

revista de **e**EDUCACIÓN

Nº 379 ENERO-MARZO 2018



Evaluación de factores relacionados con la eficacia escolar en PISA 2015. Un análisis multinivel

Assessment of factors related to school effectiveness in PISA 2015. A multilevel analysis

Adriana Gamazo
Fernando Martínez-Abad
Susana Olmos-Migueláñez
María José Rodríguez-Conde



GOBIERNO
DE ESPAÑA

MINISTERIO
DE EDUCACIÓN, CULTURA
Y DEPORTE



Evaluación de factores relacionados con la eficacia escolar en PISA 2015. Un análisis multinivel¹

Assessment of factors related to school effectiveness in PISA 2015. A multilevel analysis

DOI: 10.4438/1988-592X-RE-2017-379-369

Adriana Gamazo
Fernando Martínez-Abad
Susana Olmos-Migueláñez
María José Rodríguez-Conde

Universidad de Salamanca

Resumen

El estudio de la eficacia escolar ha ganado relevancia en las últimas décadas. La disponibilidad de datos relativos tanto al rendimiento del alumnado como a las características socio-económicas, demográficas, organizativas y educativas de los alumnos y los centros ha permitido la proliferación de estudios sobre la relación de todo tipo de variables con el rendimiento, y sobre las prácticas que resultan fundamentales para poder ofrecer una educación equitativa y de calidad.

La presente investigación se centra en el estudio de la eficacia escolar, utilizando para ello la aplicación de modelos jerárquicos lineales (multinivel) con los datos de rendimiento de matemáticas, lectura y ciencias de la muestra española de PISA 2015 con el objetivo de determinar qué factores contextuales tienen más efecto en el rendimiento de los estudiantes. El género, el nivel socio-económico, el curso, la repetición de curso y los cambios de escuela, junto con el nivel socio-económico medio del centro, son las variables que aparecen como relevantes consistentemente en los tres modelos realizados.

⁽¹⁾ Este trabajo de investigación está financiado a través del proyecto de excelencia I+D+i EDU2015-64524-P (Ministerio de Economía y Competitividad/Fondos FEDER).

Dicha información se utiliza para realizar una selección de centros de alta y baja eficacia basada en la diferencia entre la puntuación observada de los centros y su puntuación esperada en función de los factores contextuales relevantes. A partir de esta selección, se realiza un estudio de los factores no contextuales a nivel de estudiante y de centro que se encuentran relacionados con la eficacia de los centros. No se encuentran relaciones significativas con las variables del nivel de centro, aunque sí con algunas a nivel de estudiante (clima de disciplina, auto-eficacia, apoyo emocional parental o nivel de injusticia del profesorado).

Palabras clave: evaluación a gran escala, eficacia escolar, efectos contextuales, modelos jerárquicos lineales, rendimiento académico.

Abstract

The study of school effectiveness has gained relevance in the past few decades. The availability of data pertaining both to student performance and to the socio-economic, demographic, organisational and educational features of students and schools has allowed for the proliferation of studies on the relationship of all kinds of variables with student performance, and on the essential practices to provide a quality and equal education.

This research is focused on the study of school effectiveness, using to this end the application of hierarchical linear models (multilevel) to math, reading and science performance data from the Spanish sample of PISA 2015, aiming to establish which contextual factors have a larger effect on student performance. Gender, socio-economic level, grade, grade repetition and school changes, together with the school's average socio-economic level, are the variables that appeared as relevant consistently in all three models.

This information was used to select the schools with the highest and lowest levels of effectiveness based on the difference between their observed performance scores and their expected scores according to the relevant contextual factors. This selection allows for a study on which non-contextual factors (at student and school levels) are related to school effectiveness. There were no significant relationships with the school level variables, although there were some at student level (classroom discipline, self-efficacy, teacher unfairness or parental emotional support).

Keywords: large scale assessment, school effectiveness, contextual effects, hierarchical linear models, academic achievement.

Introducción

El estudio de la eficacia escolar tiene un largo recorrido en el campo de la investigación educativa. Este ámbito se inició a partir de la publicación del Informe Coleman (Coleman, 1966), cuyas conclusiones sobre la escasa incidencia de la intervención escolar sobre el rendimiento de los alumnos, frente a variables de tipo socioeconómico, produjeron el desarrollo de numerosas investigaciones que trataron de ampliar el conocimiento sobre este novedoso tema. Ejemplo de ello son los estudios de Weber (1971), que añadió variables procesuales como el clima escolar o el liderazgo a un objeto de estudio que hasta entonces se centraba en factores contextuales, los trabajos de Brookover, Beady, Flood, Schewitzer y Wisenbaker (1979), quienes ampliaron la investigación de la influencia de los procesos escolares con el empleo de grandes muestras, o los hallazgos de Aitkin y Longford (1986), a partir del uso de los modelos estadísticos multinivel, realizando contribuciones clave que ayudaron a avanzar en el estudio de la eficacia escolar, estableciendo las bases de lo que hoy es un campo consolidado de investigación en educación.

Actualmente, se entiende por escuela eficaz aquella que “consigue un desarrollo integral de todos y cada uno de sus alumnos, mayor de lo que sería esperable teniendo en cuenta su rendimiento previo y la situación social, económica y cultural de las familias” (Murillo, 2005, p.25). Esta definición sienta las bases en tres aspectos fundamentales que caracterizan la investigación en este campo: la *equidad* (se busca el desarrollo de todos los alumnos), el *desarrollo integral* de los alumnos (no se busca sólo el rendimiento académico, sino la formación en valores y el bienestar de los estudiantes), y el *valor añadido* (se incluyen los elementos contextuales en el estudio del rendimiento) (Murillo, 2003). Aunque diversos autores utilizan una definición más amplia del término “valor añadido”, que implica la necesidad de evaluar los resultados de los centros educativos una vez controlados los efectos de las variables contextuales, ya que resulta una práctica más rigurosa y equitativa (Joaristi, Lizasoain y Azpillaga, 2014), existe la recomendación de restringir el uso del término a los estudios de carácter longitudinal (OECD, 2008), por lo que en el presente estudio nos referiremos al modelo empleado como “modelo contextualizado sin ganancia” debido a la falta de datos de naturaleza longitudinal.

Este tema de estudio ha ganado mayor interés en la comunidad científica en las últimas décadas (Gamazo, Olmos-Migueláñez y Martínez-Abad, 2016), debido, en parte, a la disponibilidad de las bases de datos procedentes de evaluaciones a gran escala (OCDE, 2017). Estas pruebas no sólo proveen datos sobre el rendimiento de los alumnos en distintas competencias, sino que también ofrecen un extenso catálogo de datos contextuales, tanto de los alumnos como de los centros educativos, que sirven para realizar estudios en profundidad acerca de los factores que influyen más en el rendimiento, o en la eficacia de las escuelas. Sin embargo, algunos autores apuntan que la validez y fiabilidad de los instrumentos varían, entre las altas garantías técnicas de las pruebas de rendimiento y el deficiente diseño de los cuestionarios de contexto (De la Orden y Jornet, 2012).

Existen diversas pruebas a gran escala, aplicadas a nivel internacional, que pueden servir como base para realizar este tipo de estudios. El Estudio Internacional de Progreso en Comprensión Lectora (PIRLS) y el Estudio Internacional de Tendencias en Matemáticas y Ciencias (TIMSS), de la Asociación Internacional para la Evaluación del Rendimiento Educativo (IEA), o el Programa para la Evaluación Internacional de Alumnos (PISA), del Organismo para la Cooperación y Desarrollo Económico (OCDE), son ejemplos de pruebas estandarizadas internacionales que miden el nivel competencial de los alumnos, a la vez que recogen datos sobre su entorno personal, familiar, escolar y social.

El acceso abierto a los resultados de estas pruebas facilita el estudio del impacto de los factores culturales, económicos, sociales, educativos y personales en el rendimiento de los alumnos y, a través de la investigación acerca de estos factores, permite también el análisis de la eficacia de los centros educativos.

Factores asociados al rendimiento

Los factores que más comúnmente se asocian al rendimiento de los alumnos, y aquellos cuya información es proporcionada por los cuestionarios de contexto de las pruebas competenciales mencionadas anteriormente, son de diversa naturaleza. Varios autores como Murillo (2007) o Jornet, González-Such y Perales (2012) catalogan estos factores: de entrada (género, nivel socio-económico, lengua materna, recursos del

centro, etc.), de proceso (hábitos de estudio, expectativas académicas, apoyo familiar, clima del centro, metodología docente, etc.) y de producto (rendimiento académico). A su vez, estos factores pueden dividirse en dos niveles: alumnos y centros educativos.

Diversas investigaciones han comprobado el efecto significativo de algunos factores personales de los *alumnos*. Algunos pertenecen a la categoría de variables de entrada, también denominados factores de contexto, como el género, el estatus migratorio o la asistencia a educación infantil (Karakolidis, Pitsia y Emvalotis, 2016; Özdemir, 2016), el índice socio-económico (Cordero, Manchón y Simancas, 2014; Ehmke, Drechsel, y Carstensen, 2008), la lengua materna (Özdemir, 2016; Riederer y Verwiebe, 2015) o el nivel educativo y la ocupación de los padres (Riederer y Verwiebe, 2015; Tsai, Smith, y Hauser, 2017). Otros se clasifican como variables de proceso, o factores no contextuales, incluyendo factores emocionales y motivacionales como la auto-eficacia (Aksu y Güzeller, 2016), la ansiedad o el autoconcepto (Karakolidis et al., 2016; Risso, Peralbo y Barca, 2010), las oportunidades para aprender en casa (Liu y Whitford, 2011; Santibañez y Fagioli, 2016), los hábitos y las estrategias de estudio (Risso et al., 2010; Santos, Godás y Lorenzo, 2013) o el apoyo familiar (King et al., 2005; Santos et al., 2013).

Los factores *escolares* también se exploran en este tipo de estudios, aunque existe menos unanimidad sobre los efectos significativos de algunas variables (Choi y Calero, 2012; Martínez-Abad y Chaparro-Caso, 2017). Sin embargo, existen datos a favor de la influencia de algunos factores de entrada como el nivel socio-económico y cultural medio del centro (Perry y McConney, 2010a, 2010b), el tamaño de la escuela o la ratio profesor-alumno (Nath, 2012), así como de factores de proceso como el agrupamiento de los alumnos según su desempeño académico (Kunz, 2014; Meunier, 2011), la metodología docente (Nath, 2012; Payandeh-Najafabadi, Omid-Najafabadi, y Farid-Rohani, 2013) o el ambiente de aprendizaje (Payandeh-Najafabadi et al., 2013; Santos et al., 2013).

Al introducir los factores internos del alumno (de tipo personal, familiar, cultural, económico y social) y los externos pertenecientes al entorno escolar en un modelo explicativo del rendimiento académico, deberían estar orientados por un modelo de calidad de carácter sistémico que relacione estas dimensiones más o menos complejas, para que puedan ser sometidas a prueba con los datos empíricos recogidos (De la Orden y Jornet, 2012; Jornet, et al., 2012).

Técnicas estadísticas para la investigación de la eficacia escolar

La complejidad en el estudio de los factores asociados al rendimiento y la eficacia escolar, producida por su elevado número y la red de relaciones establecidas entre ellos (Tejedor, 2003), ha resultado en el uso de una gran variedad de técnicas estadísticas para abordar este objeto de investigación.

Por un lado está el análisis multinivel, utilizado en el curso de este estudio, cuyas características permiten diferenciar la variabilidad aportada por cada nivel de agregación existente en los datos de naturaleza jerárquica, distinción sin la cual se podría incurrir en errores de sobreestimación de coeficientes o de interpretación de los datos (Snijders y Bosker, 2012). Por ello, su uso se recomienda en los casos en los que los datos presentan una estructura anidada, como son los datos procedentes de las evaluaciones a gran escala, en los que los estudiantes se encuentran reunidos en estructuras de nivel superior (Lenkeit, 2013; Lizasoain y Angulo, 2014; Martínez-Arias, 2009; Murillo y Hernández, 2011).

Otro ejemplo de técnicas cuantitativas son los modelos de ecuaciones estructurales, que facilitan el establecimiento de relaciones entre variables predictoras y variables criterio, e introducir otras variables de tipo latente o dimensiones (análisis factoriales confirmatorios), es decir, constructos que no pueden medirse directamente, pero que pueden estudiarse a través del análisis de otras variables observables (Castro y Lizasoain, 2012). Las técnicas de minería de datos también son propicias para el análisis de datos procedentes de grandes evaluaciones, ya que permiten extraer información relevante, como patrones o relaciones significativas entre variables, de bases de datos con grandes cantidades de información (Castro y Lizasoain, 2012).

Partiendo de este estado de la cuestión, el *objetivo* de esta investigación es doble. Por una parte, se pretende estudiar el efecto de determinadas variables de entrada (contextuales) en el rendimiento, analizando la variabilidad de los centros en cuanto a su eficacia, medida a través de la diferencia entre su puntuación real y esperada en función de dichas variables. En segundo lugar, se analizarán las variables de proceso (no contextuales) que mayor poder de discriminación presenten sobre el residuo de los centros educativos.

Para cada uno de los objetivos el método analítico será distinto. En el primer caso, se aplican *modelos multinivel* para estudiar el efecto de las variables contextuales, o de entrada, en el rendimiento de los alumnos,

y seleccionar aquellos centros cuya puntuación real se encuentre significativamente por encima (o por debajo) de su puntuación esperada en función de sus características demográficas y socio-económicas. Estos centros, denominados de alto y bajo residuo o eficacia, sirven como base para la segunda parte del estudio, en la que se analiza, a través de técnicas de *regresión logística*, qué variables no contextuales, o de proceso, influyen de manera más significativa en el nivel de residuo de los centros educativos, revelando así qué factores tienen una mayor relación con la eficacia escolar.

