



VNiVERSiDAD
D SALAMANCA

Eficacia escolar y pruebas PISA. Factores asociados al rendimiento académico

TESIS DOCTORAL – Programa de Doctorado Formación en la Sociedad del Conocimiento

Autora: Cristina Frade Martínez

Directoras:

Dra. Susana Olmos Migueláñez

Dra. Adriana Gamazo García

Salamanca, febrero 2024





TESIS DOCTORAL

EFICACIA ESCOLAR Y PRUEBAS PISA. FACTORES ASOCIADOS AL RENDIMIENTO ACADÉMICO

Doctoranda:

Cristina Frade Martínez

Directoras:

Susana Olmos Migueláñez

Adriana Gamazo García

Autorización de las directoras para la presentación de la Tesis Doctoral

La **Dra. Susana Olmos Migueláñez**, Profesora Titular en el Área de Métodos de Investigación y Diagnóstico en Educación, y la **Dra. Adriana Gamazo García**, Profesora Permanente Laboral en el Área de Métodos de Investigación y Diagnóstico en Educación, ambas pertenecientes al Departamento de Didáctica, Organización y Métodos de Investigación de la Universidad de Salamanca,

Certifican:

Que **Dña. Cristina Frade Martínez** ha realizado la Tesis Doctoral titulada: “Eficacia escolar y pruebas PISA. Factores asociados al rendimiento académico” bajo su dirección, y que esta cumple con los criterios de calidad y originalidad necesarios para optar a la obtención del Grado en Doctor por la Universidad de Salamanca. Además, certifican que la doctoranda opta a la mención Doctor Internacional.

Y para que conste, se firma por las directoras de la Tesis Doctoral a los efectos oportunos en Salamanca, a 20 de febrero de 2024.

Susana Olmos Migueláñez

Adriana Gamazo García

Quisiera expresar mi sincero agradecimiento a todas las personas que han contribuido de manera significativa ($\alpha \leq 0,05$) a la realización de esta tesis doctoral.

En primer lugar, quiero agradecer a mis directoras de tesis, Susana y Adriana, por su orientación, apoyo y sabiduría a lo largo de todo este proceso. Vuestra paciencia y confianza han sido fundamentales para mi crecimiento como investigadora y para ayudarme a conocer de primera mano la realidad de nuestra carrera académica.

A María José, por confiar en mi desde el primer momento para formar parte de este proyecto de investigación y ayudar inmensamente en mi formación.

Mi agradecimiento se extiende también a mis compañeras y compañeros del Grupo GRIAL y en especial a los miembros de GE2O y de Cedetel.

A Joe y Martin por su apoyo durante mi estancia en Irlanda haciéndome sentir una más de su familia. A Ricardo por su predisposición y ayuda durante el tiempo en Leiria.

Al Ministerio de Ciencia e Innovación, que a través de una ayuda para contratos predoctorales para la formación de doctores me ha permitido dedicarme en exclusividad a esta tarea (*PRE2019-087412*).

A mi familia, quiero expresar mi amor y gratitud. Mamá, gracias por tu apoyo incondicional, tu paciencia y tu comprensión en cada paso de este viaje, sobre todo en estos momentos finales. Vuestros ánimos son fuente de fuerza y motivación.

Quiero agradecer a todos mis amigos y seres queridos que han estado a mi lado a lo largo de este largo camino, interesándose y acompañándome en cada uno de los pasos. Vuestro interés y vuestra compañía han hecho que este camino sea mucho más llevadero.

A ti, Sara, esta tesis no estaría hoy aquí sin tus revisiones, lecturas, repasos y tu infinita paciencia durante todos estos meses.

Por último, agradecer también a todas esas personas que han formado parte de esta travesía y, en especial a aquellas que han llegado en los momentos finales y sin las cuales no habría sido posible conseguir esto, Yai tú entre ellas gracias a tu constancia y fortaleza.

Este logro no habría sido posible sin la contribución de cada uno de vosotros. Estoy profundamente agradecida por formar parte de este increíble viaje académico y personal. Tan solo es el inicio de una nueva etapa.

¡Muchas gracias a todos!

Resumen

Las pruebas de evaluación a gran escala o ILSA (*International Large Scale Assessment*) son actualmente una inmensa fuente de información sobre la realidad educativa a nivel mundial. Numerosas investigaciones se han construido sobre los datos provenientes de alguna de estas pruebas de evaluación, pretendiendo conocer los factores ligados al rendimiento académico y a la eficacia escolar como parte de la corriente de Investigación en Eficacia Escolar (IEE). En el caso que se presenta, la atención se centra en una de estas pruebas: el Programa para la Evaluación Internacional de Estudiantes (PISA).

Esta Tesis Doctoral se alinea con este tipo de investigaciones con el objetivo principal de generar un marco de buenas prácticas educativas relacionadas con la eficacia escolar a nivel internacional, estudiando el caso de tres países europeos: España, Irlanda y Portugal. Para alcanzar este objetivo, se plantean los siguientes objetivos específicos: seleccionar las variables de contexto a nivel de centro educativo y de estudiante que presenten relaciones significativas en el rendimiento de los estudiantes; detectar los centros educativos de alta y baja eficacia escolar en los tres países estudiados; analizar las variables relacionadas con la eficacia escolar para determinar las relaciones existentes entre distintas variables de proceso y esta variable criterio; y conocer las buenas prácticas implementadas por los centros educativos de alta eficacia para tratar de replicarlas en los de baja eficacia. La consecución de estos objetivos hace necesaria la combinación de diferentes metodologías de investigación.

Tomando como base los datos de PISA 2018, se inicia el estudio cuantitativo aplicando la metodología de análisis multinivel para la generación de modelos jerárquico-lineales que expresen la relación existente entre las variables contextuales de dos niveles de análisis (estudiante y centro educativo) y el rendimiento académico en las tres competencias básicas estudiadas en esta prueba a gran escala (matemáticas, ciencias y comprensión lectora). Este análisis se realiza con los datos de los tres países individualmente, obteniendo un total de nueve modelos estadísticos. Ciertas variables como el nivel socioeconómico de los estudiantes, la repetición de curso o el género han resultado influir consistentemente en el rendimiento académico, aunque de diferente manera en función del contexto y de la competencia a la que se refieren.

A partir de estos resultados, se determinaron los centros de alta y baja eficacia para iniciar el estudio de la relación existente entre algunas variables no contextuales y la

eficacia escolar mediante la regresión logística binaria. Algunas de estas variables se relacionan con el autoconcepto de los estudiantes, la autoeficacia de los profesores o el clima escolar. Los factores relacionados con la eficacia escolar más comúnmente encontrados son la capacidad percibida por los estudiantes para resumir textos, su autoconcepto relacionado con la lectura en cuanto a percepción de su competencia y el uso de las TIC fuera de la escuela para la realización de tareas escolares.

Finalmente, se desgranán las principales implicaciones para la política y la práctica educativa de los resultados hallados, y se detallan una serie de limitaciones ligadas a la propia utilización de las pruebas a gran escala para estudiar el rendimiento académico y la eficacia escolar como su carácter transversal, planteando también posibles líneas de investigación como por ejemplo la realización de estudios de tendencias.

Palabras clave

Evaluación a gran escala, PISA, eficacia escolar, rendimiento académico, modelos multinivel, regresión logística binaria, evaluación educativa.

Abstract¹

ILSA (International Large-Scale Assessment) tests are currently an immense source of information on the reality of education worldwide. Numerous investigations have been built on the data from some of these assessment tests, seeking to understand the factors linked to academic performance and school effectiveness as part of the School Effectiveness Research (SER) trend. In this case, the focus is on one of these tests: the Programme for International Student Assessment (PISA).

This Doctoral Thesis is aligned with this type of research with the main objective of generating a framework of good educational practices related to school effectiveness at an international level, studying the case of three European countries: Spain, Ireland and Portugal. In order to achieve this objective, the following specific objectives are proposed: to select the context variables at school and student level that show significant relationships in student performance; to detect schools with high and low school effectiveness in the three countries studied; to analyse the variables related to school effectiveness in order to determine the relationships between different process variables and this criterion variable; and to learn about the good practices implemented by schools with high effectiveness in order to try to replicate them in those with low effectiveness. To achieve these objectives, it is necessary to combine different research methodologies.

Based on the PISA 2018 data, the quantitative study begins by applying the methodology of multilevel analysis to generate hierarchical-linear models that express the relationship between the contextual variables of two levels of analysis (student and school) and academic performance in the three basic skills studied in this large-scale test (mathematics, science and reading comprehension). This analysis is carried out with data from the three individual countries, yielding a total of nine statistical models. Certain variables such as students' socio-economic status, grade repetition and gender are consistently found to influence academic performance, although in different ways depending on the context and the competence to which they refer.

Based on these results, high and low performing schools were identified to begin the study of the relationship between some non-contextual variables and school performance by means of binary logistic regression. Some of these variables are related to student self-concept, teacher self-efficacy or school climate. The most found factors related to school

¹ An extended abstract of the Ph.D. Dissertation can be found in Annex 1 (Anexo 1)

effectiveness are students' perceived ability to summarise texts, their self-concept related to reading in terms of their perceived competence and the use of ICT outside school for homework.

Finally, the main implications for educational policy and practice of the results found are unpacked, and a few limitations linked to the use of large-scale tests to study academic performance and school effectiveness such as their cross-sectional nature are detailed, as well as possible lines of research such as trend studies.

Keywords

Large scale assessment, PISA, school effectiveness, academic achievement, multilevel models, binary logistic regression, educational assessment.

ÍNDICE

INTRODUCCIÓN.....	20
Contexto de la investigación.....	24
Objetivos.....	26
1. EFICACIA ESCOLAR Y RENDIMIENTO ACADÉMICO: RECORRIDO HISTÓRICO Y RELACIÓN CON LAS EVALUACIONES A GRAN ESCALA... 28	
1.1. Eficacia escolar.....	30
1.2. Rendimiento académico	38
1.3. Pruebas de evaluación a gran escala.....	39
1.3.1. TIMSS – Estudio Internacional de Tendencias en Matemáticas y Ciencias.....	40
1.3.2. PIRLS – Estudio Internacional de Progreso en Comprensión Lectora	42
1.3.3. PISA – Programa para la Evaluación Internacional de Alumnos.....	43
2. REVISIÓN SISTEMÁTICA DE LA LITERATURA: PRUEBAS DE EVALUACIÓN A GRAN ESCALA, EFICACIA ESCOLAR Y RENDIMIENTO ACADÉMICO	60
2.1. Metodología empleada en la revisión de la literatura.....	63
2.1.1. Preguntas de investigación.....	64
2.1.2. Cadena de búsqueda.....	65
2.1.3. Criterios de elegibilidad	65
2.1.4. Procedimiento	69
2.2. Resultados del mapeo y de la revisión sistemática de la literatura.....	70
2.2.1. Resultados del estudio de <i>mapping</i>	70
2.2.2. Resultados de la revisión sistemática de la literatura.....	79
2.3. Conclusiones derivadas de la SLR	83
3. MODELOS MULTINIVEL.....	86
3.1. Fundamentos de los modelos multinivel: supuestos básicos.....	89
3.2. Especificación del modelo.....	90
3.2.1. Paso 1. Formulación del modelo nulo.....	90
3.2.2. Paso 2. Ampliación del modelo: el modelo condicional.....	91
3.3. Los efectos escolares y el valor añadido	92
3.4. Características metodológicas	94
3.4.1. Métodos de estimación.....	95

3.4.2.	Efectos fijos y aleatorios	95
3.5.	Ventajas y limitaciones de esta técnica.....	96
4.	METODOLOGÍA	99
4.1.	Objetivos.....	102
4.2.	Diseño.....	103
4.2.1.	Población y muestra	103
4.2.2.	Instrumentos	109
4.2.3.	Consideraciones metodológicas: el uso de las ponderaciones muestrales y de los valores plausibles.....	109
4.2.4.	Variables	110
4.2.5.	Procedimiento y análisis de datos	113
4.3.	Diseño de un cuestionario para la evaluación de la eficacia auto percibida por el profesorado de secundaria	115
5.	RESULTADOS	119
5.1.	Análisis de los factores contextuales del rendimiento académico de los estudiantes españoles.....	122
5.1.1.	Consideraciones previas.....	122
5.1.2.	Competencia matemática	125
5.1.3.	Competencia científica.....	133
5.1.4.	Competencia en comprensión lectora	140
5.2.	Análisis de los factores contextuales del rendimiento académico de los estudiantes irlandeses	147
5.2.1.	Consideraciones previas.....	147
5.2.2.	Competencia matemática	150
5.2.3.	Competencia científica.....	156
5.2.4.	Competencia en comprensión lectora	162
5.3.	Análisis de los factores contextuales del rendimiento académico de los estudiantes portugueses	170
5.4.	Análisis comparativo de los modelos estadísticos multinivel finales	170
5.5.	Selección de centros de alta y baja eficacia.....	170
5.6.	Análisis del efecto de las variables procesuales en la eficacia escolar.....	171
5.6.1.	Variables procesuales con influencia en la eficacia escolar en España	171
5.6.2.	Variables procesuales con influencia en la eficacia escolar en Irlanda	174
5.6.3.	Variables procesuales con influencia en la eficacia escolar en Portugal	176

5.7.	Validación del cuestionario para la evaluación de la eficacia auto percibida por el profesorado de secundaria en activo	177
6.	CONCLUSIONES, LIMITACIONES Y PROSPECTIVA	181
6.1.	Factors linked with academic performance	184
6.1.1.	Socio-economic and cultural status	184
6.1.2.	Grade repetition, grade and age of the students	186
6.1.3.	Student Gender	188
6.1.4.	Number of school changes	188
6.1.5.	Language spoken at home and immigration status	189
6.1.6.	Teachers with doctoral studies	191
6.1.7.	Duration on Early Childhood Education and Care	191
6.2.	Factors influencing school effectiveness	192
6.2.1.	Student's Metacognition - summarising.....	193
6.2.2.	Self-concept of reading: Perception of competence.....	194
6.2.3.	Use of ICT outside of school for schoolwork activities.....	194
6.3.	Limitations.....	195
6.4.	Future research lines	199
6.5.	Dissemination and publications.....	200
	REFERENCIAS	202
	ANEXOS.....	227
	Anexo 1 – Resumen extendido en inglés.....	229
	Anexo 2 – Variables procesuales incluidas en la Correlación de Pearson.....	271
	Anexo 3 – Documento enviado a los jueces para la validación del cuestionario ...	274
	Anexo 4 – Puntuaciones otorgadas por los expertos a los diferentes ítems.....	283
	Anexo 5 – Versión final del cuestionario	285

ÍNDICE DE TABLAS

Tabla 1.1. Tipos de formato de texto en las pruebas PISA 2018	52
Tabla 1.2. Relación entre los procesos y las habilidades matemáticas básicas en PISA	55
Tabla 1.3. Relación entre los contextos y los temas de la prueba de alfabetización científica	59
Tabla 2.1. Criterios de calidad de las publicaciones.....	67
Tabla 2.2. Número de publicaciones por autor	72
Tabla 2.3. Estudios individuales por país y referencias bibliográficas.....	75
Tabla 2.4. Número de publicaciones por fuente	77
Tabla 2.5. Número de publicaciones según la metodología utilizada	78
Tabla 2.6. Empleo de valores plausibles (PV) y pesos muestrales (PM) en investigaciones con datos PISA	81
Tabla 4.1. Distribución de la muestra por Comunidades Autónomas	104
Tabla 4.2. Variables contextuales.....	111
Tabla 4.3. Codificación de las dimensiones y los ítems	116
Tabla 5.1. Correlaciones entre las variables contextuales a nivel de estudiante en España	123
Tabla 5.2. Correlaciones entre las variables contextuales a nivel de centro en España	124
Tabla 5.3. Componentes de la varianza del modelo nulo de la competencia matemática en España.....	126
Tabla 5.4. Modelo final competencia matemática en España.....	127
Tabla 5.5. Componentes de la varianza del modelo final para la competencia matemática en España	128
Tabla 5.6. Estadísticos descriptivos para el residuo de nivel 1 en matemáticas en España	129
Tabla 5.7. Estadísticos descriptivos para el residuo de nivel 2 en matemáticas en España	131
Tabla 5.8. Componentes de la varianza del modelo nulo de la competencia científica en España.....	134
Tabla 5.9. Modelo final competencia científica en España	135
Tabla 5.10. Componentes de la varianza del modelo final para la competencia científica en España	136
Tabla 5.11. Estadísticos descriptivos para el residuo de nivel 1 en ciencias en España	136
Tabla 5.12. Estadísticos descriptivos para el residuo de nivel 2 en ciencias en España	138
Tabla 5.13. Componentes de la varianza del modelo nulo de la competencia en comprensión lectora en España	141
Tabla 5.14. Modelo final competencia en comprensión lectora en España.....	142
Tabla 5.15. Componentes de la varianza del modelo final para la competencia en comprensión lectora en España	143
Tabla 5.16. Estadísticos descriptivos para el residuo de nivel 1 en comprensión lectora en España.....	143
Tabla 5.17. Estadísticos descriptivos para el residuo de nivel 2 en comprensión lectora en España.....	145
Tabla 5.18. Correlaciones entre las variables contextuales a nivel de estudiante en Irlanda	148
Tabla 5.19. Correlaciones entre las variables contextuales a nivel de centro en Irlanda	149

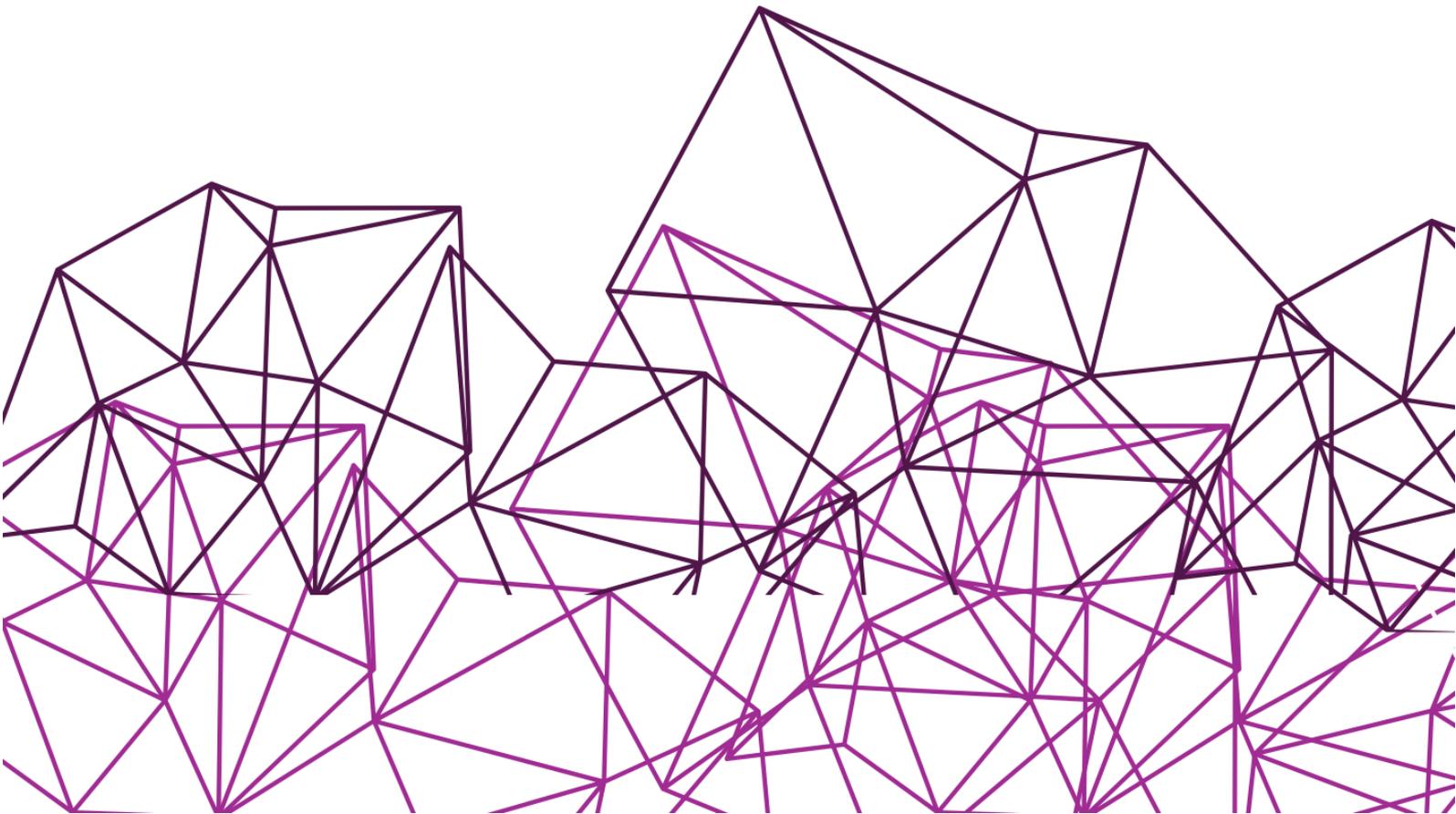
Tabla 5.20. Componentes de la varianza del modelo nulo de la competencia matemática en Irlanda	150
Tabla 5.21. Modelo final competencia matemática en Irlanda.....	151
Tabla 5.22. Componentes de la varianza del modelo final para la competencia matemática en Irlanda	152
Tabla 5.23. Estadísticos descriptivos para los residuos de nivel 1 y nivel 2 en matemáticas en Irlanda	153
Tabla 5.24. Componentes de la varianza del modelo nulo de la competencia científica en Irlanda	156
Tabla 5.25. Modelo final competencia científica en Irlanda	157
Tabla 5.26. Componentes de la varianza del modelo final para la competencia científica en Irlanda	158
Tabla 5.27. Estadísticos descriptivos del residuo de nivel 1 en ciencias en Irlanda.....	158
Tabla 5.28. Estadísticos descriptivos para el residuo de nivel 2 en ciencias en Irlanda	160
Tabla 5.29. Componentes de la varianza del modelo nulo de la competencia en comprensión lectora en Irlanda.....	162
Tabla 5.30. Modelo final competencia en comprensión lectora en Irlanda.....	163
Tabla 5.31. Componentes de la varianza del modelo final para la competencia en comprensión lectora en Irlanda.....	164
Tabla 5.32. Estadísticos descriptivos del residuo de nivel 1 en comprensión lectora en Irlanda.....	165
Tabla 5.33. Estadísticos descriptivos para el residuo de nivel 2 en comprensión lectora en Irlanda	166
Tabla 5.52. Resultados de la regresión logística binaria en España	172
Tabla 5.53. Precisión del modelo predictivo de la eficacia escolar en España.....	173
Tabla 5.54. Resultados de la regresión logística binaria en Irlanda	174
Tabla 5.55. Precisión del modelo predictivo de la eficacia escolar en Irlanda.....	175
Tabla 5.58. Estadísticos descriptivos para cada ítem y V de Aiken	178
Table 6.1. Initial CCI values for each competence in each country	192

ÍNDICE DE FIGURAS

Figura 1.1. Hitos principales de la investigación en eficacia educativa	32
Figura 1.2. Modelo dinámico de la eficacia escolar	36
Figura 1.3. Modelo de currículo planteado en TIMSS	41
Figura 1.4. Módulos de los cuestionarios de PISA 2018.....	46
Figura 1.5. Países y economías participantes en las distintas ediciones de PISA	48
Figura 1.6. Procesos del marco conceptual de comprensión lectora en PISA 2018.....	51
Figura 2.1. Diagrama de flujo con el proceso de búsqueda y selección de recursos.....	70
Figura 2.2. Número de publicaciones por año	71
Figura 2.3. Número de publicaciones por país	74
Figura 4.1. Fases del estudio	102
Figura 4.2. Características de la muestra de estudiantes	106
Figura 4.3. Distribución de la variable Nivel socioeconómico y cultural	107
Figura 4.4. Características de los centros educativos	108
Figura 4.5. Características de la muestra de profesores	108
Figura 5.1. Distribución del residuo de nivel 1 de la competencia matemática en España	130
Figura 5.2. Comprobación del supuesto de homocedasticidad del residuo de nivel 1 de la competencia matemática en España	131
Figura 5.3. Distribución del residuo de nivel 2 de la competencia matemática en España	132
Figura 5.4. Gráfico Q-Q para la distribución de residuos de nivel 2 en matemáticas en España.....	132
Figura 5.5. Comprobación del supuesto de homocedasticidad del residuo de nivel 2 en la competencia matemática en España	133
Figura 5.6. Distribución del residuo de nivel 1 de la competencia científica en España	137
Figura 5.7. Comprobación del supuesto de homocedasticidad del residuo de nivel 1 en ciencias en España	138
Figura 5.8. Distribución de los residuos de nivel 2 de la competencia científica en España	139
Figura 5.9. Comprobación del supuesto de homocedasticidad del residuo de nivel 2 en ciencias en España	139
Figura 5.10. Gráfico Q-Q para la distribución de residuos de nivel 2 en ciencias en España	140
Figura 5.11. Distribución del residuo de nivel 1 de la competencia en comprensión lectora en España	144
Figura 5.12. Gráfico Q-Q para los residuos de nivel 1 de la competencia en comprensión lectora en España	144
Figura 5.13. Comprobación del supuesto de homocedasticidad para los residuos de nivel 1 en la competencia en comprensión lectora en España.....	145
Figura 5.14. Distribución del residuo de nivel 2 de la competencia en comprensión lectora en España	146
Figura 5.15. Comprobación del supuesto de homocedasticidad del residuo de nivel 2 de la competencia en comprensión lectora en España.....	147
Figura 5.16. Distribución del residuo de nivel 1 de la competencia matemática en Irlanda	153

Figura 5.17. Comprobación del supuesto de homocedasticidad del residuo de nivel 1 de la competencia matemática en Irlanda.....	154
Figura 5.18. Distribución del residuo de nivel 2 de la competencia matemática en Irlanda	154
Figura 5.19. Gráfico Q-Q para la distribución de residuos de nivel 2 en matemáticas en Irlanda	155
Figura 5.20. Comprobación del supuesto de homocedasticidad del residuo de nivel 2 en la competencia matemática en Irlanda.....	155
Figura 5.21. Distribución del residuo de nivel 1 de la competencia científica en Irlanda	159
Figura 5.22. Comprobación del supuesto de homocedasticidad del residuo de nivel 1 en ciencias en Irlanda	159
Figura 5.23. Distribución de los residuos de nivel 2 de la competencia científica en Irlanda	160
Figura 5.24. Gráfico Q-Q para la distribución de residuos de nivel 2 en ciencias en Irlanda	161
Figura 5.25. Comprobación del supuesto de homocedasticidad del residuo de nivel 2 en ciencias en Irlanda	162
Figura 5.26. Distribución del residuo de nivel 1 en comprensión lectora en Irlanda...	165
Figura 5.27. Comprobación del supuesto de homocedasticidad del residuo de nivel 1 de la competencia en comprensión lectora en Irlanda.....	166
Figura 5.28. Distribución de las puntuaciones del residuo de nivel 2 para la competencia en comprensión lectora en Irlanda	167
Figura 5.29. Comprobación del supuesto de homocedasticidad del residuo de nivel 2 en comprensión lectora en Irlanda.....	168
Figura 5.30. Gráfico Q-Q para la distribución del residuo de nivel 2 en comprensión lectora en Irlanda	168

INTRODUCCIÓN



Introducción

A finales de la década de los años 60 del pasado siglo se publicaba *Equality of Educational Opportunity Study*, también conocido como Informe Coleman (Coleman, 1966), un documento en el que se hacía referencia a la importancia de los factores socioeconómicos y demográficos de los estudiantes como motivo de las desigualdades en su rendimiento académico, frente a la influencia de las características del centro escolar; fuente principal de inequidad hasta ese momento. Este informe desafiaba la idea extendida hasta entonces de que el reparto equitativo de los recursos escolares derivaría en la igualdad en el rendimiento. A partir de este momento, se inició una nueva corriente de investigación en el campo de la educación relacionada con la eficacia escolar (Investigación en Eficacia Escolar – IEE) para así tratar de demostrar que las escuelas sí producen efectos en el rendimiento educativo y comprender cuáles son los factores que pueden contribuir al mismo, en mayor o menor medida.

Esta corriente de investigación experimentó un gran auge en la década de los 90, surgiendo diferentes modelos para el desarrollo de este tipo de investigaciones como el modelo dinámico planteado por Creemers y Kyriakides (2008), según el cual, existen factores referidos a distintos niveles que pueden ayudar a explicar el rendimiento educativo. Desde características referidas al contexto a nivel de país, hasta el nivel del centro educativo o del propio estudiante. Este modelo ofrece una perspectiva contextualizada y detallada sobre las interacciones existentes entre los factores que afectan al proceso educativo y cómo pueden influir esas interacciones en el proceso de enseñanza-aprendizaje de los estudiantes.

El aumento de la investigación en esta temática derivó en la publicación de estudios centrados en analizar las relaciones existentes entre distintas variables y el rendimiento académico de los estudiantes, desde variables contextuales como el nivel socioeconómico y cultural, hasta otras de carácter procesual, como el liderazgo educativo (Gil-Flores & García-Gómez, 2017; Huang & Sebastian, 2015; Lizasoain & Angulo, 2014; Scheerens et al., 2015; Sirin, 2005; H. Wu, Shen, et al., 2020; Yetişir & Bati, 2021). Algunas de estas variables han demostrado consistentemente su influencia sobre el rendimiento del alumnado.

De la mano de estos avances en la IEE, se observó un aumento del interés en la medición del rendimiento educativo a través de pruebas estandarizadas de evaluación, también denominadas pruebas internacionales de evaluación a gran escala (ILSA por sus

siglas en inglés: *International Large Scale Assessments*). La naturaleza de estas evaluaciones en las que la información proviene de diferentes agentes, permite caracterizar las variables extraídas en función del nivel al que se refieren de acuerdo con lo planteado en el modelo dinámico (Creemers & Kyriakides, 2008). Las pruebas estandarizadas son, generalmente, planteadas por algún organismo internacional y pretenden presentar una visión comparativa de los sistemas educativos a nivel internacional. En este caso, se explican las características básicas de tres de ellas por su importancia para la investigación que se presenta:

1. El Programa para la Evaluación Internacional de Alumnos (PISA): propuesta por la Organización para la Cooperación y el Desarrollo Económico (OCDE) y dirigida a evaluar las habilidades de los estudiantes que se encuentran finalizando la etapa de escolarización obligatoria (15 años) en tres competencias básicas: comprensión lectora, matemáticas y ciencias.
2. El Estudio Internacional de Tendencias en Matemáticas y Ciencias (TIMSS): planteado por la International Association for the Evaluation of Educational Achievement (IEA) dirigido a los estudiantes de 4º de Educación Primaria y 2º de Educación Secundaria Obligatoria para la evaluación de las competencias cognitivas especificadas (matemáticas y ciencias).
3. El Estudio Internacional del Progreso en Comprensión Lectora (PIRLS): también planteado por la IEA dirigido a los estudiantes de 4º de Educación Primaria con el fin de evaluar su competencia en comprensión lectora.

Del mismo modo, las técnicas e instrumentos utilizados para el análisis de los resultados de estos estudios fueron sufriendo cambios debido, en parte, a las características de las ILSA. Para el estudio estadístico de las características de diferentes niveles de análisis destaca el uso de la metodología multinivel, propuesta en los años 80 (Aitkin & Longford, 1986; Raudenbush & Bryk, 1992). La presente investigación se fundamenta en esta base metodológica.

En España, al igual que en otros países, la eficacia escolar ha sido un tema de interés continuo, siendo nuestro país uno de los que mayor atención y esfuerzos ha prestado al estudio de la eficacia escolar (ver sección 2.2).

Destaca la labor desarrollada por el grupo HEPE – Evaluación de Programas, Centros y Sistemas Educativos dependiente de la Universidad del País Vasco que sentó las bases

de la IEE en España o por el Grupo de Evaluación Educativa y Orientación (GE2O)² dependiente de la Universidad de Salamanca y encargado del desarrollo de proyectos I+D centrados en la eficacia escolar y, actualmente, en la equidad educativa.

Contexto de la investigación

La tesis doctoral que se presenta se ha desarrollado dentro del Programa de Doctorado Formación en la Sociedad del Conocimiento³ en el Instituto Universitario de Ciencias de la Educación de la Universidad de Salamanca (IUCE). Este programa de doctorado presenta un carácter eminentemente interdisciplinar con hasta nueve líneas de investigación y más de cinco Grupos de Investigación Reconocidos de la Universidad de Salamanca, encontrándose este estudio enmarcado en el área de Evaluación educativa y Orientación.

El grupo que desarrolla la línea de evaluación (GRIAL-GE2O), ha llevado a cabo múltiples proyectos de I+D+i nacionales desde el año 2006 hasta la actualidad. Esta tesis se enmarca en el proyecto concedido al amparo de la convocatoria de Proyectos de Generación del Conocimiento del año 2018 bajo la financiación del Ministerio de Ciencia e Innovación del Gobierno de España titulado *School Effectiveness for the Improvement of the System – EFI-6* (PGC2018–099174–B–I00). La presente tesis doctoral ha sido además financiada mediante una Ayuda para contratos predoctorales con referencia PRE2019-087412 por MCIN/AEI/10.13039/501100011033 y por FSE invierte en tu futuro asociada al proyecto anterior.

Con anterioridad al Proyecto EFI-6, el grupo ha sido beneficiario de las siguientes subvenciones para la realización de proyectos de I+D+i financiados por distintos Ministerios dependientes del Gobierno de España:

- Modelo de evaluación y desarrollo de competencias clave implícitas en la ESO: TIC, Lectura y Convivencia Escolar (E-TELECO), referencia SEJ2006-10700; financiado por el Ministerio de Educación y Ciencia.
- Evaluación de competencias clave y formación del profesorado de educación secundaria: TIC, ALFIN y convivencia escolar (EF-TALCO), referencia

² Integrante del Grupo de Investigación Reconocido Grupo de Investigación en Interacción y e-Learning (GRIAL) y del Instituto de Ciencias de la Educación de la USAL (IUCE).

³ <https://knowledgesociety.usal.es>

EDU2009-08753; concedido en el año 2009 y financiado por el Ministerio de Ciencia e Innovación.

- Evaluación, Formación e Innovación sobre competencias clave en Educación Secundaria: TIC, Competencia Informacional y Resolución de Conflictos (EFI-CINCO), referencia EDU2012-34000; concedido en el año 2012 financiado por el Ministerio de Economía y Competitividad.
- Evaluación de Impacto del Desarrollo de Competencias Básicas sobre el Rendimiento Académico en Educación Secundaria: Propuesta de Formación e Innovación Docente (EFI-4), referencia EDU2015-64524-P; financiado por el Ministerio de Economía, Industria y Competitividad.

De manera posterior al proyecto de investigación en el que se encuadra esta tesis (PGC2018-099174-B-I00), este mismo Ministerio ha concedido al citado grupo de investigación un nuevo proyecto de investigación bajo el título *Evaluación y promoción de la equidad educativa en Educación Secundaria: Análisis secundario de las evaluaciones PISA y formación de agentes educativos claves* (PID2021-125775NB-I00) con el que se pretende continuar las investigaciones realizadas hasta el momento.

Por último, cabe mencionar que, como parte complementaria a la formación predoctoral desarrollada en este periodo (2019 – actualidad), la autora de esta Tesis ha realizado tres estancias de investigación buscando conseguir aportar una perspectiva internacional sobre el tema estudiado. La primera de ellas tuvo lugar en el Centre for Evaluation, Quality & Inspection (EQI) dependiente de la Dublin City University (DCU) en Irlanda, con el objetivo de seleccionar las escuelas de alta y baja eficacia irlandesas y obtener información sobre las prácticas desarrolladas en ellas. La segunda estancia, se desarrolló en la Escola Superior de Educação e Ciências Sociais del Politécnico de Leiria (Portugal), consiguiendo un objetivo idéntico al planteado en el caso irlandés, pero aplicado al contexto luso. Por último, se desarrolló una estancia a nivel nacional en la Universidade da Coruña (UDC) donde se seleccionaron los centros educativos de alta y baja eficacia de la comunidad gallega y se compararon sus características contextuales con las de los estos centros en Castilla y León.

Objetivos

Esta tesis plantea como objetivo principal la generación de un marco de buenas prácticas educativas relacionadas con la eficacia escolar a nivel internacional, estudiando concretamente los casos de tres países europeos: España, Irlanda y Portugal. Este objetivo general se pretende alcanzar a través de los siguientes objetivos específicos:

- Seleccionar las variables de contexto a nivel de centro educativo y de estudiante que presenten relaciones significativas en el rendimiento de los estudiantes según los datos proporcionados por las pruebas PISA 2018 aplicando la metodología de modelos multinivel.
- Detectar los centros educativos que rinden por encima y por debajo de lo esperado según sus características contextuales en estos tres países. Es decir, determinar cuáles son los centros de alta y baja eficacia escolar.
- Analizar las variables relacionadas con la eficacia escolar mediante la regresión logística binaria para determinar las relaciones existentes entre distintas variables de proceso y esta variable criterio.
- Conocer las buenas prácticas implementadas por los centros educativos con altos residuos (alta eficacia) para tratar de replicarlas en aquellos centros con valores residuales más bajos.

El documento que se presenta se estructura en torno a seis apartados. En el primero de ellos se revisa el contenido teórico sobre el que se construye esta investigación, prestando especial atención a tres conceptos clave de la misma, como son: eficacia escolar, rendimiento académico y pruebas de evaluación a gran escala.

A continuación, para profundizar en los estudios que han tomado como base las pruebas de evaluación a gran escala, en la sección 2 se presenta una revisión sistemática de la literatura donde se analizan un total de 63 artículos de investigación obteniendo información sobre la evolución de la investigación en este campo, los autores, revistas y países que más atención han prestado a esta temática, etc. También se han tenido en cuenta las metodologías de análisis de datos empleadas, así como a las variables estudiadas.

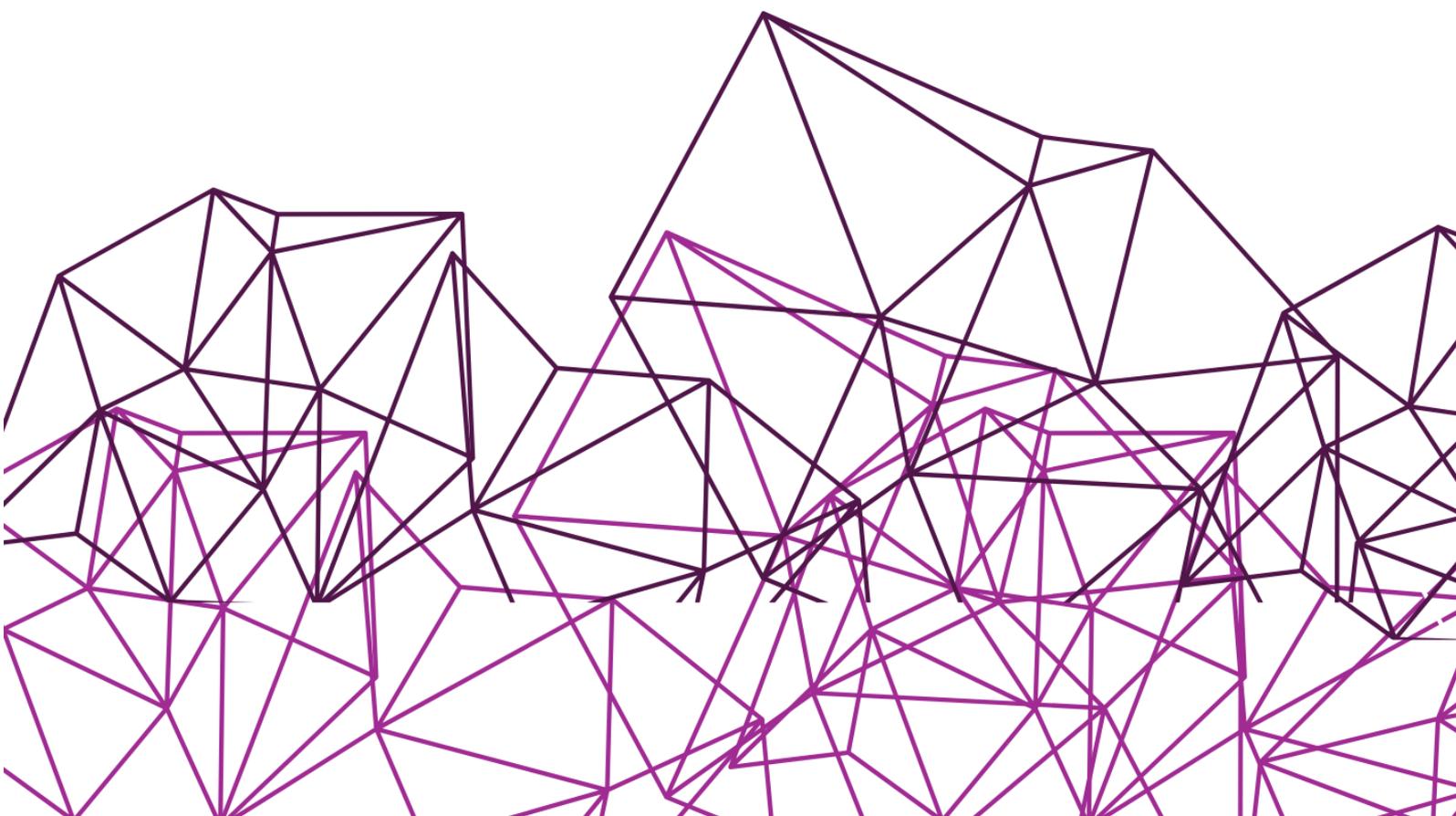
En el tercer capítulo se presenta una introducción a la metodología de análisis multinivel, incluyendo la definición y especificación de los distintos modelos, así como los supuestos básicos que deben cumplirse y los tipos existentes.

El cuarto apartado se centra en la descripción de la metodología aplicada en el desarrollo de la presente investigación detallando diseño, muestra, variables empleadas y procedimiento de análisis de datos.

Los resultados del estudio se presentan en la sección 5, donde en primer lugar, se aborda la construcción de los diferentes modelos multinivel para las tres competencias básicas estudiadas en PISA en los tres países para, a continuación, determinar las variables contextuales relacionadas con el rendimiento académico. Se prosigue con la selección y caracterización de los centros de alta y baja eficacia. Finalmente, se presentan aquellas variables de proceso con influencia en la eficacia escolar en España, Irlanda y Portugal.

Por último, se presenta la discusión de los resultados obtenidos y sus implicaciones, así como las conclusiones más relevantes alcanzadas con esta investigación, confrontándolas con las obtenidas en la literatura existente en relación con la práctica educativa.

1. EFICACIA ESCOLAR Y RENDIMIENTO ACADÉMICO: RECORRIDO HISTÓRICO Y RELACIÓN CON LAS EVALUACIONES A GRAN ESCALA



Eficacia escolar y rendimiento académico: recorrido histórico y relación con las evaluaciones a gran escala

Siguiendo lo propuesto en el apartado introductorio, esta tesis doctoral se construye integrando tres grandes áreas temáticas dentro de la investigación educativa: la eficacia escolar, el rendimiento académico y las pruebas de evaluación a gran escala. Para ello, se ha diseñado una investigación en la que se hace necesario comprender qué es lo que se entiende por eficacia escolar y rendimiento académico y cuál ha sido la evolución de la investigación en estos campos.

En este primer capítulo se propone una revisión bibliográfica centrada en mostrar un análisis general sobre la eficacia escolar, el rendimiento académico y las pruebas de evaluación a gran escala, prestando especial atención en el apartado dedicado a las pruebas de evaluación a gran escala al Programa Internacional de Evaluación de los Estudiantes (PISA por sus siglas en inglés) planteado por la OCDE. Este mayor esfuerzo dedicado a la explicación de esta prueba de evaluación se debe a que es con los datos de PISA 2018 con los que se realiza el análisis cuantitativo planteado en esta investigación.

Se presenta a continuación una revisión del estado de la cuestión sobre la eficacia escolar, el rendimiento académico y las pruebas de evaluación a gran escala desde el inicio de la investigación en estos campos hasta su situación actual, estructurado en esos tres grandes apartados: eficacia escolar, rendimiento académico y pruebas de evaluación a gran escala.

1.1. Eficacia escolar

Para analizar el origen de esta investigación debemos remontarnos a los años 60, cuando el gobierno de los Estados Unidos encargó un estudio exhaustivo sobre la situación de la igualdad de oportunidades educativas en su país a un grupo de expertos educativos. El grupo estaba dirigido por Coleman y en el citado estudio contaron con la participación de aproximadamente 600 000 estudiantes y 60 000 docentes de unos 4 000 centros educativos (Fenstermacher & Soltis, 1998). Esta investigación obtuvo como resultado el informe *Equality of Educational Opportunity* (Coleman, 1966), comúnmente conocido como Informe Coleman. En él se definieron una serie de indicadores específicos sobre distintas áreas educativas que han servido de base para la investigación educativa en las últimas décadas. Estos indicadores versaban sobre el gasto en educación, los recursos

humanos o los contextos de la educación entre otros. Las principales conclusiones expuestas en el Informe Coleman sugerían que el gasto en educación no parecía suponer grandes diferencias en el logro educativo (rendimiento académico) de los estudiantes. Por lo que se destaca la necesidad de otro tipo de variables relacionadas con las características socioeconómicas y culturales de los estudiantes, así como su entorno familiar, etc. (Coleman, 1966). Este documento se considera el origen de la investigación sobre la eficacia escolar y las pruebas de evaluación a gran escala, y derivó en multitud de investigaciones para profundizar en sus hallazgos o, incluso, desmentirlos (Fenstermacher & Soltis, 1998). Con la publicación de este informe se generó una nueva corriente de investigación en el campo de la educación denominada Investigación en Eficacia Escolar (IEE), en la que se pretendía demostrar el posible efecto de las escuelas en el rendimiento educativo de los estudiantes. Uno de los campos más desarrollados dentro de esta corriente es el dedicado a conocer los factores que influyen en la eficacia escolar en mayor o menor medida desde diferentes perspectivas.

Durante los últimos 50 años ha habido notables avances en la investigación educativa, generando nuevos conceptos, movimientos y reformas educativas. Murillo define la IEE como el

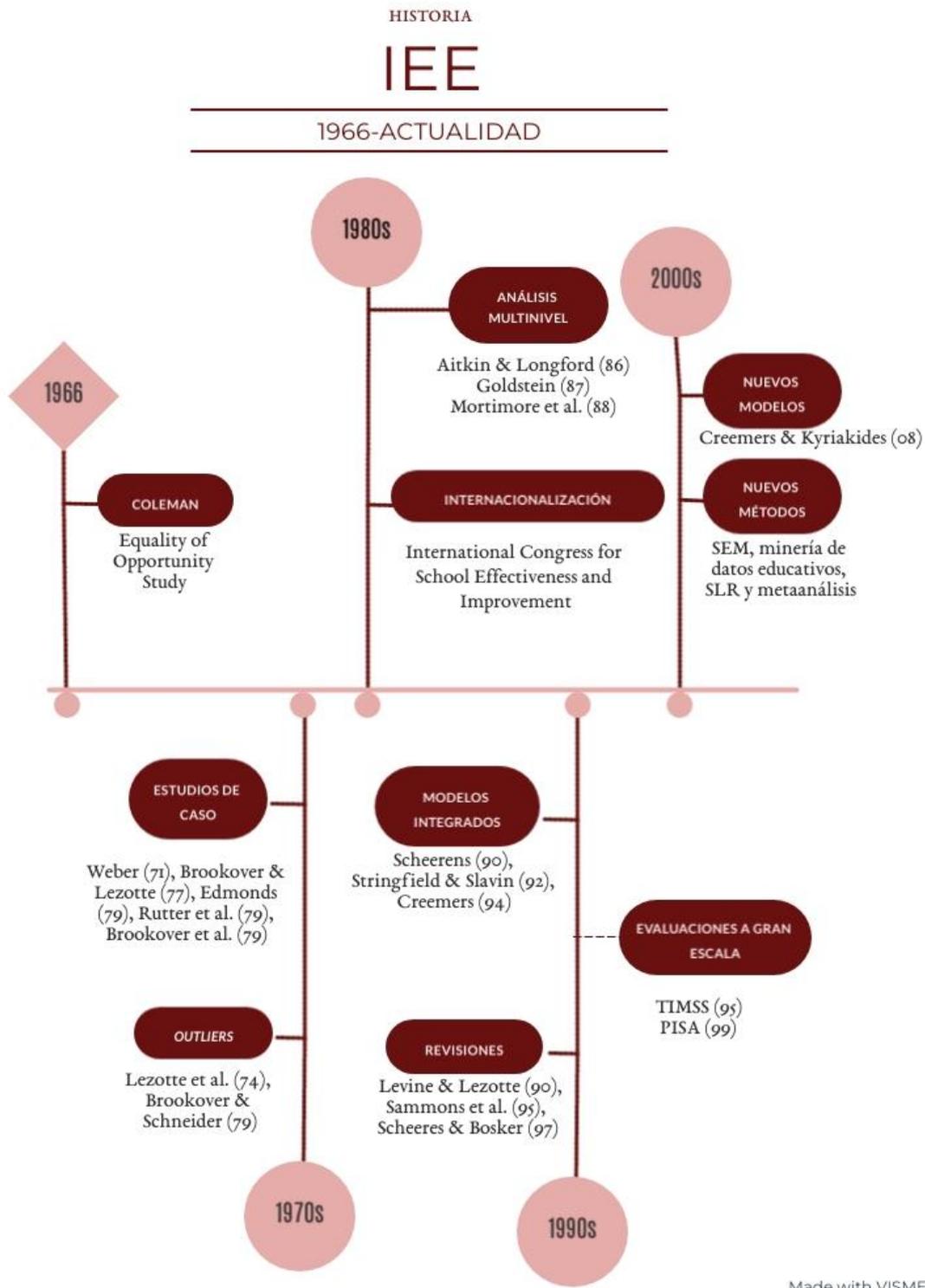
Conjunto de estudios empíricos que tienen por objeto la estimación de la magnitud de los efectos escolares y el análisis de sus propiedades científicas, y/o el estudio de los factores escolares, de aula y de contexto que caracterizan una escuela eficaz, sea cual sea el enfoque metodológico utilizado para conseguirlo (Murillo, 2005, p. 22)

A partir de esta definición, sería pertinente concretar el significado de las escuelas eficaces como aquellos centros educativos en los que los alumnos rinden por encima de lo esperado teniendo en cuenta la influencia de sus características contextuales en el rendimiento académico y logrando el desarrollo integral de los estudiantes (Murillo, 2005).

