

Determinantes del bajo rendimiento del alumnado español en comprensión lectora en PISA: un enfoque de machine learning

Determinants of Spanish Students' Low Performance in Reading Comprehension in PISA: A Machine Learning Approach

Alexander Constante-Amores^{1*}, Delia Arroyo Resino^{**}, Enrique Navarro Asencio y María Castro Morena^{**}

* Departamento de educación. Universidad Camilo José Cela (España)

** Departamento de investigación y psicología en educación. Universidad Complutense de Madrid (España)

Resumen

El bajo rendimiento en comprensión lectora es uno de los grandes problemas del sistema educativo español. Debido a esto, el objetivo de dicho trabajo es estudiar los determinantes de contexto que se asocian a este bajo desempeño lector. La muestra se encuentra conformada por un total de 35943 estudiantes españoles/as y 1089 centros educativos que participaron en PISA 2018. La variable criterio es la competencia lectora, la cual se ha dicotomizado (0 = medio y alto rendimiento y 1 = bajo rendimiento). Como variables independientes se seleccionaron un total de 721 predictores. Para el análisis de los datos se aplicó el algoritmo de machine learning Random Forest y se realizó una regresión logística binaria multinivel. Las 30 variables más importantes relacionadas con los y las estudiantes y centro escolar explican el 46% y 24% de la variable criterio, respectivamente. El modelo final (formado por ambos predictores) explican un 47%. Entre las principales conclusiones se destaca la relevancia que tienen las variables de proceso educativo y los constructos no cognitivos y meta-cognitivos en el bajo rendimiento lector. Por lo tanto, se subraya la importancia de trabajar este fenómeno educativo desde una perspectiva menos vinculada a determinantes socioeconómicos y más orientada hacia aspectos pedagógicos.

Palabras clave: PISA; machine learning; bajo rendimiento; competencia lectora.

Correspondencia: Delia Arroyo Resino (garroy01@ucm.es). Departamento de Investigación y Psicología en Educación, Facultad de Educación, Universidad Complutense de Madrid

Abstract

Low performance in reading comprehension is one of the major challenges in the Spanish educational system. Therefore, the objective of this study is to investigate the contextual determinants associated with this low reading performance. The sample consists of a total of 35,943 Spanish students and 1,089 educational centers that participated in the 2018 PISA assessment. The criterion variable is reading competence, which has been dichotomized (0 = medium and high performance, 1 = low performance). A total of 721 predictors were selected as independent variables. For data analysis, Random Forest machine learning algorithm was applied, and a multilevel binary logistic regression was conducted. The 30 most important variables related to students and school center explain 46% and 24% of the criterion variable, respectively. The final model (comprising both predictors) explains 47%. Among the main conclusions, the significance of educational process variables and non-cognitive and meta-cognitive constructs in low reading performance stands out. Therefore, the importance of addressing this educational phenomenon from a perspective less linked to socio-economic determinants and more focused on pedagogical aspects is emphasized.

Keywords: PISA, machine learning, low performance, reading competence

Introducción

Uno de los grandes problemas del sistema educativo español es el bajo nivel de rendimiento de los y las estudiantes en la enseñanza obligatoria (Antelm et al., 2018; Garrido et al., 2020). El Sistema Estatal de Indicadores de la Educación sostiene que esto propicia el abandono temprano de la educación (Ministerio de Educación y Formación Profesional, 2020a). También conlleva desmotivación, bajo autoconcepto académico (Fernández-Lasarte et al., 2019), pocas posibilidades de acceder al mundo laboral (Marchesi, 2003), mayor riesgo de exclusión social (Comisión Europea, 2016) y un menor crecimiento económico y cultural del país (Astakhova et al., 2016).

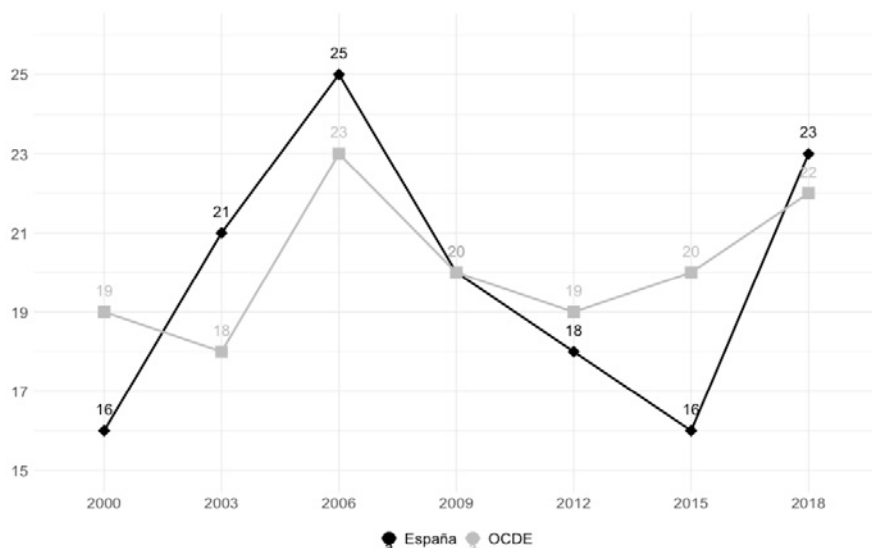
La comprensión lectora es una de las competencias con mayor tasa de estudiantes con bajo rendimiento. En PISA 2018, donde fue la competencia principal evaluada, un 23% de los estudiantes españoles mostraron un bajo nivel, siendo la media del estudiantado de 477 puntos, significativamente inferior a la media de la OCDE (487) y a la total de la Unión Europea (489) (Ministerio de Educación y Formación Profesional, 2020b).

La competencia lectora en PISA 2018 se define como “la comprensión, el uso, la evaluación, la reflexión y el compromiso con los textos para lograr objetivos propios, desarrollar el conocimiento y participar en la sociedad” (OCDE, 2018, p. 9). Se evalúa en una escala con media 500 y desviación típica 100, distribuida en seis niveles (1 a 6). El nivel 1, considerado de bajo desempeño se subdivide en estudiantes de nivel 1a que identifican información relevante en textos simples; los del 1b que comprenden textos breves y localizan fragmentos; y los del 1c que manejan vocabulario básico, pero no comprenden relaciones ni significados complejos. A partir del nivel 2 (considerado el mínimo para un desempeño adecuado y una inserción social y económica satisfactoria) se exige interpretar relaciones en el texto, mientras que los niveles superiores implican comprender textos complejos (nivel 3), extensos (nivel 4), distinguir contenido y propósito (nivel 5) y analizar textos abstractos (nivel 6) (OCDE, 2016, 2019).

En España, como se observa en la figura 1 se ha pasado de un 16% de estudiantes que se situaban en un bajo desempeño (nivel 1) en el año 2000, a un 23% en el año 2018, dato similar al promedio de los países de la OCDE (22%) (Ministerio de Educación y Formación Profesional, 2020b). Esta tendencia respecto al porcentaje de estudiantes que se encuentran en los niveles de bajo rendimiento se ve mitigada en los años 2009, 2012 y 2015, pero parece que nuevamente se está incrementando.

