

## Incident factors in the academic performance of secondary education students in Cundinamarca-Colombia

### *Factores incidentes en el desempeño académico de estudiantes de educación media en Cundinamarca-Colombia*

Lilian Daniela Suárez-Riveros<sup>1\*</sup>, Oscar Jardey Suárez<sup>2</sup>, Julio Del Carmen Lizarazo-Osorio<sup>3</sup>

<sup>1</sup>\*MSc. en Ciencias de Datos, ldsuarezriveros@gmail.com, <https://orcid.org/0000-0002-8329-0765>, IQuartil, Bogotá D.C., Colombia.

<sup>2</sup>Dr. en Educación, ojsuarez@udenar.edu.co, <https://orcid.org/0000-0001-8780-595X>, Universidad de Nariño, Boyacá, Colombia.

<sup>3</sup>Dr. en Matemáticas, julio.lizarazo@uptc.edu.co, <https://orcid.org/0000-0003-0576-1987>, Universidad Pedagógica y Tecnológica de Colombia, Tunja, Colombia.

**Forma de citar:** Suárez-Riveros, L. D., Suárez, O. J., y Lizarazo-Osorio, J. D. C. (2024). Factores incidentes en el desempeño académico de estudiantes de educación media en Cundinamarca-Colombia. *Eco Matemático*, 15 (1), 48-59. <https://doi.org/10.22463/17948231.4128>

Recepción: Agosto 27, 2023

Aprobación: Diciembre 22, 2023.

#### Palabras Clave

Aprender, calidad de la educación, estandarizadas, modelamiento, pruebas, rendimiento académico.

**Resumen:** La calidad de la educación es multifactorial y multidimensional. El presente estudio se centra en comprender las variables que inciden en el rendimiento académico, medido en las pruebas estandarizadas Saber 11, de los estudiantes de educación media en el departamento de Cundinamarca-Colombia. El enfoque del estudio es cuantitativo. El proceso en el desarrollo del trabajo se ciñe a la metodología Cross Industry Standard Process for Data Mining (Comprensión del negocio o problema, Comprensión de los datos, Preparación de los datos, Modelado, Evaluación e Implementación). Los datos utilizados son las bases de datos abiertas del Instituto Colombiano para el Fomento de la Educación Superior ICFES. El modelo utilizado para determinar las variables que tienen efecto es de regresión Multinivel. Los resultados indican que las variables con efecto fijo en la predicción del rendimiento académico son género, trabajo del estudiantado, educación de los padres, número de integrantes del núcleo familiar y recursos contextuales. A manera de conclusión, transformar las condiciones de los contextos y favorecer el tiempo en familia, en especial de la madre, al parecer pueden incidir en la mejor obtención de resultados en el rendimiento académico del estudiantado en las pruebas Saber 11.

\*Autor para correspondencia: [ldsuarezriveros@gmail.com](mailto:ldsuarezriveros@gmail.com)

<https://doi.org/10.22463/17948231.4128>

## Keywords

Learning, quality of education, standardized, modeling, tests, academic performance.

**Abstract:** The quality of education is multifactorial and multidimensional. The present study focuses on understanding the variables that affect the academic performance, measured in the Saber 11 standardized tests of secondary school students in the department of Cundinamarca-Colombia. The focus of the study is quantitative. The process in the development of the work adheres to the Cross Industry Standard Process for Data Mining methodology (Understanding the business or problem, Understanding the data, Data preparation, Modeling, Evaluation and Implementation). The data used are the open databases of the Colombian Institute for the Promotion of Higher Education ICFES. The model used to determine the variables that have an effect is Multilevel regression. The results indicate that the variables with a fixed effect in predicting academic performance are gender, student work, parental education, number of family members and contextual resources. In conclusion, transforming the conditions of the contexts and favoring time with the family, especially with the mother, seem to have an impact on obtaining better results in the academic performance of students in the Saber 11 tests.

## Introducción

Los diferentes países, a través de políticas promueven la calidad de la educación, toda vez que han encontrado que está directamente relacionada con la calidad de vida, en particular de la salud, de las comunidades. Adicionalmente, para Reyes (2020), la calidad de la educación está estrechamente relacionada con los diversos estándares necesarios para una prestación del servicio, más allá de los diferentes aspectos de orden educativo.

De igual manera, en Colombia, la cobertura y la calidad de la educación se constituye en uno de los desafíos para el crecimiento económico y la ampliación de oportunidades en pro del bienestar de la población. Además, Colombia continúa siendo uno de los países más desiguales de América Latina, sobresaliendo como una causa la insuficiente capacidad del sistema educativo para incentivar la movilidad social y aumentar la acumulación de capital humano (Delgado Barrera, 2014).

Por lo tanto, el Ministerio de Educación Nacional (MEN) asume la calidad como “...*el conjunto de atributos articulados, interdependientes, dinámicos, construidos por la comunidad académica como referentes y que responden a las demandas sociales, culturales y ambientales...*”(2019), por lo que entonces la calidad se asume desde una mirada

multidimensional, respetando la diversidad presente en Colombia.