Método

Este análisis secundario de los datos de PISA 2015 presenta un diseño de carácter ex-post-facto o no experimental, dado que no existe un control experimental sobre las variables recogidas para su estudio. En esta sección se presentan la muestra participante y los instrumentos de recogida de datos, así como las técnicas de análisis de datos empleadas.

Muestra

La muestra de este estudio, extraída de la base de datos proporcionada por la OECD (2017), está compuesta por todos aquellos alumnos de 15 años (nacidos entre enero y diciembre de 1999) que participaron en las pruebas PISA 2015 en España. Aunque esta muestra inicial estuvo compuesta por 32.330 alumnos y 976 centros, los estudiantes procedentes de centros educativos con menos de 20 participantes en PISA 2015, así como los propios centros, fueron eliminados de la muestra para asegurar el correcto análisis de las variables agregadas a nivel de centro, tal como se ha realizado en otros estudios (Joaristi, Lizasoain y Azpillaga, 2014; Martínez-Abad, Lizasoain, Castro y Joaristi, 2017; Meunier, 2011). En total, se contó con 31.273 estudiantes, de los cuales un 49,4% (15.437) eran mujeres, y un 50,6% (15.836) eran hombres. Estos estudiantes proceden de 897 centros educativos. En esta edición de la prueba, todas las Comunidades Autónomas decidieron ampliar la muestra estatal para poder comparar sus datos a nivel internacional (Ministerio de Educación, Cultura y Deporte, 2016).

La distribución de alumnos según su Comunidad Autónoma de procedencia resulta uniforme (entre un 4,5% y un 6% del total de participantes), excepto en el caso del País Vasco, que aporta un 10,7% de participantes. Los centros educativos muestran una distribución parecida.

De los 897 centros participantes, un 66.6% pertenecen a la red pública, un 28.2% a la privada concertada, y un 5.2% a la privada (exceptuando los 49 valores perdidos para esta variable).

Instrumentos

Para llevar a cabo esta investigación se recurrió a los instrumentos creados específicamente para las pruebas PISA, que son fundamentalmente de dos tipos. Por un lado, para medir el nivel competencial de los alumnos se utilizan las pruebas de evaluación que, en el caso de 2015, midieron el rendimiento de los alumnos en lectura, matemáticas y ciencias. Estas pruebas están compuestas de baterías de distintos tipos de ítems que pueden tener tres tipos de respuestas: abiertas (explicar los pasos para solucionar un problema, por ejemplo), cerradas (respuestas numéricas, o de una sola palabra) o de elección múltiple.

Por otro lado, también se utilizan los datos extraídos de los *cuestionarios de contexto* aplicados a estudiantes, padres y centros educativos. Estos cuestionarios ofrecen gran cantidad de datos acerca de cuestiones socio-económicas, culturales y demográficas, y también informan sobre otros temas de interés educativo, como el clima del centro, la motivación de los estudiantes, la formación del profesorado, o las prácticas evaluativas del centro.

Variables

Las variables *criterio* para la realización de los modelos multinivel son las puntuaciones obtenidas por los alumnos en las pruebas competenciales de lectura, matemáticas y ciencias. Estas variables se definen de la siguiente manera en los documentos guía de las pruebas PISA 2015 (OCDE, 2016):

- Lectura: “La habilidad de los estudiantes para comprender, utilizar, reflexionar e interactuar con textos escritos con el objetivo de alcanzar sus metas, desarrollar su conocimiento y potencial, y participar en la sociedad” (p. 13).
- Matemáticas: “La habilidad de los estudiantes para formular, emplear e interpretar las matemáticas en una variedad de contextos. Esto incluye el razonamiento matemático y el uso de conceptos, procedimientos, hechos y herramientas matemáticas para describir, explicar y predecir fenómenos.” (p. 13).
- Ciencias: “Una persona con competencia científica está dispuesta a comprometerse con un discurso razonado sobre ciencia y tecnología, lo cual requiere habilidades para explicar fenómenos de manera científica, evaluar y diseñar investigaciones científicas, e interpretar datos y evidencias de manera científica.” (p. 13).

Para este análisis se incluyeron los 10 valores plausibles proporcionados por alumno para cada competencia. Estos valores se obtienen a través de una metodología de imputación para estimar el nivel de rendimiento de un estudiante a partir de las puntuaciones obtenidas en los ítems. También fueron incluidas en el análisis las ponderaciones muestrales tanto de alumnos como de centros, para asegurar el correcto tratamiento de los datos y el cálculo apropiado del error de muestreo (OECD, 2012).

Por su parte, los factores contextuales extraídos de los cuestionarios de contexto fueron utilizados como variables *predictoras* de los modelos. Estas variables, cuya selección se basó en la revisión de la literatura realizada en el marco teórico, se dividen en dos grupos: nivel 1 (alumnos) y nivel 2 (centros educativos) (Tabla I). En el caso de los datos nominales u ordinales, se generaron variables dummy (tantas como el número de categorías de la variable original menos una, siendo la categoría de referencia siempre la más frecuente).

TABLA I. Variables predictoras del modelo multinivel.

	Variable	Etiqueta	Rango
Nivel 1 - Alumnos	Género	NIGEN	0: Masculino 1: Femenino
	Mes de nacimiento	NIBMONTH	1-12
	Curso	NIGRADE	1º ESO – 1º Bachillerato
	Índice socio-económico y cultural (ESCS)	NIESCS	Continua
	Estatus migratorio	NIIMMIG	0: Nativo, 1: Inmigrante de 2ª generación 2: Inmigrante de 1ª generación
	Condición de repetidor	NIREPEAT	0: No 1: Sí
	Número de cambios de centro educativo	NISCCH	0: Ningún cambio 1: Un cambio 2: Dos o más cambios
	Idioma hablado en el hogar	NIIDIOMA	0: Idioma de la prueba 1: Otro idioma
	Nivel 2 - Centros	Tamaño del centro	N2TAMESC
Tamaño de las clases		N2TAMCLS	Continua
Escasez de recursos		N2ESCREC	Continua
Escasez de profesorado		N2ESCPER	Continua
Titularidad del centro		N2TITESC	1: Privado 2: Concertado 3: Público
Ratio profesor-alumnos		N2RATIO	Continua
ESCS medio		N2ESCS	Continua
Tasa de repetidores		N2REPETI	Continua
Tasa de alumnos inmigrantes		N2IMMIG	Continua
Porcentaje de chicas		N2PCGIRL	Continua

Fuente: Elaboración propia.

Por otro lado, las variables no contextuales también fueron utilizadas en una fase posterior del análisis (ver sección sobre procedimiento). Se seleccionaron las variables ofrecidas en forma de índices (es decir, aquellas que agrupan información de varios ítems en un solo factor) tanto

del nivel de estudiantes como de centros. Las variables de nivel 1 analizadas fueron motivación, disposición al trabajo colaborativo, creencias epistemológicas, disfrute e interés por la ciencia, competencia, interés y autonomía en el uso de las nuevas tecnologías, clima de disciplina en el aula, apoyo del profesor, conciencia y optimismo medioambiental, estatus ocupacional esperado, pertenencia al centro, apoyo emocional y académico parental, retroalimentación percibida e imparcialidad del profesorado. Las variables de nivel 2 empleadas fueron liderazgo, desarrollo del currículum, desarrollo profesional, responsabilidad del centro sobre los recursos y el currículum, participación del profesorado, autonomía escolar y actividades extraescolares de carácter creativo.

Procedimiento

Para obtener los centros educativos de alta y baja eficacia, se aplicaron *modelos jerárquicos lineales* (Snijders y Bosker, 2012), que permitieron identificar el efecto de los factores contextuales sobre el rendimiento promedio de los centros. Se definió un modelo para cada competencia evaluada, manteniendo únicamente las variables predictoras significativas ($\alpha=.05$) en los dos niveles, lo cual permitió calcular la diferencia entre la puntuación real de los centros y la esperada según su nivel socio-económico y cultural, también denominada “residuo” de los centros, obtenido a través de estimadores bayesianos empíricos (Raudenbush, Bryk, Cheong, Congdon, y Du Toit, 2011). Tras el estudio previo de la colinealidad de las variables predictoras, y partiendo de las evidencias recogidas por Özdemir (2016), en las que no se localizan diferencias sustanciales entre modelos multinivel de pendientes fijas y aleatorias a partir de datos de las pruebas PISA, se decidió introducir las variables contextuales en el primer nivel como covariables de efectos fijos, sin incluir los efectos aleatorios.

Posteriormente, se elaboró un protocolo con los criterios que debían cumplir los centros para ser seleccionados como de alto o bajo residuo partiendo, por un lado, del trabajo de Joaristi et al. (2014) en el que fue fijado el percentil 80 como región límite para los 6 modelos estimados. Por otro lado, se tiene en cuenta el trabajo de Martínez-Abad et al. (2017), en el que se optó por establecer como límite inferior (alta eficacia) el

percentil 66 y como límite superior (baja eficacia) el 33, requiriendo el cumplimiento en 5 de los 8 modelos. Así, y dados los datos disponibles, el presente trabajo optó por un criterio ecléctico entre ambos procedimientos, tomando los percentiles 33 y 66 como regiones límite, pero estableciendo la necesidad de su cumplimiento en los 3 modelos obtenidos finalmente.

A continuación, se analizó la correlación entre las principales variables de proceso (nivel 1 y 2) y la variable dicotómica generada (centro de alto o bajo residuo, variable criterio) a través del coeficiente de correlación biserial puntual. En caso de alcanzarse correlaciones significativas ($\alpha=.05$), y tras descartarse efectos importantes de colinealidad entre las variables predictoras del modelo (de proceso), se procedió con la aplicación de *técnicas regresión logística* para el estudio multivariante de la relación de las variables de proceso (predictoras) con la variable anterior. Estas técnicas son las más recomendables para variables criterio de naturaleza dicotómica.

La construcción de los modelos jerárquicos lineales fue realizada con el programa HLM7², y la regresión logística se llevó a cabo con el paquete estadístico SPSS v.20³.

Resultados

Inicialmente, se calculó el modelo nulo, que consiste en un modelo incondicional sin ninguna variable predictora (Hayes, 2006; Lee, 2000). La estimación de los componentes de varianza de este modelo permite el cálculo del Coeficiente de Correlación Intraclase (CCI), que representa la proporción de varianza atribuible al segundo nivel (Snijders y Bosker, 2012). Para que se considere adecuado utilizar modelos multinivel, el valor del CCI debe ser superior al 10% (Lee, 2000), que se define por la siguiente ecuación (1):

$$CCI = \frac{\tau_{00}}{\tau_{00} + \sigma^2} \quad (1)$$

²) Licencia comercial del Grupo de Investigación en Evaluación Educativa y Orientación, Universidad de Salamanca.

³) Licencia comercial del campus de la Universidad de Salamanca.

El término τ_{00} se refiere a la variabilidad entre los centros y σ^2 representa la variabilidad entre los alumnos.

Los datos de la muestra española de PISA 2015 presentan un CCI del 12.26% en matemáticas, 12.04% en lectura y un 12.41% en ciencias, por lo que se estima oportuno el procedimiento basado en modelos multinivel.

Rendimiento en matemáticas

Tras aplicar el procedimiento descrito en la sección anterior, el modelo multinivel para el rendimiento de los alumnos en las pruebas competenciales de matemáticas queda definido por la siguiente ecuación (2):

$$PV1MATH_{ij} = \gamma_{00} + \gamma_{01} * N2ESCS_j + \gamma_{10} * N1GEN_{ij} + \gamma_{20} * N1ESCS_{ij} + \gamma_{30} * N1REPEAT_{ij} + \gamma_{40} * N1GRADE_{ij} + \gamma_{50} * N1IMM2_{ij} + \gamma_{60} * N1SCCH1_{ij} + \gamma_{70} * N1SCCH2_{ij} + u_{0j} + r_{ij} \quad (2)$$

En la tabla II se pueden observar los coeficientes, valores t y valores de significación de las variables finalmente incluidas en el modelo de la competencia matemática.

TABLA II. Estimación de los efectos fijos (con errores estándares robustos), competencia matemática.

Efecto Fijo	Coefficiente	Error estándar	t-ratio	p-valor
Para INTRCPT1, β_0				
INTRCPT2, γ_{00}	543.097	1.731	313.758	<.001
ESCS CENTROS, γ_{01}	12.986	1.906	6.813	<.001
Para GÉNERO slope, β_1				
INTRCPT2, γ_{10}	-22.599	1.972	-11.457	<.001
Para ESCS slope, β_2				
INTRCPT2, γ_{20}	9.799	1.123	8.723	<.001
Para REPETICIÓN slope, β_3				
INTRCPT2, γ_{30}	-35.721	4.972	-7.184	<.001
Para CURSO slope, β_4				
INTRCPT2, γ_{40}	34.960	3.874	9.024	<.001
Para INMIGRANTE 1ª GEN. slope, β_5				
INTRCPT2, γ_{50}	-11.785	4.863	-2.423	.026
Para CAMBIOS DE CENTRO (I) slope, β_6				
INTRCPT2, γ_{60}	-10.690	1.870	-5.716	<.001
Para CAMBIOS DE CENTRO (>I) slope, β_7				
INTRCPT2, γ_{70}	-15.512	3.303	-4.696	<.001

Fuente: Elaboración propia.