Si se pretende realizar un recorrido histórico por esta disciplina, resulta de gran utilidad la revisión realizada por Gamazo (2019) sintetizada en la figura 1.1:

Figura 1.1

Hitos principales de la investigación en eficacia educativa



Nota: Tomada de Factores asociados al rendimiento y a la eficacia escolar: un estudio basado en métodos mixtos a partir de PISA 2015 (p. 80), por Gamazo, (2019).

En el eje cronológico presentado se observan los hitos principales de la investigación en eficacia educativa, comenzando por la publicación del Informe Coleman en el año 1966 ya comentado previamente. Posteriormente, en la década de los años 70 y, a partir de los resultados presentados en el *Equality of Opportunity Study* que retrataban la escasa diferencia en el rendimiento causada por factores relacionados con las escuelas, se comenzaron a realizar estudios de caso buscando profundizar en el conocimiento de las denominadas escuelas eficaces.

La primera de las investigaciones destacadas es la desarrollada por Weber (1971). En ella se presentaba un estudio de cuatro escuelas estadounidenses que demostraban su eficacia pese a disponer de unas características socioeconómicas más desfavorecidas que la media del país, condición resaltada como factor más influyente en la eficacia escolar (Coleman, 1966). Con esta investigación se pretendía demostrar que hay factores más relevantes con influencia en la eficacia escolar que el nivel socioeconómico, concretamente estudiando variables como las metodologías aplicadas por el profesorado en las aulas, la organización del currículo, las prácticas de liderazgo o el clima escolar, es decir, factores procesuales implicados en el proceso de enseñanza-aprendizaje. De tal modo, la investigación desarrollada por Weber introdujo un cambio en la IEE al añadir otra serie de factores a los contextuales estudiados hasta el momento.

Al igual que en estudios similares de la época, se pretendía obtener conocimiento sobre la eficacia escolar mediante el estudio del rendimiento académico y su relación con distintas variables (Brookover et al., 1979; Brookover & Lezotte, 1979; Edmonds, 1979; Rutter et al., 1979), en el caso de Weber se trabajaba sobre el nivel lector de los estudiantes de las cuatro escuelas seleccionadas (1971).

En la figura se diferencian dos tipos de estudios en los años 70, los estudios de caso, comentados previamente, y los *outliers* o estudios de casos atípicos, de acuerdo con la clasificación propuesta por Purkey y Smith (1983). Dentro de los estudios *outliers*, los autores diferencian entre positivos y negativos “One major strategy of school effectiveness research has been to determine statistically highly effective schools (positive outliers) and unusually ineffective schools (negative outliers)” (Purkey & Smith, 1983, p. 430). En relación con los estudios de casos atípicos, destacan los trabajos de Brookover y Schneider (1975) y Lezotte et al. (1974) centrados en explicar las diferencias en el nivel de rendimiento entre escuelas con un estatus socioeconómico y una composición racial similares.

Durante la década de los años 80 del pasado siglo, comenzaron a surgir nuevas técnicas estadísticas más complejas que se adaptaban mejor a las características naturales de la realidad educativa, permitiendo llevar a cabo análisis más completos. Una de ellas es la metodología de análisis multinivel o modelos multinivel (metodología explicada en el capítulo 3 de este documento), que posibilitaba tener en cuenta la posible variación existente en distintos niveles de análisis junto con la independencia de las observaciones (Aitkin & Longford, 1986; Murillo, 2008; Raudenbush & Bryk, 1992). Estos niveles, en el contexto educativo, podrían abarcar, por ejemplo: el ámbito nacional, el institucional (escolar), el aula y el estudiante; cada uno con unas características comunes y otras diferenciales.

Las propuestas que más destacaron fueron las planteadas por Aitkin y Longford, (1986), Goldstein (1987) y Mortimore et al. (1989) en las que los autores aplicaban esta nueva metodología de investigación obteniendo resultados muy relevantes para la investigación en eficacia escolar. Para la correcta aplicación de esta técnica estadística determinaron la necesidad de información específica sobre variables de contexto de los distintos niveles (estudiantes, escuelas y administraciones educativas), así como datos del propio rendimiento de los estudiantes.

Hasta el momento la IEE se desarrollaba mayoritariamente en el contexto estadounidense, pero a partir de esta década se inicia un proceso de internacionalización de las experiencias de la mano de la fundación del International Congress for School Effectiveness and Improvement (ICSEI). Esta organización nació con el propósito de impulsar la calidad y la equidad de la educación promoviendo y contribuyendo a la mejora y la eficacia de las escuelas y otras instituciones educativas (International Congress for School Effectiveness and Improvement [ICSEI], 2024). Para alcanzarlo, se busca desde este organismo la colaboración internacional en el desarrollo de estudios relacionados con la eficacia escolar, así como el intercambio de experiencias mediante la creación de un congreso internacional celebrado con carácter anual desde el año 1988, y la fundación de la revista científica *School Effectiveness and School Improvement* en el año 1990.

La década de los 90 destaca por dos hitos: la elaboración de modelos integrados y el desarrollo de revisiones de la literatura sobre el estado de la cuestión en esta área (Chitty, 1997; Levine & Lezotte, 1990; Sammons et al., 1995; Scheerens & Bosker, 1997). Numerosos autores europeos, sobre todo procedentes de Países Bajos, dedicaron sus esfuerzos al estudio de los factores ligados a la eficacia escolar en su país. Los resultados de estos trabajos se describen a continuación.

El análisis realizado sobre los factores más comúnmente relacionados con la eficacia escolar se estructuró de manera sistemática diferenciando entre los factores de contexto, los factores de entrada (*inputs*), los de proceso y los resultados del propio sistema educativo (*outputs*) (Scheerens & Bosker, 1997). Este tipo de modelos se denominan modelos integrados. Con los estudios realizados durante esta época se percibió una mayor influencia del nivel del aula frente al nivel de escuela (Creemers, 1994; Teddlie & Stringfield, 1993).

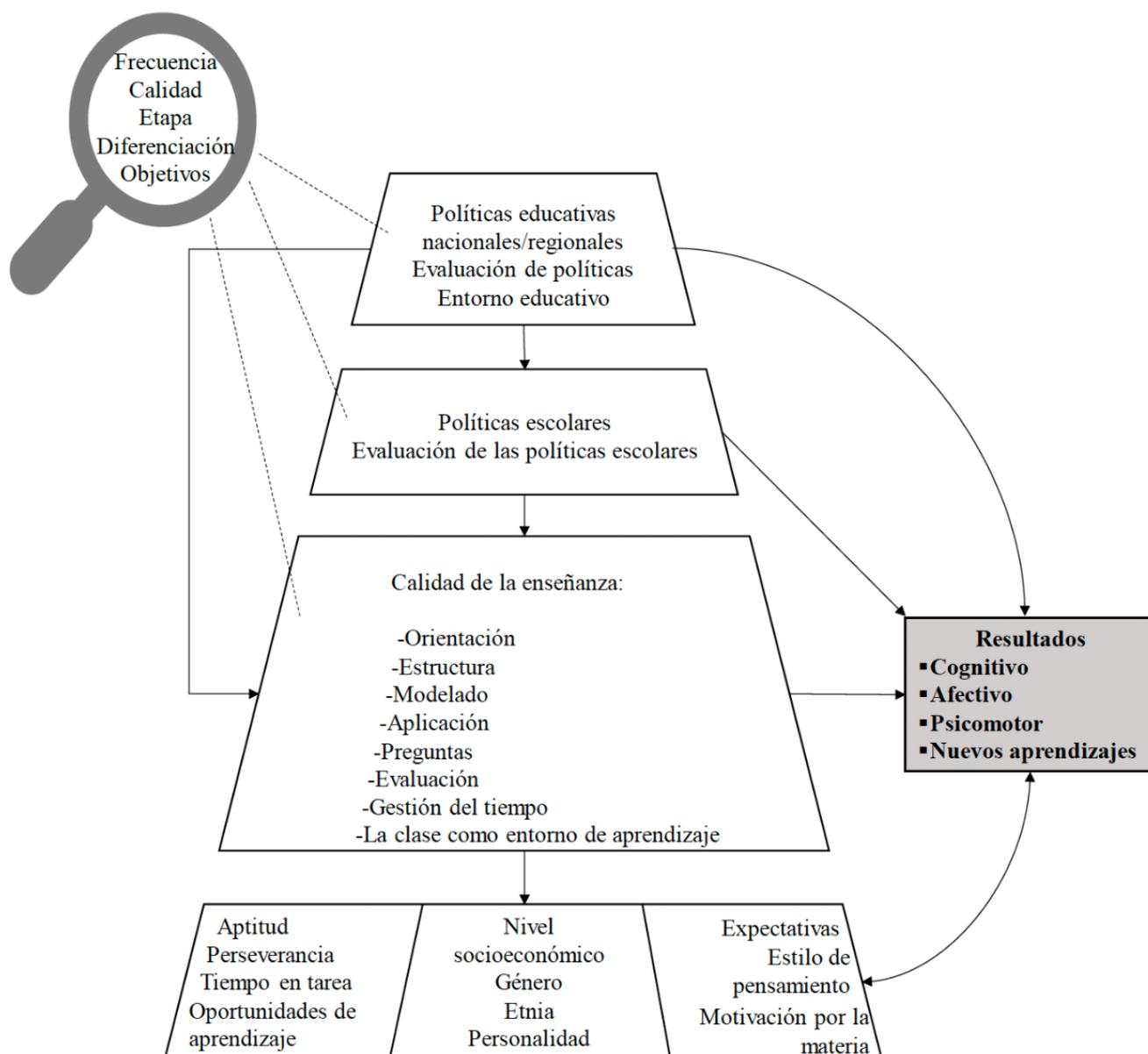
Cabe destacar en este momento la aparición de las pruebas de evaluación a gran escala, creadas con objetivos diversos para conocer el rendimiento de los estudiantes en diferentes etapas educativas (consultar en el apartado 1.3).

Junto con el desarrollo de los distintos tipos de estudios relacionados con la eficacia escolar, se elaboraron múltiples modelos teóricos que los fundamentasen como los propuestos por Scheerens (1990), Stringfield y Slavin (1992) o Creemers (1994). Estos modelos fueron evolucionando y generando otros nuevos en los años posteriores.

En la década de los años 2000 se desarrollaron nuevos enfoques para estudiar la IEE. A continuación (figura 1.2), se presenta el modelo propuesto por Creemers y Kyriakides (2008). Este modelo diferencia entre cuatro niveles con influencia en la eficacia escolar: nacional/regional, escolar, aula y estudiante. Cada uno se relaciona con unas variables concretas.

Figura 1.2

Modelo dinámico de la eficacia escolar



Nota. Tomada de *Factores asociados al rendimiento y a la eficacia escolar: un estudio basado en métodos mixtos a partir de PISA 2015* (p. 95), por Gamazo, (2019), adaptado de Creemers y Kyriakides, (2008).

En el nivel nacional/regional, los autores resaltan la falta de información causada por la limitada cantidad de investigaciones que estudien su relación con la eficacia escolar. Este nivel lo sustentan las políticas nacionales y regionales con influencia en alguno de los niveles inferiores y su evaluación y las características contextuales referidas, por ejemplo, a la financiación.

A nivel escolar, los estudios se centran en las políticas implementadas en las escuelas referidas a la docencia y al clima escolar además de en su evaluación.

En el nivel de aula indican la existencia de ocho factores implicados en la educación eficaz y ligados a la práctica docente: orientación hacia los estudiantes, estructura de los contenidos a trabajar, modelado de las tareas, aplicación de los aprendizajes obtenidos, preguntas para conseguir la participación de los estudiantes en la clase, evaluación o seguimiento de los estudiantes entendida como proceso de evaluación formativa, gestión del tiempo buscando la optimización del mismo y la clase como entorno de aprendizaje.

El último nivel indicado, el referido a los estudiantes, se caracteriza por la relación de dos tipos de variables. Por un lado, aquellas de corte contextual tan estudiadas hasta el momento y que se han mantenido relevantes a lo largo del tiempo como el género o el nivel socioeconómico. Por otro lado, las variables denominadas como psicológicas, con una tradición más reciente en el estudio de la eficacia escolar y relacionadas con aspectos como la motivación y las expectativas de los estudiantes, o su actitud hacia las tareas, aspectos que varían enormemente durante la evolución de los estudiantes y su paso por el sistema educativo.

Los autores entienden que los resultados de aprendizaje no se deben limitar a la obtención de conocimientos por parte de los estudiantes (resultados cognitivos), sino que existen otro tipo de resultados esenciales para la formación integral de las personas, relacionados con aspectos afectivos, psicomotores o con nuevos aprendizajes.

La evolución teórica de la disciplina ha ido acompañada de una evolución metodológica fundamentada en el uso de las técnicas estadísticas emergentes y del desarrollo de distintas herramientas tecnológicas que han servido para facilitar la investigación en este ámbito. Algunas de estas técnicas aplicadas al estudio de la eficacia escolar y el rendimiento académico en los últimos tiempos son la minería de datos educativos (Ayala Franco et al., 2021; Martínez-Abad et al., 2020), las revisiones sistemáticas de la literatura (SLR) (Cordero Ferrera et al., 2013), el metaanálisis o los modelos de ecuaciones estructurales (SEM) (Li, 2016; Mora-Ruano et al., 2021; Samperio Pacheco, 2019).

1.2. Rendimiento académico

Como se evidencia a lo largo de este apartado, la investigación educativa sobre el rendimiento académico goza también de una larga tradición literaria, específicamente centrada en la investigación sobre los factores asociados al mismo relacionada con la publicación del Informe Coleman.

El rendimiento académico puede definirse como el nivel de logro que un estudiante alcanza en un área o materia (Edel Navarro, 2003; Jiménez Hernández, 1994), o lo que es lo mismo, lo que el alumno ha aprendido durante el proceso formativo, “el resultado del aprendizaje suscitado por la actividad didáctica del profesor y producido en el alumno” (Lamas, 2015, p. 315). A pesar de la naturaleza multifactorial indicada y concretada en los diversos componentes que conforman este constructo, este concepto se ha visto reducido históricamente a la obtención de calificaciones por parte de los estudiantes, entendiéndose por tanto el rendimiento académico como los resultados educativos. Sin embargo, en las últimas décadas la investigación del ámbito educativo se ha centrado en explorar los factores predictores del rendimiento, ampliando de esta forma la conceptualización del rendimiento académico.

Los factores asociados al rendimiento académico son muchos y muy diversos, desde las características socioeconómicas del país hasta aspectos relacionados con la organización de las escuelas y las aulas o con las propias características de los estudiantes.

Con la revisión sistemática de la literatura que se presenta en el capítulo siguiente de este documento se pretende ahondar en este tema, presentando la evolución del estudio de los factores asociados al rendimiento académico y la eficacia escolar. Estas investigaciones gozan de una gran relevancia en el ámbito educativo, ya que se consideran herramientas esenciales para guiar las políticas educativas y las prácticas pedagógicas, además de servir como instrumentos de diagnóstico de los sistemas educativos en caso de tratarse de pruebas de evaluación estandarizadas capaces de medir esta variable (Heyneman & Lee, 2014).

Desde los años 80, los factores asociados al rendimiento académico mayoritariamente estudiados fueron aquellos de carácter procesual como el liderazgo escolar (Edmonds, 1979; Mortimore et al., 1989; H. Wu, Gao, et al., 2020), el clima escolar (Edmonds, 1979; Scheerens & Bosker, 1997; Shin et al., 2009) o la participación de los padres en la educación de sus hijos (participación educativa) (Ertem, 2021; Rodríguez Gómez, 1991) y otros tantos de carácter contextual como el entorno socioeconómico y cultural de los

estudiantes (Burns et al., 2020; Pomianowicz, 2021; Sousa et al., 2012) o este mismo aspecto, pero referido a las escuelas (Ding & Homer, 2020; Kameshwara et al., 2020; Martins & Veiga, 2010).

Es en este periodo ha aumentado el interés por la medición del rendimiento académico a través de pruebas estandarizadas, derivando en la generación e implementación de las citadas pruebas de evaluación a gran escala.

1.3. Pruebas de evaluación a gran escala

La evaluación a gran escala se refiere a la aplicación estandarizada de pruebas de evaluación a un gran número de estudiantes para comprobar cuál es el estado del sistema educativo de un grupo (país, región, etc.) en un determinado momento (Martínez Rizo, 2009) con el fin de determinar y conocer, entre otras variables, el rendimiento educativo de los estudiantes. Básicamente la necesidad de aplicación de pruebas estandarizadas o pruebas a gran escala surgió de la premisa de que las evaluaciones propuestas por los maestros tenían serias deficiencias, buscando la generación de instrumentos que permitiesen comprobar el nivel de logro educativo de los estudiantes de diferentes entornos (Martínez Rizo, 2009) bajo el respaldo de distintas técnicas estadísticas (Jiménez Moreno, 2014).

Aplicando esta definición a un contexto internacional, se obtiene un nuevo concepto: las pruebas internacionales de evaluación a gran escala (ILSA). Su objetivo principal es capturar la situación de los sistemas educativos de diferentes territorios para posteriormente establecer comparaciones entre los mismos llegando, en ciertos casos, a influir en la toma de decisiones en materia de política educativa (Jiménez Moreno, 2016).

Persiguiendo ese objetivo, diferentes organismos internacionales han dedicado sus esfuerzos al diseño y a la posterior aplicación de pruebas internacionales de evaluación a gran escala. En este caso, se presentan tres ILSA propuestas por dos organismos internacionales debido a su relevancia y amplitud de uso dentro de la investigación educativa:

1. El Estudio Internacional de Tendencias en Matemáticas y Ciencias (*Trends in International Mathematics and Sciences Study*, TIMSS), desarrollado por la Asociación Internacional para la Evaluación del Rendimiento Educativo (*International Association for the Evaluation of Educational Achievement*, IEA).

2. El Estudio Internacional de Progreso en Comprensión Lectora (*Progress in International Reading Literacy Study*, PIRLS), propuesto también por la IEA.
3. El Programa para la Evaluación Internacional de Alumnos (*Programme for International Student Assessment*, PISA), planteado por la Organización para la Cooperación y el Desarrollo Económicos (OCDE).

En definitiva, las pruebas internacionales de evaluación a gran escala proporcionan una visión a nivel de sistema que retrata los resultados de aprendizaje del estudiantado en un determinado momento de su vida escolar. En palabras de Jornet y Suárez “no es posible abordar un análisis adecuado de un sistema o un Programa si no se cuenta con pruebas estandarizadas de probada fiabilidad y validez” (1996, p. 145).

Una de las principales ventajas de este tipo de pruebas de evaluación es la comparabilidad de los resultados (Martínez Rizo, 2009) generada por su riguroso proceso de construcción. Aunque también están sujetas a multitud de limitaciones que se exponen a lo largo de este documento, se debe destacar el hecho de que en ellas no se tienen en cuenta las distintas culturas de los países participantes (Lenkeit, 2012), lo que la literatura científica ha denominado como sesgo cultural de las ILSA.

En la actualidad, las pruebas internacionales de evaluación a gran escala constituyen la manera más común de monitoreo del estatus de los sistemas educativos (Jiménez Moreno, 2016). A continuación, se exponen las características principales de las tres ILSA mencionadas.

1.3.1. TIMSS – Estudio Internacional de Tendencias en Matemáticas y Ciencias

El Estudio Internacional de Tendencias en Matemáticas y Ciencias (TIMSS – *Trends in International Mathematics and Science Study*) es una prueba de evaluación a gran escala planteada por la IEA para evaluar las competencias cognitivas en estas dos materias de los estudiantes de 4º y 8º grado (en España 4º de Educación Primaria (EP) y 2º de Educación Secundaria Obligatoria (ESO)) (Instituto Nacional de Evaluación Educativa [INEE], Ministerio de Educación y Formación Profesional, 2023).

Esta prueba se realiza desde el año 1995 cada cuatro años, siendo 4 años la diferencia existente entre las dos etapas evaluadas en cada edición. TIMSS se emplea para presentar un estudio sobre las tendencias educativas en estas dos disciplinas, analizando la evolución en el rendimiento académico de los estudiantes en cuatro años. Para ello, se

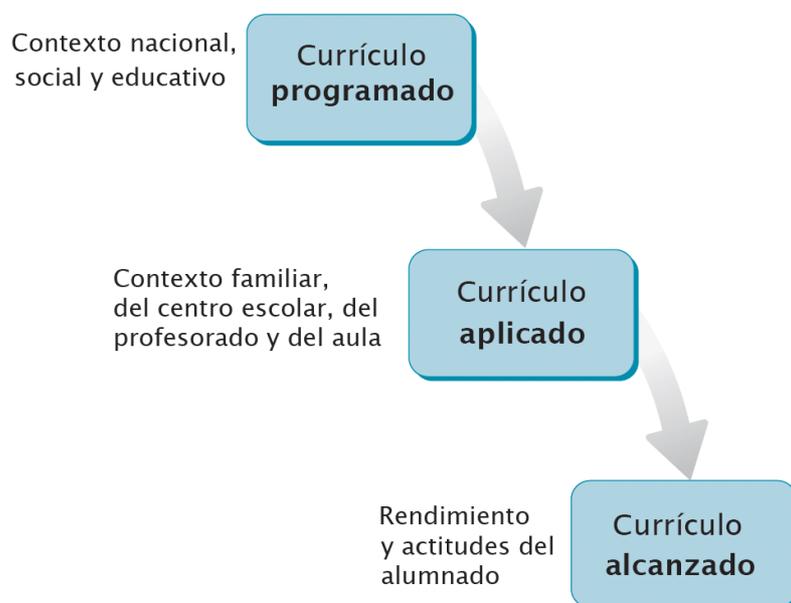
asume que la cohorte de 4º de EP de cada país en una determinada edición es prácticamente la misma que la cohorte de 2º de ESO de la edición inmediatamente siguiente. Esto hace que desde la propia IEA se caracterice este estudio bajo un diseño de tipo cuasi-longitudinal (IEA, 2023).

Para llevar a cabo la evaluación, TIMSS se construye en torno a un marco de la evaluación elaborado en colaboración con los países participantes gracias al papel que desarrollan los “Coordinadores Nacionales de Investigación” (IEA, 2023). Del mismo modo, estos Coordinadores colaboran en la aplicación de las pruebas y su evaluación, en la elaboración de los informes y en la interpretación de los resultados obtenidos. La mayor parte de esta información básica para la elaboración de la prueba se presenta en lo que la IEA denomina la *Enciclopedia de TIMSS*, que proporciona abundantes datos sobre el contexto educativo de cada país participante y su sistema educativo.

El citado marco de la evaluación se basa en tres elementos relacionados con el currículo, siendo este su concepto principal. Estos tres elementos básicos son el currículo programado, el currículo aplicado y el currículo alcanzado. En la siguiente figura se puede observar el modelo planteado por la IEA (figura 1.3):

Figura 1.3

Modelo de currículo planteado en TIMSS



Nota. Extraído de *TIMSS 2019. Marcos de la evaluación*, (p. 4), Ministerio de Educación y Formación Profesional & Subdirección General de Atención al Ciudadano, Documentación y Publicaciones, (2018).

De acuerdo con lo planteado en su marco para la evaluación (Ministerio de Educación y Formación Profesional & Subdirección General de Atención al Ciudadano, Documentación y Publicaciones, 2018), cada uno de estos tres tipos de currículo se define de la siguiente manera:

1. Currículo programado: lo que se espera que los estudiantes deben aprender en relación con los contenidos y habilidades propuestos en el currículo a nivel nacional, social y educativo, en las dos disciplinas evaluadas y cómo debe estar organizado el sistema educativo para facilitarlos.
2. Currículo aplicado: la realidad de lo enseñado en las aulas, entendiéndolo por tanto como el qué, quién y cómo se enseña. Dirigido a un contexto familiar, escolar, del profesorado y de las propias aulas.
3. Currículo alcanzado: lo que los estudiantes han aprendido, su rendimiento, sus actitudes y su percepción sobre el aprendizaje en estas materias.

Las evaluaciones de los contenidos propiamente dichas se complementan con cuestionarios contextuales con los que se recoge información relevante relacionada con la escuela, el profesorado, el alumnado y sus respectivos hogares, obteniendo datos sobre algunos factores contextuales que podrían estar influyendo en el rendimiento educativo de los estudiantes.

En el caso de los países incluidos en este estudio (España, Irlanda y Portugal), todos ellos han participado en TIMSS en las ediciones de los años 1995, 2011, 2015 y 2019. Además, cabe indicar que la Comunidad Autónoma del País Vasco participó también en las ediciones celebradas en 2003 y 2007, a pesar de no haberlo hecho el conjunto del territorio español.

La última aplicación de TIMSS ha tenido lugar en el año 2023, estando prevista la publicación de los resultados en diciembre del año 2024 aproximadamente.

1.3.2. PIRLS – Estudio Internacional de Progreso en Comprensión Lectora

El Estudio Internacional de Progreso en Comprensión Lectora (PIRLS – *Programme for International Reading Literacy Study*) es un estudio de la IEA que propone evaluar la comprensión lectora de los estudiantes de 4º grado (4º de Educación Primaria en España). PIRLS se centra en la evaluación de lo que desde la IEA denominan como *las dos finalidades generales de la lectura* (IEA, 2024), que son obtener experiencia literaria y

adquirir y utilizar la información. Concretamente, se evalúan cuatro procesos dentro de los propósitos de la lectura:

1. Centrarse en información explícita y recuperarla.
2. Hacer inferencias directas.
3. Interpretar e integrar ideas e información.
4. Evaluar y criticar el contenido y los elementos textuales.

En este caso, la evaluación se lleva a cabo con una periodicidad quinquenal. Su primera edición tuvo lugar en el año 2001. España ha participado desde la edición celebrada en el año 2006, mientras que Irlanda y Portugal se incorporaron a esta evaluación a partir de 2011.

Al igual que en el caso anterior, se elabora un marco para la evaluación en colaboración con los Coordinadores Nacionales de Investigación y se encargan de la elaboración de la *Enciclopedia de PIRLS*.

La construcción de esta prueba es similar a TIMSS, complementando la propia evaluación del nivel competencial de los alumnos en lectura con la aplicación de una serie de cuestionarios de contexto que proporcionan una valiosa información sobre distintos factores que podrían influenciar ese rendimiento (IEA, 2024), dirigidos al centro, al profesorado, al estudiantado y a sus familias.

En el caso de PIRLS 2021 (última edición celebrada), la evaluación ha evolucionado hacia una evaluación completamente digital en línea con los avances en las tecnologías de la información y la comunicación y su aplicación al mundo educativo.

Este estudio, junto con TIMSS, conforman el núcleo de las pruebas de evaluación a gran escala planteadas por la IEA para medir el rendimiento de las tres competencias básicas de los estudiantes: matemáticas, ciencias y lectura.

1.3.3. PISA – Programa para la Evaluación Internacional de Alumnos

El Programa para la Evaluación Internacional de Alumnos (PISA – *Programme for International Student Assessment*) propuesto por la OCDE tiene como objetivo evaluar el nivel competencial del estudiantado al finalizar la etapa de escolarización obligatoria en torno a tres competencias clave: comprensión lectora, matemáticas y ciencias. Es decir, lo que los jóvenes de aproximadamente 15 años saben y son capaces de hacer en relación con dichas competencias. Frente a lo planteado en otras evaluaciones internacionales, PISA no evalúa los conocimientos curriculares de los estudiantes, “se centra en el

reconocimiento y valoración de las destrezas y conocimientos adquiridos por los alumnos al llegar a sus quince años” (OECD, 2006, p. 6).

Su primera aplicación data del año 2000 y se aplica con carácter trienal, profundizando en una de las tres competencias principales en cada edición. La primera edición (PISA 2000) se centró en la comprensión lectora, la segunda edición se centraba en la competencia matemática (PISA 2003), mientras que en la tercera (PISA 2006) se profundizaba en la competencia científica, reiniciándose el procedimiento en el año 2009 centrando de nuevo la evaluación en la comprensión lectora y así sucesivamente hasta la actualidad. Las competencias clave estudiadas en PISA se complementan con la evaluación de alguna otra competencia específica o innovadora. Algunas de ellas a lo largo de los años han sido la resolución de problemas, la competencia global o el pensamiento creativo. Cabría hacer una mención especial a la competencia financiera, incluida para su evaluación en el programa en el año 2012 y evaluada en las posteriores ediciones a partir de ese momento (Ministerio de Educación y Formación Profesional, 2023).

Pese a su carácter trienal, la pandemia causada por la COVID-19 hizo que la edición prevista para el año 2021 sufriera un retraso, posponiendo la recogida de datos hasta el año 2022. Los resultados se han publicado en el mes de diciembre de 2023. Estos últimos resultados indican que el rendimiento en estas tres competencias ha descendido considerablemente desde el inicio de PISA (OECD, 2023a).

Al tratarse de una prueba estandarizada de evaluación, PISA se construye con un gran rigor metodológico, obteniendo información sobre resultados cognitivos (pruebas competenciales) y no cognitivos (autoeficacia, actitudes hacia la educación, motivación) y sobre características contextuales tanto de los estudiantes como de los propios centros educativos. Para ello, diferentes comités se encargan del diseño de los instrumentos de recogida de información en forma de cuestionarios con características bien diferenciadas. Por un lado, se encuentran las pruebas competenciales, es decir, las pruebas propiamente dedicadas a evaluar el nivel competencial de los estudiantes. Por otro lado, se sitúan los cuestionarios de contexto, “destinados a explorar los antecedentes y circunstancias de los alumnos que realizan la evaluación” (OECD, 2006, p. 28).

En el diseño de las pruebas competenciales colaboran diversos agentes tratando de paliar las diferencias culturales, lingüísticas e incluso organizativas existentes entre los países participantes. Para ello, se establecen en un primer momento los marcos conceptuales sobre los que los comités elaboran el contenido de las pruebas aportando

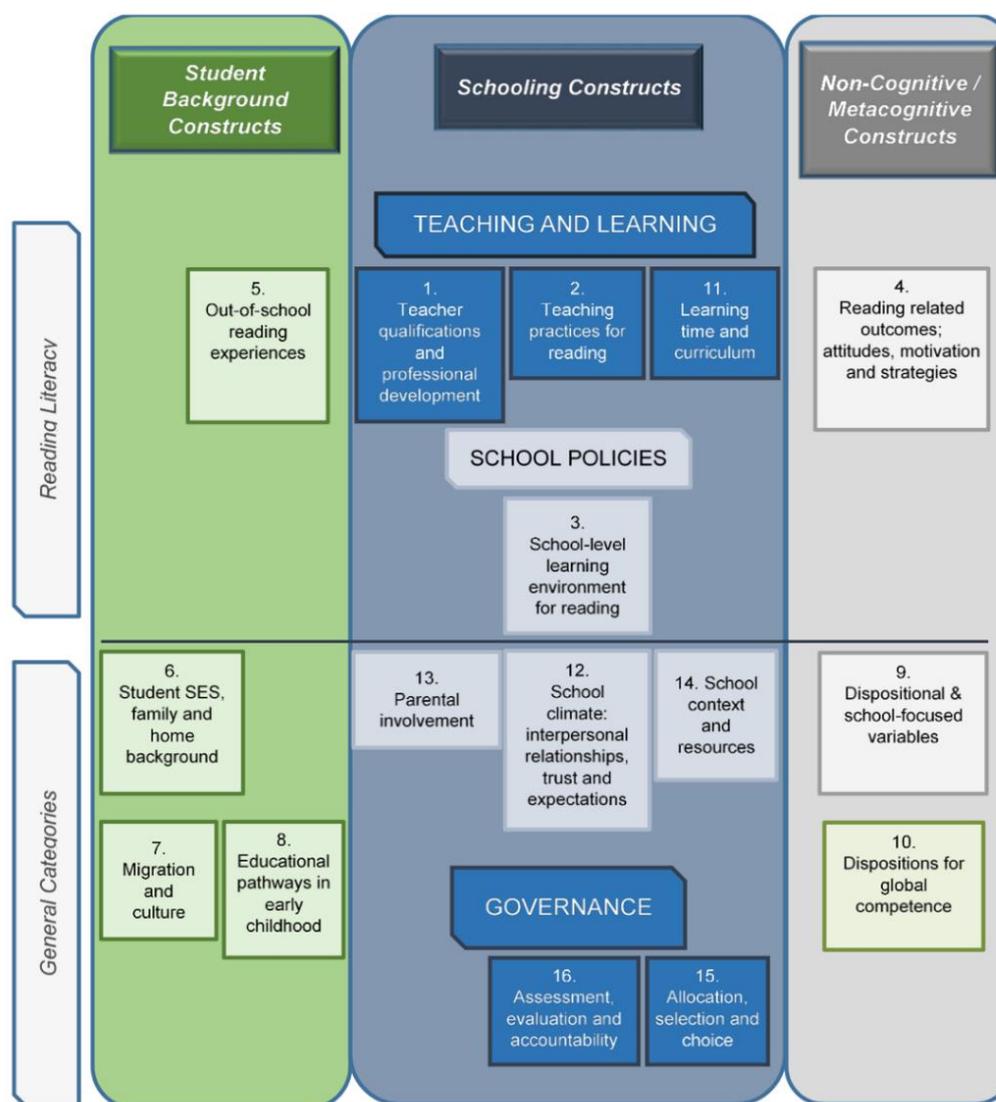
preguntas que, posteriormente, revisan distintos expertos garantizando que se abordan todos los dominios para cada una de las competencias evaluadas. Con carácter previo a la aplicación final de las pruebas, se lleva a cabo una prueba piloto en la que se testean las preguntas realizando las calibraciones oportunas. Una vez consensuada la estructura final de la prueba, esta es traducida convenientemente para su correcta aplicación asegurando “la integridad y la equivalencia lingüística de las distintas versiones” (Pajares et al., 2004, p. 11). Con ellas se pretende obtener la información referida a los resultados cognitivos de los estudiantes. Las pruebas se construyen con ítems de distinta índole según el tipo de respuesta: abierta o de respuesta múltiple.

Los expertos señalan como uno de los aspectos clave de las pruebas PISA la aplicación de los cuestionarios de contexto, ya que estos cuestionarios sirven para obtener información sobre las circunstancias familiares de los estudiantes (OECD, 2006), información contextual de los centros educativos, así como la información sobre los resultados no cognitivos citada anteriormente.

En la siguiente figura (Figura 1.4) se observa el marco utilizado para la construcción de los cuestionarios en PISA, aplicado desde el año 2012. En él se definen 16 módulos referidos a los constructos contextuales del estudiante, los constructos escolares y los no cognitivos y metacognitivos.

Figura 1.4

Módulos de los cuestionarios de PISA 2018



Nota. Extraído de *PISA 2018 Assessment and Analytical framework* (p. 220), por OECD, (2019).

La primera columna (*Student Background Constructs*) se refiere a los antecedentes familiares de los estudiantes en cuanto a nivel socioeconómico, migración y cultura, así como a la educación que han recibido hasta la fecha. Esta información la suelen proporcionar los estudiantes o sus familias.

La columna sobre constructos escolares (*Schooling Constructs*) se centra en los procesos educativos a nivel de sistema, centro y aula. Los ítems se relacionan con el contexto escolar y sus recursos, el clima escolar, las características del profesorado (formación, experiencia, etc.), entre otros. Las preguntas asociadas a estos constructos

son respondidas por los centros escolares, aunque algunas también pueden dirigirse a los estudiantes o a las familias.

La última columna (*Non-cognitive/Metacognitive Constructs*) se refiere a todos aquellos resultados relativos a constructos no cognitivos, bien específicamente relacionados con el dominio de la comprensión lectora o bien con el ámbito educativo en general.

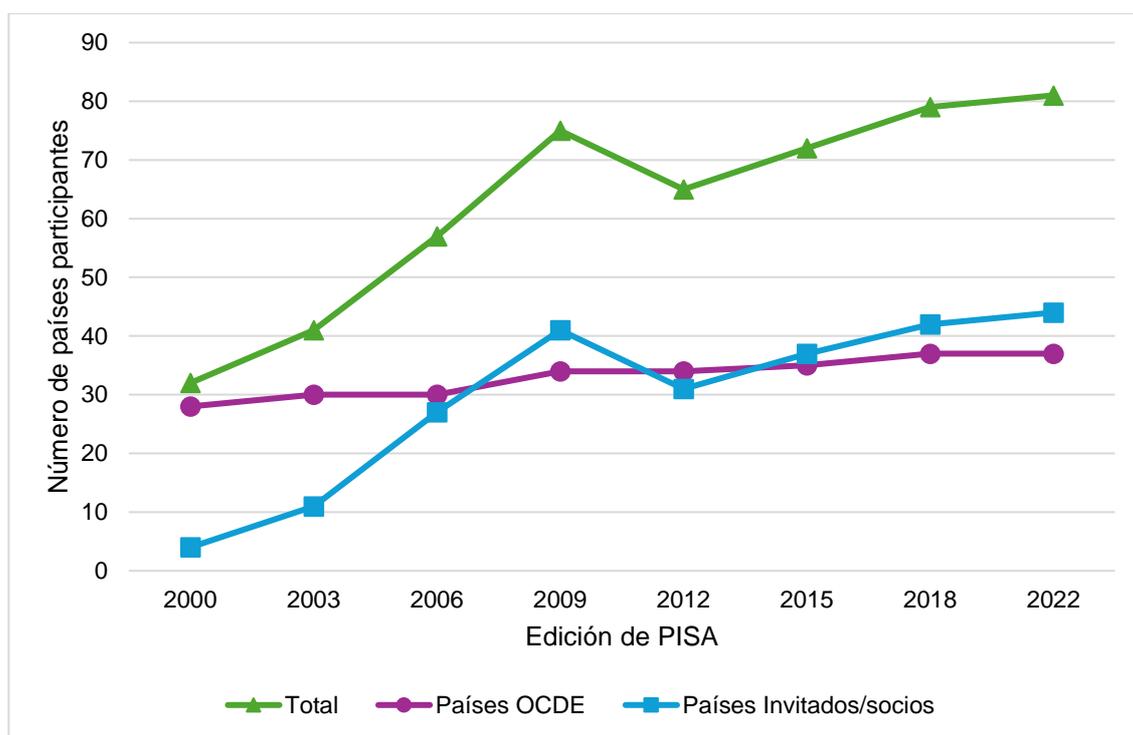
Con este marco se pretende que las conexiones entre los cuestionarios sean lo más transparentes posible, además de describir lo que se cubre conceptualmente en cada cuestionario, qué constructos se examinan a nivel de alumno y a nivel escolar, y quién responde a cada pregunta individual.

Tras haber presentado los instrumentos de recogida de información utilizados, es necesario conocer quiénes son los informantes; es decir, los agentes de los que se obtiene esa tan preciada información en PISA. Evidentemente el principal agente es el estudiantado, pero no el único. A parte de esto, los equipos directivos de los centros educativos también proporcionan una valiosa información contextual indicando cómo se organizan sus centros. En las últimas ediciones, esta información proporcionada por los directores de los centros educativos se ha complementado con otra muy relevante aportada por los padres de los alumnos, además de por los propios docentes de algunas de las economías y países participantes para conocer su formación y desarrollo profesional, sus prácticas docentes y su satisfacción laboral (OECD, 2019a), como es el caso de España.

En esta prueba de evaluación participan los países miembros de la OCDE y algunos otros interesados en conocer el nivel de adquisición de las competencias básicas de sus estudiantes independientemente de su pertenencia o no a este organismo internacional. Estos otros países participantes son denominados países invitados o socios. En la figura siguiente se puede observar la evolución en cuanto a los países participantes en PISA (Figura 1.5), comenzando por los 32 países que evaluaron el rendimiento educativo de sus estudiantes en PISA 2000 hasta las 81 economías y países participantes en la última edición (PISA 2022).

Figura 1.5

Países y economías participantes en las distintas ediciones de PISA



Como se observa en la figura anterior, el número de países y economías que han participado en PISA a lo largo de los años ha aumentado considerablemente, llegando a duplicar ampliamente la muestra inicial.

La aplicación de las pruebas, como en el caso de las otras ILSA presentadas, ha sufrido cambios desde sus inicios. Se partía de una aplicación de manera escrita, llegando en la actualidad hasta una evaluación desarrollada fundamentalmente mediante el empleo de ordenadores (desde el año 2015).

Los resultados se presentan en una escala con media 500 y desviación típica de 100 con diferentes niveles de logro. Esto supone que la mayor parte de los estudiantes, (aproximadamente dos tercios), obtienen puntuaciones entre 400 y 600 puntos.

A continuación, se presentan las características de la prueba PISA 2018, ya que los datos obtenidos tras su aplicación son los empleados para el análisis cuantitativo que fundamenta esta investigación.

1.3.3.1. *PISA 2018: marco conceptual y definición de las competencias*⁴

PISA 2018 se plantea teniendo la competencia en comprensión lectora como competencia principal, al igual que había sucedido en los años 2000 y 2009. Las tres competencias básicas evaluadas se complementaron con la evaluación de la denominada competencia global. Además, se pusieron a disposición de los países participantes otros tres cuestionarios opcionales a responder por los estudiantes: uno de ellos dedicado a conocer su familiaridad con el uso de las tecnologías de la información y la comunicación, otro para conocer la educación de los estudiantes hasta la fecha (buscando reconocer la existencia de interrupciones de cualquier tipo durante su escolarización, o si se están preparando para una futura carrera profesional y cómo lo están haciendo) y un tercero, distribuido por primera vez en PISA 2018, destinado a examinar el bienestar y la satisfacción con la vida de los estudiantes.

Al tratarse de una prueba cíclica, en cada edición se introducen cambios acordes con los cambios sociales del momento. Es por esto por lo que, dentro de los marcos conceptuales elaborados para el desarrollo de la prueba, se definen específicamente las distintas competencias. En PISA 2018 las tres competencias troncales se definen como:

Comprensión lectora: “Capacidad de un individuo para comprender, utilizar, evaluar, reflexionar y comprometerse con los textos con el fin de alcanzar sus objetivos, desarrollar sus conocimientos y su potencial, y participar en la sociedad” (OECD, 2019a, p. 28).

Competencia matemática:

Capacidad de un individuo para formular, emplear e interpretar las matemáticas en diversos contextos. Incluye el razonamiento matemático y el uso de conceptos, procedimientos, hechos y herramientas matemáticas para describir, explicar y predecir fenómenos. Ayuda a las personas a reconocer el papel que desempeñan las matemáticas en el mundo y a adoptar los juicios y decisiones fundamentados que necesitan los ciudadanos constructivos, comprometidos y reflexivos (OECD, 2019a, p. 75)

⁴En este apartado se describen las competencias principales de PISA 2018 a partir del documento *PISA 2018 Assessment and Analytical Framework* (OECD, 2019a). Por lo que, para evitar la redundancia repitiendo la misma referencia, se omitirán todas ellas exceptuando aquellos casos en los que se empleen citas textuales.

Competencia científica:

La capacidad de comprometerse con temas relacionados con la ciencia, y con las ideas de la ciencia, como un ciudadano reflexivo.

Por tanto, una persona científicamente alfabetizada está dispuesta a participar en un discurso razonado sobre la ciencia y la tecnología, lo que requiere las competencias de:

- Explicar fenómenos científicamente: reconocer, ofrecer y evaluar explicaciones de una serie de fenómenos naturales y tecnológicos.
- Evaluar y diseñar la investigación científica: describir y valorar investigaciones científicas y proponer formas de abordar cuestiones científicamente.
- Interpretar datos y pruebas científicamente: analizar y evaluar datos, afirmaciones y argumentos en una variedad de representaciones y extraer conclusiones científicas adecuadas. (OECD, 2019a, pp. 100-101)

Las definiciones anteriores van más allá de la demostración de la posesión de unos conocimientos específicos, y se refieren a las capacidades adquiridas por los estudiantes para aplicar con adecuación los conocimientos de los que disponen a la sociedad actual.

A continuación, se presentan las características básicas incluidas en los marcos conceptuales de estas tres competencias.

Competencia en comprensión lectora

La última edición en la que la comprensión lectora había sido la competencia principal en PISA se corresponde con la edición del año 2009. Desde ese momento, nuevas formas de lectura han revolucionado el mundo, especialmente la lectura en entornos digitales debido, en gran parte, a la cantidad de recursos disponibles (Burin et al., 2016; OECD, 2014). Por este motivo, se han introducido novedades en el marco conceptual de PISA 2018, donde de nuevo la comprensión lectora es la competencia principal, comenzando por la propia denominación de la competencia que ha evolucionado de “lectura” a “comprensión lectora”.

Para construir la evaluación de la comprensión lectora en esta edición se han utilizado diversos elementos, desde los procesos cognitivos implicados en la lectura hasta las distintas situaciones o contextos, pasando por el tipo de formato de texto presentado o el

tipo de texto propiamente dicho. Cada uno de estos elementos dispone de unas características concretas que se exponen a continuación.

Dentro de los procesos cognitivos se diferencian habilidades como la localización de información específica dentro de un texto, dividida en el acceso y la recuperación de la información y la búsqueda y selección de textos relevantes; la comprensión del texto, desde la representación literal del mismo hasta la construcción de una visión integrada sobre su significado; y la evaluación y reflexión sobre el texto, entendida como “la evaluación de su calidad y credibilidad y como la reflexión sobre su contenido y forma” (OECD, 2019a, p. 15).

Se podría determinar que para alcanzar la fluidez en comprensión lectora es necesario manejar diferentes procesos relacionados con el procesamiento del texto y con la aplicación de tareas independientes a la propia comprensión lectora como se ilustra en la siguiente figura (figura 1.6):

Figura 1.6

Procesos del marco conceptual de comprensión lectora en PISA 2018



Nota: Adaptado de *PISA 2018 Assessment and Analytical framework* (p. 33), por OECD, (2019).

Las situaciones de lectura se refieren al tipo de uso para el que se ha elaborado el texto, como por ejemplo un uso personal ligado al ocio, el uso público ligado a un interés más social, el uso laboral relacionado con el ámbito profesional o el uso educativo asociado a la formación.

