Revista Maestros & Pedagogía
Revista Facultad Ciencias de la Educación
Volumen 6 Número 2 de 2024
e-ISSN: 2665-5306

UN ESTUDIO DE FACTORES SOCIO-ECONÓMICOS ASOCIADOS AL DESEMPEÑO EN EL EXAMEN SABER 11 EN COLOMBIA

A STUDY OF SOCIO-ECONOMIC FACTORS ASSOCIATED WITH PERFORMANCE IN THE SABER 11 EXAM IN COLOMBIA

UM ESTUDO DE FATORES SOCIOECONÔMICOS ASSOCIADOS AO DESEMPENHO NO EXAME SABER 11 NA COLÔMBIA

DANIEL BUITRAGO ARRIA[1]
https://orcid.org/0000-0001-5709-486X
danielbuitragoa@gmail.com

LAURA CAMILA SÁNCHEZ GARZÓN[2]
https://orcid.org/0009-0001-7117-6279
laurac.sanchez@uniagustiniana.edu.co

[1]Phd. Humanidades, Humanismo y Persona. Adscripto al Centro de Estudios Andinos Ekirawa.

[2] Adscripta a la Facultad de Humanidades, Ciencias Sociales y Educación. Universitaria Agustiniana

RESUMEN

En Colombia, el puntaje global de la prueba Saber 11 es considerado como un proxy de la posibilidad que tiene un estudiante de acceder a una educación superior de calidad y de oportunidades de movilidad social. La pregunta abordada es si el obtener altos puntajes depende mayoritariamente de cada estudiante o si existen factores socio-económicos que condicionan ciertos puntajes. Para ello, se recopiló información socio-económica de estudiantes de último grado que presentaron la prueba Saber 11 entre los años 2016-2020. Se encontró que puntajes bajos estaban claramente asociados a la ausencia de recursos educativos y a un nivel educativo bajo de los padres, mientras que los puntajes altos se asociaban sistemáticamente a altos índices de vida del entorno familiar.

Palabras clave:
Escuela secundaria, desigualdad social, examen de ingreso, norma de examen.

ABSTRACT

In Colombia, the overall score of the Saber 11 test is considered a proxy for

Cómo citar:
Fecha Recibido: 01/09/2024 **Fecha Aceptado:** 12/12/2024 **Fecha Publicado:** 19/06/2025
Buitrago Arria, D. & Sánchez Garzón, L. C. (2024). Un estudio de factores socio-económicos asociados al desempeño en el examen Saber 11 en Colombia. Maestros & Pedagogía Vol. 6(2). ppt. 8-31

Esta obra puede compartirse bajo la Licencia Creative Commons Atribución-NoComercial-CompartirIgual 4.0 Internacional.

a student's possibility of accessing quality higher education and opportunities for social mobility. The question addressed is whether obtaining high scores depends mainly on each student or if there are socio-economic factors that determine certain scores. To do this, socio-economic information was collected from senior students who took the Saber 11 test between the years 2016-2020. It was found that low scores were clearly associated with the absence of educational resources and a low educational level of the parents, while high scores were systematically associated with high indices of family life.

Keywords:
High schools, social inequality, entrance examination, examination standards.

RESUMO

Na Colômbia, a pontuação geral do teste Saber 11 é considerada uma proxy da possibilidade de um estudante acessar um ensino superior de qualidade e de oportunidades de mobilidade social. A questão colocada é se a obtenção de pontuações elevadas depende sobretudo de cada aluno ou se existem factores socioeconómicos que determinam determinadas pontuações. Para isso, foram coletadas informações socioeconômicas de alunos do último ano que realizaram a prova Saber 11 entre os anos 2016-2020. Verificou-se que pontuações baixas estavam claramente associadas à ausência de recursos educacionais e ao baixo nível educacional dos pais, enquanto pontuações altas estavam sistematicamente associadas a índices elevados de convivência familiar.

Palavras-chave:
Ensino médio, desigualdade social, vestibular, normas de exames.

INTRODUCCIÓN

En Colombia, el examen de Estado Saber 11 es uno de los ejercicios de medición de la calidad educativa más importantes del país. Esta prueba es un examen de alcance nacional diseñado e implementado por el Instituto Colombiano para la Evaluación de la Educación (ICFES), adscrito al Ministerio de Educación Nacional, que busca medir el nivel alcanzado por cada estudiante de último grado escolar en términos de las competencias básicas establecidas por el Ministerio de Educación en las áreas de matemáticas, lectura crítica, sociales y ciudadanas, ciencias naturales e inglés, que son las áreas consideradas como prioritarias para el desarrollo del país de acuerdo con el Gobierno Nacional (Decreto 5014 de 2009).

Este examen está compuesto por 254 preguntas de selección múltiple con única respuesta, divididas en cinco secciones: matemáticas (50 preguntas), lectura crítica (41 preguntas), sociales y ciudadanas (50 preguntas), ciencias naturales (58 preguntas) e inglés (55 preguntas). Para calcular el puntaje total de la prueba, se expresa primero el puntaje de cada sección en una escala de 100 (es decir, se cuenta el número de respuestas correctas, se divide por el número de preguntas totales y se multiplica por 100). Posteriormente, se calcula un promedio ponderado de todos los puntajes en donde cada sección tiene un peso de 3 con excepción de inglés, que tiene un peso de 1. Finalmente, este resultado se multiplica por 500 para que el puntaje global se exprese en una escala de 0 a 500.

Los resultados de las pruebas Saber 11 han tenido múltiples usos. Durante muchos años se tuvieron en cuenta en procesos de admisión tanto a universidades privadas como oficiales (Oyola & Echeverri, 2021). También, se interpretaron como predictores del desempeño académico en programas de educación superior (Castillo & Rodríguez, 2013; Díaz *et al.*, 2020; Gabalán-Coello & Vásquez-Rizo, 2016), e incluso para estudiar brechas educativas por género, región o nivel socioeconómico en el país (Bernal & Bernal, 2016; Meisel Roca, 2019; Quintero & Vélez, 2016).

Sin embargo, los estudios de más largo alcance sobre la prueba Saber 11 la han entendido como un proxy no sólo del desempeño académico en la educación superior, sino de las condiciones laborales del graduado universitario. De esta manera, puntajes bajos en la prueba Saber 11 se han visto asociados a desempeños bajos en la educación superior y, consecuentemente, a salarios bajos en el mercado laboral (Cuenca, 2016). De aquí la motivación por estudiar los factores que son determinantes en la obtención de ciertos niveles de puntaje en las pruebas Saber 11, de tal forma que se permita una discusión de las implicaciones de estos factores en la elaboración de políticas públicas y sociales para el país.

En este ámbito del estudio de los factores que influyen en el puntaje de las pruebas Saber 11 existen numerosos estudios. Buena parte de los más recientes incluyen herramientas de Machine Learning para predecir el puntaje que van a obtener los estudiantes los estudiantes a partir de las variables socio-económicas (Peña & González, 2022; Pereira *et al.*, 2019, 2020; Vargas & Ardila, s. f.). A partir de todos estos estudios se puede afirmar que, en efecto, existe una importante asociación entre las variables socio-económicas del estudiante y el puntaje que obtendrá en la prueba Saber 11 que, como ya se mencionó, no sólo impacta su posibilidad de acceso a la educación superior, sino también su desempeño académico en ella (de poder acceder) e incluso sus futuras condiciones laborales.