Mientras que a nivel de escuela sólo se incluye el nivel socio-económico medio del centro, a nivel de alumnos hay varias variables con influencia significativa, siendo el género la variable con un valor t más alto, seguida del curso, el índice socio-económico y cultural y la condición de repetidor. La variable relativa al estatus migratorio de los alumnos indica que solo la condición de inmigrante de 1ª generación resulta significativa.

Tras la estimación del modelo, el CCI para la competencia matemática se sitúa en 4.55%, por lo que podemos afirmar que las variables incluidas explican un 7.71% de la varianza entre los centros.

Rendimiento en lectura

En el caso de la competencia en comprensión lectora, las variables que finalmente han compuesto el modelo multinivel se encuentran en la siguiente ecuación (3):

$$PVIREAD_{ij} = \gamma_{00} + \gamma_{01} * N2ESCS_j + \gamma_{02} * N2REPETI_j + \gamma_{03} * N2PCGIRL_j + \gamma_{10} * N1GEN_{ij} + \gamma_{20} * N1ESCS_{ij} + \gamma_{30} * N1REPEAT_{ij} + \gamma_{40} * N1GRADE_{ij} + \gamma_{50} * N1SCCH1_{ij} + \gamma_{60} * N1SCCH2_{ij} + \gamma_{70} * N1IDIOMA_{ij} + u_{0j} + r_{ij} \quad (3)$$

La tabla III muestra los valores de magnitud y significación de la relación entre las variables que componen el modelo y el rendimiento de los alumnos en la prueba de comprensión lectora.

TABLA III. Estimación de los efectos fijos (con errores estándares robustos), comprensión lectora.

Efecto Fijo	Coefficiente	Error estándar	t-ratio	p-valor
Para INTRCPT1, β_0				
INTRCPT2, γ_{00}	522.475	6.153	84.912	<.001
ESCS CENTROS, γ_{01}	20.293	2.658	7.633	<.001
TASA DE REPETIDORES, γ_{02}	36.502	10.829	3.371	<.001
PORCENTAJE DE ALUMNAS, γ_{03}	24.719	10.572	2.338	.020
Para GÉNERO slope, β_1				
INTRCPT2, γ_{10}	6.776	1.849	3.665	<.001
Para ESCS slope, β_2				
INTRCPT2, γ_{20}	8.299	0.949	8.748	<.001
Para REPETICIÓN slope, β_3				
INTRCPT2, γ_{30}	-32.120	5.162	-6.222	<.001
Para CURSO slope, β_4				
INTRCPT2, γ_{40}	41.092	3.495	11.757	<.001
Para CAMBIOS DE CENTRO (1) slope, β_5				
INTRCPT2, γ_{50}	-8.857	2.127	-4.164	<.001
Para CAMBIOS DE CENTRO (>1) slope, β_6				
INTRCPT2, γ_{60}	-20.642	3.224	-6.403	<.001
Para IDIOMA slope, β_7				
INTRCPT2, γ_{70}	-6.718	3.026	-2.220	.033

Fuente: Elaboración propia.

A nivel de escuela, este modelo presenta más variables predictoras que el anterior, sumándose al ESCS medio tanto el porcentaje de alumnos repetidores como el porcentaje de chicas en la escuela. En el nivel 1, las variables con un valor t más alto son el curso, el ESCS, la condición de repetidor y el hecho de haber cambiado dos o más veces de escuela.

El CCI de las puntuaciones en comprensión lectora tras la aplicación del modelo es de 5.07%, lo cual supone que el modelo es capaz de explicar casi un 7% de la varianza de segundo nivel en esta competencia.

Rendimiento en ciencias

Por último, los factores que resultaron tener una relevancia significativa con la competencia científica se recogen en la siguiente ecuación (4):

$$\begin{aligned}
 PV1SCIE_{ij} = & \gamma_{00} + \gamma_{01} * N2TAMESC_j + \gamma_{02} * N2ESCPER_j + \gamma_{03} * N2ESCS_j + \\
 & \gamma_{04} * N2REPETI_j + \gamma_{05} * N2PCGIRL_j + \gamma_{10} * N1GEN_{ij} + \gamma_{20} * N1ESCS_{ij} + \\
 & \gamma_{30} * N1BMONTH_{ij} + \gamma_{40} * N1REPEAT_{ij} + \gamma_{50} * N1GRADE_{ij} + \gamma_{60} * N1IMM2_{ij} + \\
 & \gamma_{70} * N1SCCH1_{ij} + \gamma_{80} * N1SCCH2_{ij} + u_{0j} + r_{ij} \quad (4)
 \end{aligned}$$

En la tabla IV pueden consultarse los coeficientes derivados de la aplicación del modelo.

TABLA IV. Estimación de los efectos fijos (con errores estándares robustos), competencia científica.

Efecto Fijo	Coeficiente	Error estándar	t-ratio	p-valor
Para INTRCPT1, β_0				
INTRCPT2, γ_{00}	537.733	7.298	73.683	<.001
TAMAÑO ESCUELA, γ_{01}	-0.006	0.003	-2.226	.026
ESCASEZ DE PERSONAL, γ_{02}	2.513	1.163	2.161	.031
ESCS CENTROS, γ_{03}	20.708	2.605	7.949	<.001
TASA DE REPETIDORES, γ_{04}	31.815	9.545	3.333	<.001
PORCENTAJE DE ALUMNAS, γ_{05}	31.854	11.979	2.659	.008
Para GÉNERO slope, β_1				
INTRCPT2, γ_{10}	-19.782	1.686	-11.730	<.001
Para ESCS slope, β_2				
INTRCPT2, γ_{20}	9.792	0.704	13.912	<.001
Para MES DE NACIMIENTO slope, β_3				
INTRCPT2, γ_{30}	-0.590	0.228	-2.586	.011
Para REPETICIÓN slope, β_4				
INTRCPT2, γ_{40}	-34.223	3.874	-8.833	<.001
Para CURSO slope, β_5				
INTRCPT2, γ_{50}	38.389	2.898	13.248	<.001
Para INMIGRANTE 1ª GEN. slope, β_6				
INTRCPT2, γ_{60}	-9.596	2.886	-3.325	<.001
Para CAMBIOS DE CENTRO (1) slope, β_7				
INTRCPT2, γ_{70}	-10.417	1.945	-5.356	<.001
Para CAMBIOS DE CENTRO (>1) slope, β_8				
INTRCPT2, γ_{80}	-16.847	2.631	-6.404	<.001

Fuente: Elaboración propia.

El modelo resultante contiene varias variables significativas a nivel de centro, siendo la más relevante el ESCS medio, seguido de la tasa de repetidores. Otras variables incluidas en el modelo son el porcentaje de chicas en la escuela, el tamaño del centro y la escasez de personal docente.

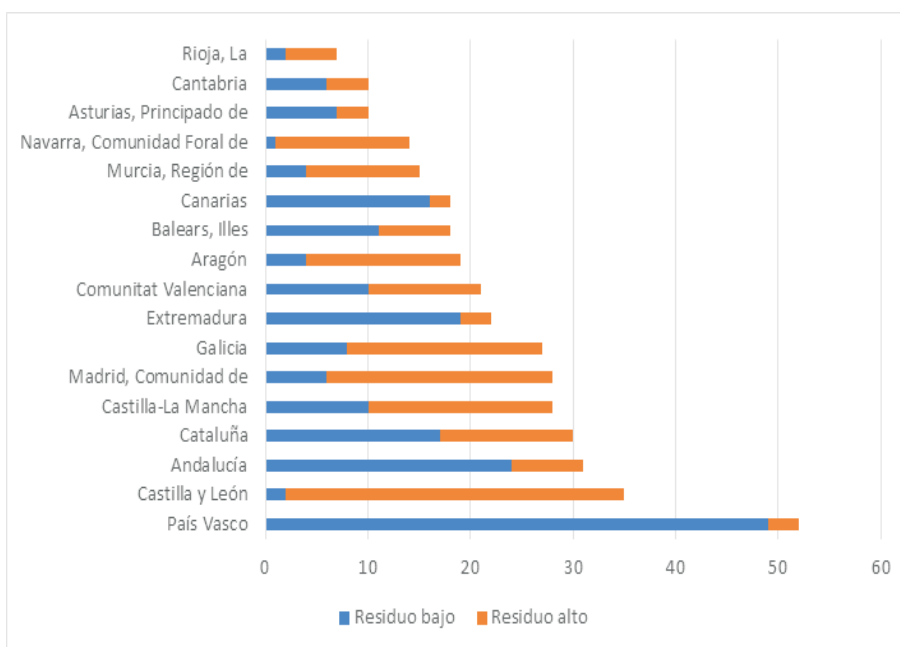
A nivel de alumnos, las variables con valores t más altos son el ESCS, el curso, el género y haber realizado cambios de escuela a lo largo de su recorrido educativo. De nuevo, la variable relativa a la inmigración indica diferencias significativas sólo para los estudiantes inmigrantes de 1ª generación.

Una vez aplicado el modelo, el cálculo del CCI del modelo final (5.6%) revela que el modelo multinivel para esta competencia ha conseguido explicar un 6.8% de la varianza de nivel 2.

Selección de centros

Para la selección de centros de alta y baja eficacia, se aplicó el procedimiento de selección expuesto anteriormente sobre las variables obtenidas como resultado de sustraer el rendimiento medio observado en los centros de la puntuación estimada por los modelos finales. Bajo este criterio fueron seleccionados 196 centros de residuo bajo y 189 de residuo alto. La figura I muestra la distribución de estos centros por Comunidad Autónoma.

FIGURA I. Distribución de los centros de alto y bajo residuo por Comunidad Autónoma.



Fuente: Elaboración propia.

Con respecto a la titularidad de los centros seleccionados, la mayoría de ellos (aproximadamente un 67%) pertenecen a la red pública, mientras que los centros privados y concertados componen alrededor de un tercio de la muestra, siendo los centros privados el grupo más reducido (tabla V). La prueba de bondad de ajuste confirma la correspondencia entre la muestra inicial y la selección ($\chi^2=0.016$; $p=.992$).

TABLA V. Ajuste de la selección de centros a la muestra inicial por titularidad.

		TITULARIDAD			Total
		Privada	Concertada	Pública	
Selección de centros	Recuento	19	102	245	366
	%	5,2%	27,9%	66,9%	100,0%
Muestra inicial	Recuento	44	239	564	847
	%	5,2%	28,2%	66,6%	100,0%

Fuente: Elaboración propia.

La tabla VI ofrece datos descriptivos del ESCS, tasa de repetidores, tasa de alumnos inmigrantes y puntuación esperada media para cada competencia, variables que caracterizan la población de centros seleccionados. Tanto los indicadores de las variables de contexto como los de las puntuaciones esperadas no difieren mucho entre ambos grupos de centros, no localizándose diferencias significativas entre las medias de ambos grupos para ninguna de las variables (prueba de t, $\alpha=.05$), excepto para la referida a la tasa de inmigrantes, que es significativamente superior en el grupo de escuelas de alto residuo ($p=.002$).

TABLA VI. Media, desviación típica, valor máximo y mínimo.

	Residuo bajo				Residuo alto			
	Med.	D.t.	Mín.	Máx.	Med.	D.t.	Mín.	Máx.
ESCS medio	-0,42	0,66	-1,92	0,97	-0,53	0,56	-1,55	1,14
T. repetidores	0,27	0,16	0,00	0,67	0,30	0,14	0,00	0,79
T. inmigrantes	0,09	0,13	0,00	0,69	0,13	0,11	0,00	0,58
P. esp. mates	537,60	8,77	517,75	556,01	536,18	7,37	522,67	558,19
P. esp. lectura	537,39	10,24	513,54	563,23	536,26	8,34	517,89	559,68
P. esp. ciencias	547,87	10,63	518,18	576,41	546,79	8,72	525,10	572,22

Fuente: Elaboración propia.

Efecto de las variables de proceso sobre la eficacia escolar (regresión logística)

El estudio correlacional inicial reveló que ninguna de las variables de proceso extraídas de los cuestionarios de contexto aplicados a los equipos directivos de los centros (nivel 2) presentaba una correlación significativa con la variable indicativa del residuo de los centros. Por este motivo, no se procedió con el análisis de regresión logística a nivel de centro.

Para el estudio a partir de las variables de proceso en el nivel de estudiante (nivel 1), las puntuaciones promedio de estas variables predictoras fueron agregadas a la base de datos de centros. En el estudio correlacional previo, buena parte de las variables presentaron correlaciones significativas con la variable agrupación del centro (en base al residuo). Así, el modelo de regresión logística obtenido, tras eliminar paso a paso las variables no significativas, quedó configurado con las variables reflejadas en la tabla VII y la ecuación 5.

TABLA VII. Regresión logística.