De acuerdo con ello, los resultados de las pruebas estandarizadas se constituyen en un elemento, no el único, que da indicios sobre los factores que afectan la calidad de la educación en un país. En especial, en Colombia, las pruebas Saber 11 son pruebas estandarizadas construidas, aplicadas y divulgadas por el Instituto Colombiano para la Evaluación de la Educación (ICFES), que dan cuenta de algunos elementos relacionados con la educación básica y mediaClick or tap here to enter text., los que podrían ser considerados para trazar políticas educativas que contribuyan en transformar su calidad de vida.

En consecuencia, el presente estudio se enfoca en Cundinamarca, un departamento de Colombia que está compuesto por 116 municipios y que es la cuarta área más poblada de Colombia, cuyo propósito es identificar las variables que afectan el rendimiento académico medido en las pruebas Saber 11, para el departamento, a través de un modelo lineal multinivel.

Por ende, tomando como referencia las bases de datos abiertas, la pregunta que este artículo se propone tratar es ¿cuáles son las variables que inciden en el rendimiento académico, medido en las pruebas estandarizadas Saber 11, de los

estudiantes de educación media en el departamento de Cundinamarca-Colombia?

Se elaboró una revisión de diferentes estudios, aplicando dos puntos de vista: identificar las técnicas estadísticas empleadas para comprender o predecir la obtención del rendimiento académico y determinar las variables estudiadas que están asociadas al rendimiento académico.

Primeramente, se identificó que los modelos de regresión lineal y logística han sido usados con mayor frecuencia para la predicción (Qiu & Wu, 2019; Rebai et al., 2019; Sbroglio-Rizzotto & Aniceto-França, 2020). A su vez, el análisis descriptivo e inferencial, empleado junto con la utilización de diversos modelos predictivos es la segunda técnica estadística de preferencia (Chacón-Vargas & Roldán-Villalobos, 2021; Rodríguez-De-Souza-Pajuelo et al., 2021; Xu et al., 2019). Seguidamente, los árboles de decisión (Hasan et al., 2018; Masci et al., 2018), las redes neuronales (Guo et al., 2015; Lau et al., 2019) y los modelos lineales multinivel (Lisboa- Bartholo & Da-Costa, 2016; Orjuela, 2014; Schuth et al., 2017) son las terceras técnicas seleccionadas para el estudio de la comprensión o explicación del rendimiento académico.

Particularmente, los modelos lineales, conocidos como regresión lineal, han sido ampliamente utilizados en diversas aplicaciones por su interpretabilidad, pero, algunas se relacionan con datos agrupados y relacionados donde este método es inadecuado, ya que uno de sus supuestos establece la independencia entre las variables (Correa Morales & Salazar Uribe, 2016).

Por consiguiente, surgieron los modelos lineales generalizados mixtos que permite incluir este comportamiento por medio de los efectos fijos, coeficientes que no alteran entre niveles, y efectos aleatorios, tomados al azar de un conjunto de datos con niveles. De acuerdo con esto, los modelos

multinivel o lineal jerárquico generalizado son un tipo de modelo generalizado mixto empleados para datos con una estructura jerárquica, por ejemplo, el sector educativo donde los puntajes de pruebas estandarizadas pueden estar relacionados con los colegios o su geolocalización (Acevedo-Álvarez, 2008).

Por otro lado, se determinaron 16 variables asociadas con el rendimiento académico en pruebas estandarizadas, en diferentes partes del mundo. Las variables son la naturaleza del colegio – pública o privada - (Contreras et al., 2019; Orjuela, 2014), el nivel socioeconómico (Ali et al., 2013; Guo et al., 2015), el género (Ibourk & Amaghous, 2014; Rodríguez-De-Souza-Pajuelo et al., 2021), la ubicación de la institución - rural o urbana - (Rebai et al., 2019; Wandera et al., 2019), la motivación (Ali et al., 2013; Froiland & Oros, 2014), los antecedentes académicos previos (Guo et al., 2015; Hasan et al., 2018), la condición académica y ocupacional de los padres (Maulida & Kariyam, 2017), los recursos materiales disponibles (Salal & Abdullaev, 2020), la dependencia espacial (Ariza et al., 2021), la relación con pares (Cvencek et al., 2017), las habilidades cognitivas (Schuth et al., 2017), la alimentación (Gaete-Rivas et al., 2021), las características de las instituciones (Masci et al., 2018), la composición familiar (Chacón-Vargas & Roldán-Villalobos, 2021), la experiencia docente (Qiu & Wu, 2019) y la edad (Shah et al., 2019).

Conforme a lo anterior, en la comunidad académica para la predicción del rendimiento académico se acude a diversas variables sin que haya acuerdo global, sin embargo, de acuerdo con la revisión bibliográfica realizada, el nivel socioeconómico a nivel de básica, media y universidad es la más estudiada, mientras que la variable alimentación es la de menor frecuencia en los estudios.

## Materiales y Métodos

El enfoque de la investigación es cuantitativo (Hernández-Sampieri et al., 2014), fundamentado en el modelo lineal multinivel (Correa Morales & Salazar Uribe, 2016; Gelman & Hill, 2006). Los momentos en la investigación son: comprensión del problema, comprensión de los datos, preparación de los datos, modelado y evaluación.