A pesar de lo avanzado en identificar estas variables socio-económicas como

determinantes en el futuro académico y laboral del estudiante colombiano, poco se ha estudiado la manera en que estos factores afectan el puntaje. En otras palabras, partiendo de la premisa de un presupuesto público limitado, ¿cuáles son las variables socio-económicas que requieren una inyección de recursos más urgente para disminuir, de manera más efectiva, la brecha del puntaje en la prueba? Esta es la pregunta que busca discutirse a partir de la presente investigación, usando para ello una pregunta más general: ¿cuáles son los factores socio-económicos que exhiben la mayor asociación tanto con los mayores como con los menores puntajes del examen Saber 11? Esta es la pregunta problema de esta investigación. Para abordarla, se tomó la información socio-económica recopilada por el ICFES en una ventana de 5 años. Posteriormente, se aplicaron las técnicas de Análisis de Correspondencias Múltiple y Análisis de Correspondencias Conjuntas. La primera es una técnica de estadística multivariada que ha sido usada en numerosos estudios para indagar las asociaciones entre múltiples variables cualitativas (Costa *et al.*, 2013). Así mismo, es una técnica muy utilizada para estudios de variables cualitativas específicas a lo largo de varios períodos de tiempo bajo enfoques longitudinales como la presente investigación (Lopes de Oliveira *et al.*, 2021). Adicionalmente, esta técnica tiene la ventaja de ser robusta en el caso de tener un número grande de variables y un número grande de registros, a diferencia de técnicas como la prueba de chi-cuadrado, que pierde confiabilidad con grandes conjuntos de datos (Lin *et al.*, 2013). Por otro lado, el Análisis de Correspondencias Conjuntas es una forma mejorada del Análisis de Correspondencias Múltiples que emplea un algoritmo que optimiza las inercias y la dimensionalidad del análisis, incrementando así la validez de las conclusiones (Greenacre, 2005).

En el ámbito educativo, aunque la función principal del Análisis de Correspondencias Múltiple es hallar asociaciones entre varias variables cualitativas, se ha usado para identificar perfiles (Ogunsakin *et al.*, 2021) o para medir el impacto de ciertos rasgos cualitativos sobre otra variable objeto (López *et al.*, 2022). En el caso colombiano, quizás el estudio más cercano que utiliza las técnicas de Análisis de Correspondencias Múltiple y Análisis de Correspondencias Conjuntas a las variables del examen Saber 11 es el de Cuenca (2016). No obstante, una ventaja del presente estudio frente al de Cuenca es que no sólo se busca establecer asociaciones entre factores socio-económicos y el desempeño académico (algo ya ampliamente estudiado, como se mencionó anteriormente), sino que se busca establecer, de manera específica, cuáles de esos factores afectan, tanto positiva como negativamente, el puntaje de la prueba. Para esto, se usará una variante de esta técnica de Análisis de Correspondencias Múltiple que permite advertir cuáles son los factores cualitativos más influyentes sobre una variable cuantitativa, en este caso, el puntaje de la prueba Saber 11, y con ello, permitir al Gobierno Nacional discutir priorizaciones en la asignación de recursos públicos orientados a disminuir la brecha del puntaje en la prueba.

MÉTODO

A. Recolección y preparación de los datos

El estudio a nivel nacional se llevó a cabo utilizando las bases de datos de los resultados de las pruebas Saber 11, disponibles en el portal del Icfes. Estas bases no sólo contienen los puntajes obtenidos por cada estudiante en cada área disciplinar, sino que también recogen información socio-económica del entorno familiar de quien presenta la prueba. En total, en cada cohorte, el Icfes recoge información de entre 74 y 78 variables entre puntajes, información del colegio, información socio-económica e identificadores de municipio, departamento, etc. Las variables de interés para el Icfes pueden cambiar de un año a otro, razón por la cual se hizo necesario un estudio de enfoque cuantitativo, de tipo longitudinal y de corte descriptivo, en donde se abarcaron las dos cohortes de presentación del examen durante los años 2016 al 2020. Para cada año, se juntaron ambas cohortes en una sola base y, partiendo de las variables recolectadas por el Icfes, se eliminaron variables de identificación y las relativas al puntaje cuantitativo obtenido por cada estudiante en la prueba, dejando así únicamente las variables de índole socio-económica (todas cualitativas) y el percentil alcanzado por el estudiante con base en el puntaje total obtenido en la prueba. A partir del percentil, se construyó una nueva variable cualitativa, el decil del puntaje, como una agrupación del percentil en conjuntos de a 10. Finalmente, dado que la base queda constituida únicamente por variables cualitativas, se reemplazaron los valores faltantes por la moda de cada variable.

B. Población y muestra

En Colombia, el calendario académico de los colegios se clasifica en A (aquellos colegios que inician actividades desde principios de febrero y finalizan los últimos días de noviembre) o B (aquellos que comienzan sus labores académicas en septiembre y terminan en junio del siguiente año). Debido a esto, quienes presentan el examen Saber 11 en la primer cohorte son, mayoritariamente, estudiantes de colegios de calendario B, que son minoría, mientras que aquellos que presentan el examen en la segundo cohorte son, en gran parte, estudiantes de colegios de calendario A. Debido a lo anterior, se reunió la información de ambas cohortes y se consolidó la información por años. Esto resultó en bases de datos muy extensas, con entre 74 y 78 variables y, en promedio, unos 549.901 registros por cada año. Para reducir la complejidad de las bases de datos y economizar en recursos computacionales, se aplicó un algoritmo de selección de variables que se explicará en la siguiente sección. Acto seguido, se aplicó un muestreo aleatorio simple sobre el total de la población de estudiantes de cada año siguiendo la metodología desarrollada en Singh (2003). El total de la población de cada año y el tamaño de muestra calculado, junto con el correspondiente margen de error, se muestran en la Tabla 1.

Tabla 1
Población y muestra de cada año de presentación de la prueba Saber 11

Año/Item	2016	2017	2018	2019	2020
Población	561 270	559 254	562 461	546 212	520 308
Tamaño muestral mínimo	9442	9442	9443	9438	9430
Tamaño muestral tomado	12 134	13 758	10 098	10 000	12 867
Error estándar relativo	1 %	1 %	1 %	1 %	1 %

Fuente: *elaboración propia.*

C. Procesamiento y análisis de los datos

A partir de la necesidad de identificar las variables de mayor cercanía con el decil del puntaje obtenido por cada estudiante, y teniendo en cuenta el gran número de variables de la base de datos, para cada población de cada año, de manera separada, se corrió, con ayuda del software Weka, el algoritmo de selección de variables Correlation-Based Feature Selection (Hall, 1999) con el propósito de hallar el conjunto de variables que guardaran la mayor asociación con la variable Decil. Las variables detectadas por este algoritmo se muestran en la Tabla 2. Posteriormente, una vez seleccionadas las muestras aleatorias para cada año, se corrió un Análisis de Correspondencias Múltiple usando los paquetes FactomineR y Factoshiny del software R con las variables seleccionadas por el algoritmo. Esto con el objeto de detectar tendencias o regularidades a lo largo de la ventana de 5 años analizada.

RESULTADOS OBTENIDOS

Una vez aplicado el Análisis de Correspondencias Múltiple, se encontró que, en la mayoría de casos, las dos primeras dimensiones recogían la mayor parte de la inercia (en total un 8,8 %), mientras que la inercia restante se repartía casi equitativamente en el resto de dimensiones (ver Tabla 3). Posteriormente, el Análisis de Correspondencias Conjuntas optimizó este esquema, mostrando que al menos el 70 % de la inercia era recogida por las primeras dos dimensiones en el período temporal estudiado.

Tabla 2

Variables seleccionadas por el algoritmo CBFS en cada año de presentación analizado de la prueba Saber 11