Figura 1

Evolución del porcentaje bajo desempeño en la comprensión lectora en España y OCDE



Nota. Elaboración propia a partir de Ministerio de Educación y Formación Profesional (2020b)

El bajo rendimiento en comprensión lectora tiene una serie de consecuencias negativas. A nivel académico, como está estrechamente vinculado al rendimiento en otras asignaturas (matemáticas, ciencias y lengua), puede repercutir negativamente en las mismas (Viramontes et al., 2019), ya que puede existir una relación positiva y significativa entre ellas (García et al., 2018). Asimismo, dificulta la posibilidad de obtener un aprendizaje significativo (Molina Ibarra, 2020). A nivel cognitivo y personal la falta de comprensión lectora se vincula con cierta incapacidad de pensamiento crítico. La comprensión lectora permite analizar, inferir y evaluar información, por lo que su ausencia limita el desarrollo de estas capacidades esenciales (Huamán Herreros, 2024). Todo ello repercute en el autoconcepto y reduce el interés por aprender (Martín-Ruiz y González-Valenzuela, 2022). A nivel laboral y social, una comprensión lectora deficiente afecta la capacidad de interpretar documentos, seguir instrucciones y comunicarse eficazmente, impactando en la empleabilidad y en el ejercicio de la ciudadanía (Molina

Ibarra, 2020), ya que como señalan García et al. (2018), la capacidad de aprender de forma autónoma a lo largo de la vida se sustenta en una adecuada comprensión lectora.

Por todo ello, la identificación temprana de los factores que contribuyen al bajo rendimiento lector es fundamental, teniendo en cuenta su carácter multidimensional (Choi y Calero, 2013; Hu et al., 2022).

Siguiendo a PISA 2018, dichos factores pueden agruparse en torno a tres áreas: antecedentes del estudiante, procesos educativos y constructos no cognitivos y meta-cognitivos (Figura 2).

Figura 2

Estructura modular de las variables de contexto en PISA 2018

	Constructos de antecedentes de los estudiantes		Constructos asociados a la escuela			Constructos no cognitivos y metacognitivos
Competencia lectora	5. Experiencia científica fuera de la escuela		ENSEÑANZAS Y APRENDIZAJE			4. Medidas relacionadas con la lectura: actitudes, motivación y estrategias
			1. Cualificación y conocimiento profesional del profesor	2. Práctica de enseñanza de las ciencias	11. Tiempo de aprendizaje y currículo	
			POLÍTICAS ESCOLARES			
			3. Entorno de aprendizaje escolar de la lectura			
Categorías generales	6. Estatus socioeconómico del estudiante y la familia	8. Recorrido educativo en la primera infancia	13. Implicación de los padres	12. Clima escolar: relaciones interpersonales, confianza y expectativas	14. Contextos y recursos escolares	9. Variables disposicionales y enfocadas en la escuela
	7. Migración y cultura		GOBERNANZA			10. Disposiciones para la competencia global
			16. Evaluación del estudiante, evaluación institucional y responsabilidad		15. Asignación, selección y elección	

Nota. OCDE (2019, p. 220)

Los *antecedentes de los y las estudiantes* destacan en la literatura como las variables más importantes en la explicación del bajo rendimiento en comprensión lectora (Calero et al., 2010; Garrido et al., 2020), sobre todo el contexto socioeconómico del alumnado (Choi y Calero, 2013). La investigación realizada por Franco et al. (2016) indican que un bajo nivel socioeconómico se vincula con puntuaciones bajas en comprensión lectora. El trabajo de Gil-Flores y García-Flores (2017) muestran la importancia del nivel socioeconómico del estudiantado y del centro en el rendimiento en PISA. Además, la OCDE (2015) destaca que los centros educativos con un alto porcentaje de estudiantes con un nivel socioeconómico desfavorecido tienen mayor probabilidad de obtener puntuaciones bajas en dicha competencia.

Otras variables relacionadas con esta área son la repetición de curso, sexo y hábitos lectores. En relación con la primera, las investigaciones de Calero y Choi (2013) y

Garrido et al. (2020) muestran que los y las estudiantes que han repetido curso son las que tienen una mayor probabilidad de obtener un bajo rendimiento en la competencia lectora. En relación con la segunda, son los varones los que tienen un mayor bajo rendimiento (Cordero et al., 2011; Instituto Nacional de Estadística, 2018; Ostria et al., 2014). Con respecto a los hábitos lectores, el trabajo realizado por Akande y Oyedapo (2018) indica que el estudiantado con un menor desempeño en comprensión lectora son los que menos tiempo dedican a la lectura. En este sentido, cuando no existen unos hábitos lectores consolidados encontramos bajos niveles de rendimiento lector (Jiménez-Pérez et al., 2020).

En relación con el área de *procesos de enseñanza y aprendizaje*, destacan un conjunto de variables que tienen un impacto directo en la comprensión lectora, como la calidad del entorno escolar (González y Jackson, 2014), incluyendo la implicación familiar (Murillo y Hernández-Castilla, 2020), la formación docente (Cutimbo, 2008; Vélez et al., 1994), los recursos disponibles (Beltrán et al., 2011; Rouse y Krueger, 2004; Albelda, 2019) y el clima institucional (Albelda, 2019), entre otras. Mejorar estos procesos es clave para transformar el rendimiento académico y avanzar hacia una escuela más inclusiva y eficaz.

Finalmente, entre los constructos no *cognitivos y metacognitivos*, destacan las aspiraciones educativas de los y las estudiantes. Garrido et al. (2020) demuestran que el alumnado que espera completar estudios universitarios son los que tienen una mayor probabilidad de obtener un alto rendimiento en la competencia lectora. Asimismo, también son relevantes la actitud y motivación del alumnado, por ello, la OCDE (2016) muestra que los y las estudiantes que presentan actitudes académicas negativas son las que tienen una mayor probabilidad de obtener un bajo rendimiento. Las investigaciones realizadas por Franco et al. (2016) manifiestan que la motivación junto con la actitud hacia la lectura y el uso del tiempo libre son fundamentales en el desarrollo de la competencia lectora. También, Guerra-García et al. (2021) indican que la desmotivación es una de las principales causas del bajo rendimiento a la hora de interpretar un texto escrito.

Por lo tanto, el rendimiento no depende solo de capacidades académicas, sino también de expectativas personales y emocionales, lo que exige una educación que fomente el compromiso, la autoestima y el sentido de propósito en los estudiantes.

Los análisis de estos factores se han abordado tradicionalmente mediante modelos multinivel (Calero et al., 2010; Choi y Calero, 2013; Garrido, 2020), valorados por su equilibrio entre parsimonia y ajuste a los datos (Pardo y Ruiz, 2013). Sin embargo, la selección de predictores en estos modelos suele apoyarse en criterios teóricos, lo que puede implicar una elección limitada de variables y la exclusión de otras potencialmente relevantes por falta de un respaldo empírico. En este contexto, cobran relevancia los enfoques de minería de datos y aprendizaje automático, que permiten identificar, entre grandes volúmenes de información, las variables con mayor capacidad predictiva (Constante-Amores et al., 2024). Esta integración metodológica contribuye a reducir errores de especificación y evita excluir variables significativas (Gaviria y Castro, 2005).