En la Figura 1 se señalan los momentos, en los que la comprensión se centra en el problema del objeto en estudio, la comprensión de datos se orienta a la descripción y concreción del conjunto de datos, la preparación de los datos implica organizar y adecuar los datos (imputar, transformar, entre otros) para ser procesados, en el modelado se aplican las técnicas de modelamiento estadístico y en la evaluación se reflexionan los resultados del modelo en el contexto del problema y del marco teórico.

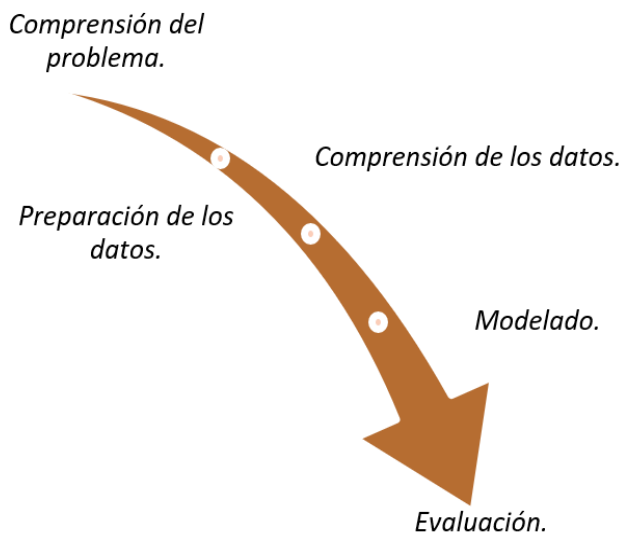


Figura 1. Momentos de la investigación

Las bases de datos utilizadas en el presente estudio son los resultados de las pruebas Saber 11, Indicadores de Infancia y adolescencia, información de catalogación de municipios por el Departamento Administrativo Nacional de Estadísticas (DANE) y consolidado de educación del Departamento de Cundinamarca durante el lustro antes de la pandemia

debido a la COVID-19. Los datos se consolidaron en 49 variables y 174.693 registros (observaciones).

A fin de la preparación de los datos, se procedió a eliminar las variables con más de 20% de datos faltantes e imputar las variables restantes por medio del algoritmo K-Nearest Neighbor (KNN). Luego, se diseñó una base de datos basados en el modelo entidad relación con la información de las diferentes fuentes primarias. Diversos estudios educativos señalan la necesidad de estudiar el contexto como un factor que integra el acto educativo de enseñar, por lo que integrar la información municipal cuantitativa disponible como los delitos, cobertura de educación superior, calidad del agua, entre otros es relevante.

Por otro lado, para el modelamiento de los datos se utilizó R®, específicamente la función lmer() de la librería “library(lme4)” para entrenar los modelos multinivel relacionados en la Tabla 1. De manera que, el conjunto de datos se distribuyó en dos partes, un conjunto que contiene el 70% de las observaciones con el cual se hace el entrenamiento del modelo y otro archivo con el 30% de las observaciones con las que se hacen las pruebas del modelo.

Una vez entrenado el modelo, se procedió a comparar el Mean Absolute Percentage Error (MAPE) y el Criterio de Información de Akaike (AIC) como criterio de calidad de medida del modelo (Touchon, 2021). Cabe mencionar que las características del equipo de cómputo para el procesamiento son: procesador Ryzen 5 (4 núcleos y 8 subprocesos o hilos), memoria RAM 20 Gb, SSD 6.0 Gb/s y tarjeta gráfica AMD Radeon®.

Tabla I. Modelos.

Modelo	Descripción
1	Dos niveles (colegios, municipios), con todas las variables del conjunto de datos.
2	Dos niveles (colegios, municipios) con las variables del conjunto de datos, excepto las que no resultaron estadísticamente significativas en el modelo 1.
3	Dos niveles (colegios, municipios), considerando que los resultados varíen por variables relacionadas con los colegios.
4	Dos niveles (colegios, municipios), considerando que los resultados varíen por variables relacionadas con los municipios.
5	Dos niveles (colegios, municipios), considerando que los resultados varíen por variables relacionadas con los municipios y los colegios, simultáneamente.

El modelo multinivel corresponde con la ampliación de los modelos de regresión lineal, es decir que en cada nivel puede haber varios modelos lineales, en los que se van relacionando los modelos del nivel 1 con un modelo del nivel 2 y sucesivamente (Murillo-Torrecilla, 2008). El modelo multinivel, en comparación con otros modelos utilizados en el presente estudio, resulta ser adecuado por la simplicidad, robustez y que es equiparable con los modelos, específicamente al considerar el Error Cuadrático Medio (Root Mean Square Error RMSE).

### Resultados y Discusión

De acuerdo con el análisis exploratorio, los municipios Chía, Facatativá, Fusagasugá, Madrid, Mosquera, Soacha y Zipaquirá, que están geográficamente cerca a Bogotá, concentran cerca del 50% de los habitantes del departamento de Cundinamarca. La Figura 2, permite observar los puntajes promedios, así como la diferencia en los municipios que tienen oferta educativa oficial y no oficial. Los municipios que no están en la figura 2 es dado que tienen un solo tipo de oferta educativa.