Año	2016	2017	2018	2019	2020	Variables en común a todos los años estudiados
Variables seleccionadas y sus categorías.	- Nivel educativo del padre (primaria incompleta, primaria completa, secundaria incompleta, secundaria completa, técnica o tecnológica incompleta, profesional incompleta, profesional completa,postgrado, ninguna, no sabe, no aplica).	- Nivel educativo del padre (primaria incompleta, primaria completa, secundaria incompleta, secundaria completa, técnica o tecnológica incompleta, profesional incompleta, profesional completa,postgrado, ninguna, no sabe, no aplica).	- Nivel educativo del padre (primaria incompleta, primaria completa, secundaria incompleta, secundaria completa, técnica o tecnológica incompleta, profesional incompleta, profesional completa,postgrado, ninguna, no sabe, no aplica).	- Nivel educativo del padre (primaria incompleta, primaria completa, secundaria incompleta, secundaria completa, técnica o tecnológica incompleta, profesional incompleta, profesional completa,postgrado, ninguna, no sabe, no aplica).	- Nivel educativo del padre (primaria incompleta, primaria completa, secundaria incompleta, secundaria completa, técnica o tecnológica incompleta, profesional incompleta, profesional completa,postgrado, ninguna, no sabe, no aplica).	- Nivel educativo del padre (primaria incompleta, primaria completa, secundaria incompleta, secundaria completa, técnica o tecnológica incompleta, profesional incompleta, profesional completa,postgrado, ninguna, no sabe, no aplica).
	- Nivel educativo de la madre (primaria incompleta, primaria completa, secundaria incompleta, secundaria completa, técnica o tecnológica incompleta, profesional incompleta, profesional completa,postgrado, ninguna, no sabe, no aplica).	- Nivel educativo de la madre (primaria incompleta, primaria completa, secundaria incompleta, secundaria completa, técnica o tecnológica incompleta, profesional incompleta, profesional completa,postgrado, ninguna, no sabe, no aplica).	- Nivel educativo de la madre (primaria incompleta, primaria completa, secundaria incompleta, secundaria completa, técnica o tecnológica incompleta, profesional incompleta, profesional completa,postgrado, ninguna, no sabe, no aplica).	- Nivel educativo de la madre (primaria incompleta, primaria completa, secundaria incompleta, secundaria completa, técnica o tecnológica incompleta, profesional incompleta, profesional completa,postgrado, ninguna, no sabe, no aplica).	- Nivel educativo de la madre (primaria incompleta, primaria completa, secundaria incompleta, secundaria completa, técnica o tecnológica incompleta, profesional incompleta, profesional completa,postgrado, ninguna, no sabe, no aplica).	- Nivel educativo de la madre (primaria incompleta, primaria completa, secundaria incompleta, secundaria completa, técnica o tecnológica incompleta, profesional incompleta, profesional completa,postgrado, ninguna, no sabe, no aplica).
	- Estrato de la vivienda (1, 2, 3, 4, 5 o 6).	- Estrato de la vivienda (1, 2, 3, 4, 5 o 6).	- Estrato de la vivienda (1, 2, 3, 4, 5 o 6).	- Estrato de la vivienda (1, 2, 3, 4, 5 o 6).	- Estrato de la vivienda (1, 2, 3, 4, 5 o 6).	- Estrato de la vivienda (1, 2, 3, 4, 5 o 6).
	- Número de libros en el hogar (de 0 a 10, de 11 a 25, de 26 a 100, más de 100).	- Número de libros en el hogar (de 0 a 10, de 11 a 25, de 26 a 100, más de 100).	- Número de libros en el hogar (de 0 a 10, de 11 a 25, de 26 a 100, más de 100).	- Número de libros en el hogar (de 0 a 10, de 11 a 25, de 26 a 100, más de 100).	- Número de libros en el hogar (de 0 a 10, de 11 a 25, de 26 a 100, más de 100).	- Número de libros en el hogar (de 0 a 10, de 11 a 25, de 26 a 100, más de 100).
	- Jornada del colegio (mañana, tarde, noche, completa, sabatina, única).	- Jornada del colegio (mañana, tarde, noche, completa, sabatina, única).	- Jornada del colegio (mañana, tarde, noche, completa, sabatina, única).	- Jornada del colegio (mañana, tarde, noche, completa, sabatina, única).	- Jornada del colegio (mañana, tarde, noche, completa, sabatina, única).	- Jornada del colegio (mañana, tarde, noche, completa, sabatina, única).

- Acceso a computador en el hogar (Si, No).	- Acceso a internet en el hogar (Si, No).	- Género (M, F).	- Tipo de labor del padre (trabajador por cuenta propia, profesional independiente, pequeño empresario, pensionado, otra actividad, hogar, empresario, empleado obrero, empleado de nivel técnico o profesional, empleado de nivel directivo, empleado de nivel auxiliar, empleado directivo).	- Acceso a internet en el hogar (Si, No).
- Ingreso familiar (Menos de 1 SMLV, entre 1 y menos de 2 SMLV, entre 2 y menos de 3 SMLV, entre 3 y menos de 5 SMLV, entre 5 y menos de 7 SMLV, entre 7 y menos de 10 SMLV, 10 o más SMLV).	- Acceso a horno microondas o gas (Si, No).	- Acceso a internet en el hogar (Si, No).	- Tipo de labor de la madre (trabajador por cuenta propia, profesional independiente, pequeño empresario, pensionado, otra actividad, hogar, empresario, empleado obrero, empleado de nivel técnico o profesional, empleado de nivel directivo, empleado de nivel auxiliar, empleado directivo).	- Acceso a computador en el hogar (Si, No).
	- Consumo de leche y derivados (Nunca o rara vez comemos eso, 1 o 2 veces por semana, 3 a 5 veces por semana, Todos o casi todos los días).	- Acceso a computador en el hogar (Si, No).	- Acceso a internet en el hogar (Si, No).	- Consumo de leche y derivados (Nunca o rara vez comemos eso, 1 o 2 veces por semana, 3 a 5 veces por semana, Todos o casi todos los días).

- Consumo de proteína de origen animal (Nunca o rara vez comemos eso, 1 o 2 veces por semana, 3 a 5 veces por semana, Todos o casi todos los días).

- Tipo de labor del padre (trabajador por cuenta propia, profesional independiente, pequeño empresario, pensionado, otra actividad, hogar, empresario, empleado obrero, empleado de nivel técnico o profesional, empleado de nivel directivo, empleado de nivel auxiliar, empleado directivo).

- Tipo de labor de la madre (trabajador por cuenta propia, profesional independiente, pequeño empresario, pensionado, otra actividad, hogar, empresario, empleado obrero, empleado de nivel técnico o profesional, empleado de nivel directivo, empleado de nivel auxiliar, empleado directivo).

- Consumo de leche y derivados (Nunca o rara vez comemos eso, 1 o 2 veces por semana, 3 a 5 veces por semana, Todos o casi todos los días).

- Acceso a computador en el hogar (Si, No).

- Consumo de leche y derivados (Nunca o rara vez comemos eso, 1 o 2 veces por semana, 3 a 5 veces por semana, Todos o casi todos los días).

- Consumo de proteína de origen animal (Nunca o rara vez comemos eso, 1 o 2 veces por semana, 3 a 5 veces por semana, Todos o casi todos los días).

- Horas de dedicación al internet (30 minutos o menos, entre 30 y 60 minutos, entre 1 y 3 horas, más de 3 horas, no navega en internet).

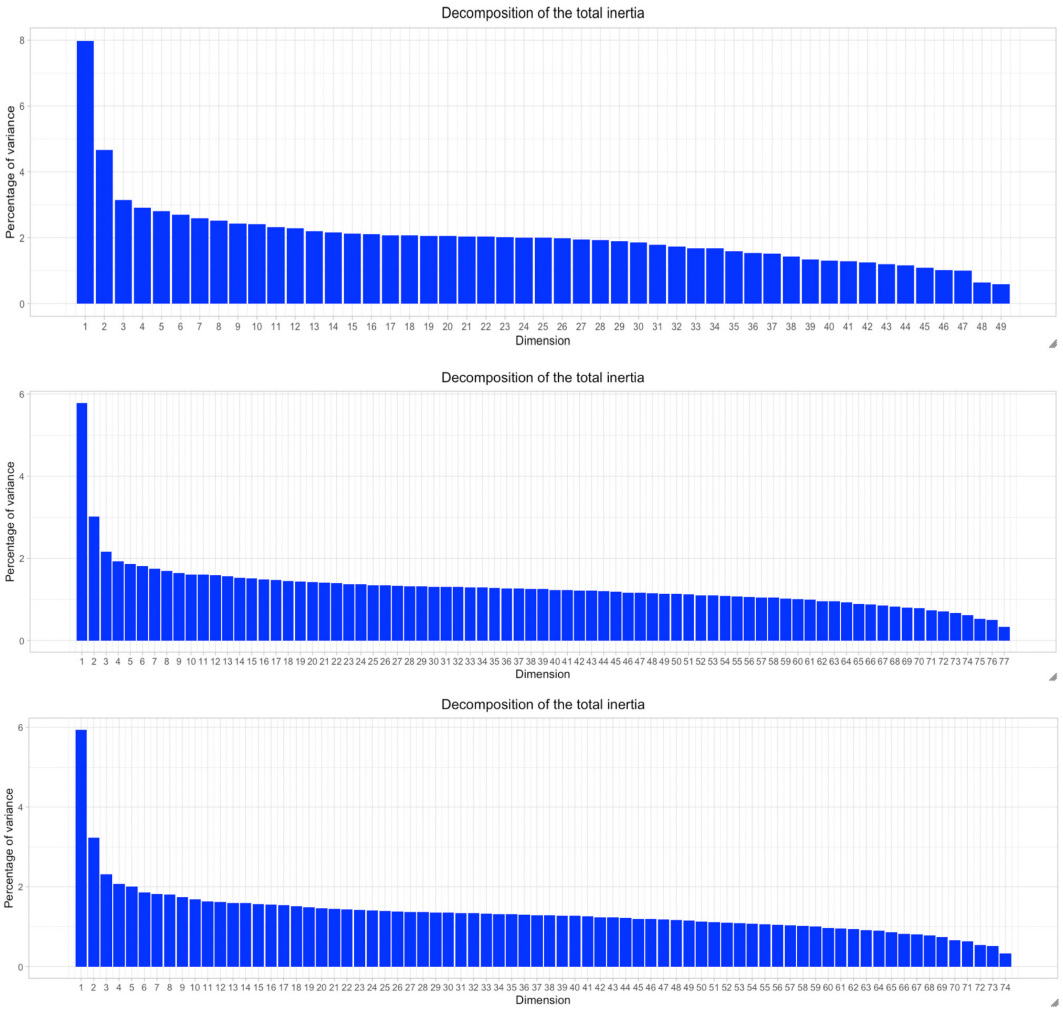
- Naturaleza del colegio (oficial, no oficial).

