

***DETERMINANTES DE LAS DIFERENCIAS EN LOS RESULTADOS DE LAS
PRUEBAS ACADÉMICAS DE ESTADO. EL CASO DE LA EDUCACIÓN
MEDIA OFICIAL EN BOGOTÁ.***

CLASIFICACIÓN JEL: I20, I21.

ANDRÉS MAURICIO MUÑOZ SÁNCHEZ

Escuela Colombiana de Ingeniería Julio Garavito

Facultad de Economía

Bogotá D.C., Colombia, noviembre de 2017

DETERMINANTES DE LAS DIFERENCIAS EN LOS RESULTADOS DE LAS PRUEBAS ACADÉMICAS DE ESTADO. EL CASO DE LA EDUCACIÓN MEDIA OFICIAL EN BOGOTÁ.

CLASIFICACIÓN JEL: I20, I21.

ANDRÉS MAURICIO MUÑOZ SÁNCHEZ

Director del trabajo de Grado: ALVARO HERNANDO CHAVES

Resumen:

Existen diferencias entre los resultados de las pruebas SABER 11 que obtienen estudiantes de colegios públicos y privados y aunque estas diferencias se manifiestan en todo Colombia se puede advertir de forma más evidente cuando centramos el análisis en su capital, Bogotá. Se da a conocer el caso de la educación media oficial en Bogotá y sus diferencias con la educación privada mediante un análisis descriptivo a nivel de estudiantes y de colegios y la implementación de la metodología de los modelos multinivel. Este trabajo estima los determinantes de las diferencias en los resultados en la prueba SABER 11 presentada en el segundo semestre del año 2016.

Palabras clave: Educación oficial y privada, calidad de la educación, Logro escolar, efecto plantel, modelo multinivel, SABER 11, Bogotá.

Abstract:

There are differences between the results of the SABER 11 tests obtained by public and private schools students, these differences are evident throughout Colombia but they become much clearer while we focus the analysis on its capital city, Bogotá. Showing this particular case through a descriptive analysis to the student's and school's results and using a multi-level modelling to estimate their achievement's determinants. This paper reports about the differences in results in the SABER 11 test taked in 2016-2.

Keywords: public and private educational programs, quality of education, multi-level modelling, school achievement, SABER 11, Bogotá.

Dedicatoria

Dedico este trabajo de tesis a mi madre María Isabel a quien le debo todo lo que soy y seré.

A mi esposa Diana quien siempre estuvo ahí para mí en todo momento y no me dejó caer por duro que pareciera ser este camino.

A mi hija Juanita Luna quien es mi motivo para levantarme y seguir soñando con un siguiente día.

Para ellas tres dedico todo mi esfuerzo aquí plasmado, ya que es a ellas a quienes siempre llevo en mis más profundos pensamientos.

Agradecimientos

En primer lugar, le doy gracias a la Escuela Colombiana de Ingeniería Julio Garavito y en particular a su facultad de Economía que me otorgó la posibilidad de comenzar este bello proyecto de vida.

Al profesor Álvaro Chaves por todo el esfuerzo y dedicación que me brindó en la dirección de este trabajo y a las profesoras María Constanza e Irma quienes realizaron aportes valiosos y sin los cuales no hubiera sido posible concluir este trabajo de tesis.

Por último, doy gracias a todas las personas que aportaron directa o indirectamente con la consecución de mi trabajo de grado y que en este momento no me ha sido posible nombrarlas. Muchas gracias.

Tabla de contenido

1. Introducción	6
Pregunta de investigación	8
Hipótesis.....	8
Objetivo general:.....	8
Objetivos específicos:.....	8
2. Revisión bibliográfica	9
3. Marco Metodológico.....	14
3.1 Datos y análisis de estadísticas descriptivas	15
3.1.1 Cambios en la estructura de los datos y muestras.....	15
3.1.2 Características de los planteles educativos.....	15
3.1.3 Análisis de las características de los estudiantes	18
3.2 Modelo multinivel	26
3.2.1 Modelo Nulo (modelo I)	27
3.2.2 Modelo con las variables de ajuste (modelo II)	29
3.2.3 Modelo Final (modelo III)	30
3.2.4 Verificación del cumplimiento de los supuestos.....	32
3.2.5 El poder explicativo del Modelo.....	34
4 Conclusiones.....	37
Bibliografía	39
Anexos	42
Anexo 1. Modelo nulo (Colegios oficiales)	42
Anexo 2. Modelo con las variables de ajuste (Colegios oficiales)	42
Anexo 3. likelihood ratio test: significancia conjunta entre el nulo y el 1ro (Colegios oficiales)	43
Anexo 4. Modelo final (Colegios oficiales)	43
Anexo 5. likelihood ratio test: significancia conjunta entre el nulo y el 1ro (Colegios oficiales)	43
Anexo 6. Corrección del Modelo final (Colegios oficiales).....	44
Anexo 7. likelihood ratio test: significancia conjunta entre el 1ro y el 2do (Colegios oficiales)	44
Anexo 8. Modelo nulo (Colegios “No” oficiales).....	45
Anexo 9. Modelo con las variables de ajuste (Colegios “No” oficiales).....	45
Anexo 10. Corrección del Modelo con las variables de ajuste eliminando variable Xgene (Colegios “No” oficiales).....	46
Anexo 11. likelihood ratio test: significancia conjunta entre el nulo y el 1ro (Colegios “No” oficiales)	46

Anexo 12. Modelo final (Colegios “No” oficiales)	46
Anexo 13. likelihood ratio test: significancia conjunta entre el 1ro y el 2do (Colegios “No” oficiales)	47
Anexo 14. Modelo nulo (Colegios de “Ambas” Naturalezas).....	47
Anexo 15. Modelo con las variables de ajuste (Colegios de “Ambas” Naturalezas).....	48
Anexo 16. likelihood ratio test: significancia conjunta entre el nulo y el 1ro (Colegios de “Ambas” Naturalezas)	48
Anexo 17. Modelo final (Colegios de “Ambas” Naturalezas).....	49
Anexo 18. likelihood ratio test: significancia conjunta entre el 1ro y el 2do (Colegios de “Ambas” Naturalezas)	49
Anexo 19. Tabla de Variables	50
Anexo 20. Cálculo de la correlación intraclase para el modelo nulo (Modelo I)	50
Anexo 21. Cálculo de la correlación intraclase para el modelo con variables de ajuste (Modelo II).....	50
Anexo 22. Cálculo de la correlación intraclase para el modelo final (Modelo III)	50
Anexo 23– Cambios en la estructura de la prueba SABER 11	50
Anexo 24 – Cambios en la metodología clasificación de las instituciones educativas	51
Anexo 25 – Clasificación antigua de las instituciones educativas según categoría.	51
Anexo 26 – Clasificación vigente de las instituciones educativas según categoría.	51
Anexo 27 – Cálculo del porcentaje de la varianza explicada.	51
Anexo 28 – Cálculo del porcentaje de la varianza total explicada.	52

1. Introducción

Existen diferencias entre los resultados de las pruebas SABER 11 que obtienen estudiantes de colegios públicos y privados y aunque esta disparidad se manifiesta en los resultados de todo Colombia, las diferencias se pueden advertir de forma más evidente cuando centramos el análisis en su capital, Bogotá.

Tanto la presentación del examen como también la calificación obtenida en las pruebas SABER 11 tienen una gran importancia para el desarrollo personal de los estudiantes colombianos y sus metas de vida a corto y largo plazo, ya sea porque es necesario haber presentado el examen para acceder al mercado laboral formal o para acceder a las oportunidades académicas que ofrece la educación superior, debido a que las universidades en su gran mayoría hacen uso de los componentes evaluados en las pruebas SABER 11 para seleccionar estudiantes en sus procesos de admisión. Por esto, la importancia de la calidad de educación que reciben los distintos habitantes de un país ha venido aumentando, tanto en la búsqueda de un respaldo del crecimiento y desarrollo económico de un país, como también en la búsqueda de garantías personales para afrontar sus decisiones de plan de vida.

El Instituto Colombiano para la Evaluación de la Educación tiene la responsabilidad de evaluar la calidad de la educación que ofrecen los establecimientos educativos del país, por medio de sus pruebas SABER que pretenden medir el desempeño de sus estudiantes en ciertas competencias básicas definidas por los requerimientos del Ministerio de Educación Nacional de Colombia. Es por eso que, como se muestra en el cuadro 1, la prueba SABER 11 tiene una relevancia significativa para los individuos que la presentan como también para las instituciones de control y evaluación para poder tomar decisiones de política.

Cuadro 1. Importancia de la Prueba SABER 11

ESTUDIANTES	INSTITUCIONES
<ul style="list-style-type: none">• Realizar una autoevaluación de las competencias adquiridas y así formar una perspectiva de su proyecto profesional junto con sus metas de vida a corto y largo plazo.• Acceso al mercado laboral formal.• La presentación del examen de Estado SABER 11 no es requisito para graduarse de bachiller, educación media.• La presentación del examen de Estado SABER 11 es requisito indispensable para ingresar a un programa de educación superior, así como para la obtención del título, según las Leyes 30 de 1992 (Artículos 14 –literal a) y 1324 de 2009 (Artículo 7°).	<ul style="list-style-type: none">• Evaluar la calidad de la educación que ofrecen los establecimientos educativos del país. Según el desempeño de sus estudiantes en unas competencias básicas definidas por los requerimientos del MinEdu.• Construir indicadores de calidad de la educación, para el control y supervisión del servicio público educativo.• Proveer información pertinente a los establecimientos educativos para autoevaluarse y gestionar políticas particulares de mejora de la calidad del establecimiento.• Proveer información pertinente al Ministerio de Educación Nacional y a las entidades territoriales que posteriormente establecerán bancos de proyectos de mejoramiento de la calidad de la educación, y podrán destinar recursos para financiarlos.

Fuente: Elaboración propia.

Pese a que ya existe un acervo oportuno de trabajos que realizan esfuerzos por identificar y estimar los determinantes de la calidad de la educación con base en

los resultados académicos que obtienen los estudiantes en las pruebas estatales y que previamente aquí se ha planteado como tema central de estudio y que posteriormente se va a desarrollar el análisis, es preciso resaltar que este estudio posee bondades que se originan en el aprovechamiento de la riqueza de las bases de datos disponibles para la última prueba de Estado SABER 11 y la clasificación de establecimientos y sedes SABER 11 para el 2016 que cuentan con nuevos componentes evaluados y que esta inclusión de componentes pueden aportar mejores herramientas de estudio y de esta forma ampliar la perspectiva con la que este trabajo abordará el análisis de este fenómeno y por último, la necesidad de actualizar los resultados de estudios anteriores debido a los recientes ajustes que se han realizado tanto a la estructura del examen como a la tendencia contemporánea de diferenciación de los resultados de estudiantes de colegios oficiales y privados.

En las últimas décadas se ha evidenciado un aumento sostenido en la cobertura y avances en el acceso a la educación básica y media, sin embargo, las diferencias entre los estudiantes de distintas situaciones socioeconómicas y características relacionadas con distintos entornos educativos se han hecho más visibles. La distinción de las cualidades entre los docentes (especialización, experiencia y salario) y las diferencias entre los planteles educativos (área rural o urbana, pensión e infraestructura) repercuten en diferencias en los logros académicos obtenidos en la prueba SABER 11.

Con el propósito de poder proporcionar un informe del comportamiento en el 2016-2 de los distintos determinantes de las diferencias en los resultados de las pruebas SABER 11, tanto a nivel estudiante como a nivel del plantel educativo, se realiza este trabajo académico que agrupa una revisión de literatura y herramientas estadísticas y econométricas para entender el panorama de la efectividad de la educación pública en Bogotá.

Este trabajo de grado se compone por cuatro apartados, de los cuales esta introducción es la primera. En el segundo apartado, se presenta la revisión bibliográfica en forma de estado del arte de los avances teóricos y empíricos por identificar y estimar los determinantes de la calidad de la educación en Colombia y un repaso general de la perspectiva internacional acerca de este interesante fenómeno. En el tercero, se desarrolla el marco metodológico en el que en primera instancia se introducen las estadísticas descriptivas de las variables objeto de estudio, con el fin de analizar previamente la perspectiva de esta problemática haciendo uso de las dos bases de datos disponibles para este estudio (la base de establecimientos educativos y la de estudiantes), para luego poder desarrollar la metodología empírica y econométrica por medio de un modelo multinivel, con el cual se analizarán los determinantes en los rendimientos educativos y su significancia dentro del modelo. En el cuarto apartado, se culmina este trabajo con la exposición de las principales conclusiones de este estudio sobre los determinantes de las diferencias en los resultados de las pruebas académicas de estado, el caso de la educación media oficial en Bogotá.

Pregunta de investigación

¿Existen diferencias en los resultados de las pruebas de Estado para el ingreso a la educación superior a nivel de colegios oficiales y privados en Bogotá para el año 2016?

Hipótesis

Las diferencias en los resultados de las pruebas SABER 11 están determinados tanto por factores socioeconómicos como por factores relacionados con el colegio, la jornada de estudio, ubicación del colegio, valor de la pensión y factores inherentes al tipo de colegio.

Objetivo general:

- Analizar las posibles diferencias en los resultados de las pruebas de estado SABER 11 a nivel de colegios públicos y privados para Bogotá en el año 2016.

Objetivos específicos:

- Realizar un análisis descriptivo de los factores que explican las diferencias en los resultados de las pruebas Saber 11.
- Estimar econométricamente la incidencia del tipo de colegio y de los factores socioeconómicos sobre el resultado en el SABER 11 en Bogotá para el año 2016.

2. Revisión bibliográfica

Para poder identificar los determinantes de la efectividad de los planteles educativos en los resultados del aprendizaje y pruebas estatales, como la prueba SABER 11, en las últimas décadas se han venido realizando acercamientos empíricos hacia estimar los determinantes de estos resultados y así poder guiar las prácticas de aprendizaje y la eficiencia escolar, procurando entender los factores que tienen mayor efecto sobre los estudiantes para obtener mejores resultados educativos en comparación con otros alumnos (Zambrano, 2012). Por consiguiente, a partir de la revisión de la literatura que se establecerán cuáles son los determinantes sugiera la teoría y que pueden ser susceptibles de estimar econométricamente.

La visión de una educación arte en el Mercantilismo, se puede apreciar en ideas de pensadores de la época y su posición sobre cómo y para que las personas deben ser formados.

“...Según Martínez (1997), antes del siglo XVIII, el término con el que se hacía referencia a las destrezas, habilidades y adiestramiento técnico, era arte. En diferentes obras, los mercantilistas mencionan que uno de los propósitos más importantes de la política estatal consistía en aumentar el arte de la nación, a fin de que ésta pudiera exportar mercancías, y con esto, obtener una balanza comercial favorable que se traduciría en metales preciosos, expresión de la riqueza de un país.”

Cardona, M., Montes, I., Vásquez, J., Villegas, M. y Brito, T. (2007)

Para Bernard Mandeville, médico, filósofo y sobre todo un pensador de la moral y economía de finales del siglo XVII, la educación tenía un cometido social que todos los individuos deben formarse como compromiso del individuo con el colectivo “...a lo largo de la evolución de las sociedades se logra la incorporación de la obligatoriedad social en los individuos, gracias a la educación entendida como la formación (Bildung) del hombre en la cultura y a un aprendizaje mecánico, que por la práctica se convierten en normas de convivencia.”¹, sin embargo, también en sus escritos expone un pensamiento característico de la época planteando que existe una diferencia entre formación académica y formación (experiencia) laboral y además de resaltar que ambos tipos de formación tienen un resultado diferente en población con distintas condiciones socioeconómicas, de donde destacan afirmaciones como la siguiente:

“Going to school in comparison to working is idleness, and the longer boys continue in this easy sort of life, the more unfit they'll be when grown up for...labor...”²

La idea, excesiva, de que los pobres se deben dedicar a formarse para el mundo laboral lo más pronto posible y que si se dedicaban a la formación académica sería un tiempo mal empleado de forma ociosa y que esa formación no laboral, como leer, puede crear falsas perspectivas en los pobres y distorsionar el buen entendimiento y funcionamiento de la sociedad.

Con el paso de las décadas y desde un punto de vista mucho más social, Hume afirmaba que la educación familiar y comunitaria tiene la función de fortalecer convicciones morales y para ello es importante que ambas educaciones

¹ Rios, M. (2007)

² Rios, M. (2002)

muestren una tendencia convergente en sus intereses más valiosos en común³. Por su parte, Smith le daba una mayor importancia a la formación personal, para él la educación hace que las personas se logren diferenciar más allá de las cualidades que la naturaleza les ha brindado y así poder desempeñarse en distintos entornos laborales.⁴ Malthus por otro lado prestaba mucha más atención a la educación como elemento sustituto de las ayudas de carácter benéfico a los pobres y como una necesidad prioritaria para que ellos mismos solucionen su situación precaria.⁵

A puertas del siglo XX existía una noción pre-capital humano, donde toda inversión en educación se llegaba a considerar como una simple decisión de consumo, que normalmente revelaba expectativas de satisfacción social inmediata para poder permitirle al individuo tener un acervo de habilidades sociales y políticas que se declaran como determinantes del equilibrio de una sociedad pacífica y ordenada.

Pasada la mitad del siglo XX, se empezó a desarrollar en la Escuela de Chicago el concepto que hoy conocemos como Capital Humano, con economistas que por sus aportes en la economía siguen siendo recordados en nuestros días como Schultz, Mincer y Becker entre otros.