En cuanto a los formatos de texto, en las pruebas competenciales se emplean textos de los siguientes tipos (tabla 1.1):

Tabla 1.1

Tipos de formato de texto en las pruebas PISA 2018

De fuente única	De fuente múltiple
Estáticos	Dinámicos
Continuos (oraciones y párrafos)	No continuos (listas, gráficos, etc.)
Textos mixtos	

Los tipos de formato de texto se emparejan de tal modo que se diferencia entre:

- Textos de fuente única y de fuente múltiple: se refieren a aquellos que provienen de una única referencia o de la combinación de varias, respectivamente.
- Textos estáticos y dinámicos: los primeros son aquellos construidos en base a una organización simple y con un menor número de herramientas de navegación, entendiéndose por herramientas de navegación por ejemplo la barra de desplazamiento en los entornos digitales o los hipervínculos, mientras que los textos dinámicos se refieren a los construidos de manera más compleja y que requieren de la utilización de diferentes herramientas de navegación para su comprensión.
- Textos continuos y no continuos: los textos continuos son aquellos construidos en torno a frases que se organizan a su vez en párrafos. En cambio, los textos no continuos se refieren a aquellos formados sobre listas que pueden aparecer de manera interrelacionada. Algunos ejemplos de textos no continuos son las listas, las tablas, gráficos, índices, etc.
- Textos mixtos: básicamente se refieren a aquellos textos en los que se combinan diferentes tipologías de las detalladas previamente, como por ejemplo un gráfico con su respectiva leyenda o una tabla con una explicación en prosa asociada.

Los textos utilizados según su propósito pueden ser de tipo descriptivo, narrativo, expositivo, argumentativo, instructivo o transaccional.

El rendimiento de los estudiantes puede diferir en función del tipo de situación de lectura planteada, ya que cada una de ellas se relaciona con una serie de tareas o procesos implicados en la lectura, por lo que durante la prueba se incluyen distintas opciones para su evaluación de modo que se abarque la mayor cantidad de combinaciones posible.

Competencia matemática

La evaluación de la competencia matemática sigue el marco conceptual desarrollado en ediciones anteriores, concretamente en el año 2012 donde esta competencia fue la principal.

Como en el caso de la competencia en comprensión lectora, la competencia matemática se basa en una serie de constructos, utilizando tres elementos básicos interrelacionados:

- Los procesos y las capacidades matemáticas básicas.
- La forma de organización del conocimiento del contenido matemático.
- Los contextos en los que los estudiantes se enfrentan a retos matemáticos.

El primer elemento se refiere a los procesos matemáticos, entendidos como lo que realizan los estudiantes para conectar el contexto del problema con las matemáticas y así poder resolverlo, así como las capacidades que subyacen a esos procesos. PISA diferencia entre tres tipos de procesos matemáticos: la formulación de las situaciones matemáticas, el empleo de los conceptos matemáticos y la interpretación, aplicación y evaluación de los resultados matemáticos.

La formulación de situaciones matemáticas se entiende como la capacidad de un sujeto de reconocer e identificar oportunidades para la utilización de las matemáticas y la provisión de una estructura matemática a un determinado problema para su resolución. En él pueden estar implicadas diferentes actividades como el reconocimiento de la estructura matemática, la identificación del problema, etc.

El empleo de los conceptos matemáticos está relacionado con la aplicación de los citados conceptos en la propia resolución de problemas formulados matemáticamente. Para ello, se llevan a cabo los procedimientos matemáticos necesarios como pueden ser la realización de cálculos aritméticos o la resolución de ecuaciones, entre otros. De nuevo este proceso matemático se asocia a unas actividades concretas como el uso de herramientas matemáticas o la elaboración de diagramas matemáticos.

Por último, la interpretación, aplicación y evaluación de los resultados matemáticos es entendida, en esencia, como la “capacidad de las personas para reflexionar sobre soluciones, resultados o conclusiones matemáticas e interpretarlos en el contexto de problemas de la vida real” (OECD, 2019a, p. 79). En este caso, las actividades relacionadas con este proceso matemático son, entre otras: interpretar un resultado matemático en el contexto del mundo real, explicar por qué un resultado tiene o no sentido

en el contexto de un determinado problema o criticar e identificar los límites del modelo utilizado para la resolución de un problema.

PISA ha definido siete habilidades matemáticas básicas en este marco conceptual: comunicación, matematización, representación, razonamiento y argumentación, diseño de estrategias para la resolución de problemas, utilización de lenguajes simbólicos, formales o técnicos y operaciones y uso de herramientas matemáticas. Estas siete capacidades matemáticas, se interrelacionan con los tres procesos matemáticos básicos definidos tal y como se explica en la tabla 1.2 (OECD, 2019a, p. 82):

Tabla 1.2

Relación entre los procesos y las habilidades matemáticas básicas en PISA

	Formulación matemática de situaciones	Empleo de conceptos, hechos, procedimientos y razonamiento matemáticos	Interpretar, aplicar y evaluar los resultados matemáticos
Comunicación	Leer, decodificar y dar sentido a afirmaciones, preguntas, tareas, objetos o imágenes para formular un modelo mental de la situación	Articular una solución, mostrar el trabajo necesario para llegar a una solución y/o resumir y presentar resultados matemáticos intermedios	Construir y comunicar explicaciones y argumentos en el contexto del problema
Matematización	Identificar las variables y estructuras matemáticas subyacentes en problemas del mundo real, y hacer suposiciones para poder utilizarlas	Utilizar la comprensión del contexto para guiar o acelerar el proceso de resolución matemática, por ejemplo, trabajando con un nivel de precisión adecuado al contexto	Comprender el alcance y los límites de una solución matemática que son consecuencia del modelo matemático empleado
Representación	Crear una representación matemática de información del mundo real	Comprender, relacionar y utilizar diversas representaciones al interactuar con un problema	Interpretar resultados matemáticos en una variedad de formatos en relación con una situación o uso; comparar o evaluar dos o más representaciones en relación con una situación
Razonamiento y argumentación	Explicar, defender o justificar la representación identificada o ideada de una situación del mundo real	Explicar, defender o justificar los procesos y procedimientos utilizados para determinar un resultado o solución matemática Conectar elementos de información para llegar a una solución matemática, hacer generalizaciones o crear un argumento de varios pasos	Reflexionar sobre soluciones matemáticas y crear explicaciones y argumentos que apoyen, refuten o califiquen una solución matemática a un problema contextualizado

	Formulación matemática de situaciones	Empleo de conceptos, hechos, procedimientos y razonamiento matemáticos	Interpretar, aplicar y evaluar los resultados matemáticos
Diseño de estrategias para la resolución de problemas	Seleccionar o elaborar un plan o estrategia para replantear matemáticamente problemas contextualizados	Activar mecanismos de control eficaces y sostenidos a lo largo de un procedimiento de varios pasos que conduzca a una solución, conclusión o generalización matemática	Idear y aplicar una estrategia para interpretar, evaluar y validar una solución matemática a un problema contextualizado
Uso de operaciones y lenguajes simbólicos, formales y técnicos	Utilizar variables, símbolos, diagramas y modelos estándar apropiados para representar un problema del mundo real utilizando un lenguaje simbólico/formal	Comprender y utilizar construcciones formales basadas en definiciones, reglas y sistemas formales, así como emplear algoritmos	Comprender la relación entre el contexto del problema y la representación de la solución matemática. Utilizar esta comprensión para ayudar a interpretar la solución en su contexto y calibrar la viabilidad y las posibles limitaciones de la solución
Uso de herramientas matemáticas	Utilizar herramientas matemáticas para reconocer estructuras matemáticas o representar relaciones matemáticas	Conocer y ser capaz de utilizar adecuadamente diversas herramientas que puedan ayudar en la aplicación de procesos y procedimientos para determinar soluciones matemáticas	Utilizar herramientas matemáticas para determinar la razonabilidad de una solución matemática y los límites y restricciones de dicha solución, teniendo en cuenta el contexto del problema

Nota: Adaptado de *PISA 2018 Assessment and Analytical framework* (p. 82), por OECD, (2019).

Para PISA 2018 se han establecido una serie de categorías entendidas como forma de organización del contenido matemático. En ellas se abarcan las distintas vertientes del currículo matemático que deben dominar los estudiantes para demostrar su alfabetización matemática. Estas categorías son: cambio y relaciones, espacio y forma, cantidades, e incertidumbre y datos.

La primera categoría se refiere a los contenidos relacionados con el cambio y las relaciones. Concretamente, se busca incluir en esta categoría la capacidad de los alumnos para percatarse de las relaciones existentes entre diferentes variables y cómo esas relaciones pueden variar según distintos cambios. En un sentido puramente matemático, esto supone “modelizar el cambio y las relaciones con funciones y ecuaciones adecuadas, así como crear, interpretar y traducir entre representaciones simbólicas y gráficas de las relaciones” (OECD, 2019a, p. 84).

La categoría de espacio y forma presenta una relación directa con el apartado dedicado en el currículo matemático al área de la geometría, pero aumentando sus límites hacia un carácter más aplicado al mundo real recurriendo al conocimiento de otras áreas como álgebra, medición o visualización espacial.

En tercer lugar, en la categoría de cantidades se incluyen todos los contenidos relacionados con el número en su más amplia acepción. Es por eso por lo que es considerada como una categoría omnipresente y esencial en la interacción con el mundo, componiéndola aspectos variados como la cuantificación de objetos, relaciones, situaciones o entidades, la comprensión de diversas representaciones de esas cuantificaciones y la valoración de las interpretaciones y argumentos basados en la cantidad. Para ello se hace necesaria la comprensión de las medidas, las magnitudes, los recuentos, etc. Esta categoría podría ser resumida en el conocimiento del número y las operaciones numéricas en diferentes contextos.

Por último, la categoría de incertidumbre y datos se relaciona con el reconocimiento de la variabilidad y su cuantificación, de la propia incertidumbre como un hecho en la ciencia y la tecnología y del error de medida en los procesos matemáticos. En este caso, se relacionaría esta categoría con las áreas del currículo dedicadas a la estadística y la probabilidad.

Del mismo modo que en lo propuesto en la evaluación de la comprensión lectora de los estudiantes, y de acuerdo con la variedad de contextos en los que se puede requerir del empleo del conocimiento matemático, para la evaluación de esta competencia se

determinan cuatro contextos (personal, ocupacional, social o científico) persiguiendo abarcar el máximo de posibilidades de los estudiantes.

Competencia científica

La última edición desarrollada de esta prueba de evaluación a gran escala, PISA 2022, toma como competencia principal la competencia científica (OECD, 2023a). Se adopta la definición propuesta en PISA 2018, ya que es de esta edición de la que se han extraído los datos para elaborar esta tesis doctoral. Según el marco conceptual propuesto en PISA 2018, una buena alfabetización científica necesita de la demostración del dominio de los distintos elementos interrelacionados que componen esta competencia. Estos elementos son tres: competencias, conocimiento y contextos.

Se comienza definiendo las tres competencias básicas de la ciencia. En primer lugar, se sitúa la competencia de explicar fenómenos de manera científica y sus implicaciones para la sociedad. En segundo lugar, se encuentra la competencia para evaluar y diseñar investigaciones científicas implicando los propios conocimientos previos y la comprensión de la investigación científica. En tercer lugar, se encuentra la interpretación de datos y pruebas de manera científica, donde aparece de nuevo la importancia del reconocimiento del fenómeno de la incertidumbre en la ciencia y la capacidad de juzgar si esos datos son adecuados y justifican las interpretaciones y deducciones derivadas.

Cada una de estas tres competencias básicas de la ciencia se relaciona con los distintos tipos de conocimiento científico, en los que se definen los conocimientos del contenido de la ciencia, los conocimientos procedimentales y los conocimientos epistémicos, estos últimos referidos a “la comprensión de los fundamentos de las prácticas habituales de la investigación científica”(OECD, 2019a, p. 99). Según lo planteado por PISA, la primera competencia básica de la ciencia (explicar fenómenos de manera científica) se relaciona mayoritariamente con los conocimientos del contenido científico; mientras que las dos competencias restantes requieren del dominio de los conocimientos procedimentales y epistémicos.

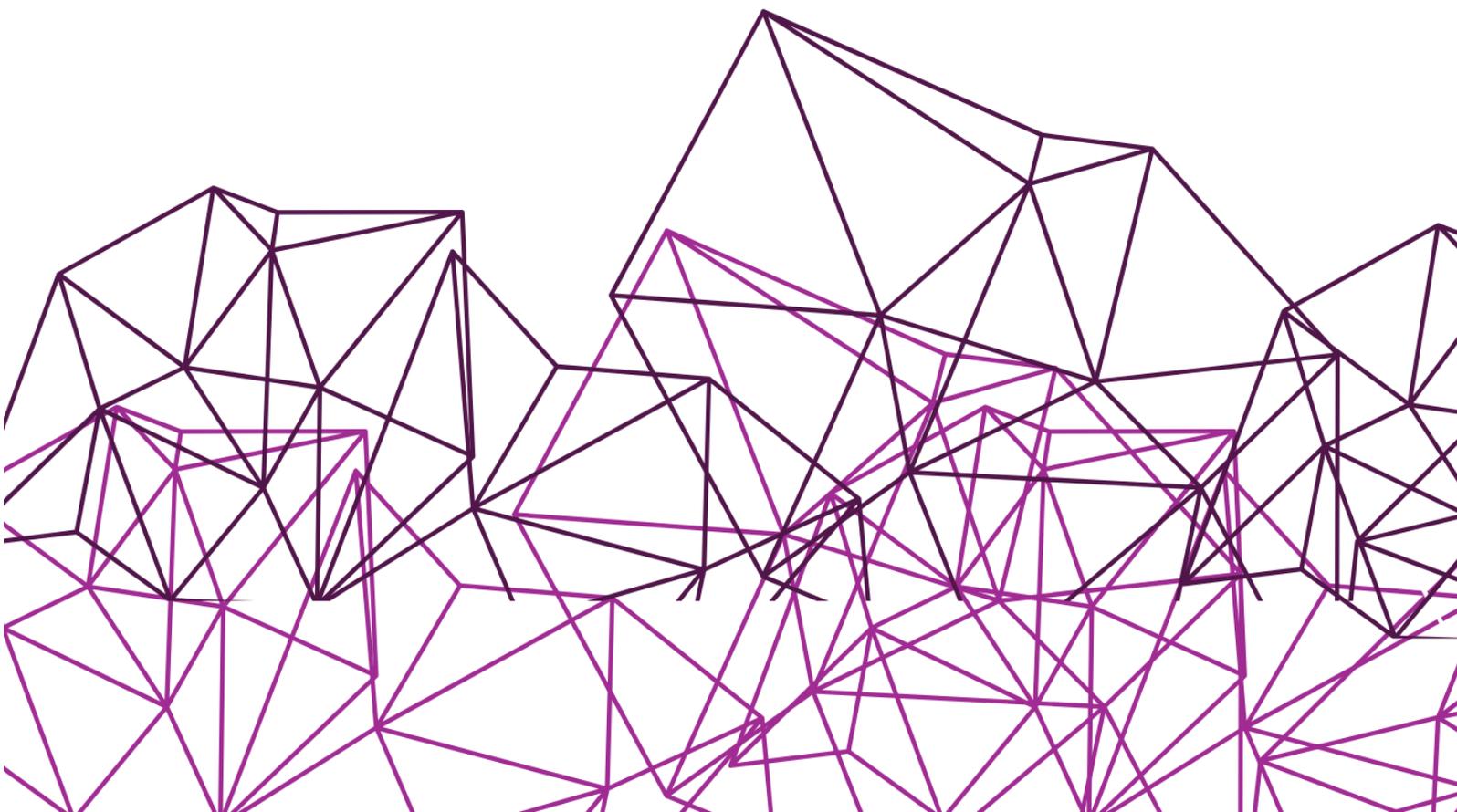
En cuanto a los contextos planteados para la evaluación de la alfabetización científica, se proponen situaciones científicas enmarcadas en tres ámbitos: personal, local/nacional o global. Del mismo modo, estos contextos se han categorizado en torno a cinco posibles aplicaciones de la ciencia y la tecnología, pudiendo consultar la interrelación entre contextos y aplicaciones en la tabla 1.3 (OECD, 2019a, p. 103):

Tabla 1.3*Relación entre los contextos y los temas de la prueba de alfabetización científica*

	Personal	Local/Nacional	Global
Salud y enfermedad	Mantenimiento de la salud, accidentes, nutrición	Control de enfermedades, opciones alimentarias, salud comunitaria	Epidemias, propagación de enfermedades infecciosas
Recursos naturales	Consumo personal de materiales y energía	Mantenimiento de las poblaciones humanas, calidad de vida, seguridad, producción y distribución de comida, suministros de energía	Sistemas naturales renovables y no renovables, crecimiento demográfico, uso sostenible de las especies
Calidad medioambiental	Acciones respetuosas con el medioambiente, uso y desecho de materiales y dispositivos	Distribución de la población, desecho de residuos, impacto medioambiental	Biodiversidad, sostenibilidad ecológica, control de la polución, producción y pérdida del suelo/biomasa
Riesgos	Evaluación de riesgos de los estilos de vida	Cambios rápidos (por ejemplo, terremotos, condiciones meteorológicas adversas), cambios lentos y progresivos (por ejemplo, erosión costera, sedimentación), evaluación de riesgos	Cambio climático, impacto de las comunicaciones modernas
Fronteras de la ciencia y la tecnología	Aspectos científicos de las aficiones, tecnología personal, música y actividades deportivas	Nuevos materiales, dispositivos y procesos, modificaciones genéticas, tecnología de la salud, transporte	Extinción de las especies, exploración del espacio, origen y estructura del universo

Nota: Adaptado de *PISA 2018 Assessment and Analytical framework* (p. 103), por OECD, (2019).

2. REVISIÓN SISTEMÁTICA DE LA LITERATURA: PRUEBAS DE EVALUACIÓN A GRAN ESCALA, EFICACIA ESCOLAR Y RENDIMIENTO ACADÉMICO



Revisión Sistemática de la Literatura: Pruebas de evaluación a gran escala, eficacia escolar y rendimiento académico

En el marco de la investigación planteada y con el objetivo de conocer los estudios, sobre todo de carácter secundario, realizados sobre las pruebas de evaluación internacional a gran escala (ILSA) y su relación con la eficacia escolar y el rendimiento académico, se realizó una revisión sistemática de la literatura. Concretamente esta revisión se centra en tres de estas pruebas de evaluación: PISA, TIMMS y PIRLS, tratando de comprender cuáles son los objetivos planteados en estos estudios, así como la finalidad y el uso que se ha dado a los resultados obtenidos en cada caso y las principales limitaciones detectadas.

Por lo tanto, el principal objetivo de la revisión se ha centrado en conocer la evolución y las características de los análisis secundarios de las tres ILSA mencionadas, así como su relación con la eficacia escolar y el rendimiento académico, tratando así de valorar las posibles mejoras a la hora de realizar estudios similares, así como definir nuevas líneas de investigación a modo de prospectiva en este campo.

Para ello, se ha optado por la combinación de dos metodologías de revisión que aportan un cierto valor añadido a esta investigación y permiten la extracción de información de manera precisa y planificada facilitando la replicación de la revisión planteada. Estas dos metodologías son el *mapping* o mapeo (Kitchenham et al., 2011; Torres-Carrion et al., 2018) y la revisión sistemática de la literatura (*Systematic Literature Review* o SLR) (Kitchenham, 2004).

Con la primera de las metodologías (*mapping*), y de acuerdo con lo propuesto por Kitchenham et al. (2011), se pretende proporcionar una visión general sobre el tema abordado. Mientras que la SLR se emplea con el fin de recopilar y sintetizar información de relevancia de manera rigurosa que sirva para comprender en profundidad distintos aspectos relacionados con los estudios secundarios basados en ILSA. Para la realización de esta revisión sistemática de la literatura se siguieron las directrices planteadas en la Declaración PRISMA 2020 (Page et al., 2021) otorgándole por tanto ese carácter sistematizado al proceso.

En función de la metodología empleada, se han formulado una serie de preguntas de investigación a resolver gracias a la aplicación de un proceso organizado en el que se registra cada una de las fases a seguir. Se comienza la revisión de la literatura por las

preguntas de investigación planteadas para el *mapping* (concretadas más adelante en esta sección), teniendo en cuenta que aportan mayor generalidad y una visión universal del tipo de investigaciones realizadas hasta el momento. En cuanto a la SLR, se plantean una serie de preguntas de investigación más específicas sobre el contenido y la construcción de las investigaciones en este campo.

La búsqueda de información se ha llevado a cabo en tres repositorios: Web of Science, Scopus y ERIC y se centra en los estudios que proponen análisis secundarios de datos procedentes de las tres evaluaciones a gran escala seleccionadas que se hayan desarrollado a partir del año 2000. El objetivo principal debe ser el análisis de cuestiones relacionadas con los factores predictores del rendimiento académico y/o la eficacia escolar.

Al no haber históricamente una medida única para la eficacia escolar, en los diferentes estudios los investigadores adoptan como variable criterio mayoritariamente el rendimiento académico. Es por este motivo por el que es este concepto sobre el que se ha construido el concepto de escuelas eficaces que, en el caso de la investigación planteada en esta tesis doctoral, se refiere a las escuelas que obtienen rendimientos por encima de lo esperado teniendo en cuenta sus características contextuales.

Esta sección se construye en torno a tres apartados: metodología, resultados y conclusiones. En el primero de ellos se describen las dos metodologías de investigación empleadas, así como el proceso de investigación desarrollado desde el planteamiento de las preguntas de investigación hasta la selección final de artículos para esta revisión de la literatura. A continuación, se presentan los resultados que responden a las preguntas de investigación formuladas, tanto para el *mapping* como para la SLR. Finalmente, se incluye un análisis de las semejanzas, diferencias y limitaciones de los estudios seleccionados a modo de conclusión.

2.1. Metodología empleada en la revisión de la literatura

Según lo mencionado, se comienza explicando los pasos a seguir para la correcta aplicación de cada una de las metodologías utilizadas en este proceso de búsqueda de información.

2.1.1. Preguntas de investigación

La combinación de las metodologías de *mapping* y SLR hace necesaria la formulación de diferentes preguntas de investigación a responder con cada una de ellas. Por tanto, en primer lugar, se definen dichas preguntas y se concreta la cadena de búsqueda.

Como se aplican dos metodologías diferentes para la obtención de información, se detallan a continuación las preguntas de investigación del *mapping* (M) y las preguntas de investigación de la SLR (PI).

Se comienza esta revisión de la literatura empleando la metodología de mapeo de la literatura definiendo las preguntas más globales (M) en busca de obtener esa información general sobre la materia:

- M1: ¿Cuántos estudios se han publicado en revistas científicas relacionando alguna prueba estandarizada (PISA, TIMMS o PIRLS) con la eficacia escolar?
- M2: ¿Cuál ha sido la evolución temporal de las publicaciones en este tema? ¿Cuándo ha experimentado su mayor auge la investigación en esta temática?
- M3: ¿Quiénes son los autores de referencia en este campo?
- M4: ¿Qué países realizan más estudios de este tipo?
- M5: ¿En qué revistas científicas se publican más estudios relacionados con las pruebas a gran escala y la eficacia escolar?
- M6: ¿Cuáles son las metodologías más empleadas para el análisis de datos secundario de las pruebas internacionales de evaluación a gran escala?

En cuanto a las PI, referidas al contenido y a la construcción de las investigaciones, se plantearon otras seis cuestiones:

- PI1: ¿Cuáles son las variables predictoras más utilizadas en la investigación sobre rendimiento y eficacia escolar?
- PI2: ¿Se ha centrado la investigación basada en los resultados de las pruebas a gran escala en el rendimiento académico o en la eficacia escolar?
- PI3: ¿Los autores tienen en cuenta la estructura jerárquica de los datos de las pruebas a gran escala (estudiante, aula, escuela, región, país, etc.) en sus análisis?
- PI4: ¿Se trabaja con los valores plausibles y los pesos muestrales⁵ de las diferentes jerarquías (alumnos y escuelas) en la estimación de parámetros?

⁵ Consultar capítulo 4 para más información sobre estos elementos.

- PI5: ¿Se detectan muchas limitaciones a este tipo de estudios? ¿Cuáles son las más comunes?
- PI6: Generalmente, ¿cuáles son las implicaciones que tienen en la sociedad los estudios basados en pruebas de evaluación a gran escala?

2.1.2. Cadena de búsqueda

Una vez determinadas las preguntas de investigación, se seleccionan las bases de datos en las que realizar la búsqueda de información. Debido al carácter puramente educativo de la investigación, y teniendo en cuenta las características y posibilidades de los diferentes buscadores, se seleccionan tres grandes repositorios para la búsqueda: *Web of Science (WoS)*, Scopus y ERIC, la última de ellas por su marcado carácter educativo.

Las características de estas tres bases de datos permitieron acotar la búsqueda a artículos científicos que hubieran sido publicados a partir del año 2000, considerando estos dos criterios estrictamente necesarios por dos razones: la primera, los artículos son sometidos a la evaluación por pares con carácter previo a su publicación y la segunda, en el año 2000 tuvo lugar la primera edición de PISA, prueba de evaluación a gran escala sobre la que se construye la investigación presentada en esta tesis doctoral.

En base a las particularidades de cada uno de los repositorios escogidos y tras la realización de una primera metabúsqueda en la que se probaron diferentes combinaciones a modo de cadena de búsqueda que incluyeran estudios basados en las tres grandes ILSA seleccionadas (PISA, TIMMS y PIRLS), se establecieron tres cadenas de búsqueda diferentes, una para cada una de las bases de datos. Estas cadenas de búsqueda son:

- WoS → *(PISA OR TIMMS OR PIRLS) (Topic) and "school effectiveness" (Topic)*
- Scopus → *TITLE-ABS-KEY (pisa OR timms OR pirls) AND "school effectiveness"*
- ERIC → *abstract:(PISA or TIMMS or PIRLS) AND "school effectiveness"*

2.1.3. Criterios de elegibilidad

Posteriormente, se establecen los criterios de selección para determinar la inclusión o exclusión de los diferentes recursos en esta revisión de la literatura según su relevancia para la investigación. Concretamente los siete criterios de inclusión (CI) y los cuatro criterios de exclusión (CE) definidos son:

- CII: La publicación está relacionada con la temática investigada (eficacia escolar, rendimiento académico o pruebas de evaluación a gran escala).

- CI2: El artículo está en castellano o inglés.
- CI3: Se utilizan los datos de las pruebas internacionales a gran escala como PISA, TIMMS o PIRLS.
- CI4: Se hace referencia a los factores predictores del rendimiento académico o la eficacia escolar.
- CI5: Se trata de artículos empíricos sobre la eficacia escolar o el rendimiento académico.
- CI6: La variable criterio del estudio es el rendimiento académico o a la eficacia escolar.
- CI7: La metodología empleada es de corte cuantitativo.
- CE1: La temática no tiene relación con nuestro tema de investigación.
- CE2: El artículo está en algún idioma diferente del castellano o el inglés.
- CE3: Los datos en los que se basa el artículo no pertenecen a pruebas estandarizadas internacionales.
- CE4: Se trata de un artículo meramente teórico.

En este caso, se determinó la necesidad de cumplimiento de todos los criterios de inclusión en su conjunto para poder acceder a la fase de análisis en la que se revisaron los documentos. En el caso de los criterios de exclusión, bastó con que se admitiera al menos uno de ellos para ser descartado del análisis posterior.

De manera complementaria a este primer filtrado de datos, se establecen una serie de criterios de calidad de las publicaciones. Estos criterios de calidad se basan en los delimitados en investigaciones similares a la que se presenta (Gamazo, 2019), pretendiendo dar respuesta a 10 preguntas con tres opciones de respuesta cada una de ellas en función de su nivel de cumplimiento: 0 si no se cumplía, 0.5 si se cumplía parcialmente y 1 si el criterio se cumplía en su totalidad. Para su valoración, se elaboró una rúbrica (Tabla 2.1) combinando las diferentes preguntas y respuestas para así facilitar la aplicación de estos criterios a los documentos seleccionados, estableciendo la necesidad de alcanzar al menos una nota de 8.5 para ser considerados para la fase final de extracción de información y recapitulación de los resultados.

Tabla 2.1

Criterios de calidad de las publicaciones

Cuestión	No	Parcial	Sí
1. ¿Existen preguntas de investigación cualitativas y cuantitativas claras (u objetivos)?	No hay objetivos/preguntas de investigación claros, ni se intenta explicarlos en el texto.	Los objetivos/preguntas no se especifican como tales, pero se deducen claramente del texto	Los objetivos/preguntas están claramente identificados o etiquetados como tales en el texto
2. ¿Permiten los datos recogidos al autor o autores abordar la pregunta de investigación (objetivo)?	Los datos recogidos son insuficientes para abordar los objetivos/preguntas	Los datos recogidos permiten responder parcialmente a los objetivos/preguntas	Los datos recogidos permiten al autor o autores responder con éxito a todos los objetivos/preguntas
3. ¿Utiliza una muestra completa de PISA o de las otras pruebas de evaluación a gran escala incluidas en este análisis (es decir, un muestreo realizado por expertos)?	Utiliza una submuestra ad hoc y no está claro si la muestra es representativa de la población estudiada	Utiliza una submuestra ad hoc, pero su composición y representatividad se explican claramente	Utiliza una muestra realizada originalmente por los expertos de PISA
4. ¿Son adecuadas las mediciones (origen claro, o validez conocida, o instrumento estándar) y están claramente definidas?	El origen de las variables no es claro/válido/fiable, y/o las variables no están claramente definidas	El origen de las variables es claro, válido y fiable, pero su definición carece de claridad	El origen de las variables es claro, válido y fiable y están claramente definidas
5. ¿Se explican claramente los métodos?	El método no se explica claramente en el texto	El método se explica un poco en el texto	El método se explica claramente en el texto
6. ¿Son los métodos utilizados adecuados para responder a las preguntas/objetivos?	Los métodos no arrojan los resultados necesarios para responder a las preguntas de la investigación	Los métodos permiten al autor o autores responder a las preguntas de la investigación, pero no son los mejores métodos para hacerlo	Los métodos utilizados son los más adecuados para responder a las preguntas de investigación planteadas por los autores

Cuestión	No	Parcial	Sí
7. ¿Se responden todas las preguntas de investigación?	Al menos dos preguntas de investigación no tienen respuesta	Se responde a todas las preguntas de la investigación menos a una	Se responde a todas las preguntas de la investigación
8. ¿Se basan las conclusiones en los datos?	Las conclusiones no se basan en los datos presentados en el documento	Las conclusiones se basan en cierta medida en los resultados obtenidos, pero el autor o autores hacen afirmaciones que no se basan en los datos	Todas las conclusiones se basan en los resultados presentados en el documento
9. ¿Se explica el uso de valores plausibles?	No se menciona el uso de valores plausibles/No hay uso de valores plausibles/Hay un uso incorrecto de valores plausibles	El uso de valores plausibles se menciona, pero no se explica claramente	El uso de valores plausibles es adecuado y está correctamente explicado
10. ¿En qué consiste el uso de las ponderaciones muestrales explicadas?	No se menciona el uso de las ponderaciones muestrales/No hay uso de las ponderaciones muestrales/Hay un uso incorrecto de las ponderaciones muestrales	Se menciona el uso de pesos de muestreo, pero no se explica claramente	El uso de las ponderaciones muestrales es adecuado y está correctamente explicado

2.1.4. Procedimiento

La búsqueda y selección de los documentos se realizó en torno a tres fases. La primera consistió en aplicar las cadenas de búsqueda en los repositorios, posteriormente, se incluyeron todos los registros en un documento de Excel para facilitar la gestión de la información.

Seleccionados los artículos, se eliminaron aquellos que estaban duplicados. A continuación, se aplicaron los criterios de elegibilidad (detallados en la sección 2.1.3 de este mismo capítulo), diferenciando dos momentos:

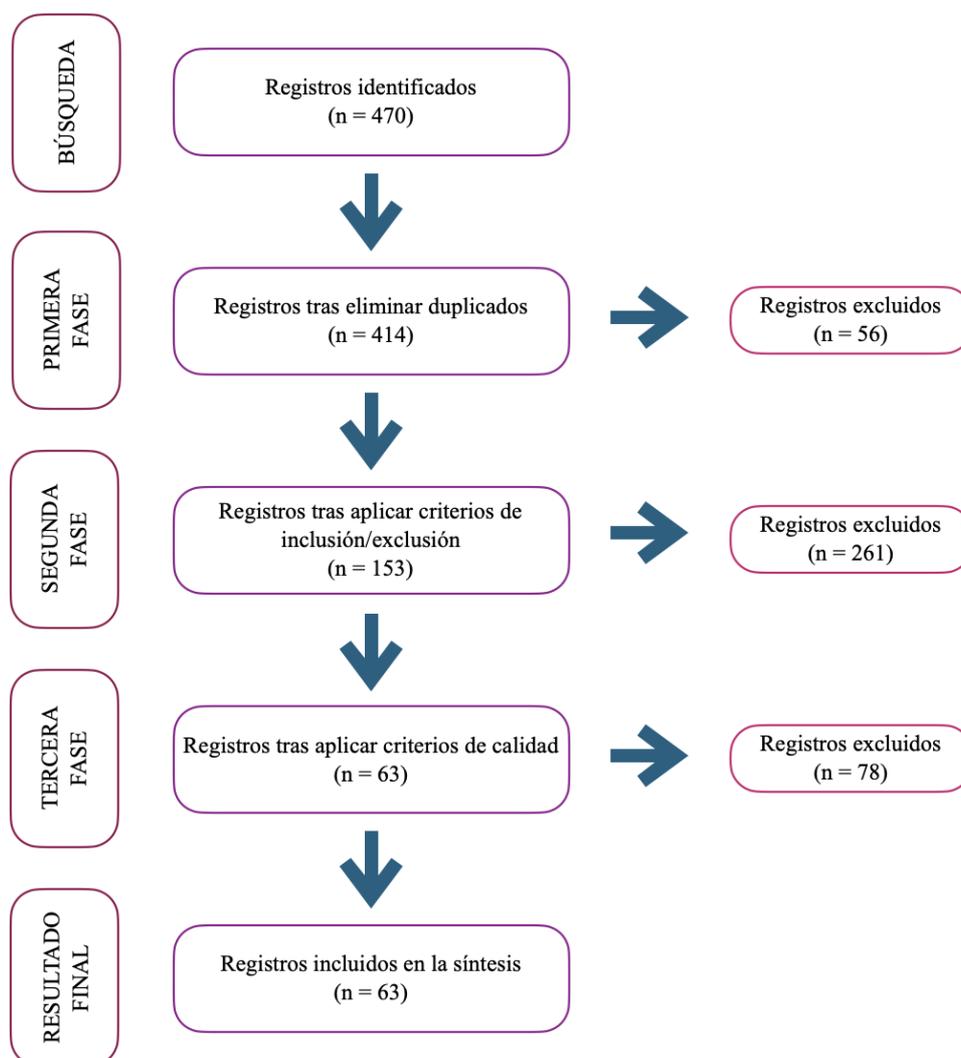
1. Aplicación de los criterios de inclusión y exclusión al título, resumen y palabras clave de los diferentes artículos.
2. Aplicación de los criterios de calidad al texto completo de los artículos que hayan superado el filtro anterior.

Tras la aplicación de los criterios de calidad se procede con el proceso de extracción de información con el ánimo de dar respuesta a las diferentes preguntas de investigación planteadas, tanto del mapeo como de la revisión sistemática de la literatura.

El número total de registros tras la búsqueda en los tres repositorios fue de 470 artículos que, tras eliminar los duplicados, se redujeron hasta 414 (382 de Scopus, 24 de ERIC y 382 de WoS). Aplicando los criterios de inclusión y exclusión la muestra disminuyó de 414 artículos a 153. Estos 153 artículos fueron revisados a texto completo aplicando la rúbrica con los criterios de calidad (tabla 2.1), de los cuales superaron la nota de corte establecida un total de 63 artículos que son los que componen finalmente la muestra para proceder a la extracción de datos necesaria para poder responder adecuadamente a las preguntas de investigación propuestas. En la figura incluida a continuación (figura 2.1) se pueden consultar los resultados de este procedimiento de acuerdo con las indicaciones de la Declaración PRISMA (Page et al., 2021):

Figura 2.1

Diagrama de flujo con el proceso de búsqueda y selección de recursos



2.2. Resultados del mapeo y de la revisión sistemática de la literatura

A continuación, se presenta la respuesta a las preguntas de investigación planteadas para el *mapping* y, posteriormente, se contestan las preguntas de investigación de la SLR.

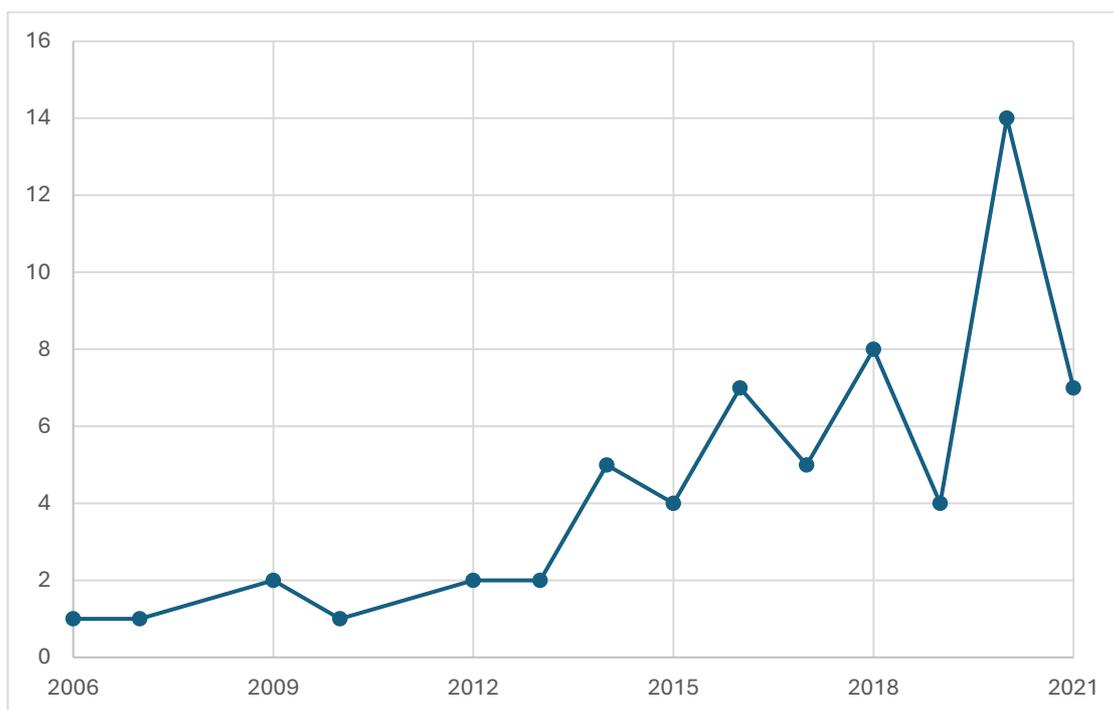
2.2.1. Resultados del estudio de *mapping*

La respuesta a la primera pregunta planteada para el *mapping* la tenemos con el número total de artículos incluidos en la revisión final, 63.

Para responder a la segunda pregunta se analizaron el número de publicaciones seleccionadas por año, obteniendo los siguientes resultados (figura 2.2):

Figura 2.2

Número de publicaciones por año



Desde el inicio de la investigación en esta temática se ha observado, a grandes rasgos, una tendencia creciente en la realización de estudios secundarios basados en los datos procedentes de las pruebas de evaluación internacional a gran escala (M2). Aplicando los criterios de elegibilidad, la primera publicación relevante fue la elaborada por Ruiz de Miguel y Castro Morera (2006) en el año 2006, pero no ha sido hasta el año 2020 cuando se hace visible el auge de esta corriente de investigación, llegando hasta 14 publicaciones de relevancia. Esta primera publicación se centraba en el análisis de los factores relacionados con la eficacia escolar en matemáticas en una muestra compuesta por diferentes países participantes en PISA 2003, primer año en que la competencia matemática fue la competencia principal (Ruiz de Miguel & Castro Morera, 2006). En cuanto a la publicación más reciente, nos encontramos con la investigación planteada por Pomianowicz (2021), en la que la autora estudia las diferencias en el rendimiento académico en comprensión lectora en función del índice de inmigración de los diferentes países participantes en PISA 2018, siendo la comprensión lectora la competencia principal de esa edición de PISA.

De entre los diferentes autores incluidos en la muestra final (n=133), destaca Martínez-Abad por la cantidad de publicaciones sobre el tema, con 4 artículos. En la

mayor parte de las publicaciones ha colaborado con Gamazo, situándose esta autora en segundo lugar junto con otros investigadores. Dos de las tres publicaciones en colaboración entre estos autores destacan por el empleo de la metodología de minería de datos (*data mining*) para estudiar las variables relacionadas con el rendimiento académico y la eficacia escolar (Gamazo & Martínez-Abad, 2020; Martínez-Abad et al., 2020). La tercera publicación emplea en cambio la regresión logística binaria para el estudio de la relación entre la eficacia escolar y diferentes variables (Gamazo et al., 2018).

Cabe destacar también la colaboración entre otros tres autores: Van Damme, Van Den Noortgate y Gielen, profesores de la universidad belga de KU Leuven que presentaron tres artículos en los que investigaban: los factores asociados al rendimiento académico en matemáticas de los países miembros de la OCDE que habían participado en PISA 2003 (Liu et al., 2014); la influencia del clima disciplinario en el rendimiento en lectura en los países participantes en PISA 2009 (Ning et al., 2015); y las diferencias en el rendimiento en lectura en las regiones belgas de nuevo en base a datos procedentes de PISA 2009 (Ning et al., 2016).

Por último, habría que mencionar los trabajos realizados por Ruiz de Miguel, otra autora que destaca por sus tres publicaciones seleccionadas, dos de ellas centradas en el estudio de los factores asociados al rendimiento académico en alguna de las competencias principales de PISA mediante modelos jerárquico lineales (Blanco Blanco et al., 2014; Ruiz de Miguel, 2009) y la restante dedicada al estudio del rendimiento académico en matemáticas en PISA 2003 con tres niveles de análisis: estudiante, escuela y país (Ruiz de Miguel & Castro Morera, 2006).

En la tabla 2.2 se refleja el número de publicaciones por autor, detallando aquellos autores con al menos dos publicaciones seleccionadas:

Tabla 2.2

Número de publicaciones por autor

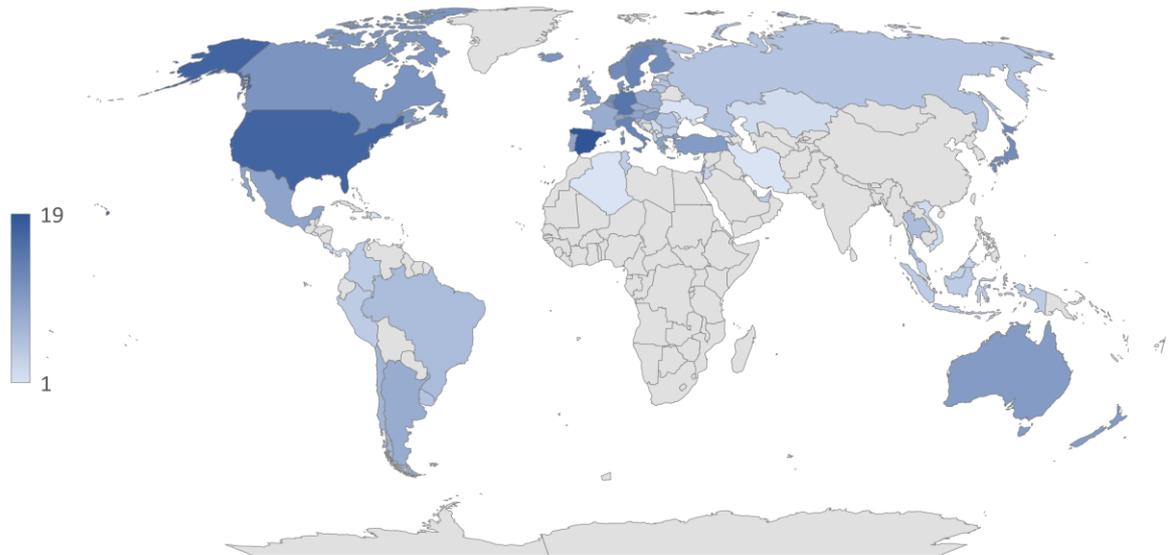
Autor	Número de publicaciones
Martínez-Abad, F.	4
Gamazo, A.	3
Van Damme, J.	3
Van Den Noortgate, W.	3
Gielen, S.	3

Autor	Número de publicaciones
Ruiz de Miguel, C.	3
Agasisti, T.	2
Costa, P.	2
Formichella, M. M.	2
Johansson, S.	2
Krüger, N.	2
Longobardi, S.	2
Myrberg, E.	2
Ning, B.	2
Rodríguez-Conde, M. J.	2
Shen, J.	2
Tan, C. Y.	2
Wu, H.	2

Para dar respuesta a la tercera pregunta planteada, se analiza la muestra incluida en cada uno de los artículos seleccionados, determinando finalmente el número de publicaciones por país, entendiéndolos como los estudios realizados tanto a nivel individual como conjunto. En el gráfico presentado, se han incluido los datos de aquellos países que aparecieran en al menos dos de los 63 artículos incluidos en la síntesis. Básicamente, se comprobó en cuántos estudios se incluía información referida a cada país (figura 2.3):

Figura 2.3

Número de publicaciones por país



Analizando el mapa coroplético anterior, se observa que los países más estudiados han sido España (20) y Estados Unidos (17), seguidos por Suecia (16) y Alemania (15). Esto supone que las regiones en las que más estudios relacionados se han llevado a cabo son Europa y Norteamérica, destacando la baja participación asiática y africana.

De los 63 artículos que componen la selección final, en 32 de ellos se realizan análisis en los que la muestra la componen datos de diferentes países tratando de establecer comparaciones entre ellos a partir de los resultados de las ILSA. Las investigaciones restantes ponen el foco de atención en un único país, siendo de nuevo el más estudiado España, con 8 publicaciones, seguido en este caso por Argentina y Estados Unidos con 4 investigaciones respectivamente (tabla 2.3).

Tabla 2.3*Estudios individuales por país y referencias bibliográficas*

País	Referencia
España	Blanco, A., 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. <i>Revista de educación</i> , 365, 122-149.
	Burger, K., & Walk, M. (2016). Can children break the cycle of disadvantage? Structure and agency in the transmission of education across generations. <i>Social Psychology of Education</i> , 19(4), 695-713. https://doi.org/10.1007/s11218-016-9361-y
	Doncel Abad, D., & Cabrera Álvarez, P. (2020). Comunidades Autónomas bilingües, identidades y desempeño educativo según PISA 2015. <i>Revista de educación</i> , 387, 163-188. https://doi.org/10.4438/1988-592X-RE-2020-387-443
	Gamazo, A., Martínez-Abad, F., Olmos-Migueláñez, S., & Rodríguez-Conde, M. J. (2018). Evaluación de factores relacionados con la eficacia escolar en PISA 2015. Un análisis multinivel. <i>Revista de educación</i> , 379, 56-84.
	Gil-Flores, J. G., & García-Gómez, S. (2017). Importancia de la actuación docente frente a la política educativa regional en la explicación del rendimiento en PISA. <i>Revista de educación</i> , 378, 52-77.
	Gómez-Fernández, N., & Mediavilla, M. (2021). Exploring the relationship between Information and Communication Technologies (ICT) and academic performance: A multilevel analysis for Spain. <i>Socio-Economic Planning Sciences</i> , 77, 101009. https://doi.org/10.1016/j.seps.2021.101009
	Martínez-Abad, F. (2019). Identification of Factors Associated With School Effectiveness With Data Mining Techniques: Testing a New Approach. <i>Frontiers in Psychology</i> , 10, 2583. https://doi.org/10.3389/fpsyg.2019.02583
Martínez-Abad, F., Gamazo, A., & Rodríguez-Conde, M.-J. (2020). Educational Data Mining: Identification of factors associated with school effectiveness in PISA assessment. <i>Studies in Educational Evaluation</i> , 66, 100875. https://doi.org/10.1016/j.stueduc.2020.100875	
Estados Unidos	Li, H. (2016). How is formative assessment related to students' reading achievement? Findings from PISA 2009. <i>Assessment in Education: Principles, Policy & Practice</i> , 23(4), 473-494. https://doi.org/10.1080/0969594X.2016.1139543
	Liu, Y. (2021). Distributed Leadership Practices and Student Science Performance through the Four-Path Model: Examining Failure in

País	Referencia
	<p>Underprivileged Schools. <i>Journal of Educational Administration</i>, 59(4), 472-492. https://doi.org/10.1108/JEA-07-2020-0159</p> <p>Wu, H., Gao, X., & Shen, J. (2020). Principal leadership effects on student achievement: A multilevel analysis using Programme for International Student Assessment 2015 data. <i>Educational Studies</i>, 46(3), 316-336. https://doi.org/10.1080/03055698.2019.1584853</p> <p>Wu, H., Shen, J., Zhang, Y., & Zheng, Y. (2020). Examining the effect of principal leadership on student science achievement. <i>International Journal of Science Education</i>, 42(6), 1017-1039. https://doi.org/10.1080/09500693.2020.1747664</p>
Argentina	<p>Formichella, M. M., & Krüger, N. S. (2017). Reconociendo el carácter multifacético de la educación: Los determinantes de los logros cognitivos y no-cognitivos en la escuela media argentina. <i>El Trimestre Económico</i>, 84(333), 165-191.</p> <p>Ibañez Martín, M. M., & Formichella, M. M. (2017). Logros educativos: ¿Es relevante el género de los estudiantes? <i>Education Policy Analysis Archives</i>, 25(3), 1-32. https://doi.org/10.14507/epaa.25.2520</p> <p>Krüger, N. S. (2013). Segregación social y desigualdad de logros educativos en Argentina. <i>Education Policy Analysis Archives</i>, 21(86), 1-30.</p> <p>Quiroz, S. S., Dari, N. L., & Cervini, R. A. (2018). Nivel Socioeconómico y Brecha entre Educación Secundaria Pública y Privada en Argentina. Los Datos de PISA 2015. <i>REICE. Revista Iberoamericana sobre Calidad, Eficacia y Cambio en Educación</i>, 16(4), 79-97. https://doi.org/10.15366/reice2018.16.4.005</p>

La quinta pregunta de investigación formulada para el mapeo se centraba en las revistas científicas en las que se han publicado los artículos. Se puede afirmar que de la gran variedad de revistas (45), destacan dos en las que se concentran más de una quinta parte de las investigaciones: *Revista de Educación* y *School Effectiveness and School Improvement*. La primera de ellas dependiente del Ministerio de Educación y Formación Profesional del Gobierno de España y la segunda dependiente de la editorial Taylor&Francis y fundada por el ICSEI.