Tras lo expuesto, el objetivo general de esta investigación es analizar, en el contexto español, los determinantes (antecedentes del estudiante, procesos educativos y constructos no cognitivos y meta-cognitivos) asociados al bajo rendimiento en comprensión

lectora en el marco de la evaluación PISA 2018. Este objetivo general se concreta en los siguientes específicos:

- Identificar las variables del/a estudiante y centro escolar que tienen una mayor influencia con el bajo rendimiento en comprensión lectora.
- Determinar las variables que más inciden en la probabilidad de que un/a estudiante tenga un bajo rendimiento lector teniendo en cuenta la estructura jerárquica de los datos (nivel 1: estudiante y nivel 2: centro escolar).

Método

Diseño

Para el estudio de los factores que inciden en el bajo rendimiento en comprensión lectora se llevó a cabo un análisis secundario de la evaluación PISA 2108, ya que fue la materia central de esa edición. La metodología de investigación se caracteriza por ser de carácter cuantitativo tratándose de un diseño no experimental de corte transversal y predictivo.

Participantes

La muestra está compuesta por 35943 estudiantes españoles y 1089 centros. La edad media de los sujetos fue de 15 años ($DE = 0.28$), de los cuales el 49.95% son mujeres y el 50.04% son hombres.

Variables

La variable respuesta es la comprensión lectora en PISA 2018, la cual se ha dicotomizado (0 = medio y alto rendimiento y 1 = bajo rendimiento). Se considera que un/a estudiante obtiene un bajo rendimiento cuando no alcanza el nivel 2 en comprensión lectora.

Como variables predictoras se utilizaron los cuestionarios administrados a los/as estudiantes (estudiante, uso de las tecnologías de la información y la comunicación, bienestar del estudiante, recorrido educativo) y a los y las directoras del centro (cuestionario de centro escolar). No se incluyeron las variables del cuestionario de competencia financiera, ya que se centran en conocimientos que no se encuentran relacionados con el rendimiento lector. De las 739 variables predictoras iniciales, se eliminaron 18 por superar el 70 % de valores perdidos. Finalmente, se trabajó con 558 variables del/a estudiante y 163 del centro. Además, los predictores se mantuvieron en su nivel original, sin agregación ni desagregación entre niveles.

Procedimiento y análisis de los datos

Para dar respuesta al primer objetivo de investigación, orientado a identificar las variables del/a estudiante y de centro que tienen una mayor asociación con el bajo rendimiento en comprensión lectora se utilizó el algoritmo de clasificación de *Random*

Forest de Breiman (2001). En primer lugar, se realizó un preprocesado de los datos tal siguiendo a Arroyo et al. (2024a). La base de datos del/a estudiante y del centro se dividieron aleatoriamente en dos conjuntos: uno de entrenamiento (70%), utilizado para ajustar el modelo, y otro de validación (30%), empleado para evaluar su rendimiento. A continuación, se optimizó de manera conjunta los hiperparámetros de *Random Forest* más importantes (número de árboles, número de variables en cada árbol y máxima profundidad de los árboles) en la muestra de entrenamiento mediante la validación cruzada de 10 pliegues. En la implementación del algoritmo no se consideró la estructura jerárquica de los datos, ya que la versión multinivel de Random Forest para clasificación aún está en desarrollo. Por ello, no se excluyeron centros con menos de 20 estudiantes, criterio que resulta fundamental al construir modelos multinivel (Gaviria y Castro, 2005). Tampoco se aplicaron ponderaciones según el peso muestral, ya que se trata de un análisis exploratorio.

Al implementar un algoritmo de aprendizaje automático, no existe un criterio unificado para el manejo de los Valores Plausibles (VP). Por ejemplo, el estudio de Hu et al. (2022) utilizó el primer VP. Por su parte, los trabajos de Arroyo et al. (2024a) y Arroyo et al. (2024b) optaron por seleccionar el VP que consideraron más preciso (menor error de predicción). La presente investigación siguió la metodología adoptada por estos últimos autores, pero en esta investigación al tratarse de una variable categórica se utilizó aquel valor que clasificó correctamente un mayor porcentaje de casos en ambas categorías (0 = no bajo rendimiento y 1 = bajo rendimiento) tanto para la base de datos del estudiante como la de centro. En la Tabla 1 se presentan los VP más precisos junto con sus respectivos hiperparámetros, optimizados de forma conjunta en la muestra de entrenamiento. Observándose un excelente rendimiento de los modelos en la muestra de validación.

Tabla 1

Hiperparámetros y accuracy de Random Forest (nivel 1: estudiante y nivel 2: centro)

	Hiperparámetros (muestra de entrenamiento)			Accuracy (muestra de validación)
	Número de variables	Número de árboles	Máxima profundidad	
VP 6 (nivel 1: estudiante)	5	1000	20	86%
VP 5 (nivel 2: centro)	45	1000	10	81%

Una vez identificados los VP más precisos se seleccionaron las variables más importantes. Una de las limitaciones de Random Forest y en general, de los algoritmos de *machine learning* es que no existe un punto de corte específico a la hora de seleccionar los predictores más relevantes (Raschka y Mirjalili, 2019). Por ello, siguiendo a Goros-

tiaga y Rojo-Álvarez (2016), se evaluaron tres conjuntos de variables (15, 20 y 30) para determinar el número óptimo de predictores. El análisis mostró que el conjunto de 15 variables, seleccionadas mediante el índice de Gini, ofrecía la mayor precisión tanto en la base de datos del estudiante como la de centro.

Para dar respuesta al segundo objetivo, que consistía en identificar las variables que inciden en la probabilidad de que un/a estudiante obtenga un bajo desempeño se aplicaron ponderaciones según los pesos muestrales, tal y como recomienda la OCDE para no obtener estimaciones sesgadas (OCDE, 2025) utilizando la técnica de regresión logística multinivel, teniendo en cuenta la estructura jerárquica de los datos (nivel 1: estudiante y nivel 2 = centro escolar), ya que la varianza aleatoria entre centros del modelo nulo resultó estadísticamente significativa y el Coeficiente de Correlación Intraclass era superior al 10% (Lee, 2000). Previamente a la realización de los modelos, se correlacionaron las variables para eliminar aquellas con una correlación superior a 0.80 (Kassambara, 2018). En este sentido, los predictores, *número de chicas* y *número de chicos* en el centro educativo tuvieron una alta relación, por lo que se procedió a eliminar este último ya que tenía una menor influencia en nuestra variable de interés. También se eliminó la variable *¿Qué crees que harás dentro de 5 años?*, ya que aportaba la misma información que *¿esperas completar estudios universitarios?* y *¿esperas completar bachillerato o grado medio?* A continuación, para garantizar un análisis multinivel adecuado, se eliminaron los centros con menos de 20 estudiantes (Gaviria y Castro, 2005), lo que supuso la exclusión de 1532 sujetos de 133 centros. La muestra final quedó conformada por 34411 estudiantes de 976 centros. El análisis se realizó siguiendo las recomendaciones metodológicas de Laukaityte y Wiberg (2017) para el uso de modelos multinivel empleando los 10 VP.

Se estimaron un total de cuatro modelos. El modelo 0 es el modelo nulo (sin predictores). El modelo 1 incluye únicamente las variables del/a estudiante y el modelo 2 solo los predictores de centro. De esta manera se puede observar la contribución de cada conjunto de variables por separado. Y, por último, en el modelo 3 se incluyen todas las variables. Cabe destacar, que el orden a la hora de introducir las variables independientes en el modelo predictivo ha sido el obtenido con el algoritmo de Random Forest. La interpretación de los resultados se basó en la *odds ratio*, mientras que la capacidad explicativa del modelo se evaluó mediante el R^2 marginal y condicional. El primero refleja la varianza explicada por los efectos fijos y el segundo, la del modelo completo, incluyendo efectos aleatorios, lo que permite valorar tanto la contribución de los predictores como la estructura jerárquica (Nakagawa y Schielzeth, 2013).