Para Becker (1962), la inversión en capital humano en formación académica y profesionalización tiene unos efectos en el aumento de los ingresos salariales de quienes la realizan, que son más notorios sobre todo cuando los individuos son ya mayores, puesto que ya no tiene que hacerse cargo de estos costos de su educación. Sin embargo, Becker realiza un esfuerzo para resaltar que las diferencias entre los ingresos se deben en gran parte a las diferencias entre clases de capital humano⁶. Además, Becker, et al. (1990) aprovechan para resaltar que así como Malthus menospreció la importancia del crecimiento de la población sobre el crecimiento económico, de igual forma el modelo neoclásico de crecimiento ha menospreciado el aporte del capital humano sobre el crecimiento económico y formulan que existen dos estados estacionarios estables, el primero para poblaciones con familias grandes y baja acumulación de capital humano y el segundo para familias pequeñas y un alto capital humano y físico.⁷

De acuerdo con Schultz (1962), la inversión en educación se empieza a entender como una elección de inversión en sí mismos, para poder acumular conocimientos que luego van a tener un valor significativamente alto en el mercado laboral al mejorar sus capacidades como productores y consumidores.⁸ Además de la importancia de la inclusión del concepto del capital humano en los modelos de crecimiento y poder escapar de la productividad marginal decreciente que ofrecen los capitales físicos, mientras que por su parte el capital humano tiene una productividad marginal creciente por lo tanto ve en

³ Gonzalez, A. (2013)

⁴ Cardona, et al. (2007)

⁵ Cardona, et al. (2007)

⁶ Cardona, et al. (2007)

⁷ Becker, et al. (1990)

⁸ Schultz, T. (1962)

la educación un gran aporte al crecimiento económico, que ha venido aumentando su aporte a través de las décadas.⁹

Mincer (1975) logra encontrar con sus estimaciones que la desigualdad entre los ingresos personales se debe a una diferencia entre las habilidades individuales, al intentar desarrollar la paradoja de Pigou respecto a cómo se puede reconciliar una distribución normal de capacidades y una distribución considerablemente sesgada de ingresos. Además, de que el nivel de escolaridad y la experiencia laboral (entrenamiento formal y un mayor rango laboral) son factores determinantes para su capacidad productiva, del ingreso de las personas y por medio del aumento de la productividad se obtiene una conexión con el crecimiento económico. Siguiendo el trabajo de Mincer, Klevmarken, A. y Quigley, J. (1976) apuntaron con su modelo a distinguir el tipo de capital humano que tiene una mayor influencia en los ingresos de las personas y encontraron que tanto la edad como la experiencia son determinantes significativos para los distintos perfiles de inversión en capital humano.¹⁰

Uno de los avances significativos que relacionan la disciplina económica con la educación, lo realizaron Alexander, L. y Simmons, J. (1975), quienes empiezan a plantear una función de producción de educación como herramienta analítica micro-fundamentada para exponer una relación entre el gasto en educación y la eficiencia de la educación. Desde aquí ya se empieza a generar interés en otros posibles determinantes de la eficiencia de la educación como los son el contexto familiar e individual, el contexto escolar y otras influencias externas como el contexto de la comunidad donde interactúa cada estudiante. Tampoco dudan en distinguir que en los países desarrollados existen determinantes muy fuertes que por ejemplo en los países pobres no lo son tanto, poniendo como ejemplo que para los países pobre la habilidad de aprender lenguajes extranjeros va acompañada de buenos resultados académicos en general. La importancia de la certificación adecuada de profesores sobre todo en los grados de bachillerato.

Por su lado, Vélez y Psacharopoulos (1998) proponen que los estudiantes que cargan consigo condiciones socioeconómicas desfavorables tienen una tendencia predominante de no invertir suficientes años a su formación educativa como si lo harían los estudiantes con condiciones socioeconómicas mucho más estables y acomodadas.

Hanushek y Kimko (2000) presentan en su estudio que un aumento de las habilidades en matemáticas y ciencias tiene un impacto mucho mayor sobre el crecimiento económico que un aumento en la escolaridad *per se*. Nuevamente, Hanushek pero esta vez junto Wößmann (2007) proponen que para mejorar la calidad de la educación debe existir una mayor responsabilidad de parte de los individuos, los colegios y las instituciones que corresponda a una alta competencia y conciencia a la hora de tomar acciones en el rendimiento académico de los estudiantes.

Magdalena Claro (2011), en América Latina los programas de integración de las TIC a la educación están aumentando, sin embargo, escasean las estimaciones sobre sus efectos. Resalta la importancia no sólo de tener acceso y usar de forma efectiva las TIC como método de integración a la sociedad y cultura

⁹ Schultz, T. (1962)

¹⁰ Klevmarken, A. y Quigley, J. (1976)

digitales, por lo que habrá grupos de población marginados ya sea por la falta de acceso a estas tecnologías o por no saber darle un buen uso a estas.

Gamboa, L. y Krüger, N. (2016), mediante un método no paramétrico, basado en la generación de contrafactuales que descompone las brechas entre factores observables y no observables, para analizar las diferencias en los resultados de las pruebas PISA en América Latina durante los años 2009 y 2012, concluye que las diferencias en la prueba son determinados significativamente por la diferencia entre la edad de acceso a la educación inicial y el tiempo de permanencia. A su vez resalta que América Latina tiene un panorama heterogéneo tanto en la cobertura de educación inicial como en su agenda política, la región ha venido tomando compromisos de política de aumento del acceso a preescolar impulsados por el consenso político y académico.

Para Vegas y Petrow (2008), las particularidades individuales, familiares, las características de los planteles educativos y la interacción entre el contexto económico, político y social, determinan la calidad del aprendizaje de los estudiantes. Palos, R. et al (2011), mediante un estudio piloto para identificar variables motivacionales y cognitivas que puedan impactar sus desempeños académicos, encuentran que los estudiantes que mejores resultados académicos obtienen se deben a que desarrollan estrategias de aprendizaje autorreguladas por ellos mismos, todo esto debido a la necesidad y la motivación de un ambiente de aprendizaje debidamente estructurado impulsado individualmente. En el estudio de Elias, H. et al (2011) se identifican seis variables motivacionales que influyen fuertemente en el desempeño académico de los estudiantes de último año: la perspectiva del futuro, la necesidad por alcanzar resultados, metas de aprendizaje (desempeño y titulación), valoración y expectativa, autoeficacia y autodeterminación.

Por su parte, Singh Dhillon y Kaur Johal (2016), para estudiar los determinantes del logro académico es necesario acercarse a factores cognitivos y no cognitivos como la inteligencia, la forma de pensar, motivación, nivel de aspiración, interés, ansiedad, entorno familiar, entorno escolar, entorno social, hábitos de estudio, optimismo, esperanza de alcanzar los logros etc (Vendal, 1994), sin embargo, el desempeño académico suele ser determinado significativamente por la inteligencia básica y la disciplina para preparar los exámenes. De acuerdo con Nuñez, J. (2007), es importante para desarrollar cualquier conocimiento y disciplina “poder” (capacidades, conocimientos, estrategias y destrezas necesarias) y “querer” (tener la disposición, intención y motivación), por lo que el desempeño depende tanto de factores cognitivos como de factores motivacionales.

En los últimos 30 años la estimación del efecto plantel ha tomado mucha más fuerza debido al incremento de herramientas teóricas y empíricas, como lo es la metodología del modelaje multinivel, que aportan un mayor valor a los análisis realizados. Por ejemplo, al emplear una modelación multinivel, Castaño (1998), encontró que el 29% de la variabilidad del rendimiento de los estudiantes es explicada por los factores de contexto a nivel colegio. Con base en los resultados de las pruebas del ICFES para el año 1999 en Bogotá, Gaviria, A. y Barrientos, J. (2001) encuentran que el efecto de *la educación de los padres* se transfiere por medio de *la calidad de los planteles* educativos que escogen para sus hijos

y la incidencia de los planteles sobre el rendimiento escolar. Adicionalmente, Núñez et al. (2002) encuentran que no existe ninguna diferencia significativa entre un colegio público que no es administrado por el gobierno y uno privado. Por medio de un análisis multinivel, Correa, J. (2004) estima que las variables asociadas al plantel tienen un efecto significativo sobre el logro académico de la prueba ICFES, para 16998 alumnos en 427 colegios de la ciudad de Cali. Mediante un logit ordenado generalizado Chica, S., Galvis, D. y Ramírez A. (2010) pretenden identificar los determinantes de los rendimientos en la prueba ICFES del 2009, para Colombia, y encuentran que los factores relacionados con los padres, tal como el grado de formación y su nivel de ingreso, tienen un impacto alto de forma significativa sobre el logro académico. Por su parte, Rodríguez, G. (2014) por medio de un modelo multinivel, busca identificar factores a nivel individual y del plantel universitario que determinen los resultados obtenidos en las pruebas SABER PRO del 2009 en la región caribe, como resultado de su estudio encuentra un débil efecto de variables socioeconómicas y un fuerte impacto de factores a nivel del plantel, definido como el efecto universidad. Para el caso de Cali, Zambrano (2012) realiza un acercamiento académico por medio de la metodología multinivel para determinar los factores escolares, familiares, socioeconómicos, algunas prácticas y métodos pedagógicos que inciden los resultados que obtienen en el área de matemáticas, utilizando los datos de las pruebas TIMSS 2007 y encuentra que no solo el plantel educativo sino también los gustos individuales por las áreas de estudio tienen una incidencia significativa sobre los logros escolares de los estudiantes. Por último, Hernández (2015) encuentra por medio, también, de un modelo multinivel, que el plantel educativo tiene un efecto significativo sobre el rendimiento alcanzado por el estudiante en las áreas de matemáticas y lenguaje, en Cundinamarca para las pruebas de SABER 11 presentadas en el 2014-2.

3. Marco Metodológico

Para poder empezar con el análisis descriptivo y posteriormente econométrico de los determinantes de los rendimientos en la prueba SABER 11 corresponde exponer los motivos por los cuales se ha escogido la metodología propuesta en este trabajo.

El presente trabajo utiliza las bases de datos disponibles para la última prueba de Estado SABER 11 y la Clasificación de establecimientos y sedes SABER 11, recuperadas del sitio FTP (<ftp://ftp.icfes.gov.co/>) del ICFES, con el fin de captar características socioeconómicas a nivel de estudiante, junto con características propias de las instituciones educativas. La publicación de las bases de datos se realiza aproximadamente tres meses después de que los estudiantes la hubiesen presentado, tal como se aprecia en la siguiente tabla, que contiene las fechas de presentación y publicación de las bases de datos utilizadas en este trabajo de tesis.

Tabla 1. Presentación y publicación de las bases de datos.

	Presentación	Publicación
Calendario A	08/27/16	11/29/16
Calendario B	03/02/17	06/21/17

Fuente: Elaboración propia.

La base presenta una estructura de datos jerárquica, donde cada muestra se puede agrupar en diferentes niveles (ver cuadro 2), un nivel correspondiente a las características propias de los estudiantes y otro nivel que contiene las características de las instituciones educativas, que particularmente en este trabajo se quieren contrastar.

Tabla 2. Estructura jerárquica propuesta en este estudio.



Fuente: Elaboración propia.

Debido a la estructura anidada de los datos la posibilidad de una mayor homogeneidad entre los individuos de un mismo grupo o subgrupo discrepa con

el supuesto de independencia y hace necesario buscar una alternativa al uso de los modelos de Mínimos cuadrados ordinarios tradicional.

De este modo, para poder realizar el análisis de manera adecuada de los determinantes de los rendimientos en la prueba SABER 11 se emplea primero un análisis descriptivo sobre las características de los planteles educativos y sus estudiantes aprovechando las bondades de las dos bases disponibles, para luego emplear una metodología que tiene algo más de treinta años posicionándose como la herramienta más adecuada para analizar el fenómeno de los determinantes de los rendimientos académicos, los Modelos Multinivel.

3.1 Datos y análisis de estadísticas descriptivas

3.1.1 Cambios en la estructura de los datos y muestras

A partir del segundo semestre del 2014 se realizó un cambio a la estructura de las pruebas SABER 11, como un esfuerzo del Gobierno nacional de Colombia para consolidar el sistema de evaluación estandarizada de la educación en el país¹¹. El ICFES ha realizado el esfuerzo por incluir nuevas variables de observación al SABER 11 y también revisar el progreso del desempeño de los estudiantes desde su primaria con las nuevas pruebas debido a la necesidad de medir mejor las competencias de los bachilleres en Colombia y así poder aportar a la elección de mejores estrategias para mejorar la calidad de la educación en Colombia.

Los cambios que nos interesan en este trabajo se pueden diferenciar en los cambios que afectan directamente a la calificación de los estudiantes (Ver anexo 23) y los cambios que afectan directamente a la clasificación de las instituciones educativa (Ver anexo 24 y 25).

La nueva metodología de clasificación de las de establecimientos y sedes SABER 11, realizó cambios que buscan premiar la homogeneidad en las instituciones como también los puntajes más altos que se obtienen en la prueba SABER 11 (Ver anexo 24). Y pasa de clasificar estos establecimientos educativos según su Índice General en 7 categorías a calificarlos en tan solo 5 categorías, como se puede apreciar en el anexo 26.

3.1.2 Características de los planteles educativos

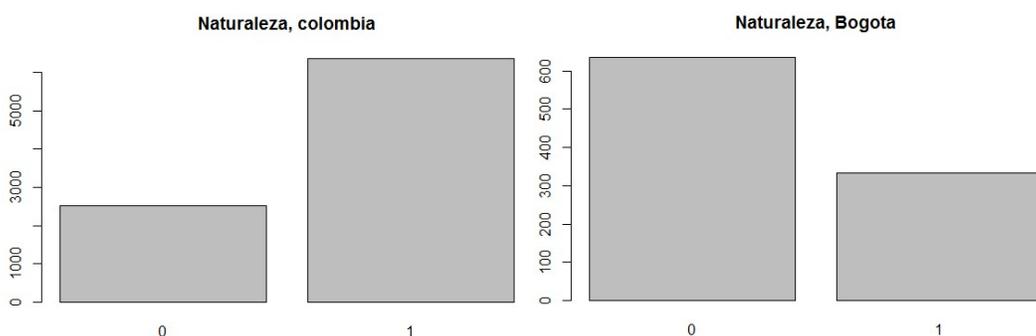
A continuación, se realiza un análisis de las características de los planteles educativos.

Cuando empezamos a utilizar base de datos de planteles educativos del ICFES, empezamos encontrando que para Colombia en su totalidad la composición según naturaleza es de un 28.3% (2511) de colegios “No oficiales” y de un 71.7% (6361) de colegios “No oficiales” con un total de 8872 Colegios registrados en la muestra para el semestre 2016-2. Para Bogotá se aprecia un cambio contundente en la composición según naturaleza comparado a con los resultados obtenidos a nivel nacional que mostraba una mayoría de los colegios oficiales y para el caso bogotano el dominio es para los colegios -no oficiales- donde un 65.5% (635) de colegios “No oficiales” y de un 34.5% (334) de colegios

¹¹ Mineducacion.gov.co. (2014).

“No oficiales” con un total de 969 Colegios registrados en la muestra para el semestre 2016-2.

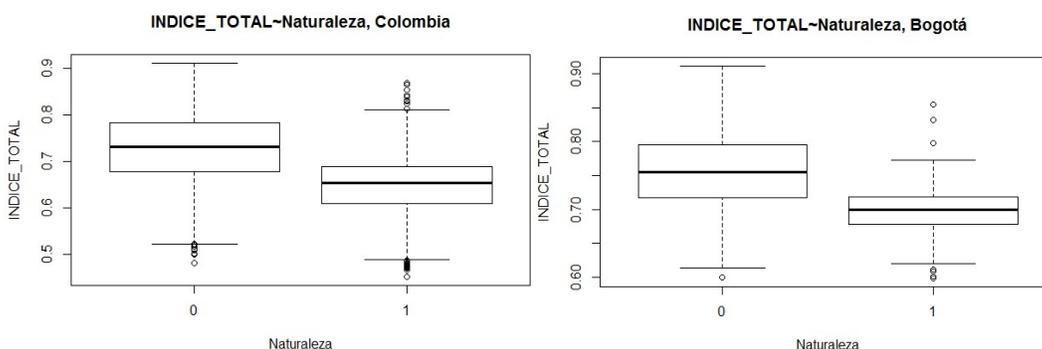
Grafico 1.



Fuente: Clasificación de establecimientos y sedes Saber 11, 2016. Elaboración propia.

El Índice general del establecimiento educativo mide como se está comportando el grupo de estudiantes de un colegio en las pruebas del ICFES. (Ver gráfico 2)

Grafico 2.



Fuente: Clasificación de establecimientos y sedes Saber 11, 2016. Elaboración propia.

El grafico (2) de caja revela los primeros indicios de que hay diferencias entre los colegios de naturaleza oficial y los no oficiales, dado que observamos que tienen una media diferente (tanto para el caso bogotano, como también para el colombiano) y que los máximos para los colegios públicos son significativamente más bajos que los máximos para el caso de los colegios privados, sobre todo para nuestro caso de estudio bogotano.

Tabla 3.

	Colombia		Bogotá	
	No Oficial	Oficial	No Oficial	Oficial
Media (Índice Total)	0.727	0.649	0.758	0.698

Fuente: Clasificación de establecimientos y sedes Saber 11, 2016. Elaboración propia.