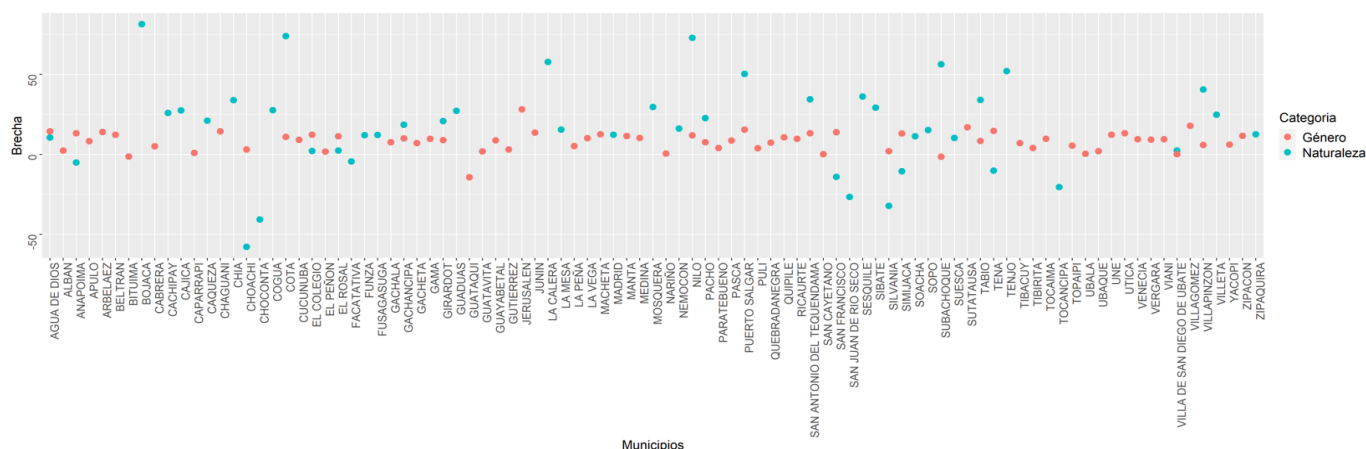


Figura 2. Diferencia de los puntajes por género, naturaleza del colegio y municipio.

Por otro lado, el modelo 5 reportado es el de menor AIC es 1.223.714, con un MAPE de entrenamiento de 11,56% y el MAPE de prueba 11,72%. Además, en la Tabla II se evidencia las variables que no son estadísticas significancias en el modelo 1 como por ejemplo la cantidad de delitos en el municipio, la cobertura en educación superior en los diferentes niveles y la tasa de muertes en accidentes de tránsito, que no se tuvieron en cuenta

en los posteriores modelos. Cabe resaltar que el ordenador utilizado demoro en promedio tres horas cada vez que se corrieron los datos para obtener cada modelo relacionados en la Tabla I.

**Tabla II.** Variables que no tienen significancia estadística.

Variable	t value
CANTIDADDELITOS	-0,69
MEDIAB	-1,65
Coberturaeducacionsuperior	1,09
CoberturaEducacionTecnologica	-0,76
Coberturaacueducto	-1,75
Cobeturabrutamedia	-1,31
Tasahomicidios1217	1,14
Tasadesercionmedia	0,98
Tasarepitenciamedia	-0,11
Prevalenciaexcesopesoninoadolescente	0,30
Porcentajeadolescentesconflictoarmado1217	0,19
Tasamuertesaccidentestransito1217	1,79
Porcentajeemujeresmadresembarazadas1019	1,31
Tasasuicidios1217	0,53
Tasaviolencia1217	-0,29
Tasaviolenciaparejamen18	-0,34
ESTUTIENEETNIASI	-1,88

Resultado del procesamiento del modelo, en la Tabla III están las variables que tienen efectos fijos en el rendimiento académico en las pruebas Saber 11 a nivel departamental, la Tabla IV están las variables que tienen efectos aleatorios a nivel colegio y la Tabla V están las variables que tienen efectos aleatorios a nivel municipal en el rendimiento académico.

**Tabla III.** Variables del modelo multinivel con los efectos fijos.

Variable referencia	Efectos fijos	
	Variables	Estimate
No aplica	(Intercept)	243,27
YEAR2015	YEAR2016	9,36
	YEAR2017	9,76
	YEAR2018	4,11
ESTUGENEROF	ESTUGENEROM	11,23
ESTUHORASTRABAJONO	ESTUHORASTRABAJOSi; 20 horas o más a la semana	-12,18
	ESTUHORASTRABAJOSi; menos de 20 horas a la semana	-8,74
ESTUAPOYOGBIERNOSNO	ESTUAPOYOGBIERNOSI	6,49
No aplica	ESTUINSEINDIVIDUAL	6,59
FAMIEDUCACIONPADRE Educación profesional completa	FAMIEDUCACIONPADREEducación profesional incompleta	4,75
	FAMIEDUCACIONPADREninguno	-10,27
	FAMIEDUCACIONPADRENo sabe	-2,44
	FAMIEDUCACIONPADREPostgrado	5,92
	FAMIEDUCACIONPADREPrimaria completa	-6,81
	FAMIEDUCACIONPADREPrimaria incompleta	-6,79
	FAMIEDUCACIONPADRESecundaria (Bachillerato)completa	-4,13
	FAMIEDUCACIONPADRESecundaria (Bachillerato)incompleta	-5,61
	FAMIEDUCACIONPADRETécnica o tecnológica completa	2,26
FAMIEDUCACIONMADRE Educación profesional incompleta	FAMIEDUCACIONMADREEducación profesional incompleta	3,01
	FAMIEDUCACIONMADREninguno	-15,22