En la tabla 2.4 se observa el número de artículos publicados en cada una de las fuentes, con al menos 2 publicaciones, de acuerdo con el criterio empleado en la presentación de resultados de la pregunta anterior, incluyendo también información relativa a dos índices de impacto propuestos por *SCIMAGO* (SJR) y *Web of Science* (JIF–SSCI o JCI–ESCI).

Tabla 2.4*Número de publicaciones por fuente*

Revista	N° de pub.	Scimago		Web of Science	
		Impacto	Cuartil	Impacto	Cuartil
Revista de Educación	5	0.345	Q3	-	-
School Effectiveness and School Improvement	5	1.017	Q1	3.000	Q2
Education Policy Analysis Archives	3	0.41	Q2	0.4*	Q4**
Compare	2	0.719	Q1	1.8	Q3
Educational Studies	2	0.605	Q2	1.7	Q3
Frontiers in Education	2	0.661	Q2	2.3*	Q2**
Frontiers in Psychology	2	0.891	Q2	3.8	Q1
International Journal of Educational Research	2	1.017	Q1	3.2	Q2
International Journal of Science Education	2	1.003	Q1	2.3	Q3
Large-Scale Assessments in Education	2	0.996	Q1	3.1*	Q1**
Scandinavian Journal of Educational Research	2	0.82	Q1	1.9	Q3

Nota. Estos índices y cuartiles han sido consultados con fecha 4 de julio de 2023 referidos a los rankings del año 2022. *revistas indexadas en el Emerging Sources Citation Index. ** revistas en las que el cuartil se basa en JCI en lugar de JIF.

Finalmente, la respuesta de la última pregunta del *mapping* M6 se construye en torno a la clasificación de las metodologías propuesta por Liou y Hung (2015) y ampliada por Gamazo (2019), donde se diferenciaba entre:

- Avanzado: correlación canónica, análisis factorial, análisis de discriminación, análisis clúster, SEM (*Structural Equation Modeling*), regresión logística, modelos multinivel, TRI (Teoría de Respuesta al Ítem), BD/ML (*Big Data y Machine Learning*) y Otros.
- Intermedio: regresión múltiple y comparación múltiple post hoc.
- Básico: análisis descriptivos, correlacionales, ANOVA, chi-cuadrado y prueba t de Student.

A partir de esta clasificación, se revisaron las metodologías empleadas en cada una de las investigaciones, obteniendo como resultado lo expuesto en la tabla 2.5:

Tabla 2.5*Número de publicaciones según la metodología utilizada*

Nivel	Técnica	n
Avanzado	Correlación canónica	1
	Análisis factorial	4
	Análisis de discriminación	-
	Análisis clúster	3
	SEM	6
	Regresión logística	3
	Modelos multinivel	55
	TRI	3
	BD/ML	-
	Otros	7
Intermedio	Regresión múltiple	6
	Comparación múltiple post hoc	-
Básico	Descriptivo	24
	Correlacional	6
	ANOVA	6
	Chi-cuadrado	-
	Prueba t	1

Uno de los criterios de inclusión fue que los estudios empleasen una metodología de corte cuantitativo. Por este motivo no es de extrañar que en la inmensa mayoría de las investigaciones incluidas en esta revisión la metodología de análisis multinivel sea una de las más utilizadas (55), puesto que tiene en cuenta la estructura anidada de los datos de las pruebas de evaluación a gran escala. Aunque en algunos casos se emplea de manera exclusiva (16 estudios), lo más frecuente es combinarla con otras metodologías, sobre todo con análisis descriptivos centrados particularmente en exponer las características de las variables incluidas posteriormente en la modelización multinivel (Caponera & Losito, 2016; Doncel Abad & Cabrera Álvarez, 2020; Laukaityte & Rolfsman, 2020; Sortkær & Reimer, 2018).

2.2.2. Resultados de la revisión sistemática de la literatura

Se procede a continuación a dar respuesta a las preguntas de investigación planteadas para la SLR.

En primer lugar, se planteaba la cuestión de cómo ha sido la evolución de la investigación sobre las pruebas a gran escala. Concretamente, la pregunta se refería a cuáles han sido las variables predictoras históricamente más utilizadas en la investigación sobre rendimiento educativo y eficacia escolar basada en datos de las ILSA y si estas variables se han ido modificando a lo largo del tiempo (en función del año de publicación). En la respuesta a la M6 se observa que la mayor parte de las investigaciones incluidas en este estudio empleaban metodologías que consideraban la estructura anidada de los datos, diferenciando las variables a nivel de alumno, escuela e incluso país, en aquellos estudios comparativos entre países, o aula, en estudios basados sobre todo en PIRLS.

Para dar respuesta, por tanto, a esta primera pregunta de investigación (PI1) y partiendo de la gran cantidad de variables consideradas en las diferentes pruebas internacionales de evaluación a gran escala, se presentan aquellas más frecuentemente utilizadas como predictoras del rendimiento académico o de la eficacia escolar de los estudiantes diferenciando entre el nivel de alumno y el de centro educativo:

- Nivel 1 (alumno): género, nivel socioeconómico y cultural, lengua hablada en el hogar, índice de inmigración y repetición de curso (Gamazo et al., 2018; Krüger, 2013; Martínez-Abad et al., 2020; H. Wu, Shen, et al., 2020).
- Nivel 2 (centro educativo): nivel socioeconómico medio del centro, localización (urbana/rural), titularidad, escasez de profesorado en el centro y escasez de recursos educativos en el centro (Gómez-Fernández & Mediavilla, 2021; Pomianowicz, 2021; Sousa et al., 2012; van Hek et al., 2018; H. Wu, Gao, et al., 2020).

En los estudios en los que se ha tenido en cuenta como tercer nivel el país, las variables empleadas para la comparación por los diferentes autores han sido: Producto Interior Bruto (PIB) del país, PIB per cápita, gasto en educación o índice Gini (Gamazo & Martínez-Abad, 2020; Gil-Flores & García-Gómez, 2017; M. Lee et al., 2021; Liu et al., 2014; Ruiz de Miguel, 2009; Ruiz de Miguel & Castro Morera, 2006; Tan & Liu, 2018).

En cuanto a la evolución en el tiempo de las variables empleadas en estos estudios secundarios, sobre todo de PISA, se puede determinar que con el paso del tiempo ha

habido una serie de variables contextuales que han permanecido como el género o el nivel socioeconómico de los estudiantes y el tamaño o la titularidad de la escuela, mientras que otras variables han adquirido mayor relevancia con el paso del tiempo como las relacionadas con el clima disciplinario o el tipo de liderazgo ejercido por los equipos directivos de los centros educativos.

La segunda pregunta de investigación se refiere a la variable criterio de los estudios. Concretamente, si se trataba del rendimiento académico o de la eficacia escolar. Tras revisar los artículos seleccionados, tan solo uno de ellos difería del resto, siendo el rendimiento académico la variable criterio en 62 de los 63 artículos. En la mayoría de los casos basados en datos de PISA, este rendimiento académico estudiado era el referido a la competencia principal de acuerdo con la edición de la prueba.

Para responder a la PI3 se centró la atención en la estructura jerárquica o anidada de los datos. De nuevo, de las 63 investigaciones incluidas en la revisión, la inmensa mayoría (60), tenían en cuenta esta estructura en sus análisis, diferenciando sobre todo dos niveles: estudiante y centro educativo. En cambio, ya se ha mencionado que algunos estudios como el de Lee et al. (2021) o el de Tan y Liu (2018) incluían también un tercer nivel, país, con el fin de establecer comparaciones entre países en base a variables como el PIB o el gasto medio en educación.

Centrándonos ahora en los estudios contruidos sobre datos de PIRLS y TIMSS seleccionados en esta revisión, el nivel de aula entra en juego al disponer de este tipo de información. De este modo, se incluyen variables referidas al número de horas de docencia semanal en matemáticas o en comprensión lectora, el tiempo medio invertido en la realización de deberes de matemáticas o el nivel socioeconómico medio de los alumnos de una determinada clase, entre otras (Bokhove et al., 2019; Grilli et al., 2016).

Como conclusión, y a modo de respuesta a esta tercera pregunta de investigación, se puede afirmar que, como norma general, los investigadores tienen en cuenta la estructura jerárquica de los datos de las pruebas internacionales de evaluación a gran escala en sus estudios.

Además de la estructura jerárquica de los datos, es importante conocer cómo se lleva a cabo el tratamiento y el empleo de los valores plausibles y de los pesos muestrales (PI4) en las investigaciones (aspectos metodológicos detallados en el apartado 4.2.3). La mayor parte de estudios incluidos en esta revisión (63 artículos) se construyen sobre datos procedentes de PISA. De entre ellos, en 48 se tienen en cuenta los valores plausibles siguiendo las recomendaciones de la OCDE en relación con su utilización. En tres

investigaciones no se consideran dichos valores (Ertem, 2021; Sousa et al., 2012; Tan & Dimmock, 2022), mientras que en otras dos sí se utilizaron, aunque no de la forma adecuada según los organismos internacionales. En un caso aplicando la media de los valores (M. Lee et al., 2021) lo que supone una pérdida de la variabilidad y, en otro caso, mediante el uso de tan solo el primero de los valores plausibles (Agasisti et al., 2021).

De nuevo analizando los estudios secundarios de PISA, en 27 publicaciones se incluyen los pesos muestrales de los dos niveles de análisis: centros educativos y estudiantes. En cambio, en otros 12 artículos se utiliza tan solo el peso muestral del nivel de estudiante, mientras que en los 14 restantes no se emplea ninguno de los pesos muestrales.

Combinando el uso de los pesos muestrales y los valores plausibles en los estudios basados en PISA, se obtienen los siguientes resultados (tabla 2.6):

Tabla 2.6

Empleo de valores plausibles (PV) y pesos muestrales (PM) en investigaciones con datos PISA

	PV		PM
		Sí	24
Sí	48	No	12
		El de los estudiantes	12
No	3	Sí	1
		No	2
Más o menos	2	Sí	2

Nota. “Más o menos” dentro de “PV” se refiere a que no se utilizan los valores plausibles siguiendo las recomendaciones de la IEA o la OCDE; “El de los estudiantes” dentro de “PM” se refiere a que únicamente se utiliza el peso muestral del nivel de estudiantes.

Prestando atención a las publicaciones construidas sobre datos de TIMSS o de PIRLS, en todas ellas se hace un uso adecuado tanto de los pesos muestrales como de los valores plausibles de acuerdo con las indicaciones de la IEA, organismo internacional encargado del desarrollo de TIMSS y PIRLS.

En cuanto a las limitaciones encontradas o derivadas de las investigaciones (PI5), no se muestran en todos los estudios. La gran limitación ligada a las pruebas a gran escala es su carácter eminentemente transversal que lleva implícito que los resultados de este tipo de investigaciones no puedan ser considerados como afirmaciones causales,

sugiriendo por lo tanto la necesidad de realizar estudios longitudinales (Bokhove et al., 2019; Costa & Araújo, 2018; Li, 2016; H. Wu, Gao, et al., 2020; H. Wu, Shen, et al., 2020). Por la naturaleza de las ILSA, dirigidas a una población específica en un momento determinado (estudiantes de 15 años en PISA, por ejemplo), los estudios longitudinales resultan imposibles de realizar. Algunos autores han indicado que sería interesante disponer de información sobre el rendimiento previo de los estudiantes, puesto que podría facilitar la realización de comparaciones en base, por ejemplo, al capital cultural de las familias (Danhier, 2018; Huang & Sebastian, 2015; Marteleto & Andrade, 2014; Myrberg et al., 2019). Asimismo, el hecho de que existan diferencias en escalas e indicadores entre las distintas ediciones de una misma prueba en función de las tendencias socioeducativas del momento dificulta esa realización de estudios longitudinales tan demandada. En palabras de Martínez-Abad et al. existe una “falta de sistematicidad en la perspectiva de las evaluaciones PISA” (2020, p. 8).

A estas limitaciones se debe sumar el posible sesgo derivado de la extracción de datos puesto que se recogen mediante autoinformes de profesores, estudiantes y equipos directivos de los centros, limitando de este modo la objetividad de los mismos (Danhier, 2018; H. Wu, Gao, et al., 2020; H. Wu, Shen, et al., 2020).

Algunos autores indican otras limitaciones ligadas al compromiso metodológico que implica la utilización de la media de los valores plausibles (M. Lee et al., 2021), en lugar de emplearlos como recomiendan los organismos internacionales.

También se hace referencia a que no se disponga de datos a nivel de aula (Burns et al., 2020; He & Fischer, 2020; H. Wu, Gao, et al., 2020), lo que podría proporcionar información importante sobre las prácticas desarrolladas por el profesorado.

Otras investigaciones sugieren el alto número de valores perdidos en ciertas variables o países (Guo et al., 2018) que, junto con los sesgos culturales de las propias pruebas (Lenkeit, 2012), derivaría en la imposibilidad de extrapolar las conclusiones obtenidas en este tipo de estudios comparativos.

En este momento, cabe hacer referencia a una última limitación referida a la deseabilidad social a la hora de responder a los cuestionarios por parte de los estudiantes y, como planteaban Lopes et al., su tendencia a “give definite answers and use the scales’ upper ends” (2022, p. 35).

Finalmente, y en respuesta a la última pregunta propuesta (PI6), las implicaciones sociales de este tipo de investigaciones han consistido en la formulación de una serie de recomendaciones adaptadas a cada país con el fin de paliar las deficiencias detectadas en

los diferentes estudios con relación a los factores asociados al rendimiento académico. Es el caso de investigaciones como la propuesta por Yetişir y Bati (2021) que señalan la necesidad de elaborar en Turquía reformas educativas a largo plazo persiguiendo aumentar la calidad de las prácticas docentes, a partir de la asociación resaltada entre la escasez de recursos educativos y de profesorado y el rendimiento educativo de los estudiantes turcos. En cambio, otros autores señalan que una posible línea de investigación podría ser la exploración de determinados contextos, como por ejemplo los centros educativos con resultados por encima de lo esperado (escuelas de alta eficacia), a través de la recogida de datos de carácter más cualitativo que permitieran profundizar en el conocimiento de los factores de contexto mediante la aplicación de métodos mixtos en la investigación (Burns et al., 2020; Ertem, 2021; H. Wu, Shen, et al., 2020).

2.3. Conclusiones derivadas de la SLR

Esta revisión pretende servir para conocer los estudios de carácter secundario realizados sobre las pruebas de evaluación internacional a gran escala y su relación con la eficacia escolar y el rendimiento académico. Antes de centrarnos en las conclusiones que se pueden extraer de este estudio, se discuten algunas de las limitaciones encontradas.

Se mencionaba con anterioridad el sesgo derivado de la extracción de datos de las ILSA, pero es necesario indicar que la propia realización de esta revisión sistemática de la literatura se asocia a otras dos grandes limitaciones: la subjetividad del investigador y el sesgo de publicación. La primera de estas limitaciones se deriva de los intereses y objetivos de la investigación, puesto que, en función de estos, se han planteado determinados criterios de elegibilidad que evidentemente condicionan la futura selección de artículos. En cuanto al sesgo de publicación, se entiende como el riesgo de que exista una mayor representación de artículos con resultados positivos derivados del sistema de publicación actual. Estas dos limitaciones se podría decir que son paliadas en gran medida por la transparencia que aporta al proceso la sistematicidad y la exposición clara de los diferentes criterios empleados en el propio procedimiento, sobre todo en relación con la primera de ellas.

Diversos autores aluden al carácter transversal de las pruebas de evaluación a gran escala como una de las mayores limitaciones asociadas a este tipo de estudios (Bokhove et al., 2019; Costa & Araújo, 2018; Li, 2016; H. Wu, Gao, et al., 2020; H. Wu, Shen, et al., 2020). Si bien es cierto que el muestreo de estudiantes realizado en este tipo de

pruebas suele ser bi-etápico, seleccionando en un primer momento una muestra representativa de centros educativos según su región o país y su titularidad para, a continuación, seleccionar a los estudiantes que participarán en la prueba. A pesar de que con las pruebas de evaluación a gran escala no es posible la realización de estudios longitudinales, puesto que se dirigen a una población determinada en un momento concreto (por ejemplo, estudiantes de 15 años en el momento de la evaluación en el caso de PISA), este tipo de muestreo lleva implícita la representatividad de la población, por lo que se podrían desarrollar investigaciones aproximadas mediante la utilización de metodologías como los estudios de tendencias. Los estudios de tendencias son una variable de los estudios por cohortes en los que gracias a la medición de diferentes poblaciones en diferentes momentos se busca retratar los cambios producidos o existentes, esto supone que los sujetos de los que se recoge la información varían en cada momento (Bisquerra, 2004; Cohen et al., 2017; Sullivan & Calderwood, 2017). Sin embargo, esta podría ser una metodología aceptable para comprobar los cambios educativos puesto que, teóricamente, la población a la que representan los estudiantes es el mismo grupo, aunque en momentos diferentes.

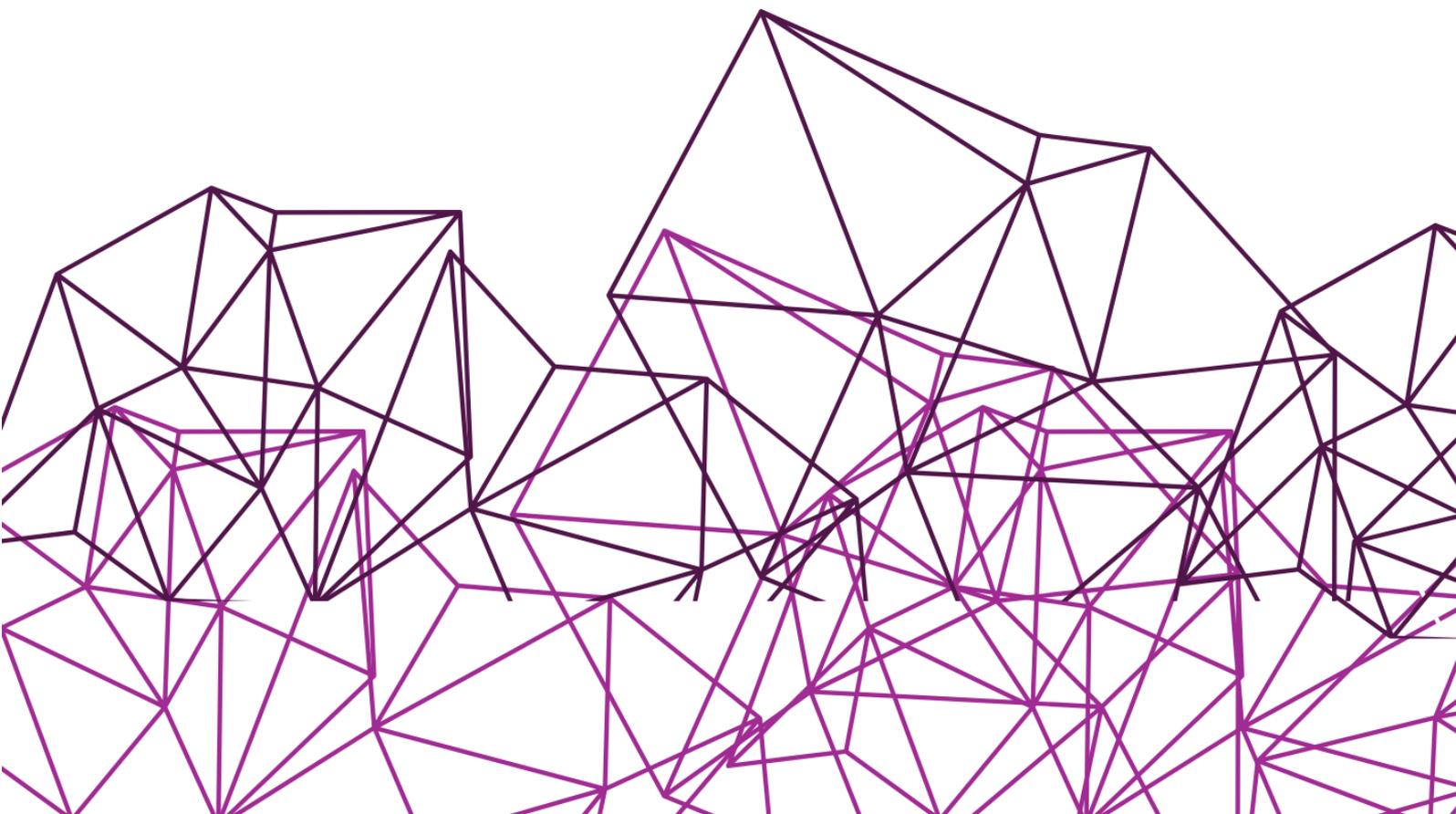
Con esta revisión se ha observado un interés creciente en esta temática, destacando un notable aumento de las publicaciones en los últimos años concentrándose más del 75% de los artículos seleccionados en el período comprendido entre el año 2015 y la actualidad.

Del mismo modo, hay que destacar el aseguramiento de las garantías estadísticas en el empleo de la metodología de análisis multinivel teniendo en cuenta tanto los pesos muestrales como los valores plausibles según lo planteado por los organismos internacionales en 34 de las 63 publicaciones. Esto podría ser considerado de manera positiva, puesto que en más de la mitad de los artículos seleccionados se emplea la citada metodología, pero también se puede entender como una posible mejora al haber otras 29 publicaciones en las que no se hace un uso correcto de estos dos elementos.

En cuanto a las otras limitaciones y críticas indicadas por algunos autores hacia las propias ILSA en relación con su construcción, objetivos o aplicación destaca que estas pruebas, por su carácter estandarizado, no tienen en cuenta las características diferenciales de las regiones o las diferentes culturas y, además, promueven la competitividad entre los países, haciendo que se deban plantear dos cuestiones fundamentales: ¿son las ILSA la herramienta más adecuada para el diagnóstico de los sistemas educativos? ¿Son válidas las comparaciones que se establecen entre los países o

regiones basadas en los resultados de estas pruebas? Se discutirán estas y otras cuestiones en la sección dedicada a las conclusiones de este mismo trabajo, presentando también las posibles líneas futuras de investigación en este ámbito.

3. MODELOS MULTINIVEL



Modelos multinivel

Históricamente, la técnica cuantitativa más empleada en la investigación educativa para el estudio de las variables explicativas de una variable criterio (como el rendimiento educativo, la eficacia escolar o la equidad educativa) ha sido la regresión lineal. Sin embargo, el uso de esta metodología está supeditado al cumplimiento de un supuesto previo fundamental: la independencia de las observaciones (Aitkin & Longford, 1986; Murillo, 2008; Raudenbush & Bryk, 2002).

Hace más de tres décadas la investigación educativa experimentó una revolución por las aportaciones de diferentes autores que evidenciaron que, en este ámbito, las observaciones no pueden ser consideradas independientes (Aitkin & Longford, 1986; Raudenbush & Bryk, 2002). Por tanto, se hacía necesaria la aplicación de otras técnicas estadísticas para el tratamiento de estos datos, proponiendo la utilización de los modelos multinivel o modelos jerárquico-lineales, entre otras.

La metodología multinivel es una técnica de regresión estadística que se construye en la organización anidada de los datos en unidades de orden superior (estudiantes en aulas, aulas en escuelas, escuelas en países, etc.) y sirven para tratar de explicar la varianza existente entre los distintos niveles de análisis de manera simultánea. Son, por tanto, ampliaciones de los modelos de regresión lineal clásicos en las que se elaboran varios modelos de regresión para cada nivel de análisis (Murillo, 2008). Cada modelo expresa la relación entre las variables del mismo nivel y cómo estas tienen influencia en las relaciones existentes en los otros niveles (Murillo, 2008). Es decir, los modelos multinivel presentan una combinación lineal de distintas variables predictoras con una variable criterio, cada una de ellas asociada a un coeficiente que indica cuánta variación en la variable predictora afecta a la variable criterio (Gamazo, 2019; Hayes, 2006).

La complejidad de la realidad educativa radica en la agrupación de los estudiantes en aulas, estas en escuelas, estas en localidades, etc.; este hecho supone que entre los distintos grupos existan una serie de características comunes que apoyan esa no independencia de los datos. Por tanto, se puede concluir que esta realidad es multivariada y multinivel.

Según Lizasoain y Angulo (2014):

Este conjunto de técnicas estadísticas respetan la estructura anidada de los datos que es la habitual en educación (...) y permiten estudiar conjuntamente los efectos

de las variables de nivel individual y del contextual facilitando el estudio de los efectos de las covariables en cada nivel de anidamiento. (p. 18)

Se ha comprobado en la investigación educativa que el uso de los modelos multinivel ha aumentado con el paso de los años, empleando esta metodología para la realización de investigaciones con dos (Gamazo et al., 2018; Yetişir & Bati, 2021), tres (Gaviria Soto & Castro Morera, 2004; M. Lee et al., 2021; Tan & Liu, 2018) o incluso más niveles de análisis (Murillo & Hernández-Castilla, 2011). En el caso que nos ocupa, se expone cómo se debe realizar la correcta construcción de los modelos multinivel de dos niveles (estudiante y centro educativo), puesto que la base de datos proporcionada por la OCDE sobre las pruebas PISA no incluye información referida al nivel de aula.

3.1. Fundamentos de los modelos multinivel: supuestos básicos

El uso adecuado de los modelos multinivel para el análisis de datos se asocia con la necesidad de cumplimiento de una serie de supuestos básicos sobre la distribución de los elementos que los componen y las relaciones que se establecen entre ellos, en línea con los supuestos básicos de los modelos de regresión tradicionales que se indican a continuación:

Dependencia de las observaciones

Las técnicas de estadística inferencial se construyen en base a la independencia existente entre las distintas observaciones de una muestra, pero, en el caso de los modelos multinivel se hace necesaria la existencia de dependencia entre las observaciones de los distintos niveles. Este grado de dependencia se mide con el cálculo del Coeficiente de Correlación Intraclase (CCI), que indica el porcentaje de variabilidad entre las puntuaciones de los estudiantes que es atribuible al nivel de centro (ver sección 3.2.1).

Linealidad de la relación entre las variables predictoras y la variable criterio

En el caso de este supuesto, las técnicas de regresión multinivel coinciden con la regresión tradicional asumiendo la naturaleza lineal de la relación entre las variables predictoras y la variable criterio.

Distribución y varianza de los residuos

Cuando aludimos a residuos se entiende como la diferencia existente entre la puntuación real y la puntuación estimada para una determinada unidad según las variables que conforman el modelo estadístico (Joaristi et al., 2014).

Se asume que los residuos se distribuyen normalmente con media 0 y varianza constante. Del mismo modo, se asume el supuesto de homocedasticidad referido a la igualdad de la varianza del residuo.

3.2. Especificación del modelo

Un modelo jerárquico lineal de dos niveles es una herramienta estadística utilizada para analizar datos que tienen una estructura jerárquica. En este caso, el primer nivel de análisis es el estudiantado (nivel 1), y el segundo nivel de análisis es la escuela (nivel 2). Se presenta a continuación la descripción teórica de los pasos a seguir para la especificación o construcción de los modelos mediante la técnica de análisis multinivel partiendo del modelo incondicional o nulo hasta alcanzar el modelo condicional final.

3.2.1. Paso 1. Formulación del modelo nulo

La construcción de un modelo multinivel comienza por la especificación del modelo nulo o incondicional en el que no existe ninguna variable predictora. Este modelo se concreta en la siguiente ecuación (ecuación 3.1):

$$Y_{ij} = \gamma_{00} + u_{0j} + r_{ij} \quad (\text{Ecuación 3.1})$$

Donde:

Y_{ij} se refiere a la puntuación de la variable criterio del alumno i en la escuela j ;

γ_{00} se refiere a la media de todos los centros en la variable criterio;

u_{0j} se refiere a la diferencia entre la puntuación del centro j en la variable criterio menos la media de todos los centros;

r_{ij} se refiere al residuo de nivel 1.

Para comprobar la dependencia entre las mediciones y más específicamente cómo difieren las puntuaciones en la variable criterio entre las unidades de nivel 2 se calcula el

CCI del modelo nulo gracias a la estimación de los componentes de la varianza de ambos niveles de análisis (ecuación 3.2):

$$CCI = \frac{\tau_{00}}{\tau_{00} + \delta^2}$$

(Ecuación 3.2)

Donde:

τ_{00} es el componente de la varianza de nivel 2, y

δ^2 es el componente de la varianza de nivel 1.

Una correlación baja o cercana a 0 indica que las observaciones de nivel 1 no dependen del nivel 2, es decir, que no existen similitudes entre los sujetos de un mismo grupo. En el caso presentado, entre los estudiantes de una misma escuela. Por lo tanto, de acuerdo con lo propuesto por Lee (2000), el valor del CCI debe ser de al menos un 10% para considerar adecuada la aplicación de esta metodología.

3.2.2. Paso 2. Ampliación del modelo: el modelo condicional

Comprobado el primer supuesto de dependencia de las observaciones, se puede proceder con el ajuste del modelo, o lo que es lo mismo, la inclusión de las variables predictoras sobre las que se quiere comprobar la relación con la variable criterio estudiada.

Se presentan, por tanto, las ecuaciones de ambos niveles de análisis con el fin de facilitar la comprensión de la composición del modelo condicional (Raudenbush & Bryk, 2002). En estas ecuaciones el subíndice i se refiere a los estudiantes, mientras que j se refiere a las escuelas.

La ecuación del modelo de primer nivel es la siguiente (ecuación 3.3):

$$\begin{aligned} Y_{ij} &= \beta_{0j} + \beta_{1j}X_{1ij} + \beta_{2j}X_{2ij} + \dots + \beta_{Qj}X_{Qij} + r_{ij} \\ &= \beta_{0j} + \sum_{q=1}^Q \beta_{qj}X_{qij} + r_{ij} \end{aligned}$$

(Ecuación 3.3)

En la cual:

β_{qj} ($q = 0, 1, \dots, Q$) son los coeficientes de nivel 1;

X_{qij} es el predictor q del nivel 1 para el caso i en la unidad j ;

r_{ij} es el efecto aleatorio del nivel 1 con varianza δ^2 .

De acuerdo con lo expuesto anteriormente, se asume que $r_{ij} \sim N(0, \delta^2)$.

Los coeficientes de nivel 1 (β_{qj}) se transforman en variables de resultado en el modelo de nivel 2 (ecuación 3.4):

$$\begin{aligned}\beta_{qj} &= \gamma_{q0} + \gamma_{q1}W_{1j} + \gamma_{q2}W_{2j} + \dots + \gamma_{qS_q}X_{S_qj} + u_{qj} \\ &= \gamma_{0j} + \sum_{s=1}^{S_q} \gamma_{qs}W_{sj} + \mu_{qj}\end{aligned}$$

(Ecuación 3.4)

Donde:

γ_{qs} ($q = 0, 1, \dots, S_q$) son los coeficientes de nivel 2;

W_{sj} es un predictor del nivel 2;

μ_{qj} es el efecto aleatorio del nivel 2.

Al igual que en el caso de la ecuación referida al primer nivel, se asume que en cada unidad j , cada elemento de μ_{qj} tiene una media de 0 y una varianza de τ_{qq} .

3.3. Los efectos escolares y el valor añadido

Con la aplicación de los modelos multinivel para la realización de estudios sobre el rendimiento académico se obtiene información referida a la magnitud y dirección de la relación entre las distintas variables independientes y esta variable criterio, pero también son de utilidad para comprobar cuál es la contribución que realiza el centro educativo a esa variable criterio. Esa aportación del centro educativo se puede comprender a través de dos conceptos: los efectos escolares y el valor añadido.

Se toma como base la definición propuesta por Murillo, según la cual los efectos escolares son “la capacidad de los centros educativos para influir en los resultados de sus alumnos” (2007, p. 24). La magnitud de los efectos escolares se mide en los modelos multinivel gracias a los residuos de segundo nivel.

Existen varios tipos de efectos escolares, los primeros han sido denominados en la literatura como efectos de tipo A y de tipo B (Raudenbush & Willms, 1995). Los efectos de tipo A se refieren al efecto institucional, es decir, la variación en el rendimiento de los estudiantes que asisten a diferentes escuelas, pero con similares características contextuales a nivel individual, concretamente hablan de que el efecto de tipo A “is the difference between a child's actual performance and the performance that would have

been expected if that child had attended a "typical school" (Raudenbush & Willms, 1995, p. 309). Es decir, la diferencia entre el rendimiento real de un niño y el que cabría esperar si ese niño hubiera asistido a una "escuela típica". En cambio, los efectos escolares de tipo B son los debidos a las diferencias individuales entre los alumnos, independientemente de la institución a la que asisten, es decir, se trata de la variabilidad en el rendimiento que no está relacionada con la escuela, según señalan Raudenbush y Willms: "is the difference between a child's performance in a particular school and the performance that would have been expected if that child had attended a school with identical context but with practice of "average" effectiveness" (1995, p. 310), o sea, la diferencia entre el rendimiento de un niño en una escuela determinada y el rendimiento que cabría esperar si ese niño hubiera asistido a una escuela con el mismo contexto pero con prácticas de eficacia "medias".

Casi una década después y prosiguiendo con la investigación relacionada con los efectos escolares, Keeves et al. (2005) definieron un segundo (o tercer) tipo de efectos escolares: los efectos de tipo X, referidos a la inclusión de características relacionadas con la política educativa de los centros que estos no pueden controlar como el tamaño de la propia escuela, su localización o su titularidad.

Un último tipo de efectos escolares son los efectos tipo Z propuestos por Darmawan y Keeves (2006), que se refieren también a aspectos relacionados con las políticas educativas de los centros, pero, en este caso, aquellos en los que sí pueden influir los propios centros, como la agrupación de los estudiantes en clases en función de distintas características.

La idea principal sobre la que se construyen los modelos de valor añadido es discernir cuánta proporción del rendimiento de los estudiantes se debe a las características de la institución en comparación con las características individuales de los estudiantes. Estos efectos son esenciales para comprender el impacto de las instituciones educativas en el rendimiento estudiantil y para realizar evaluaciones más precisas de la eficacia de las escuelas en un contexto multinivel.

A pesar de que para el cálculo de estos efectos escolares se puede utilizar la información proporcionada por estudios de carácter transversal (una única medición en un momento específico), en la IEE se resalta la importancia de la realización también de estudios longitudinales en los que existan dos o más mediciones de los mismos sujetos para establecer las comparaciones oportunas y determinar la magnitud de los efectos escolares según el progreso académico de los estudiantes. Esta cuestión ha constituido

una de las mayores críticas y limitaciones planteadas para la realización de las pruebas de evaluación a gran escala (como se ha mencionado y se mencionará más adelante en este mismo documento), generando por tanto el planteamiento de diferentes tipos de estudios de los efectos escolares, así como el término valor añadido. Se entiende que el indicador del valor añadido de un centro es “la diferencia entre la puntuación obtenida y la esperada en cada centro escolar” (Lizasoain et al., 2015, p. 81). En ese caso, para Muñoz Martín (2006):

La idea básica del concepto de valor añadido (en el aprendizaje del alumno) es la diferenciación y el aislamiento de los factores puramente escolares que influyen en los resultados escolares, separándolos de los factores individuales y sociales que irremediablemente también intervienen. (p. 5)

Es sobre este término sobre el que se ha construido la investigación en eficacia escolar. La OCDE (2011) recomienda la utilización del término valor añadido para aquellos estudios longitudinales centrados en el cambio y el crecimiento a lo largo del tiempo, en el caso de que el estudio planteado no sea de carácter longitudinal, sugieren emplear el término modelos transversales contextualizados (Joaristi et al., 2014).

En esta tesis doctoral, se ubica la investigación dentro de los modelos transversales contextualizados, puesto que se construye sobre una medida específica (PISA 2018) en la que se tienen en cuenta los efectos del contexto en el rendimiento de los estudiantes. Al no disponer de datos longitudinales, no puede considerarse como valor añadido, pero sí nos permite acercarnos a la determinación de la aportación específica de un centro educativo al rendimiento de sus estudiantes controlando los efectos del contexto mediante la estimación de los valores residuales de las escuelas (diferencia entre el rendimiento real de un centro educativo y el esperado una vez controladas las características contextuales).

3.4. Características metodológicas

La construcción de modelos multinivel supone la toma de diferentes decisiones metodológicas en función de los datos de los que se dispone y de los objetivos de investigación.

3.4.1. Métodos de estimación

Existen distintos métodos para la estimación de parámetros cuando se aplica la metodología de análisis multinivel: máxima verosimilitud o máxima verosimilitud restringida para los coeficientes del modelo, empírica bayesiana para los residuos y estimación robusta para los errores.

La estimación de parámetros mediante la utilización de las técnicas de máxima verosimilitud (ML) y máxima verosimilitud restringida (REML), como su propio nombre indica, maximizan la función de verosimilitud, que es la encargada de medir la probabilidad de observar los datos según los parámetros del modelo (Gaviria Soto & Castro Morera, 2005). Estas dos técnicas son similares en cuanto a la estimación de la parte fija del modelo, pero para la estimación de la parte aleatoria es más oportuno el empleo de la REML, puesto que puede proporcionar una estimación más precisa al tener en cuenta la variabilidad de los efectos aleatorios (Snijders & Bosker, 2011), sobre todo en muestras pequeñas inferiores a 30 unidades (Gaviria Soto & Castro Morera, 2005).

La estimación bayesiana se utiliza para enfrentar la diferencia de varianza en los residuos de los modelos multinivel, permitiendo que esta varíe entre los grupos. Las ventajas de estos estimadores son la alta correlación con la estimación del efecto verdadero y la compensación de estimaciones espurias por escasez de datos (Sanders & Wright, 2008) gracias a la aplicación de un factor de corrección (*shrinkage*) en función de la fiabilidad o el tamaño de la muestra (Sanders & Wright, 2008; Snijders & Bosker, 2011).

Las técnicas de máxima verosimilitud dan como resultado, generalmente, la estimación de los errores estándar a través de la matriz de información de Fischer (Snijders & Bosker, 2011). Sin embargo, ante la aparición de datos atípicos o ante desviaciones de la normalidad de las distribuciones, una estimación robusta para los errores supone la minimización del impacto de esos datos atípicos en las estimaciones de los parámetros, como en el caso de los modelos multinivel. Estas estimaciones robustas también se conocen como estimadores *sándwich*, con la característica principal de tener en cuenta la estructura anidada de los datos presente en las estructuras jerárquicas.

3.4.2. Efectos fijos y aleatorios

Los coeficientes de regresión que intervienen en los modelos multinivel (intercepto y covariables) pueden o bien tomar valores diferentes entre unas y otras unidades de análisis

(efectos aleatorios) o bien mantenerse estables adoptando el mismo valor para todas las unidades de análisis (efectos fijos). Determinar el comportamiento de estos coeficientes de regresión es otra de las decisiones importantes en el análisis multinivel de datos educativos.

Concretamente, el intercepto en la investigación educativa se refiere a los valores de las puntuaciones de los centros educativos, por lo tanto, sí suele entenderse como un elemento aleatorio al variar entre las unidades de segundo nivel según su media en la variable dependiente (Hayes, 2006). En cambio, los efectos de las covariables que componen los modelos estadísticos se pueden interpretar como efectos fijos o variables. Conjugar estas dos características aportaría como resultado dos tipos de modelado distintos, los modelos de efectos mixtos donde el intercepto es aleatorio y las pendientes son fijas, y los modelos completamente aleatorios con intercepto y pendientes ambos aleatorios.

En el modelo de efectos mixtos las pendientes se mantienen constantes independientemente de la unidad de nivel superior a la que se refieran, mientras que los efectos aleatorios (intercepto) indica la variabilidad entre los grupos no explicada por los efectos fijos, permitiendo así tener en cuenta las diferencias intergrupales. Suelen utilizarse cuando “el objetivo de la investigación es la inferencia sobre aquellos grupos que conforman la muestra” (Gamazo, 2019, p. 124).

En estudios previos al presentado en esta tesis doctoral (Gamazo, 2019) se calcularon los modelos estadísticos utilizando tanto modelos de efectos mixtos como modelos completamente aleatorios no encontrando grandes diferencias entre la aplicación de una u otra técnica. De acuerdo con lo propuesto por distintos autores (Özdemir, 2016; Martínez-Garrido, 2015), esta decisión sobre la elección del uso de pendientes fijas o aleatorias puede tomarse en base al principio de parsimonia, optimizando el modelo la utilización de pendientes fijas limitando los parámetros empleados en los mismos. Es por este motivo por el cual en este estudio se ha optado por el empleo de modelos mixtos con intercepto aleatorio y pendientes fijas.

3.5. Ventajas y limitaciones de esta técnica

La ventaja fundamental que aporta la utilización de modelos multinivel en la investigación educativa es que consideran la estructura jerárquica de los datos,

permitiendo la introducción de covariables de distintos niveles en el análisis y el estudio de los efectos cruzados entre los niveles (Gaviria Soto & Castro Morera, 2005).

Parece necesario, por tanto, indicar también alguna de las limitaciones planteadas por el uso de esta técnica estadística como la interpretación de sus resultados o la necesidad de muestras grandes para su adecuada aplicación. Pero, previamente, se presentan algunas ventajas y limitaciones de otras técnicas de regresión para la realización de estudios basados en datos jerarquizados como puede ser la regresión lineal.

Generalmente, para el tratamiento de datos anidados mediante una regresión lineal se pueden utilizar dos procedimientos: agregación y desagregación. El primer procedimiento consiste en agregar los datos del primer nivel según las unidades de segundo nivel (o sucesivos) utilizando en los análisis los valores medios de las covariables de nivel inferior junto con las variables propiamente del nivel superior analizado. Esta metodología implica desventajas a la hora de interpretar los resultados ya que se pierde entre un 80% y un 90% de la variabilidad debida a las diferencias individuales (Raudenbush & Bryk, 1992). Se incurriría en una falacia a la hora de interpretar los resultados denominada en la literatura como falacia ecológica (Hox et al., 2010), y que consiste en la extracción de conclusiones sobre los individuos que conforman los grupos sin tener en cuenta las características individuales, si no tan solo las del propio grupo. En el caso de la investigación educativa algunos ejemplos son el género o el nivel socioeconómico de los estudiantes de diferentes centros educativos como variables de primer nivel, que serían agregadas en función de la escuela de pertenencia eliminando de este modo la variabilidad dentro de cada escuela.

En cuanto a la desagregación, el procedimiento sería el inverso al explicado en el párrafo anterior. Las variables de nivel superior se desagregan a nivel individual eliminando la posible variación que pudiera existir entre los grupos (Woltman et al., 2012). En este caso se podría incurrir en otra falacia de interpretación de los resultados: la falacia atomista, generalizando conclusiones sobre los grupos a partir de características individuales (Hox et al., 2010).

El análisis multinivel permite esquivar estas limitaciones gracias al respeto hacia las características de los distintos grupos y niveles y a la posibilidad de analizarlos de manera conjunta estudiando las relaciones entre sí (Paterson & Goldstein, 1991).

A continuación, se exponen las limitaciones existentes en la utilización del análisis multinivel. La primera de ellas es la necesidad de disponer de muestras amplias en las unidades de nivel más altas para obtener resultados con las garantías estadísticas

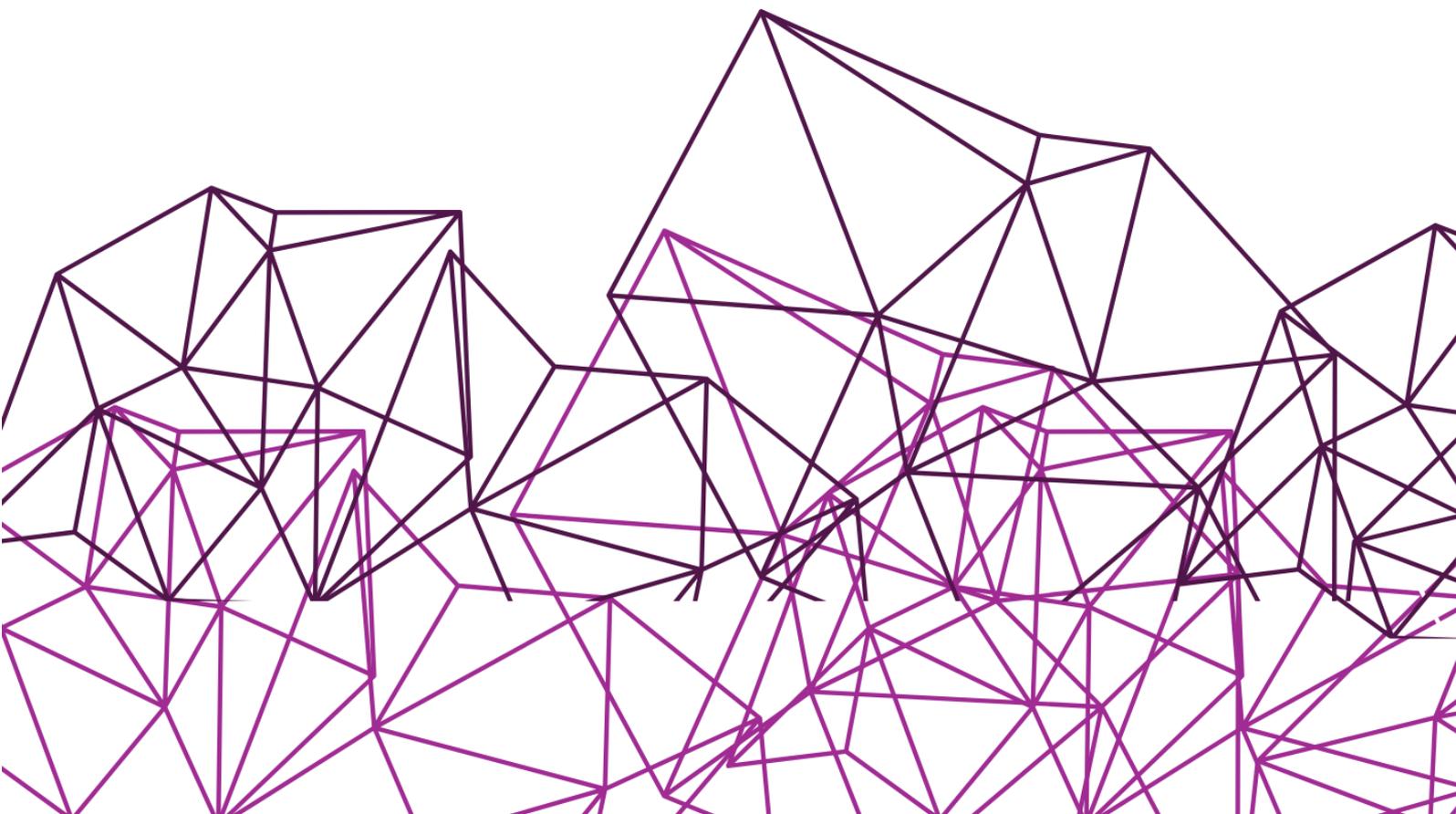
adecuadas. En cuanto al mínimo de unidades de mayor nivel necesarias, Maas y Hox (2005) sugieren un rango de entre 50 y 100 unidades para los modelos según su complejidad, puesto que menos de 50 unidades conducen a estimaciones sesgadas de los errores típicos del segundo nivel.