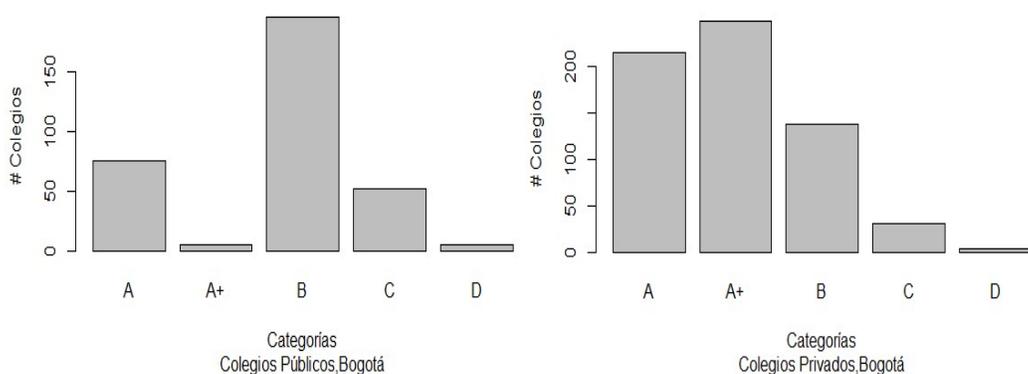
La tabla 3 presenta los resultados de un test de diferencia de medias para el índice total entre los colegios de naturaleza y oficial y los no oficiales. Aplicando la prueba t para las dos muestras ($H_0: \mu_{priv} - \mu_{pub} = 0$; $H_1: \mu_{priv} - \mu_{pub} \neq 0$) muestran que se logra rechazar la hipótesis nula y podemos observar que hay una diferencia significativa entre las medias del Índice total a favor de los colegios

privados frente a los colegios oficiales tanto en el caso bogotano ($t = 20.844$, $df = 953.34$, $p\text{-value} < 2.2e-16$) así como para Colombia en general ($t = 20.844$, $df = 953.34$, $p\text{-value} < 2.2e-16$).

Al utilizar la categorización que los colegios reciben según los resultados que obtienen sus estudiantes nos permite ver como los resultados que obtienen los estudiantes en sus pruebas SABER van otorgando una calificación de cada establecimiento educativo y así mismo, ver desde otra perspectiva el efecto plantel sobre los rendimientos en las pruebas SABER 11.

Con respecto a la categoría que recibe cada colegio, se debe resaltar que esta clasificación pretende premiar tanto los puntajes superiores obtenidos como de igual forma también la homogeneidad de la población de estudiantes que posee el colegio.

Grafico 3. Colegios según Categorías y Naturaleza



	A+	A	B	C	D
Públicos	5	76	196	52	5
Privados	248	214	138	31	4

Fuente: Clasificación de establecimientos y sedes Saber 11, 2016. Elaboración propia.

Por su parte, el grafico 3 muestra la existencia de diferencias entre colegios oficiales y no-oficiales. Una primera lectura nos permite notar que la mayoría de colegios privados se encuentran entre las categorías A+ (248), A (214) y B (138), en cambio, los colegios oficiales se encuentran entre las categorías A (76), B (196) y C (52). Una segunda lectura nos permite notar que muy pocos de los colegios bogotanos en general se encuentran en la última categoría, D, con 5 colegios públicos y otros 4 colegios privados. Por último, en nuestra tercera lectura del gráfico de torta encontramos que la mayoría de los colegios de la máxima categoría, A+, son colegios privados, con 248 colegios y por el otro lado, solo encontramos 5 colegios públicos.

Con respecto al género de la población que atienden los planteles educativos en Colombia y Bogotá, podemos afirmar la existencia de una tendencia a transformar en mixtos los colegios y por ello, son escasos los planteles que atienden a un solo género en nuestro estudio y anticipando que con el tiempo las diferencias en los resultados no tendrán mayor relación con el género de la población que atiende un colegio puesto que los colegios empiezan a inclinarse

por atender a los dos géneros por igual. y por lo anterior, el género de la población que atienden los establecimientos educativos no es una variable a tomar en cuenta para nuestro estudio econométrico debido a que las diferencias que puedan generar en los resultados están mucho más atadas a características como la naturaleza de los colegios o características socioeconómicas de los escasos planteles que quedan que atienden exclusivamente a un solo género.

Tabla 4.

	Colombia			Bogotá		
	F	M	MI	F	M	MI
Oficiales	77	13	6271	5	3	326
No Oficiales	125	43	2343	41	15	579
Total	202	56	8614	46	18	905

Fuente: Clasificación de establecimientos y sedes Saber 11, 2016. Elaboración propia.

3.1.3 Análisis de las características de los estudiantes

a. Estudiantes según la naturaleza del colegio

Para la propuesta central de este informe pretendemos mostrar que existen diferencias entre los grupos de estudiantes que presentan la prueba SABER 11, que son determinadas significativamente por la naturaleza del colegio donde está inscrito el estudiante al momento de presentar estas pruebas y las características intrínsecas a esta naturaleza oficial y no-oficial de los establecimientos educativos.

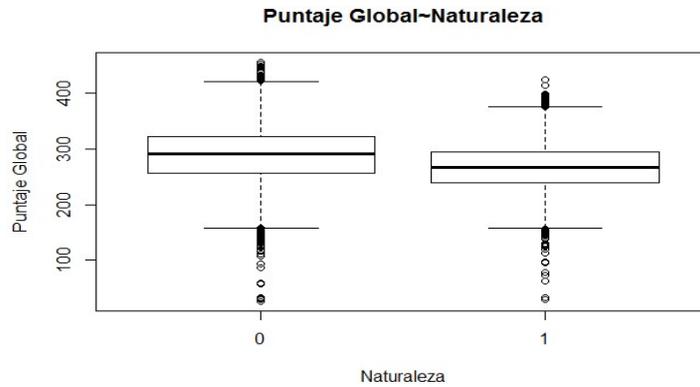
La primera noción que nos interesa exponer con el propósito de sustentar diferencias dentro del grupo de colegios de naturaleza oficial y no-oficial está relacionada con la cobertura y su capacidad. Para el caso bogotano, encontramos que la mayoría de los estudiantes que presentaron la prueba SABER 11 en 2016-2 provienen de colegios oficiales (49705 inscritos de colegios oficiales frente a 41384 de colegios privados). Con números grandes no parece quizás ser una diferencia alarmante, sin embargo, si se compara la cantidad de estudiantes que presentaron la prueba con la cantidad de colegios según su naturaleza, podemos advertir que si existe una diferencia significativa en la cantidad de estudiantes por colegio según naturaleza, donde cada colegio oficial atendió en promedio 148.82 estudiantes mientras que un colegio privado en promedio presentó 65.17 estudiantes, donde al parecer la sobrepoblación es una de las características intrínsecas de la educación oficial a resaltar en el caso bogotano.

El gráfico (4) de caja nos muestra, prematuramente, diferencias entre los resultados en la prueba SABER 11 que obtiene el grupo de estudiantes de colegios públicos frente al grupo de estudiantes de colegios privados. como era de esperarse, podemos observar que los máximos para el grupo de estudiantes de colegios públicos son más bajos que los máximos para el grupo de estudiantes de colegios privados, presentando una diferencia en su media de puntaje global.

A continuación, se presentan los resultados del test de diferencia de medias ($t = 77.128$, $df = 81258$, $p\text{-value} < 2.2e-16$), que muestra se logra rechazar la

hipótesis nula¹² y por consiguiente, se puede validar que hay una diferencia significativa entre las medias del puntaje global a favor de los colegios privados, con una media de 289.09, frente a los colegios oficiales, con media de 266.48.

Grafico 4.



Fuente: SABER 11 (2016-2), 2016. Elaboración propia.

b. Estudiantes según su género

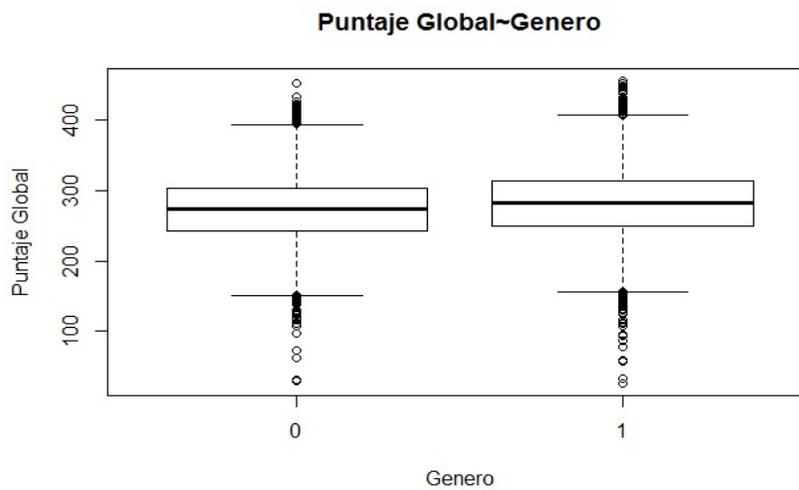
La tabla 5 presenta la cantidad de estudiantes que presentaron la prueba según su género y la naturaleza del colegio del cual proviene cada estudiante sin mostrar a primera vista una mayor diferencia entre géneros.

Tabla 5.

	F	M	Total
Oficiales	22566	27139	49705
No Oficiales	20258	21126	41384
Total	42824	48265	

Fuente: SABER 11 (2016-2), 2016. Elaboración propia.

Grafico 5.



Fuente: SABER 11 (2016-2), 2016. Elaboración propia.

¹² $H_0: \mu_{priv} - \mu_{pub} = 0$; $H_1: \mu_{priv} - \mu_{pub} \neq 0$.

No obstante, el gráfico (5) de caja nos permite evidenciar diferencias significativas ($t = -28.66$, $df = 89502$, $p\text{-value} < 2.2e-16$) en los resultados que obtiene el grupo de estudiantes de mujeres (con una media de 281.26) con respecto al de grupo de hombres (con una media de 272.76).

c. *Estudiantes según si tienen acceso a computador y/o internet en su hogar.*

El acceso a las TIC y el buen manejo de las mismas representa uno de los mayores retos para las políticas públicas educativas, para de esta forma poner el mundo de la informática al servicio de los estudiantes (Magdalena Claro, 2011). Por lo anterior es necesario realizar un análisis sobre dos de las preguntas relacionadas con las TIC en el formulario de ingreso para presentar la prueba SABER 11, relacionadas con el acceso a computador y/o internet en su hogar.

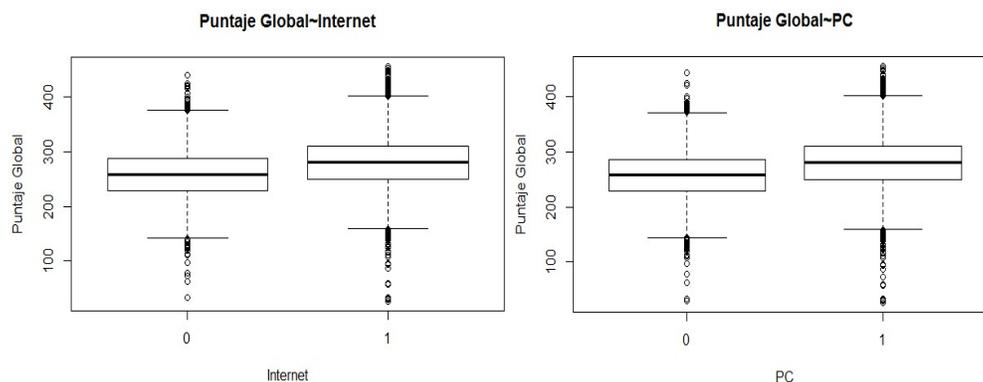
Tabla 6.

	Oficiales		No Oficiales	
	No PC	Si PC	No PC	Si PC
No Internet	8.25%	4.35%	2.43%	1.86%
Si Internet	4.13%	37.84%	1.57%	39.58%

Fuente: SABER 11 (2016-2), 2016. Elaboración propia.

De la cual se puede concluir que en su gran mayoría (77,42%)¹³ los estudiantes que presentaron la prueba 2016-2 cuentan con acceso tanto al computador como al internet en su hogar, sin presentar mayor diferencia entre los grupos de estudiantes de colegios públicos y privados.

Gráfico 6.



Fuente: SABER 11 (2016-2), 2016. Elaboración propia.

El gráfico 6 muestra la existencia de diferencias significativas en los resultados que obtiene el grupo de estudiantes que tiene acceso a computador en su hogar, con una media de 280.48, con respecto al grupo de estudiantes no lo tiene, con una media de 257.73 ($t = -60.183$, $df = 22048$, $p\text{-value} < 2.2e-16$). Y de igual forma, se obtienen resultados similares para el grupo de estudiantes que tiene

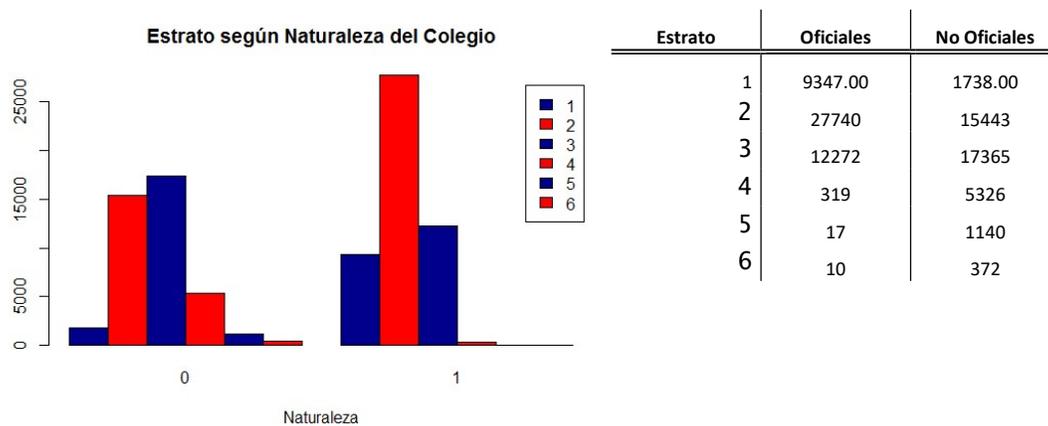
¹³ 77.42% son el resultado de sumar 37.84% de los estudiantes que cuentan con acceso en su hogar a PC e Internet de colegios oficiales más el 39.58% que también cuentan con PC e internet de colegios no oficiales.

acceso a internet en su hogar, con una media de 280.38 y los que no, con una media de 258.91 ($t = -57.044$, $df = 22890$, $p\text{-value} < 2.2e-16$).

d. Estudiantes según el estrato económico

El estrato socioeconómico de los estudiantes nos puede proporcionar herramientas para analizar la existencia de condiciones básicas para poder sacar adelante logros académicos, y para este caso un buen resultado en las pruebas SABER 11.

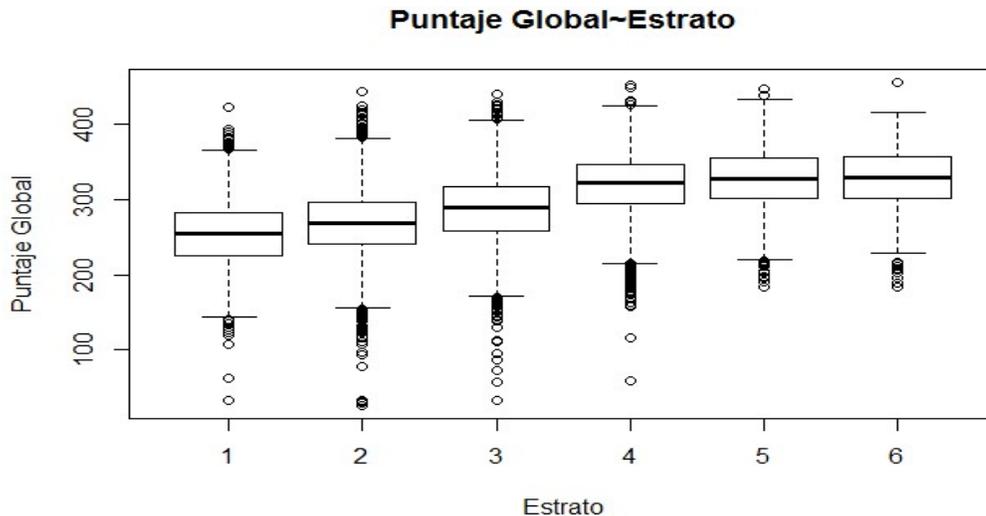
Grafico 7.



Fuente: SABER 11 (2016-2), 2016. Elaboración propia.

Para este estudio, nos brinda la posibilidad de acercarnos y contemplar no solo el panorama socioeconómico bogotano, sino que, al ser contrastado junto con la naturaleza de los colegios (centro de nuestro estudio) podemos obtener una visión sobre cómo se conforman estos grupos de estudiantes de colegios públicos, donde predominan el estrato 1,2 y 3, y el grupo de estudiantes de colegios privados, donde existe una concentración entre el estrato 2 y 3.

Grafico 8.



Fuente: SABER 11 (2016-2), 2016. Elaboración propia.