FAMIEDUCACIONMADREEducación profesional completa	FAMIEDUCACIONMADRENo sabe	-7,78
	FAMIEDUCACIONMADREPostgrado	6,52
	FAMIEDUCACIONMADREPrimaria completa	-8,72
	FAMIEDUCACIONMADREPrimaria incompleta	-9,90
	FAMIEDUCACIONMADRESecundaria (Bachillerato)completa	-4,62
	FAMIEDUCACIONMADRESecundaria (Bachillerato)incompleta	-5,73
	FAMIEDUCACIONMADRETécnica o tecnológica completa	1,72
FAMITRABAJAMADREEmpleado de nivel auxiliar o administrativo	FAMITRABAJAMADREEmpleado obrero; conductoru operario	2,96
	FAMITRABAJAMADREEmpresario	-3,38
	FAMITRABAJAMADREHogar; no trabaja o estudia	2,19
	FAMITRABAJAMADRENo sabe	-4,18
	FAMITRABAJAMADREOtra actividad u ocupación	1,27
	FAMITRABAJAMADREPensionado	-3,60
	FAMITRABAJAMADREPequeño empresario	3,05
	FAMITRABAJAMADRETrabajador por cuenta propia	1,31
FAMITRABAJAPADREEmpleado de nivel auxiliar o administrativo	FAMITRABAJAPADREEmpleado de nivel técnico o profesional	2,50
	FAMITRABAJAPADREEmpleado obrero; conductoru operario	3,61
	FAMITRABAJAPADREEmpresario	-1,99
	FAMITRABAJAPADREHogar; no trabaja o estudia	-1,75
	FAMITRABAJAPADRENo aplica	4,61
	FAMITRABAJAPADRENo sabe	3,90
	FAMITRABAJAPADREOtra actividad u ocupación	4,06
	FAMITRABAJAPADREPensionado	2,60
	FAMITRABAJAPADREPequeño empresario	5,95
	FAMITRABAJAPADREProfesional independiente	4,17
	FAMITRABAJAPADRETrabajador por cuenta propia	5,06
FAMIPERSONASHOGAR1 a 2	FAMIPERSONASHOGAR3 a 4	1,29
	FAMIPERSONASHOGAR9 o más	-1,91
	FAMICUARTOSHOGARCuatro	3,05
	FAMICUARTOSHOGARDiez o más	-2,03
	FAMICUARTOSHOGARDos	8,32
	FAMICUARTOSHOGARTres	6,31
	FAMICUARTOSHOGARUno	5,80
FAMITIENEINTERNETNo	FAMITIENEINTERNETSí	-0,71
FAMITIENECOMPUTADORNo	FAMITIENECOMPUTADORSí	1,45
FAMITIENELAVADORANo	FAMITIENELAVADORASí	-1,55
FAMITIENEAUTOMOVILNo	FAMITIENEAUTOMOVILSí	-3,25
FAMITELEVISORNo	FAMITELEVISORSí	-3,33
FAMITIENEMICROOHORNONo	FAMITIENEMICROOHORNOSí	-3,16
FAMINUMLIBROS1 A 10 LIBROS	FAMINUMLIBROS11 A 25 LIBROS	6,04
	FAMINUMLIBROS26 A 100 LIBROS	11,89
	FAMINUMLIBROSMÁS DE 100 LIBROS	14,94

Tabla IV. Muestra de las variables del modelo multinivel con los efectos aleatorios relacionado con los colegios.

Colegios	(Intercept)	NATURALEZA OFICIAL	CALENDARIO B	CALENDARIO O	ÁREA UBICACION URBANO
125001000231	12,61	-6,79	17,38	-19,07	-10,86
125035000159	8,48	-4,57	11,69	-12,83	-7,31
125040000396	-29,64	15,95	-40,84	44,81	25,52
125040000582	9,23	-4,97	12,72	-13,95	-7,95

**Tabla V.** Muestra de las variables del modelo multinivel con los efectos aleatorios relacionado con los Municipios

Municipios	(Intercept)	Calidadagua	Tasa exámenes delito sexual 1217	Porcentaje adolescentes desplazamiento forzado 1217
25001	-6,01	0,4	0,64	0,21
25035	-2,04	0,49	-0,43	0,83
25040	0,82	-0,06	-0,3	0,01

El modelamiento estadístico utilizado para establecer el efecto fijo en la predicción del rendimiento académico en las pruebas Saber 11, resultó ser simple y ajustado en concordancia con otros estudios que se han aproximada comprender y explicar resultados de pruebas estandarizadas (Benito et al., 2014; Froiland & Oros, 2014; Lisboa-Bartholo & Da-Costa, 2016; Murillo & Carrillo, 2021; Schuth et al., 2017; Yopasá & Valbuena, 2019; Zhang & Campbell, 2014).