Finalmente, para comprobar el ajuste de los modelos se utilizaron los índices AIC, BIC y Deviance, también, con el fin de comparar modelos anidados, se calculó la significatividad de este estadístico de reducción y se estimó el porcentaje de varianza reducido () (Cameron y Windmeijer, 1997). Todos los análisis se realizaron con el programa R, versión 4.3.0.

Resultados

A continuación, se presentan los resultados en función de los objetivos planteados.

Objetivo 1: Identificación de los predictores que más influyen en el bajo desempeño

En la tabla 2, se presentan las variables del/a estudiante más importantes. Como se observa, el predictor que tiene mayor influencia sobre el bajo rendimiento en comprensión lectora, es la repetición de curso, seguido de las aspiraciones educativas.

Tabla 2

Relación de variables más asociadas con el bajo nivel de desempeño en Compresión Lectora (nivel estudiante)

Variable independiente	Índice Gini
1 Haber repetido algún curso en primaria.	497.431
2 Expectativas del estudiante: ¿esperas completar estudios universitarios?	353.285
3 Haber repetido algún curso en secundaria.	288.373
4 Expectativas del estudiante: ¿esperas completar bachillerato o grado medio?	244.538
5 ¿Qué crees que harás dentro de 5 años?	225.329
6 ¿Asistes actualmente a lecciones complementarias en la asignatura de Lengua?	221.352
7 Verifico cuidadosamente si los hechos más importantes están representados cuando realizo un resumen.	214.545
8 En la prueba de PISA, ¿cómo te sientes con las tareas de lectura?: muchos textos me resultaron difíciles.	204.985
9 Considero que internet es un gran recurso para obtener información que me interesa (noticias, deportes, diccionario, etc.).	182.776
10 ¿Cuántos libros tienes en el hogar?	181.399
11 En la prueba de PISA, ¿cómo te sientes con las preguntas de lectura?: había muchas palabras que no podía entender.	175.301
12 ¿Qué tan informado estás sobre el cambio climático y el calentamiento global?	166.685
13 Verifico la dirección de correo electrónico del remitente.	163.458
14 Intento copiar con precisión tantas frases como sea posible cuando realizo un resumen.	147.937
15 ¿Cuánto tiempo, en promedio, duran tus clases?	145,35

En relación con las variables del centro (ver Tabla 3), se aprecia que los predictores que tienen una mayor influencia en el bajo desempeño son las vinculadas con la participación de los progenitores y la formación docente.

Tabla 3

Relación de variables más asociadas al riesgo de bajo nivel de desempeño en C. Lectora (nivel centro escolar)

	Variables independientes	Índice Gini
1	Proporción de padres: voluntarios en actividades físicas o extracurriculares.	409.943
2	Docentes con una titulación de máster o equivalente a tiempo completo.	369.328
3	Docentes con una titulación de máster o equivalente a tiempo parcial	297.286
4	Docentes con una titulación de licenciaturas o equivalente a tiempo completo.	287.791
5	Docentes con una titulación de licenciaturas o equivalente a tiempo parcial.	230.467
6	Docentes certificados a tiempo completo	228.013
7	Porcentaje de estudiantes procedentes de hogares socioeconómicamente desfavorecidos	228.013
8	Financiamiento escolar para el año escolar: tasas de estudiantes o cargos escolares pagados por los padres	198.487
9	Porcentaje de estudiantes cuyo idioma de origen es diferente del castellano.	197.764
10	Porcentaje de estudiantes con necesidades educativas especiales	197.535
11	Docentes certificados a tiempo parcial	191.345
12	En el último curso académico, ¿qué proporción de estudiantes dejaron la escuela sin el título de Educación Secundaria Obligatoria?	143.522
13	Número total de estudiantes	132.686
14	Número de chicas	117.493
15	Número de chicos	106.28

Objetivo II: Probabilidad de obtener un bajo desempeño

Una vez identificado las variables que tienen un mayor impacto en el bajo desempeño se realizaron los modelos de regresión logística binaria multinivel, pero previamente se dicotomizaron las variables independientes de naturaleza categórica, tal y como recomienda Pardo y Ruiz (2013) (ver Tabla 4).

Tabla 4

Variables independientes categóricas del modelo de regresión logística multinivel

Etiqueta	Variable independiente	Valores recodificados
Repetición primaria	Haber repetido algún curso en primaria.	0 = No
Repetición secundaria	Haber repetido algún curso en secundaria.	1 = Si, una vez o más
Estudios universitarios	Expectativas del estudiante: ¿esperas completar estudios universitarios?	0 = No 1 = Si
Bachillerato o grado medio	Expectativas del estudiante: ¿esperas completar bachillerato o grado medio?	
Asistencia a lecciones complementarias	¿Asistes actualmente a lecciones complementarias en la asignatura de Lengua?	
Verifico resumen	Verifico cuidadosamente si los hechos más importantes están representados cuando realizo un resumen.	0 = Poco útil, casi nada útil y nada útil 1 = Útil, bastante útil y muy útil
Copiar resumen	Intento copiar con precisión tantas frases como sea posible cuando realizo un resumen.	
Tareas de lectura	En la prueba de PISA, ¿cómo te sientes con las tareas de lectura?: muchos textos me resultaron difíciles.	0 = Totalmente en desacuerdo o en desacuerdo
Internet	Considero que internet es un gran recurso para obtener información que me interesa (noticias, deportes, diccionario, etc.).	1= Totalmente de acuerdo o de acuerdo
Lectura	En la prueba de PISA, ¿cómo te sientes con las preguntas de lectura?: había muchas palabras que no podía entender.	
Libros en el hogar	¿Cuántos libros tienes en el hogar?	0 = 25 libros o menos 1 = 26 libros o más
Verifico correo	Verifico la dirección de correo electrónico del remitente.	0 = Poco apropiado, casi nada apropiado y nada apropiado 1 = Apropiado, bastante apropiado y muy apropiado
Ciencias	¿Qué tan informado estás sobre el cambio climático y calentamiento global?	0 = Nunca he oído hablar de esto o he oído hablar de esto, pero, no sabría explicar de qué se trata realmente. 1 = Sé explicarlo o estoy familiarizado con esto

En la tabla 5 se muestran los resultados de los cuatro modelos estimados. En el modelo 1 todas las variables han resultado estadísticamente significativas y explican el 43% de la varianza de la variable respuesta. Los predictores que tiene un mayor tamaño del efecto son la *repetición de curso en la educación primaria*, *repetición de curso en educación secundaria*, *expectativas educativas* y *asistencia a lecciones complementarias en lengua*. En el modelo 2 las variables de centro explican un 14% de la varianza, la variable independiente con un mayor efecto es el *porcentaje de estudiantes procedentes de hogares socioeconómicamente desfavorecidos*. En el modelo 3, donde se incluyen todas las variables, se observa que los todos los predictores del nivel de estudiante siguen siendo significativas, sin embargo, la única que es significativa del nivel 2 es el *porcentaje de estudiantes procedentes de hogares socioeconómicamente desfavorecidos*. Las variables con un mayor efecto son la *repetición de curso en la educación primaria* y la *repetición de curso en la educación secundaria*. Concretamente, los y las estudiantes que han repetido al menos un curso tienen un 216% y un 113% más de probabilidades de pertenecer al grupo de bajo rendimiento lector en educación primaria y secundaria, respectivamente. El porcentaje de varianza explicado al introducir las variables del centro aumentan un 1%.