El gráfico de caja 8 muestra la existencia de una relación directa entre el estrato socioeconómico y los resultados que obtiene el grupo de estudiantes. Podemos observar que a medida que aumenta el estrato socioeconómico del hogar del estudiante la media en los resultados obtenidos en el examen SABER 11 también tiende a aumentar.

e. Estudiantes según el valor de la pensión pagada por el estudiante en el último año

Una de las características que nos permite reconocer la base de datos de los estudiantes que presentaron la prueba SABER 11 es el valor de la pensión pagada por el estudiante en el último año. El cuadro 2 nos muestra los valores posibles que puede tomar la variable pensión en nuestro modelo, del uno al 6, y su correspondiente valor real de pensión.

Cuadro 2.

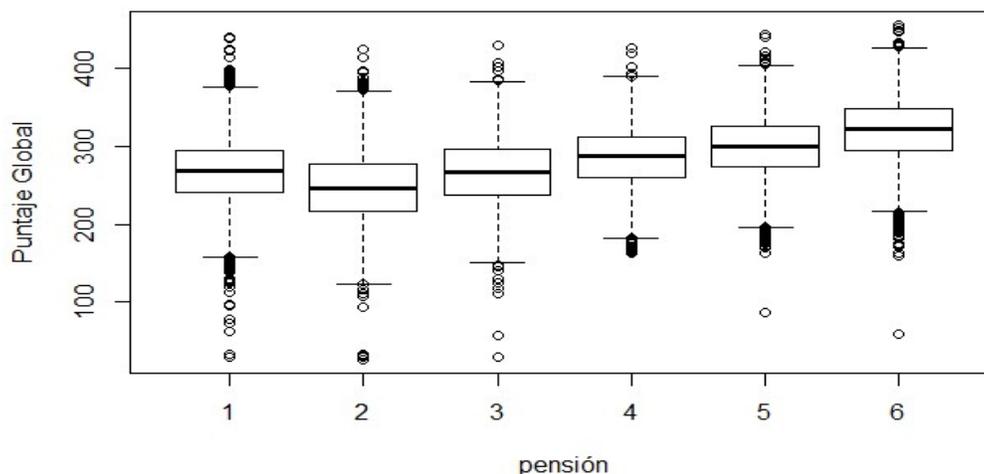
Valores posibles que puede tomar la variable pensión en el modelo	"X" Valor de la pensión pagada por el estudiante en el último año
1	$X = \$0$
2	$\$0 < X < \$87,000$
3	$\$87,000 \leq X < \$120,000$
4	$\$120,000 \leq X < \$150,000$
5	$\$150,000 \leq X < \$250,000$
6	$X > \$250,000$

Fuente: SB11-2016-2-Diccionario_de_variables-V1-0, 2016. Elaboración propia.

Para lo cual realizamos un gráfico de la caja 9 que de igual forma que en el caso del estrato socioeconómico podemos evidenciar la existencia de una relación directa entre el valor de la pensión pagada por el estudiante en el último año y los resultados en las pruebas SABER 11.

Gráfico 9.

Puntaje Global~pensión



Fuente: SABER 11 (2016-2), 2016. Elaboración propia.

Luego de advertir indicios de una relación directa entre pensión y resultados, pasamos a exponer la participación de estudiantes según el valor de la pensión pagada por el estudiante en el último año y la naturaleza del colegio donde curso

su último año. Para así lograr notar que, como era de esperar, la gran mayoría de estudiantes de colegios públicos (48929) no pagan pensión alguna (ver tabla 7), mientras que para los estudiantes de colegios privados está algo más pareja la situación, ya que 20445 estudiantes pagan una pensión de menos \$150.000 mientras que 20939 si logran pagar más de \$150.000 por su pensión mensual.

Tabla 7.

Pensión	Oficiales	No Oficiales
1	48929	3765
2	449	5575
3	64	6293
4	40	4812
5	47	10382
6	176	10557

Fuente: SABER 11 (2016-2), 2016. Elaboración propia.

f. Estudiantes según la jornada de la sede a la que asiste el estudiante

La jornada es una de las características más interesantes de nuestro análisis, puesto que veremos a continuación que es uno de los determinantes que más golpean los resultados de los estudiantes y que a su vez está fuertemente ligada con la naturaleza de los colegios.

Para entender la siguiente tabla es necesario explicar que la variable jornada toma valor de 1 para indicar que el estudiante asiste a una jornada completa u ordinaria y toma valor de 0 para indicar todo caso diferente al anterior, donde podría ser jornada mañana, jornada tarde, jornada noche o jornada sabatina-dominical.

Tabla 8.

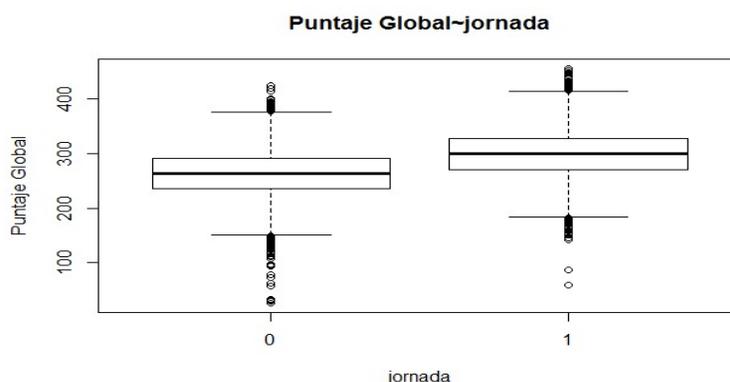
Jornada	Oficiales	No Oficiales
0	46001	10794
1	3704	30590

Fuente: SABER 11 (2016-2), 2016. Elaboración propia.

Como ya se había anticipado anteriormente la jornada está estrechamente relacionada con la naturaleza del colegio, con unas cuantas excepciones podríamos agrupar a casi todos los estudiantes de colegios oficiales en jornadas que no son completas, pues llegan a ser el 92.55%. Mientras que casi tres cuartas partes (73.92%) de los estudiantes de colegios privadas asisten a colegios que otorgan educación con jornada completa.

Ahora bien, a partir del gráfico de caja 10 obtenemos la existencia de una diferencia significativa ($t = -121.56$, $df = 70498$, $p\text{-value} < 2.2e-16$) en los resultados que obtiene el grupo de estudiantes asiste a una jornada completa (con una media de 289.45) con respecto al de grupo de estudiantes que asiste a una jornada parcial (con una media de 263.65), además de que tanto sus máximos como sus mínimos son igualmente diferenciables.

Grafico 10.

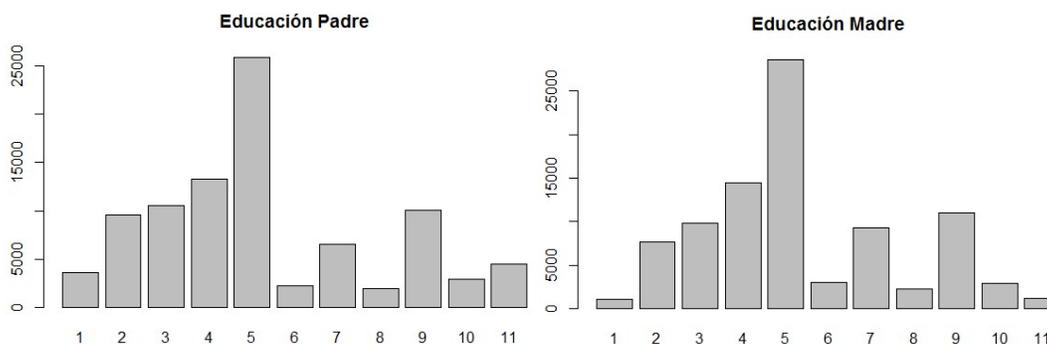


Fuente: SABER 11 (2016-2), 2016. Elaboración propia.

g. Estudiantes según el nivel educativo del padre y de la madre

Para el análisis sobre el nivel educativo del padre y de la madre de los estudiantes que presentaron la prueba nos pudimos encontrar con comportamientos bastante similares. La variable de la formación de los padres va tomando valor de 1 para indicar que el nivel educativo del padre/madre del estudiante es ninguno, 2 para indicar que el nivel educativo del padre/madre del estudiante es primaria incompleta, 3 para indicar que el nivel educativo del padre/madre del estudiante es primaria completa, 4 para indicar que el nivel educativo del padre/madre del estudiante es secundaria (bachillerato) incompleta, 5 para indicar que el nivel educativo del padre/madre del estudiante es secundaria (bachillerato) completa, 6 para indicar que el nivel educativo del padre/madre del estudiante es educación técnica o tecnológica incompleta, 7 para indicar que el nivel educativo del padre/madre del estudiante es educación técnica o tecnológica completa, 8 para indicar que el nivel educativo del padre/madre del estudiante es educación profesional incompleta, 9 para indicar que el nivel educativo del padre/madre del estudiante es educación profesional completa, 10 para indicar que el nivel educativo del padre/madre del estudiante es postgrado y por ultimo 11 para indicar que el estudiante es no sabe cuál es el nivel de formación de su padre/madre.

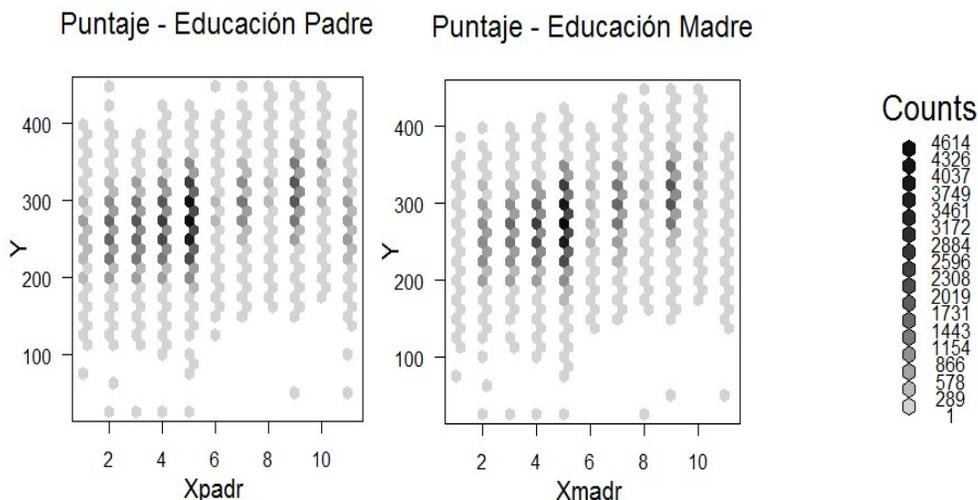
Grafico 11.



Fuente: SABER 11 (2016-2), 2016. Elaboración propia.

Tanto en el gráfico de barras 11 como en el análisis del gráfico de dispersión 12 para el nivel de formación del padre y el de la madre, muestran una concentración en el grupo de estudiantes en que el nivel educativo del padre y de la madre es secundaria (bachillerato) completa.

Grafico 12.

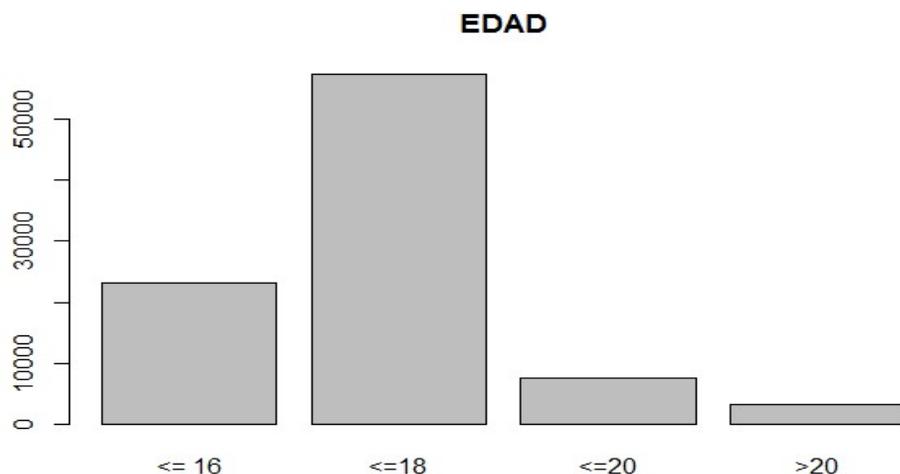


Fuente: SABER 11 (2016-2), 2016. Elaboración propia.

h. Estudiantes según la edad de los estudiantes

Para el análisis de la edad con la cual presentan los estudiantes la prueba SABER 11, se llevó a cabo una agregación en 4 grupos, para poder hacer más consistente el análisis de esta cualidad.

Grafico 13.

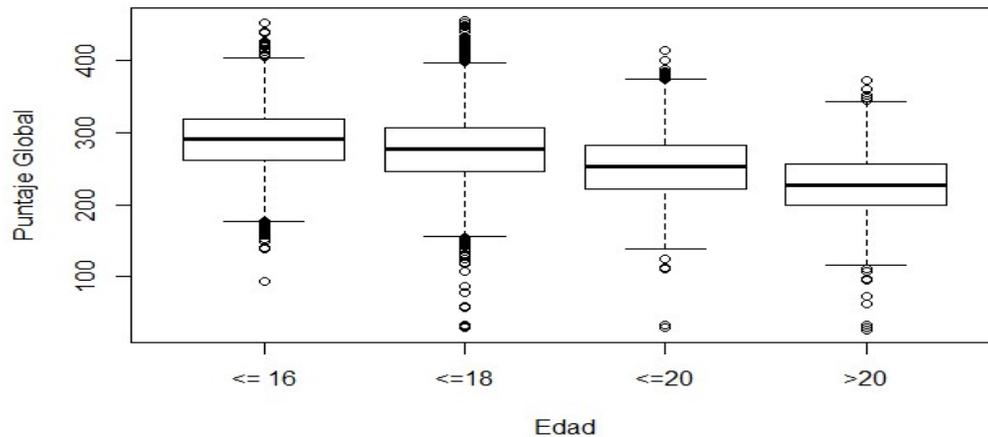


Fuente: SABER 11 (2016-2), 2016. Elaboración propia.

Para el grupo de menores de 16 años hay 23075 estudiantes, entre 16 y 18 encontramos el mayor rubro con 57357, para el grupo de entre 18 y 20 años hay 7504 y para los mayores de 20, el más pequeño de todos, 3153 estudiantes.

Grafico 14.

Puntaje Global~Edad



Fuente: SABER 11 (2016-2), 2016. Elaboración propia.

El grafico de caja 14, muestra que a medida que aumenta la edad disminuye la media de los resultados obtenidos por los estudiantes.

3.2 Modelo multinivel

En esta sección se presenta el procedimiento utilizado en la estimación econométrica de los modelos multinivel. Luego de obtener los coeficientes de regresión en un nivel, se regresan en un siguiente nivel de variables explicativas y así se continua sucesivamente para los niveles restantes. (Gaviria y Castro, 2005).

De este procedimiento se pueden destacar por lo menos tres etapas de modelaje. La primera es el desarrollo del Modelo nulo que no tiene en cuenta ninguna variable diferente a la dependiente, luego un Modelo con solo las variables de ajuste y por último el Modelo final que ya tiene en consideración la parte aleatoria.

En esta metodología encontramos tres conceptos nucleares (Murillo Torrecilla, F. J., 2008) sobre los cuales es necesario profundizar para poder empezar a entender y luego desarrollar el modelo multinivel.

- El coeficiente de correlación intraclase (IC) que se presenta como la medida del grado de dependencia de los individuos dentro de una misma agrupación, por lo que la obtención de un IC muy alto nos da a entender que los individuos dentro de su agrupación son tan distintos entre ellos que no es necesario llevar a cabo dicha agrupación o su agrupación no forma un grupo homogéneo y sus individuos se comportan de forma independiente (Murillo Torrecilla, F. J., 2008). El coeficiente de correlación intraclase se define de la siguiente forma:

$$(1) \quad IC = \frac{(\sigma_u)^2}{(\sigma_u)^2 + (\sigma_e)^2}$$

$(\sigma_u)^2$: Varianza en el nivel de colegios

$(\sigma_e)^2$: Varianza en el nivel de los estudiantes

- Los Coeficientes fijos y los coeficientes aleatorios son uno de los aportes más significativos de esta metodología multinivel, debido a que permite obtener coeficientes de regresión para un primer nivel que luego van a variar en un segundo nivel que abre la posibilidad de que el comportamiento de los grupos pueda desviarse de la solución central del modelo. Esto nos lleva a obtener una parte fija que contiene los coeficientes comunes a todo los contextos y una parte aleatoria que afecta según va cambiando el contexto (Murillo Torrecilla, F. J., 2008).
- La interacción internivel entre las variables medidas según los diferentes niveles que tiene una estructura jerárquica de datos (Murillo Torrecilla, F. J., 2008).

$$(2) \quad y_{[i]j} = \beta_{0j} + \sum_{k=1}^n \beta_{kj} x_{k[i]j} + \varepsilon_{[i]j}$$
$$\beta_{0j} = \beta_0 + u_{0j}, \beta_{1j} = \beta_1 + u_{1j}, \dots, \beta_{nj} = \beta_n + u_{nj}$$

$y_{[i]j}$ es el puntaje global en la prueba SABER 11
 β_{0j} es el promedio de y del colegio j – ésimo
 u_{1j} es el efecto aleatorio asociado al colegio j – ésimo
 $\varepsilon_{[i]j}$ es el error del nivel

Para poder implementar la metodología teórica sugerida por Murillo Torrecilla, F. (2008) a partir de la base de datos de los estudiantes del SABER 11 se empleó la metodología empírica propuesta por Torres-Reyna, O. (2013) para modelación multinivel, junto con el trabajo de Fox, J. y Weisberg, S. (2015) para la modelación de efectos mixtos en el paquete econométrico (estadístico) R. Para este estudio se utilizó íntegramente el programa estadístico R, que cuenta con un ambiente de programación a la disposición de procedimientos estadísticos en varias disciplinas y con por lo menos tres características diferenciadoras de sus competidores en el mercado que son: su gran capacidad de procesamiento, su mayor control sobre los procesos estadísticos aplicados a los datos y sobre todo su gratuidad¹⁴.