Con respecto a las variables que inciden en el rendimiento académico, el hecho de que los padres (madre o padre) hayan terminado la educación superior en alguno de sus niveles, trabajen independiente o sean microempresarios tiene un efecto positivo en la obtención de mejores puntuaciones en las pruebas Saber 11; lo anterior, quizá obedece a la incidencia de la crianza en lo que significa ser independiente, autonomía y responsabilidad de su proceso educativo. Por otra parte, si la educación de los padres es hasta básica primaria, básica secundaria o educación media tiene una influencia negativa en las puntuaciones de las pruebas estandarizadas en Colombia, quizá porque en el contexto del estudiantado, sus padres como referentes, no han dado la importancia y el sentido necesario a lo que significa estudiar. En ese mismo sentido, el contexto físico contemporáneo tiene un efecto en las puntuaciones, así por ejemplo tener internet y un computador en casa influye positivamente en las puntuaciones de las pruebas de estado.

De igual manera, los hombres tienen mayores puntuaciones que las mujeres, lo que indica que al parecer el tipo de género tiene una incidencia en las puntuaciones obtenidas en las pruebas Saber 11, hecho que se confirma tanto en Colombia como en

otras regiones (Abdul-Aziz et al., 2015; Ali et al., 2013; Chacón-Vargas & Roldán-Villalobos, 2021; Guo et al., 2015; Kumari et al., 2018; Lau et al., 2019; Lisboa-Bartholo & Da-Costa, 2016; Orjuela, 2014).

Por otro lado, el trabajo de los jóvenes tiene un efecto negativo sobre los puntajes en las pruebas Saber 11, más aún cuando dedican más de 20 horas por semana a la actividad laboral y el número de integrantes de la familia tiene un efecto inverso, es decir, entre más integrantes tenga la familia menor son las puntuaciones logradas por los estudiantes. Lo que coincide con resultados de otras investigaciones en Colombia (Chacón-Vargas & Roldán-Villalobos, 2021) y Europa (Masci et al., 2018), algunos específicamente en el área de la matemática (Lorna et al., 2014). Asimismo, los recursos (libros, computador, lavadora, microondas, televisor, automóvil, internet, entre otros) que disponen los estudiantes tiene un efecto positivo en las puntuaciones en las pruebas Saber 11.

A nivel municipal, se obtuvo que el desplazamiento forzado tiene un efecto negativo sobre el rendimiento académico en algunos municipios, como Ospina-Gómez (2018) que menciona las implicaciones que se presenta a raíz del conflicto armado en la educación, presentándose en la población de estudio manifestaciones de aislamiento, deserción, repitencia y frustración. En la misma línea, se puede evidenciar que la naturaleza del colegio es oficial y la ubicación de este es área rural tiene un efecto negativo sobre el rendimiento escolar, hecho que coincide con otras investigaciones (Ariza et al., 2021; Froiland & Oros, 2014; Lau et al., 2019; Murillo & Carrillo, 2021; Rebai et al., 2019; Wandera et al., 2019).



## Conclusiones

El trabajo logra identificar, a partir de los datos que reposan en bases de datos abiertas, que las variables que inciden en la obtención de mejores puntuaciones en las pruebas Saber 11 están estrechamente relacionadas con las interacciones y condiciones en el contexto que tienen los estudiantes durante su formación. Así como que, los padres hayan terminado algún nivel de la educación superior tiene un efecto positivo en las puntuaciones; o el caso en que el trabajo de la madre esté asociado a una labor independiente o de microempresario tiene un efecto negativo en las valoraciones de las pruebas Saber 11, quizá es una convocatoria para reflexionar la forma en cómo la sociedad distribuye las responsabilidades a la familia.

De otro lado, el hecho de que los hombres obtengan un mayor puntaje en las pruebas Saber 11, llama la atención y se constituyen un argumento para reflexionar algunos aspectos de orden social, pero en especial la necesidad de continuar promoviendo programas orientados a que los géneros, distintos al hombre, mejoren sus condiciones en sus diferentes dimensiones que les permitan obtener mejores puntuaciones.

Finalmente, el modelo multinivel resulta ser más adecuado por su simplicidad y debido a la estructura con datos agrupados y relacionados en el contexto de la educación, lo que permitió revisar el efecto negativo de la ubicación – rural y urbano – de los colegios, el desplazamiento forzado y la naturaleza de los colegios – oficial y no oficial.

## Agradecimientos

La primera autora agradece a la Universidad Escuela Colombiana Ingeniería Julio Garavito, el segundo autor a la Universidad de Nariño y el tercer autor a la Universidad Pedagógica y Tecnológica de Colombia.