Por último, en lo que concierne al ajuste de los modelos, se observa que el modelo con un menor AIC y BIC es el modelo compuesto por todos los predictores (modelo 3), donde se consigue una reducción significativa de la varianza equivalente a un del 28%, respecto al modelo 1. También fue significativa la reducción de la varianza del modelo 1 respecto al modelo 0 (27%) y del modelo 2 respecto al nulo (1%).

Tabla 5

Probabilidad (odds ratio) de obtener un bajo rendimiento en la comprensión lectora

	Modelo 0 (nulo)	Modelo 1 (estudiantes)	Modelo 2 (centro escolar)	Modelo 3 (todas)
Repetición primaria		3.133 ***		3.158 ***
Estudios universitarios		0.560 ***		0.559 ***
Repetición secundaria		2.132 ***		2.134 ***
Bachillerato o grado medio		0.609 ***		0.607 ***
Asistencia a clases complementarias		2.086 ***		2.084 ***
Verifico resumen		0.502 ***		0.503 ***
Tareas de lectura		1.808 ***		1.809 ***
Internet		0.573 ***		0.575 ***
Libros en el hogar		0.659 ***		0.660 ***
Lectura		1.666 ***		1.667 ***
Cambio climático y calentamiento global		0.718 ***		0.714 ***
Verifico correo		0.584 ***		0.584 ***
Copiar resumen		1.738 ***		1.739 ***

	Modelo 0 (nulo)	Modelo 1 (estudiantes)	Modelo 2 (centro escolar)	Modelo 3 (todas)
Tiempo		1.006 ***		1.006 ***
Implicación de los padres en actividades escolares			1.001	1.000
Docentes a tiempo completo (Máster)			0.999	1.000
Docentes a tiempo parcial (Máster)			1.010	1.004
Docentes a tiempo completo (Licenciatura)			1.000	1.001
Docentes a tiempo parcial (Licenciatura)			0.973	0.967
Docentes certificados a tiempo completo			1.001	1.000
Porcentaje de estudiantes procedentes de hogares socio-económicamente desfavorecidos			1.001***	1.003***
Financiación escolar			0.998***	1.000
Porcentaje de estudiantes cuyo idioma de origen es diferente del castellano.			0.992	0.967
Estudiantes con necesidades especiales			1.029	1.000
Docentes certificados a tiempo parcial			1.011	0.997
Sin título de educación secundaria			1.001	1.004
Número total de estudiantes			1.000	0.999
Número de chicas			0.996**	0.992
Intercepto	0.223***	0.744*	0.225***	0.149***
Varianza	0.9718	0.715	0.970	0.709
R2 marginal		35%	1%	38%
R2 condicional	23%	46%	24%	47%
Ajuste				
AIC	383286.414	281850.241	383025,585	281450,708
BIC	383303.301	281985.379	383160,723	281704,092
Log likelihood	191641.1245	-140909.121	-191496,792	-140695,354
Desviianza	383282.125	281818.214	382994.457	275718.114
Pr(>Chisq)		0.000	0.000	0.000

Discusión

Debido al impacto que el bajo rendimiento lector puede tener en la vida académica, social y laboral de los y las estudiantes, en este trabajo se propuso como *primer objetivo* identificar las variables del/a estudiante y centro que tienen una mayor relación con el bajo desempeño en comprensión lector.

En relación con el/la estudiante se observó que las variables con más peso fueron: *haber repetido un curso en primaria* (antecedentes del estudiante) y las *expectativas del estudiante respecto a completar estudios universitarios* (medidas no cognitivas). Respecto a la primera variable, las investigaciones de Choi y Calero (2013) y Garrido et al. (2020) indican que el estudiantado que ha repetido curso es el que tiene una mayor probabilidad de obtener un bajo rendimiento en la competencia lectora, además se relaciona estrechamente con el abandono educativo temprano y con el fracaso escolar (Crespo, 2018). En cuanto a la segunda variable, parece que los y las estudiantes que tienen expectativas altas y positivas tienen un nivel alto de rendimiento (Gonzalo et al., 2022), concretamente Garrido et al. (2020) destacan que el alumnado que espera completar estudios universitarios es el que tiene una mayor probabilidad de obtener un alto rendimiento en la competencia lectora.

Por lo tanto, esta realidad exige una mirada crítica al sistema educativo, donde para prevenir la repetición de curso se pueden establecer planes de refuerzo individualizados a partir de los resultados obtenidos en evaluaciones diagnósticas periódicas, contar con docentes de apoyo, utilizar metodologías inclusivas (aprendizaje cooperativo, enseñanza diferenciada, etc.) y por supuesto, contar con la participación de las familias para que exista una adecuada comunicación escuela-familia. Esta participación también es fundamental para cultivar altas expectativas académicas en los y las estudiantes fomentando desde la escuela un autoconcepto académico positivo, generando un clima escolar motivador, donde se ayude al alumnado a establecer metas claras y accesibles.

Respecto a las variables del centro se aprecia que los predictores que tienen una mayor influencia, son, primero *la proporción de padres: voluntarios en actividades físicas o extracurriculares*, en este sentido, González y Jackson (2014) hallan que la participación de los progenitores en el centro afectan positivamente y significativamente en lectura, asimismo Murillo y Hernández-Castilla (2020) demuestran que la participación de los padres y madres en actividades extracurriculares tiene un impacto positivo en la lectura. La segunda variable que tiene un mayor impacto es la referida a *docentes con una titulación de máster o equivalente a tiempo completo* (proceso de enseñanza-aprendizaje), en este sentido, diversas investigaciones concluyen que existe una correlación positiva y significativa entre el nivel de capacitación docente y el rendimiento académico del estudiantado (Cutimbo, 2008; Vélez et al., 1994).

Ante estos resultados, parece que la participación de las familias no solo fortalece el vínculo escuela-hogar, sino que se traduce en mejoras significativas en comprensión lectora. Por lo tanto, es importante invitar a las familias a colaborar en diferentes proyectos, ofreciendo horarios flexibles e incluso facilitando la participación virtual. Del mismo modo, también es importante proporcionar y facilitar a los y las docentes una formación calidad, donde se les incentive a seguir actualizándose, ya que esto repercute en el desempeño académico de sus estudiantes. Estas conclusiones subrayan

la necesidad de políticas que promuevan la implicación familiar y la capacitación continua del profesorado como pilares para mejorar los resultados educativos.

Respecto a este primer objetivo, de las 30 variables más importantes relacionadas con el alumnado y el centro educativo, 10 son de antecedentes del estudiante, 12 de proceso de enseñanza-aprendizaje y 8 medidas no cognitivas. Estos resultados distan de una amplia literatura que aboga por la importancia que tienen los predictores de antecedentes del estudiante en el rendimiento en comprensión lectora, frente al resto de áreas (Barrera et al., 2019; Choi y Calero, 2013; Franco et al., 2016; Guio y Choi, 2014). De esta manera, este estudio aporta evidencia empírica que cuestiona la visión tradicional de la literatura, ya que los resultados muestran que, aunque las variables vinculadas con los antecedentes son relevantes, las relacionadas con el proceso de enseñanza-aprendizaje y las medidas no cognitivas tienen un peso incluso mayor. Por lo tanto, sería conveniente considerar factores escolares y motivacionales, fácilmente modificables, en las políticas educativas para mejorar la comprensión lectora.

