5,8 %	5,1 %
Lectura Crítica: 1	4,5 %	6,4 %	3,8 %	2,5 %	Educación Primaria	27,2 %	52,7 %	20,1 %	3,7 %
Lectura Crítica: 2	43 %	53,7 %	40,8 %	30,7 %	Educación Secundaria	43 %	33,7 %	54,8 %	30,1 %
Lectura Crítica: 3	45 %	36,7 %	47,6 %	52,7 %	Educación Técnica	12,8 %	3 %	13,4 %	26,2 %
Lectura Crítica: 4	7,5 %	3,2 %	7,8 %	14,1 %	Educación Superior	9,9 %	1 %	5,9 %	34,9 %
Matemáticas: 1	12,2 %	17 %	10,9 %	7,2 %	Colegio (%)				
Matemáticas: 2	43 %	50 %	41,8 %	34,1 %	Jornada Mañana	64,3 %	63,7 %	64,1 %	65 %
Matemáticas: 3	42 %	32 %	44,4 %	53,1 %	Jornada Tarde	13,8 %	11,7 %	15,5 %	14,3 %
Matemáticas: 4	2,8 %	1 %	2,9 %	5,6 %	Jornada Completa	10 %	14,5 %	8,1 %	7,1 %
Ciencias Naturales: 1	23,6 %	30,6 %	22 %	15,6 %	Jornada Única	11,9 %	10,1 %	12,3 %	13,6 %
Ciencias Naturales: 2	52,1 %	54,9 %	52,4 %	47,4 %	ISCE (Promedio)	5,60	5,17	5,58	6,24
Ciencias Naturales: 3	22,9 %	14 %	24 %	33,9 %	Docente: 1	30,3 %	28,4 %	29,5 %	34,1 %
Ciencias Naturales: 4	1,4 %	0,5 %	1,6 %	3,1 %	Municipio (%)				
Sociales y Ciudadanas: 1	32,9 %	41,9 %	31,2 %	22,6 %	Región Caribe	29 %	33,3 %	29,3 %	22,4 %
Sociales y Ciudadanas: 2	43,9 %	44,2 %	43,4 %	42,1 %	Región Centro Oriente	20,4 %	20,2 %	20,8 %	20 %
Sociales y Ciudadanas: 3	21,3 %	13,3 %	22,2 %	31,4 %	Región Centro Sur	9,2 %	11,3 %	8,2 %	8,2 %
Sociales y Ciudadanas: 4	1,9 %	0,6 %	3,2 %	3,9 %	Región Eje Cafetero	21,6 %	14,8 %	22,4 %	29,7 %
Desempeño Académico (%)					Región Llano	5,4 %	5,1 %	5,4 %	5,7 %
Bueno	32 %	21,7 %	33,3 %	48,3 %	Región Pacífico	14,4 %	15,3 %	13,9 %	14 %
Malo	68 %	78,3 %	66,7 %	51,7 %	Vereda	21,5 %	37,2 %	16,5 %	8,8 %
Personales					Capital	31,3 %	16,3 %	34,7 %	46 %
Edad (promedio años)	17,9	17,47	17,71	17,39	Municipio	47,2 %	46,5 %	48,8 %	45,2 %
Mujer (%)	43,14 %	45,10 %	44,44 %	47,2 %	No Certificado	50,1 %	70 %	45 %	31,9 %

Fuente: Cálculos propios a partir de Saber 11 (2018), Educación Formal (2018), y SICOLE (2019)

B. Número óptimo de cluster en INSE



(a) Coeficiente de silueta



(b) With in Sum of Squares

Figura B-1.: Número óptimo de cluster

Bibliografía

- Accinelli, E., Brida, J. G., and London, S. (2007). Crecimiento económico y trampas de pobreza: ¿cuál es el papel del capital humano? Technical report.
- Aldana, E., Rodríguez, P., and Hederich, C. (1989). El logro educativo en matemáticas y lenguaje de los alumnos de tercer grado.
- Arias Gaviria, J. (2017). Problemas y retos de la educación rural colombiana. *Educación y Ciudad*, pages 53–62.
- Baker, D. P., Goesling, B., and LeTendre, G. K. (2002). Socioeconomic Status, School Quality, and National Economic Development: A Cross - National Analysis of the Heyneman - Loxley Effect on Mathematics and Science Achievement. *Comparative Education Review*, 46(3):291–312.
- Banco Mundial (2009). *La calidad de la educación en Colombia: un análisis y algunas opciones para un programa de política*. Washington D.C.
- Barón, J. and Bonilla, L. (2014). Desempeño relativo de los graduados en el área de educación en el examen de Estado del Icfes. In *Educación y desarrollo regional en Colombia*. Banco de la República, Bogotá.
- Barón, J. D. and Bonilla, L. (2011). La calidad de los maestros en Colombia: Desempeño en el examen de Estado del ICFES y la probabilidad de graduarse en el área de educación. Technical report.
- Becker, G. S. G. S. (1993). *Human capital : a theoretical and empirical analysis, with special reference to education*. The University of Chicago Press.
- Benabou, R. (1996). Equity and Efficiency In Human Capital Investment: The Local Connection. Technical report.
- Blanco, E. (2011). Efectos escolares sobre los aprendizajes en México: una perspectiva centrada en la interacción escuela-entorno. *Papeles de población*, 17(69):219–256.
- Bonilla-Mejía, L., Londoño, E., Cardo-Sosa, L., and Trujillo-Escalante, D. (2018). ¿Quiénes son los docentes en Colombia? Características generales y brechas regionales.
- Colbert, V. (1999). Mejorando el acceso y la calidad de la educación para el sector rural pobre. El caso de la Escuela Nueva en Colombia. *OEI: 50 años de cooperación*.