La metodología multinivel ha sido aplicada de forma tal en la que se estima el efecto plantel en Bogotá en tres casos de los cuales el primero tiene en consideración solo estudiantes de colegios oficiales, el segundo solo considera estudiantes de colegios no oficiales y el ultimo, para todos los estudiantes que presentaron la prueba SABER 11 indiscriminadamente. Para posteriormente poder compararlos y concluir si es significativamente determinante la naturaleza de los colegios en los resultados que obtienen estos estudiantes en la prueba SABER 11 en Bogotá, para el 2016-2.

3.2.1 Modelo Nulo (modelo I)

Este primer modelo es un modelo vacío, puesto que no contiene ninguna de las variables predictoras solo se modela con la variable respuesta, que para nuestros tres casos será el puntaje global obtenido por los estudiantes “Y”. Este

¹⁴ “R studio” ofrece tanto opciones gratuitas como también opciones de pago según las necesidades comerciales de las empresas que lo soliciten.

modelo nulo se realiza con la intención de poder analizar el aporte a la varianza de los demás modelos que vayamos obteniendo (Murillo Torrecilla, F. J., 2008).

$$(3) \quad y_{[i]j} = \beta_{0j} + \varepsilon_{[i]j}$$

Aquí, $y_{[i]j}$ es el puntaje global en la prueba SABER 11, que obtiene el estudiante i en el colegio j , y donde $\varepsilon_{[i]j}$ es el error que se distribuye de forma normal con una varianza σ^2 constante¹⁵. Por lo tanto, β_{0j} es entendida como el promedio de y del colegio j -ésimo. Para este modelo nulo también se estiman los efectos en los dos niveles, donde el gran promedio para la población de estudiantes que presentaron la prueba sería $\beta_{0j} = u_{yj}$, con $\beta_{0j} = \beta_0 + u_{0j}$ y donde u_{0j} es el efecto aleatorio asociado al colegio j -ésimo, con media 0 y una varianza τ_{00} .

Tabla 9.

Modelo Nulo (modelo I)

	Oficial			No Oficial			Ambas		
Parte Fija									
	Estimate	Std. Error	t value	Estimate	Std. Error	t value	Estimate	Std. Error	t value
(Intercept)	263.8105	0.6868	384.1	284.577	1.091	260.9	275.3314	0.7335	375.4
Parte Aleatoria									
Groups	Variance	Std.Dev.		Variance	Std.Dev.		Variance	Std.Dev.	
Xcolicfe(Intercept)	280.1	16.74		944.1	30.73		757.3	27.52	
Residual	1335.4	36.54		1192.1	34.53		1270.5	35.64	

Fuente: SABER 11 (2016-2), 2016. Elaboración propia.

Una vez estimado el modelo nulo para los tres casos, podemos realizar el cálculo de la correlación Intraclase que nos muestra el grado de dependencia de los estudiantes por compartir un mismo colegio. Como se había explicado anteriormente en el caso de que la correlación intraclase tendiera a 0, no tendría sentido agrupar en las muestras por colegios debido a que cada estudiante dentro de un mismo colegio sería tan diferente como entre estudiantes de distinto colegio.¹⁶

Tabla 10.

Oficial	No oficial	Ambos
$IC = 0.174$	$IC = 0.442$	$IC = 0.374$

Fuente: SABER 11 (2016-2), 2016. Elaboración propia.

La medición del IC para los tres casos, nos indica que si es pertinente agrupar por colegio y de esta forma, procedemos con la siguiente etapa del modelo. La tabla 9 nos permite corroborar los resultados que se obtuvieron previamente con

¹⁵ Las pruebas que se le aplican al modelo para contrastar estos supuestos se muestran en el apartado 3.2.4 - Verificación del cumplimiento de los supuestos.

¹⁶ Murillo Torrecilla, F. J. (2008). Los modelos multinivel como herramienta para la investigación educativa. pp.48

el test de diferencias de medias. Es decir, que en promedio los estudiantes de los colegios oficiales alcanzan un puntaje global mucho mas bajo comparado con el puntaje que logran los estudiantes de colegios privados.

3.2.2 Modelo con las variables de ajuste (modelo II)

En esta segunda etapa de estimación se procede a incluir las de variables de ajuste del modelo que están descritas en la Tabla de Variables (ver Anexo 19).

La variable de naturaleza (Xcolnatu) se incorpora como variable de entrada solo para el caso que considera estudiantes de ambas naturalezas (oficial y no oficial). Ya que para el caso “oficial” y el “no oficial” esta variable no tiene ningun efecto en la estimacion al ser una constante que toma valor de 1 en caso oficial y 0 en el caso de privado.

Tabla 11.

Modelo con variables de ajuste (modelo II)

	Oficial			No Oficial			Ambas		
Parte Fija									
	Estimate	Std. Error	t value	Estimate	Std. Error	t value	Estimate	Std. Error	t value
(Intercept)	235.11	1.11	212.13	220.49	1.79	123.01	219.76	1.52	144.83
Xgene	10.55	0.33	31.97	8.16	0.36	22.44	9.51	0.24	38.89
Xpadr	0.98	0.08	12.77	1.15	0.09	13.32	1.06	0.06	18.49
Xmadr	2.06	0.10	21.08	1.76	0.10	17.42	1.94	0.07	27.58
Xestr	3.22	0.31	10.24	0.93	0.30	3.05	2.14	0.22	9.74
Xinte	2.38	0.41	5.86				2.03	0.34	6.02
Xingr	1.65	0.20	8.10	2.01	0.17	11.79	1.87	0.13	14.26
Xcoljorn	18.26	2.61	7.00	31.55	1.73	18.22	29.66	1.41	21.11
Xcolnatu							13.06	1.45	9.01
Xcolpens	-2.54	0.49	-5.21	2.35	0.26	9.16	1.48	0.22	6.62
Parte Aleatoria									
Groups	Variance	Std.Dev.		Variance	Std.Dev.		Variance	Std.Dev.	
Xcolicfe (Intercept)	191.10	13.82		389.20	19.73		304.50	17.45	
Residual	1271.40	35.66		1148.70	33.89		1217.50	34.89	

Fuente: SABER 11 (2016-2), 2016. Elaboración propia.

Como resultado interesante de la inclusion de las variables de ajustes, la tabla 11 muestra que la media del resultado global que obtienen los estudiantes de colegios oficiales ha pasado a ser mayor que el de los estudiantes de colegios privados (235.11>22.49), asimismo de la magnitud del efecto que tiene tanto el genero del estudiante como la jornada del colegio sobre el resultado global.

Por medio de la prueba t ($t_{cal} = \frac{\widehat{\beta}_i}{\sigma_{\widehat{\beta}_i}}$) de significancia individual se determina si los coeficientes estimados tienen una aportación significativa ¹⁷ para el modelo¹⁸ y delimitarlo solo a las variables que realmente aporten al análisis del fenómeno de nuestro interés. Una vez realizada esta estrategia estadística se calculó nuevamente el IC para el modelo corregido, cuyos resultados se presentan en la tabla 12.

Tabla 12.

Oficial	No oficial	Ambos
IC = 0.13	IC = 0.25	IC = 0.20

Fuente: SABER 11 (2016-2), 2016. Elaboración propia.

Aunque el IC para todos los casos presentó una disminución en esta nueva etapa el valor obtenido para cada uno de los casos sigue señalando que aún es pertinente la agrupación por colegio.

Desde esta etapa, se empieza a usar la prueba de razón de verosimilitud (lrtest: likelihood-ratio test) que nos permite comparar entre dos modelos por máxima verosimilitud y así probar si son significativamente diferentes¹⁹. En este caso la prueba se aplicó para estimar en cada uno de los tres casos (Ver tabla 13) entre el modelo nulo anteriormente estimado y el modelo II de esta segunda etapa.

Tabla 13.

Oficial	No oficial	Ambos
$\Pr(> \chi^2) = 2.2e - 16 < 0.05$	$\Pr(> \chi^2) = 2.2e - 16 < 0.05$	$\Pr(> \chi^2) = 2.2e - 16 < 0.05$
Se rechaza H_0	Se rechaza H_0	Se rechaza H_0

Fuente: SABER 11 (2016-2), 2016. Elaboración propia.

Para los tres casos se rechaza la hipótesis nula debido y así mismo podemos concluir que si hay diferencia estadísticamente significativa entre los modelos nulos y los modelos de la segunda etapa de estimación (con variables de ajuste) y que los coeficientes aleatorios del modelo II se ajustan mejor.

3.2.3 Modelo Final (modelo III)

En esta etapa del modelo se integran la variables de la parte aleatoria, en las cuales se empezará a variar su pendiente y así poder estimar los efectos fijos y aleatorios del modelo conjuntamente.

¹⁷ $H_0: \widehat{\beta}_i = 0$; $H_1: \widehat{\beta}_i \neq 0$; Si $|t_{cal}| > |1.96|$ entonces se rechaza H_0 y por lo tanto, $\widehat{\beta}_i \neq 0$

¹⁸ Tabla 11 es resultado de la corrección del modelo, luego de encontrar que para el caso de los colegios no oficiales Xinte (Internet) no tenía un aporte significativo al modelo. (Ver anexo 9.)

¹⁹ H_0 : No hay diferencias significantes entre los dos modelos y H_1 : Hay diferencias significantes entre los dos modelos

Tabla 14.

Modelo final

	Oficial			No Oficial			Ambas		
Parte Fija									
	Estimate	Std. Error	t value	Estimate	Std. Error	t value	Estimate	Std. Error	t value
(Intercept)	235.01	1.12	210.77	220.68	1.79	123.55	220.90	1.65	133.74
Xgene	10.68	0.36	30.04	7.96	0.40	19.81	9.53	0.24	39.00
Xpadr	0.98	0.08	12.77	1.15	0.09	13.31	1.05	0.06	18.41
Xmadr	2.06	0.10	21.09	1.77	0.10	17.44	1.94	0.07	27.58
Xestr	3.23	0.31	10.27	0.93	0.30	3.05	2.15	0.22	9.78
Xinte	2.38	0.41	5.86				2.12	0.34	6.28
Xingr	1.65	0.20	8.10	2.02	0.17	11.82	1.82	0.13	13.83
Xcoljorn	18.42	2.60	7.09	31.47	1.72	18.31	29.79	1.56	19.04
Xcolnatu							12.64	1.54	8.20
Xcolpens	-2.54	0.49	-5.20	2.37	0.26	9.24	1.17	0.23	5.12
Parte Aleatoria									
Groups Name	Variance	Std.Dev.		Variance	Std.Dev.		Variance	Std.Dev.	
Xcolicfe (Intercept)	200.41	14.16		405.06	20.13		400.00	20.00	
Xcolnatu							550.10	23.46	
Xcoljorn				445.21	21.10		454.40	21.32	
Xgene	8.64	2.94		13.86	3.72				
Residual	1269.33	35.63		1145.91	33.85		1217.10	34.89	

Fuente: SABER 11 (2016-2), 2016. Elaboración propia.

Para el caso “oficial” y el “no oficial” se incluyen en la parte aleatoria las variables relacionadas con las características del colegio donde estudió su último año escolar cada alumno, con la finalidad de estimar el efecto del plantel sobre los estudiantes y para el caso que considera tanto estudiantes de ambas naturalezas (oficial y no oficial) se tiene la variable de naturaleza (Xcolnatu) ya no solo como variable de entrada sino también como variable de contexto.

En este modelo final²⁰ ya podemos empezar a notar cambios en los efectos que tiene cada determinante escogido sobre el puntaje global en las pruebas SABER 11.

En efecto, para el caso de los estudiantes de colegios oficiales encontramos que la variable del género del estudiante tiene un aporte significativo alto sobre sus resultados en las pruebas, pero podemos apreciar que el impacto más significativamente alto sobre los resultados de la prueba SABER 11 lo tiene la variable jornada, que destaca sobre el impacto que pueden generar las demás en este caso. Para el caso que solo considera los estudiantes de colegios privados observamos que el impacto más alto sigue siendo el de la jornada, que

²⁰ La Tabla 14 es resultado de la corrección del modelo final, luego de encontrar que para el caso de los colegios oficiales (XColjorn) la jornada no tenía un aporte significativo al modelo como variable de contexto. (Ver anexo 4 y 5)

en este caso es aún mayor y que la variable genero sigue generando un impacto relativamente mayor que las demás variables²¹. Para el caso general (que considera tanto estudiantes de colegios públicos como estudiantes de colegios privados) al incluir la variable que discrimina la naturaleza del colegio podemos observar el gran aporte que tienen las características del plantel sobre los resultados en las pruebas SABER 11.

Aplicando nuevamente la prueba t de significancia individual determinamos que todos los coeficientes estimados para los tres casos tienen una aportación significativa y se decide mantenerlos en el modelo.

Tabla 15.

Oficial	No oficial	Ambos
$IC = 0.14$	$IC = 0.43$	$IC = 0.54$

Fuente: SABER 11 (2016-2), 2016. Elaboración propia.

Ahora, con los cálculos realizados del IC como se debe realizar en cada avance de etapa, el coeficiente de correlación intraclase (ver tabla 15) nos permite notar que al agregar encada caso las variables de contexto el IC para todos los casos presentó un aumento con respecto al valor de los IC obtenido para cada uno de los casos en el Modelo II (anteriormente estimado solo con variables de ajuste). Por lo que podemos llegar a la conclusión que con la inclusión de las variables de contexto es pertinente la agrupación por colegio.

Tabla 16.

Oficial	No oficial	Ambos
$\Pr(> \chi^2) = 2.2e - 16 < 0.05$	$\Pr(> \chi^2) = 2.2e - 16 < 0.05$	$\Pr(> \chi^2) = 2.2e - 16 < 0.05$
Se rechaza H_0	Se rechaza H_0	Se rechaza H_0

Fuente: SABER 11 (2016-2), 2016. Elaboración propia.

Al usar la prueba de razón de verosimilitud comparamos por máxima verosimilitud los dos modelos últimos estimados para cada uno de los tres casos (Ver tabla 16) entre el modelo nulo anteriormente estimado y el modelo II de esta segunda etapa podemos al rechazar la hipótesis nula (para todos los casos) y concluir que los coeficientes aleatorios del modelo final se ajustan mejor que los del modelo II.

3.2.4 Verificación del cumplimiento de los supuestos

Para poder culminar correctamente la aplicación de la metodología de Los Modelos Multinivel, es necesario realizar unas mínimas pruebas que nos permitan contrastar los supuestos que hacen parte del maderamen teórico y empírico anteriormente expuesto y desarrollado. El cumplimiento de estos supuestos cuestiona el valor descriptivo de las estimaciones obtenidas a lo largo

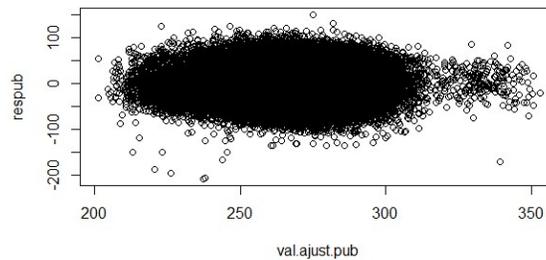
²¹ Un resultado interesante de la comparativa entre los modelos para el caso oficial y el no oficial es que la variable pensión (XColpens) impacta negativamente el resultado global (Y) obtenido por de los estudiantes de colegios oficiales y, por el contrario, el impacto de la pensión sobre el resultado de los estudiantes de colegios no oficiales si tiene una dirección positiva.

de todo nuestro trabajo. A continuación, se presentan las pruebas realizadas sobre los errores de los modelos estimados para cada caso, realizadas sobre los residuos (Murillo Torrecilla, F., 2008).

El primer supuesto a medir es la homocedasticidad del error, por lo que se pretende estimar que el error tiene media nula y varianza constante. El segundo supuesto a contrastar es la independencia de los errores y junto al tercer supuesto, la normalidad del error, suelen ser abarcados por medio de análisis gráficos (Murillo Torrecilla, F., 2008).

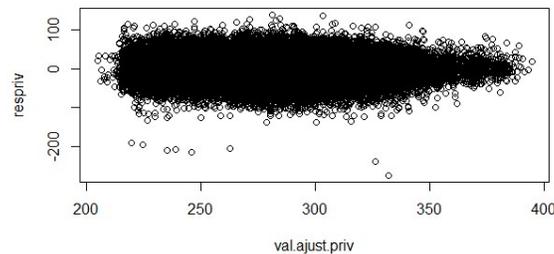
Para el segundo supuesto, la independencia del error, se realiza un gráfico de dispersión que nos permite observar la relación entre los errores y así poder verificar si se aprecia alguna relación entre ellos.