Variable	Contenido de la variable	B	Sig.	Odds Ratio
DISCLISCI	Clima de disciplina (ciencias)	1,435	,001	4,199
TEACHSUP	Apoyo del profesor (ciencias)	-1,868	,014	,154
ENVAWARE	Conciencia medioambiental	1,443	,018	4,235
SCIEEFF	Auto-eficacia (ciencias)	1,781	,001	5,939
EPIST	Creencias epistemológicas	2,017	,007	7,514
BSMJ	Estatus ocupacional esperado	-,091	,008	,913
BELONG	Pertenencia al centro	-1,628	,003	,196
COOPERATE	Disfrute de la cooperación	1,628	,029	5,095
EMOSUPS	Apoyo emocional parental	-3,606	,000	,027
PERFEED	Retroalimentación percibida	-1,652	,006	,192
ADINST	Adaptación de la instrucción	2,466	,004	11,770
USESCH	Uso de TIC en el centro	-1,330	,001	,264
AUTICT	Autonomía en el uso de las TIC	2,768	,001	15,920
SOIAICT	TIC como tema de interacción social	-2,569	,005	,077
unfairteacher	Injusticia del profesorado	-,882	,000	,414
Constant	Constante	14,223	,000	1502425,583

Fuente: Elaboración propia.

$$\hat{Y} = 14.223 + 1.435 * DISCLISCI - 1.868 * TEACHSUP + 1.443 * ENVAWARE + 1.781 * SCIEEFF + 2.017 * EPIST - 0.091 * BSMJ - 1.628 * BELONG + 1.628 * COOPERATE - 3.606 * EMOSUPS - 1.652 * PERFEED + 2.466 * ADINST - 2.569 * SOIAICT - 0.882 * unfairteacher \quad (5)$$

Entre las variables incluidas en el modelo, las que han resultado más significativas con un efecto positivo en la eficacia de los centros, es decir, aquellas cuyo incremento produce una mayor probabilidad de pertenecer al grupo de centros de residuo alto, han sido el clima de disciplina, la auto-eficacia, la autonomía en el uso de las TIC y la adaptación de la instrucción, mientras que aquellas que han presentado un efecto negativo y un alto nivel de significatividad han sido el nivel de injusticia del profesorado, el apoyo emocional parental, el uso de TIC en el centro, el sentimiento de pertenencia al centro y la retroalimentación percibida.

Cabe destacar que el modelo alcanza un buen ajuste, obteniéndose un $R^2=.527$ (Índice de Nagelkerke). Por otro lado, la precisión del modelo predictivo alcanza un 80,52% de clasificaciones correctas de centros de alto y bajo residuo, prediciendo correctamente un 79,59% en el caso de centros de bajo residuo y un 81.48% de los de alto residuo.

Discusión y conclusiones

La presente investigación se planteó con dos objetivos principales. El primero de ellos ha producido como resultado la construcción de tres modelos que permitieron detectar la magnitud y significación de la relación entre las variables de entrada tanto de los alumnos como de las escuelas en el rendimiento de los alumnos.

Las variables de nivel 1 incluidas en los modelos coinciden con los resultados de estudios previos, siendo una de las más influyentes el género (Karakolidis et al., 2016; Özdemir, 2016; Stoet y Geary, 2014), que en el caso de las matemáticas y las ciencias favorece a los varones, y en la comprensión lectora favorece a las mujeres, aunque en menor medida. Otras variables relevantes son el nivel socio-económico (Risso et al., 2010), la condición de repetidor (Choi y Calero, 2013; Ehmke et al.2008), o el número de cambios de escuela que haya sufrido un alumno a lo largo de su historial académico, variable cuya influencia se encuentra poco estudiada. Además, dos de los modelos (matemáticas y ciencias) apuntan al estatus migratorio como variable relevante, pero sólo para los estudiantes inmigrantes de 1ª generación, no encontrando diferencias entre los inmigrantes de 2º generación y los nativos. Esta discrepancia podría deberse a factores contextuales más amplios, como el declive en la inmigración de países hispanohablantes y el aumento de inmigración

de otras nacionalidades (Riederer y Verwiebe, 2015). Entre las variables de nivel 2 introducidas en los modelos, la única que ha resultado significativa de manera consistente en los tres modelos ha sido el ESCS medio del centro, variable cuya relevancia ha sido destacada por otros estudios (Perry y McConney, 2010a, 2010b). Además, se encontró que el impacto de las variables de entrada presenta diferencias según la variable de rendimiento utilizada. Por ejemplo, el género influye de manera diferente en lectura y matemáticas (Stoet y Geary, 2013), así como el estatus migratorio (Meunier, 2011) o el nivel socioeconómico medio del centro (Perry y McConney, 2010a) no afectan por igual a las tres competencias, por lo que las diferencias en este sentido encontradas en este estudio presentan precedentes en la literatura.

La distribución de los centros seleccionados a partir de estos modelos se ajusta adecuadamente a la muestra original en casi todas las variables de caracterización controladas. Sin embargo, la distribución según su Comunidad Autónoma no se ajusta del todo, sino que hay algunas Comunidades Autónomas infrarrepresentadas (Asturias, Cantabria, La Rioja) y otras, por el contrario, presentan una proporción de centros más alta en la selección que en la muestra de PISA (Andalucía, Castilla y León, País Vasco). Además, en algunas Comunidades se ha detectado un desequilibrio notable entre la proporción de centros de alto y de bajo residuo, indicando la pertinencia de profundizar en el estudio particular de estos casos para explorar qué factores regionales pueden causar estas desigualdades.

El segundo objetivo de la investigación era el análisis de las variables no contextuales en función de esta selección previa de centros, el cual reveló que no existían diferencias significativas en las puntuaciones de las variables procesuales obtenidas por los centros de alto y bajo residuo. Esta conclusión contrasta con algunos de los estudios cualitativos que concluyen que existen numerosas variables de proceso relacionadas de manera importante con la eficacia de los centros, como las investigaciones de Lizasoain y Angulo (2014), realizadas con centros de alta eficacia, o Murillo (2007), llevadas a cabo con escuelas de alto, medio y bajo residuo. Las causas para esta importante discrepancia pueden ser de diversa naturaleza. De la Orden y Jornet (2012) destacan la deficiencia de los cuestionarios que emplea PISA para recoger los datos de carácter contextual, lo cual podría dificultar un correcto análisis de la realidad educativa y conducir a conclusiones erróneas. Esta apreciación podría

apuntar la necesidad de buscar fuentes alternativas para los datos contextuales que doten a la investigación de un mayor grado de validez interna.

Las variables de nivel 1 que resultaron significativas en el modelo de regresión logística indican algunas cuestiones transversales, como el clima de disciplina en las aulas, la auto-eficacia, la adaptación de la instrucción o la autonomía en el uso de las TIC por parte de los alumnos, que sería conveniente trabajar desde las escuelas, ya que los análisis indican que las percepciones de los estudiantes en estas cuestiones están íntimamente relacionadas con el residuo positivo en la medición de la eficacia escolar. Aunque este análisis agregado presenta la desventaja de eliminar la varianza inter-centro, permite comparar las escuelas a nivel global a partir de las características de sus alumnos.

Las principales fortalezas de esta investigación son el empleo todas las garantías estadísticas recomendadas para la aplicación de modelos multinivel (consideración de la varianza a ambos niveles de análisis, uso de valores plausibles de la variable criterio e introducción de los pesos muestrales de alumnos y centros), la selección sistemática de los centros de alto y bajo residuo a través de las puntuaciones en las tres competencias básicas, y el alto porcentaje explicativo de los modelos finales, ya que los tres resultaron en una explicación de más del 50% de la CCI del modelo nulo. Además, la regresión logística presenta unos valores de bondad de ajuste muy altos.

Por otro lado, también presenta algunas debilidades, como la falta de sistematicidad en la selección de las variables para la realización de los modelos, la aparición de variables con un efecto opuesto al esperado en la regresión logística, o la falta de evidencias métricas suficientes de la fiabilidad y validez de los cuestionarios de contexto, una de las fuentes principales de información.

Los resultados obtenidos en este estudio sugieren la necesidad de profundizar en la investigación de los factores relacionados con la eficacia escolar en nuestro país, siempre desde la comparación de los centros de alta y baja eficacia con el objetivo de eliminar factores que puedan estar produciéndose en ambos ambientes y sean, por tanto, irrelevantes. Se propone, por tanto, el establecimiento de nuevas líneas de investigación que, partiendo de la selección de centros realizada en el presente estudio, se valgan de fuentes alternativas de datos no contextuales, como las técnicas de investigación cualitativas, para determinar los factores

escolares relevantes en el estudio de la eficacia escolar y poder sacar conclusiones pertinentes para la práctica educativa.

Referencias bibliográficas

- Aitkin, M., y Longford, N. (1986). Statistical modelling issues in school effectiveness studies. *Journal of the Royal Statistical Society. Series A (General)*, 149(1), 1-43.
- Aksu, G., y Güzeller, C. O. (2016). Classification of PISA 2012 Mathematical Literacy Scores Using Decision-Tree Method: Turkey Sampling. *Egitim ve Bilim*, 41(185), 101-122.
- Blanco-Blanco, Á., López Martín, E., y Ruiz de Miguel, C. (2014). Aportaciones de los modelos jerárquico-lineales multivariados a la investigación educativa sobre el rendimiento. Un ejemplo con datos del alumnado español en PISA 2009. *Revista de Educación*, 365, 122-149.
- Brookover, W.B., Beady, C., Flood, P., Schewitzer, J. y Wisenbaker, J. (1979). *School social systems and student achievement: schools can make a difference*. New York: Praeger.
- Castro, M., y Lizasoain, L. (2012). Las técnicas de modelización estadística en la investigación educativa: Minería de datos, modelos de ecuaciones estructurales y modelos jerárquicos lineales. *Revista Española de Pedagogía*, 70, 131-148.
- Choi, A., y Calero, J. (2013). Determinantes del riesgo de fracaso escolar en España en PISA-2009 y propuestas de reforma. *Revista De Educación*, 362, 562-593.
- Coleman, J. S. (1966). *Equality of educational opportunity* (Vol. 2). Washington, DC: US Department of Health, Education, and Welfare, Office of Education.
- Cordero Ferrera, J. M., Manchón López, C. y Simancas Rodríguez, R. (2014). La repetición de curso y sus factores condicionantes en España. Repetition and explanatory factors in Spain. *Revista de Educacion*, 365, 12-37.
- Creswell, J. W. (2015). Revisiting mixed methods and advancing scientific practices. En Hesse-Biber, S. N., y Johnson, R. B. (Eds.) *The Oxford*

- handbook of multimethod and mixed methods research inquiry*. (pp. 57-71). Oxford: Oxford University Press.
- De la Orden, A. y Jornet, J. M. (2012). La utilidad de las evaluaciones de sistemas educativos: el valor de la consideración del contexto. *Bordón*, 64(2), 69-88.
- Demir, I, Kılıç, S., y Ünal, H. (2010). Effects of students' and schools' characteristics on mathematics achievement: Findings from PISA 2006. *Procedia-Social and Behavioral Sciences*, 2(2), 3099-3103.
- Ehmke, T., Drechsel, B., y Carstensen, C. H. (2008). Klassenwiederholen in PISA-I-plus: Was lernen sitzenbleiber in mathematik dazu? *Zeitschrift Für Erziehungswissenschaft*, 11(3), 368-387.
- Gamazo, A., Olmos-Migueláñez, S., y Martínez-Abad, F. (2016, November). Multilevel models for the assessment of school effectiveness using PISA scores. In *Proceedings of the Fourth International Conference on Technological Ecosystems for Enhancing Multiculturalism* (pp. 1161-1166). ACM.
- Hayes, A. F. (2006). A primer on multilevel modeling. *Human Communication Research*, 32(4), 385-410.
- Joaristi, L., Lizasoain, L., y Azpillaga, V. (2014). Detección y caracterización de los centros escolares de alta eficacia de la Comunidad Autónoma del País Vasco mediante Modelos Transversales Contextualizados y Modelos Jerárquicos. *Estudios sobre Educación*, 27, 37-61.
- Jornet, J. M., González-Such, J., y Perales, M. J. (2012). Diseño de cuestionarios de contexto para la evaluación de sistemas educativos: optimización de la medida de constructos complejos. *Bordón*, 64(2), 89-110.
- Karakolidis, A., Pitsia, V., & Emvalotis, A. (2016). Examining students' achievement in mathematics: A multilevel analysis of the Programme for International Student Assessment (PISA) 2012 data for Greece. *International Journal of Educational Research*, 79, 106-115.
- King, G., McDougall, J., DeWit, D., Hong, S., Miller, L., Offord, D., . . . LaPorta, J. (2005). Pathways to children's academic performance and prosocial behaviour: Roles of physical health status, environmental, family, and child factors. *International Journal of Disability, Development and Education*, 52, 313-344.
- Kunz, J. S. (2014). Analyzing Educational Achievement Differences between Second Generation Immigrants: Comparing Germany and

- German Speaking Switzerland. *German Economic Review*, 17(1), 61-91.
- Lee, V. E. (2000). Using hierarchical linear modeling to study social contexts: The case of school effects. *Educational Psychologist*, 35(2), 125-141.
- Lenkeit, J. (2013). Effectiveness measures for cross-sectional studies: A comparison of value-added models and contextualised attainment models. *School Effectiveness and School Improvement*, 24(1), 1-25.
- Liu, X., y Whitford, M. (2011). Opportunities-to-learn at home: Profiles of students with and without reaching science proficiency. *Journal of Science Education and Technology*, 20(4), 375-387.
- Lizasoain, L., y Angulo, A. (2014). Buenas prácticas de escuelas eficaces del País Vasco. Metodología y primeros resultados. *Participación Educativa*, 3(4), 17-27.
- Martínez-Abad, F., y Chaparro-Caso, A. A. (2017). Data-mining techniques in detecting factors linked to academic achievement. *School Effectiveness and School Improvement*, 28(1), 39-55.
- Martínez-Abad, F., Lizasoain, L., Castro, M., & Joaristi, L. (2017). Selección de escuelas de alta y baja eficacia en Baja California (México). *REDIE. Revista Electrónica de Investigación Educativa*, 19(2), 38-53.
- Martínez-Arias, M. R. (2009). Usos, aplicaciones y problemas de los modelos de valor añadido en educación. *Revista De Educación*, 348, 217-252.
- Meunier, M. (2011). Immigration and student achievement: Evidence from Switzerland. *Economics of education review*, 30(1), 16-38.
- Ministerio de Educación, Cultura y Deporte (2016). *PISA 2015. Programa para la Evaluación Internacional de los Alumnos. Informe Español*. Madrid: Secretaría General Técnica.
- Murillo, F. J. (2003). El movimiento de investigación de Eficacia Escolar. En F.J. Murillo (Coord.), *La investigación sobre Eficacia Escolar en Iberoamérica. Revisión internacional del estado del arte*. Bogotá: Convenio Andrés Bello.
- Murillo, F. J. (2005). *La Investigación sobre Eficacia Escolar*. Barcelona: Octaedro.
- Murillo, F. J. (Coord.) (2007). *Investigación Iberoamericana sobre Eficacia Escolar*. Bogotá: Convenio Andrés Bello.