En segundo lugar, el análisis de diversos niveles de manera simultánea dificulta la consecución de modelos parsimoniosos (Diez-Roux, 2000). Debido a la complejidad de los análisis, en el planteamiento de las relaciones entre las variables y en la propia interpretación de los resultados.

Además, Hox y Kreft (1994) indican una tercera y muy importante limitación de los modelos multinivel: no permiten el estudio de más de una variable criterio a la vez. Ante esta limitación, se están produciendo una serie de avances metodológicos relacionados con la elaboración de modelos de ecuaciones estructurales de múltiple indicador y múltiple causa (MIMIC) que sí que permiten este uso de diferentes variables criterio (Tsai et al., 2017).

Del mismo modo, distintos autores aluden al hecho de que los modelos multinivel proporcionan “un entorno natural en el que se pueden expresar y comparar las teorías acerca de las relaciones estructurales entre variables de cada uno de los niveles en una jerarquía organizativa” (Ruiz de Miguel & Castro Morera, 2006, p. 6) de acuerdo con lo propuesto por Draper (1995), siendo este uno de los puntos fuertes de esta metodología.

4. METODOLOGÍA



Metodología

Esta investigación se construye sobre un enfoque de análisis cuantitativo con el fin de conocer aquellas variables predictoras de los dos fenómenos que se estudian: el rendimiento académico y la eficacia escolar. Bisquerra (2004) indica cómo en estas investigaciones se tiende a fragmentar la realidad educativa y se trabaja con variables muy específicas que pueden ser cuantificadas con la pretensión de obtener una mayor objetividad, así como la generalización de los resultados.

Pero la educación es un fenómeno complejo, y por consiguiente la investigación en educación también lo es. Esta complejidad dificulta el diseño de las investigaciones y ha supuesto que para alcanzar los diferentes objetivos planteados en un mismo estudio se haga necesario el empleo de diferentes metodologías dentro un mismo paradigma o incluso la combinación de metodologías provenientes de los paradigmas cuantitativo y cualitativo, constituyendo así una vía metodológica conocida bajo el nombre de Métodos Mixtos (Nunez Moscoso, 2017).

La investigación realizada propone el estudio de diferentes aspectos relacionados con la realidad educativa mediante la utilización de distintas metodologías de análisis cuantitativo.

Se divide la investigación en tres grandes fases: en primer lugar, se presenta un análisis secundario de las pruebas PISA centrado en la determinación de los factores contextuales asociados al rendimiento académico de los estudiantes en las tres competencias principales estudiadas en esta prueba: comprensión lectora, matemáticas y ciencias utilizando una metodología de análisis multinivel (Gaviria Soto & Castro Morera, 2005; Raudenbush & Bryk, 1992). A continuación, y partiendo de los resultados obtenidos, se seleccionaron los centros educativos de alta y baja eficacia siguiendo el procedimiento empleado en investigaciones previas (Gamazo, 2019; Lizasoain & Angulo, 2014); para, finalmente, proceder con el estudio de aquellas variables procesuales significativamente relacionadas con la eficacia escolar según los datos obtenidos en PISA utilizando la regresión logística binaria. De acuerdo con lo planteado por Sabariego Puig “existe la necesidad de contar con un pluralismo de enfoques que combinen datos, métodos y técnicas de investigación (...) para ampliar nuestra forma de pensar sobre temas tan complejos como los educativos” (2004, p. 78), un pluralismo de enfoques como el planteado en este estudio.

En la siguiente figura se observan las tres fases principales que componen este estudio (figura 4.1), proporcionando información sobre los objetivos específicos de cada una de ellas y las metodologías utilizadas:

Figura 4.1

Fases del estudio



Una vez finalizadas estas tres primeras fases y con el objetivo de profundizar en los factores asociados a la eficacia escolar y obtener una información diferente proveniente de los propios agentes que intervienen en el proceso educativo, se procedió a la elaboración y posterior validación de un cuestionario dirigido a los profesores en activo de los centros educativos de alta y baja eficacia sobre la eficacia auto percibida en su centro de trabajo.

4.1. Objetivos

El objetivo general de esta investigación consiste en generar un marco común de buenas prácticas educativas relacionadas con la eficacia de las escuelas a nivel internacional. Para alcanzarlo, se plantean una serie de objetivos específicos que son:

1. Seleccionar las variables de contexto relacionadas con el rendimiento académico de los estudiantes españoles, portugueses e irlandeses participantes en PISA 2018.
2. Detectar las escuelas de alta y baja eficacia a partir de los resultados de PISA 2018 en España, Portugal e Irlanda.

3. Generar un modelo de análisis de la eficacia escolar que sea replicable en un contexto internacional.
4. Obtener un marco de referencia sobre las buenas prácticas realizadas en los centros de alta eficacia para poder replicarlas en los centros con una eficacia menor.

Para la consecución de los objetivos se hace necesario el planteamiento de metodologías de investigación cuantitativa de diferente índole de acuerdo con lo expuesto previamente.

4.2. Diseño

Esta investigación presenta un diseño no experimental *ex post facto* (Bisquerra, 2004) al tratarse de un análisis secundario de los datos proporcionados por las pruebas PISA en su edición del año 2018, análisis por tanto realizado a posteriori de la recogida de datos. Como se trata de una medición realizada en un momento determinado en el tiempo, se podría caracterizar también este estudio como transversal (Bisquerra, 2004; Cvetkovic-Vega et al., 2021).

4.2.1. Población y muestra

La muestra del estudio aquí presentado está formada por aquellos estudiantes que participaron en PISA en la edición del año 2018 escolarizados en algún centro educativo perteneciente a España, Portugal o Irlanda. La selección de estos países vino determinada por su similitud en cuanto a sus características socioeconómicas y por formar parte del grupo de países del sur de Europa denominados en la jerga financiera y económica como PIIGS (Portugal, Italia, Irlanda, Grecia y España) a raíz de las dificultades que tuvieron para la recuperación económica tras la crisis de 2008 (Bellod Redondo, 2016).

Concretamente, la muestra española la componían 35943 estudiantes pertenecientes a 1089 centros educativos, así como 21621 profesores. La muestra irlandesa la formaban 5577 estudiantes de 157 centros educativos. Por último, la muestra portuguesa estaba formada por 5932 estudiantes de 276 centros educativos y 5452 profesores.

Siguiendo el procedimiento empleado por diversos autores en estudios similares, se eliminaron del análisis aquellos centros educativos con menos de 20 estudiantes evaluados (Liu et al., 2014; Martínez-Abad et al., 2020) para así poder “asegurar un

correcto análisis de las variables agrupadas a nivel de centro” (Gamazo, 2019, p. 135). Esto supuso que finalmente la muestra del estudio la conformaran 44650 estudiantes (34411 españoles, 5551 irlandeses y 4688 portugueses), 1309 centros educativos (976 españoles, 155 irlandeses y 178 portugueses) y 24052 profesores (19991 españoles y 4061 portugueses).

Centrándonos en el caso español, y siguiendo con la trayectoria iniciada en la edición anterior, se dispone de datos representativos a nivel de todas las comunidades autónomas, permitiendo la comparación estadística de los resultados entre el alumnado de las diferentes regiones. Esta distribución de los estudiantes y centros por comunidades autónomas se puede observar en la Tabla 4.1, en la que se excluyeron los datos relativos a los centros educativos que participaron con menos de 20 estudiantes, de acuerdo con el criterio indicado anteriormente.

Tabla 4.1

Distribución de la muestra por Comunidades Autónomas

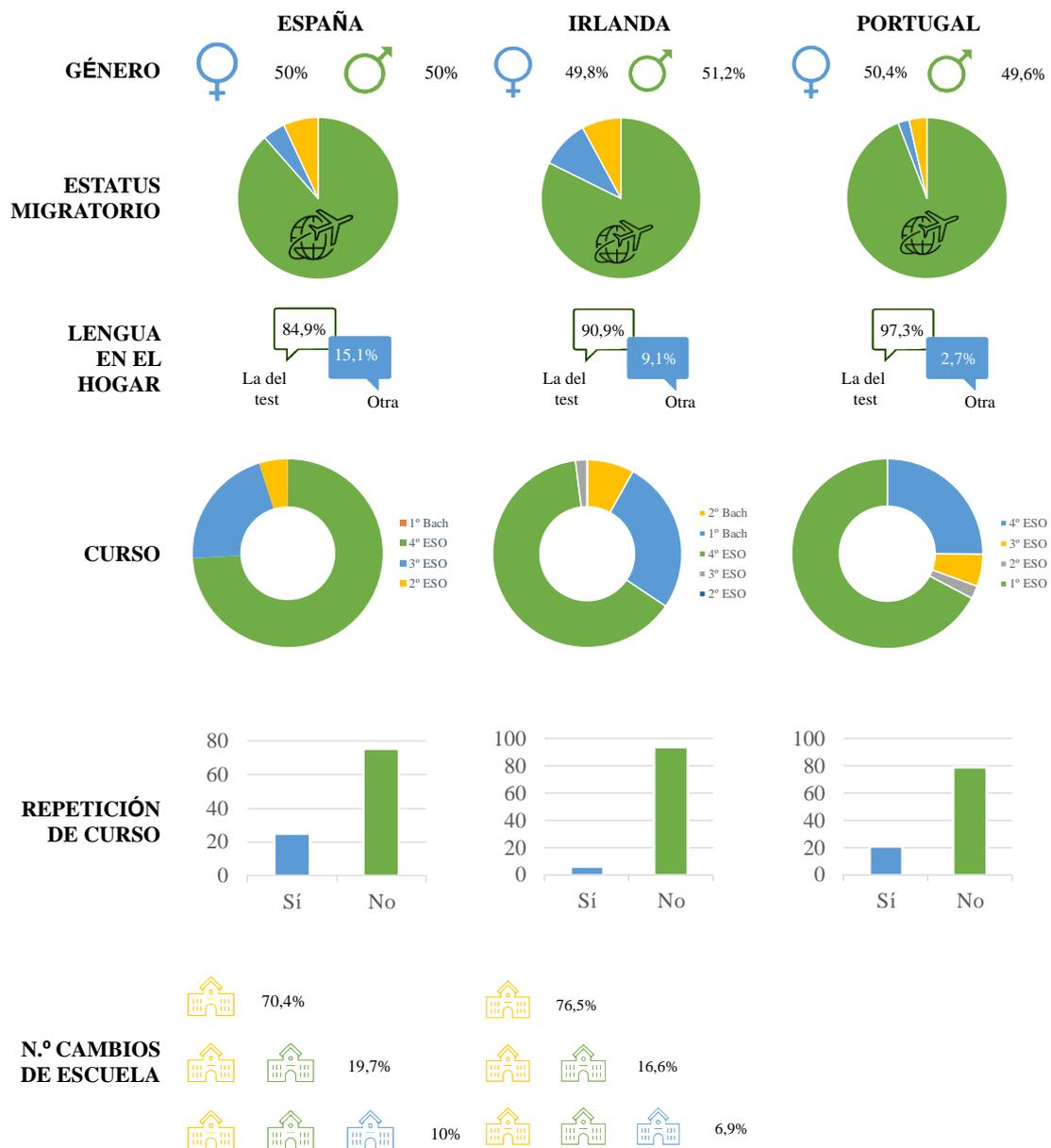
CC AA	Estudiantes		Centros	
	n	%	n	%
Andalucía	1732	5.033	50	5.123
Aragón	1753	5.094	49	5.020
Asturias	1825	5.304	50	5.123
Cantabria	1768	5.137	49	5.020
Castilla -La Mancha	1818	5.283	52	5.328
Castilla y León	1775	5.159	51	5.225
Cataluña	1626	4.725	46	4.713
Ceuta	370	1.075	11	1.127
Extremadura	1770	5.144	51	5.225
Galicia	1800	5.231	50	5.123
Islas Baleares	1632	4.742	47	4.816
Islas Canarias	1749	5.083	51	5.225
La Rioja	1427	4.147	40	4.098
Madrid	4926	14.315	137	14.037

CC AA	Estudiantes		Centros	
	n	%	n	%
Melilla	279	0.811	8	0.820
Murcia	1623	4.717	48	4.918
Navarra	1665	4.839	45	4.611
País Vasco	3207	9.32	94	9.631
Comunidad Valenciana	1666	4.841	47	4.816

Las características de la muestra de estudiantes, tomando como referencia diferentes variables como el curso del alumno, la lengua hablada en el hogar, el estatus migratorio, el género, la repetición de curso o el número de cambios de escuela (todas ellas variables contextuales referidas a las características de los alumnos), son las siguientes en cada uno de los tres países estudiados (Figura 4.2).

Figura 4.2

Características de la muestra de estudiantes



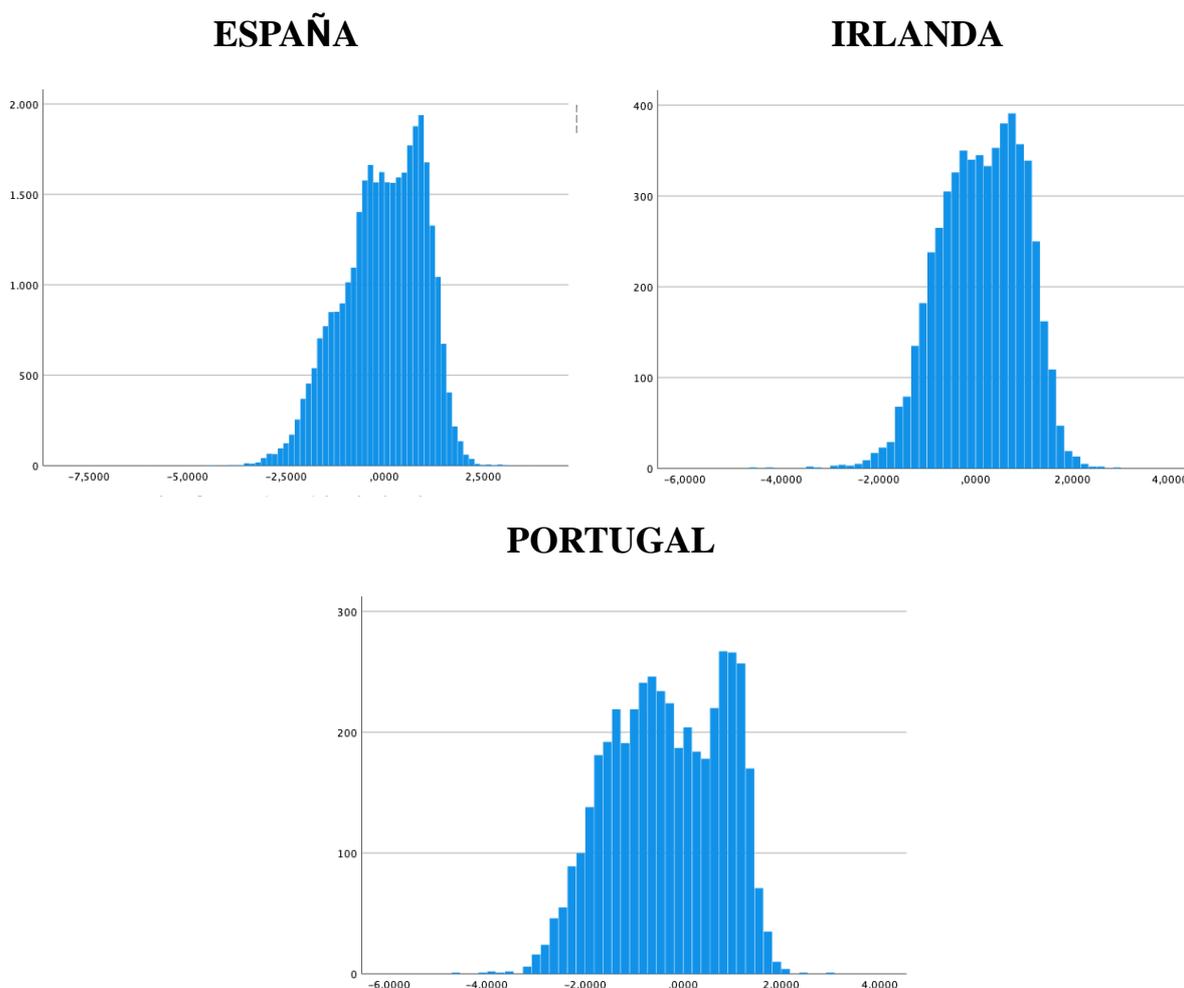
Nota. En la variable “Estatus Migratorio” el color verde se identifica con la proporción de estudiantes nativos, el color azul con los estudiantes considerados inmigrantes de primera generación y el color amarillo con los estudiantes inmigrantes de segunda generación.

Una de las variables que ha sido objeto de un mayor número de estudios es la referida al índice socioeconómico y cultural (ESCS). En la figura 4.3 se observa la distribución de esta variable en cada uno de los tres países estudiados según la información proporcionada por el estudiantado. En España el ESCS tiene una media de -0,34 y una

desviación típica de 1,03. En Irlanda aparece un ESCS medio de 0,13 con una desviación típica de 0,86. Mientras que en Portugal estos valores son de -0,33 de media y 1,14 de desviación típica.

Figura 4.3

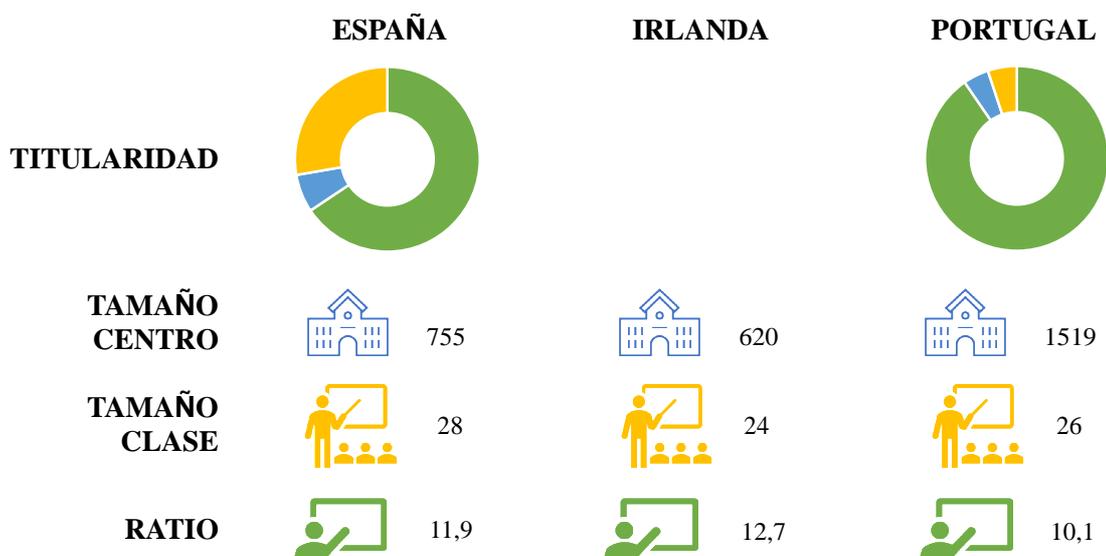
Distribución de la variable Nivel socioeconómico y cultural



En cuanto a las características de los centros educativos que conforman esta muestra (Figura 4.4), se observa que los centros son mayoritariamente públicos (no se dispone de información referida a esta variable en el caso irlandés), con clases con tamaños medios comprendidos entre 24 y 28 estudiantes y un tamaño medio de los centros de 755, 620 y 1519 alumnos aproximadamente en España, Irlanda y Portugal respectivamente. Además, la ratio profesor-estudiante promedio oscila entre los 10,16 estudiantes de Portugal hasta los 12,71 de Irlanda, pasando por los 11,88 de España.

Figura 4.4

Características de los centros educativos

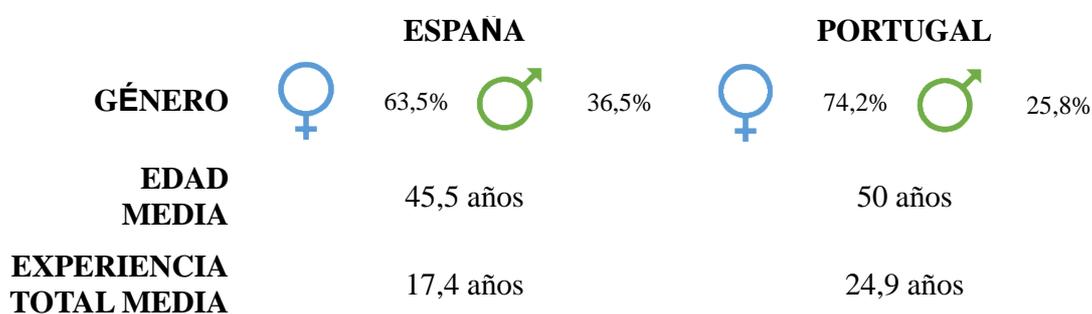


Nota. En la variable “Titularidad” el color verde indica titularidad pública, el color azul titularidad privada y el color amarillo se refiere a los centros concertados.

En el caso de España y Portugal, se dispone de información proporcionada por un tercer agente, el profesorado. Es por esto por lo que, para la determinación de los factores asociados al rendimiento académico, se han incluido algunas variables contextuales de los mismos. Un resumen de las características del profesorado de ambos países se puede consultar en la figura 4.5:

Figura 4.5

Características de la muestra de profesores



4.2.2. Instrumentos

Se utiliza información proveniente de dos de los instrumentos desarrollados por la OCDE para la obtención de datos en las pruebas PISA del año 2018. Concretamente, se trata de los cuestionarios para la evaluación de las competencias de los estudiantes en matemáticas, ciencias y comprensión lectora, y de los cuestionarios de contexto dirigidos a la recopilación de información no cognitiva, información sobre las características individuales tanto del entorno educativo como socioeconómico de estudiantes y centros y para la obtención de información relacionada con las características organizativas y procesuales del centro educativo (aplicados a alumnado y centros educativos).

4.2.3. Consideraciones metodológicas: el uso de las ponderaciones muestrales y de los valores plausibles

Una vez realizado un análisis exhaustivo del origen, concepto y evolución de las pruebas PISA, así como su construcción (Capítulo 1), se hace necesaria la consideración de una serie de elementos propios del diseño metodológico de PISA que garanticen tanto la representatividad de la muestra como la relevancia de los resultados obtenidos.

El tipo de muestreo aplicado en PISA es el muestreo estratificado bi-etápico según el cual en un primer momento se seleccionan los centros educativos de entre aquellos susceptibles de participar en el estudio teniendo en cuenta características como la ubicación en la que se encuentre o la titularidad. En esta primera etapa deben seleccionarse al menos 150 centros educativos por país o economía participante. A continuación, se lleva a cabo la selección aleatoria de al menos 42 estudiantes por centro que son quienes deberán responder a las pruebas (Schleicher, 2019).

Para garantizar la representatividad de los datos obtenidos, se hace necesario emplear las ponderaciones muestrales en los análisis derivados de los resultados de las pruebas PISA. Las ponderaciones o pesos muestrales son, en esencia, una ponderación asignada a cada estudiante y a cada centro para representar a la cohorte elegible en la edición de PISA correspondiente. Como se plantea en el Manual de Análisis de Datos de PISA (OECD, 2009), la ponderación consiste en reconocer que algunas unidades de la muestra son más importantes que otras y deben contribuir más que las restantes al cálculo de cualquier estimación poblacional.

Por otro lado, los valores plausibles reflejan el rango de habilidades que un estudiante podría tener razonablemente según sus respuestas a los ítems (Wu, 2005), debido a que

los alumnos no responden a la totalidad de las preguntas formuladas en las pruebas competenciales. En consecuencia, sirven para facilitar el abordaje de problemas relacionados con los sesgos derivados de la estimación de los parámetros poblacionales. En esta edición se generaron un total de 10 valores plausibles para cada una de las competencias estudiadas, valores que han sido tenidos en cuenta en los análisis estadísticos realizados.

4.2.4. Variables

Esta investigación se construye en torno a diferentes fases bien diferenciadas. Cada una de ellas, por la naturaleza del objetivo a alcanzar y la metodología a utilizar, supone el empleo de diferentes tipos de variables.

Se presenta un estudio de carácter no experimental *ex post facto*, con lo cual se trabajará con variables predictoras y criterio.

El primero de los análisis presentados toma como variable criterio el nivel competencial de los estudiantes (españoles, irlandeses y portugueses) en cada una de las tres competencias estudiadas; es decir, los resultados de los alumnos en las pruebas PISA, lo que es denominado en la investigación educativa como rendimiento académico. Estas puntuaciones se construyen en torno a los 10 valores plausibles generados para aumentar la precisión de las estimaciones de los diferentes niveles de competencia empleados en este análisis.

Como variables predictoras se dispone, en este primer momento, de aquellas variables contextuales referidas a los dos niveles estudiados, alumnos y centro educativo, que podrían presentar una relación de algún tipo con las variables criterio y que han sido extraídas de los diferentes cuestionarios de contexto de PISA 2018. La mayor parte de estas variables derivan de un único ítem, sin embargo, se han incluido en la tabla otro tipo de variables generadas según la información de diferentes ítems que se construyen como índices: el índice del nivel socioeconómico y cultural (ESCS), la escasez de personal en el centro educativo (STAFFSHORT) y la escasez de recursos educativos en el centro (EDUSHORT).

En el segundo nivel de análisis se han añadido a las variables propias de los centros educativos otras referidas a las características de los estudiantes y profesores de cada uno de ellos y que, para poder ser incorporadas a este segundo nivel, han sido agregadas en función del centro educativo de referencia. Estas variables son: género, edad, formación

y experiencia del profesorado del centro, y género, nivel socioeconómico, repetición de curso y estatus migratorio de los estudiantes del centro.

En la tabla 4.2 se presentan las diferentes variables contextuales, así como sus posibles opciones de respuesta:

Tabla 4.2

Variables contextuales

	Variable	Etiqueta	Valores
Nivel 1 - Alumnos	Índice del nivel socioeconómico y cultural	ESCS	Continua {-6,07 – 3,72}
	Género	GENDER	0: Masculino 1: Femenino
	Edad	AGE	Continua
	Curso	GRADE	1º ESO – 1º Bachillerato
	Estatus migratorio	IMMIG	0: Nativo, 1: Inmigrante de 1ª generación 2: Inmigrante de 2ª generación
	Condición de repetidor	REPEAT	0: No 1: Sí
	Número de cambios de centro educativo	SCCHANGE	0: Sin cambios 1: Uno 2: Dos o más cambios
	Número de años cursados en Educación Infantil	DURECEC	Continua {<1 año – 8 años}
	Lengua hablada en el hogar	LANG	0: Idioma de la prueba 1: Otro idioma
Nivel 2 - Centros	Tamaño del centro	SCHSIZE	Continua {37 – 2698}
	Tamaño de las clases	CLSSIZE	Continua {13 – 53}
	Escasez de recursos	EDUSHORT	Continua {-1,42 – 2,96}
	Escasez de personal	STAFFSHORT	Continua {-1,45 – 4,04}
	Titularidad del centro	SCHOWN	1: Privado 2: Concertado 3: Público

Variable	Etiqueta	Valores
Ratio profesor- alumnos	RATIO	Continua {1 – 51,58}
Localización del centro	LOCAT	1: Área rural (menos de 3000 habitantes) 2: Pueblo pequeño (3000 – 15000 habitantes) 3: Pueblo (15000 – 100000 habitantes) 4: Ciudad (100000 – 1000000 habitantes) 5: Ciudad grande (Más de 1000000 habitantes)
Proporción de docentes con Máster	PROAT5AM	Continua {0 - 1}
Proporción de docentes con Doctorado	PROAT6	Continua {0 - 1}
Género del profesorado*	TCH_GEN	Continua {0 – 3}
Edad del profesorado*	TCH_AGE	Continua {33,79 – 55,55}
Experiencia del profesorado en ese centro*	TCH_EXPSCH	Continua {1 – 39}
Experiencia total del profesorado*	TCH_EXPTOT	Continua {6,78 – 39}
Formación del profesorado*	OTT1	Continua {0 – 1}
Formación continua del profesorado*	OTT2	Continua {0 – 1}
Género medio de los estudiantes*	N2GEN	Continua {1 – 2}
ESCS medio de los estudiantes*	N2ESCS	Continua {-1,69 – 1,46}
Proporción de estudiantes repetidores*	N2REPEAT	Continua {0 – 1}
Proporción de alumnado inmigrante*	N2INMIG	Continua {1 – 2,43}

A partir de los resultados de esta primera fase de análisis, se generó una variable referida a la eficacia escolar que caracterizaba a los centros educativos como de alta o baja eficacia, o centros promedio, teniendo en cuenta la diferencia entre las puntuaciones esperadas en cada centro educativo según sus características contextuales y las puntuaciones obtenidas por sus estudiantes.

En aras de alcanzar el tercer objetivo específico, se modificaron las variables de trabajo. En este caso, la variable criterio es esa variable referida a la eficacia escolar, mientras que las variables predictoras son una serie de variables no contextuales extraídas de las diferentes bases de datos de PISA de estudiantes, profesorado y centros (que pueden ser consultadas en el Anexo 2). Cabe destacar que, como en la primera fase del estudio cuantitativo, las variables referidas al estudiantado y al profesorado se agregaron según el centro educativo para proceder con el análisis correspondiente. Estas variables están relacionadas con aspectos como: las actividades extraescolares, el comportamiento de estudiantes y profesorado, la opinión del director sobre las creencias multiculturales e igualitarias del profesorado, la opinión de los profesores sobre la escasez de recursos y de personal, la colaboración entre los profesores, su coordinación para la enseñanza, la satisfacción del profesorado con su profesión y con su entorno laboral, diferentes aspectos sobre la percepción de su autoeficacia, el uso de las TIC, el clima disciplinario en las aulas, la formación del profesorado, la evaluación de los estudiantes, la formación del profesorado, etc., así como otras muchas variables de proceso construidas en base a las opiniones del estudiantado.

4.2.5. Procedimiento y análisis de datos

Con el fin de detectar los factores contextuales relacionados con el rendimiento académico de los estudiantes pertenecientes a los tres países europeos seleccionados al finalizar la educación secundaria, se ha empleado un análisis de regresión multinivel, también conocido como Modelos Jerárquicos Lineales (Raudenbush & Bryk, 1992; Snijders & Bosker, 2011), gracias al cual se ha diseñado un modelo estadístico para cada una de las tres competencias estudiadas en cada uno de los tres países en el que se conjugaron las variables que presentaban una relación significativa con el rendimiento académico en esa competencia en particular ($\alpha = .05$). La elección de esta técnica para el análisis de los datos permitió considerar el efecto de variables contextuales pertenecientes a diferentes niveles respetando la estructura jerárquica de los datos provenientes de PISA,

facilitando, de este modo, la eliminación de sesgos provenientes del uso de otro tipo de técnicas de agregación o desagregación de datos (Gaviria Soto & Castro Morera, 2005; Raudenbush & Bryk, 1992).

Estos modelos sirvieron como base para el cálculo de la diferencia entre la puntuación real obtenida en cada centro educativo y la estimada según sus características contextuales, es decir, aquellas variables que resultaron significativas en los modelos estadísticos finales. Al resultado de esta diferencia se le denomina residuo (Lizasoain et al., 2015).

Siguiendo estudios anteriores que demostraban la no existencia de diferencias sustanciales entre los modelos de pendientes fijas y aleatorias (Gamazo, 2019), se introdujeron las variables contextuales en el primer nivel como covariables de efectos fijos, sin incluir los efectos aleatorios. Además, de acuerdo con las recomendaciones de la OCDE se incluyeron en todos los modelos los 10 valores plausibles proporcionados en las bases de datos referidas a los resultados educativos, así como los pesos muestrales de los estudiantes y los centros.

Utilizando los residuos de las tres competencias y un protocolo similar al planteado por Gamazo (2019) y Lizasoain y Angulo (2014) se procedió a la selección de los centros. De este modo, se ordenaron los centros educativos en función de los valores de los correspondientes residuos para cada una de las competencias estudiadas en una escala ascendente. Se entiende entonces que, en el primer puesto se sitúa el centro que obtiene una puntuación residual muy por debajo de lo que cabría esperar según sus características contextuales, no la escuela con los resultados más bajos. Esto mismo se aplica al caso opuesto, situándose en el percentil 100 aquel centro en el que los estudiantes hayan rendido muy por encima de lo esperado. A partir de esta caracterización, aquellos centros educativos situados entre los percentiles 100 y 66 en matemáticas, ciencias y comprensión lectora se caracterizaron como centros de alta eficacia, mientras que los situados entre los percentiles 33 y 1 en todas las competencias fueron considerados de baja eficacia. A partir de esta clasificación, se generó una nueva variable dicotómica según la cual se caracterizaban los centros educativos como de alta o baja eficacia.

Posteriormente, con el fin de comprobar la relación entre las diferentes variables de proceso elegidas y esta nueva variable generada, alta o baja eficacia escolar, se procedió al cálculo de la correlación de Pearson. En caso de que el resultado de este coeficiente indicase la existencia de correlaciones significativas ($\alpha = .05$), se analizaría esta

correlación mediante técnicas de regresión logística binaria para poder así determinar las relaciones existentes entre las distintas variables de proceso y la eficacia escolar.

En este análisis cuantitativo se emplearon diferentes softwares estadísticos. Por un lado, para la elaboración de los Modelos Jerárquicos Lineales se utilizó el programa HLM7, mientras que para el cálculo de las correlaciones y de la regresión logística se empleó SPSS en su versión 23.

4.3. Diseño de un cuestionario para la evaluación de la eficacia auto percibida por el profesorado de secundaria

Al inicio de esta sección se planteaba la creación de un instrumento de recogida de información dirigido al profesorado en activo de educación secundaria de los centros educativos considerados de alta y baja eficacia. Se muestra a continuación cómo ha sido su diseño.

Con el fin de conocer las percepciones del profesorado sobre las condiciones contextuales que favorecen el aprendizaje del estudiantado y así valorar la eficacia auto percibida por el profesorado en sus centros a partir tanto de los resultados obtenidos en la regresión logística desarrollada con anterioridad (factores procesuales relacionados con la eficacia escolar), como de la revisión de la literatura relacionada con esta temática, se establecieron una serie de categorías de información relacionadas con algunos de esos factores que fueron posteriormente concretados en diferentes ítems. Estas categorías de información fueron, en un primer momento, las siguientes:

1. Proyectos, planes y formación.
2. Metodologías y materiales de enseñanza.
3. Atención a la diversidad y seguimiento de los estudiantes.
4. Evaluación del alumnado y gestión del tiempo.
5. Liderazgo y gestión del centro.
6. Coordinación, implicación, clima y convivencia.
7. Evaluación de los docentes, del centro y de programas.
8. Infraestructura y proyección del centro en el entorno.

Definidas esas 8 primeras dimensiones, se plantearon los ítems que compusieron el primer borrador del cuestionario. En la Tabla 4.3 se puede observar la codificación

realizada para los ítems y dimensiones del cuestionario enviado para su validación a los jueces:

Tabla 4.3

Codificación de las dimensiones y los ítems

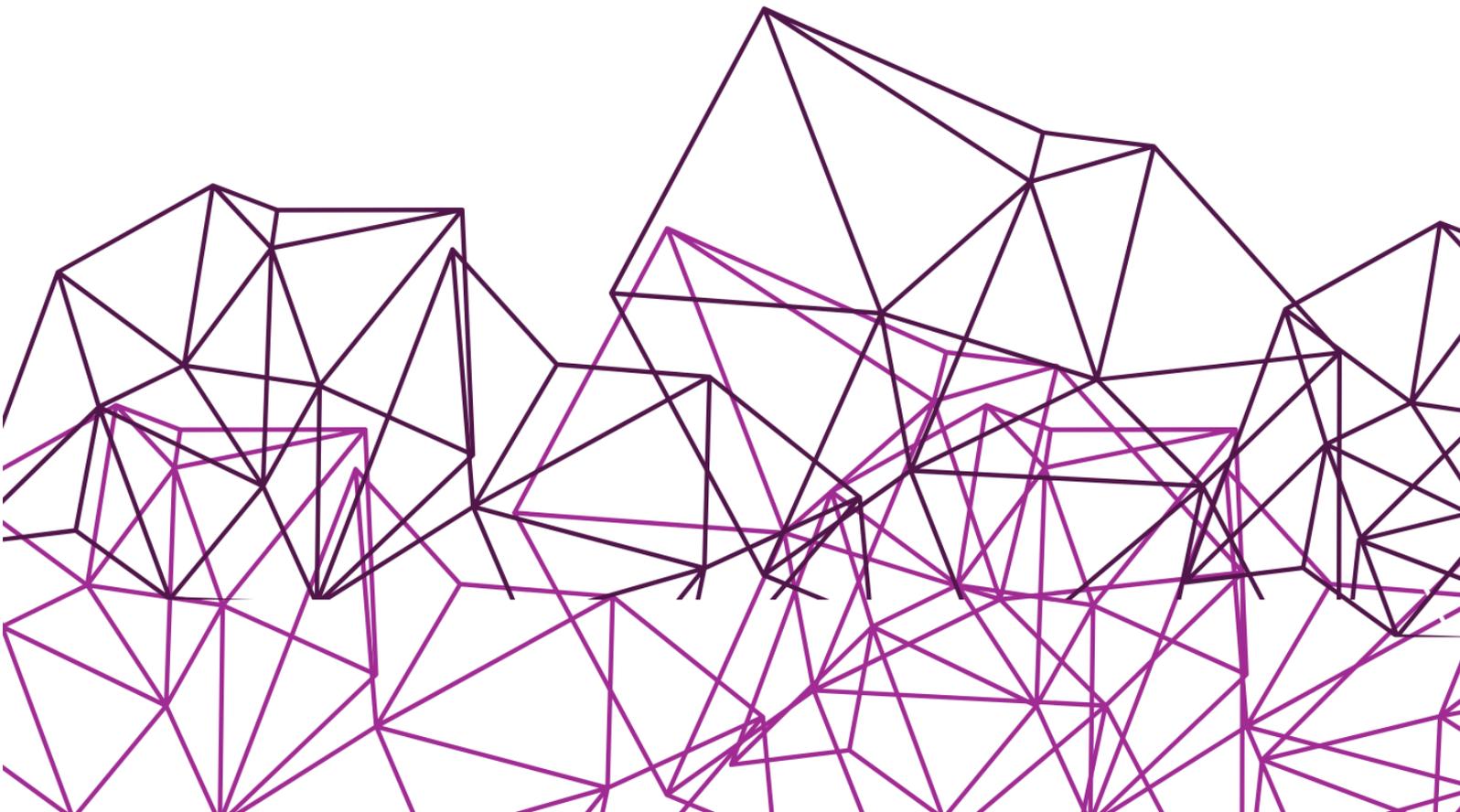
CÓDIGO	EQUIVALENCIA
D1	<u>Proyectos, planes y formación</u>
1.1	1. Este centro educativo diseña y desarrolla habitualmente proyectos educativos de interés para toda la comunidad educativa.
1.2	2. Este centro presta especial atención a la formación permanente y al desarrollo profesional del profesorado para la mejora del centro.
1.3	3. Las actividades de formación docente ofertadas en el centro tienen en cuenta las necesidades formativas expresadas por el profesorado.
1.4	4. Los proyectos de innovación educativa iniciados en el centro suelen cumplir con la duración y los objetivos establecidos en su comienzo.
D2	<u>Metodologías y materiales de enseñanza</u>
2.1	1. La mayoría de los profesores del centro utilizamos una metodología didáctica basada en la clase magistral y el uso del libro de texto.
2.2	2. El centro dedica especial atención a trabajar la competencia matemática y/o científica.
2.3	3. La mayoría de los profesores utilizamos habitualmente tecnologías digitales para apoyar el aprendizaje de los estudiantes.
2.4	4. La comprensión escrita y expresión oral son las competencias principales a las que atendemos los profesores de este centro.
D3	<u>Atención a la diversidad y seguimiento de los estudiantes</u>
3.1	1. El claustro del centro es sensible y aplica medidas de refuerzo y apoyo a los alumnos con necesidades educativas especiales
3.2	2. En general, el profesorado utiliza metodologías diferenciales para atender a la diversidad en las aulas.
3.3	3. El plan de orientación del centro se elabora conjuntamente entre los distintos implicados (orientador, profesores y familias).
D4	<u>Evaluación del alumnado y gestión del tiempo</u>
4.1	1. En el centro se presta especial atención a la evaluación de los alumnos como proceso educativo, más que sumativo.
4.2	2. En general, los profesores compartimos criterios comunes en el control del rendimiento de los alumnos en un mismo grupo.
4.3	3. El centro participa habitualmente en Evaluaciones Externas (PISA, de nivel, etc.)
4.4	4. En el centro se aplican sistemas de garantía de calidad, reconocidos externamente.

CÓDIGO	EQUIVALENCIA
4.5	5. En este centro se insiste en la puntualidad y en el respeto al horario.
D5	<u>Liderazgo y gestión del centro</u>
5.1	1. El modo de actuar de este equipo directivo es más democrático que jerárquico.
5.2	2. Considero que el tipo de liderazgo del equipo directivo tiene una gran influencia en el clima del centro.
5.3	3. Nuestro equipo directivo ha planteado unas metas claras y conoce el funcionamiento del centro.
5.4	4. Generalmente, en este centro el equipo apoya al profesorado implicándole en las decisiones de organización.
5.5	5. En este centro, en general, la plantilla de profesorado es estable.
5.6	6. La organización del centro permite dedicar tiempo a la formación permanente.
D6	<u>Coordinación, implicación, clima y convivencia</u>
6.1	1. Es un centro caracterizado por un buen clima de trabajo.
6.2	2. El centro dispone de un plan de convivencia que se aplica y funciona eficazmente.
6.3	3. Habitualmente, no existen conflictos graves de relación entre el personal del centro.
6.4	4. Las reuniones de coordinación son planificadas con anterioridad, planteando los objetivos a conseguir en cada caso.
6.5	5. Como profesor, me identifico con y creo en el proyecto del centro.
6.6	6. La dirección del centro fomenta la participación, la implicación y el compromiso del profesorado y del resto de la comunidad educativa en la gestión total del centro.
D7	<u>Evaluación de los docentes, del centro y de programas</u>
7.1	1. En este centro, habitualmente, compartimos experiencias sobre la mejora de la práctica docente en el aula
7.2	2. Además, los profesores hacemos el seguimiento de los progresos y los cambios que introducimos
7.3	3. El equipo de docentes, a partir de los resultados de las evaluaciones de los estudiantes y otras evidencias, elaboramos planes de mejora.
D8	<u>Infraestructura y proyección del centro en el entorno</u>
8.1	1. Las instalaciones e infraestructuras del centro posibilitan un buen desarrollo de la labor educativa.
8.2	2. La imagen del centro ante la comunidad próxima es óptima.
8.3	3. Como profesor, percibo que se necesitaría una mayor implicación de las familias del alumnado para conseguir una mejora del centro educativo.
8.4	4. En este centro se ha establecido una comunicación eficaz con toda la comunidad educativa.
OTROS	<u>Ítems cualitativos del cuestionario</u>

CÓDIGO	EQUIVALENCIA
O1	1. Como profesor, considero que las condiciones en las que este centro desarrolla su proyecto educativo tienen una influencia significativa en la mejora del aprendizaje de todo su alumnado, contribuyendo así a la equidad educativa.
O2	2. A modo de RESUMEN u OBSERVACIONES: Por favor, exponga en un breve texto, cómo cree que es considerado su centro por la comunidad en la que está asentado y por qué razones.

Con el fin de aplicar el cuestionario, se llevó a cabo la oportuna validación. Los resultados de esta validación se incluyen en la sección 5.7.

5. RESULTADOS



Resultados

En esta sección, se presentan los modelos jerárquicos lineales elaborados para analizar el impacto de las variables contextuales en el rendimiento académico de los estudiantes españoles, irlandeses y portugueses. Se seleccionan los centros educativos de alta y baja eficacia en cada uno de estos países y se analiza el efecto de las variables procesuales en la eficacia escolar utilizando diversas técnicas estadísticas. Además, se incluye un apartado dedicado a presentar la validación del instrumento de recogida de información elaborado como parte de esta investigación y dirigido al profesorado.

Para ello, se ha organizado este capítulo en siete secciones. En las tres primeras se aborda la construcción de los modelos multinivel para cada competencia en cada uno de los tres países estudiados hasta alcanzar los correspondientes modelos finales. De este modo, se pueden observar las variables asociadas al rendimiento académico de los estudiantes en cada país. En la cuarta sección se presenta un análisis comparativo de los resultados obtenidos resaltando sus semejanzas y diferencias para cada competencia estudiada en cada país y que se ha caracterizado como análisis comparado. La quinta sección se centra en la selección de los centros de alta y baja eficacia partiendo de los modelos multinivel finales y teniendo en cuenta los residuos. A continuación, en la sección seis se analizan los efectos de las variables de proceso en la eficacia escolar de los centros españoles, irlandeses y portugueses de manera independiente. Por último, se incluyen los resultados del proceso de validación del cuestionario.

Para conocer los factores de contexto asociados al rendimiento académico de los estudiantes se han analizado los resultados de las tres competencias principales estudiadas en PISA: matemáticas, ciencias y comprensión lectora. Ya se ha mencionado y explicado en capítulos anteriores el uso de los 10 valores plausibles propuestos por la OCDE para cada competencia como variable criterio de este análisis. Del mismo modo, los pesos muestrales de cada uno de los niveles de análisis han sido tenidos en cuenta para los cálculos presentados en este apartado centrados en la construcción de los diferentes modelos jerárquico-lineales. Se sigue, por tanto, un procedimiento basado en el planteado en estudios similares (Gamazo, 2019) de acuerdo con el siguiente esquema:

1. Cálculo del modelo nulo y de su correspondiente Coeficiente de Correlación Intraclase (CCI) para así comprobar el supuesto de dependencia, es decir, la existencia de variación en las puntuaciones debida a la existencia del segundo nivel de análisis.

2. Comprobación de la colinealidad para adecuar la selección de las variables a incluir en el modelo condicional, garantizando la ausencia de relaciones importantes estadísticamente entre pares de variables.
3. Cálculo del modelo condicional mediante una modelización con diseño mixto donde las pendientes son fijas y el intercepto es aleatorio y presentación del modelo estadístico final. El hecho de que el intercepto sea aleatorio permite que la media de la variable dependiente pueda tener valores diferentes en función del centro educativo al que se refiere. Además, que los valores de las pendientes sean fijos supone que los efectos de las covariables incluidas en el modelo se mantienen fijos independientemente del centro educativo.
4. Cálculo del estadístico *Pseudo R*² para demostrar la proporción de varianza explicada por el modelo condicional final en cuanto a los dos niveles estudiados.
5. Comprobación de los supuestos relacionados con los residuos de ambos niveles del modelo final: distribución normal, media 0 y homocedasticidad.

5.1. Análisis de los factores contextuales del rendimiento académico de los estudiantes españoles

Se comienza, por tanto, presentando la elaboración de los modelos multinivel para las tres competencias en España. Cabe destacar que, de las 19 variables de segundo nivel seleccionadas inicialmente para la realización del análisis multinivel⁶, la proporción de profesorado con estudios de Máster (PROAT5AM) presentaba un alto porcentaje de valores perdidos (28,2%), por lo que, de acuerdo con el criterio adoptado por Gamazo y Martínez-Abad (2020) según el cual las variables con un porcentaje de valores perdidos igual o superior al 20% deben ser suprimidas del análisis, se procedió a su eliminación.

5.1.1. Consideraciones previas

La correcta aplicación de los modelos multinivel toma como base el cumplimiento de una serie de requisitos estadísticos previos relacionados con la colinealidad existente entre las diferentes variables de estudio a fin de no incluir variables estrechamente relacionadas en un mismo modelo, como se ha indicado previamente. Se han definido con anterioridad

⁶ Que se pueden consultar en la tabla 4.2 del capítulo 4 de este estudio.

las variables de contexto a incluir en los distintos modelos y, puesto que son las mismas para las tres competencias estudiadas en este país, se incluyen a continuación como consideraciones previas.