Grafico 15. Caso de colegios oficiales.



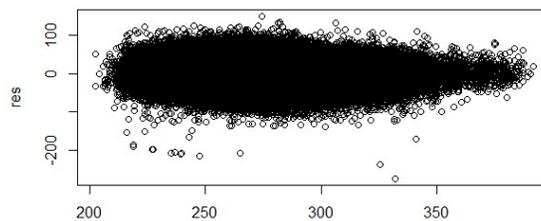
Fuente: SABER 11 (2016-2), 2016. Elaboración propia.

Grafico 16. Caso de colegios privados.



Fuente: SABER 11 (2016-2), 2016. Elaboración propia.

Grafico 17. Caso de colegios de ambas naturalezas.

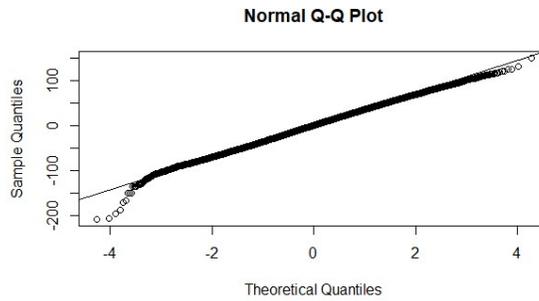


Fuente: SABER 11 (2016-2), 2016. Elaboración propia.

Mediante los gráficos 15, 16 y 17 para cada caso, respectivamente, podemos apreciar la relación entre el valor previsto y los residuales muestran un comportamiento independiente del error.

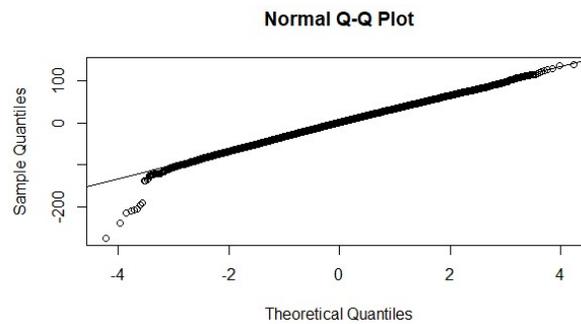
Para el tercero, la normalidad del error, por medio de un gráfico QQ (QQ-plot). Donde el eje de y representa los residuos escolares estandarizados mientras que el eje x representa el valor esperado respectivo de la distribución Normal.

Grafico 18. Caso de colegios públicos



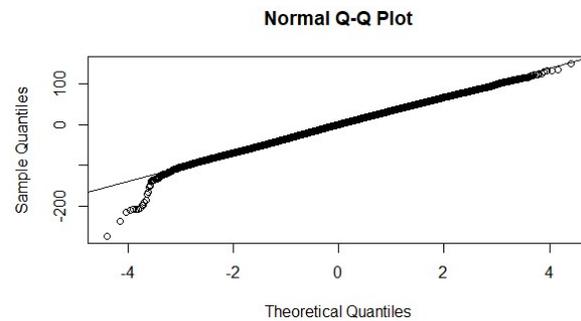
Fuente: SABER 11 (2016-2), 2016. Elaboración propia.

Grafico 19. Caso de colegios privados



Fuente: SABER 11 (2016-2), 2016. Elaboración propia.

Grafico 20. Caso de colegios de ambas naturalezas.



Fuente: SABER 11 (2016-2), 2016. Elaboración propia.

Mediante los gráficos QQ 18, 19 y 20 se demuestra que para los modelos estimados en todos los errores se distribuyen normalmente.

3.2.5 El poder explicativo del Modelo

Por otro lado, para evaluar la calidad del modelo final, se verifica por medio de la estimación del coeficiente de determinación R^2 , para cada uno de los tres casos respectivamente, cuanta varianza del nivel de los estudiantes y del nivel

de los colegios es explicada por nuestro modelo y así entender su capacidad explicativa (Murillo Torrecilla, F., 2008).

“... Coeficiente de determinación R^2 (Longford, 1993). Si el intercepto apenas tiene varianza aleatoria la varianza total será la suma de las varianzas de los niveles 1 y 2 ($\sigma_e^2 + \sigma_{u0}^2$). De esta forma, podremos estimar el coeficiente de determinación total R^2 ... con la siguiente

$$\text{fórmula: } R^2 = 1 - \frac{\text{Var}(final)_n}{\text{Var}(nulo)} \text{ (Murillo Torrecilla, F., 2008)}$$

Por consiguiente, tomamos la varianza residual del modelo final (ver tabla 14) para el caso público, el privado y el que considera estudiantes de colegios de ambas naturalezas y la varianza del modelo nulo (Ver tabla 9) para poder calcular el coeficiente de determinación R^2 .

Tabla 17.

Oficial	No oficial	Ambos
$R^2 = 1 - \frac{1269.33}{1335.4} = 0.0494$	$R^2 = 1 - \frac{1145.91}{1192.1} = 0.0387$	$R^2 = 1 - \frac{1217.1}{1270.5} = 0.042$

Fuente: SABER 11 (2016-2), 2016. Elaboración propia.

La tabla 17 muestra los cálculos del coeficiente de determinación R^2 permitiendo apreciar que los tres modelos finales estimados tienen un gran poder explicativo.

Para estimar el porcentaje de la varianza dentro del modelo, tenemos que apreciar la parte aleatoria de los modelos estimados. La parte aleatoria del modelo final muestra la varianza explicada por la inclusión de las variables de contexto por cada nivel y junto con la varianza no explicada del modelo (parte aleatoria del modelo nulo) podemos estimar el porcentaje de la varianza explicada entre los estudiantes, entre los colegios y el porcentaje de varianza total explicada respectivamente (Hernández Angulo, O. 2015). Esta relación se representa en las fórmulas 4, 5 y 6.

$$(4) \quad \% \text{ de varianza explida entre los estudiantes} = \frac{\varepsilon_{[ij]}(nulo) - \varepsilon_{[ij]}(final)}{\varepsilon_{[ij]}(final)}$$

$$(5) \quad \% \text{ de varianza explida entre los colegios} = \frac{u_{yj}(nulo) - u_{yj}(final)}{u_{yj}(final)}$$

$$(6) \quad \% \text{ de varianza total explida} = \frac{(\varepsilon_{[ij]}(nulo) + u_{yj}(nulo)) - (u_{yj}(final) + \varepsilon_{[ij]}(final))}{(u_{yj}(final) + \varepsilon_{[ij]}(final))}$$

Tabla 18.

% Varianza explicada...	Oficial	No oficial	Ambos
Entre estudiantes	5.21%	4.03%	4.39%
Entre colegios	33.9%	9.25%	46.08%
Total	9.27%	6.27%	22.65%

Fuente: SABER 11 (2016-2), 2016. Elaboración propia.

La tabla 18 (para el cálculo ver Anexo 27 y 28) muestra el que el porcentaje de la varianza explicada entre los estudiantes para el caso de estudiantes de colegios públicos es del 33.9%, para el caso de colegios privados es del 9.25% y para el caso que considera estudiantes de colegios de ambas naturalezas es

del 46.08%. Así mismo, muestra que la varianza explicada entre colegios para el caso público es del 5.21%, para el caso privado es del 4.03% y para el caso que considera estudiantes de colegios de ambas naturalezas es del 4.39%. Por su parte, el porcentaje de varianza total explicada para el modelo del caso oficial es el 9.27%, para el caso de naturaleza no oficial es de 6.27% y para el caso considera ambas naturalezas es 22.65%.

4 Conclusiones

Mediante la estrategia de análisis estadístico descriptivo pudimos prever resultados que luego fueron estimados en el desarrollo de la metodología de los modelos multinivel, donde la diferencia entre las medias estimadas por género, jornada, estrato económico, el acceso en el hogar a internet y un computador ya mostraban un significativo distanciamiento entre los grupos de estudiantes que asisten a un colegio de naturaleza oficial y los que lo hacen en colegios de naturaleza privada, de manera específica para el caso de Bogotá.

Particularmente, en este trabajo se realizó también un análisis sobre la base de datos de los establecimientos académicos que nos mostró como los resultados que van obteniendo los estudiantes le van dando cuerpo al contexto escolar. Ya que cada colegio realiza un aporte significativo a los estudiantes según sus características propias como la jornada, su planta docente, la planta física y su metodología entre otras singularidades del colegio, que tienen un efecto diferenciador sobre sus estudiantes y consecuentemente ellos con su rendimiento en las pruebas de estado terminan retribuyendo dicho aporte en la clasificación de planteles. De esta clasificación de planteles para el caso bogotano, encontramos como resultado destacado que una gran parte de los planteles educativos oficiales de Bogotá lograron situarse en la categoría media (B), mientras que en su gran mayoría los colegios que sí lograron ubicarse en la más alta categoría (A+) son de naturaleza privada.

De igual forma, al comparar los tres casos de estudio mediante la estimación del modelo multinivel pudimos cuantificar cual es el efecto, tanto la magnitud como la dirección, que tienen estos determinantes sobre los resultados en la prueba SABER 11 que obtienen los estudiantes. Los resultados muestran el gran impacto que tienen las variables de contexto escogidas como el género del estudiante, la jornada del establecimiento educativo y su naturaleza. La jornada es uno de los determinantes más significativos estimados en este trabajo, por lo que, es necesario destacar que está fuertemente relacionada con la naturaleza de los colegios (ver anexos 6, 12 y 17) y como muestra la tabla 8, la mayoría de los estudiantes matriculados en colegios oficiales asisten a un tipo de jornada diferente a la jornada única o completa, mientras que la mayoría de los estudiantes matriculados en colegios privados si asisten a jornadas completas de estudio.

Respecto al poder explicativo del modelo final se demostró que este explica el 9% de la varianza total para el caso de los estudiantes de colegios oficiales, el 34% de la varianza entre colegios y apenas el 5% de la varianza entre estudiantes. Para el caso de los estudiantes de colegios no oficiales el modelo explica el 6%, apenas explica el 9% de la varianza entre colegios y el 5% de la varianza entre los estudiantes. Sin embargo, para el caso que considera los estudiantes de colegios de ambas naturalezas el modelo logra explicar el 23% de la varianza total, el 46% de la varianza entre colegios y el 4% de la varianza entre estudiantes. Por consiguiente, los factores de contexto del colegio en el modelo logran explicar razonablemente la varianza total obtenida en los modelos

estimados para cada caso y corroboramos la hipótesis teórica propuesta en este trabajo de grado.

En la estimación del modelo final, el coeficiente de correlación intraclase (ver tabla 15), no solo nos mostró que, si es necesaria la agrupación por colegios para poder evaluar el efecto de los determinantes sobre el rendimiento académico de los estudiantes, sino también que al no desconocer el efecto del plantel y sus características que dan contexto al rendimiento de los estudiantes en las pruebas SABER 11 se obtiene una mayor capacidad explicativa en la estimación de estos determinantes.

Finalmente, este trabajo nos brinda una perspectiva sobre la educación pública en Bogotá, donde se han venido realizando grandes esfuerzos en la cobertura de la educación pública por medio de políticas distritales que aumentaron la cantidad de planteles educativos, la planta de docentes y una mejor dotación para los estudiantes en la alimentación, útiles escolares y acceso a medios educativos como libros académicos, bibliotecas, laboratorios, computadores y acceso a internet (Pérez Martínez, A., 2016). Sin embargo, aún hay esfuerzos pendientes por realizar para acotar aún más las diferencias en los resultados académicos que puedan ser generadas por diferencias socioeconómicas y las también relacionadas con el efecto plantel.

Bibliografía

1. Alexander, L. y Simmons, J. (1975). The Determinants of School Achievement in Developing Countries: the educational production function. Staff working paper No. 201. Washington: International Bank for reconstruction and development.
2. Becker, G. (1962). Investment in human capital: A theoretical analysis. Vol. 70, No. 5, Part 2: Investment in human beings (Oct., 1962), pp. 9-49. The University of Chicago Press.
3. Becker, G., Murphy, K. y Tamura, R. (1990) Human Capital, Fertility, and Economic. Journal of Political Economy, The University of Chicago. 1990, vol. 98, no. 5, pt. 2
4. Cardona, M., Montes, I., Vásquez, J., Villegas, M. y Brito, T. (2007). Capital Humano: Una mirada desde la educación y la experiencia laboral. Semillero de Investigación en Economía de EAFIT, Grupo de Estudios Sectoriales y Territoriales. Medellín, Col
5. Castaño, E. (1998). El efecto colegio sobre la variabilidad del rendimiento en matemáticas. Lecturas de Economía. Universidad de Antioquia
6. Cavieres, E. (2014). La calidad de la educación como parte del problema educación escolar y desigualdad en Chile. Revista Brasileira de Educação. v. 19 n. 59
7. Chica, S., Galvis, D. y Ramirez A. (2010) Determinantes del rendimiento académico en Colombia: pruebas ICFES Saber 11º, 2009. EAFIT
8. Claro, M. (2011). El papel de las tecnologías de la información y las comunicaciones en la educación inclusiva. Naciones Unidas y CEPAL
9. Correa, J. (2004). Determinantes del rendimiento educativo de los estudiantes de secundaria en Cali: un análisis multinivel. Revista Sociedad y Economía, núm. 6, abril, 2004, pp. 81-105. Universidad del Valle Cali, Colombia
10. Cox, C., Schiefelbein, E., Lemaitre, M., Hopenhayn, M. y Himmel, E. (1995). Calidad y equidad de la educación media en Chile: Rezagos estructurales y criterios emergentes. Series políticas sociales 8. CEPAL.
11. Elias, H., Syed, S., Roslan, S. y Mohd, S. (2011) Motivational Predictors of Academic Performance in End Year Examination. Procedia - Social and Behavioral Sciences 29 (2011) 1179 – 1188. International Conference on Education and Educational Psychology
12. Farooq, M., Chaudhry, A., Shafiq, M. y Berhanu, G. (2011) Factors affecting students' quality of academic performance: a case of secondary school level. Journal of Quality and Technology Management. Volume VII, Issue II, December, 2011, Page 01 - 14
13. Fox, J. y Weisberg, S. (2015). Mixed-Effects Models in R. An Appendix to An R Companion to Applied Regression. 2da Ed. Recuperado de: <https://socialsciences.mcmaster.ca/jfox/Books/Companion/appendix/Appendix-Mixed-Models.pdf>
14. Gamboa, L. y Krüger, N. (2016) ¿Existen diferencias en América Latina en el aporte de la educación preescolar al logro educativo futuro?: pisa 2009-2012. Revista CEPAL 118. Abril 2016
15. Gandhi, M. (Sin fecha). "Learn As If You Were to Live Forever", Political Poster Print, Amazon. Recuperado de <https://www.amazon.com/Mahatma-Gandhi-Forever-Political-36-Inch/dp/B000Y30IIS>
16. Gaviria, A. y Barrientos, J. (2001). Calidad de la educación y rendimiento académico en Bogotá. Coyuntura Social, Fedesarrollo, (24), 111-126.
17. Gaviria, J. y Castro, M. (2005). "Modelos jerárquicos lineales", Cuadernos de Estadística 29. Madrid, Editorial La Muralla S.A.
18. Gelman, A. y Hill, J. (2006). Data analysis using regression and multilevel/hierarchical models. Cambridge: Cambridge University Press.