- Murillo, F. J., y Hernández, R. (2011). Efectos escolares de factores socio-afectivos. Un estudio multinivel para Iberoamérica. *Revista De Investigación Educativa*, 29(2), 407-427.
- Nath, S. R. (2012). Factors influencing primary students' learning achievement in Bangladesh. *Research in Education*, 88, 50-63.
- Organisation for Economic Co-operation and Development (OECD) (2008). *Measuring improvements in learning outcomes: best practices to assess the value-added of schools*. Paris: OECD Publishing.
- Organisation for Economic Co-operation and Development (OECD) (2014). *PISA 2012 Technical Report*. Paris: OECD Publishing
- Organisation for Economic Co-operation and Development (OECD) (2016). *PISA 2015 Assessment and Analytical Framework: Science, Reading, Mathematics and Financial Literacy*. Paris: OECD Publishing.
- Organisation for Economic Co-operation and Development (OECD) (2017), *PISA: Programme for International Student Assessment, OECD Education Statistics (database)*. DOI: <http://dx.doi.org/10.1787/data-00365-en>
- Özdemir, C. (2016). Equity in the Turkish education system: A multilevel analysis of social background influences on the mathematics performance of 15-year-old students. *European Educational Research Journal*, 15(2), 193-217.
- Payandeh-Najafabadi, A. T., Omidi-Najafabadi, M., y Farid-Rohani, M. R. (2013). Factors contributing to academic achievement: A Bayesian structure equation modelling study. *International Journal of Mathematical Education in Science and Technology*, 44(4), 490-500.
- Perry, L., y McConney, A. (2010a). Does the SES of the school matter? an examination of socioeconomic status and student achievement using PISA 2003. *Teachers College Record*, 112(4), 1137-1162.
- Perry, L., y McConney, A. (2010b). School socio-economic composition and student outcomes in Australia: Implications for educational policy. *Australian Journal of Education*, 54(1), 72-85.
- Raudenbush, S. W., Bryk, A. S., Cheong, Y. F., Congdon, R., & Du Toit, M. (2011). *Hierarchical linear and nonlinear modeling (HLM7)*. Lincolnwood, IL: Scientific Software International.
- Riederer, B., y Verwiebe, R. (2015). Changes in the educational achievement of immigrant youth in Western societies: The contextual effects of national (educational) policies. *European Sociological Review*, 31(5), 628-642.

- Risso, A., Peralbo, M., y Barca, A. (2010). Cambios en las variables predictoras del rendimiento escolar en Enseñanza Secundaria. *Psicothema*, 22(4), 790–796.
- Santibañez, L., y Fagioli, L. (2016). Nothing succeeds like success? Equity, student outcomes, and opportunity to learn in high-and middle-income countries. *International Journal of Behavioral Development*, 40(6), 517-525.
- Santos, M. A., Godás, A., y Lorenzo, M. (2013). Rendimiento académico y diversidad cultural: El eje lingüístico. *Revista Española de Pedagogía*, 256, 461–478.
- Snijders, T., y Bosker, R. J. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modeling* (2nd Edition). London: Sage Publications.
- Stoet, G., y Geary, D. C. (2013). Sex differences in mathematics and reading achievement are inversely related: Within-and across-nation assessment of 10 years of PISA data. *PLoS One*, 8(3), e57988.
- Tejedor, F. J. (2003). Poder explicativo de algunos determinantes del rendimiento en los estudios universitarios. *Revista Española de Pedagogía*, 61(224), 5–32.
- Tsai, S. L., Smith, M. L., & Hauser, R. M. (2017). Families, Schools, and Student Achievement Inequality: A Multilevel MIMIC Model Approach. *Sociology of Education*, 90(1), 64-88.
- Weber, G. (1971). *Inner-city children can be taught to read: four successful schools*. Washington, DC: Council for Basic Education.

Información de contacto: Adriana Gamazo. Universidad de Salamanca, Facultad de Educación, Departamento de Didáctica, Organización y Métodos de Investigación. Paseo de Canalejas 169, 37008, Salamanca E-mail: adrianagamazo@usal.es.

Assessment of factors related to school effectiveness in PISA 2015. A multilevel analysis¹

Evaluación de factores relacionados con la eficacia escolar en PISA 2015. Un análisis multinivel

DOI: 10.4438/1988-592X-RE-2017-379-369

Adriana Gamazo
Fernando Martínez-Abad
Susana Olmos-Migueláñez
María José Rodríguez-Conde

Universidad de Salamanca

Abstract

The study of school effectiveness has gained relevance in the past few decades. The availability of data pertaining both to student performance and to the socioeconomic, demographic, organisational and educational characteristics of students and schools has allowed for the proliferation of studies on the relationship of all kinds of variables with student performance, and on the essential practices necessary to provide a quality and equal education.

This research is focused on the study of school effectiveness. To this end, hierarchical linear models (multilevel) are implemented with math, reading and science performance data from the Spanish sample of PISA 2015, aiming to establish which contextual factors have a larger effect on student performance. Gender, socio-economic level, grade, grade repetition, and school changes, together with the school's average socio-economic level, are the variables that consistently appeared as relevant in all three models.

¹ Research work funded through the project EDU2015-64524-P (Ministry of Economy and Competitiveness/FEDER)

This information was used to select the schools with the highest and lowest levels of effectiveness, calculated as the difference between their observed performance scores and their expected scores according to the relevant contextual factors. This selection allows for a study on which non-contextual factors (at student and school levels) are related to school effectiveness. There were no significant relationships with the school level variables, although there were some at student level (classroom discipline, self-efficacy, teacher unfairness or parental emotional support).

Keywords: large scale assessment, school effectiveness, contextual effects, hierarchical linear models, academic achievement.

Resumen

El estudio de la eficacia escolar ha ganado relevancia en las últimas décadas. La disponibilidad de datos relativos tanto al rendimiento del alumnado como a las características socio-económicas, demográficas, organizativas y educativas de los alumnos y los centros ha permitido la proliferación de estudios sobre la relación de todo tipo de variables con el rendimiento, y sobre las prácticas que resultan fundamentales para poder ofrecer una educación equitativa y de calidad.

La presente investigación se centra en el estudio de la eficacia escolar, utilizando para ello la aplicación de modelos jerárquicos lineales (multinivel) con los datos de rendimiento de matemáticas, lectura y ciencias de la muestra española de PISA 2015 con el objetivo de determinar qué factores contextuales tienen más efecto en el rendimiento de los estudiantes. El género, el nivel socio-económico, el curso, la repetición de curso y los cambios de escuela, junto con el nivel socio-económico medio del centro, son las variables que aparecen como relevantes consistentemente en los tres modelos realizados.

Dicha información se utiliza para realizar una selección de centros de alta y baja eficacia basada en la diferencia entre la puntuación observada de los centros y su puntuación esperada en función de los factores contextuales relevantes. A partir de esta selección, se realiza un estudio de los factores no contextuales a nivel de estudiante y de centro que se encuentran relacionados con la eficacia de los centros. No se encuentran relaciones significativas con las variables del nivel de centro, aunque sí con algunas a nivel de estudiante (clima de disciplina, auto-eficacia, apoyo emocional parental o nivel de injusticia del profesorado).

Palabras clave: evaluación a gran escala, eficacia escolar, efectos contextuales, modelos jerárquicos lineales, rendimiento académico.

Introduction

The study of school effectiveness has a long history within the field of educational research. Its scientific exploration began with the publication of the Coleman Report (Coleman, 1966), whose conclusions on the small impact of school intervention on student performance in comparison with socioeconomic variables promoted the development of numerous studies that aimed to expand the knowledge basis of this new subject. Some examples of the studies that emerged in the wake of this report are those conducted by Weber (1971), who introduced process variables such as school climate or leadership to a study field that, up until that point, only considered context factors, the works by Brookover, Beady, Flood, Schewitzer and Wisenbaker (1979), which expanded the research by employing large samples, or the findings of Aitkin and Longford (1986), obtained through the use of multilevel statistical analyses. All these authors provided the field of school effectiveness with key contributions, and laid the foundations for what is nowadays considered a well-established topic within educational research.

Currently, a school is defined as effective when it “achieves a comprehensive and integral development of each and every one of its students, even higher than it would be expected taking into consideration their previous performance and the social, financial and cultural situation of their families” (Murillo, 2005, p.25). This definition lays out three key aspects that characterise the research in this field: *equity* (every student’s development is sought), *integral development* of the students (we do not only seek to promote academic performance, but also values and student well-being), and *added value* (contextual elements are included in the study of performance) (Murillo, 2003). Several authors employ a broader definition of the term “value added”, which advocates the need to assess the results of the schools once the effects of the contextual variables have been controlled, as it results in a more equitable and rigorous practice (Joaristi, Lizasoain & Azpillaga, 2014). However, it is recommended to limit the use of this term to longitudinal studies (OECD, 2008), therefore the present study refers to the model employed as “contextualised model without gain” due to the lack of longitudinal data.

This study subject has achieved a growing interest in the scientific community in the last few decades (Gamazo, Olmos-Migueláñez & Martínez-Abad, 2016), due, in part, to the availability of datasets from

large-scale assessments (OCDE, 2017). Not only do these tests provide us with data on the performance of the students in different competences, but they also offer a broad catalogue of contextual data, both from students and schools, which enables in-depth studies on the factors that most influence performance, or even school effectiveness. However, some authors suggest that the reliability and validity of the instruments vary, ranging from the high technical guarantees of performance tests to the flawed design of the context questionnaires (De la Orden & Jornet, 2012).

There are many large-scale assessments conducted at an international level which can serve as a basis to carry out this kind of studies. The Progress in international Reading Literacy Study (PIRLS) and the Trends in International Mathematics and Science Study (TIMSS), both from the International Association for the Evaluation of Educational Achievement (IEA), or the Programme for international Student Assessment (PISA) from the Organisation for Economic Co-operation and Development (OECD), are a few examples of international standardised tests that measure the level of competence development among students, and that also gather data on their personal, family, school and social backgrounds.

The open access to the results of these tests enables the study of the impact of the cultural, financial, social, educational and personal factors on student performance, and, through the research of these factors, it also allows us to analyse the effectiveness of schools.

Factors related to student performance

The factors that are most commonly linked to student performance, and those whose information is also provided by the context questionnaires of the abovementioned competence tests, vary in nature. Several authors, such as Murillo (2007) or Jornet, González-Such and Perales (2012) classify these factors in three categories: input (gender, socioeconomic level, mother tongue, school resources, etc.), process (study habits, academic expectations, family support, school climate, teaching methodology, etc.) and product (academic performance). In turn, these factors can also be divided in two levels: student and school.

Several studies have verified the significant effect of some personal *student* factors. Some of them belong to the category of input factors, also referred to as contextual factors, such as gender, migratory status, pre-school

education attendance (Karakolidis, Pitsia & Emvalotis, 2016; Özdemir, 2016), socioeconomic index (Cordero, Manchón & Simancas, 2014; Ehmke, Drechsel, & Carstensen, 2008), mother tongue (Özdemir, 2016; Riederer & Verwiebe, 2015) or the education and occupation of the parents (Riederer & Verwiebe, 2015; Tsai, Smith, & Hauser, 2017). Others are classified as process, or non-contextual factors, including emotional and motivational variables such as self-efficacy (Aksu & Güzeller, 2016), anxiety or self-concept (Karakolidis et al., 2016; Risso, Peralbo & Barca, 2010), the opportunities to learn at home (Liu & Whitford, 2011; Santibañez & Fagioli, 2016), study habits and strategies (Risso et al., 2010; Santos, Godás & Lorenzo, 2013) or family support (King et al., 2005; Santos et al., 2013).

School factors are also explored in these kinds of studies, although there is less consensus about the significant effect of these variables (Choi & Calero, 2012; Martínez-Abad & Chaparro-Caso, 2017). However, there is data that supports the influence of some input factors such as the average socioeconomic and cultural level of the school (Perry & McConney, 2010a, 2010b), the school size, or the teacher to student ratio (Nath, 2012), as well as some process factors like student grouping according to their academic performance (Kunz, 2014; Meunier, 2011), teaching methodology (Nath, 2012; Payandeh-Najafabadi, Omid-Najafabadi, & Farid-Rohani, 2013) or learning environment (Payandeh-Najafabadi et al., 2013; Santos et al., 2013).

The introduction of internal student factors (personal, family, cultural, financial and social) and external school factors in an academic performance model should be guided by a systemic quality model that relates these complex dimensions, so that they can be tested with the empirical data gathered (De la Orden & Jornet, 2012; Jornet, et al., 2012).