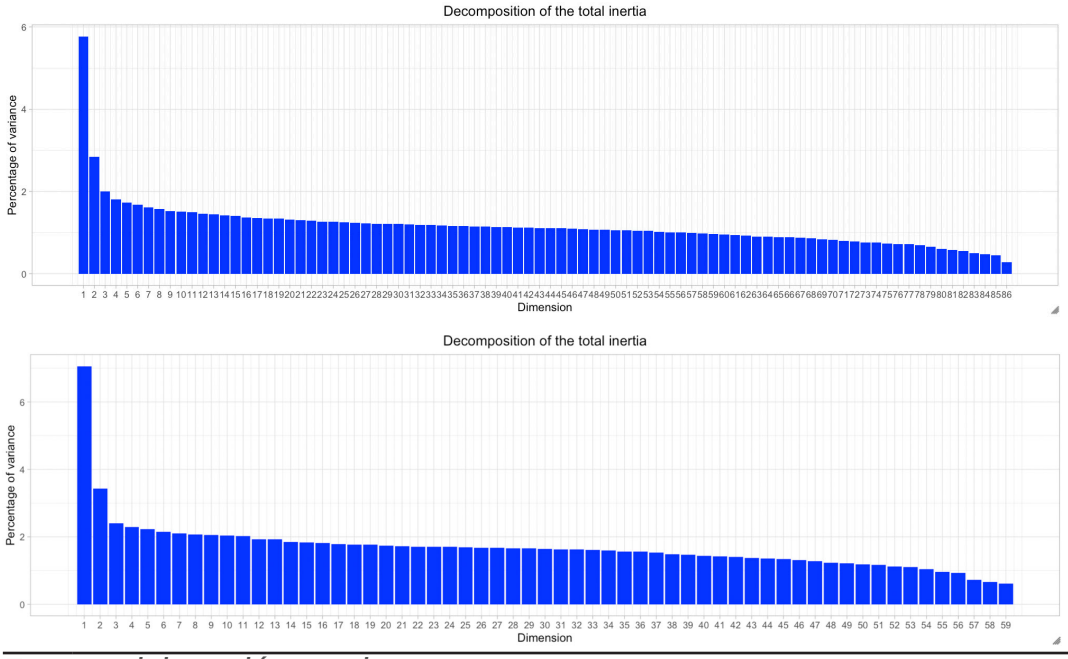
	- Horas de dedicación al internet (30 minutos o menos, entre 30 y 60 minutos, entre 1 y 3 horas, más de 3 horas, no navega en internet). - Naturaleza del colegio (oficial, no oficial).

Fuente: *elaboración propia.*

Tabla 3

Descomposición de la inercia por año

Año	Descomposición de la inercia con ACM
	



Fuente: *elaboración propia.*

Tabla 4

Descomposición de la inercia por año usando la técnica ACC

Año	Descomposición de la inercia con ACC
2016	Principal inertias (eigenvalues): dim value % cum% scree plot 1 0.172860 68.7 68.7 ****************** 2 0.033896 13.5 82.1 **** 3 0.005951 2.4 84.5 * 4 0.003687 1.5 86.0
2017	Principal inertias (eigenvalues): dim value % cum% scree plot 1 0.098607 68.5 68.5 ****************** 2 0.014465 10.1 78.6 *** 3 0.003617 2.5 81.1 * 4 0.001947 1.4 82.5
2018	Principal inertias (eigenvalues): dim value % cum% scree plot 1 0.115325 66.9 66.9 ****************** 2 0.019321 11.2 78.1 *** 3 0.005013 2.9 81.0 * 4 0.002854 1.7 82.6 5 0.002311 1.3 84.0
2019	Principal inertias (eigenvalues): dim value % cum% scree plot 1 0.080104 68.9 68.9 ******************** 2 0.010624 9.1 78.0 *** 3 0.002692 2.3 80.4 * 4 0.001545 1.3 81.7

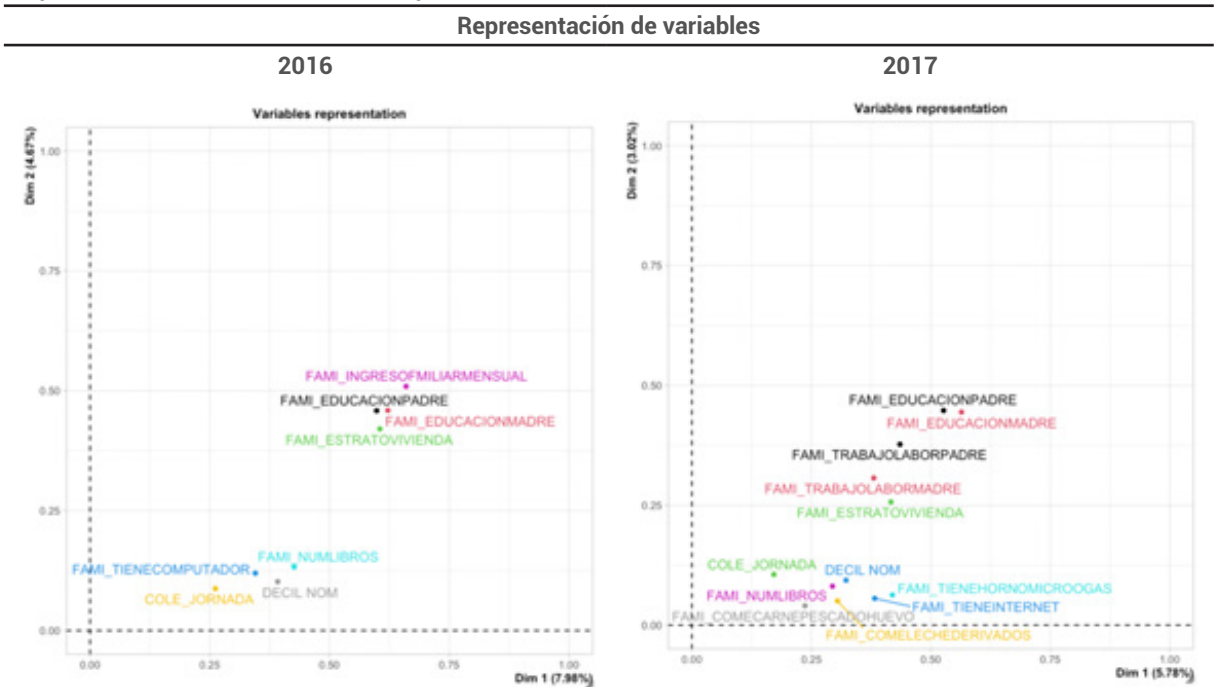
2020	Principal inertias (eigenvalues): dim value % cum% scree plot 1 0.082856 75.6 75.6 ********** 2 0.008703 7.9 83.6 ** 3 0.001450 1.3 84.9

Fuente: *elaboración propia.*

Por otro lado, la representación gráfica de las variables sugiere que la variable DECIL NOM (el decil del puntaje de la prueba Saber 11) guarda, sistemáticamente, una asociación con el número de libros en el hogar y la jornada del colegio (ver Figura 1).