Según el tipo de variables se ha calculado el índice de correlación correspondiente⁷ tanto para las variables de nivel 1 (Tabla 5.1) como para las variables de nivel 2 (Tabla 5.2):

Tabla 5.1

Correlaciones entre las variables contextuales a nivel de estudiante en España

	1	2	3	4	5	6	7	8	9
1. ESCS	-								
2. GENDER	0.029	-							
3. AGE	0.006	0.014	-						
4. GRADE	0.334	0.101	0.066	-					
5. IMMIG	-0.241	-0.001	-0.005	0.217	-				
6. REPEAT	-0.35	-0.096	-0.057	0.68	0.183	-			
7. SCCHANGE	-0.073	-0.025	-0.008	0.067	0.234	0.215	-		
8. DURECEC	0.095	0.035	0.024	-0.211	-0.105	-0.062	-0.043	-	
9. LANG	-0.085	-0.014	-0.002	0.058	0.26	0.048	0.101	-0.022	-

⁷ Coeficiente Fi para los pares de variables dicotómicas, correlación biserial-puntual para pares de variables dicotómicas y cuantitativas, coeficiente de contingencia para pares de variables dicotómica-politómica, politómica-politómica, dicotómica-ordinal y politómica-ordinal, coeficiente de Spearman para pares de variables ordinales y ordinal-cuantitativa, y coeficiente de Pearson para pares de variables cuantitativas.

Tabla 5.2

Correlaciones entre las variables contextuales a nivel de centro en España

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	
1. SCHSIZE	-																		
2. CLSSIZE	0.208	-																	
3. EDUSHORT	-0.122	-0.002	-																
4. STAFFSHORT	-0.009	0.007	0.523	-															
5. SCHOWN	0.138	0.062	-0.184	-0,183	-														
6. RATIO	0.545	0.22	-0.112	-0.02	0.033	-													
7. LOCAT	0.395	0.181	-0.081	0.011	0.281	0.371	-												
8. PROAT6	0.073	-0.86	0.051	0.011	-0.028	0.01	0.076	-											
9. TCH_GEN	0.009	-0.01	-0.006	0	0.023	0.023	-0.032	0.022	-										
10. TCH_AGE	0.031	-0.045	0.033	0.002	-0.028	0.028	0.022	0.032	0.058	-									
11. TCH_EXPSCH	0.011	-0.008	-0.013	-0.016	0.018	0.041	-0.035	-0.03	-0.1	0.174	-								
12. TCH_EXPTOT	0.059	-0.016	-0.002	-0.021	0.005	0.035	0.036	0.033	0.012	0.84	0.44	-							
13. TCH_OTT1	-0.032	0.013	0.067	-0.012	-0.02	-0.014	-0.016	0.004	-0.07	-0.136	-0.051	-0.192	-						
14. TCH_OTT2	-0.022	0.045	0.076	0.028	-0.056	0.027	-0.008	-0.002	-0.025	-0.048	-0.036	-0.072	0.786	-					
15. N2GEN	-0.046	-0.054	-0.047	-0.006	0.074	-0.029	-0.01	0.008	-0.052	0.052	0.026	0.068	0.019	0.029	-				
16. N2ESCS	0.022	0.049	-0.047	-0.003	-0.031	0.018	-0.04	0.011	0.026	-0.032	0.14	0.004	0.023	0.023	-0.04	-			
17. N2REPEAT	-0.058	0.021	0.002	-0.026	0.043	-0.018	0.045	-0.031	-0.044	0.028	-0.129	-0.021	-0.016	-0.019	0.146	-0.723	-		
18. N2INMIG	-0.054	-0.035	-0.025	0.012	0.003	-0.026	-0.008	-0.046	-0.031	0.075	-0.066	0.037	0.021	0.045	0.083	-0.372	0.44	-	

Como norma general no se observan variables con correlaciones excesivamente elevadas que impidan su inclusión en los análisis o dificulten los cálculos planteados, sobre todo en relación con las variables del primer nivel, donde la única correlación con un valor medio-alto es la existente entre el curso y el índice de repetición de los estudiantes. Centrándonos en las variables del segundo nivel, existe una relación alta o media-alta entre los siguientes pares de covariables: el tamaño de la clase y la proporción de profesores con doctorado en el centro, la edad de los profesores y su experiencia total como docentes, la formación del profesorado en sus dos vertientes (OTT1 y OTT2) y el nivel socioeconómico medio de los estudiantes del mismo centro y el valor medio de la repetición de curso. Teniendo en cuenta que las correlaciones calculadas no presentan valores demasiado elevados, se ha tomado la decisión de incluirlas en la construcción de los modelos, siendo fundamental la interpretación que se haga de los resultados en caso de la aparición de alguno de estos pares de variables en los modelos finales de una misma competencia.

5.1.2. Competencia matemática

Para el correcto cálculo del modelo multinivel final, es necesario presentar en primer lugar el modelo nulo, es decir el modelo incondicional que nos indique cuánto varían las puntuaciones en matemáticas a nivel de centro educativo. Este modelo nulo seguiría la Ecuación 5.1:

$$\text{Matemáticas}_{ij} = y_{00} + u_{0j} + r_{ij} \quad (\text{Ecuación 5.1})$$

Donde:

Matemáticas_{ij} se refiere a la puntuación media en matemáticas de cada centro;

y_{00} se refiere a la media de todos los centros españoles en matemáticas;

u_{0j} se refiere a la diferencia entre la puntuación del centro en matemáticas menos la puntuación media de todos los centros;

r_{ij} se refiere al residuo de nivel 1.

El modelo nulo permite el cálculo del Coeficiente de Correlación Intraclase (CCI) a partir de los resultados referidos a los componentes de la varianza de los dos niveles de análisis estudiados y que se pueden observar en la Tabla 5.3:

Tabla 5.3*Componentes de la varianza del modelo nulo de la competencia matemática en España*

Efecto aleatorio	σ	Componentes de la varianza	g. l.	χ^2	p-valor
Nivel 2	31.90600	1017.99298	975	6024.22339	<0.001
Nivel 1	82.04445	6731.29256			

A partir de los resultados de estos componentes de la varianza y aplicando la Ecuación 5.2 se calculó el CCI inicial, obteniendo el siguiente resultado:

$$CCI_{matemáticas} = \frac{\tau_{00}}{\tau_{00} + \delta^2} = \frac{1017,99298}{1017,99298 + 6731,29256} = 0,1314$$

(Ecuación 5.2)

Donde:

τ_{00} es el componente de la varianza de nivel 2, y

δ^2 es el componente de la varianza de nivel 1.

El resultado de este coeficiente indica que la variabilidad entre centros para el rendimiento en la competencia matemática de los estudiantes españoles es del 13,14%. Esto supone la existencia de dependencia entre las observaciones y admite la aplicación del análisis multinivel ya que, como se mencionó con anterioridad en el apartado dedicado a los modelos multinivel, el valor del CCI debe ser superior al 10% para considerar adecuada la aplicación de los modelos jerárquico lineales (V. E. Lee, 2000).

Tras esta comprobación, se procedió al ajuste del modelo. Para esto se incluyeron las variables de primer y segundo nivel eliminándolas de manera individual y consecutiva según el valor de su significación en el mismo hasta llegar al modelo final. Este modelo, compuesto por las variables que presentaban una relación significativa con los resultados en matemáticas, sigue un diseño de efectos mixtos con pendientes fijas e intercepto aleatorio.

Finalmente, el modelo estuvo compuesto por nueve variables, como puede observarse en la Ecuación 5.3:

Matemáticas_{ij}

$$\begin{aligned}
 &= \gamma_{00} + \gamma_{01} * SCHSIZE_j + \gamma_{02} * STAFFSHORT_j + \gamma_{03} * PROAT6_j \\
 &+ \gamma_{10} * ESCS_{ij} + \gamma_{20} * GENDER_{ij} + \gamma_{30} * REPEAT_{ij} + \gamma_{40} \\
 &* GRADE_{ij} + \gamma_{50} * AGE_{ij} + \gamma_{60} * SCCHANGE_{ij} + u_{0j} + r_{ij}
 \end{aligned}$$

(Ecuación 5.3)

Donde:

Matemáticas_{ij} se refiere a la puntuación media en matemáticas de cada centro;

γ_{00} se refiere a la media de todos los centros españoles en matemáticas;

$\gamma_{01} - \gamma_{03}$ son las covariables de nivel 2;

$\gamma_{10} - \gamma_{60}$ son las covariables de nivel 1;

u_{0j} es la diferencia entre la puntuación del centro en matemáticas menos la media de todos los centros;

r_{ij} se refiere al residuo de nivel 1.

En la Tabla 5.4 se pueden observar las variables que presentan una relación significativa con nuestra variable criterio y que conforman el modelo final para la competencia matemática.

Tabla 5.4

Modelo final competencia matemática en España

Efecto Fijo	Coefficiente	Error típico	t-ratio	g. l.	p-valor
Intercepto, γ_{00}	350.194552	43.211894	8.104	72	<0.001
SCHSIZE, γ_{01}	0.010999	0.002687	4.094	511	<0.001
STAFFSHORT, γ_{02}	-4.614822	1.356288	-3.403	320	<0.001
PROAT6, γ_{03}	56.384914	26.729524	2.109	971	0.035
ESCS, γ_{10}	13.236273	0.968905	13.661	34	<0.001
GENDER, γ_{20}	-17.004681	2.164621	-7.856	24	<0.001
REPEAT, γ_{30}	-57.323130	7.347867	-7.801	21	<0.001
GRADE, γ_{40}	25.908413	4.743159	5.462	38	<0.001
AGE, γ_{50}	10.444725	2.745175	3.805	66	<0.001
SCCHANGE, γ_{60}	-12.188733	1.845310	-6.605	19	<0.001

Según el valor de t-ratio de las variables, se observa que el índice del nivel socioeconómico y cultural de los estudiantes (ESCS) presenta un claro efecto positivo en el rendimiento en matemáticas. En el lado opuesto de la balanza se encuentran los efectos

negativos similares por parte de dos variables: el género de los estudiantes (GENDER) y la repetición de curso (REPEAT). En el caso de la variable género y puesto que se trata de una variable de tipo categórico, es preciso indicar que esta relación negativa con el rendimiento se produce cuando las estudiantes han indicado género femenino, suponiendo por tanto que el hecho de pertenecer a este grupo se asocia con la obtención de peores resultados en esta competencia.

A partir de los resultados obtenidos para la estimación de los componentes de la varianza del modelo final (Tabla 5.5), se ha calculado el estadístico *Pseudo R*² propuesto por Raudenbush y Bryk (2002), para interpretar la proporción de varianza explicada con este modelo condicional final en cada nivel.

Tabla 5.5

Componentes de la varianza del modelo final para la competencia matemática en España

Efecto aleatorio	σ	Componentes de la varianza	g.l.	χ^2	p-valor
Nivel 2	18.92874	358.29726	971	3093.74266	<0001
Nivel 1	69.80097	4872.17512			

Para el cálculo del estadístico *Pseudo R*² se tomaron como base los componentes de la varianza del modelo incondicional y del modelo estadístico final y se aplicaron las siguientes ecuaciones (Ecuación 5.4 y 5.5) expresadas siguiendo la notación de Hayes (2006):

$$Pseudo R^2_{Nivel1} = \frac{\delta^2_{Nulo} - \delta^2_{Final}}{\delta^2_{Nulo}} \quad \text{(Ecuación 5.4)}$$

$$Pseudo R^2_{Nivel2} = \frac{\tau_{00Nulo} - \tau_{00Final}}{\tau_{00Nulo}} \quad \text{(Ecuación 5.5)}$$

Donde:

δ^2 es la variabilidad entre las unidades de nivel 1; y

τ_{00} es la variabilidad entre las unidades de nivel 2.

En el caso de la competencia matemática en España, el porcentaje de la varianza explicada en el primer nivel de análisis fue de 27,62%. En cuanto al segundo nivel, el valor de *Pseudo R²* fue de 64,8%.

Una vez obtenido el modelo final, hay que proceder con la comprobación de los supuestos referidos a sus residuos, tanto del nivel 1 (Tabla 5.6) como del nivel 2 (Tabla 5.7). Estos supuestos son tres: que la distribución de los residuos sea normal (o se aproxime a este tipo de distribución), que su media sea 0 (o cercana) y que se cumpla el supuesto de homocedasticidad, es decir, que se pueda asumir la igualdad en la dispersión de las puntuaciones de los residuos.

Los residuos de nivel 1 para matemáticas presentan una media de aproximadamente 3 puntos en una escala con una amplitud de más de 700 puntos, pudiendo afirmar que la media se aproxima a 0, concretamente 2,86, junto con una desviación típica de 69,93.

Tabla 5.6

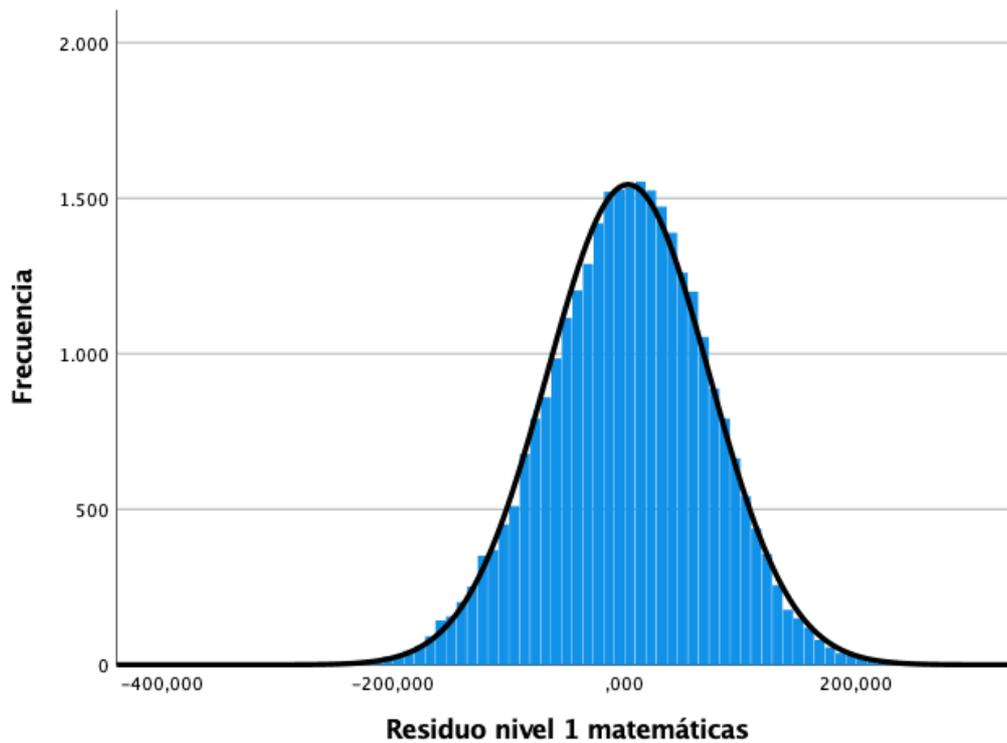
Estadísticos descriptivos para el residuo de nivel 1 en matemáticas en España

	N	Mín.	Máx.	Media	σ	Asimetría		Curtosis	
						Valor	Error típico	Valor	Error Típico
Residuo N1	29774	-348.09	273.92	2.86	69.93	-0.12	0.01	0.08	0.03

Según las puntuaciones del residuo, se puede determinar que se asemeja a la distribución normal (Figura 5.1) aunque presenta una asimetría negativa y un nivel de significación de la prueba de Kolmogorov-Smirnov <0,001.

Figura 5.1

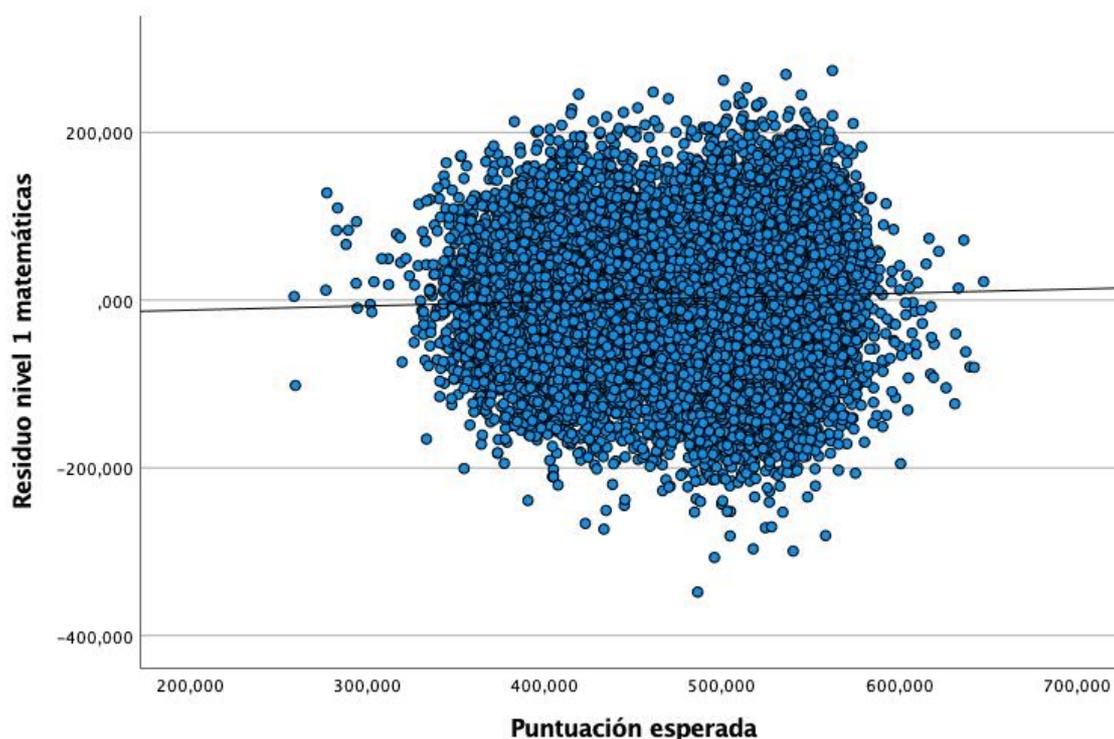
Distribución del residuo de nivel 1 de la competencia matemática en España



En cuanto al supuesto de homocedasticidad del residuo de este nivel, se ha comprobado analizando la dispersión de los valores según la puntuación esperada para cada estudiante. En este caso, como se observa en la Figura 5.2, este supuesto se cumple, obteniendo un valor de R^2 de 0,001.

Figura 5.2

Comprobación del supuesto de homocedasticidad del residuo de nivel 1 de la competencia matemática en España



A continuación, se presentan los estadísticos descriptivos para el residuo de nivel 2 (Tabla 5.7), donde se observa una media cercana a 0, concretamente con un valor de 1,63 y una desviación típica de 17,1.

Tabla 5.7

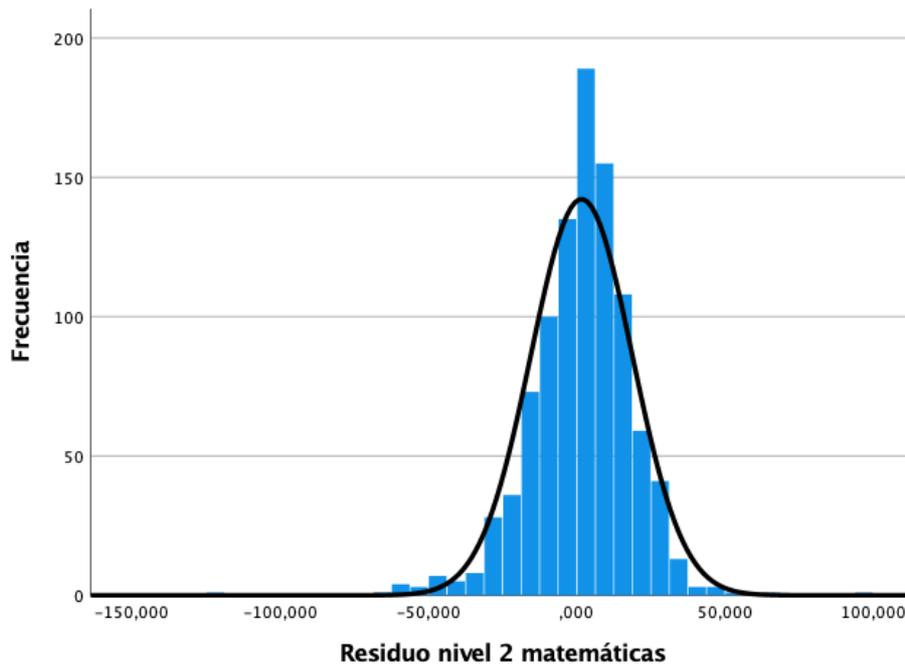
Estadísticos descriptivos para el residuo de nivel 2 en matemáticas en España

	N	Mín.	Máx.	Media	σ	Asimetría		Curtosis	
						Valor	Error típico	Valor	Error Típico
Residuo N2	975	-122.43	96.4	1.72	17.1	-0.66	0.08	4.55	0.16

La distribución seguida por estos residuos de segundo nivel se asemeja de nuevo a una distribución normal (Figura 5.3), obteniendo un valor de significación en la prueba de Kolmogorov-Smirnov $<0,001$, aunque con asimetría negativa y un marcado carácter leptocúrtico.

Figura 5.3

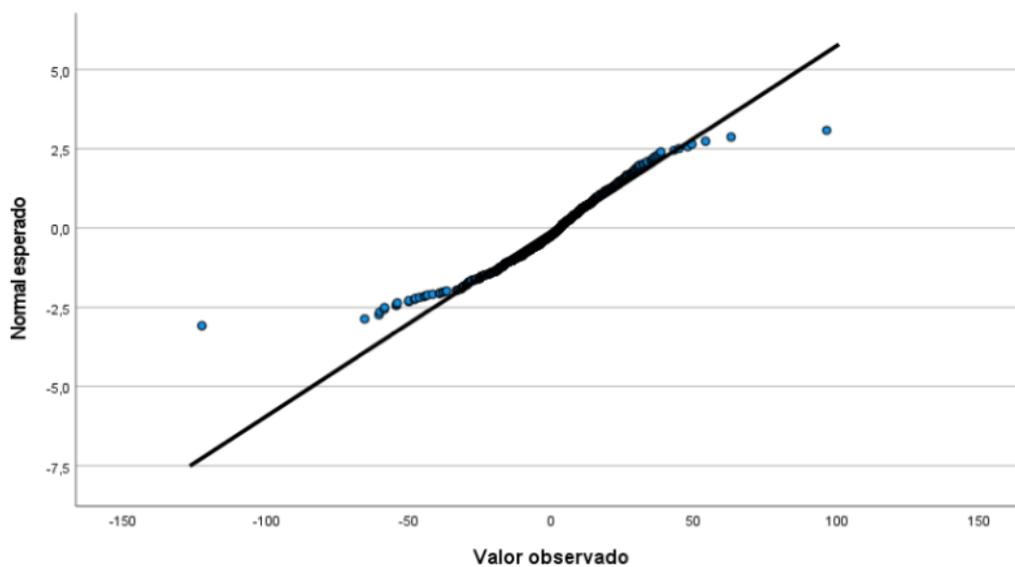
Distribución del residuo de nivel 2 de la competencia matemática en España



En el gráfico Q-Q del residuo de nivel 2 en matemáticas de los estudiantes españoles (Figura 5.4) se puede ver que la mayoría de los valores se ajustan a la curva normal, apareciendo algunos *outliers* en los extremos de la distribución que podrían ser la causa de la desviación observada.

Figura 5.4

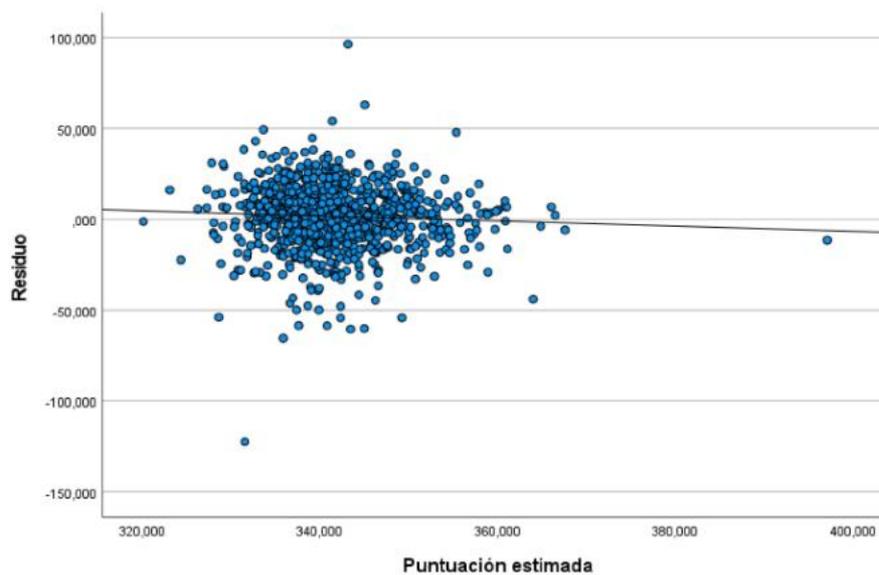
Gráfico Q-Q para la distribución de residuos de nivel 2 en matemáticas en España



El supuesto de homocedasticidad de los residuos de nivel 2 en la competencia matemática de los estudiantes españoles, comprobado mediante la Figura 5.5, se cumple con un valor de R^2 de 0,003.

Figura 5.5

Comprobación del supuesto de homocedasticidad del residuo de nivel 2 en la competencia matemática en España



Tras la comprobación de los supuestos relacionados con los residuos, se confirma el modelo presentado en la Tabla 5.4 y en la ecuación 5.3 como modelo final para la competencia matemática.

5.1.3. Competencia científica

Como en el caso de la competencia matemática, se presenta en primer lugar el modelo nulo para la competencia científica (Ecuación 5.6).

$$Ciencias_{ij} = y_{00} + u_{0j} + r_{ij}$$

(Ecuación 5.6)

Donde:

$Ciencias_{ij}$ se refiere a la puntuación media en ciencias de cada centro;

y_{00} se refiere a la media de todos los centros españoles en ciencias;

u_{0j} se refiere a la diferencia entre la puntuación del centro en ciencias menos la puntuación media de todos los centros;

r_{ij} se refiere al residuo de nivel 1.

Del mismo modo, para el estudio del CCI se calcularon los componentes de la varianza de los dos niveles de análisis (Tabla 5.8).

Tabla 5.8

Componentes de la varianza del modelo nulo de la competencia científica en España

Efecto aleatorio	σ	Componentes de la varianza	g. l.	χ^2	p-valor
Nivel 2	29.92805	895.68801	975	5304.07298	<0.001
Nivel 1	83.47139	6967.47286			

A partir de estos datos, se obtuvo el siguiente resultado para el CCI (Ecuación 5.7):

$$CCI_{ciencias} = \frac{\tau_{00}}{\tau_{00} + \delta^2} = \frac{895,68801}{895,68801 + 6967,47286} = 0,1139$$

(Ecuación 5.7)

En la que:

τ_{00} es el componente de la varianza de nivel 2, y

δ^2 es el componente de la varianza de nivel 1.

Este coeficiente de correlación intraclase supone que la variabilidad correspondiente al nivel de centro para el rendimiento de los estudiantes españoles en ciencias es del 11,39%.

A continuación, se inició el ajuste del modelo (con intercepto aleatorio y pendientes fijas) y, tras la eliminación una a una de las variables no significativas, se obtuvo el modelo final (Ecuación 5.8) compuesto por once variables, cuatro de ellas a nivel de centro y las siete restantes relacionadas con las características contextuales de los estudiantes:

$$\begin{aligned}
Ciencias_{ij} = & \gamma_{00} + \gamma_{01} * OTT1_j + \gamma_{02} * SCHSIZE_j + \gamma_{03} * STAFFSHORT_j + \gamma_{04} \\
& * PROAT6_j + \gamma_{10} * ESCS_{ij} + \gamma_{20} * GENDER_{ij} + \gamma_{30} * REPEAT_{ij} \\
& + \gamma_{40} * GRADE_{ij} + \gamma_{50} * AGE_{ij} + \gamma_{60} * LANG_{ij} + \gamma_{70} * SCCHANGE_{ij} \\
& + u_{0j} + r_{ij}
\end{aligned}$$

(Ecuación 5.8)

Donde:

$Ciencias_{ij}$ se refiere a la puntuación media en ciencias de cada centro;

γ_{00} se refiere a la media de todos los centros españoles en ciencias;

$\gamma_{01} - \gamma_{04}$ son las covariables de nivel 2;

$\gamma_{10} - \gamma_{70}$ son las covariables de nivel 1;

u_{0j} es la diferencia entre la puntuación del centro en ciencias menos la media de todos los centros;

r_{ij} se refiere al residuo de nivel 1.

En la siguiente tabla (Tabla 5.9) se incluye información sobre la relación entre las variables que han resultado significativas y el rendimiento en esta competencia que conforman el modelo final para la competencia en ciencias en España:

Tabla 5.9

Modelo final competencia científica en España

Efecto fijo	Coefficiente	Error típico	t-ratio	g. l.	p-valor
Intercepto, γ_{00}	421.753215	40.088739	10.520	77	<0.001
OTT1, γ_{01}	-20.114711	9.332314	-2.155	970	0.031
SCHSIZE, γ_{02}	0.007455	0.002636	2.828	605	0.005
STAFFSHORT, γ_{03}	-3.984929	1.532136	-2.601	245	0.010
PROAT6, γ_{04}	56.897067	27.248689	2.088	970	0.037
ESCS, γ_{10}	11.619450	0.964503	12.047	38	<0.001
GENDER, γ_{20}	-12.047715	2.067586	-5.827	39	<0.001
REPEAT, γ_{30}	-55.750288	4.999875	-11.150	84	<0.001
GRADE, γ_{40}	22.022909	3.886366	5.667	93	<0.001
AGE, γ_{50}	6.966728	2.513184	2.772	87	0.007
LANG, γ_{60}	-9.416640	3.044682	-3.093	39	0.004
SCCHANGE, γ_{70}	-10.320797	1.877325	-5.498	19	<0.001

Fijándonos en los valores de t-ratio y como en el caso de la competencia matemática, la variable con una mayor influencia en el rendimiento en ciencias de los estudiantes españoles es el ESCS, de nuevo con un claro efecto positivo. En cambio, prestando atención a la variable con una influencia negativa más alta destaca la repetición de curso sobremanera en relación con el resto de covariables.

Al igual que en el caso de la competencia matemática, el género presenta una relación significativa negativa con el rendimiento en ciencias, repitiéndose por tanto el patrón anterior según el cual al pertenecer al género femenino es más probable que se obtengan peores resultados en esta competencia.

De este modelo, se extrajo la estimación de los componentes de la varianza (Tabla 5.10) para poder calcular la varianza explicada por las variables del modelo (*Pseudo R*²).

Tabla 5.10

Componentes de la varianza del modelo final para la competencia científica en España

Efecto aleatorio	σ	Componentes de la varianza	g.l.	χ^2	p-valor
Nivel 2	19.72179	388.94919	970	3099.33596	<0.001
Nivel 1	73.12534	5347.31467			

Para el primer nivel estudiado, el valor del estadístico *Pseudo R*² fue del 23,25%; siendo la proporción de varianza explicada por las variables de segundo nivel del modelo de 56,57%.

Finalmente, se comprobaron los supuestos referidos a los residuos de este modelo de ambos niveles. En la Tabla 5.11 se observan los estadísticos descriptivos correspondientes a los residuos de nivel 1 de esta competencia:

Tabla 5.11

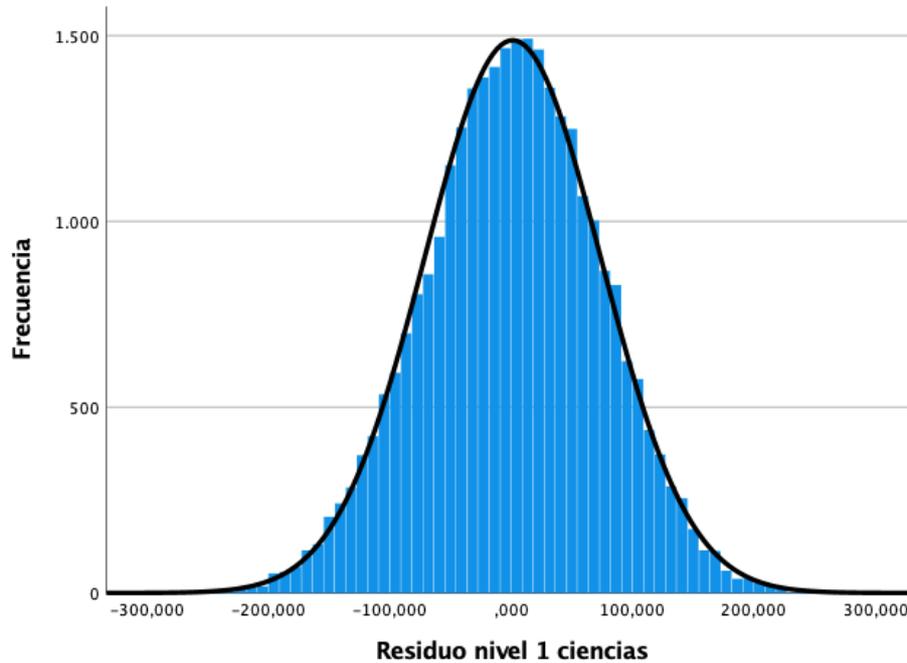
Estadísticos descriptivos para el residuo de nivel 1 en ciencias en España

	N	Mín.	Máx.	Media	σ	Asimetría		Curtosis	
						Valor	Error típico	Valor	Error Típico
Residuo N1	29774	-294.26	294.55	1.29	72.58	-0.06	0.01	-0.05	0.03

A continuación, se observa en la Figura 5.6 la distribución de estos residuos de nivel 1 para ciencias, que se asemeja a la distribución normal, con un valor de significación de la prueba Kolmogorov-Smirnov $<0,001$ junto con una media de 1,29 y una desviación típica de 72,58.

Figura 5.6

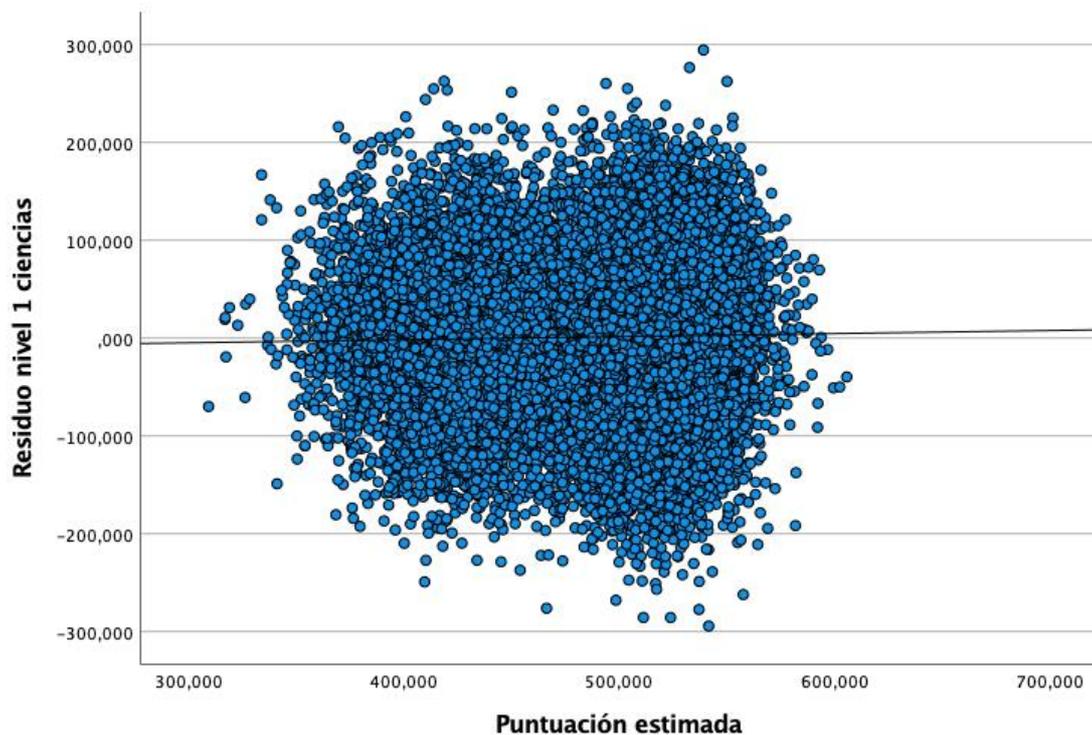
Distribución del residuo de nivel 1 de la competencia científica en España



Para proceder a la comprobación del supuesto de homocedasticidad de los valores del residuo de nivel 1 se analizó el gráfico de dispersión de sus puntuaciones (Figura 5.7), obteniendo un valor de R^2 de 4,069E-4.

Figura 5.7

Comprobación del supuesto de homocedasticidad del residuo de nivel 1 en ciencias en España



En cuanto a los residuos de nivel 2, en la Tabla 5.12 se pueden observar sus estadísticos descriptivos, presentando de nuevo una media cercana a 0 (0,86) y una desviación típica de 14,51 dentro de un rango de aproximadamente 130 puntos:

Tabla 5.12

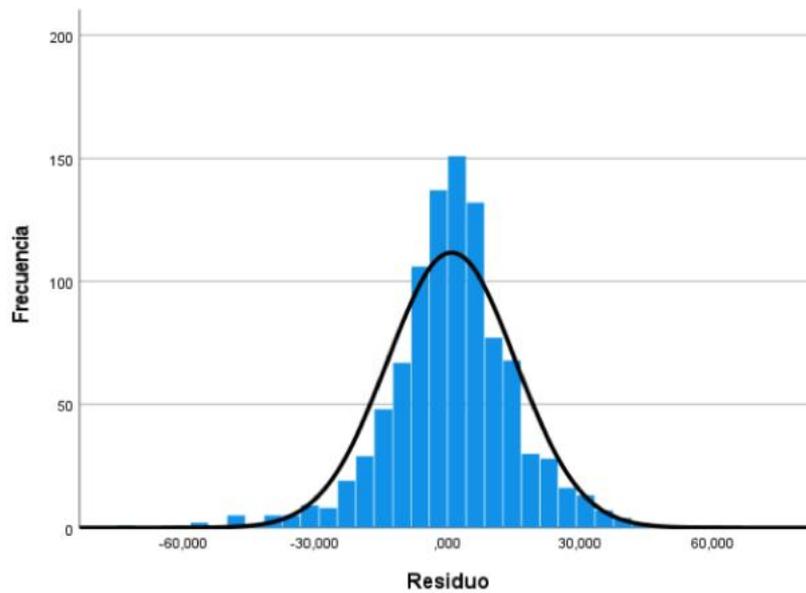
Estadísticos descriptivos para el residuo de nivel 2 en ciencias en España

	N	Mín.	Máx.	Media	σ	Asimetría		Curtosis	
						Valor	Error típico	Valor	Error Típico
Residuo N2	975	-72.75	61.91	0.86	14.51	-0.36	0.08	-2.42	0.16

En la siguiente figura (Figura 5.8), se observa que la distribución de estos residuos se aproxima a la normal, obteniendo un valor de $p < 0,001$ en la prueba de Kolmogorov-Smirnov.

Figura 5.8

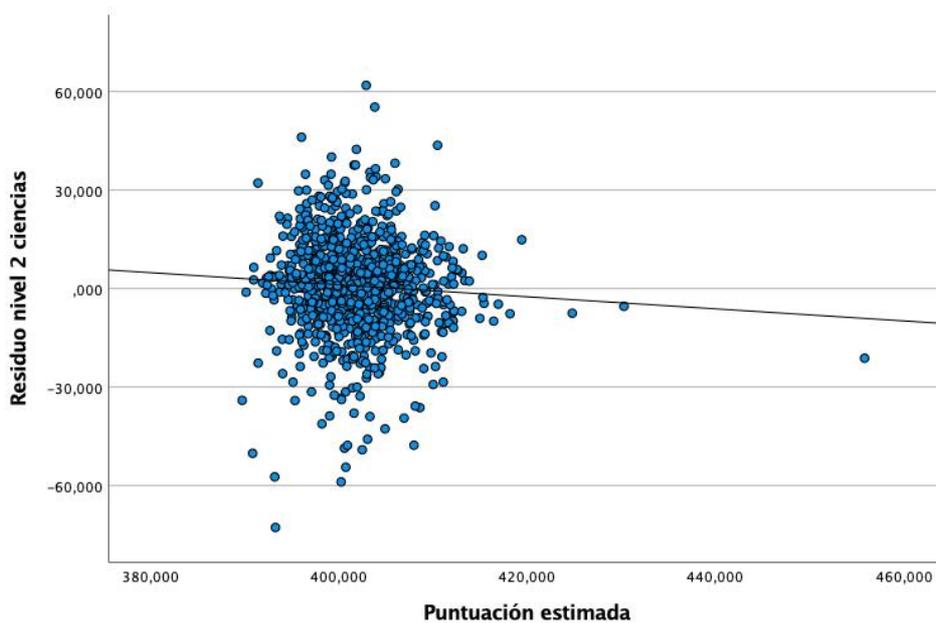
Distribución de los residuos de nivel 2 de la competencia científica en España



Para comprobar que la igualdad en la dispersión de los valores del residuo de nivel 2 en ciencias se cumple (supuesto de homocedasticidad), se ha empleado el gráfico siguiente (Figura 5.9), obteniendo además un valor de R^2 de 0,004:

Figura 5.9

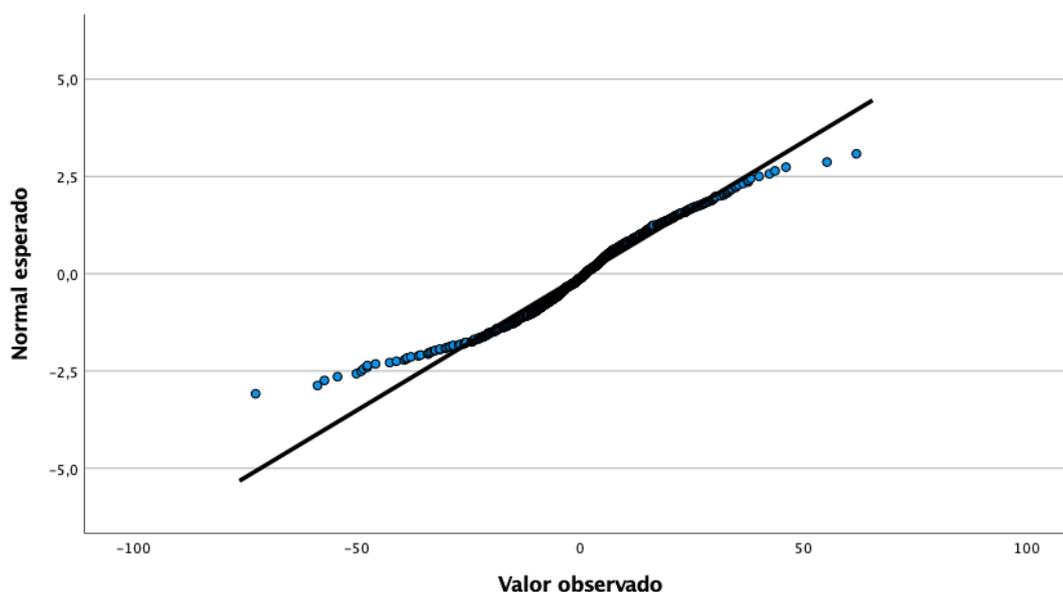
Comprobación del supuesto de homocedasticidad del residuo de nivel 2 en ciencias en España



Además, también se deduce que la mayoría de las puntuaciones se asemejan a una distribución normal en la figura incluida a continuación (Figura 5.10):

Figura 5.10

Gráfico Q-Q para la distribución de residuos de nivel 2 en ciencias en España



Tras haber comprobado el cumplimiento de estos supuestos, se confirma el modelo estadístico presentado en la Tabla 5.9 y en la ecuación 5.8 como modelo final para la competencia científica en España.

5.1.4. Competencia en comprensión lectora

Se inicia la construcción del modelo multinivel para la competencia en comprensión lectora presentando su modelo nulo (Ecuación 5.9).

$$Lectura_{ij} = y_{00} + u_{0j} + r_{ij}$$

(Ecuación 5.9)

Donde:

$Lectura_{ij}$ se refiere a la media en comprensión lectora de cada centro;

y_{00} se refiere a la media de todos los centros españoles en comprensión lectora;

u_{0j} se refiere a la diferencia entre la puntuación del centro en comprensión lectora menos la puntuación media de todos los centros;

r_{ij} se refiere al residuo de nivel 1.

Del mismo modo, para poder proceder con el cálculo del CCI se obtuvieron los valores de los componentes de la varianza de los dos niveles de análisis (Tabla 5.13).

Tabla 5.13

Componentes de la varianza del modelo nulo de la competencia en comprensión lectora en España

Efecto aleatorio	σ	Componentes de la varianza	g. l.	χ^2	p-valor
Nivel 2	34.62875	1199.15064	975	6523.26562	<0.001
Nivel 1	85.07013	7236.92640			

Con esta información, se calculó el CCI del modelo inicial (Ecuación 5.10):

$$CCI_{lectura} = \frac{\tau_{00}}{\tau_{00} + \delta^2} = \frac{1199.15064}{1199.15064 + 7236.92640} = 0.1421$$

(Ecuación 5.10)

En la que:

τ_{00} es el componente de la varianza de nivel 2, y

δ^2 es el componente de la varianza de nivel 1.

Este resultado indica que la variabilidad correspondiente al nivel de centro para el rendimiento de los estudiantes españoles en comprensión lectora es del 14,21%.

En cuanto al ajuste del modelo, de nuevo con intercepto aleatorio y pendientes fijas, tras la eliminación individual y consecutiva de las variables no significativas, se alcanzó el modelo final (Ecuación 5.11) compuesto por once variables, cuatro de ellas de segundo nivel y las siete restantes de primer nivel:

$$\begin{aligned} Lectura_{ij} = & \gamma_{00} + \gamma_{01} * OTT2_j + \gamma_{02} * SCHSIZE_j + \gamma_{03} * STAFFSHORT_j + \gamma_{04} \\ & * LOCAT_j + \gamma_{10} * ESCS_{ij} + \gamma_{20} * GENDER_{ij} + \gamma_{30} * REPEAT_{ij} + \gamma_{40} \\ & * GRADE_{ij} + \gamma_{50} * AGE_{ij} + \gamma_{60} * LANG_{ij} + \gamma_{70} * SCCHANGE_{ij} + u_{0j} \\ & + r_{ij} \end{aligned}$$

(Ecuación 5.11)

En donde:

$Lectura_{ij}$ se refiere a la puntuación media en comprensión lectora de cada centro;

y_{00} se refiere a la media de todos los centros españoles en comprensión lectora;

$y_{01} - y_{04}$ son las covariables de nivel 2;

$y_{10} - y_{70}$ son las covariables de nivel 1;

u_{0j} es la diferencia entre la puntuación del centro en comprensión lectora menos la media de todos los centros;

r_{ij} se refiere al residuo de nivel 1.

En la Tabla 5.14 se observan las variables que componen el modelo final para la competencia en comprensión lectora en España:

Tabla 5.14

Modelo final competencia en comprensión lectora en España

Efectos fijos	Coefficiente	Error típico	t-ratio	g.l.	p-valor
Intercepto, y_{00}	387.063426	37.030460	10.453	357	<0.001
OTT2, y_{01}	-29.686920	12.372569	-2.399	970	0.017
SCHSIZE, y_{02}	0.007459	0.003463	2.154	970	0.031
STAFFSHORT, y_{03}	-4.225187	1.549762	-2.726	970	0.007
LOCAT, y_{04}	3.458251	1.408653	2.455	970	0.014
ESCS, y_{10}	11.004325	0.861783	12.769	137	<0.001
GENDER, y_{20}	17.296984	1.701709	10.164	110	<0.001
REPEAT, y_{30}	-55.058085	4.418017	-12.462	660	<0.001
GRADE, y_{40}	23.023086	3.372247	6.827	1719	<0.001
AGE, y_{50}	7.870401	2.299009	3.423	302	<0.001
LANG, y_{60}	-9.408444	2.670744	-3.523	135	<0.001
SCCHANGE, y_{70}	-9.043357	1.252330	-7.221	133	<0.001

Analizando los valores de t-ratio, la variable con mayor influencia en el rendimiento en comprensión lectora de los estudiantes españoles es el ESCS, de nuevo con un claro efecto positivo. En contraposición, la variable con mayor influencia negativa es de nuevo la repetición de curso destacando sobre el resto de covariables.