Statistical techniques for school effectiveness research

The complexity of the study of the factors related to academic performance and school effectiveness, which is caused by their large number and the network of relationships established among them (Tejedor, 2003), has resulted in the use of a broad variety of statistical techniques to address this research subject.

One of these techniques is multilevel analysis, which is used during the course of this study. Its characteristics allow researchers to

differentiate the variability contributed by each of the aggregation levels encountered in the hierarchical data, a distinction without which we could incur the over-estimation of coefficients or data interpretation errors (Snijders & Bosker, 2012). For this reason, its use is recommended for cases where the data presents a nested structure, such as the data from large-scale assessments, where students are gathered in higher level structures (Lenkeit, 2013; Lizasoain & Angulo, 2014; Martínez-Arias, 2009; Murillo & Hernández, 2011). Another example of quantitative techniques is structural equation modelling, which facilitates the establishment of relationships between predicting and criterion variables. It also enables the introduction of other variables of a latent nature, or dimensions (factorial confirmation analysis), which are constructs that cannot be directly measured, but which can be studied through the analysis of other observable variables (Castro & Lizasoain, 2012). Data mining techniques are also suitable for the analysis of data from large-scale assessments, given that they allow the researcher to extract relevant information, such as patterns or significant relationships among variables, from datasets with high amounts of information (Castro & Lizasoain, 2012).

Taking this framework into consideration, the *aim* of this research is twofold. On the one hand, we intend to study the effect of certain input variables (contextual) on student performance, thus analysing the variability of the schools according to their effectiveness, which is measured through the difference between their actual score and the score they would be expected to get according to said variables. On the other hand, we will analyse the process variables (non-contextual) that present a higher discrimination power over the residual of the schools.

For each of the aims, the analysis method will differ. For the first aim, *multilevel models* will be used to study the effect of contextual variables on student performance, and to select the schools whose actual score is significantly above (or below) their expected score according to their demographic and socioeconomic characteristics. These schools, referred to as high and low residual or effectiveness, serve as a basis for the second part of the study, in which *logistic regression* techniques are used to figure out which non-contextual variables most influence the level of the schools' residual, revealing which factors have a significant relationship with school effectiveness.

Method

This secondary analysis of PISA 2015 data is of a non-experimental ex-post-facto nature, due to the lack of experimental control over the variable collection. This section offers information of the participating sample and the data-collecting instruments, as well as the data analysis techniques employed.

Sample

The sample for this study was extracted from the dataset provided by the OECD (2017), and it is composed of all 15-year-old students (born between January and December 1999) who participated in the 2015 PISA test in Spain. Although the initial sample was composed of 32,330 students and 976 schools, the students from schools with less than 20 participating students, along with their schools, were removed from the sample in order to ensure the correct analysis of the variables aggregated at school level, as has been done in other similar studies (Joaristi, Lizasoain & Azpillaga, 2014; Martínez-Abad, Lizasoain, Castro & Joaristi, 2017; Meunier, 2011). This study had a final sample of 31,273 students, where 49.4% (15,437) were female and 50.6% (15,836) were male. These students were enrolled in 897 schools. In this edition of the test, all Autonomous Communities decided to broaden their sample so that their data could be compared at an international level (Ministerio de Educación, Cultura y Deporte, 2016).

The distribution of students according to their Autonomous Community is even (between 4.5% and 6% of the total population), except for the Basque Country, which represents 10.7% of the participants. The schools show a similar distribution.

Out of the 897 schools participating, 66.6% are public schools, 28.2% are publicly-funded private schools, and 5.2% are private schools (excluding the 49 missing values of this variable).

Instruments

In order to conduct this research, the specific instruments created for the PISA test were used. There are two main kinds of instruments. On the

one hand, the competence assessment tests are used to measure the competence level of the students. In 2015, these tests measured student performance in reading, mathematics and science. They are composed of sets of different items that can have three types of answers: open (e.g. explaining the necessary steps to solve a problem), closed (numerical or one-word answers) or multiple choice.

On the other hand, the study also uses information from the context questionnaires administered to students, parents and schools. These questionnaires provide a large amount of information on socioeconomic, cultural and demographic questions, and they also report on other topics of educational interest, such as school climate, student motivation, teacher training, or school assessment practices.

Variables

The criterion variables used to construct the multilevel models are the scores obtained by the students in the competence tests (reading, mathematics, science). These variables are defined by the steering documents of PISA 2015 (OECD, 2016) as follows:

- Reading: “An individual’s capacity to understand, use, reflect on and engage with written texts, in order to achieve one’s goals, to develop one’s knowledge and potential, and to participate in society” (p. 13).
- Mathematics: “An individual’s capacity to formulate, employ and interpret mathematics in a variety of contexts. It includes reasoning mathematically and using mathematical concepts, procedures, facts and tools to describe, explain and predict phenomena” (p. 13).
- Science: “A scientifically literate person is willing to engage in reasoned discourse about science and technology, which requires the competencies to explain phenomena scientifically, evaluate and design scientific enquiry, and interpret data and evidence scientifically” (p. 13).

This analysis includes the 10 plausible values provided for each student and each competence. These values are obtained through an imputation method in order to estimate the performance level of a student from the scores obtained in the items. The sampling weights of both students and schools were also included in the analysis to ensure the correct treatment of the data and the proper calculation of the sampling error (OECD, 2012).

For their part, the context factors extracted from the context questionnaires were used as predicting variables for the models. The selection of these variables was based on the literature review conducted for the theoretical framework, and they are divided in two groups: level 1 (students) and level 2 (schools) (Table I). Where nominal or ordinal data was encountered, dummy variables were generated (in a number equal to the number of categories of the original variable minus one, with the most frequent category being the reference).

TABLE I. Predicting variables for the multilevel model.

	Variable	Label	Range
Level 1 - Students	Gender	NIGEN	0: Male 1: Female
	Birth month	NIBMONTH	1 (Jan) – 12 (Dec)
	Grade	NIGRADE	7 th – 11 th
	Economic, social and cultural status (ESCS)	NIESCS	Continuous
	Migratory status	NIIMMIG	0: Native 1: 2nd generation immigrant 2: 1st generation immigrant
	Grade repetition	NIREPEAT	0: No 1: Yes
	Number of school changes	NISCCH	0: No change 1: One change 2: Two or more changes
	Language spoken at home	NIIDIOMA	0: Language of the test 1: Other language
	Level 2 - Schools	School size	N2TAMESC
Class size		N2TAMCLS	Continuous
Resource shortage		N2ESCPRES	Continuous
Staff shortage		N2ESCPER	Continuous
School ownership		N2TITESC	1: Private 2: Publicly-funded private 3: Public
Teacher-student ratio		N2RATIO	Continuous
Average ESCS		N2ESCS	Continuous
Rate of repeaters		N2REPETI	Continuous
Rate of immigrant students		N2IMMIG	Continuous
Percentage of girls		N2PCGIRL	Continuous

Source: Elaborated by the authors.

On the other hand, the non-contextual variables were used in a later stage of the study (see section on procedure). The variables were selected on the basis of whether they were provided in the form of indices (i.e. those grouping information from several items in one factor) both at student and school levels. The level-1 variables analysed were motivation, disposition for collaborative work, epistemological beliefs, enjoyment and interest in science, competence, interest and autonomy in the use of new technologies, climate of discipline in the classroom, teacher support, environmental awareness and optimism, expected occupational status, sense of belonging to the school, emotional and academic parental support, perceived feedback and teacher fairness. The level-2 variables employed were leadership, curriculum development, professional development, school responsibility over resources and curriculum, teacher participation, school autonomy and creative extra-curricular activities.

Procedure

To obtain the high and low effectiveness schools we used *hierarchical linear models* (Snijders & Bosker, 2012), which enabled the identification of the effect of contextual factors on the schools' average performance. A model was defined for each competence assessed, keeping only the significant predicting variables ($\alpha=.05$) at both levels, which allowed us to calculate the difference between the actual school score and the expected score according to the socioeconomic and cultural variables, also referred to as school "residual", obtained through empirical Bayes estimators (Raudenbush, Bryk, Cheong, Congdon, & Du Toit, 2011). After a previous collinearity study of the predicting variables, and based on the evidence gathered by Özdemir (2016), who did not appreciate substantial differences between fixed and random slope models built with PISA data, a decision was made to introduce the level-1 contextual variables as fixed effects covariables, without including the random effects.

Afterwards, a protocol was set up with a series of criteria that the schools must meet to be selected as high or low residual schools. The protocol was based in two main research works. The first one is the work of Joaristi et al. (2014), where the 80th percentile was chosen as the cut-off point for the 6 estimated models. The second work is the research

conducted by Martínez-Abad et al. (2017), where the lower limit was set at the 66th percentile (high effectiveness) and the upper limit (low effectiveness) was the 33rd percentile, and a school had to comply with the inclusion criterion in 5 of the 8 models estimated in order to be selected. Given the data at our disposal, we opted for an eclectic criterion located somewhere between the abovementioned two. The 33rd and 66th percentiles are taken as limits for the selection, but a school must be included within the limit regions in all three of the estimated models.

Then, we proceeded with the calculation of the correlation between the main process variables (levels 1 and 2) and the dichotomous variable generated (criterion variable: high or low residual school) through a point-biserial correlation. After we ruled out any relevant collinearity effects among the predicting variables of the model, if significant correlations ($\alpha=.05$) were found, we applied *logistic regression techniques*. These techniques are particularly recommended for dichotomous criterion variables.

The construction of the hierarchical linear models was carried out with the statistical software HLM7², and the logistic regression was conducted with SPSS v.20³.

Results

The first step was to calculate the null model, which is an unconditional model without any predicting variables (Hayes, 2006; Lee, 2000). The estimation of the variance components of this model allows for the calculation of the Intraclass Correlation Coefficient (ICC), which represents the proportion of variance attributable to the second level (Snijders & Bosker, 2012). In order for the use of multilevel models to be deemed a suitable alternative, the value of the ICC must be over 10% (Lee, 2000). The ICC is defined by the following equation (1):

$$CCI = \frac{\tau_{00}}{\tau_{00} + \sigma^2} \quad (1)$$

²) Commercial licence of the Research Group of Educational Assessment and Guidance, University of Salamanca.

³) Commercial licence of the University of Salamanca.

The term τ_{00} refers to the variability among schools, and σ^2 represents the variability among students.

The data from the Spanish sample of PISA 2015 presents an ICC of 12.26% in mathematics, 12,04% in reading, and 12,41% in science, thus making it pertinent to use a procedure based on multilevel analysis.

Performance in mathematics

After applying the procedure described in the previous section, the multilevel model for the performance of the students in the mathematical competence tests is defined by the following equation (2):

$$\begin{aligned}
 PV1MATH_{ij} = & \gamma_{00} + \gamma_{01} * N2ESCS_j + \gamma_{10} * N1GEN_{ij} + \gamma_{20} * N1ESCS_{ij} + \\
 & \gamma_{30} * N1REPEAT_{ij} + \gamma_{40} * N1GRADE_{ij} + \gamma_{50} * N1IMM2_{ij} + \gamma_{60} * N1SCCH1_{ij} + \\
 & \gamma_{70} * N1SCCH2_{ij} + u_{0j} + r_{ij} \quad (2)
 \end{aligned}$$

Table II shows the coefficients, t-values and signification values of the variables that were eventually included in the mathematical competence model.

TABLE II. Final estimation of fixed effects with robust standard errors, mathematics.

Fixed Effect	Coefficient	Standard error	t-ratio	p-value
For INTRCPT1, β_0				
INTRCPT2, γ_{00}	543.097	1.731	313.758	<.001
SCHOOL ESCS, γ_{01}	12.986	1.906	6.813	<.001
For GENDER slope, β_1				
INTRCPT2, γ_{10}	-22.599	1.972	-11.457	<.001
For ESCS slope, β_2				
INTRCPT2, γ_{20}	9.799	1.123	8.723	<.001
For REPEAT GRADE slope, β_3				
INTRCPT2, γ_{30}	-35.721	4.972	-7.184	<.001
For GRADE slope, β_4				
INTRCPT2, γ_{40}	34.960	3.874	9.024	<.001
For 1st GEN IMMIGRANT slope, β_5				
INTRCPT2, γ_{50}	-11.785	4.863	-2.423	.026
For SCHOOL CHANGES (1) slope, β_6				
INTRCPT2, γ_{60}	-10.690	1.870	-5.716	<.001
For SCHOOL CHANGES (2+) slope, β_7				
INTRCPT2, γ_{70}	-15.512	3.303	-4.696	<.001

Source: Elaborated by the authors.

While the only school factor included is the average ESCS, at student level there are several factors with a significant influence. Gender is the variable with the highest t-value, followed by grade, ESCS and repeating a grade. The variable related to migratory status indicates that only being a 1st generation immigrant is significant.

After estimating the model, the ICC for the mathematical competence is 4.55%, which enables us to confirm that the variables included in the model explain 7.71% of the variance among schools.

Performance in reading

In the case of reading comprehension skills, the variables that finally composed the model are found in the following equation (3):

$$PV1READ_{ij} = \gamma_{00} + \gamma_{01} * N2ESCS_j + \gamma_{02} * N2REPETI_j + \gamma_{03} * N2PCGIRL_j + \gamma_{10} * N1GEN_{ij} + \gamma_{20} * N1ESCS_{ij} + \gamma_{30} * N1REPEAT_{ij} + \gamma_{40} * N1GRADE_{ij} + \gamma_{50} * N1SCCH1_{ij} + \gamma_{60} * N1SCCH2_{ij} + \gamma_{70} * N1IDIOMA_{ij} + u_{0j} + r_{ij} \quad (3)$$

Table III shows the magnitude and signification values of the relationship between the variables that compose the model and the performance of the students in the reading comprehension test.