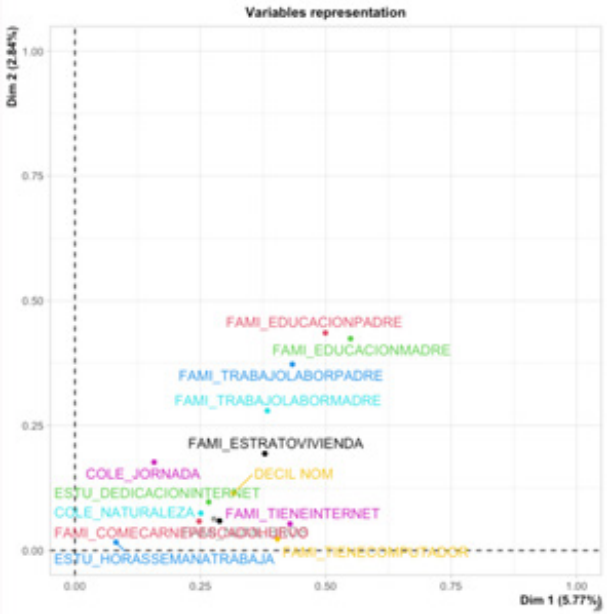
Figura 1

Representación de variables por año

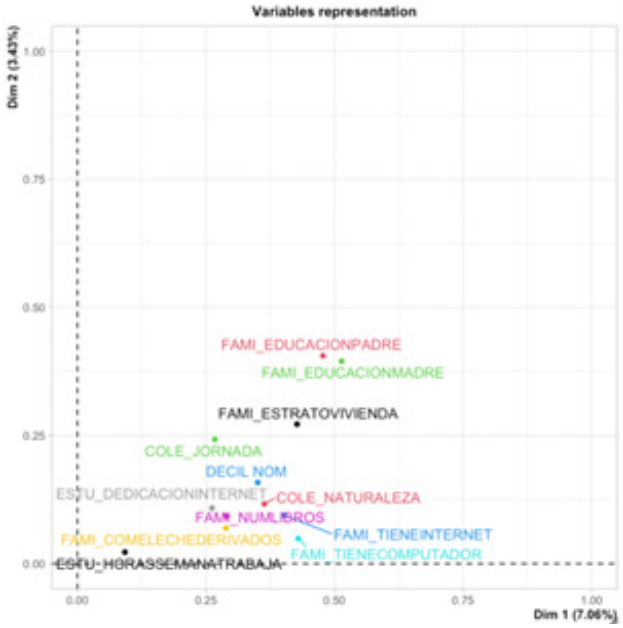


2018

2019



2020



Fuente: *elaboración propia con el paquete Factoshiny de R.*

Para profundizar en las asociaciones encontradas, se estudiaron, a continuación, los diagramas de correspondencias múltiple de cada año. Se encontró que la distribución de los valores de las variables seguía, siempre, una forma de U, lo que es consistente con la distribución de la inercia en los ejes, donde la primera dimensión (la horizontal en los diagramas) captura la mayor parte de la inercia (casi el doble que la que captura la segunda dimensión). Adicionalmente, se

encontraron patrones de asociaciones entre los valores de las variables en todos los años estudiados. En primer lugar, se encuentra persistentemente una asociación entre los deciles más bajos (primer y segundo decil) y un bajo nivel educativo de los padres (primaria incompleta o completa), como se puede ver en la Figura 2.

Otro patrón significativo y constante encontrado en los diagramas fue la asociación entre el número de libros en el hogar con el decil del puntaje. Un bajo número de libros se asocia con un decil bajo, mientras que el decil más alto se encontró sistemáticamente asociado con estudiantes que reportan tener más de 100 libros en el hogar. Con respecto al estrato, el estrato 1 aparece asociado al decil más bajo, mientras que el estrato 3 se asocia al decil 9 y el estrato 4 se muestra como el más cercano al decil 10. Finalmente, se observó en los diagramas que los puntajes en el decil más alto se asociaban a colegios de jornada completa.

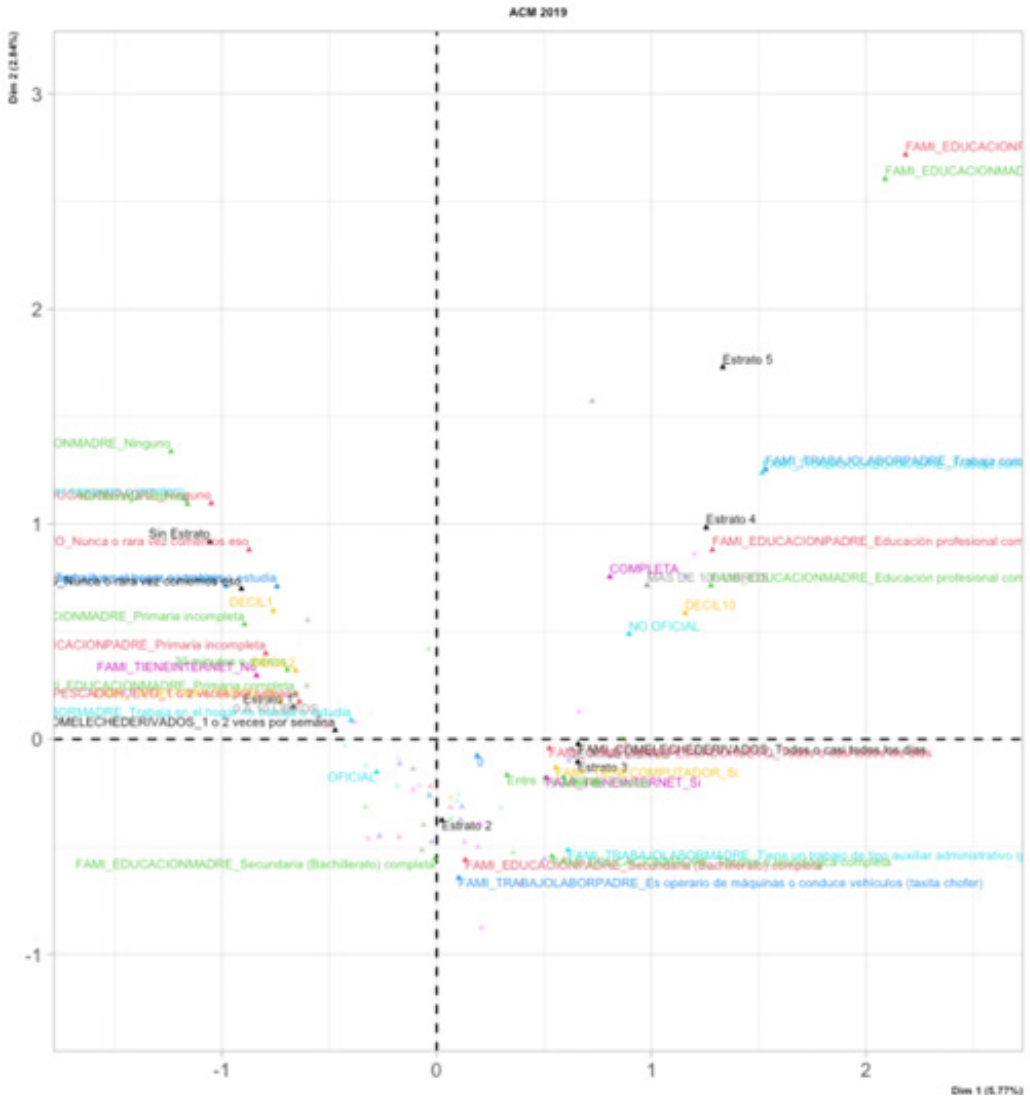
Una de las ventajas que tiene el análisis de correspondencias múltiple es que, más allá de identificar asociaciones entre variables, permite identificar exactamente qué valores de cuál variable se asocian con otros valores de otras variables. Esto, a su vez, permite construir perfiles alrededor de cúmulos de asociaciones, como los descritos anteriormente. De esta manera, a partir de lo observado en los diagramas de todos los años bajo estudio se pueden identificar claramente tres perfiles de estudiantes: un primer tipo de estudiante que es hijo de padres con bajos niveles educativos (a lo sumo, primaria incompleta) carece de recursos de estudio como computador o internet, no estudia en colegios de jornada completa y posee máximo 10 libros en su hogar. Este estudiante es el que obtiene los puntajes más bajos en la prueba Saber 11.