Si se analiza la relación entre la variable género y el rendimiento en comprensión lectora, se observa que, a diferencia de los casos de las otras dos competencias estudiadas, aparece una relación positiva con el rendimiento en esta competencia. Esto implica que

pertenecer al género femenino está asociado con la obtención de mejores resultados en esta competencia.

Los valores de los componentes de la varianza (Tabla 5.15) se utilizaron para el cálculo del estadístico *Pseudo R*²:

Tabla 5.15

Componentes de la varianza del modelo final para la competencia en comprensión lectora en España

Efecto aleatorio	σ	Componentes de la varianza	g.l.	χ^2	<i>p</i> -valor
Nivel 2	24.16347	583.87330	970	4066.06312	<0.001
Nivel 1	73.63180	5421.64134			

El cálculo del mencionado estadístico arrojó un resultado para la varianza explicada al nivel 1 por este modelo del 25,08% y para el nivel 2 del 51,31%.

A continuación, se comprobaron los supuestos referidos a los residuos de este modelo. La Tabla 5.16 contiene los estadísticos descriptivos correspondientes al residuo de nivel 1 de la competencia en comprensión lectora en España:

Tabla 5.16

Estadísticos descriptivos para el residuo de nivel 1 en comprensión lectora en España

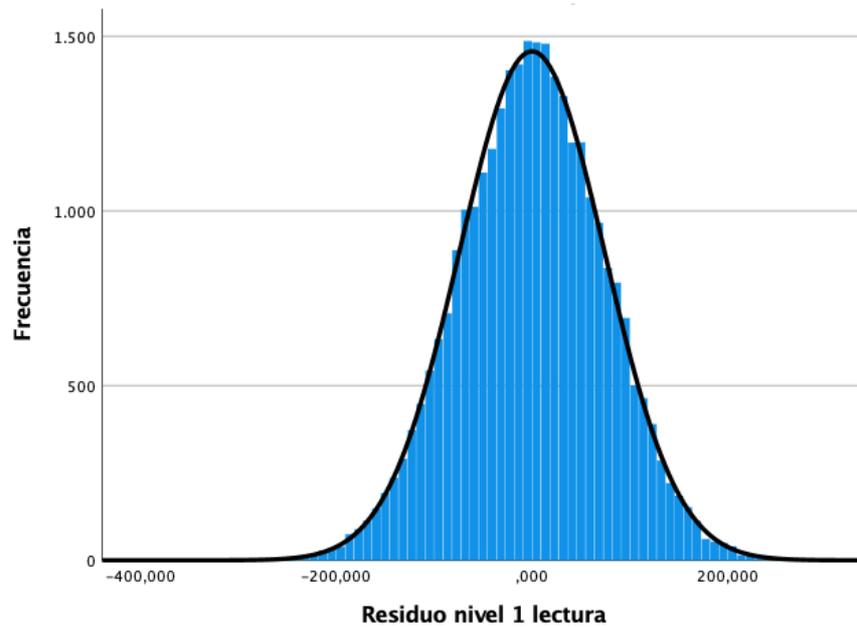
	N	Mín.	Máx.	Media	σ	Asimetría		Curtosis	
						Valor	Error típico	Valor	Error Típico
Residuo N1	29774	-327.13	270.98	-0.13	74.1	-0.05	0.01	-0.02	0.03

En este caso, el residuo presenta una media de -0,13 y una desviación típica de 74,1.

En la Figura 5.11 se presenta la distribución del residuo de nivel 1 en comprensión lectora, que de nuevo se asemeja a una distribución normal, obteniendo un valor de la prueba Kolmogorov-Smirnov de 0,04.

Figura 5.11

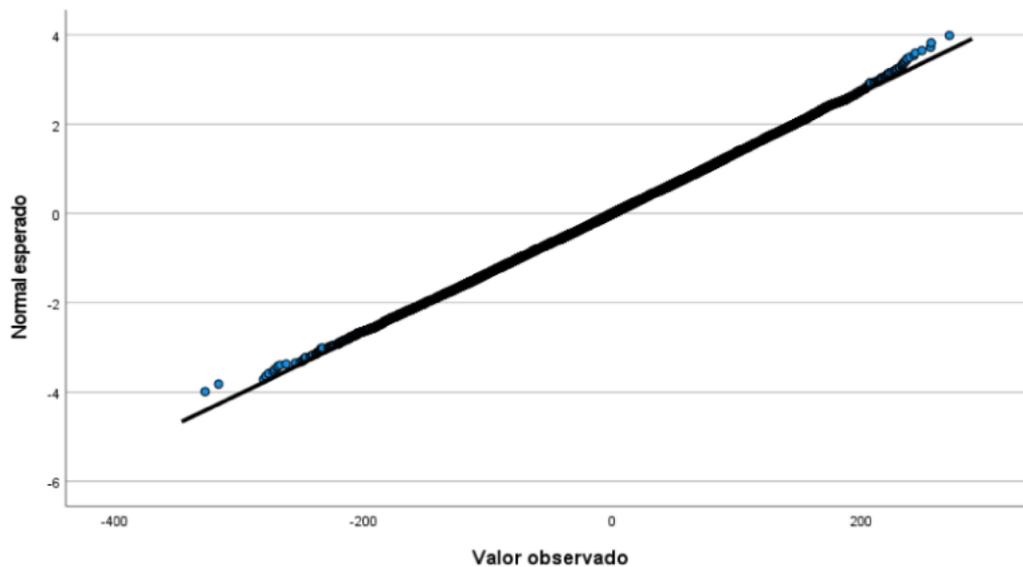
Distribución del residuo de nivel 1 de la competencia en comprensión lectora en España



Como se puede observar, esta distribución presenta una ligera desviación causada por una asimetría negativa debida a la existencia de algunos valores extremos que pueden observarse en la Figura 5.12, además esta distribución es leptocúrtica (Figura 5.11).

Figura 5.12

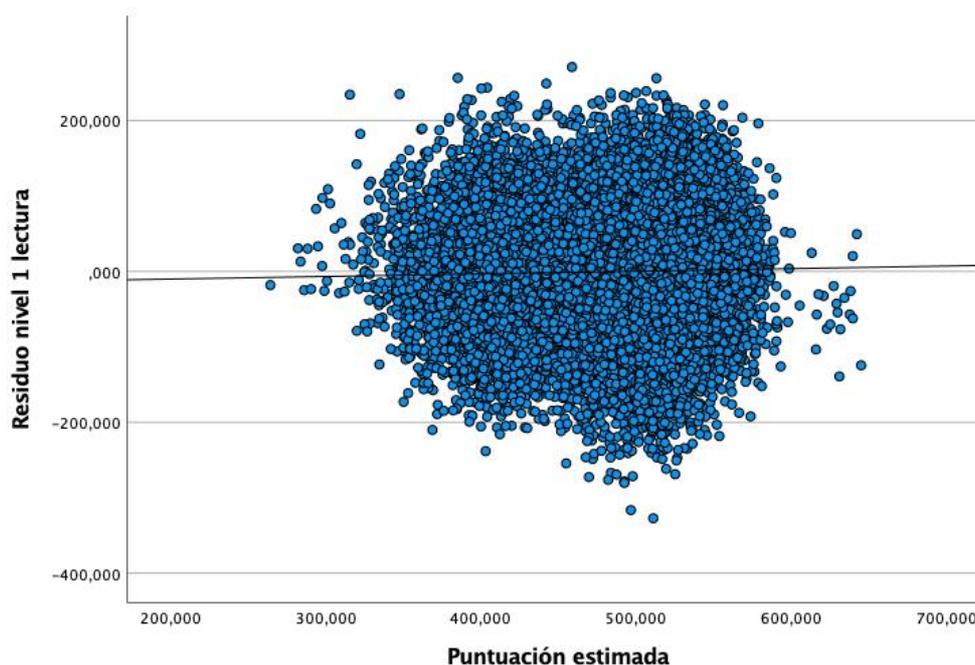
Gráfico Q-Q para los residuos de nivel 1 de la competencia en comprensión lectora en España



Centrándonos en el supuesto de homocedasticidad (Figura 5.13), de nuevo se cumple, obteniendo un valor de R^2 de 5,492E-4.

Figura 5.13

Comprobación del supuesto de homocedasticidad para los residuos de nivel 1 en la competencia en comprensión lectora en España



En cuanto a los residuos de nivel 2 de la competencia en comprensión lectora de los estudiantes españoles, estos presentan una media muy próxima a 0 con un valor de 0,35 y una desviación típica de 20,09. En este caso, la amplitud de la distribución es de casi 210 puntos. En la Tabla 5.17 se pueden consultar los estadísticos descriptivos de este residuo:

Tabla 5.17

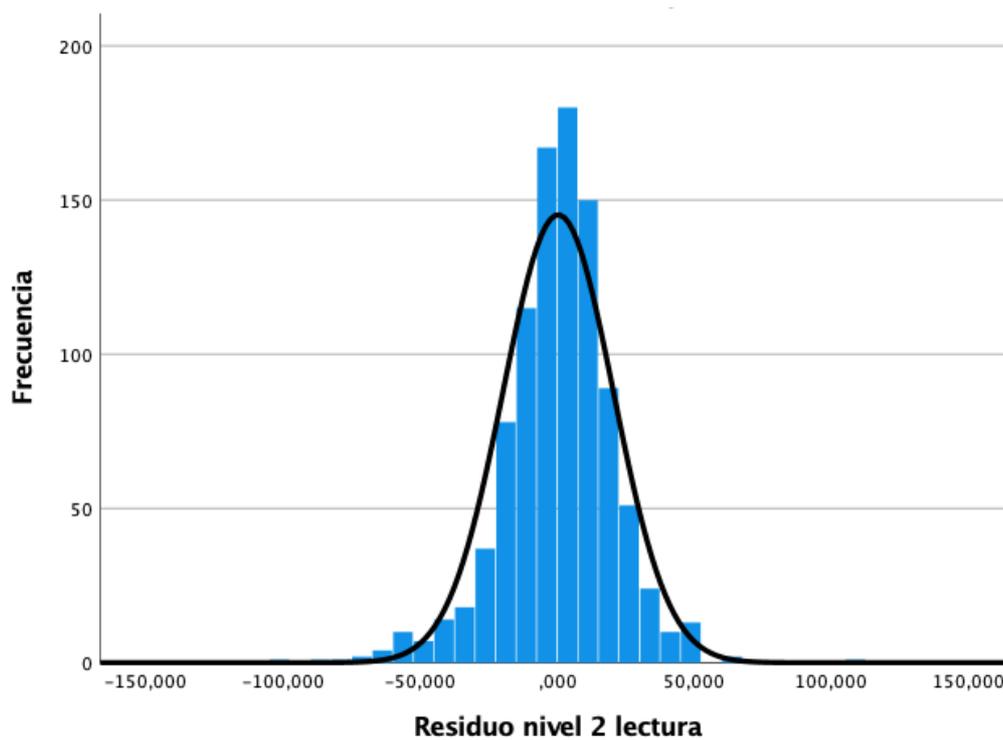
Estadísticos descriptivos para el residuo de nivel 2 en comprensión lectora en España

	N	Mín.	Máx.	Media	σ	Asimetría		Curtosis	
						Valor	Error típico	Valor	Error Típico
Residuo N2	975	-102.36	106.71	0.35	20.09	-0.45	0.08	2.69	0.16

Analizando el siguiente histograma (Figura 5.14), se puede determinar que la distribución de estos residuos se aproxima de nuevo a una distribución normal, obteniendo un valor de $p < 0,001$ en la prueba de Kolmogorov-Smirnov y presentando repetidamente una distribución leptocúrtica.

Figura 5.14

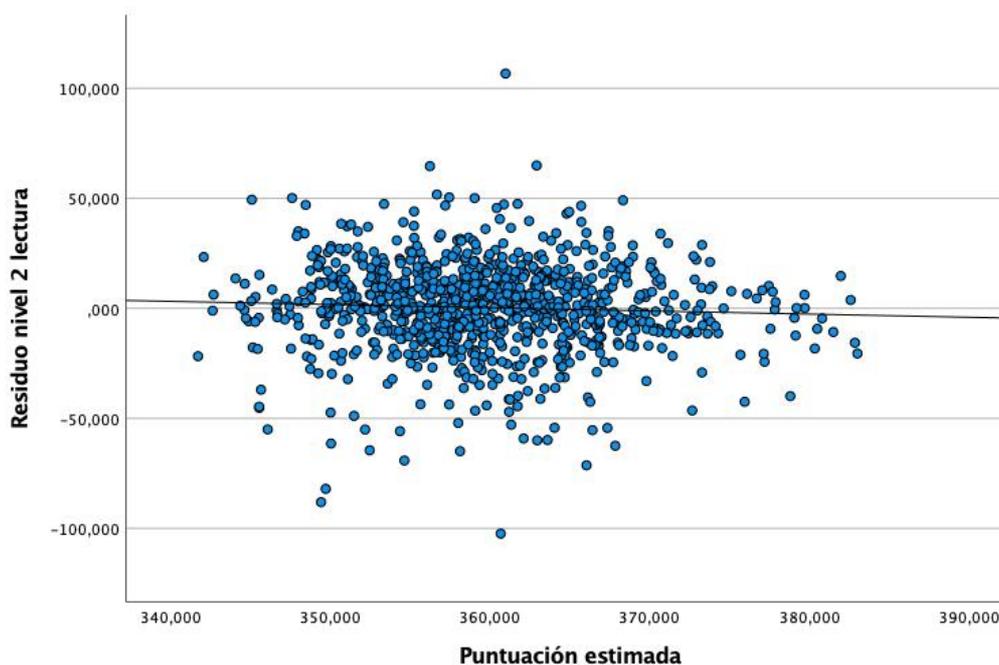
Distribución del residuo de nivel 2 de la competencia en comprensión lectora en España



En cuanto al supuesto de igualdad en la dispersión de las puntuaciones, presentado en la Figura 5.15, este se cumple, presentando un valor de R^2 de 0,003.

Figura 5.15

Comprobación del supuesto de homocedasticidad del residuo de nivel 2 de la competencia en comprensión lectora en España



Por tanto, al cumplirse todos los supuestos relacionados con los residuos de manera aceptable, el modelo presentado en la ecuación 5.11 y en la Tabla 5.16 queda determinado como modelo estadístico final para la competencia en comprensión lectora del estudiantado español.

5.2. Análisis de los factores contextuales del rendimiento académico de los estudiantes irlandeses

Se presenta a continuación el estudio de las variables contextuales asociadas al rendimiento académico en matemáticas, ciencias y comprensión lectora de los estudiantes irlandeses siguiendo el esquema del apartado anterior.

5.2.1. Consideraciones previas

En el caso de Irlanda y como se ha detallado en el apartado dirigido a la explicación de la metodología a utilizar en este estudio, las pruebas PISA 2018 no proporcionan información proveniente del profesorado, por lo que aquellas variables extraídas a partir de esta información no se han incluido en el estudio multinivel. Del mismo modo,

tampoco se reportaba información sobre la titularidad de los centros educativos participantes, por lo que la variable SCHOWN tampoco se ha tenido en cuenta.

El cálculo de los modelos jerárquico-lineales se inicia con la comprobación de las correlaciones existentes entre las diferentes variables en ambos niveles de análisis calculando el índice correspondiente en función del tipo de variable de que se trate (según lo indicado en la sección 5.1.1 de este mismo capítulo).

En la Tabla 5.18 se observan los índices de correlación referidos a las variables de nivel 1 (estudiante).

Tabla 5.18

Correlaciones entre las variables contextuales a nivel de estudiante en Irlanda

	1	2	3	4	5	6	7	8	9
1. ESCS	-								
2. GENDER	0.03	-							
3. AGE	-0.001	0.01	-						
4. GRADE	-0.027	0.08	0.597	-					
5. IMMIG	-0.018	0.024	-0.016	0.147	-				
6. REPEAT	-0.066	-0.016	0.043	0.305	0.038	-			
7. SCCHANGE	-0.028	-0.027	0.014	0.067	0.278	0.182	-		
8. DURECEC	0.048	-0.049	-0.012	-0.097	0.114	-0.01	0.083	-	
9. LANG	-0.075	-0.001	0.011	0.156	0.548	0.033	0.222	0.14	-

En la siguiente tabla (Tabla 5.19) se pueden consultar los índices de correlación para las variables de segundo nivel (centros):

Tabla 5.19*Correlaciones entre las variables contextuales a nivel de centro en Irlanda*

	1	2	3	4	5	6	7	8	9	10	11	12
1. SCHSIZE	-											
2. CLSSIZE	0.459	-										
3. EDUSHORT	-0.145	-0.117	-									
4. STAFFSHORT	-0.216	-0.207	0.473	-								
5. RATIO	0.476	0.491	-0.065	-0,042	-							
6. LOCAT	0.133	0.09	0.061	-0.071	-0.014	-						
7. PROAT5AM	-0.027	-0.025	-0.107	0.067	0.006	0.233	-					
8. PROAT6	0.021	0.071	-0.085	-0.045	0.017	0.185	0.304	-				
9. N2GEN	0.068	0.104	-0.103	0.058	0.172	0.023	0.11	0.179	-			
10. N2ESCS	0.291	0.352	-0.176	-0.174	0.44	0.036	0.036	0.115	0.143	-		
11. N2REPEAT	-0.396	-0.299	0.077	0.09	-0.442	-0.212	-0.011	-0.058	-0.052	-0.283	-	
12. N2INMIG	0.008	-0.038	0.065	-0.011	-0.094	0.224	0.013	0.024	0.062	-0.231	0.14	-

Se observa que no existen variables que correlacionen en exceso en ninguno de los dos niveles, siendo los valores más altos los existentes entre la edad y el curso (0,597) y la inmigración y la lengua hablada en el hogar (0,548), ambos pares de variables referidos al primer nivel de estudio.

A continuación, se detalla la elaboración de los modelos correspondientes a las tres variables estudiadas en el caso de Irlanda.

5.2.2. Competencia matemática

El modelo nulo o incondicional para la competencia matemática sigue lo planteado en la Ecuación 5.1. Tomando como base este modelo nulo se ha calculado el CCI inicial con los valores correspondientes a los componentes de la varianza de ambos niveles de estudio, que pueden ser consultados en la Tabla 5.20:

Tabla 5.20

Componentes de la varianza del modelo nulo de la competencia matemática en Irlanda

Efecto aleatorio	σ	Componentes de la varianza	g. l.	χ^2	p-valor
Nivel 2	31.87376	1015.93626	154	1183.19969	<0.001
Nivel 1	72.26304	5221.94632			

A partir de esta información, y aplicando la Ecuación 5.12 se ha obtenido el siguiente resultado:

$$CCI_{matemáticas} = \frac{\tau_{00}}{\tau_{00} + \delta^2} = \frac{1015,93626}{1015,93626 + 5221,94632} = 0,1629$$

(Ecuación 5.12)

Donde:

τ_{00} es el componente de la varianza de nivel 2, y

δ^2 es el componente de la varianza de nivel 1.

Este resultado nos indica que la variabilidad entre centros en esta competencia en Irlanda es de un 16,29%, suponiendo la existencia de dependencia entre las observaciones y aceptando la aplicación de la modelización multinivel.

A continuación, se procedió con el ajuste del modelo hasta llegar al modelo final siguiendo, como en el caso de los cálculos de los modelos para España, un diseño donde el intercepto es aleatorio y las pendientes son fijas. En la siguiente ecuación (Ecuación 5.13) se pueden apreciar las variables que conforman el modelo final para la competencia matemática en Irlanda:

Matemáticas_{ij}

$$= \gamma_{00} + \gamma_{01} * N2ESCS_j + \gamma_{10} * ESCS_{ij} + \gamma_{20} * GENDER_{ij} + \gamma_{30} * DURECEC_{ij} + \gamma_{40} * REPEAT_{ij} + \gamma_{50} * AGE_{ij} + \gamma_{60} * LANG_{ij} + \gamma_{70} * INM2_{ij} + \gamma_{80} * SCCHANGE_{ij} + u_{0j} + r_{ij}$$

(Ecuación 5.13)

Donde:

Matemáticas_{ij} se refiere a la puntuación media en matemáticas de cada centro;

γ_{00} se refiere a la media de todos los centros irlandeses en matemáticas;

γ_{01} es la covariable de nivel 2;

$\gamma_{10} - \gamma_{80}$ son las covariables de nivel 1;

u_{0j} es la diferencia entre la puntuación del centro en matemáticas menos la media de todos los centros;

r_{ij} se refiere al residuo de nivel 1.

La Tabla 5.21 incluye información relativa a las variables que han presentado una relación significativa con la variable criterio estudiada (rendimiento en matemáticas) y que componen el modelo final para la competencia matemática en Irlanda:

Tabla 5.21

Modelo final competencia matemática en Irlanda

Efectos fijos	Coefficiente	Error típico	t-ratio	g. l.	p-valor
Intercepto, γ_{00}	209.192181	71.867914	2.911	153	0.004
N2ESCS, γ_{01}	46.316745	4.180229	11.080	153	<0.001
ESCS, γ_{10}	20.829723	1.754272	11.874	319	<0.001
GENDER, γ_{20}	-11.475024	3.277676	-3.501	84	<0.001
DURECEC, γ_{30}	-5.187889	1.543995	-3.360	151	<0.001
REPEAT, γ_{40}	-36.039016	5.865523	-6.144	198	<0.001
AGE, γ_{50}	19.189583	4.598107	4.173	206	<0.001

Efectos fijos	Coeficiente	Error típico	t-ratio	g. l.	p-valor
LANG, y_{60}	16.994840	6.215276	2.734	68	0.008
INM_2, y_{70}	-4.575120	0.941155	-4.861	76	<0.001
SCCHANGE, y_{80}	-6.716187	2.585708	-2.597	114	0.011

Se presentaba en la Ecuación 5.14 cómo las variables con influencia en el rendimiento en matemáticas se relacionan fundamentalmente con las características individuales de los estudiantes (nivel 1), siendo la única variable de segundo nivel que aparece en el modelo final el índice del nivel socioeconómico y cultural medio del centro.

Según los valores de t-ratio de estas variables, la mayor parte de esta varianza se debe tanto al nivel socioeconómico y cultural medio del estudiantado de los centros como a este mismo índice a nivel individual, presentando una relación positiva con el rendimiento en matemáticas. Situándonos en el lado opuesto y prestando atención a la variable con mayor influencia negativa en esta competencia se encuentra la repetición de curso de los estudiantes, suponiendo que un aumento de una unidad en esta variable supone la reducción de 6,144 puntos en el rendimiento en matemáticas del estudiante.

Al proceder con el cálculo de *Pseudo R*² para conocer la proporción de varianza explicada por el modelo estadístico final a partir de los valores de los componentes de la varianza (Tabla 5.22), se obtuvo un resultado para el primer nivel de un 10,84%, mientras que la proporción de varianza explicada a segundo nivel fue del 79,64%.

Tabla 5.22

Componentes de la varianza del modelo final para la competencia matemática en Irlanda

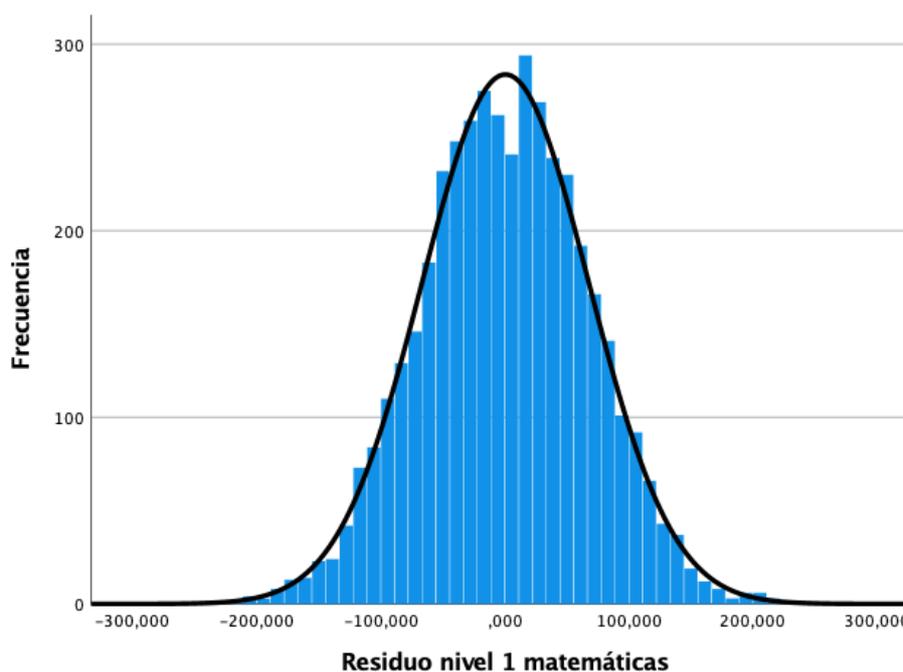
Efecto aleatorio	σ	Componentes de la varianza	g.l.	χ^2	p-valor
Nivel 2	14.38037	206.79501	153	350.62353	<0001
Nivel 1	68.23449	4655.94517			

A continuación, se procedió a la comprobación de los supuestos referidos a los residuos de este modelo final, incluyendo los estadísticos descriptivos de los residuos de los dos niveles de estudio en la Tabla 5.23.

Tabla 5.23*Estadísticos descriptivos para los residuos de nivel 1 y nivel 2 en matemáticas en Irlanda*

	N	Mín.	Máx.	Media	σ	Asimetría		Curtosis	
						Valor	Error típico	Valor	Error Típico
Residuo N1	4305	-268.62	219.77	0.5	67.23	-0.05	0.04	0.03	0.08
Residuo N2	155	-31.65	30.48	0.21	10.33	-0.09	0.19	0.5	0.39

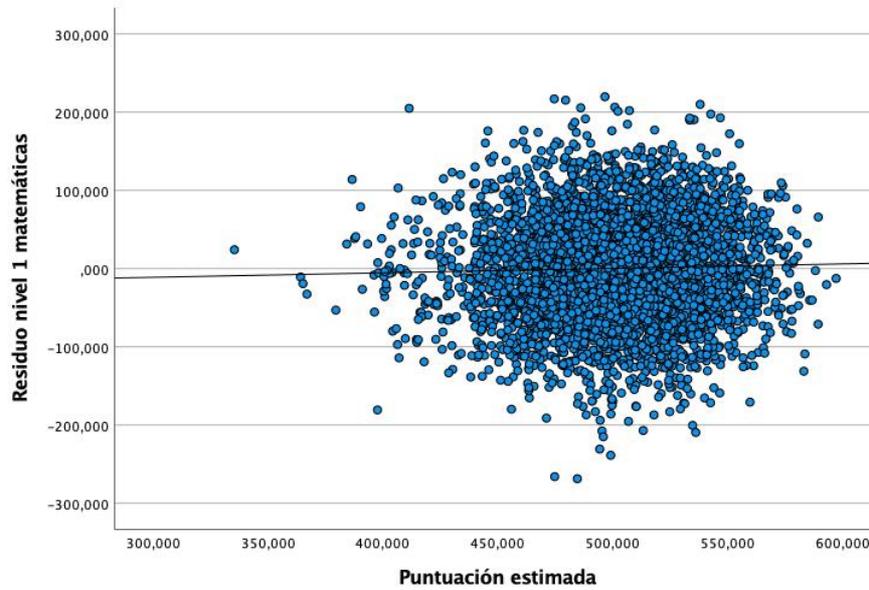
Las puntuaciones del residuo de nivel 1 presentan una distribución con una media próxima a 0 y una desviación típica de 67,23 comprendidos entre los valores -268,62 y 219,77. La distribución de este residuo sigue una distribución normal en la siguiente figura (Figura 5.16), habiendo obtenido una significación en la prueba de Kolmogorov-Smirnov de 0,2.

Figura 5.16*Distribución del residuo de nivel 1 de la competencia matemática en Irlanda*

Para comprobar el supuesto de homocedasticidad se utiliza el gráfico de dispersión incluido a continuación (Figura 5.17), que junto con el valor de R^2 de 8,187E-4 nos indica que el residuo de nivel 1 de la competencia matemática en Irlanda cumple con este requisito.

Figura 5.17

Comprobación del supuesto de homocedasticidad del residuo de nivel 1 de la competencia matemática en Irlanda

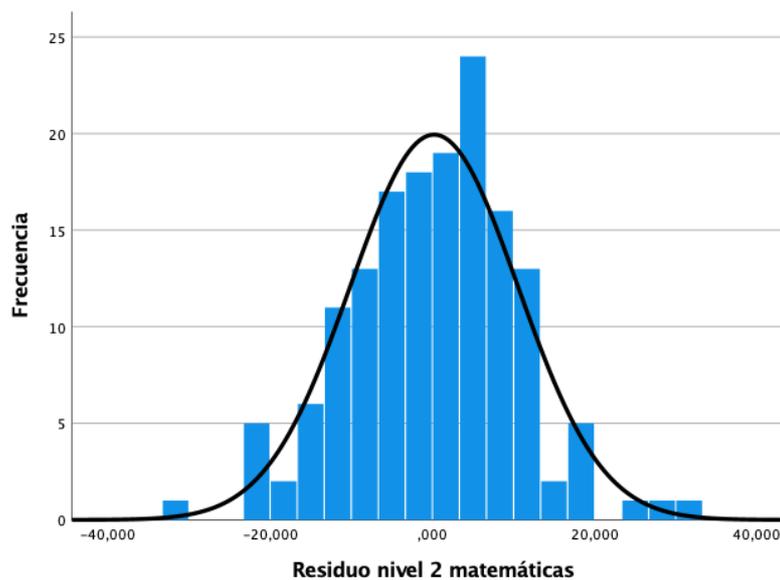


Los estadísticos descriptivos del residuo de nivel 2 de esta competencia nos indican de nuevo una media muy cercana a 0 (0,21) y una desviación típica de 67,23.

La distribución de las puntuaciones correspondientes al residuo de nivel 2 es normal con un valor de significación de Kolmogorov-Smirnov de 0,2 (Figura 5.18):

Figura 5.18

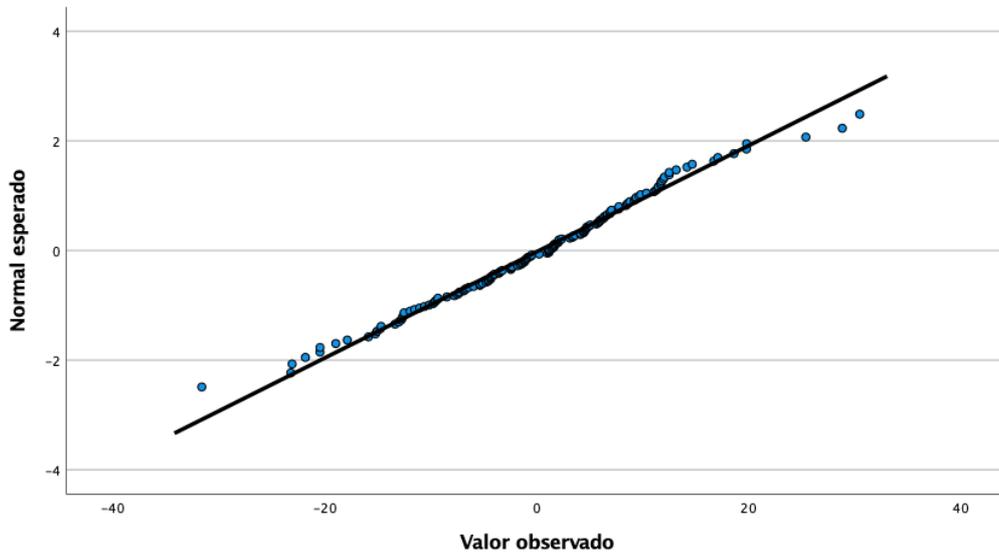
Distribución del residuo de nivel 2 de la competencia matemática en Irlanda



El gráfico incluido a continuación (Figura 5.19) presenta la distribución de los valores de este residuo, apareciendo, como cabía esperar según la figura anterior (Figura 5.18), algunos *outliers* en sus extremos.

Figura 5.19

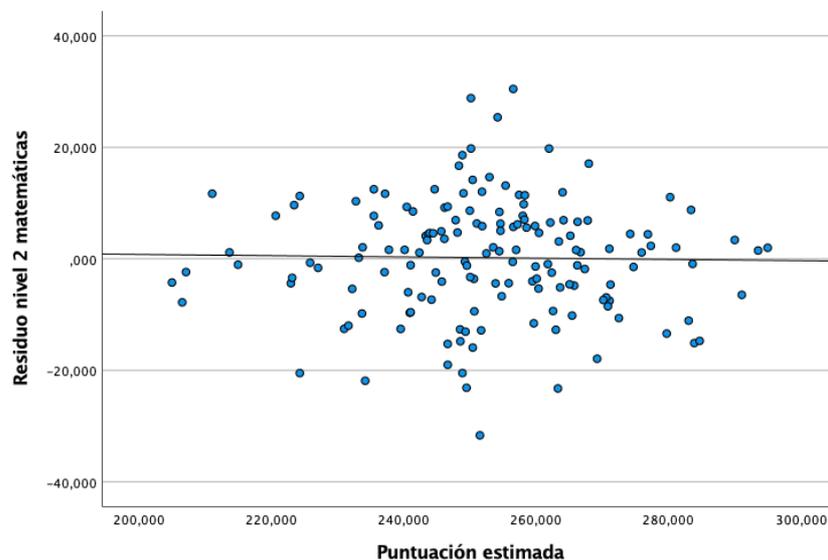
Gráfico Q-Q para la distribución de residuos de nivel 2 en matemáticas en Irlanda



En cuanto al supuesto de homocedasticidad, de nuevo se cumple, obteniendo un valor de R^2 de 3,621E-4, como se observa en la Figura 5.20:

Figura 5.20

Comprobación del supuesto de homocedasticidad del residuo de nivel 2 en la competencia matemática en Irlanda



Todo esto supone que el modelo estadístico incluido en la Ecuación 5.13 y en la Tabla 5.21 se trata del modelo final para la competencia matemática en Irlanda.

5.2.3. Competencia científica

El modelo nulo para la competencia científica de los estudiantes irlandeses es el mismo que el presentado en el caso de los estudiantes españoles (ecuación 5.6). A partir de este modelo incondicional, se han obtenido los componentes de la varianza de ambos niveles de estudio (Tabla 5.24) para posibilitar el cálculo del coeficiente de correlación intraclase.

Tabla 5.24

Componentes de la varianza del modelo nulo de la competencia científica en Irlanda

Efecto aleatorio	σ	Componentes de la varianza	g. l.	χ^2	p-valor
Nivel 2	34.10661	1163.26073	154	1050.96047	<0.001
Nivel 1	82.88297	6869.58644			

Con esta información se calculó el CCI inicial del modelo incondicional aplicando la fórmula correspondiente (Ecuación 5.14) y obteniendo un resultado de 14,48%.

$$CCI_{ciencias} = \frac{\tau_{00}}{\tau_{00} + \delta^2} = \frac{1163,26073}{1163,26073 + 6869,58644} = 0,1448$$

(Ecuación 5.14)

En la que:

τ_{00} es el componente de la varianza de nivel 2, y

δ^2 es el componente de la varianza de nivel 1.

El valor del índice calculado supone que la variación en las puntuaciones en ciencias de los estudiantes irlandeses debidas al segundo nivel de análisis son de un 14,48%, aceptándose por tanto el valor mínimo establecido del 10% para proceder a la aplicación de los modelos multinivel.

Con el fin de alcanzar el modelo condicional final, se procedió con su ajuste mediante la inclusión de las covariables oportunas con peso significativo, manteniendo las pendientes fijas y el intercepto aleatorio. Este modelo final lo compusieron un total de

cinco variables, cuatro asociadas al primer nivel y la restante referida al nivel socioeconómico medio del centro educativo (Ecuación 5.15).

$$Ciencias_{ij} = \gamma_{00} + \gamma_{01} * N2ESCS_j + \gamma_{10} * ESCS_{ij} + \gamma_{20} * REPEAT_{ij} + \gamma_{30} * AGE + \gamma_{40} * INM_2_{ij} + u_{0j} + r_{ij}$$

(Ecuación 5.15)

En esta ecuación:

$Ciencias_{ij}$ se refiere a la puntuación media en ciencias de cada centro;

γ_{00} se refiere a la media de todos los centros irlandeses en ciencias;

γ_{01} es la covariable de nivel 2;

$\gamma_{10} - \gamma_{40}$ son las covariables de nivel 1;

u_{0j} es la diferencia entre la puntuación del centro en ciencias menos la media de todos los centros;

r_{ij} se refiere al residuo de nivel 1.

La relación existente entre las variables que componen el modelo estadístico multinivel final de la competencia científica en Irlanda se puede apreciar en la tabla incluida a continuación (Tabla 5.25):

Tabla 5.25

Modelo final competencia científica en Irlanda

Efectos fijos	Coficiente	Error típico	t-ratio	g.l.	p-valor
Intercepto, γ_{00}	271.126877	74.250670	3.652	153	<0.001
N2ESCS, γ_{01}	42.978556	5.776254	7.441	153	<0.001
ESCS, γ_{10}	23.644826	1.980467	11.939	210	<0.001
REPEAT, γ_{20}	-36.821218	5.438212	-6.771	263	<0.001
AGE, γ_{30}	14.085468	4.719389	2.985	607	0.003
INM_2, γ_{40}	-5.906658	0.863670	-6.839	609	<0.001

Según los valores de t-ratio, se observa de nuevo que la mayor influencia en el rendimiento académico en esta competencia depende del nivel socioeconómico de los estudiantes a ambos niveles, siendo esta relación positiva. Esto supone que, teóricamente, cuanto más alto sea el nivel socioeconómico más alto será este rendimiento. En cuanto a influencia negativa en el rendimiento en ciencias de los estudiantes irlandeses, tanto el

hecho de que el estudiante sea inmigrante de segunda generación como que haya repetido curso presentan valores similares cercanos a los 7 puntos, siendo las dos únicas variables del modelo que influyen negativamente en los resultados en esta competencia.

A partir de este modelo condicional, se extrajeron de nuevo los componentes de la varianza (Tabla 5.26) para proceder al cálculo de *Pseudo R*².

Tabla 5.26

Componentes de la varianza del modelo final para la competencia científica en Irlanda

Efecto aleatorio	σ	Componentes de la varianza	g.l.	χ^2	p-valor
Nivel 2	17.03482	290.18505	153	410.10478	<0.001
Nivel 1	79.51730	6323.00030			

Utilizando los componentes de la varianza del modelo final para esta competencia y los del modelo nulo, se calculó el valor de *Pseudo R*² para cada uno de los dos niveles de análisis, obteniendo valores de 7,96% y 75,05% de varianza explicada en el nivel 1 y nivel 2 respectivamente.

A continuación, se procedió con la comprobación de los supuestos relacionados con los residuos de los dos niveles de análisis incluidos en este cálculo, pudiendo observar en la Tabla 5.27 los estadísticos descriptivos del residuo de nivel 1 de la competencia científica en Irlanda, con una media de 0,24 y una desviación típica de 78,56 en un rango de aproximadamente 550 puntos:

Tabla 5.27

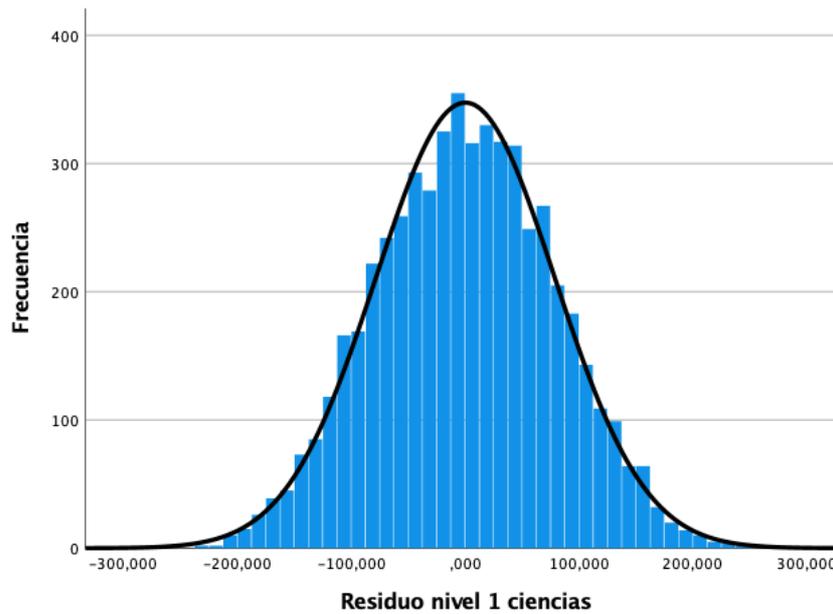
Estadísticos descriptivos del residuo de nivel 1 en ciencias en Irlanda

	N	Mín.	Máx.	Media	σ	Asimetría		Curtosis	
						Valor	Error típico	Valor	Error Típico
Residuo N1	5476	-278.7	272.75	0.24	78.56	-0.01	0.03	-0.26	0.07

En el siguiente histograma (Figura 5.21) se puede observar que la distribución de estas puntuaciones se asemeja a una distribución normal con una media muy cercana a 0, obteniendo un valor de significación en la prueba de Kolmogorov-Smirnov de 0,02.

Figura 5.21

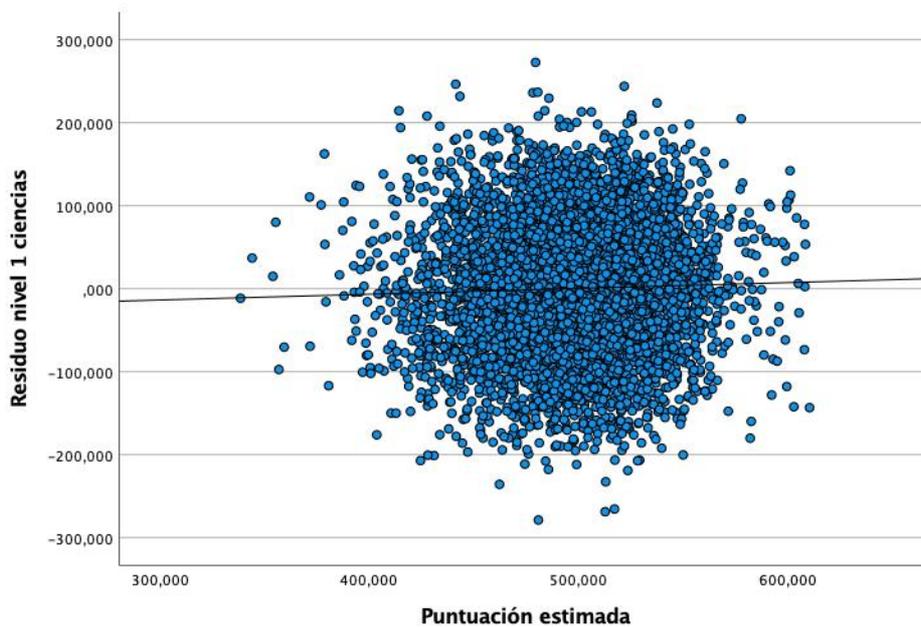
Distribución del residuo de nivel 1 de la competencia científica en Irlanda



En relación con el supuesto de homocedasticidad de los valores de la distribución se analizó su gráfico de dispersión (Figura 5.22) confirmando su cumplimiento y obteniendo además un valor de R^2 de 9,888E-4.

Figura 5.22

Comprobación del supuesto de homocedasticidad del residuo de nivel 1 en ciencias en Irlanda



Para finalizar, se procedió a la comprobación de los supuestos del residuo de nivel 2, siendo sus estadísticos descriptivos los incluidos en la Tabla 5.28:

Tabla 5.28

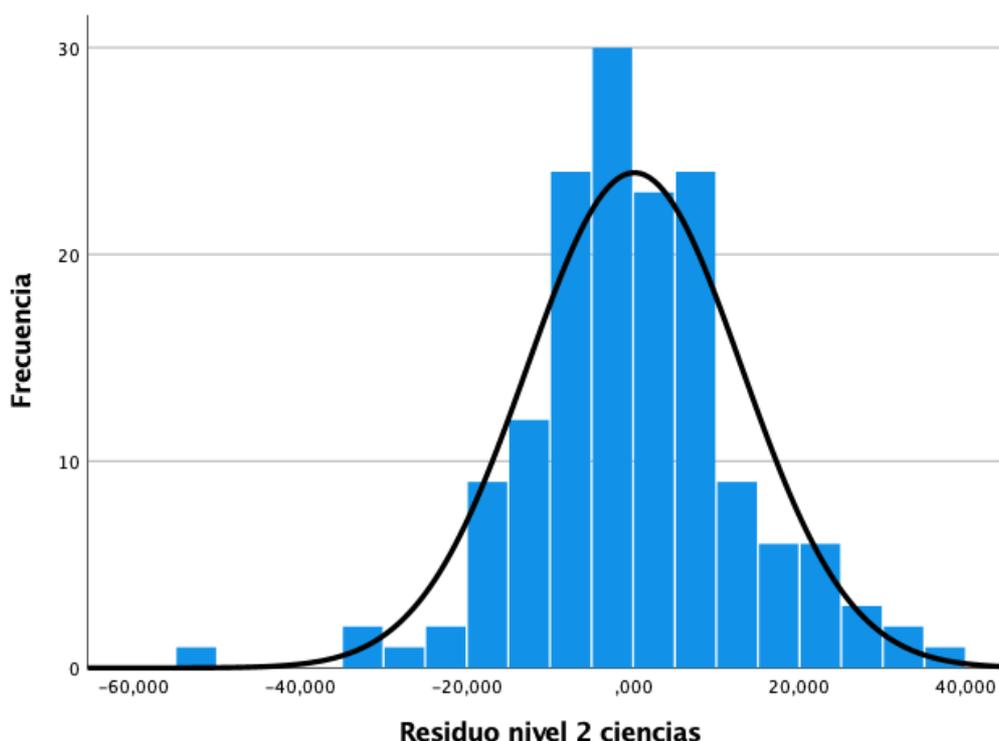
Estadísticos descriptivos para el residuo de nivel 2 en ciencias en Irlanda

	N	Mín.	Máx.	Media	σ	Asimetría		Curtosis	
						Valor	Error típico	Valor	Error Típico
Residuo N2	155	-52.02	36.69	0.15	12.91	-0.2	0.19	1.79	0.39

En este caso, las puntuaciones presentan una media nuevamente próxima a 0 con un valor de 0,15 y una desviación típica de 12,91. Siendo los valores mínimo y máximo de -52,02 y 36,69 respectivamente. En la siguiente figura, se observa la distribución de las puntuaciones del residuo de nivel 2 en ciencias en Irlanda (figura 5.23), donde aparece una figura con carácter leptocúrtico y asimetría negativa.