## Referencias

- Abdul-Aziz, A., Hafieza-Ismail, N., Ahmad, F., & Hassan, H. (2015). A framework for students' academic performance analysis using naïve bayes classifier. *Jurnal Teknologi*, 75(3), 13–19. <https://doi.org/10.11113/jt.v75.5037>
- Acevedo-Álvarez, R. (2008). *Los modelos jerárquicos lineales: fundamentos básicos para su uso y aplicación*. Universidad de Costa Rica.
- Ali, S., Haider, Z., Munir, F., Khan, H., & Awais, A. (2013). Factors Contributing to the Students Academic Performance: A Case Study of Islamia University Sub-Campus. *American Journal of Educational Research*, 1(8), 283–289. <https://doi.org/10.12691/education-1-8-3>
- Ariza, J., Saldarriaga, J., Reinoso, K., & Tafur, C. (2021). Tecnologías de la información y la comunicación y desempeño académico en la educación media en Colombia. *Lecturas de Economía*, 94, 47–86. <https://doi.org/10.17533/udea.le.n94a338690>
- Benito, R., Alegre, M., & González-Balletbò, I. (2014). School Segregation and Its Effects on Educational Equality and Efficiency in 16 OECD Comprehensive School Systems. *Comparative Education Review*, 58(1), 104–134. <https://doi.org/10.1086/672011>
- Chacón-Vargas, É., & Roldán-Villalobos, G. (2021). Factores que inciden sobre el rendimiento académico de los estudiantes de primer ingreso del curso Matemática General del Instituto Tecnológico de Costa Rica. *Uniciencia*, 35(1), 265–283. <https://doi.org/10.15359/ru.35-1.16>
- Contreras, D., Delgadillo, J., & Riveros, G. (2019). Is home overcrowding a significant factor in children's academic performance? Evidence from Latin America. *International Journal of Educational Development*, 67, 1–17. <https://doi.org/10.1016/j.ijedev.2019.05.001>

- org/10.1016/j.ijedudev.2019.01.006
- Correa Morales, J. C., & Salazar Uribe, J. C. (2016). *Introducción a los modelos mixtos* (1st ed.). Universidad Nacional de Colombia.
- Cvencek, D., Fryberg, S., Covarrubias, R., & Meltzoff, A. (2017). Self-Concepts, Self-Esteem, and Academic Achievement of Minority and Majority North American Elementary School Children. *Child Development*, 89(4), 1099–1109. <https://doi.org/10.1111/cdev.12802>
- Delgado Barrera, M. (2014). *La educación básica y media en Colombia: Retos en equidad y calidad*.
- Froiland, J., & Oros, E. (2014). Intrinsic motivation, perceived competence and classroom engagement as longitudinal predictors of adolescent reading achievement. *Educational Psychology*, 34(2), 119–132. <https://doi.org/10.1080/01443410.2013.822964>
- Gaete-Rivas, D., Olea, M., Meléndez-Illanes, L., Granfeldt, G., Sáez, K., Zapata-lamana, R., & Cigarroa, I. (2021). Hábitos alimentarios y rendimiento académico en escolares chilenos de quinto a octavo año básico. *Revista Chilena de Nutrición*, 48(1), 41–50. <https://doi.org/10.4067/S0717-75182021000100041>
- Gelman, A., & Hill, J. (2006). *Data Analysis Using Regression and Multinivel/Hierarchical Models*. Cambridge University Press.
- George-Reyes, C. (2020). Pruebas Estandarizadas Y Calidad De La Educacion En México. *Universidad y Sociedad Revista Científica de La Universidad de Cienfuegos*, 12(4), 418–425.
- Guo, B., Zhang, R., Xu, G., Shi, C., & Yang, L. (2015). Predicting Students Performance in Educational Data Mining. *International Symposium on Educational Technology*, 125–128. <https://doi.org/10.1109/ISET.2015.33>
- Hasan, R., Palaniappan, S., Rafiez-Abdul, A., Mahmood, S., & Uddin-Sarker, K. (2018). Student Academic Performance Prediction by using Decision Tree Algorithm. *4th International Conference on Computer and Information Sciences (ICCOINS)*, 1–5. <https://doi.org/10.1109/ICCOINS.2018.8510600>
- Hernández-Sampieri, R., Fernández-Collado, C., & Baptista-Lucio, M. (2014). *Metodología de la investigación* (J. Mares-Chacon, Ed.). McGraw Hill Intereamericana Editores S.A.
- Ibourk, A., & Amaghous, J. (2014). The performance of educational system in Morocco: A spatial analysis. *Regional and Sectoral Economic Studies*, 14(2), 109–128.
- Kumari, P., Jain, P., & Pamula, R. (2018). An Efficient use of Ensemble Methods to Predict Students Academic Performance. *4th Int'l Conf. on Recent Advances in Information Technology*. <https://doi.org/10.1109/RAIT.2018.8389056>
- Lau, E. T., Sun, L., & Yang, Q. (2019). Modelling, prediction and classification of student academic performance using artificial neural networks. *SN Applied Sciences*, 1(982). <https://doi.org/10.1007/s42452-019-0884-7>
- Lisboa- Bartholo, T., & Da-Costa, M. (2016). Evidence of a school composition effect in Rio de Janeiro public schools. *Ensaio*, 24(92), 498–521. <https://doi.org/10.1590/S0104-40362016000300001>
- López-Vera, C., Vargas-Peñaloza, M., Gómez-Rodríguez, F., Rico-Marin, J., & Escandón-Wittsack, J. (2020). *Informe Nacional de resultados del examen Saber 11 2020* (M. Bravo-Osorio & P. Cifuentes-Velasquez, Eds.). Instituto Colombiano para el Fomento de la