El segundo tipo de estudiante vive en sectores de estrato 3, tiene acceso a recursos educativos como computador e internet y a mayor número de libros en casa. Este tipo de estudiantes obtienen puntajes en la prueba Saber 11 dentro del decil 9.

Finalmente, el tercer tipo de estudiante vive en sectores de estrato 4, es hijo de padres con formación universitaria completa, estudia en colegios de jornada completa y tiene a su disposición una cantidad importante de libros en casa. Este estudiante obtiene los más altos puntajes en la prueba Saber 11.

Figura 2

Diagrama del Análisis de Correspondencias Múltiple para el año 2019

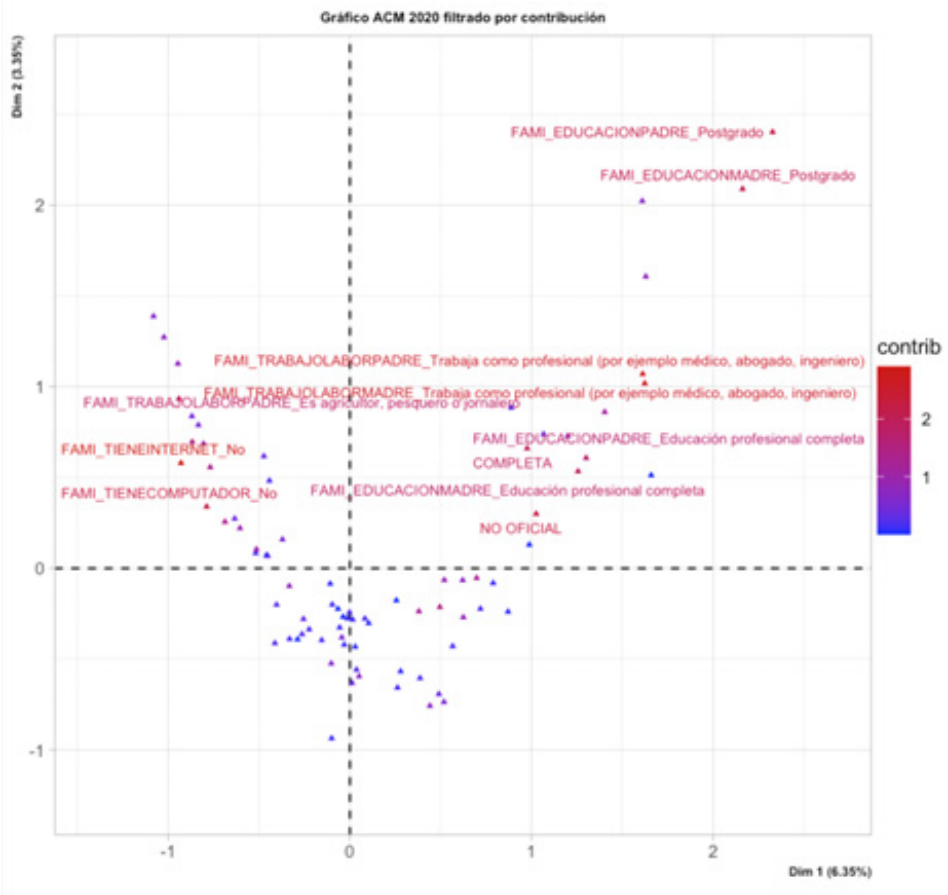


Fuente: *elaboración propia usando el paquete Factoshiny de R.*

Este perfilamiento fue complementado corriendo nuevamente otro análisis de correspondencias múltiple tomando como variable suplementaria el puntaje global (Figura 4). Esto con el propósito de identificar, puntualmente, cuáles factores contribuyen positivamente al puntaje y cuáles negativamente. El gráfico de este análisis puede verse en la Figura 3. Allí, los factores que aparecen en el cuadrante derecho son aquellos que contribuyen positivamente al puntaje de la prueba, y cuanto más cálido el color, más fuerte su contribución. Por otro lado, los factores que aparecen en el cuadrante izquierdo son los que contribuyen negativamente al puntaje.

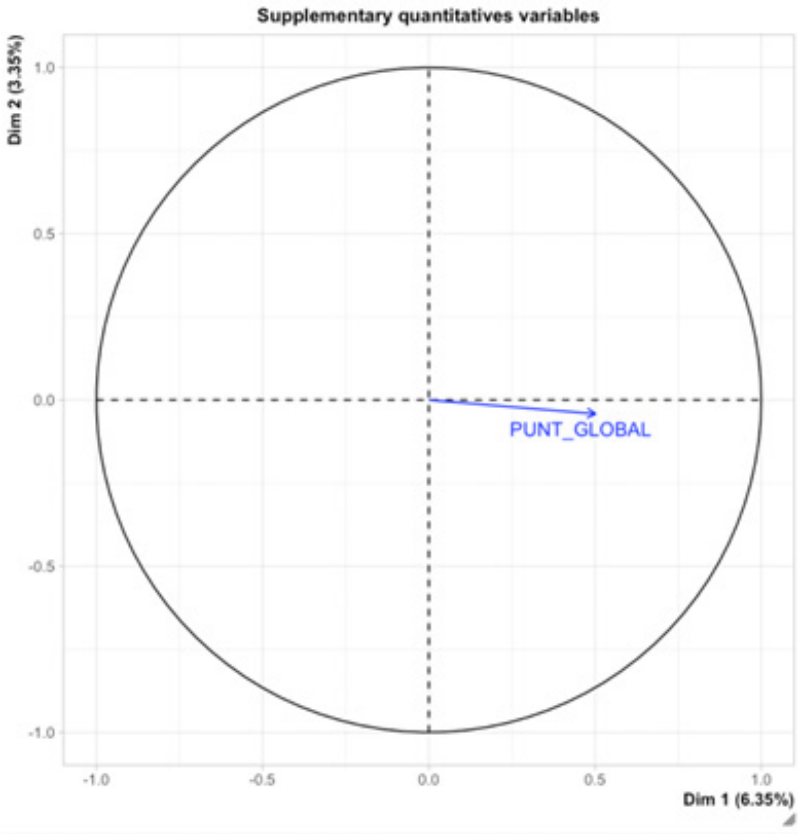
Figura 3

Gráfico del análisis de correspondencias múltiple tomando el puntaje global como variable suplementaria



Fuente: *elaboración propia usando el software R.*

Figura 4
Diagrama del puntaje como variable suplementaria



Fuente: *elaboración propia a partir del paquete Factoshiny de R.*

Para sintetizar estos resultados, se construyó una tabla (ver Tabla 5) en la que se muestran, en orden descendente a su contribución, los factores que impactan tanto positiva como negativamente en el puntaje de la prueba. Allí se puede observar que el factor que más negativamente impacta el puntaje de la prueba es el bajo nivel educativo de los padres, seguido del tipo de labor que desempeñan y el acceso limitado a recursos educativos como internet o libros. Por otro lado, el factor que más impacta positivamente el puntaje es, de manera similar, el alto nivel educativo de los padres, el tipo de labor, seguido del alto estrato socio-económico de la vivienda y el amplio acceso a recursos educativos.