- Coleman, J. S., Campbell, E., Hobson, C., McPartland, J., Mood, A., Weinfeld, F., and York, R. (1966). *Equality of Educational Opportunity*. US Department of Health, Education & Welfare, Office of Education., Washington, D.C.
- Congreso de la República de Colombia (1994). Ley 115.
- DANE (2019). Metodología General de Educación Formal - EDUC. Technical report, Bogotá.
- Duraisamy, P., James, E., Lane, J., and Tan, J. P. (1998). Is there a quantity-quality trade-off as pupil-teacher ratios increase? Evidence from Tamil Nadu, India. *International Journal of Educational Development*, 18(5):367–383.
- Eide, E. and Showalter, M. H. (1998). The effect of school quality on student performance: A quantile regression approach. Technical report.
- Garces, J. D. and Montes, I. (2015). Rendimiento académico: ¿qué papel juegan los factores institucionales?
- Gardner, H. (1995). Reflections on multiple intelligences: Myths and messages. *Phi Delta Kappan*.
- Gaviria, A. and Barrientos, J. (2001a). Calidad de la educación y rendimiento académico en Bogotá. *Fedesarrollo*.
- Gaviria, A. and Barrientos, J. H. (2001b). Determinantes de la calidad de la educación en Colombia.
- Gertel, H., Giuliadori, R., Herrero, V., and Fresoli, D. (2007). El rendimiento escolar de la población de estudiantes de la educación básica en Argentina: ¿cómo contribuye la gestión de la escuela? *IX Jornadas Argentinas de Estudios de Población*.
- Goldstein, H. (2011). *Multilevel Statistical Models*. University of Bristol, UK, 4th edition.
- Gordillo, D. and Ariza, N. (2005). Evolución de los resultados de la educación en Colombia (1997 - 2003).
- Hadfield, J. (2010). MCMC Methods for Multi-response Generalized Linear Mixed Models: The MCMCglmm R Package. Technical report.
- Hanushek, E. (1986). The Economics of Schooling: Production and Efficiency in Public Schools. *Journal of Economic Literature*, XXIV.
- Hanushek, E. A. (2013). Economic growth in developing countries: The role of human capital. *Economics of Education Review*, 37:204–212.
- Hastie, T., Tibshirani, R., and Friedman, J. (2001). *The Elements of Statistical Learning Data Mining, Inference, and Prediction*. Springer Series in Statistics, Berlín, 2th edition.

- ICFES (2010). Metodología de construcción del índice de nivel socioeconómico de los estudiantes -INSE- y de la clasificación socioeconómica -CSE- de los colegios. page 210.
- ICFES (2016). Reporte de resultados individuales Saber 11. Technical report, Bogotá.
- ICFES (2018). Guía de orientación Saber 11.º para instituciones educativas. page 49.
- Iregui, A. M., Melo, L., and Ramos, J. (2006). La educación en Colombia: análisis del marco normativo y de los indicadores sectoriales. *Revista de Economía del Rosario*.
- John, R. (2010). LA ECONOMÍA Y LA FUNCIÓN DE PRODUCCIÓN EN EDUCACIÓN. Technical report.
- Manski, C. F. (1993). Identification of Endogenous Social Effects: The Reflection Problem. *The Review of Economic Studies*, 60(3):531.
- Mina, A. (2004). Factores asociados al logro educativo a nivel municipal. *Universidad de los Andes*.
- Ministerio de Educación Nacional de Colombia (2018). Estándares Básicos de Competencia.
- OCDE (2016). PISA 2015: Resultados Clave. Technical report.
- Patrinos, H. A., Moreno, V. G., and Alvarez, J. (2007). *Institutional effects as determinants of learning outcomes: exploring state variations in Mexico*. Policy Research Working Papers. The World Bank.
- Ramírez, M. and Téllez, J. (2006). La educación primaria y secundaria en Colombia en el siglo XX. Technical report, Banco de la República.
- Raudenbush, S. W. and Bryk, A. S. (2002). *Hierarchical linear models : applications and data analysis methods*. Sage Publications, Thousand Oaks, 2nd ed. edition.
- Restrepo, P. and Alviar, M. (2004). El logro académico y el efecto colegio en las pruebas Icfes en Antioquia. *Lecturas de Economía*, (60):67–95.
- Rodríguez, M. and Correa, J. (2019). *Impacto del contexto municipal sobre el desempeño académico individual*. Number 90.
- Rozo Alzate, 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.
- Soares, J. F. and Alves, M. T. G. (2013). Efeitos de escolas e municípios na qualidade do ensino fundamental. *Cadernos de Pesquisa*, 43(149):492–517.
- Twisk, J. W. R. (2006). *Applied multilevel analysis : a practical guide*. Cambridge University Press.

-
- Villegas, M. G., Restrepo, J. R. E., Ángel, F. J., and Heredia, J. D. P. (2013). *Separados y desiguales. Educación y clases sociales en Colombia*. Centro de Estudios de Derecho, Justicia y Sociedad, Dejusticia, Santafe de Bogotá.
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.