19. Gómez Silva, Carlos Alberto. (2016). Clasificación de colegios según las pruebas Saber 11 del ICFES: un análisis usando modelos marginales (MM). *Sociedad y Economía*, (30), 69-89. Recuperado de http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=
20. Gonzalez, A. (2013) *Sociedad civil y normatividad. La teoría social de David Hume*. Dykinson; Edición: 1
21. Hanushek, E. y Wößmann, L. (2007). *The Role of Education Quality in Economic Growth*. Documento de Trabajo de Investigación de Políticas, Banco Mundial, Washington, D.C.
22. Hernández Angulo, O. (2015). *Determinantes del rendimiento académico en la educación media de Cundinamarca*. [Monografía]. Recuperado de <http://repositorio.escuelaing.edu.co/handle/001/349>
23. Instituto Colombiano para la Evaluación de la Educación – ICFES-. *Bases de datos pruebas Saber (2017) Clasificación de establecimientos y sedes SABER 11*.
24. Instituto Colombiano para la Evaluación de la Educación – ICFES-. *Bases de datos pruebas Saber (2017) Diccionario Clasificación Planteles SABER 11, 2016*. Recuperado de <ftp://ftp.icfes.gov.co/SABER11/SB11-CLASIFICACION-PLANTELES/Diccionario%20Clasificac>
25. Instituto Colombiano para la Evaluación de la Educación – ICFES-. *Bases de datos pruebas Saber (2017) Guía Clasificación Planteles SABER 11, 2012*. Recuperado de <ftp://ftp.icfes.gov.co/SABER11/SB11-CLASIFICACION-PLANTELES/Guia%20Clasificacion%20Plantele>
26. Instituto Colombiano para la Evaluación de la Educación – ICFES-. *Bases de datos pruebas Saber (2017) SB11-2016-2-Diccionario_de_variables-V1-0, SABER11*. Recuperado de ftp://ftp.icfes.gov.co/SABER11/SB11-BASES%20DE%20DATOS/SB11-2016-2-Diccionario_de_va
27. Instituto Colombiano para la Evaluación de la Educación – ICFES-. *Bases de datos pruebas Saber (2017) SB11-20162-RGSTRO-CLFCCN-V1-0, SABER 11*. Recuperado de <ftp://ftp.icfes.gov.co/SABER11/SB11-BASES%20DE%20DATOS/SB11-20162-RGSTRO-CLFCCN-V1-0.zip>
28. Klevmarken, A. y Quigley, J. (1976) *Age, Experience, Earnings, and Investments in Human Capital*. *Journal of Political Economy*, The University of Chicago. 1976, vol. 84, no. 1
29. Lepp, A., Barkley, J., y Karpinski, A. (2015) *The Relationship Between Cell Phone Use and Academic Performance in a Sample of U.S. College Students*. January-March 2015: 1–9
30. Mincer, J. (1975). *Education, experience, and the distribution of earnings and employment: An overview*. En F. Thomas Juster (ed.), *Education, Income, and Human Behavior* (pp. 71 - 93). Cambridge, MA, USA: NBER
31. Mineduccion.gov.co. (2014). *El ICFES cambia la clasificación de resultados - Centro Virtual de Noticias de Educación*. Mineduación. Recuperado de <http://www.mineduccion.gov.co/cvn/1665/w3-article-343956.html>
32. Murillo Torrecilla, F. J. (2008). *Los modelos multinivel como herramienta para la investigación educativa*. *Magis, Revista Internacional de Investigación en Educación*, 1, 45-62.
33. Nuñez, J. (2007) *Motivación, aprendizaje y rendimiento académico*. *Actas do X Congresso Internacional GalegoPortuguês de Psicopedagogia*. Braga: Universidade do Minho, 2009 ISBN- 978-972-8746-71-1
34. Palos, R., Munteanu, A., Costea, I. y Macsinga, I. (2011) *Motivational and cognitive variables with impact on academic performance Preliminary study*. *Procedia Social and Behavioral Sciences* 15 (2011) 138–142

35. Pérez Martínez, A. (2016). Educación en Bogotá: lo que falta. Razón Publica. Recuperado de: <https://www.razonpublica.com/index.php/economia-y-sociedad/9258-educacion-en-bogota-lo-que-falta>
36. Rios, M. (2002) Fundamentación ética del mercantilismo. Bernard Mandeville: la paradoja del vicio en la sociedad. Publicaciones Cruz O., S.A-CUM. Centro Universitario México. Division de estudios superiores.
37. Rios, M. (2007) Bernard Mandeville: La ética del mercado y la desigualdad social como base del progreso moderno. EN-CLAVES del pensamiento, vol. I, núm. 1, junio 2007. pp. 13-38 Recuperado de: <http://www.scielo.org.mx/pdf/enclav/v1n1/v1n1a2.pdf>
38. Rodriguez, G. (2014) Determinantes del desempeño académico universitario. El caso de la Región Caribe colombiana. ICFES. Estudios sobre calidad de la educación en Colombia. Working paper
39. Schultz, T. (1962). Reflections on investment in man, *Journal of Political Economy* 70, no. 5, Part 2 (Oct., 1962)
40. Sen, A. (1998). Capital humano y capacidad humana. En Cuadernos de Economía. (pp. 67-72). Bogotá, Colombia: Facultad de Ciencias Económicas Unal
41. Singh Dhillon, J. y Kaur johal, S. (2016) Intelligence, motivation and study habits as predictors of scholastic achievement in science among secondary school boys. *IMPACT: International Journal of Research in Humanities, Arts and Literature (IMPACT: IJ*
42. Steiner, R. Núñez, J. Cadena, X. y Pardo, R. (2002). ¿Cuáles colegios ofrecen mejor educación en Colombia? Coyuntura Social. Fedesarrollo. Recuperado de http://www.repository.fedesarrollo.org.co/bitstream/11445/1057/2/Co_So_Mayo_2002_Steiner_et_al.pdf
43. Torres-Reyna, O. (2013). Multilevel Analysis. DSS Online Training Section. Recuperado de: <http://www.princeton.edu/~otorres/Multilevel101.pdf>
44. Trucco, D. (2014). Educación y desigualdad en América Latina. Naciones Unidas, Santiago de Chile.
45. Trucco, D. y Nieves, M. (2013). La educación secundaria: un desafío para el ejercicio de derechos en la adolescencia. Los adolescentes y el derecho a la educación. División de Desarrollo Social, CEPAL.
46. Vegas, E. y Petrow, J. (2008). Incrementar el aprendizaje estudiantil en América Latina el desafío para el siglo XXI. Bogotá: Mayol ediciones.
47. Zambrano Jurado, J. (2012). Análisis multinivel del rendimiento escolar en matemáticas para grado cuarto de educación básica primaria en Colombia.

Anexos

Anexo 1. Modelo nulo (Colegios oficiales)

```

Linear mixed model fit by REML ['lmerMod']
Formula: Y ~ 1 + (1 | xcolicfe)
Data: Dat_Estu_pub

REML criterion at convergence: 500517.2

Scaled residuals:
  Min      1Q  Median      3Q      Max
-5.8425 -0.6876  0.0072  0.6838  4.2865

Random effects:
 Groups      Name      Variance Std.Dev.
xcolicfe (Intercept) 280.1   16.74
Residual          1335.4   36.54
Number of obs: 49705, groups: xcolicfe, 646

Fixed effects:
              Estimate Std. Error t value
(Intercept) 263.8105    0.6868   384.1

```

Anexo 2. Modelo con las variables de ajuste (Colegios oficiales)

```

Linear mixed model fit by REML ['lmerMod']
Formula: Y ~ Xgene + Xpadr + Xmadr + Xestr + Xinte + Xingr + Xestr + Xcoljorn +
  Xcolpens + (1 | Xcolicfe)
Data: Dat_Estu_pub

REML criterion at convergence: 497876.6

Scaled residuals:
  Min      1Q  Median      3Q      Max
-5.8717 -0.6800  0.0135  0.6836  4.2153

Random effects:
 Groups      Name      Variance Std.Dev.
xcolicfe (Intercept) 191.1   13.82
Residual          1271.4   35.66
Number of obs: 49705, groups: xcolicfe, 646

Fixed effects:
              Estimate Std. Error t value
(Intercept) 235.11172    1.10836   212.13
Xgene         10.55224    0.33008    31.97
Xpadr          0.97580    0.07640    12.77
Xmadr          2.05506    0.09747    21.08
Xestr          3.21549    0.31398    10.24
Xinte          2.37642    0.40535     5.86
Xingr          1.64950    0.20375     8.10
Xcoljorn      18.26319    2.60800     7.00
Xcolpens      -2.54054    0.48754    -5.21

Correlation of Fixed Effects:
      (Intr) Xgene  Xpadr  Xmadr  Xestr  Xinte  Xingr  Xcljrn
Xgene  -0.082
Xpadr  -0.106 -0.006
Xmadr  -0.178 -0.025 -0.358
Xestr  -0.448 -0.022 -0.066 -0.081
Xinte  -0.133 -0.027 -0.015 -0.050 -0.089
Xingr  -0.244 -0.050 -0.038 -0.105 -0.120 -0.154

```

Xcoljorn	-0.106	0.001	-0.001	-0.015	0.018	-0.013	-0.005
Xcolpens	-0.448	0.000	0.003	0.008	-0.006	0.009	-0.020 -0.031

Anexo 3. likelihood ratio test: significancia conjunta entre el nulo y el 1ro (Colegios oficiales)

```

Likelihood ratio test

Model 1: Y ~ 1 + (1 | xcolicfe)
Model 2: Y ~ Xgene + Xpadr + Xmadr + Xestr + Xinte + Xingr + Xestr + Xcoljorn +
  Xcolpens + (1 | Xcolicfe)
#Df LogLik Df Chisq Pr(>Chisq)
1 3 -250259
2 11 -248938 8 2640.6 < 2.2e-16 ***
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Anexo 4. Modelo final (Colegios oficiales)

```

Linear mixed model fit by REML ['lmerMod']
Formula: Y ~ Xgene + Xpadr + Xmadr + Xestr + Xinte + Xingr + Xestr + Xcoljorn +
  Xcolpens + (1 + Xcoljorn | Xcolicfe)
Data: Dat_Estu_pub

REML criterion at convergence: 497875.7

Scaled residuals:
   Min       1Q   Median       3Q      Max
-5.8722 -0.6798  0.0135  0.6835  4.2154

Random effects:
 Groups   Name                Variance Std.Dev. Corr
xcolicfe (Intercept) 188.4    13.73
          Xcoljorn      122.1    11.05  -0.22
Residual                1271.3   35.66
Number of obs: 49705, groups: xcolicfe, 646

Fixed effects:
              Estimate Std. Error t value
(Intercept) 235.16485    1.10702  212.43
Xgene        10.54998    0.33007   31.96
Xpadr         0.97551    0.07640   12.77
Xmadr         2.05505    0.09747   21.08
Xestr         3.21110    0.31397   10.23
Xinte         2.38132    0.40534    5.87
Xingr         1.64709    0.20375    8.08
Xcoljorn     18.25518    2.90725    6.28
Xcolpens     -2.57629    0.48818   -5.28

Correlation of Fixed Effects:
          (Intr) Xgene Xpadr Xmadr Xestr Xinte Xingr xcljrn
Xgene    -0.083
Xpadr    -0.107 -0.006
Xmadr    -0.178 -0.024 -0.358
Xestr    -0.449 -0.022 -0.066 -0.081
Xinte    -0.133 -0.027 -0.015 -0.050 -0.089
Xingr    -0.244 -0.050 -0.038 -0.105 -0.120 -0.154
Xcoljorn -0.093  0.001 -0.001 -0.014  0.017 -0.012 -0.004
Xcolpens -0.450  0.000  0.003  0.008 -0.005  0.009 -0.020 -0.028

```

Anexo 5. likelihood ratio test: significancia conjunta entre el nulo y el 1ro (Colegios oficiales)

```

Likelihood ratio test

```

```

Model 1: Y ~ Xgene + Xpadr + Xmadr + Xestr + Xinte + Xingr + Xestr + Xcoljorn +
  Xcolpens + (1 | Xcolicfe)
Model 2: Y ~ Xgene + Xpadr + Xmadr + Xestr + Xinte + Xingr + Xestr + Xcoljorn +
  Xcolpens + (1 + Xcoljorn | xcolicfe)
#Df LogLik Df Chisq Pr(>Chisq)
1 11 -248938
2 13 -248938 2 0.8504 0.6536

```

Anexo 6. Corrección del Modelo final (Colegios oficiales)

```

Linear mixed model fit by REML ['lmerMod']
Formula: Y ~ Xgene + Xpadr + Xmadr + Xestr + Xinte + Xingr + Xestr + Xcoljorn +
  Xcolpens + (1 + Xgene | Xcolicfe)
Data: Dat_Estu_pub

REML criterion at convergence: 497868.1

Scaled residuals:
  Min      1Q  Median      3Q      Max
-5.8434 -0.6802  0.0134  0.6829  4.2072

Random effects:
 Groups Name          Variance Std.Dev. Corr
Xcolicfe (Intercept) 200.408 14.157
          Xgene        8.636  2.939  -0.30
Residual          1269.327 35.628
Number of obs: 49705, groups: Xcolicfe, 646

Fixed effects:
              Estimate Std. Error t value
(Intercept) 235.01163    1.11502  210.77
Xgene        10.68133    0.35556   30.04
Xpadr         0.97530    0.07640   12.77
Xmadr         2.05551    0.09745   21.09
Xestr         3.22562    0.31395   10.27
Xinte         2.37578    0.40532    5.86
Xingr         1.64979    0.20373    8.10
Xcoljorn     18.41792    2.59735    7.09
Xcolpens     -2.53636    0.48750   -5.20

Correlation of Fixed Effects:
          (Intr) Xgene Xpadr Xmadr Xestr Xinte Xingr Xcljrn
Xgene    -0.130
Xpadr    -0.106 -0.005
Xmadr    -0.177 -0.023 -0.358
Xestr    -0.445 -0.021 -0.066 -0.081
Xinte    -0.132 -0.026 -0.015 -0.050 -0.089
Xingr    -0.242 -0.047 -0.038 -0.105 -0.120 -0.154
Xcoljorn -0.104 -0.006 -0.001 -0.015  0.019 -0.013 -0.005
Xcolpens -0.445  0.001  0.003  0.008 -0.006  0.009 -0.020 -0.031

```

Anexo 7. likelihood ratio test: significancia conjunta entre el 1ro y el 2do (Colegios oficiales)

```

Likelihood ratio test

Model 1: Y ~ Xgene + Xpadr + Xmadr + Xestr + Xinte + Xingr + Xestr + Xcoljorn +
  Xcolpens + (1 | Xcolicfe)
Model 2: Y ~ Xgene + Xpadr + Xmadr + Xestr + Xinte + Xingr + Xestr + Xcoljorn +
  Xcolpens + (1 + Xgene | Xcolicfe)
#Df LogLik Df Chisq Pr(>Chisq)
1 11 -248938
2 13 -248934 2 8.444 0.01467 *

```

 Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Anexo 8. Modelo nulo (Colegios "No" oficiales)

```
Linear mixed model fit by REML ['lmerMod']
Formula: Y ~ 1 + (1 | Xcolicfe)
Data: Dat_Estu_priv

REML criterion at convergence: 413396.9

Scaled residuals:
  Min      1Q  Median      3Q      Max
-7.7824 -0.6587  0.0237  0.6688  4.1589

Random effects:
 Groups   Name      Variance Std.Dev.
Xcolicfe (Intercept)  944.1    30.73
Residual                1192.1    34.53
Number of obs: 41384, groups: Xcolicfe, 836

Fixed effects:
              Estimate Std. Error t value
(Intercept)  284.577      1.091    260.9
```

Anexo 9. Modelo con las variables de ajuste (Colegios "No" oficiales)

```
Linear mixed model fit by REML ['lmerMod']
Formula: Y ~ Xgene + Xpadr + Xmadr + Xestr + Xinte + Xingr + Xestr + Xcoljorn +
  Xcolpens + (1 | Xcolicfe)
Data: Dat_Estu_priv

REML criterion at convergence: 411202

Scaled residuals:
  Min      1Q  Median      3Q      Max
-8.0730 -0.6543  0.0207  0.6740  4.0627

Random effects:
 Groups   Name      Variance Std.Dev.
Xcolicfe (Intercept)  388.6    19.71
Residual                1148.7    33.89
Number of obs: 41384, groups: Xcolicfe, 836

Fixed effects:
              Estimate Std. Error t value
(Intercept)  219.8518      1.8290   120.20
Xgene         8.1453      0.3636    22.40
Xpadr         1.1489      0.0865    13.28
Xmadr         1.7583      0.1014    17.35
Xestr         0.8980      0.3050     2.94
Xinte         1.0910      0.6319     1.73
Xingr         1.9839      0.1715    11.57
Xcoljorn     31.4732      1.7312    18.18
Xcolpens      2.3438      0.2567     9.13

Correlation of Fixed Effects:
      (Intr) Xgene Xpadr Xmadr Xestr Xinte Xingr Xcljrn
Xgene  -0.082
Xpadr  -0.068  0.000
Xmadr  -0.101 -0.019 -0.400
Xestr  -0.259 -0.006 -0.062 -0.056
Xinte  -0.201 -0.022 -0.020 -0.038 -0.059
Xingr  -0.041 -0.034 -0.081 -0.123 -0.220 -0.097
```

```
Xcoljorn -0.574 0.020 -0.013 -0.022 -0.015 -0.025 -0.025
Xcolpens -0.314 -0.008 -0.023 -0.060 -0.105 -0.021 -0.089 -0.201
```

Anexo 10. Corrección del Modelo con las variables de ajuste eliminando variable Xgene (Colegios “No” oficiales)

```
Linear mixed model fit by REML ['lmerMod']
Formula: Y ~ Xgene + Xpadr + Xmadr + Xestr + Xingr + Xestr + Xcoljorn +
  Xcolpens + (1 | Xcolicfe)
Data: Dat_Estu_priv

REML criterion at convergence: 411205.9

Scaled residuals:
  Min      1Q  Median      3Q      Max
-8.0733 -0.6548  0.0202  0.6736  4.0662

Random effects:
 Groups Name          Variance Std.Dev.
Xcolicfe (Intercept) 389.2    19.73
Residual             1148.7   33.89
Number of obs: 41384, groups: Xcolicfe, 836

Fixed effects:
              Estimate Std. Error t value
(Intercept) 220.49474    1.79244  123.01
Xgene        8.15961     0.36357   22.44
Xpadr        1.15179     0.08649   13.32
Xmadr        1.76484     0.10129   17.42
Xestr        0.92840     0.30446    3.05
Xingr        2.01221     0.17068   11.79
Xcoljorn     31.55136     1.73174   18.22
Xcolpens     2.35164     0.25665    9.16

Correlation of Fixed Effects:
      (Intr) Xgene Xpadr Xmadr Xestr Xingr Xcljrn
Xgene  -0.088
Xpadr  -0.073  0.000
Xmadr  -0.111 -0.020 -0.401
Xestr  -0.277 -0.008 -0.063 -0.058
Xingr  -0.062 -0.036 -0.083 -0.127 -0.227
Xcoljorn -0.591  0.019 -0.014 -0.023 -0.017 -0.027
Xcolpens -0.325 -0.008 -0.024 -0.061 -0.106 -0.091 -0.202
```

Anexo 11. likelihood ratio test: significancia conjunta entre el nulo y el 1ro (Colegios “No” oficiales)