Tabla 5
Factores que impactan positiva y negativamente el puntaje de la prueba Saber 11

Factores que más influyen negativamente en el puntaje global de la prueba Saber11	Factores que más influyen positivamente en el puntaje global de la prueba Saber11
• Bajo nivel educativo de los padres (primaria o ninguno). • Tipo de labor que desempeñan los padres (agricultor, jornalero). • No tener acceso a internet. • Bajo consumo de proteína animal o derivados lácteos. • No tener acceso a computador o libros. • Jornada del colegio (noche o sabatina)	• Alto nivel educativo de los padres (postgrado). • Tipo de labor que desempeñan los padres (profesional) • Estrato de la vivienda (5 y 6). • Tener acceso a libros (Más de 100). • Tipo de colegio privado. • Jornada del colegio (completa).

Fuente: *elaboración propia.*

DISCUSIÓN

Entender qué factores socioeconómicos afectan los resultados de las pruebas Saber 11 es sustancial para determinar cuál es la guía de acción que deberán tomar las políticas públicas si es que se quiere generar una disminución en las brechas educativas en Colombia. Sin embargo, los resultados que arroja la presente investigación indican que la problemática en cuestión no obedece simplemente a falencias conceptuales que exhiben estudiantes de último grado de escolaridad y que puedan solventarse de manera particular e individual, sino que las diferencias en los puntajes de la prueba son más bien el reflejo de una problemática estructural de naturaleza socio-económica del país. Esto implica que una política que impacte de manera significativa en la disminución de la brecha educativa en Colombia necesariamente debe apuntarle al mejoramiento de las condiciones socio-económicas de las poblaciones más vulnerables del territorio en aspectos clave como el acceso a recursos educativos y a una jornada escolar completa.

En este sentido, los factores encontrados en el presente estudio, que están asociados con ciertos puntajes de la prueba Saber 11, se pueden agrupar fundamentalmente en dos ejes no necesariamente disyuntos: (1) capital económico y cultural y (2) jornada escolar. A continuación se discuten brevemente estos ejes.

1. Capital económico y cultural

Como se puede observar en la Figura 2, un alto nivel educativo de los padres está asociado con un alto estrato socioeconómico. Esto puede explicarse desde la Teoría del Capital Humano, en donde se predice que a mayor nivel educativo mayores ingresos salariales. Estos mayores ingresos familiares permiten que los hijos de estas familias no sólo tengan acceso a una jornada escolar completa en colegios privados, sino que probablemente tendrán acceso a una mayor cantidad de libros en casa producto del alto nivel cultural de sus padres y a recursos educativos en general (como computador personal e internet de alta velocidad). Sumado a lo anterior, un alto nivel educativo de los padres abre la posibilidad de una mayor transmisión cultural que se materializa en información que se transfiere a los hijos (Varvarigos, 2020), como conocimiento sobre métodos efectivos de estudio o técnicas para el dominio de ciertas temáticas, o sobre la estructura de la prueba Saber 11, etc. Por ende, un alto nivel educativo en los padres es indicador de un alto nivel de capital cultural, que se materializa en mayores oportunidades para que un estudiante logre un alto puntaje en las pruebas de estado. A este respecto, cabe añadir que, según lo encontrado en el estudio, los niveles educativos del padre y de la madre suelen estar asociados, es decir, por lo general ambos exhiben el mismo nivel educativo (véase Figura 3).

Con respecto a la labor desempeñada por los padres, parece existir una diferenciación ordinal a partir de la cualificación de los oficios. De esta manera, hijos de padres que se dedican a labores no cualificadas como la agricultura tendrán menor probabilidad de alcanzar puntajes altos. Por un lado, esto puede tener que ver con la marginalización socio-económica que impacta en la calidad educativa. En Colombia, históricamente las zonas rurales del país han tenido múltiples dificultades en el acceso a la educación y, en general, a los recursos educativos (Echavarría, 2018). Si el estudiante, hijo de agricultores, vive en una zona rural, no sólo encontrará dificultades para desplazarse a un centro educativo, sino que, además, en muchas ocasiones la necesidad de ayudar a sus padres en las labores del campo supera incluso la voluntad de asistir a las clases. De ahí los altos índices de deserción en las escuelas rurales (Benitez *et al.*, 2020). Por otro lado, la precariedad financiera y la incertidumbre económica de labores como la agricultura en Colombia obstaculizan el acceso privado a recursos educativos como computadores, libros o internet, tal como se constata en la Figura 3.

2. Jornada escolar

El análisis elaborado en el presente estudio arrojó que la jornada escolar es otro factor determinante en el puntaje de la prueba Saber 11 (ver Figura 2). En particular, el análisis muestra que estudiantes que provienen de un colegio con jornada escolar completa (al menos 35 horas semanales para estudiantes de básica secundaria y media) tienden a alcanzar mayores puntajes en el examen de Estado. Esto en contraste con otro tipo de jornadas de menor intensidad

como la media jornada matutina o vespertina (30 horas semanales) o la nocturna (17,5 horas semanales) o la sabatina-dominical (10 horas semanales), que no parecen tener una contribución significativa ni a altos ni a bajos puntajes de la prueba (ver Figura 3). Sin embargo, es preciso notar que la contribución de la jornada completa a los altos puntajes de la prueba aparece muy cercana a otro factor determinante: el carácter no oficial de las instituciones (ver Figura 3). Esto no es casual. De hecho, es consistente con otras investigaciones como la de Bonilla-Mejía (2014), en la que se muestra que la jornada completa es predominante en las instituciones no oficiales (privadas). Incluso, la jornada en algunos colegios privados suele superar en intensidad a la jornada completa de los colegios oficiales, llegando en algunos casos hasta 40 horas semanales. Esto nuevamente pone sobre la mesa la importancia del papel de la estructura socio-económica en la brecha educativa, esto en la medida en que, en Colombia, los colegios privados son un bien que no está al alcance de buena parte de la población.

CONCLUSIONES

El presente estudio tenía dos objetivos centrales: el primero, determinar hasta qué punto los puntajes de la prueba Saber 11 podían estar asociados a factores socio-económicos del núcleo familiar del estudiante y, de ser así, el segundo objetivo le apuntaba a determinar cuáles son los factores más influyentes tanto en los puntajes altos como en los bajos del examen de Estado. Con esto, lo novedoso que aporta el presente estudio es el esclarecimiento de esos factores puntuales que están impactando con mayor fuerza tanto positivamente como negativamente los puntajes de la prueba Saber 11. Este esclarecimiento permite optimizar la asignación de recursos públicos orientados a la disminución de la brecha educativa en Colombia.

Frente a los objetivos planteados, se encontró que, como lo sostienen muchas otras investigaciones, efectivamente existen factores socio-económicos que están asociados a los puntajes de la prueba de Estado. Pero, adicionalmente, se lograron determinar aquellos factores puntuales que impactan en mayor medida estos puntajes. Dichos factores son: el nivel educativo de los padres (entendido como el máximo nivel de escolaridad alcanzado tanto por el padre como por la madre), el tipo de labor u oficio de los padres, el acceso a recursos educativos en casa (como computador, internet o libros) y el tipo de jornada escolar (matutina, vespertina, nocturna, sabatina o completa). De manera general, se encontró que aquellos estudiantes que viven en sectores de estratos 5 y 6, en hogares con más de 100 libros, estudian en colegios privados de jornada completa, y cuyos padres tienen altos niveles educativos (profesional o posgrado), trabajan en labores profesionales (como médicos, abogados o ingenieros), tienen mayores probabilidades de obtener altos puntajes en la prueba Saber 11 (es decir, de ubicarse dentro del 10 % más alto). En cambio, aquellos estudiantes que no

tienen acceso a recursos educativos en casa, como computador, internet o libros; estudian en jornada nocturna o sabatina y tienen padres con bajos niveles educativos (a lo sumo, básica primaria) que trabajan en labores de agricultura o como jornaleros, tienen altas probabilidades de obtener bajos puntajes en la prueba Saber 11 (es decir, de ubicarse en el 10 % más bajo).