TABLE III. Final estimation of fixed effects with robust standard errors, reading.

Fixed Effect	Coefficient	Standard error	t-ratio	p-value
For INTRCPT1, β_0				
INTRCPT2, γ_{00}	522.475	6.153	84.912	<.001
SCHOOL ESCS, γ_{01}	20.293	2.658	7.633	<.001
REPEATER RATE, γ_{02}	36.502	10.829	3.371	<.001
GIRL RATE, γ_{03}	24.719	10.572	2.338	.020
For GENDER slope, β_1				
INTRCPT2, γ_{10}	6.776	1.849	3.665	<.001
For ESCS slope, β_2				
INTRCPT2, γ_{20}	8.299	0.949	8.748	<.001
For REPEAT GRADE slope, β_3				
INTRCPT2, γ_{30}	-32.120	5.162	-6.222	<.001
For GRADE slope, β_4				
INTRCPT2, γ_{40}	41.092	3.495	11.757	<.001
For SCHOOL CHANGES (1) slope, β_5				
INTRCPT2, γ_{50}	-8.857	2.127	-4.164	<.001
For SCHOOL CHANGES (2+) slope, β_6				
INTRCPT2, γ_{60}	-20.642	3.224	-6.403	<.001
For LANGUAGE slope, β_7				
INTRCPT2, γ_{70}	-6.718	3.026	-2.220	.033

Source: Elaborated by the authors.

At school level, this model presents more predicting variables than the previous one, with the percentage of repeating students and the percentage of girls joining the average ESCS of the school. At level 1, the variables with a highest t-value are grade, ESCS, grade repetition, and school changes (two or more times).

The ICC of the scores obtained in reading after the application of the model is 5.07%, which means that the model was able to explain roughly 7% of the variance of the second level in this competence.

Performance in science

Lastly, the factors that resulted in a significant relationship with the scientific competence are reflected in the following equation (4):

$$\begin{aligned}
 PV1SCIE_{ij} = & \gamma_{00} + \gamma_{01} * N2TAMESC_j + \gamma_{02} * N2ESCPER_j + \gamma_{03} * N2ESCS_j + \\
 & \gamma_{04} * N2REPETI_j + \gamma_{05} * N2PCGIRL_j + \gamma_{10} * N1GEN_{ij} + \gamma_{20} * N1ESCS_{ij} + \\
 & \gamma_{30} * N1BMONTH_{ij} + \gamma_{40} * N1REPEAT_{ij} + \gamma_{50} * N1GRADE_{ij} + \gamma_{60} * N1IMM2_{ij} + \\
 & \gamma_{70} * N1SCCH1_{ij} + \gamma_{80} * N1SCCH2_{ij} + u_{0j} + r_{ij} \quad (4)
 \end{aligned}$$

Table IV shows the coefficients derived from the implementation of the model.

The resulting model contains several significant school-level variables, with the most relevant being the average ESCS, followed by the rate of students who have repeated a grade. Other variables included in the model are the percentage of girls in the school, school size and the shortage of teaching staff.

At student level, the variables with the highest t-values are ESCS, grade, gender and school changes. Again, the variable related to migratory status indicates significant differences only for 1st generation immigrant students.

Once the model was applied, the calculation of the ICC of the final model (5.6%) revealed that the model for this competence has managed to explain 6.8% of the level-2 variance.

TABLA IV. Final estimation of fixed effects with robust standard errors, science.

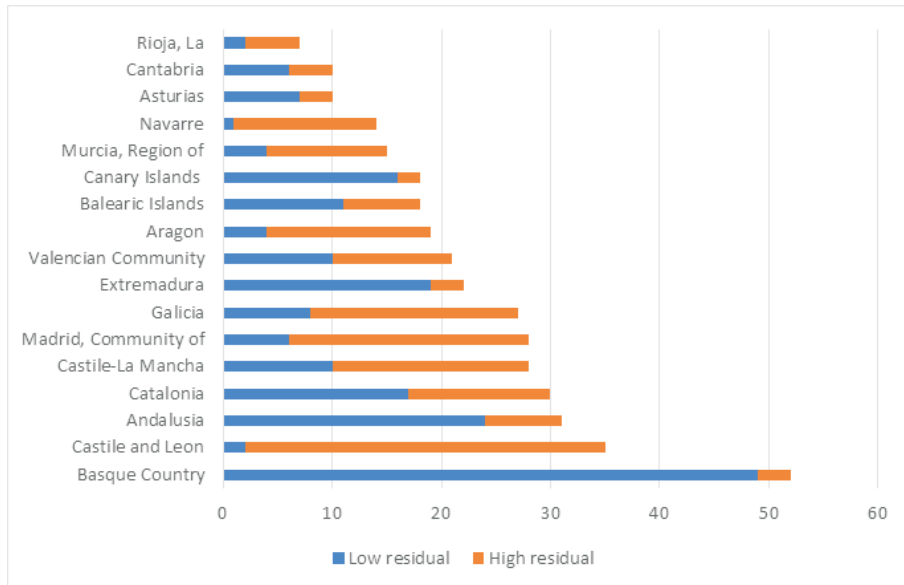
Fixed Effect	Coefficient	Standard error	t-ratio	p-value
For INTRCPT1, β_0				
INTRCPT2, γ_{00}	537.733	7.298	73.683	<.001
SCHOOL SIZE, γ_{01}	-0.006	0.003	-2.226	.026
STAFF SHORTAGE, γ_{02}	2.513	1.163	2.161	.031
SCHOOL ESCS, γ_{03}	20.708	2.605	7.949	<.001
REPEATER RATE, γ_{04}	31.815	9.545	3.333	<.001
GIRL RATE, γ_{05}	31.854	11.979	2.659	.008
For GENDER slope, β_1				
INTRCPT2, γ_{10}	-19.782	1.686	-11.730	<.001
For ESCS slope, β_2				
INTRCPT2, γ_{20}	9.792	0.704	13.912	<.001
For BIRTH MONTH slope, β_3				
INTRCPT2, γ_{30}	-0.590	0.228	-2.586	.011
For REPEAT GRADE slope, β_4				
INTRCPT2, γ_{40}	-34.223	3.874	-8.833	<.001
For GRADE slope, β_5				
INTRCPT2, γ_{50}	38.389	2.898	13.248	<.001
For 1st GEN IMMIGRANT slope, β_6				
INTRCPT2, γ_{60}	-9.596	2.886	-3.325	<.001
For SCHOOL CHANGES (1) slope, β_7				
INTRCPT2, γ_{70}	-10.417	1.945	-5.356	<.001
For SCHOOL CHANGES (2+) slope, β_8				
INTRCPT2, γ_{80}	-16.847	2.631	-6.404	<.001

Source: Elaborated by the authors.

School selection

In order to select high and low residual schools, we implemented the previously explained selection procedure with the variables obtained as a result of the difference between the observed score of the schools and the score estimated by the final model of each competence. With this criterion, 196 low-residual schools and 189 high-residual schools were selected. Figure I shows the distribution of these schools by Autonomous Community.

FIGURE I. Distribution of high and low residual schools by Autonomous Community.



Source: Elaborated by the authors.

Regarding the ownership of the selected schools, most of them belonged to the public network (roughly 67%), while private and publicly-funded private schools represented about a third of the sample, with private schools being the smallest group (table V). The goodness-of-fit test show the correspondence between the initial sample and the selection ($\chi^2=0.016$; $p=.992$).

TABLE V. Fit of the school selection to the initial sample, by school ownership.

		Ownership			Total
		Private	Publicly-funded private	Public	
School selection	Count	19	102	245	366
	%	5,2%	27,9%	66,9%	100,0%
Initial sample	Count	44	239	564	847
	%	5,2%	28,2%	66,6%	100,0%

Source: Elaborated by the authors.

Table VI offers descriptive information on the ESCS, rate of repeaters, rate of immigrant students and expected score for each competence, which are variables that characterise the sample of selected schools. Both the indicators of the context variables and the expected scores are fairly similar, with no significant differences being found between the means of both groups for any of the variables (t-test, $\alpha=.05$), except for the one referred to the rate of immigrants, which was significantly higher in the group of high-residual schools ($p=.002$).

TABLE VI. Mean, standard deviation, maximum value, minimum value.

	Low residual				High residual			
	Avg.	S.d.	Min.	Max.	Avg.	S.d.	Min.	Max.
Average ESCS	-0,42	0,66	-1,92	0,97	-0,53	0,56	-1,55	1,14
Repeater rate	0,27	0,16	0,00	0,67	0,30	0,14	0,00	0,79
Immigrant rate	0,09	0,13	0,00	0,69	0,13	0,11	0,00	0,58
Exp. Sc. math	537,60	8,77	517,75	556,01	536,18	7,37	522,67	558,19
Exp. Sc. reading	537,39	10,24	513,54	563,23	536,26	8,34	517,89	559,68
Exp. Sc. science	547,87	10,63	518,18	576,41	546,79	8,72	525,10	572,22

Source: Elaborated by the authors.

Effect of the process variables on school effectiveness (logistic regression)

The initial correlational study revealed that none of the process variables extracted from the context questionnaires administered to the leadership teams of the schools (level 2) presented a significant correlation with the school selection variable. For this reason, we did not proceed with the logistic regression analysis at school level.

In order to study the process variables at student level (level 1), the average scores of these variables were aggregated to the school database. In the correlational study, a good part of the variables presented significant correlations with the residual-based school selection variable. Thus, the logistic regression model, obtained after eliminating one by one the variables that were not significant, was composed of the variables reflected in table VII and equation 5.

TABLE VII. Logistic regression.

Variable	Content of the variable	B	Sig.	Odds Ratio
DISCLISCI	Discipline climate (science)	1,435	,001	4,199
TEACHSUP	Teacher support (science)	-1,868	,014	,154
ENVAWARE	Environmental awareness	1,443	,018	4,235
SCIEEFF	Self-efficacy (science)	1,781	,001	5,939
EPIST	Epistemological beliefs	2,017	,007	7,514
BSMJ	Expected occupational status	-,091	,008	,913
BELONG	Belonging to the school	-1,628	,003	,196
COOPERATE	Enjoyment of cooperation	1,628	,029	5,095
EMOSUPS	Emotional parental support	-3,606	,000	,027
PERFEED	Perceived feedback	-1,652	,006	,192
ADINST	Adaptation of instruction	2,466	,004	11,770
USESCH	Use of ICT in the school	-1,330	,001	,264
AUTICT	Autonomy in the use of ICT	2,768	,001	15,920
SOIAICT	ICT as a topic of social interaction	-2,569	,005	,077
unfairteacher	Teacher unfairness	-,882	,000	,414
Constant	Constant	14,223	,000	1502425,583

Source: Elaborated by the authors.

$$\hat{Y} = 14.223 + 1.435 * DISCCLISCI - 1.868 * TEACHSUP + 1.443 * ENVAWARE + 1.781 * SCIEEFF + 2.017 * EPIST - 0.091 * BSMJ - 1.628 * BELONG + 1.628 * COOPERATE - 3.606 * EMOSUPS - 1.652 * PERFEED + 2.466 * ADINST - 2.569 * SOIAICT - 0.882 * unfairteacher \quad (5)$$

Among the variables included in the model, the most significant ones with a positive effect in school effectiveness, i.e. those variables whose increment produces a higher probability of the school to belong to the high-residual group, were climate discipline, self-efficacy, autonomy in the use of ICT and the adaptation of instruction. The variables that present a highly significant negative effect were the level of teacher unfairness, parental emotional support, ICT use in school, sense of belonging to the school, and perceived feedback.

It is worth noting that the model achieves a good fit, obtaining a $R^2=.527$ (Nagelkerke index). On the other hand, the accuracy of the predicting model reaches an 80.52% of correct classifications of high and low residual schools, correctly predicting 79.59% for low-residual schools and 81.48% in high residual schools.

Discussion and conclusions

This research had two main aims. The first one resulted in the construction of three models that allowed us to detect the size and significance of the relationship between input variables, both from students and schools, and student performance.

The level-1 variables included in the models match the results of previous studies, with gender being one of the most influencing factors (Karakolidis et al., 2016; Özdemir, 2016; Stoet & Geary, 2014). In mathematics and science, male students outperform females, while in reading the opposite is true, although to a lesser extent. Other relevant variables are the socioeconomic level (Risso et al., 2010), the fact of repeating a grade (Choi & Calero, 2013; Ehmke et al.2008), or the number of school changes in the academic history of a student, a variable which is largely unexplored. Moreover, two of the models (mathematics and science) suggest that migratory status is a relevant variable, but only for 1st generation immigrants, since no differences were found between 2nd

generation immigrants and native students. This discrepancy might be due to broader contextual factors, such as the decline in immigration from Spanish-speaking countries and the rise in immigration from other countries (Riederer & Verwiebe, 2015). Among the level-2 variables introduced in the models, the only one that was significant in all of them was the average ESCS of the school, a factor whose relevance has been underlined by other studies (Perry & McConney, 2010a, 2010b). Furthermore, we found that the impact of the input variables presents some differences depending on the performance variable used to make the model. For example, gender has a different influence in reading and mathematics (Stoet & Geary, 2013), and migratory status (Meunier, 2011) or the school's socioeconomic level (Perry & McConney, 2010a) do not influence the three competences to the same extent, so the differences found in this respect have a precedent in the scientific literature.