En cuanto al *segundo objetivo*, basado en determinar las variables que más inciden en la probabilidad de que un/a estudiante tenga un bajo desempeño en comprensión lectora teniendo en cuenta la estructura jerárquica de los datos. Encontramos que en el modelo 1 (variables del estudiante) todas han resultado estadísticamente significativas. En relación con los coeficientes, los que tienen un mayor efecto son los vinculados con los antecedentes del/a estudiante, concretamente la *repetición de curso*. Así, los y las estudiantes que repiten un curso en Educación Primaria tienen un 216% más de probabilidad de obtener un bajo rendimiento en la comprensión lectora en comparación con aquellos/as que nunca han repetido, reduciéndose dicha probabilidad en la siguiente etapa educativa a un 113%. En esta línea Asensio et al. (2018) con datos de PISA 2015, encuentran que efectivamente, la repetición se asocia con el bajo rendimiento y Hattie (2009), como síntesis de 800 metaanálisis llegó a la conclusión de que la repetición es una de las variables más importantes en los niveles de logro, siendo su efecto de un 16%. Otra de las variables con mayor impacto, en este caso del área de procesos de enseñanza y aprendizaje, se refiere a la *asistencia a clases complementarias de lengua*, aquellos/as estudiantes que asisten tienen un 108% de obtener una puntuación baja en PISA con respecto a los que no asisten. Esto puede ser debido a que los y las estudiantes que asisten a estas clases son aquellos que necesitan refuerzo u apoyo en dicha materia, en este sentido, cuando no existen unos hábitos lectores consolidados encontramos bajos niveles de rendimiento en esta competencia (Jiménez-Pérez et al., 2020). Respecto al modelo 2, referido a las variables de centro, la variable con mayor peso que resultó significativa pertenece al área de los antecedentes del estudiantado y se refiere al *porcentaje de estudiantes procedentes de hogares socioeconómicamente desfavorecidos y sin título de educación secundaria*. En este sentido, por cada estudiante más en el centro escolar *procedentes de hogares socioeconómicamente desfavorecidos*, la probabilidad de obtener un bajo rendimiento aumenta en aproximadamente un 1%. Esto se encuentra en sintonía con lo establecido por Barrera et al. (2019) quienes demuestran que los y las estudiantes que pertenecen a un nivel socioeconómico bajo logran desempeños cognitivos inferiores en tareas como la comprensión lectora. Asimismo, Franco et al. (2016) evidencia que un bajo nivel socioeconómico se vincula con puntuaciones bajas en comprensión lectora. Finalmente, en el modelo 3, donde se introduce tanto las variables

del/a estudiante como las de centro, se observa que mientras todas las variables del/a estudiante son significativas, a nivel de centro solo lo fue la referida al *Porcentaje de estudiantes procedentes de hogares socioeconómicamente desfavorecidos*, de esta manera, es mayor el porcentaje de varianza explicado por las variables del/a estudiante que del centro (46% y 24%, respectivamente).

Los resultados confirman que las variables del/a estudiante, sobre todo, la repetición de curso, son los predictores más determinantes del bajo rendimiento en comprensión lectora, lo que evidencia la persistencia de un modelo educativo que no logra compensar las desigualdades iniciales. Sin embargo, también se observa que factores del centro, como la concentración de alumnado socioeconómicamente desfavorecido, influyen en el desempeño, aunque en menor medida. Este hallazgo plantea la necesidad de replantear las políticas de repetición, sustituyéndolas por estrategias preventivas y de apoyo temprano, así como de fortalecer la equidad entre centros mediante recursos adicionales y programas compensatorios. Solo una intervención integral que combine acciones sobre el/la estudiante y el contexto escolar permitirá reducir las brechas en comprensión lectora.

Así pues, siguiendo la clasificación realizada por PISA 2018, parece que son las variables relacionadas con los antecedentes del/la estudiante las que tienen un mayor impacto en la variable respuesta, especialmente la repetición de curso y las expectativas académicas, ambas variables inversamente relacionadas ya que como indican Arroyo et al. (2019) se observa que la probabilidad de repetir es menor cuando los y las estudiantes tienen aspiraciones educativas más altas. A partir de todo lo comentado, sería adecuado que desde los centros educativos se trabajara en la mejora del autoconcepto estudiantil, ya que el mismo repercute en las expectativas académicas de los estudiantes y por ende en el rendimiento de diferentes materias, tal y como demuestran Carrillo-López et al. (2022). También sería interesante desarrollar actividades que refuercen la motivación de los y las estudiantes hacia la lectura con el fin de mejorar el rendimiento en comprensión lectora y, así, mitigar el efecto que la falta de dominio de dicha competencia puede tener sobre otras materias y a la hora de desenvolverse en la sociedad.

Además de las aportaciones educativas, la presente investigación también se caracteriza por su aportación metodológica, ya que se analiza la influencia de 721 variables independientes en el bajo rendimiento lector en PISA mediante el algoritmo de machine learning Random Forest, mientras que tradicionalmente, este fenómeno se ha estudiado a través de las técnicas de regresión logística multinivel (Calero et al., 2010; Choi y Calero, 2013; Garrido, 2020). En este sentido, dicho trabajo puede sentar las bases para futuros estudios secundarios de PISA que utilicen este enfoque metodológico caracterizado por su precisión.

En relación con las limitaciones del estudio, hay que destacar que, aunque en este estudio no se realizan comparaciones internacionales y la finalidad del primer análisis era puramente exploratorio, una limitación metodológica es no haber aplicado los pesos muestrales al identificar qué variables del alumnado y del centro escolar se asocian con el bajo rendimiento en comprensión lectora, lo que podría generar cierto sesgo en las estimaciones. Asimismo, por su carácter transversal, PISA no permite establecer relaciones causales, limitando la inferencia sobre los factores asociados al rendimiento. Para establecer conclusiones de este tipo, sería necesario aplicar modelos de ecuaciones

estructurales que permitan contrastar hipótesis sobre los efectos directos e indirectos entre las variables. Por otro lado, el diseño transversal de este estudio impide evaluar la evolución de los resultados a lo largo del tiempo, lo cual representa una limitación metodológica propia de los estudios PISA.

Como prospectiva de investigación, sería interesante determinar los predictores asociados al bajo rendimiento en comprensión lectora de manera internacional con el propósito de comprobar si en los demás países europeos las variables de proceso y medidas cognitivas y metacognitivas son tan determinantes como en esta investigación.