Ante esto, en el apartado de discusión se analizó cómo estos hallazgos hablan más de problemáticas a nivel estructural que de deficiencias conceptuales particulares, y de cómo, a pesar de que estos factores se pueden categorizar en dos ejes centrales de análisis: capital económico y cultural, y jornada escolar, ambos ejes se devuelven, en últimas, sobre las condiciones socio-económicas del núcleo familiar del estudiante. Esto obliga a conducir la reflexión en dos vías: la primera, que cualquier discusión sobre políticas públicas que le apunte a reducir la brecha educativa en Colombia, necesariamente debe pasar por un estudio de reforma estructural de mejoramiento de las condiciones socio-económicas de las familias que se busca impactar, al menos en lo que tiene que ver con el acceso a recursos educativos. En segundo lugar, este estudio plantea un cuestionamiento al papel de la prueba Saber 11, ya que, al tener presente de que se parte de condiciones socio-económicas asimétricas, que impactarán en el resultado de la prueba, hasta qué punto el examen de Estado deja de ser, como se anuncia, un termómetro de unas competencias estandarizadas de la educación escolar en Colombia, y más bien se convierte en un instrumento de validación de una clase social privilegiada, reforzando con ello la estratificación social en el país.

REFERENCIAS BIBLIOGRÁFICAS

Bernal, L. K. A.-G., & Bernal, G. (2016). Brechas de género en el rendimiento escolar a lo largo de la distribución de puntajes: Evidencia pruebas Saber 11. Universidad Javeriana-Bogotá. https://ideas.repec.org/p/col/000416/015301.html

Bonilla-Mejía, L. (2014). Doble jornada escolar y calidad de la educación en Colombia. Capítulo 1. Doble jornada escolar y calidad de la educación en Colombia. Pág.: 3-56. https://repositorio.banrep.gov.co/handle/20.500.12134/6705

Castillo, A. D., & Rodríguez, J. G. O. (2013). Pruebas ICFES Saber 11 y su relación con el desempeño académico en estudiantes de primer semestre de psicología. *Cuadernos hispanoamericanos de psicología*, 13(1), 26-35. https://doi.org/10.18270/chps.. v13i1.1355

Costa, P. S., Santos, N. C., Cunha, P., Cotter, J., & Sousa, N. (2013). The use of multiple correspondence analysis to explore associations between categories of qualitative variables in healthy ageing. *Journal of aging research*, 2013. https://doi.org/10.1155/2013/302163

Cuenca, A. (2016). Desigualdad de oportunidades en Colombia: Impacto del origen social sobre el desempeño académico y los ingresos de graduados universitarios. *Estudios pedagógicos (Valdivia)*, 42(2), 69-93. http://dx.doi.org/10.4067/S0718-07052016000200005

Díaz, L. P. A., Carvajal, R. R., & da Costa, U. (2020). Análisis de la relación entre el resultado de lectura crítica en las pruebas Saber 11, el promedio académico y el desempeño en tareas de atención y memoria en estudiantes universitarios. *Revista Temas: Departamento de Humanidades Universidad Santo Tomás Bucaramanga*, 14, 99-108. https://doi.org/10.15332/rt.v0i14.2461

Echavarría, J. J. (2018). La calidad de vida en el sector agropecuario colombiano: Una mirada a la población rural. Capítulo 3. La calidad de vida en el sector agropecuario colombiano: una mirada a la población rural. Pág.: 73-91. https://repositorio.banrep.gov.co/handle/20.500.12134/9645

Gabalán-Coello, J., & Vásquez-Rizo, F. E. (2016). SABER 11 y rendimiento universitario: Un análisis del progreso en el plan de estudios. C*iencia, docencia y tecnología*, 53, 135-161. http://ref.scielo.org/cgkk3p

Greenacre, M. (2005). From correspondence analysis to multiple and joint correspondence analysis. Available at SSRN 847664. https://dx.doi.org/10.2139/ssrn.847664

Hall, M. (1999). Correlation-based Feature Selection for Machine Learning. University of Waikato. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.444.2856&rep=rep1&type=pdf

Lin, M., Lucas, H., & Shmueli, G. (2013). Too Big to Fail: Large Samples and thep-Value Problem. *Information Systems Research*, Articles in Advance, 1-12. doi: 10.1287/isre.2013.0480

Lopes de Oliveira, T., Oliveira, R. V. C. de, Griep, R. H., Moreno, A. B., Almeida, M. da C. C. de, Almquist, Y. B., & Fonseca, M. de J. M. da. (2021). Brazilian Longitudinal Study of Adult Health (ELSA-Brasil) participant's profile regarding self-rated health: A multiple correspondence analysis. *BMC Public Health*, 21(1), 1-9. https://doi.org/10.1186/s12889-021-11760-2

López, D. E. T., Blanquicett, E. J. G., Ramos, M. C. M., Tovar, A. C. M., & Morinson, M. L. V. (2022). Actividades Complementarias-Extraescolares y el aprendizaje en Estudiantes de Básica Primaria: Un Estudio con Análisis en Correspondencias Múltiples: Impact of Supplementary and Extracurricular Activities to Students Learning of Primary Basic: A Study with Multiple Correspondence Analysis. *STUDIES IN ENGINEERING AND EXACT SCIENCES*, 3(1), 137-155. https://doi.org/10.54018/seesv3n1-013

Meisel Roca, A. (2019). ¿Atrapados en la periferia? Brechas de calidad en la educación en Colombia: Pruebas Saber 11 (2000-2018). Universidad del Norte. https://EconPapers.repec.org/RePEc:col:000383:018078

Ogunsakin, R. E., Moyo, S., Olugbara, O., & Israel, C. (2021). Relating student engagement indicators to academic performance using multiple correspondence analysis. *Cybernetics and Information Technologies*, 21(1), 87-102. https://doi.org/10.2478/cait-2021-0007

Oyola, J., & Echeverri, H. (2021). Análisis comparativo de las políticas de admisión de la Universidad de Córdoba: Un caso de estudio en el programa de Ingeniería Industrial en el semestre 2019-I. *Revista Aglala*, 12(1). https://revistas.curn.edu.co/index.php/aglala/article/view/1824

Peña, Y., & González, J. (2022). Modelo de predicción de los resultados de la prueba icfes saber 11 en el área de matemáticas a partir de variables socioeconómicas. *Studies in Engineering and Exact Sciences*, 3(1), 31-37. https://doi.org/10.54021/seesv3n1-006

Pereira, R. T., Caicedo, X., & Troya, A. H. (2019). Árboles de decisión para predecir factores asociados al desempeño académico de estudiantes de bachillerato en las pruebas Saber 11. *Revista de Investigación Desarrollo e Innovación: RIDI*, 9(2), 12. Doi: 10.19053/20278306.v9.n2.2019.9184

Pereira, R. T., Troya, A. H., & Caicedo, X. (2020). Factores asociados al desempeño académico en Lectura Crítica en las pruebas Saber 11 con árboles de decisión. *Investigación e Innovación en Ingenierías*, 8(3), 29-37. https://doi.org/10.17081/ invinno.8.3.4701

Quintero, O. L. L., & Vélez, D. H. (2016). Un estudio de las brechas municipales en calidad educativa en Colombia: 2000-2012. *Ensayos sobre política económica*, 34(79), 3-20. https://doi.org/10.1016/j.espe.2016.01.001

Ricardo Benitez, W. V., Tunjo Prada, C. J., & Soto Ortiz, L. P. (2020). Prevención frente a los factores escolares asociados a la deserción escolar. Universidad El Bosque. http://hdl.handle.net/20.500.12495/4047

Singh, S. (2003). Simple Random Sampling. En S. Singh (Ed.), *Advanced Sampling Theory with Applications: How Michael ' selected' Amy Volume I* (pp. 71-136). *Springer Netherlands*. https://doi.org/10.1007/978-94-007-0789-4_2

Vargas, V., & Ardila, L. (s. f.). Predicción del desempeño en las pruebas Saber 11 utilizando variables del contexto socio-económico de los aplicantes mediante un análisis estadístico con técnicas de machine learning. ICFES. https://www.icfes.gov.co/

ents/20143/2157860/04+Policy+Brief+Vladimir+Vargas+y+Leonel+Ardila.pdf

Varvarigos, D. (2020). Cultural transmission, education-promoting attitudes, and economic development. *Review of Economic Dynamics*, 37, 173-194. https://doi.org/10.1016/j.red.2019.11.005