Educación Superior ICFES.

reving.2013.2.a04

- Lorna, A., Donders, G., Simon, E., López, O., Madden, C., Morrone, A., Puddephatt, A., Throsby, D., & Wagner, A. (2014). *Indicadores de cultura para el desarrollo. Manual metodológico. Patrimonio, Relevancia de la dimensión para la cultura y el desarrollo* (G. A. y M. Medici, Ed.). Organización de las Naciones Unidas para la Educación.
- Masci, C., Johnes, G., & Agasisti, T. (2018). Student and school performance across countries: A machine learning approach. *European Journal of Operational Research*, 269(3), 1072–1085. <https://doi.org/10.1016/j.ejor.2018.02.031>
- Maulida, J., & Kariyam. (2017). Students academic performance based on behavior. *AIP Conference Proceedings, 1911*(December 2017). <https://doi.org/10.1063/1.5016003>
- Ministerio de Educación Nacional de Colombia. (2019). Decreto 1330 de 2019. *Ministerio de Educación Nacional*, 32.
- Murillo-Torrecilla, F. (2008). Los modelos multinivel como herramienta para la investigación educativa. *Magis revista internacional de Investigación en Educación*, 1(1), 45-62. <https://www.redalyc.org/pdf/2810/281021687004.pdf>
- Murillo, J., & Carrillo, S. (2021). Incidencia de la Segregación Escolar por Nivel Socioeconómico en el Rendimiento Académico. Un Estudio desde Perú. *Archivos Analíticos de Políticas Educativas*, 29(49), 3–11. <https://doi.org/10.14507/epaa.29.5129>
- Orjuela, J. (2014). Análisis del Desempeño Estudiantil en las Pruebas de Estado para Educación Media en Colombia mediante Modelos Jerárquicos Lineales. *Ingeniería*, 18(2). <https://doi.org/10.14483/udistrital.jour>
- Ospina-Gómez, M. C. (2018). *La escuela y su implicación en el conflicto armado en Colombia*. Universidad Católica de Manizales.
- Qiu, X., & Wu, S. sheng. (2019). Contextual variables of student math proficiency and their geographic variations in Missouri. *Applied Geography*, 109, 102040. <https://doi.org/10.1016/j.apgeog.2019.102040>
- Rebai, S., Ben Yahia, F., & Essid, H. (2019). A graphically based machine learning approach to predict secondary schools performance in Tunisia. *Socio-Economic Planning Sciences*, 70(August 2018), 100724. <https://doi.org/10.1016/j.seps.2019.06.009>
- Rodríguez-De-Souza-Pajuelo, A. A., Tarazona-Luján, A. F., & Reyes-Bossio, M. (2021). Physical activity enjoyment and self-efficacy in school performance of 11-17-year-old students at educational institutions in Lima. *Journal of Physical Education and Sport*, 21(3), 2183–2189. <https://doi.org/10.7752/jpes.2021.s3278>
- Salal, Y., & Abdullaev, S. (2020). Deep Learning based Ensemble Approach to Predict Student Academic Performance: Case Study. In *2020 3rd International Conference on Intelligent Sustainable Systems (ICISS)*. <https://doi.org/10.1109/ICISS49785.2020.9316044>
- Sbroglia-Rizzotto, J., & Aniceto-França, M. (2020). Does Bullying Affect the School Performance of Brazilian Students? An Analysis Using Pisa 2015. *Child Indicators Research*, 14, 1027–1053. <https://doi.org/10.1007/s12187-020-09790-0>
- Schuth, E., Köhne, J., & Weinert, S. (2017). The influence of academic vocabulary knowledge on school performance. *Learning and Instruction*, 49, 157–165. <https://doi.org/10.1016/j>

learninstruc.2017.01.005

Shah, M., Kaistha, M., & Gupta, Y. (2019). Student Performance Assessment and Prediction System using Machine Learning. *4th International Conference on Information Systems and Computer Networks*, ISCON 2019, 386–390. <https://doi.org/10.1109/ISCON47742.2019.9036250>

Touchon, J. C. (2021). Applied Statistics with R. In *Oxford Scholarship*. <https://doi.org/10.1093/oso/9780198869979.001.0001>

Wandera, H., Marivate, V., & Sengeh, M. (2019). Predicting national school performance for policy making in South Africa. *6th International Conference on Soft Computing and Machine Intelligence, ISCFMI 2019*, 23–28. <https://doi.org/10.1109/ISCFMI47871.2019.9004323>

Xu, X., Wang, J., Peng, H., & Wu, R. (2019). Prediction of academic performance associated with internet usage behaviors using machine learning algorithms. *Computers in Human Behavior*, 98(April), 166–173. <https://doi.org/10.1016/j.chb.2019.04.015>

Yopasá, A., & Valbuena, F. (2019). *Resultados de las Pruebas iCFES en Ciencias sociales a través del análisis espacial*.

Zhang, D., & Campbell, T. (2014). An examination of the impact of teacher quality and “Opportunity Gap” on student Science Achievement in China. *International Journal of Science and Mathematics Education*, 13(3), 489–513. <https://doi.org/10.1007/s10763-013-9491-z>