Referencias

- Arroyo Resino, D.; Constante Amores, I. A., y Asensio Muñoz, I. (2019). La repetición de curso a debate: un estudio empírico a partir de PISA 2015. *Educación XX1*, 22(2), 69-92. <https://doi.org/10.5944/educxx1.22479>
- Akande, S. O., y Oyedapo, R. O. (2018). Developing the Reading Habits of Secondary School Students in Nigeria: The Way Forward. *International Journal of Library Science*, 7(1), 15-20. Developing the Reading Habits of Secondary School Students in Nigeria: The Way Forward
- Albelda Esteban, B. (2019). Contribución de las bibliotecas escolares a la adquisición de competencias en comprensión lectora en educación primaria en España: una aproximación a partir de los datos del estudio PIRLS 2016. *Revista de educación*, 384, 1-9. <https://doi.org/10.4438/1988-592X-RE-2019-384-408>
- Antelm Lanzat, A. M., Gil López, A. J., Cacheiro González, M. L., y Pérez Navío, E. (2018). Causas del fracaso escolar: Un análisis desde la perspectiva del profesorado y del alumnado. *Enseñanza & Teaching: Revista Interuniversitaria de Didáctica*, 36(1), 129-149. <https://doi.org/10.14201/et2018361129149>
- Arroyo Resino, D., Constante-Amores, A., Castro, M., y Navarro, E. (2024b) School effectiveness and high reading achievement of Spanish students in PISA 2018: a machine learning approach. *Educación XX1*, 27(2), 223-251. <https://doi.org/10.5944/educxx1.38634>
- Arroyo Resino, D., Constante-Amores, A., Gil-Madróna, P., y Carrillo-López, P. J. (2024a). Student well-being and mathematical literacy performance in PISA 2018: a machine-learning approach. *Educational Psychology*, 44(3), 340-357. <https://doi.org/10.1080/01443410.2024.2359104>
- Asensio Muñoz, I., Carpintero Molina, E., Expósito Casas, E., y López Martín, E. (2018). ¿Cuánto oro hay entre la arena? Minería de datos con los resultados de España en PISA 2015. *Revista Española de Pedagogía (REP)*, 76(270), 225-245. <https://doi.org/10.22550/REP76-2-2018-02>
- Astakhova, K. V., Korobeev, A. I., Prokhorova, V. V., Kolupaev, A. A., Vorotnoy, M. V., y Kucheryavaya, E. R. (2016). The Role of Education in Economic and Social Development of the Country. *International Review of Management and Marketing*, 6(1S), 53-58. <https://econjournals.com/index.php/irmm/article/view/1865>

- Barrera, J. E. C., Polanco, J. G., y Acosta, J. D. (2019). Comprensión lectora de estudiantes universitarios. Factores asociados y mecanismos de acción. *Revista Venezolana de Gerencia*, 24(87), 874-889. <https://www.redalyc.org/articulo.oa?id=29060499015>
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5-32. <https://doi.org/10.1023/A:1010933404324>
- Beltrán, A. C., Seinfeld, J. C., Narro Llacza, O., y Lisboa Vásquez, C. (2011). *Hacia una educación de calidad: La importancia de los recursos pedagógicos en el rendimiento escolar*. Universidad del Pacífico, CIES.
- Calero, J., Choi, A., y Waisgrais, S. (2010). Determinantes del riesgo de fracaso escolar en España: una aproximación a través de un análisis logístico multinivel aplicado a PISA-2006. *Revista de Educación, numero extraordinario*, 225-256. <https://www.educacionfpydeportes.gob.es/revista-de-educacion/numeros-revista-educacion/numeros-antteriores/2010/re2010/re2010-09.html>
- Choi de Mendizábal, Á., y Calero Martínez, J. (2013). Determinantes del riesgo de fracaso escolar en España en PISA-2009 y propuestas de reforma. *Revista de Educación*, 362, 562-593. <https://doi.org/10.4438/1988-592X-RE-2013-362-242>
- Constante-Amores, A., Arroyo-Resino, D., Sánchez-Munilla, M., y Asencio-Muñoz, I. (2024). Contribution of machine learning to the analysis of grade repetition in Spain: A study based on PISA data [Contribución del machine learning al análisis de la repetición escolar en España: un estudio con datos PISA]. *Revista Española de Pedagogía*, 82 (289), 539-562. <https://doi.org/10.22550/2174-0909.4014>
- Cameron, A. C., y Windmeijer, F. A. (1997). An R-squared measure of goodness of fit for some common nonlinear regression models. *Journal of econometrics*, 77(2), 329-342. [https://doi.org/10.1016/S0304-4076\(96\)01818-0](https://doi.org/10.1016/S0304-4076(96)01818-0)
- Carrillo-López, P. J., Constante-Amores, A., Arroyo-Resino, D., y Sánchez-Munilla, M. (2022). Self-concept and academic achievement in primary school: A predictive study. *International Journal of Education in Mathematics, Science, and Technology (IJEMST)*, 10(4), 1057-1073. <https://doi.org/10.46328/ijemst.2303>
- Comisión Europea. (2016). *Abandono Escolar. Fichas temáticas del semestre europeo*. Consejería de Educación. european-semester_thematic-factsheet_early-school-leavers_es.docx
- Crespo Cebada, E. (2018). *Fracaso escolar español: Una comparativa internacional con datos de PISA*. Universidad de Extremadura, Servicio de Publicaciones. <http://hdl.handle.net/10662/20543>
- Cordero, J. M., Crespo, E., y Pedraja, F. (2011). Rendimiento educativo y determinantes según PISA: Una revisión de la literatura en España. *Revista de Educación*, 362, 361-388. <http://dx.doi.org/10.4438/1988-592X-RE-2011-362-161>
- Cutimbo Estrada, P. M. (2008). *Influencia del nivel de capacitación docente en el rendimiento académico de los estudiantes del Instituto Superior Pedagógico Público de Puno: caso de la Especialidad de Educación Primaria IX Semestre-2008* [Trabajo de fin de Grado, Universidad Nacional Mayor de San Marcos]. <https://hdl.handle.net/20.500.12672/2395>

- Fernández-Lasarte, O., Goñi, E., Camino, I., y Zubeldia, M. (2019). Ajuste escolar y autoconcepto académico en la Educación Secundaria. *Revista de Investigación Educativa*, 37(1), 163-179. <http://dx.doi.org/10.6018/rie.37.1.308651>
- Franco Montenegro, M. P., Cárdenas Rodríguez, R., y Santrich Sánchez, E. R. (2016). Factores asociados a la comprensión lectora en estudiantes de noveno grado de Barranquilla. *Psicogente*, 19(36), 296-310. <https://doi.org/10.17081/psico.19.36.1299>
- García, M. A.; Arévalo D., M. A., y Hernández, C. A. (2018). La comprensión lectora y el rendimiento escolar. *Cuadernos de Lingüística Hispánica*, (32), 155-174. <https://doi.org/10.19053/0121053X.n32.2018.8126>
- Garrido Yserte, R., Gallo-Rivera, M. T., y Martínez-Gautier, D. (2020). ¿Cuáles son y cómo operan los determinantes del fracaso escolar? Replantando las políticas públicas para el caso de España y sus regiones. *Revista Internacional de Ciencias del Estado y de Gobierno*, 1(4), 509-540.
- Gaviria Soto, J. L., y Castro Morera, M. (2005). *Modelos jerárquicos lineales*. La Muralla
- GilFlores, J., y GarcíaGómez, S. (2017). *Importancia de la actuación docente frente a la política educativa regional en la explicación del rendimiento en PISA*. *Revista de Educación*, 378, 187210. <http://dx.doi.org/10.4438/1988-592x-RE-2017-378-361>
- González, R. L., y Jackon, C. L. (2014). Engaging with parents: The relationship between school engagement efforts, social class, and learning. *School Effectiveness and School Improvement*, 24(3), 225-316. <http://dx.doi.org/10.1080/09243453.2012.680893>
- Gorostiaga, A., y Rojo-Álvarez, J. L. (2016). On the use of conventional and statistical-learning techniques for the analysis of PISA results in Spain. *Neurocomputing*, 171, 625-637. <https://doi.org/10.1016/j.neucom.2015.07.001>
- Guerra-García, J., Saldívar-Llanos, A., y Sandria-López, S. (2021). Evaluación de comprensión lectora, uso de estrategias y su relación con variables académicas y sociodemográficas en estudiantes universitarios. *Revista Innova Educación*, 3(2), 360-373. <https://doi.org/10.35622/j.rie.2021.02.005>
- Guio, J. M., y Choi, A. (2014). Evolución del riesgo de fracaso escolar en España durante la década del 2000: Análisis de los resultados de PISA con un modelo logístico de dos niveles. *Estudios Sobre Educación*, (26), 33-63. <https://doi.org/10.15581/004.26.33-62>
- Hattie, J. (2009). *Visible Learning: A Synthesis of Over 800 Meta-Analyses Relating to Achievement*. Taylor & Francis. <https://doi.org/10.4324/9780203887332>
- Hu, J., Peng, Y., y Ma, H. (2022). Examining the contextual factors of science effectiveness: a machine learning-based approach. *School Effectiveness and School Improvement*, 33, 21-50. <https://doi.org/10.1080/09243453.2021.1929346>
- Huamán Herreros, J. A. (2024). *Comprensión lectora y su relación con el rendimiento académico en estudiantes de secundaria*. *Revista Iberoamericana para la Investigación y el Desarrollo Educativo (RIDE)*, 15(29), e743. <https://doi.org/10.23913/ride.v15i29.2109>