```
Likelihood ratio test

Model 1: Y ~ 1 + (1 | Xcolicfe)
Model 2: Y ~ Xgene + Xpadr + Xmadr + Xestr + Xingr + Xestr + Xcoljorn +
  Xcolpens + (1 | Xcolicfe)
#Df LogLik Df Chisq Pr(>Chisq)
 1   3 -206698
 2  10 -205603  7  2191 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Anexo 12. Modelo final (Colegios “No” oficiales)

```
Linear mixed model fit by REML ['lmerMod']
Formula: Y ~ Xgene + Xpadr + Xmadr + Xestr + Xingr + Xestr + Xcoljorn +
  Xcolpens + (1 + Xcoljorn + Xgene | Xcolicfe)
Data: Dat_Estu_priv
```

REML criterion at convergence: 411189.5

Scaled residuals:

Min	1Q	Median	3Q	Max
-8.0882	-0.6549	0.0205	0.6746	4.0721

Random effects:

Groups	Name	Variance	Std.Dev.	Corr
xcolicfe	(Intercept)	405.06	20.126	
	xcoljorn	445.21	21.100	-0.56
	xgene	13.86	3.723	-0.37 0.54
Residual		1145.91	33.851	

Number of obs: 41384, groups: xcolicfe, 836

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	220.67686	1.78620	123.55
xgene	7.96416	0.40194	19.81
xpadr	1.15089	0.08647	13.31
xmadr	1.76576	0.10126	17.44
xestr	0.92940	0.30431	3.05
xingr	2.01707	0.17065	11.82
xcoljorn	31.47198	1.71861	18.31
xcolpens	2.36712	0.25627	9.24

Correlation of Fixed Effects:

	(Intr)	xgene	xpadr	xmadr	xestr	xingr	xcljrn
xgene	-0.142						
xpadr	-0.073	0.000					
xmadr	-0.112	-0.017	-0.401				
xestr	-0.278	-0.007	-0.063	-0.058			
xingr	-0.061	-0.033	-0.084	-0.127	-0.228		
xcoljorn	-0.591	0.100	-0.014	-0.022	-0.016	-0.028	
xcolpens	-0.325	-0.008	-0.024	-0.061	-0.107	-0.092	-0.203

Anexo 13. likelihood ratio test: significancia conjunta entre el 1ro y el 2do (Colegios "No" oficiales)

Likelihood ratio test

Model 1: $Y \sim xgene + xpadr + xmadr + xestr + xingr + xestr + xcoljorn + xcolpens + (1 | xcolicfe)$

Model 2: $Y \sim xgene + xpadr + xmadr + xestr + xingr + xestr + xcoljorn + xcolpens + (1 + xcoljorn + xgene | xcolicfe)$

#Df LogLik Df Chisq Pr(>Chisq)

1 10 -205603

2 15 -205595 5 16.424 0.005732 **

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Anexo 14. Modelo nulo (Colegios de "Ambas" Naturalezas)

Linear mixed model fit by REML ['lmerMod']

Formula: $Y \sim 1 + (1 | xcolicfe)$

Data: Dat_Estu

REML criterion at convergence: 914493.2

Scaled residuals:

Min	1Q	Median	3Q	Max
-7.5338	-0.6739	0.0163	0.6760	4.3852

Random effects:

Groups	Name	Variance	Std.Dev.
Xcolicfe	(Intercept)	757.3	27.52
	Residual	1270.5	35.64

Number of obs: 91089, groups: xcolicfe, 1482

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	275.3314	0.7335	375.4

Anexo 15. Modelo con las variables de ajuste (Colegios de "Ambas" Naturalezas)

Linear mixed model fit by REML ['lmerMod']
 Formula: Y ~ Xgene + Xpadr + Xmadr + Xestr + Xinte + Xingr + Xestr + Xcoljorn + Xcolnatu + Xcolpens + (1 | Xcolicfe)
 Data: Dat_Estu

REML criterion at convergence: 909417.8

Scaled residuals:

Min	1Q	Median	3Q	Max
-7.8572	-0.6685	0.0173	0.6790	4.3093

Random effects:

Groups	Name	Variance	Std.Dev.
xcolicfe	(Intercept)	304.5	17.45
	Residual	1217.5	34.89

Number of obs: 91089, groups: xcolicfe, 1482

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	219.76399	1.51744	144.83
Xgene	9.50516	0.24444	38.89
Xpadr	1.05869	0.05726	18.49
Xmadr	1.93951	0.07032	27.58
Xestr	2.14055	0.21974	9.74
Xinte	2.03269	0.33759	6.02
Xingr	1.87472	0.13144	14.26
Xcoljorn	29.66101	1.40516	21.11
Xcolnatu	13.06002	1.45002	9.01
Xcolpens	1.48017	0.22366	6.62

Correlation of Fixed Effects:

	(Intr)	Xgene	Xpadr	Xmadr	Xestr	Xinte	Xingr	xcljrn	xcolnt	
Xgene		-0.060								
Xpadr		-0.058	-0.003							
Xmadr		-0.091	-0.022	-0.377						
Xestr		-0.237	-0.014	-0.066	-0.070					
Xinte		-0.109	-0.027	-0.015	-0.045	-0.072				
Xingr		-0.072	-0.040	-0.060	-0.117	-0.178	-0.121			
Xcoljorn		-0.571	0.014	-0.010	-0.021	-0.009	-0.015	-0.024		
Xcolnatu		-0.813	0.008	0.007	0.007	0.030	0.002	-0.004	0.595	
Xcolpens		-0.394	-0.001	-0.016	-0.038	-0.082	0.004	-0.083	-0.189	0.315

Anexo 16. likelihood ratio test: significancia conjunta entre el nulo y el 1ro (Colegios de "Ambas" Naturalezas)

Likelihood ratio test

Model 1: Y ~ 1 + (1 | Xcolicfe)
 Model 2: Y ~ Xgene + Xpadr + Xmadr + Xestr + Xinte + Xingr + Xestr + Xcoljorn + Xcolnatu + Xcolpens + (1 | Xcolicfe)

#Df	LogLik	Df	Chisq	Pr(>Chisq)
1	3	-457247		

```

2 12 -454709 9 5075.4 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Anexo 17. Modelo final (Colegios de "Ambas" Naturalezas)

```

Linear mixed model fit by REML ['lmerMod']
Formula: Y ~ Xgene + Xpadr + Xmadr + Xestr + Xinte + Xingr + Xestr + Xcoljorn +
  Xcolnatu + Xcolpens + (1 + Xcolnatu + xcoljorn | Xcolicfe)
Data: Dat_Estu

```

REML criterion at convergence: 909351.4

Scaled residuals:

Min	1Q	Median	3Q	Max
-7.8589	-0.6680	0.0170	0.6787	4.3120

Random effects:

Groups	Name	Variance	Std.Dev.	Corr
Xcolicfe	(Intercept)	400.0	20.00	
	xcolnatu	550.1	23.46	-0.80
	xcoljorn	454.4	21.32	-0.53 0.17
Residual		1217.1	34.89	

Number of obs: 91089, groups: Xcolicfe, 1482

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	220.90077	1.65169	133.74
Xgene	9.53378	0.24447	39.00
Xpadr	1.05408	0.05724	18.41
Xmadr	1.93945	0.07031	27.58
Xestr	2.14519	0.21944	9.78
Xinte	2.11929	0.33741	6.28
Xingr	1.82025	0.13159	13.83
Xcoljorn	29.78625	1.56474	19.04
Xcolnatu	12.63606	1.54159	8.20
Xcolpens	1.16916	0.22853	5.12

Correlation of Fixed Effects:

	(Intr)	Xgene	Xpadr	Xmadr	Xestr	Xinte	Xingr	xcljrn	xcolnt
Xgene	-0.054								
Xpadr	-0.055	-0.003							
Xmadr	-0.085	-0.022	-0.377						
Xestr	-0.223	-0.014	-0.065	-0.070					
Xinte	-0.099	-0.027	-0.016	-0.045	-0.073				
Xingr	-0.071	-0.041	-0.059	-0.116	-0.175	-0.121			
Xcoljorn	-0.606	0.012	-0.009	-0.019	-0.009	-0.014	-0.023		
Xcolnatu	-0.870	0.008	0.007	0.007	0.031	0.001	-0.002	0.660	
Xcolpens	-0.379	-0.002	-0.014	-0.034	-0.071	0.003	-0.076	-0.172	0.303

Anexo 18. likelihood ratio test: significancia conjunta entre el 1ro y el 2do (Colegios de "Ambas" Naturalezas)

Likelihood ratio test

Model 1: Y ~ Xgene + Xpadr + Xmadr + Xestr + Xinte + Xingr + Xestr + Xcoljorn + Xcolnatu + Xcolpens + (1 | Xcolicfe)

Model 2: Y ~ Xgene + Xpadr + Xmadr + Xestr + Xinte + Xingr + Xestr + Xcoljorn + Xcolnatu + Xcolpens + (1 + Xcolnatu + xcoljorn | Xcolicfe)

#Df	LogLik	Df	Chisq	Pr(>Chisq)
1	12	-454709		
2	17	-454676	5	66.473 5.544e-13 ***

```

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Anexo 19. Tabla de Variables

Variable	DESCRIPCION
<i>Xedad</i>	Edad del estudiante a la hora de presentar el examen.
<i>Xgene</i>	Género del estudiante inscrito.
<i>Xpadr</i>	Nivel educativo de la madre.
<i>Xpadr</i>	Nivel educativo del padre
<i>Xestr</i>	Estrato socioeconómico de la residencia del estudiante según factura de energía.
<i>Xinte</i>	Servicios con los que cuenta en su hogar: Internet.
<i>Xcomp</i>	Bienes que posee su hogar: Computador.
<i>Xingr</i>	Estrato socioeconómico de la residencia del estudiante según factura de energía.
<i>Xcoljorn</i>	Jornada de la sede.
<i>Xcolnatu</i>	Naturaleza de la institución educativa, puede ser pública o privada.
<i>Xcolpens</i>	Valor de la pensión pagada por el estudiante en el último año.
<i>PUNT_GLOBAL</i>	Puntaje total obtenido por el evaluado.

Anexo 20. Cálculo de la correlación intraclase para el modelo nulo (Modelo I)

Naturaleza	IC
<i>Oficial</i>	$IC = \frac{(16.74)^2}{(16.74)^2 + (36.54)^2} = 0.174$
<i>No oficial</i>	$IC = \frac{(30.73)^2}{(30.73)^2 + (34.53)^2} = 0.442$
<i>Ambos</i>	$IC = \frac{(27.52)^2}{(27.52)^2 + (35.64)^2} = 0.374$

Anexo 21. Cálculo de la correlación intraclase para el modelo con variables de ajuste (Modelo II)

Naturaleza	IC
<i>Oficial</i>	$IC = \frac{(13.82)^2}{(13.82)^2 + (35.66)^2} = 0.13$
<i>No oficial</i>	$IC = \frac{(19.73)^2}{(19.73)^2 + (33.89)^2} = 0.25$
<i>Ambos</i>	$IC = \frac{(17.45)^2}{(17.45)^2 + (34.89)^2} = 0.20$

Anexo 22. Cálculo de la correlación intraclase para el modelo final (Modelo III)

Naturaleza	IC
<i>Oficial</i>	$IC = \frac{(14.16)^2 + (2.94)^2}{(14.16)^2 + (2.94)^2 + (35.63)^2} = 0.14$
<i>No oficial</i>	$IC = \frac{(20.13)^2 + (21.10)^2 + (3.72)^2}{(20.13)^2 + (21.10)^2 + (3.72)^2 + (33.85)^2} = 0.43$
<i>Ambos</i>	$IC = \frac{(20.00)^2 + (23.46)^2 + (21.32)^2}{(20.00)^2 + (23.46)^2 + (21.32)^2 + (34.89)^2} = 0.54$

Anexo 23– Cambios en la estructura de la prueba SABER 11

Hasta primer semestre del 2014	A partir del segundo semestre del 2014
<ul style="list-style-type: none"> • Un Núcleo Común, con ocho pruebas, que todas las personas deben presentar. • Un Componente Flexible, en el que cada persona selecciona o una prueba de profundización o una prueba interdisciplinar, de acuerdo con sus intereses. 	<ul style="list-style-type: none"> • Conformado por cinco pruebas, que todas las personas deben presentar.
<ul style="list-style-type: none"> • Núcleo Común: <ul style="list-style-type: none"> ○ Lenguaje ○ Matemáticas ○ Biología ○ Física ○ Química ○ Ciencias Sociales ○ Filosofía 	<ul style="list-style-type: none"> • Pruebas: <ul style="list-style-type: none"> ○ Lectura Crítica ○ Matemáticas ○ Ciencias Naturales ○ Sociales y Ciudadanas ○ Inglés • Sub pruebas: <ul style="list-style-type: none"> ○ Razonamiento cuantitativo

<ul style="list-style-type: none"> ○ Inglés • Componente Flexible: <ul style="list-style-type: none"> ○ Profundización en alguna de las siguientes áreas: Biología, Ciencias Sociales, Lenguaje o Matemáticas. ○ Interdisciplinar en alguno de los siguientes temas: Violencia y Sociedad o Medio Ambiente 	<ul style="list-style-type: none"> ○ Competencias ciudadanas
<ul style="list-style-type: none"> • Únicamente preguntas de selección múltiple 	<ul style="list-style-type: none"> • Preguntas de selección múltiple • Preguntas abiertas
<ul style="list-style-type: none"> • Puntajes en cada una de las pruebas del Núcleo Común • Puntaje y nivel de desempeño en las competencias evaluadas • Puntaje y nivel de desempeño en los componentes evaluados • Nivel de desempeño en la prueba de inglés • Puesto 	<ul style="list-style-type: none"> • Puntaje global (de 0 a 500) • Puntaje en cada prueba • Puntaje en las sub-pruebas de Razonamiento Cuantitativo, y Competencias Ciudadanas. • Nivel de desempeño. En 2014 únicamente para la prueba de inglés. A partir del segundo semestre de 2015, para todas las pruebas • Puesto

Fuente: Mineducacion.gov.co, 2014.

Anexo 24 – Cambios en la metodología clasificación de las instituciones educativas

Hasta primer semestre del 2014	A partir del segundo semestre del 2014
Se tenía en cuenta los resultados de los dos últimos años.	Se tendrán en cuenta los resultados de los tres últimos años.
La categoría se estimaba con la totalidad de los resultados	Para estimar la categoría se excluye prueba a prueba el 20% de los resultados más bajo.
El número mínimo de evaluados en los últimos dos años debía ser igual o superior a 6 estudiantes	El número mínimo de evaluados en los últimos tres años debía ser igual o superior a 9 estudiantes.
Se asignaban 7 categorías: muy superior, superior, alto, medio, bajo, inferior y muy inferior.	Se asignarán 5 categorías: A+, A, B, C y D
La clasificación se publicaba una vez al año.	La clasificación se actualizará dos veces al año.

Fuente: Mineducacion.gov.co, 2014.

Anexo 25 – Clasificación antigua de las instituciones educativas según categoría.

Categoría	Muy inferior	Inferior	Bajo	Medio	Alto	Superior	Muy superior
Límite inferior	0.00	0.3955	0.5109	0.5973	0.6778	0.7450	0.8318
Límite superior	0.3955469	0.510984	0.5973	0.6778	0.7450	0.8318	100.00

Fuente: Guía Clasificación Planteles Saber11 2012, 2012.

Anexo 26 – Clasificación vigente de las instituciones educativas según categoría.

Categoría	D	C	B	A	A+
Rango del Índice General	$0 \leq I_g \leq 0.62$	$0.62 < I_g \leq 0.67$	$0.67 < I_g \leq 0.72$	$0.72 < I_g \leq 0.77$	$I_g > 0.77$

Fuente: Clasificación de establecimientos y sedes Saber 11, 2014.

Anexo 27 – Cálculo del porcentaje de la varianza explicada.

% Varianza explicada...	Entre colegios	Entre estudiantes
<i>Oficial</i>	$\frac{280.1 - (200.41 + 8.64)}{200.41 + 8.64} = 0.3399$	$\frac{1335.4 - 1269.33}{1269.33} = 0.0521$
<i>No oficial</i>	$\frac{944.1 - (405.06 + 445.21 + 13.86)}{405.06 + 445.21 + 13.86} = 0.0925$	$\frac{1192.1 - 1145.91}{1145.91} = 0.0403$
<i>Ambos</i>	$\frac{757.3 - (400 + 550.1 - 454.4)}{400 + 550.1 - 454.4} = -0.4608$	$\frac{1270.5 - 1217.1}{1217.1} = 0.0439$

Anexo 28 – Cálculo del porcentaje de la varianza total explicada.

Naturaleza	% Varianza total explicada
<i>Oficial</i>	$\frac{(280.1 + 1335.4) - ((200.41 + 8.64) + 1269.33)}{((200.41 + 8.64) + 1269.33)} = 0.0927$
<i>No oficial</i>	$\frac{(944.1 + 1192.1) - ((405.06 + 445.21 + 13.86) + 1145.91)}{((405.06 + 445.21 + 13.86) + 1145.91)} = 0.0627$
<i>Ambos</i>	$\frac{(757.3 + 1270.5) - ((400 + 550.1 - 454.4) + 1217.1)}{((400 + 550.1 - 454.4) + 1217.1)} = -0.2265$