The distribution of the schools selected with the help of these models presents a good fit with the original sample in almost all the characterisation variables controlled. However, the distribution according to their Autonomous Community is not entirely adjusted to the sample; some Autonomous Communities are underrepresented (Asturias, Cantabria, La Rioja), while others have a higher percentage of representation in the selected sample than in the PISA sample (Andalusia, Castile and Leon, Basque Country). Some of these Communities present a noticeable unbalance between the proportion of high and low residual schools, which suggests the pertinence of an in-depth study of these particular cases to explore which regional factors might be causing these inequalities.

The second aim was to analyse the non-contextual variables according to the previous selection of schools. This analysis revealed that there were no statistically significant differences in the scores of the process variables obtained by high and low residual schools. This conclusion differs from some qualitative studies which concluded that there are many process variables related to school effectiveness in a relevant way, such as the research by Lizasoain and Angulo (2014), conducted with highly effective schools, or the work of Murillo (2007) with high, medium and low residual schools. The causes for this relevant discrepancy may vary in nature. De la Orden and Jornet (2012) highlight the shortcomings presented by the PISA questionnaires to properly measure the contextual factors, which could hinder a correct analysis of the educational reality and lead to wrong conclusions. This issue might point to the necessity to

search for alternative sources for contextual data which could give the research a higher degree of internal validity.

The level-1 variables that resulted significant in the logistical regression model indicate some transversal issues, such as the disciplinary climate in the classroom, student self-efficacy, the adaptation of instruction or the autonomy in the use of ICT, which would be beneficial to promote within the schools, since the analyses indicate that the perception of the students about these questions is closely related to a positive residual in the measurement of school effectiveness. Although this aggregated analysis has the elimination of inter-school variance as a disadvantage, it allows us to compare the schools at a global level according to the characteristics of their students.

The key strengths of this research are the employment of all the statistical guarantees recommended for the application of multilevel models (consideration of the variance at both levels of analysis, use of plausible values, and introduction of sampling weights), the selection of high and low residual schools through the scores in the three measured competences, and the high explaining percentage of the final models, given that all three of them explained more than 50% of the ICC of the null model. Furthermore, the logistic regression presents very high goodness-of-fit values.

On the other hand, the research also presents some shortcomings, such as the lack of systematicity in the selection of the variables, the occurrence of variables with an opposite effect to that expected in the logistic regression, or the lack of sufficient evidence of the validity and reliability of the context questionnaires, which were one of the main sources of information.

The results obtained in this study suggest the need to encourage in-depth research about the factors related to school effectiveness in our country. This study should always be based on the comparison between high and low residual schools so as to make it possible to eliminate those factors that occur in both types of school, thus making them irrelevant. Therefore, we propose the establishment of new lines of research which, starting from the school selection presented in this study, make use of alternative sources of non-contextual data, such as qualitative research techniques, in order to determine the school factors that are relevant for the study of school effectiveness, thus being able to draw relevant conclusions for educational policies and practices.

References

- Aitkin, M., & Longford, N. (1986). Statistical modelling issues in school effectiveness studies. *Journal of the Royal Statistical Society. Series A (General)*, 149(1), 1-43.
- Aksu, G., & Güzeller, C. O. (2016). Classification of PISA 2012 Mathematical Literacy Scores Using Decision-Tree Method: Turkey Sampling. *Egitim ve Bilim*, 41(185), 101-122.
- Blanco-Blanco, Á., López Martín, E., & Ruiz de Miguel, C. (2014). Aportaciones de los modelos jerárquico-lineales multivariados a la investigación educativa sobre el rendimiento. Un ejemplo con datos del alumnado español en PISA 2009. *Revista de Educación*, 365, 122-149.
- Brookover, W.B., Beady, C., Flood, P., Schewitzer, J. & Wisenbaker, J. (1979). *School social systems and student achievement: schools can make a difference*. New York: Praeger.
- Castro, M., & Lizasoain, L. (2012). Las técnicas de modelización estadística en la investigación educativa: Minería de datos, modelos de ecuaciones estructurales y modelos jerárquicos lineales. *Revista Española de Pedagogía*, 70, 131-148.
- Choi, A., & Calero, J. (2013). Determinantes del riesgo de fracaso escolar en España en PISA-2009 y propuestas de reforma. *Revista De Educación*, 362, 562-593.
- Coleman, J. S. (1966). *Equality of educational opportunity* (Vol. 2). Washington, DC: US Department of Health, Education, and Welfare, Office of Education.
- Cordero Ferrera, J. M., Manchón López, C. & Simancas Rodríguez, R. (2014). La repetición de curso y sus factores condicionantes en España. Repetition and explanatory factors in Spain. *Revista de Educacion*, 365, 12-37.
- Creswell, J. W. (2015). Revisiting mixed methods and advancing scientific practices. In Hesse-Biber, S. N., & Johnson, R. B. (Eds.) *The Oxford handbook of multimethod and mixed methods research inquiry*. (pp. 57-71). Oxford: Oxford University Press.
- De la Orden, A. & Jornet, J. M. (2012). La utilidad de las evaluaciones de sistemas educativos: el valor de la consideración del contexto. *Bordón*, 64(2), 69-88.

- Demir, İ., Kılıç, S., & Ünal, H. (2010). Effects of students' and schools' characteristics on mathematics achievement: Findings from PISA 2006. *Procedia-Social and Behavioral Sciences*, 2(2), 3099-3103.
- Ehmke, T., Drechsel, B., & Carstensen, C. H. (2008). Klassenwiederholen in PISA-I-plus: Was lernen sitzenbleiber in mathematik dazu? *Zeitschrift Für Erziehungswissenschaft*, 11(3), 368-387.
- Gamazo, A., Olmos-Migueláñez, S., & Martínez-Abad, F. (2016, November). Multilevel models for the assessment of school effectiveness using PISA scores. In *Proceedings of the Fourth International Conference on Technological Ecosystems for Enhancing Multiculturality* (pp. 1161-1166). ACM.
- Hayes, A. F. (2006). A primer on multilevel modeling. *Human Communication Research*, 32(4), 385-410.
- Joaristi, L., Lizasoain, L., & Azpillaga, V. (2014). Detección y caracterización de los centros escolares de alta eficacia de la Comunidad Autónoma del País Vasco mediante Modelos Transversales Contextualizados y Modelos Jerárquicos. *Estudios sobre Educación*, 27, 37-61.
- Jornet, J. M., González-Such, J., & Perales, M. J. (2012). Diseño de cuestionarios de contexto para la evaluación de sistemas educativos: optimización de la medida de constructos complejos. *Bordón*, 64(2), 89-110.
- Karakolidis, A., Pitsia, V., & Emvalotis, A. (2016). Examining students' achievement in mathematics: A multilevel analysis of the Programme for International Student Assessment (PISA) 2012 data for Greece. *International Journal of Educational Research*, 79, 106-115.
- King, G., McDougall, J., DeWit, D., Hong, S., Miller, L., Offord, D., . . . LaPorta, J. (2005). Pathways to children's academic performance and prosocial behaviour: Roles of physical health status, environmental, family, and child factors. *International Journal of Disability, Development and Education*, 52, 313-344.
- Kunz, J. S. (2014). Analyzing Educational Achievement Differences between Second Generation Immigrants: Comparing Germany and German Speaking Switzerland. *German Economic Review*, 17(1), 61-91.
- Lee, V. E. (2000). Using hierarchical linear modeling to study social contexts: The case of school effects. *Educational Psychologist*, 35(2), 125-141.

- Lenkeit, J. (2013). Effectiveness measures for cross-sectional studies: A comparison of value-added models and contextualised attainment models. *School Effectiveness and School Improvement*, 24(1), 1-25.
- Liu, X., & Whitford, M. (2011). Opportunities-to-learn at home: Profiles of students with and without reaching science proficiency. *Journal of Science Education and Technology*, 20(4), 375-387.
- Lizasoain, L., & Angulo, A. (2014). Buenas prácticas de escuelas eficaces del País Vasco. Metodología y primeros resultados. *Participación Educativa*, 3(4), 17-27.
- Martínez-Abad, F., & Chaparro-Caso, A. A. (2017). Data-mining techniques in detecting factors linked to academic achievement. *School Effectiveness and School Improvement*, 28(1), 39-55.
- Martínez-Abad, F., Lizasoain, L., Castro, M., & Joaristi, L. (2017). Selección de escuelas de alta y baja eficacia en Baja California (México). *REDIE. Revista Electrónica de Investigación Educativa*, 19(2), 38-53.
- Martínez-Arias, M. R. (2009). Usos, aplicaciones y problemas de los modelos de valor añadido en educación. *Revista De Educación*, 348, 217-252.
- Meunier, M. (2011). Immigration and student achievement: Evidence from Switzerland. *Economics of education review*, 30(1), 16-38.
- Ministerio de Educación, Cultura y Deporte (2016). *PISA 2015. Programa para la Evaluación Internacional de los Alumnos. Informe Español*. Madrid: Secretaría General Técnica.
- Murillo, F. J. (2003). El movimiento de investigación de Eficacia Escolar. In F.J. Murillo (Coord.), *La investigación sobre Eficacia Escolar en Iberoamérica. Revisión internacional del estado del arte*. Bogotá: Convenio Andrés Bello.
- Murillo, F. J. (2005). *La Investigación sobre Eficacia Escolar*. Barcelona: Octaedro.
- Murillo, F. J. (Coord.) (2007). *Investigación Iberoamericana sobre Eficacia Escolar*. Bogotá: Convenio Andrés Bello.
- Murillo, F. J., & Hernández, R. (2011). Efectos escolares de factores socio-afectivos. Un estudio multinivel para Iberoamérica. *Revista De Investigación Educativa*, 29(2), 407-427.
- Nath, S. R. (2012). Factors influencing primary students' learning achievement in Bangladesh. *Research in Education*, 88, 50-63.
- Organisation for Economic Co-operation and Development (OECD) (2008). *Measuring improvements in learning outcomes: best practices to assess the value-added of schools*. Paris: OECD Publishing.

- Organisation for Economic Co-operation and Development (OECD) (2014). *PISA 2012 Technical Report*. Paris: OECD Publishing
- Organisation for Economic Co-operation and Development (OECD) (2016). *PISA 2015 Assessment and Analytical Framework: Science, Reading, Mathematics and Financial Literacy*. Paris: OECD Publishing.
- Organisation for Economic Co-operation and Development (OECD) (2017), PISA: *Programme for International Student Assessment, OECD Education Statistics (database)*. DOI: <http://dx.doi.org/10.1787/data-00365-en>
- Özdemir, C. (2016). Equity in the Turkish education system: A multilevel analysis of social background influences on the mathematics performance of 15-year-old students. *European Educational Research Journal*, 15(2), 193-217.
- Payandeh-Najafabadi, A. T., Omidi-Najafabadi, M., & Farid-Rohani, M. R. (2013). Factors contributing to academic achievement: A Bayesian structure equation modelling study. *International Journal of Mathematical Education in Science and Technology*, 44(4), 490-500.
- Perry, L., & McConney, A. (2010a). Does the SES of the school matter? an examination of socioeconomic status and student achievement using PISA 2003. *Teachers College Record*, 112(4), 1137-1162.
- Perry, L., & McConney, A. (2010b). School socio-economic composition and student outcomes in Australia: Implications for educational policy. *Australian Journal of Education*, 54(1), 72-85.
- Raudenbush, S. W., Bryk, A. S., Cheong, Y. F., Congdon, R., & Du Toit, M. (2011). *Hierarchical linear and nonlinear modeling (HLM7)*. Lincolnwood, IL: Scientific Software International.
- Riederer, B., & Verwiebe, R. (2015). Changes in the educational achievement of immigrant youth in Western societies: The contextual effects of national (educational) policies. *European Sociological Review*, 31(5), 628-642.
- Risso, A., Peralbo, M., & Barca, A. (2010). Cambios en las variables predictoras del rendimiento escolar en Enseñanza Secundaria. *Psicothema*, 22(4), 790-796.
- Santibañez, L., & Fagioli, L. (2016). Nothing succeeds like success? Equity, student outcomes, and opportunity to learn in high-and middle-income countries. *International Journal of Behavioral Development*, 40(6), 517-525.

- Santos, M. A., Godás, A., & Lorenzo, M. (2013). Rendimiento académico y diversidad cultural: El eje lingüístico. *Revista Española de Pedagogía*, 256, 461–478.
- Snijders, T., & Bosker, R. J. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modeling* (2nd Edition). London: Sage Publications.
- Stoet, G., & Geary, D. C. (2013). Sex differences in mathematics and reading achievement are inversely related: Within-and across-nation assessment of 10 years of PISA data. *PloS One*, 8(3), e57988.
- Tejedor, F. J. (2003). Poder explicativo de algunos determinantes del rendimiento en los estudios universitarios. *Revista Española de Pedagogía*, 61(224), 5–32.
- Tsai, S. L., Smith, M. L., & Hauser, R. M. (2017). Families, Schools, and Student Achievement Inequality: A Multilevel MIMIC Model Approach. *Sociology of Education*, 90(1), 64-88.
- Weber, G. (1971). *Inner-city children can be taught to read: four successful schools*. Washington, DC: Council for Basic Education.

Contact address: Adriana Gamazo. Universidad de Salamanca, Facultad de Educación, Departamento de Didáctica, Organización y Métodos de Investigación. Paseo de Canalejas 169, 37008, Salamanca E-mail: adrianagamazo@usal.es