- Instituto Nacional de Estadística (2018). *Panorama educativo de México. Indicadores del Sistema Educativo Nacional 2017. Educación básica y media superior*. Instituto nacional para la evaluación de la educación. Pnorama_educativo.pdf
- Jiménez-Pérez, E. D., Martínez León, N., y Cuadros Muñoz, R. (2020). La influencia materna en la inteligencia emocional y la competencia lectora de sus hijos. *Revista de Estudios Sobre Lectura*, 19(1), 80-89. https://doi.org/10.18239/ocnos_2020.19.1.2187
- Kassambara, A. (2018). *Machine learning essentials: Practical guide in R*. STHDA.
- Laukaityte, I., y Wiberg, M. (2017) Using plausible values in secondary analysis in large-scale assessments. *Communications in Statistics - Theory and Methods*, 46(22), 11341-11357. <https://doi.org/10.1080/03610926.2016.1267764>
- Lee, V. E. (2000). Using hierarchical linear modeling to study social contexts: The case of school effects. *Educational Psychologist*, 35(2), 125-141. https://doi.org/10.1207/S15326985EP3502_6
- Marchesi, A. (2003). *El fracaso escolar en España*. Fundación Alternativas-Documento de Trabajo 11/2003. <https://fundacionalternativas.org/wp-content/uploads/2022/07/xmlimport-ZPRlx1.pdf>
- Martín-Ruiz, I., y González-Valenzuela, M. J. (2022). *Autoconcepto académico, motivación y rendimiento escolar en estudiantes de Educación Secundaria Obligatoria*. *Anales de Psicología*, 38(2), 251-259. <https://dx.doi.org/10.6018/analesps.419111>
- Ministerio de Educación y Formación Profesional. (2020a). *El sistema estatal de indicadores de la Educación*. Instituto Nacional de Evaluación Educativa. https://www.libreria.educacion.gob.es/libro/sistema-estatal-de-indicadores-de-la-educacion-2020_180617/
- Ministerio de Educación y Formación Profesional. (2020b). *PISA 2018. Resultados de lectura en España*. Instituto Nacional de Evaluación Educativa. https://www.libreria.educacion.gob.es/libro/pisa-2018-resultados-de-lectura-en-espana_181801/
- Molina Ibarra, C. A. (2020). Comprensión lectora y rendimiento escolar. *Revista Boletín Redipe*, 9(1), 121-131. <https://doi.org/10.36260/rbr.v9i1.900>
- Murillo, F. J., y Hernández-Castilla, R. (2020). ¿La implicación de las familias influye en el rendimiento? Un estudio en educación primaria en América Latina. *Revista de Psicodidáctica*, 25(1), 13-22. <https://doi.org/10.1016/j.psicod.2019.10.002>
- Nakagawa, S., y Schielzeth, H. (2013). A general and simple method for obtaining R^2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, 4(2), 133-142. <https://doi.org/10.1111/j.2041-210x.2012.00261.x>
- Organización para la Cooperación y el Desarrollo Económicos (OCDE). (2015). *Education Policy Outlook 2015: Making Reforms Happen*. OECD Publishing. [http:// dx.doi.org/10.1787/9789264225442-en](http://dx.doi.org/10.1787/9789264225442-en)
- Organización para la Cooperación y el Desarrollo Económicos (OCDE). (2016). *PISA. Low-Performing Students. Why they fall behind and how to help them succeed*. OECD Publishing. <https://doi.org/10.1787/9789264250246-en>

- Organización para la Cooperación y el Desarrollo Económicos (OCDE). (2018). *Marco teórico de lectura PISA 2018*. Instituto Nacional de Evaluación Educativa, Ministerio de Educación, Cultura y Deporte, España
- Organización para la Cooperación y el Desarrollo Económicos (OCDE). (2019). *PISA 2018 assessment and analytical framework*. OECD Publishing. <https://doi.org/10.1787/b25efab8-en>
- Organización para la Cooperación y el Desarrollo Económicos (OCDE). (2025). *PISA 2022 Technical Report*. OECD Publishing. <https://doi.org/10.1787/01820d6d-en>
- Ostria Baltazar, C. E., López Padilla M. G., y Valenzuela González, J.R. (2014). Caracterización de las competencias transversales. La competencia lectora: identificación del nivel de logro. *Investigaciones Sobre Lectura*, 2, 32-43. <http://hdl.handle.net/10481/33864>
- Pardo, M., y Ruiz, M. (2013). *Análisis de Datos en Ciencias Sociales y de la Salud III*. Síntesis.
- Raschka, S., y Mirjalili, V. (2019). *Python Machine Learning: aprendizaje automático y aprendizaje profundo con Python. Scikit-Learn y TensorFlow*. Marcombo.
- Rouse, C. E., y Krueger, A. B. (2004). Putting computerized instruction to the test: A randomized evaluation of a “scientifically based” reading program. *Economics of Education Review*, 23(4), 323-338. <https://doi.org/10.1016/j.econedurev.2003.10.005>
- Vélez, E., Schiefelbein, E., y Valenzuela, J. (1994). Factores que afectan el rendimiento académico en la educación primaria: Revisión de la literatura de América Latina y el Caribe. *Revista latinoamericana de Innovaciones Educativas*, 17, 1-16. <https://hdl.handle.net/20.500.12799/4317>
- Viramontes, E., Amparán, A., y Núñez, L. D. (2019). Comprensión lectora y el rendimiento académico en Educación Primaria. *Investigaciones Sobre Lectura*, 12, 65-82. <http://hdl.handle.net/10481/60379>

Fecha de recepción: 20 febrero, 2025

Fecha de revisión: 10 marzo, 2025

Fecha de aceptación: 20 febrero, 2026