Figura 5.23

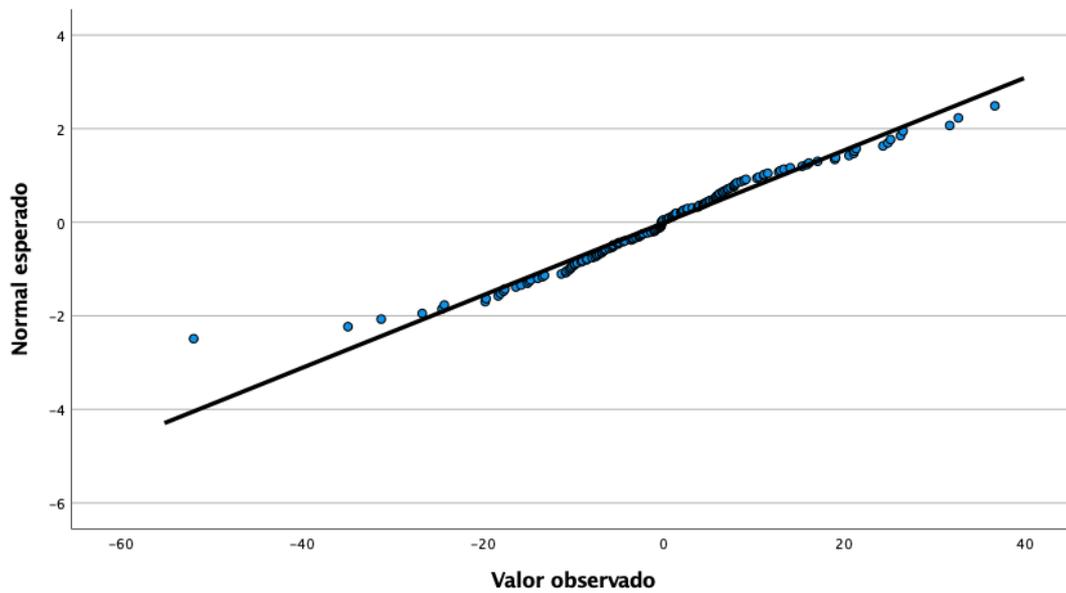
Distribución de los residuos de nivel 2 de la competencia científica en Irlanda



En cuanto a la bondad de ajuste del modelo presentado, se ha obtenido un valor de Kolmogorov-Smirnov de 0,04, asemejándose por tanto esta distribución a una distribución normal, aunque con la aparición de algunos *outliers* que pueden ser observados en el gráfico siguiente (Figura 5.24):

Figura 5.24

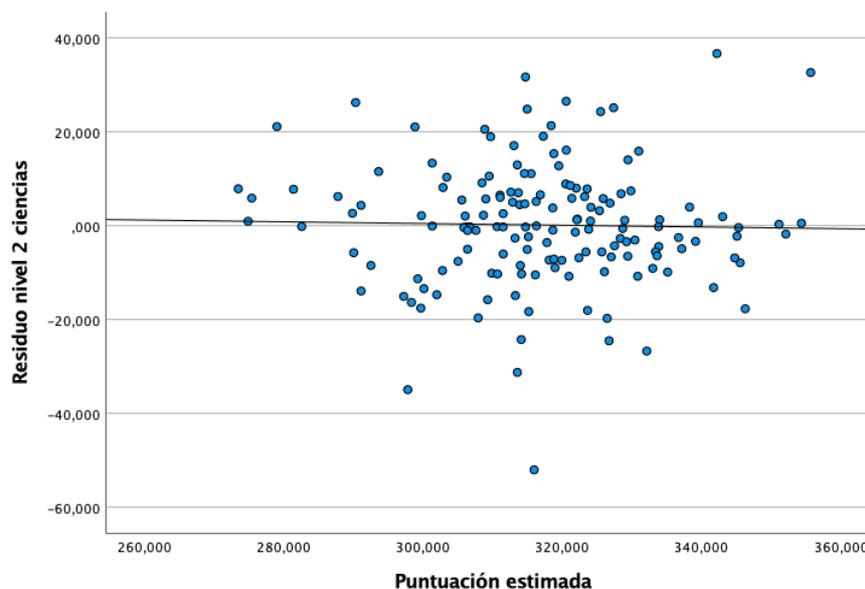
Gráfico Q-Q para la distribución de residuos de nivel 2 en ciencias en Irlanda



Por último, se comprobó el cumplimiento del supuesto de homocedasticidad, calculando el valor de R^2 con un resultado de 5,266E-4, siendo la dispersión de las puntuaciones la siguiente (Figura 5.25) y pudiendo asumir la igualdad en esta dispersión:

Figura 5.25

Comprobación del supuesto de homocedasticidad del residuo de nivel 2 en ciencias en Irlanda



Tras la comprobación de estos supuestos estadísticos relacionados con los residuos de primer y segundo nivel, se concluye que el modelo estadístico presentado con anterioridad (Ecuación 5.15 y Tabla 5.25) conforma el modelo final para la competencia científica en Irlanda.

5.2.4. Competencia en comprensión lectora

Para iniciar la construcción del modelo multinivel para la competencia en comprensión lectora se parte del modelo nulo (Ecuación 5.9). A continuación, se presentan los componentes de la varianza de este modelo incondicional (Tabla 5.29) utilizados para calcular el CCI inicial (Ecuación 5.16) del modelo:

Tabla 5.29

Componentes de la varianza del modelo nulo de la competencia en comprensión lectora en Irlanda

Efecto aleatorio	σ	Componentes de la varianza	g. l.	χ^2	<i>p</i>-valor
Nivel 2	36.64497	1342.85416	154	1138.31637	<0.001
Nivel 1	84.44702	7131.29943			

$$CCI_{lectura} = \frac{\tau_{00}}{\tau_{00} + \delta^2} = \frac{1342,85416}{1342,85416 + 7131.29943} = 0,1585$$

(Ecuación 5.16)

Siendo:

τ_{00} el componente de la varianza de nivel 2, y

δ^2 el componente de la varianza de nivel 1.

El resultado de la ecuación anterior nos indica que la variación en las puntuaciones en comprensión lectora en Irlanda debida al segundo nivel de análisis es del 15.85%.

A partir de estos datos, se ha ajustado el modelo incorporando las covariables con influencia significativa en el citado rendimiento académico, quedando compuesto finalmente el modelo condicional por siete variables (Ecuación 5.17).

$$\begin{aligned} Lectura_{ij} = & \gamma_{00} + \gamma_{01} * N2ESCS_j + \gamma_{10} * ESCS_{ij} + \gamma_{20} * GENDER_{ij} + \gamma_{30} \\ & * REPEAT_{ij} + \gamma_{40} * AGE_{ij} + \gamma_{50} * INM_2_{ij} + \gamma_{60} * SCCHANGE_{ij} \\ & + u_{0j} + r_{ij} \end{aligned}$$

(Ecuación 5.17)

En la cual:

$Lectura_{ij}$ se refiere a la puntuación media en comprensión lectora de cada centro;

γ_{00} se refiere a la media de todos los centros irlandeses en comprensión lectora;

γ_{01} es la covariable de nivel 2;

$\gamma_{10} - \gamma_{60}$ son las covariables de nivel 1;

u_{0j} es la diferencia entre la puntuación del centro en comprensión lectora menos la media de todos los centros;

r_{ij} se refiere al residuo de nivel 1.

La Tabla 5.30 presenta las características de las diferentes covariables que componen del modelo final de esta competencia en Irlanda:

Tabla 5.30

Modelo final competencia en comprensión lectora en Irlanda

Efectos fijos	Coefficiente	Error típico	t-ratio	g. l.	p-valor
Intercepto, γ_{00}	256.645451	73.776135	3.479	153	<0.001
N2ESCS, γ_{01}	47.865185	5.114160	9.359	153	<0.001

Efectos fijos	Coefficiente	Error típico	t-ratio	g. l.	p-valor
ESCS, y_{10}	22.500407	1.929044	11.664	3604	<0.001
GENDER, y_{20}	15.872162	2.984483	5.318	2002	<0.001
REPEAT, y_{30}	-37.556016	5.162031	-7.275	3573	<0.001
AGE, y_{40}	16.049949	4.689439	3.423	801	<0.001
INM_2, y_{50}	-4.723502	0.824796	-5.727	679	<0.001
SCCHANGE, y_{60}	-7.023743	2.422461	-2.899	2265	0.004

En el caso de los resultados en la competencia en comprensión lectora de los estudiantes irlandeses, la mayor influencia positiva sobre el rendimiento se encuentra en el nivel socioeconómico, de nuevo a ambos niveles. Por el contrario, presenta una elevada relación negativa con este rendimiento la repetición de curso, al igual que en las competencias anteriores en este mismo país.

Utilizando los componentes de la varianza calculados según este modelo final (Tabla 5.31), se obtuvieron valores de *Pseudo R*² que suponían un 8,46% de varianza explicada en el primer nivel, frente a un 81,97% de varianza explicada en el segundo nivel.

Tabla 5.31

Componentes de la varianza del modelo final para la competencia en comprensión lectora en Irlanda

Efecto aleatorio	σ	Componentes de la varianza	g. l.	χ^2	p-valor
Nivel 2	15.56141	242.15733	153	358.52271	<0.001
Nivel 1	80.79419	6527.70176			

Para determinar la aceptación del modelo condicional, se comprobó el cumplimiento de los tres supuestos indicados relacionados con los residuos a ambos niveles. En la Tabla 5.32 se detallan los estadísticos descriptivos del residuo de primer nivel en comprensión lectora.

Tabla 5.32

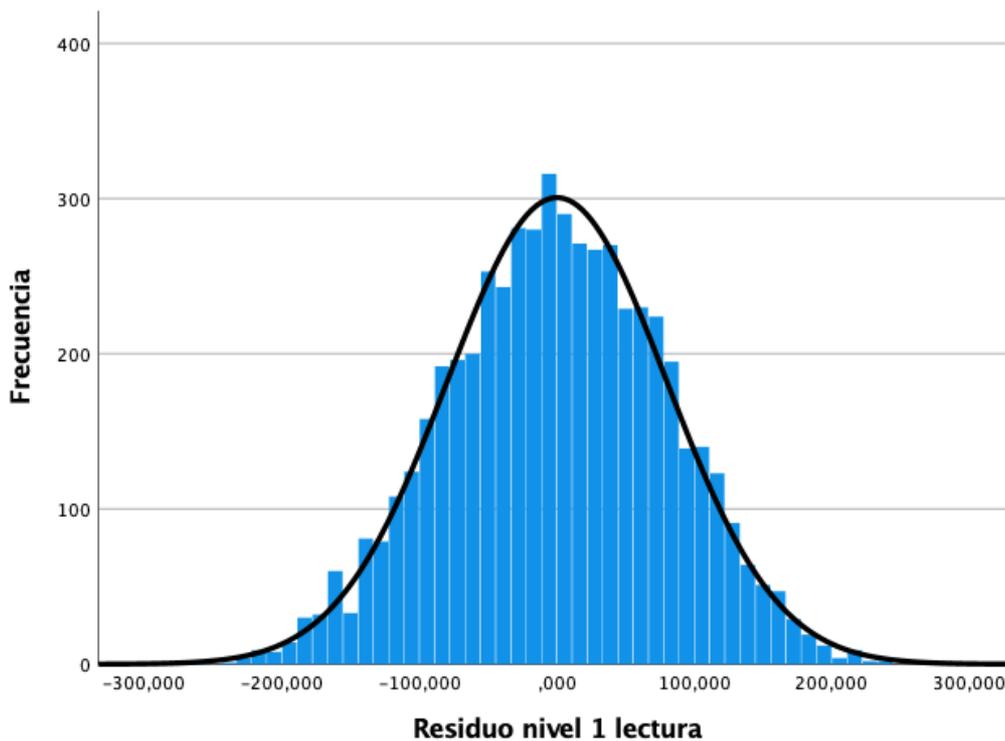
Estadísticos descriptivos del residuo de nivel 1 en comprensión lectora en Irlanda

	N	Mín.	Máx.	Media	σ	Asimetría		Curtosis	
						Valor	Error típico	Valor	Error Típico
Residuo N1	5412	-256.78	256.31	0.18	79.79	-0.04	0.03	-0.28	0.07

La distribución de estas puntuaciones se aproxima a una distribución normal, constando de una media de 0,18, una desviación típica de 79,79 y un valor de significación en la prueba de Kolmogorov-Smirnov de 0,04. Esta distribución se observa en el gráfico siguiente (Figura 5.26):

Figura 5.26

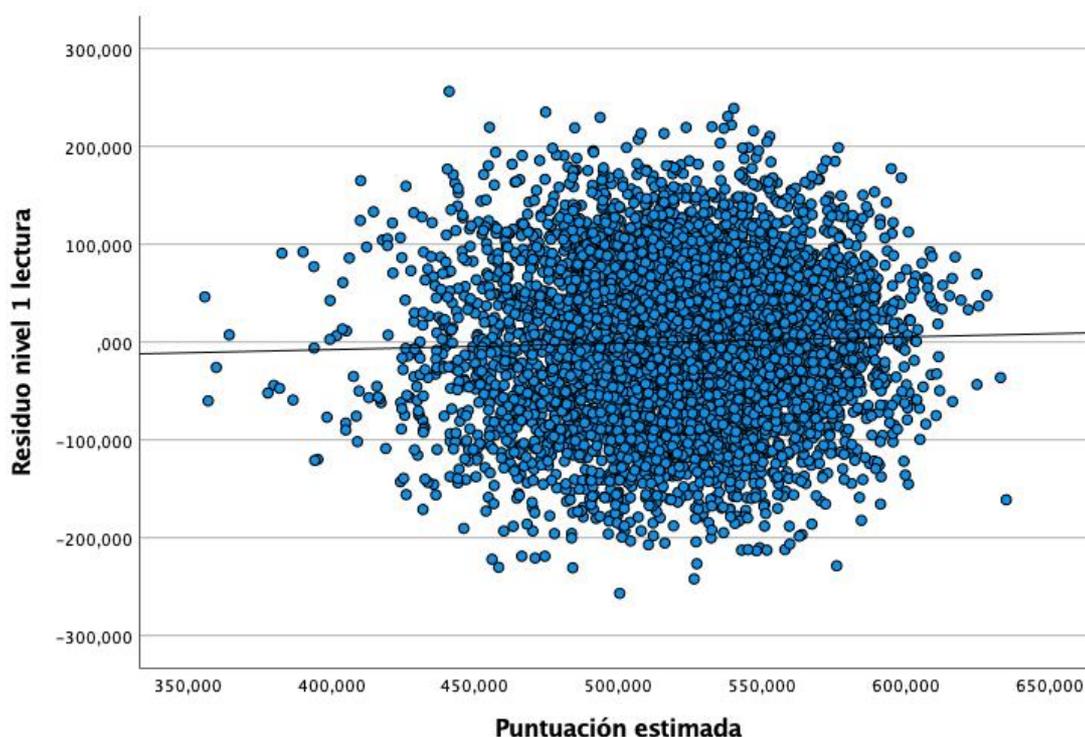
Distribución del residuo de nivel 1 en comprensión lectora en Irlanda



Se incluye la Figura 5.27 para facilitar la comprobación del supuesto de homocedasticidad, pudiendo ratificar su cumplimiento gracias a la misma junto con el valor de R^2 de 9,517E-4.

Figura 5.27

Comprobación del supuesto de homocedasticidad del residuo de nivel 1 de la competencia en comprensión lectora en Irlanda



Las puntuaciones del residuo de nivel 2 presentan una media muy cercana a 0 junto con una desviación típica de 11,22 en una distribución con un valor máximo de 31,08 y un mínimo de -39,14 (Tabla 5.33).

Tabla 5.33

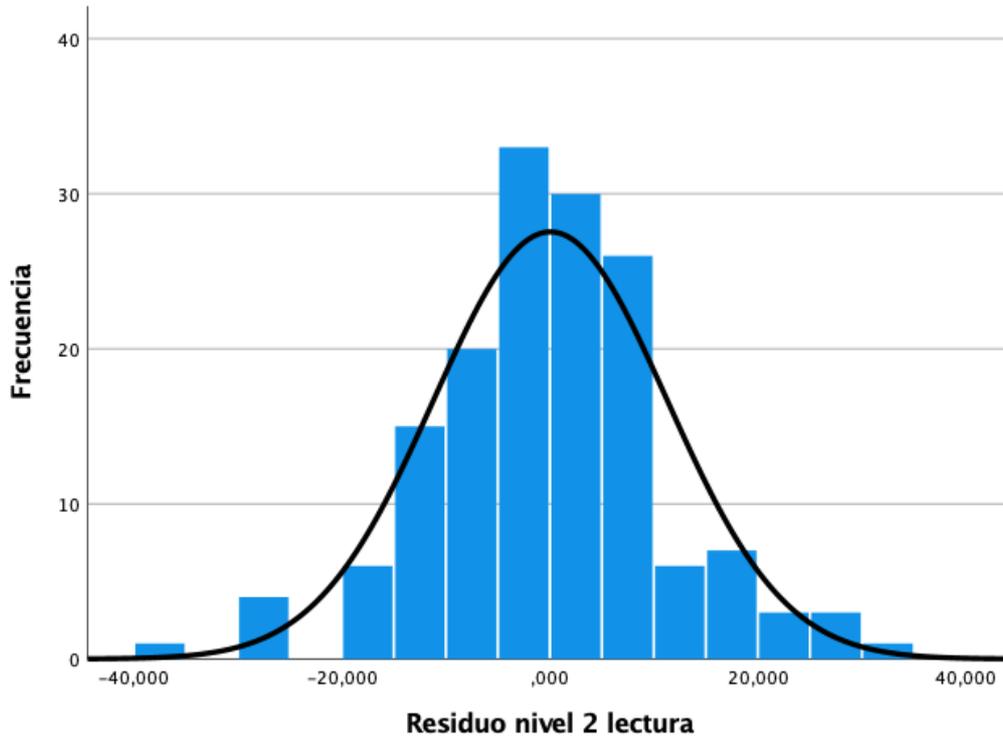
Estadísticos descriptivos para el residuo de nivel 2 en comprensión lectora en Irlanda

	N	Mín.	Máx.	Media	σ	Asimetría		Curtosis	
						Valor	Error típico	Valor	Error Típico
Residuo N2	155	-39.14	31.08	0.04	11.22	-0.05	0.19	1.15	0.39

En cuanto a la normalidad de la distribución, estas puntuaciones se aproximan a la misma presentando un valor en la prueba de Kolmogorov-Smirnov de 0,049 distribuidas conforme a lo retratado en el siguiente gráfico (Figura 5.28):

Figura 5.28

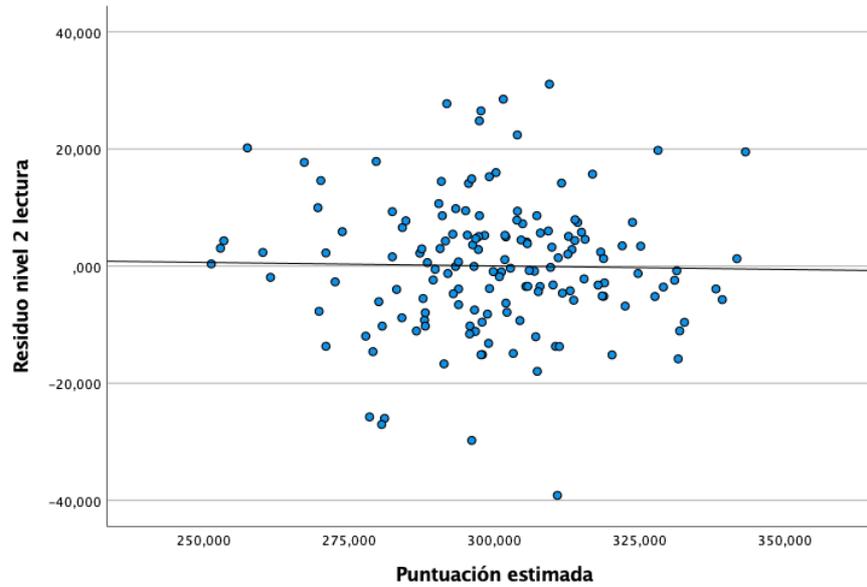
Distribución de las puntuaciones del residuo de nivel 2 para la competencia en comprensión lectora en Irlanda



La igualdad en la dispersión de las puntuaciones se ha comprobado gracias a la Figura 5.29, que, acompañada del valor calculado de R^2 con resultado 3,716E-4, nos permite asumir esta igualdad en la dispersión de los valores.

Figura 5.29

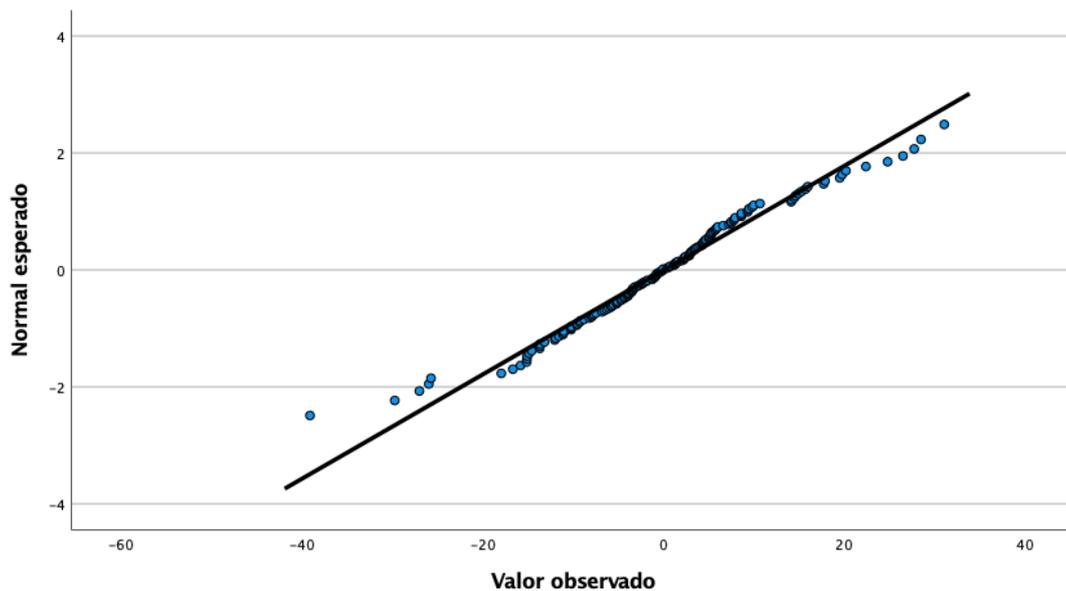
Comprobación del supuesto de homocedasticidad del residuo de nivel 2 en comprensión lectora en Irlanda



A pesar de eso, existen algunos valores extremos en la distribución de este residuo que son fácilmente detectables en el gráfico siguiente (Figura 5.30).

Figura 5.30

Gráfico Q-Q para la distribución del residuo de nivel 2 en comprensión lectora en Irlanda



En función de lo expuesto anteriormente, el modelo estadístico final para la competencia en comprensión lectora en Irlanda es el presentado en la Tabla 5.30 y expresado en la Ecuación 5.17.

- 5.3. Análisis de los factores contextuales del rendimiento académico de los estudiantes portugueses**

- 5.4. Análisis comparativo de los modelos estadísticos multinivel finales**

- 5.5. Selección de centros de alta y baja eficacia**

5.6. Análisis del efecto de las variables procesuales en la eficacia escolar

Para conocer aquellas variables que pueden estar asociadas a la eficacia escolar se inicia el análisis de su posible efecto en la variable criterio comprobando la correlación existente entre las distintas variables de manera previa a la realización de la regresión logística binaria. Esta técnica estadística es la mayoritariamente recomendada para el estudio sobre variables criterio dicotómicas (Kleinbaum & Klein, 2010). Se tuvieron en cuenta tres niveles diferentes de estudio basándonos en información referida a:

1. Características de los centros escolares.
2. Características del profesorado.
3. Características del estudiantado.

Es necesario indicar que las variables referidas a profesorado y estudiantado se agregaron en función de su centro de pertenencia.

El estudio correlacional se desarrolló mediante el cálculo del coeficiente de correlación de Pearson entre las variables procesuales y los residuos de las tres competencias. La realización de este estudio previo sirve de base para la regresión logística binaria, teniendo como consecuencia la inclusión en este segundo procedimiento de aquellas variables con correlaciones estadísticamente significativas con los tres residuos.

La interpretación de los resultados de la regresión logística binaria se construye sobre la magnitud de la relación existente entre una variable predictora y la variable criterio, entendiendo que un aumento de una unidad en la variable predictora supone mayor probabilidad de que el centro sea de alta eficacia.

El estudio del efecto de las variables de proceso en la eficacia escolar se ha realizado de manera individual para cada uno de los tres países en los que se centra esta investigación, por lo que se presenta a continuación este análisis para España, Irlanda y Portugal.

5.6.1. Variables procesuales con influencia en la eficacia escolar en España

El primer análisis realizado para el estudio del efecto de las variables procesuales en la eficacia escolar en España es el cálculo de las correlaciones entre las variables procesuales y los residuos de las tres competencias. De este modo, resultaron presentar

una relación estadísticamente significativa 40 de las 76 variables incluidas en el estudio preliminar, seis derivadas de la base de datos del profesorado y las restantes referidas a variables de proceso de los estudiantes.

Por lo tanto, se procedió con la realización de la regresión logística binaria con esas 40 variables. Siguiendo el mismo procedimiento que en el cálculo de los modelos multinivel, se comenzó el ajuste del modelo eliminando de forma individual y consecutiva las variables que no obtuvieran valores de significación menores de 0,05 hasta alcanzar el modelo final, que puede ser consultado en la siguiente tabla (Tabla 5.52):

Tabla 5.34

Resultados de la regresión logística binaria en España

	B	Odds ratio	Odds ratio convertida	p-valor
Metacognición: resumir	4.581	97.619		<0.001
Metacognición: evaluar la credibilidad	4.538	93.488		<0.001
Percepción de cooperación en la escuela	2.534	12.604		<0.001
Autoconcepto de la lectura: percepción de la competencia	2.186	8.9		0.038
Autoeficacia del profesor en la gestión del aula	1.645	5.181		0.023
Autoeficacia del profesor para establecer relaciones positivas con los alumnos	-2.018	0.133	7.519	0.003
Autoconcepto de la lectura: percepción de la dificultad	-2.711	0.066	15.152	0.009
Uso de las TIC fuera de la escuela para la realización de tareas escolares	-3.376	0.034	29.412	<0.001
Bienestar subjetivo: afecto positivo	-3.438	0.032	31.25	<0.001

Nota. La columna *Odds ratio convertida* incluye la probabilidad de las variables con una OR negativa para facilitar su comparación (1/OR).

Este modelo lo componen nueve variables referidas a aspectos relacionados con los profesores (2) y los estudiantes (7). Las dos variables que presentan una relación más alta con la probabilidad de pertenecer a un centro de alta eficacia son las relacionadas con la metacognición de los estudiantes, destacando entre el grupo de variables con valores positivos. Analizando el valor obtenido de *Odds Ratio* de estas dos variables, se determina que un aumento de una unidad en estas variables se asocia con un aumento de más de 90

veces en la probabilidad de la alta eficacia. Del mismo modo, el resto de las variables que presentan una relación significativa positiva con la eficacia escolar se refieren a la percepción de cooperación en la escuela, al autoconcepto de los estudiantes sobre su competencia en comprensión lectora y a la eficacia auto percibida por el profesor en la gestión del aula.

En cambio, las variables con una relación negativa con la eficacia escolar se refieren a aspectos como la autoeficacia del profesor, en este caso en cuanto al establecimiento de relaciones positivas con los estudiantes; el autoconcepto de la lectura de los estudiantes en cuanto a la percepción de la dificultad; el uso de TIC fuera de la escuela para tareas escolares y su bienestar subjetivo. La variable que presenta una *Odds Ratio* más elevada en este grupo es esta última referida al bienestar subjetivo de los estudiantes, suponiendo que cuanto mayor sea ese bienestar subjetivo percibido por el estudiantado, mayor probabilidad tendrá ese centro de presentar una baja eficacia.

A continuación, se procedió a la comprobación de la bondad de ajuste del modelo y de su precisión a la hora de predecir la clasificación de los centros educativos. Se puede afirmar que el ajuste del modelo es bastante bueno, R^2 de Nagelkerke = 0,740, lo que supone que el modelo presentado explica aproximadamente el 74% de la varianza en los datos. En cuanto a la precisión del modelo predictivo (Tabla 5.53) también se obtiene un valor más que adecuado, prediciendo correctamente un 88,9% de los casos, prediciendo con acierto el 89,8% de los centros de baja eficacia y el 87,8% de los centros de alta eficacia.

Tabla 5.35

Precisión del modelo predictivo de la eficacia escolar en España

		Predicho		% correcto
		Baja eficacia	Alta eficacia	
Observado	Baja eficacia	177	20	89.8
	Alta eficacia	22	159	87.8
% global				88.9

5.6.2. Variables procesuales con influencia en la eficacia escolar en Irlanda

La correlación de Pearson calculada con los datos relativos a Irlanda informaba de que tan solo ocho variables de las 54 incluidas inicialmente⁸ presentaban una relación significativa con los residuos de las tres competencias, todas ellas pertenecientes al grupo de variables referidas al estudiantado. Estas variables han sido las empleadas para el cálculo de la regresión logística binaria.

El modelo final lo compusieron tres variables, dos de ellas relacionadas con la probabilidad de que se tratara de un centro de alta eficacia y la restante asociada a la baja eficacia. Para comprobar el efecto de esas variables con la eficacia de los centros se puede consultar la Tabla 5.54.

Tabla 5.36

Resultados de la regresión logística binaria en Irlanda

	B	Odds ratio	Odds ratio convertida	p-valor
Autoconcepto de la lectura: percepción de competencia	7.278	1447.561		0.003
Autonomía percibida en relación con el uso de las TIC	5.127	168.9497		0.031
Uso de las TIC fuera de la escuela para la realización de tareas escolares	-5.929	0.003	333.333	0.005

Nota. La columna *Odds ratio convertida* incluye la probabilidad de las variables con una OR negativa para facilitar su comparación (1/OR).

En este caso, se observa que la variable referida al autoconcepto en la lectura de los estudiantes irlandeses específicamente en la percepción de su nivel de competencia tiene una enorme influencia en la probabilidad de un centro de pertenecer al grupo de alta eficacia en comparación con la probabilidad de pertenecer al grupo opuesto. Centrándonos en el uso de TIC fuera de la escuela para la realización de los deberes, el valor de *Odds ratio convertida* nos indica una disminución significativa de 333,333 veces de pertenecer al grupo de centros de alta eficacia. Se puede concluir que todas las variables incluidas en el modelo presentan fuertes asociaciones con la eficacia escolar.

⁸ Se excluyen las referidas al profesorado.

En el caso del modelo predictivo de la eficacia escolar en Irlanda, se obtuvo un valor de R^2 de Nagelkerke = 0,657, lo que indica que el ajuste del modelo es adecuado explicando aproximadamente un 65,7% de la varianza en los datos.

Para comprobar la precisión del modelo predictivo se propone la tabla siguiente (Tabla 5.55), según la cual se predice con acierto la eficacia escolar del 87,9% de los centros educativos.

Tabla 5.37

Precisión del modelo predictivo de la eficacia escolar en Irlanda

		Predicho		% correcto
		Baja eficacia	Alta eficacia	
Observado	Baja eficacia	25	4	86.2
	Alta eficacia	3	26	89.7
% global				87.9

5.6.3. Variables procesuales con influencia en la eficacia escolar en Portugal

5.7. Validación del cuestionario para la evaluación de la eficacia auto percibida por el profesorado de secundaria en activo

Algunas investigaciones sugieren la combinación de métodos cuantitativos y cualitativos a la hora de realizar un estudio de validez de contenido en profundidad (Pedrosa et al., 2013). Por tanto, para la validación del cuestionario elaborado se diferenciaron dos fases centradas cada una de ellas en un aspecto: comprobar la validez de contenido del cuestionario y comprobar su fiabilidad. Cabe en este momento aclarar estos dos conceptos, entendiendo por validez del instrumento el “grado en que el instrumento mide lo que se pretende medir” (Lacave Rodero et al., 2015, p. 136) y por fiabilidad “la confianza que se concede a los datos que se obtienen con el mismo (...) relacionada con la coherencia o consistencia interna y la precisión de las medidas recopiladas” (Lacave Rodero et al., 2015, p. 136). Estas dos fases se corresponden con:

1. Evaluación por parte de expertos (jueces) – para asegurar la validez de contenido.
2. Aplicación piloto del cuestionario en diferentes centros de educación secundaria – con el fin de analizar la fiabilidad del instrumento diseñado.

En el estudio de la validez de contenido del cuestionario centrado en la valoración por parte de jueces expertos se establecieron tres criterios a valorar para cada ítem de cada dimensión: coherencia, relevancia y claridad, con cuatro categorías de respuesta. Del mismo modo, a la hora de valorar las dimensiones en su conjunto se plantearon otros dos criterios: suficiencia de los ítems y relevancia de la dimensión para el conjunto del cuestionario; también con cuatro categorías de respuesta. El documento de validación en el que se detallan todos estos criterios se envió a los jueces para que valorasen la construcción del cuestionario y se puede consultar en el Anexo 3. En esta validación participaron un total de 15 expertos, pudiendo consultar las puntuaciones otorgadas por los jueces a cada uno de los ítems en cada criterio en el Anexo 4.

Además, buscando dotar de ese corte cualitativo a la validación, se incluyó un apartado de comentarios o sugerencias para que los evaluadores pudieran incluir ese tipo de información tanto sobre los ítems como sobre la propia construcción de las dimensiones y del instrumento en sí mismo.

Para comprobar la validez de contenido del cuestionario y tras realizar la oportuna revisión de la literatura, se optó por utilizar el índice de validez propuesto por Aiken, conocido como V de Aiken (Pedrosa et al., 2013), para determinar el nivel de acuerdo o

desacuerdo entre los jueces. Este índice tiene en cuenta el número de categorías ofrecidas a los expertos, así como el número de expertos participantes. La fórmula matemática empleada para su cálculo es la siguiente (Ecuación 5.30):

$$V = \frac{\chi - l}{k}$$

(Ecuación 5.30)

Donde:

V es el valor del Coeficiente de la V de Aiken.

χ es el promedio de las calificaciones de todos los jueces.

l es la calificación mínima.

k es el rango de calificaciones (calificación máxima menos calificación mínima).

La interpretación de los resultados de este coeficiente se hizo estableciendo 0,8 como punto de corte, considerando aquellos valores inferiores a 0,8 como indicadores de un índice de desacuerdo importante en el que centrar nuestra atención a la hora de modificar el instrumento teniendo en cuenta las sugerencias facilitadas por los jueces.

En la tabla siguiente (Tabla 5.58), se presentan los estadísticos descriptivos de cada ítem en cada criterio según las valoraciones de los expertos participantes:

Tabla 5.38

Estadísticos descriptivos para cada ítem y V de Aiken

		COHERENCIA			RELEVANCIA			CLARIDAD		
		Media	σ	V	Media	σ	V	Media	σ	V
D1	1.1	3,93	0,25	0,98	3,87	0,34	0,96	3,73	0,44	0,91
	1.2	3,8	0,54	0,93	3,6	0,71	0,87	3,53	0,88	0,84
	1.3	3,93	0,25	0,98	4	0	1	3,93	0,25	0,98
	1.4	3,47	0,81	0,82	3,6	0,61	0,87	3,2	0,91	0,73*
D2	2.1	3,53	0,88	0,84	3,4	1,02	0,8	3,13	1,09	0,71*
	2.2	3,29	0,96	0,76*	3,21	0,94	0,74*	3,5	0,91	0,83
	2.3	3,64	0,81	0,88	3,57	0,82	0,86	3,43	0,9	0,81
	2.4	3,36	0,97	0,79*	3,29	0,88	0,76*	3,5	0,91	0,83
D3	3.1	3,93	0,26	0,98	4	0	1	3,79	0,56	0,93
	3.2	3,93	0,26	0,98	3,64	0,81	0,88	3,69	0,46	0,9
	3.3	3,14	1,06	0,71*	3,36	1,11	0,79*	3,53	0,62	0,84
D4	4.1	4	0	1	3,93	0,25	0,98	3,33	0,94	0,78*

		COHERENCIA			RELEVANCIA			CLARIDAD		
		Media	σ	V	Media	σ	V	Media	σ	V
	4.2	3,93	0,26	0,98	3,93	0,26	0,98	3,86	0,52	0,95
	4.3	3,43	0,9	0,81	3,43	1,05	0,81	3,86	0,35	0,95
	4.4	3,4	0,95	0,8	3,64	0,61	0,88	3,64	0,61	0,88
	4.5	3,53	1,02	0,84	3,6	0,88	0,87	3,67	0,87	0,89
D5	5.1	3,87	0,34	0,96	3,73	0,57	0,91	3,67	0,6	0,89
	5.2	3,8	0,75	0,93	3,8	0,75	0,93	3,73	0,77	0,91
	5.3	4	0	1	4	0	1	3,87	0,5	0,96
	5.4	3,93	0,25	0,98	4	0	1	3,67	0,6	0,89
	5.5	3,47	0,72	0,82	3,4	0,8	0,8	3,87	0,34	0,96
	5.6	3,73	0,68	0,91	3,73	0,77	0,91	3,67	0,87	0,89
D6	6.1	3,93	0,26	0,98	3,87	0,34	0,96	3,73	0,68	0,91
	6.2	3,93	0,26	0,98	3,8	0,54	0,93	3,87	0,34	0,96
	6.3	3,5	0,91	0,83	3,57	0,9	0,86	3,33	1,19	0,78*
	6.4	3,43	0,82	0,81	3,53	0,72	0,84	3,6	0,61	0,87
	6.5	3,5	0,73	0,83	3,71	0,45	0,9	3,79	0,41	0,93
	6.6	3,73	0,57	0,91	3,71	0,59	0,9	3,67	0,7	0,89
D7	7.1	3,93	0,25	0,98	3,93	0,25	0,98	3,87	0,34	0,96
	7.2	3,8	0,54	0,93	3,87	0,34	0,96	3,4	0,71	0,8
	7.3	4	0	1	4	0	1	3,93	0,25	0,98
D8	8.1	3,93	0,25	0,98	3,87	0,5	0,96	4	0	1
	8.2	3,73	0,77	0,91	3,67	0,87	0,89	3,47	0,88	0,82
	8.3	3,73	0,77	0,91	3,67	0,79	0,89	3,87	0,34	0,96
	8.4	3,8	0,75	0,93	3,73	0,77	0,91	3,73	0,77	0,91
OTROS	O1	3,86	0,35	0,95	3,85	0,36	0,95	3,6	0,88	0,87
	O2	3,86	0,35	0,95	3,86	0,35	0,95	3,8	0,4	0,93

Nota. σ se refiere a los valores de la desviación típica de las puntuaciones; V se refiere al valor del Coeficiente V de Aiken. * se utiliza para resaltar los valores de V de Aiken inferiores a 0,8.

Como se observa en la tabla anterior, son pocos los casos en los que el índice de acuerdo interjueces es inferior a 0,8. Esto se debe a que la mayoría de los jueces han considerado que los ítems cumplían con la función planteada teóricamente y, además, el acuerdo entre los diferentes expertos fue muy elevado.

Algunos de los expertos que colaboraron en la validación del cuestionario incorporaron sugerencias u observaciones en alguno de los apartados. Estas sugerencias se tuvieron en cuenta, especialmente en el caso de los ítems en los que el Coeficiente V

de Aiken presentaba valores por debajo de 0,8 en alguno de los tres criterios establecidos. Estos ítems fueron: 1.4, 2.1, 2.2, 2.4, 3.3, 4.1 y 6.3. Los cambios propuestos supusieron la modificación del instrumento, haciendo que finalmente constara de 35 ítems agrupados en 7 dimensiones a valorar en una escala tipo Likert con 5 posibilidades de respuesta siendo 1 “totalmente en desacuerdo”, 2 “parcialmente en desacuerdo”, 3 “término medio”, 4 “parcialmente de acuerdo” y 5 “totalmente de acuerdo”, seguidos por una última pregunta de categoría abierta. El cuestionario final con las correcciones oportunas puede consultarse en el Anexo 5.

A continuación, se procedió con la aplicación del cuestionario a diferentes profesores de educación secundaria que se prestaron a colaborar a modo de prueba piloto del mismo con el fin de poder realizar un Análisis Factorial Exploratorio (AFE). Se obtuvieron 104 respuestas.

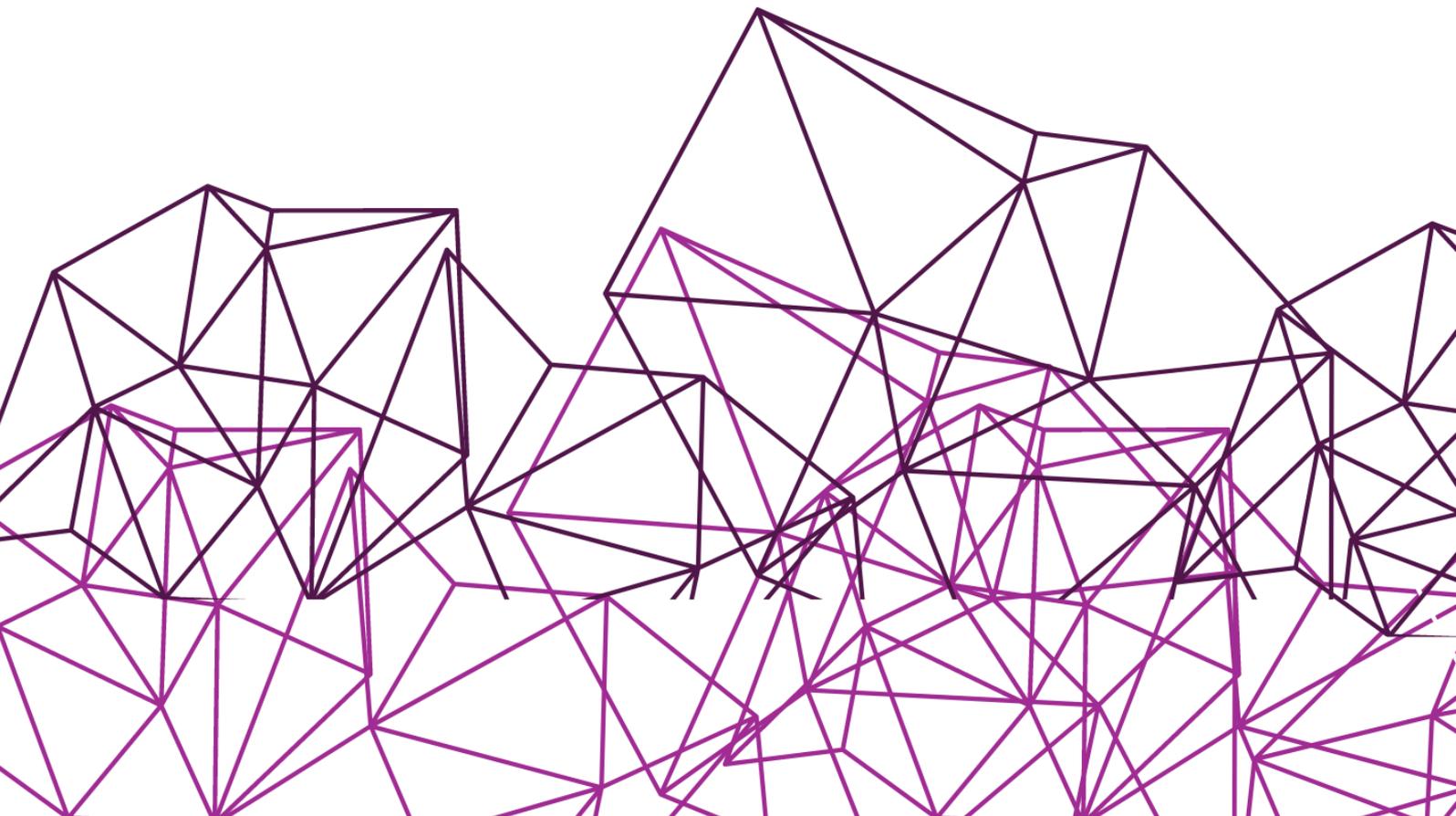
La revisión realizada por Vargas y Mora (2017) en relación con los tamaños de muestra necesarios para poder llevar a cabo este tipo de análisis (AFE) sugería diferentes criterios a cumplir para garantizar la validez de los mismos. Estos criterios podían depender de:

1. La cantidad absoluta de casos, en los que sería necesario que $n \geq 200$ siguiendo lo propuesto por diferentes autores (Catena Martínez et al., 2003; Hair et al., 2014; Stevens, 2009).
2. El número de casos por parámetro, para lo cual $n = 20$ según lo planteado por Kline (2011).

En este caso, no se cumple con ninguno de estos dos pretextos, puesto que la muestra estaba compuesta por 104 sujetos, siendo $n < 200$ en cuanto a cantidad absoluta de casos y, según el número de casos por parámetro, no se alcanzaban tampoco los 140 casos necesarios para la realización del AFE (partiendo de 7 dimensiones, con 20 casos por dimensión). La validación del instrumento quedó limitada en términos de fiabilidad e imposibilitó la aplicación del cuestionario asegurando las garantías estadísticas mínimas.

Partiendo de lo expuesto con anterioridad, se procedió a calcular el valor del estadístico Alfa de Cronbach para comprobar la fiabilidad teórica del instrumento en función de las respuestas obtenidas en la prueba piloto (insuficientes para garantizar la completa fiabilidad del cuestionario). El resultado fue de 0,969, lo que indicaba una alta fiabilidad del instrumento en su totalidad a pesar de las limitaciones indicadas.

6. CONCLUSIONES, LIMITACIONES Y PROSPECTIVA



Conclusions, limitations and prospects

In this last section, the discussion and conclusions are presented, taking as a starting point the research objective of generating a framework of good educational practices related to school effectiveness at the international level. Similarly, limitations and possible areas for future research are presented.

Contextual variables related to academic performance were studied by analysing the results provided by the PISA 2018 tests for the three basic skills of 15-year-old students (science, mathematics and reading comprehension) in Spain, Ireland, and Portugal. In this study, we used the multilevel analysis methodology, thanks to which we determined the existence of two variables with a consistent influence on the academic performance of students regardless of the context analysed: their socio-economic and cultural level and grade repetition. The first of these was always positively related to performance, while the other indicated a negative influence on performance in the three countries studied. In addition, other variables with an influence on students' academic performance were also detected. In this way, the first of the specific objectives could be achieved: to select the context variables at school and student level that have significant relationships with academic performance.

Thanks to the second-level residuals generated from the final hierarchical-linear models, the high- and low-effectiveness schools were selected in response to the second of the specific objectives.

Based on this selection and characterisation of schools in terms of their school effectiveness, the relationship between various variables extracted from the PISA 2018 questionnaires referring to students, teachers and schools and their school effectiveness was studied. This third phase of the study was used to achieve the third specific objective: to analyse the variables related to school effectiveness to determine the relationships between these variables and the criterion variable (school effectiveness).

In conclusion, an analysis of the results obtained is presented, comparing them with the existing literature and organising the presentation in four sections. It begins with a reflection on the variables that have shown a relationship of some kind with academic performance studied in more than one country. Once these results have been discussed, the detection of high- and low-effectiveness schools is analysed, followed by a discussion of those variables related to school effectiveness. Finally, the limitations detected in this research at both the theoretical and methodological levels and the possible lines of future

research derived mostly from these difficulties encountered, as well as from the results obtained, are presented.

6.1. Factors linked with academic performance

According to what has been presented in this document, it has been possible to see how educational research has been based on the study of the factors associated with students' academic performance. In section 5.4.1 of this document, the variables with a significant influence on the performance analysed were presented. It also indicated, as considerations prior to the elaboration of the statistical models, the existing correlations between the different variables to properly interpret the results obtained.

After the multilevel modelling, and considering the correlations between variables, some conclusions derived from the analysis carried out and their relationship with the current research are set out below, specifically presenting those variables that have been found to be significant in the models of at least two of the countries studied.

6.1.1. Socio-economic and cultural status

According to research such as that presented in the systematic literature review (Chapter 2), socio-economic status has been shown to be a key factor in academic performance. This phenomenon has been studied internationally (Lenkeit, 2012; Martins & Veiga, 2010; Pomianowicz, 2021; Sousa et al., 2012; Yetişir & Bati, 2021) and could be confirmed as one of the determinants of achievement in the study at hand.

This factor appears as a significant variable in the three countries studied, although with a different contribution depending on the context. In the case of Spain, it is of great positive importance if it refers to the individual socio-economic level of the students, specifically being the variable with the greatest influence on performance in the three competences. Comparing this result with those of other studies (Gamazo, 2019), it is striking that the average socio-economic status of schools is not a significant factor in the PISA 2018 data. However, this is in direct contrast to the results obtained in both Ireland and Portugal, where the socio-economic status variable is shown to have a strong relationship with academic performance at both levels (student and school). In the case of Ireland, as in the case of Spain, the socio-economic status of students is the variable with the strongest influence on performance in all three competences. In Portugal, on the

other hand, although both variables have a positive influence on performance, the variable that proves to have the greatest weight on all three competences is the students' grade.

Education systems should try to mitigate or compensate for the existing differences between students with higher and lower family incomes by promoting a more equitable education system. Equity in education is understood as “la medida en que las condiciones familiares y sociales del alumno afectan a su éxito escolar en un sistema educativo” (Gortázar, 2019, p. 17). To achieve an equitable system, it is necessary to develop educational policies that ensure equal opportunities for students regardless of their contextual characteristics.

In Spain, the promotion of equity is not a new element in education policy. The existence of a long tradition of research on the effect of socio-economic level on achievement has meant that educational equity has occupied and continues to occupy an important place in the national public debate. Work is currently being carried out in the search for a more equitable education system both through research linked to equity (Fernández García & Andrés Cabello, 2019; García-Herrero et al., 2024), and through the formulation of educational policies that take into account student diversity and guarantee equal opportunities. Some of these educational policies are proposed by the Ministry of Education, Vocational Training and Sport of the Spanish Government and include: guaranteeing early education, providing adequate resources, both human and material, to schools according to their specific needs or providing greater autonomy to schools so that they are able to implement projects appropriate to their environment and needs (INEE, 2022). There are various factors that can help to promote an equitable education system, such as those related to the quality of teachers (training, involvement in the teaching-learning process, etc.) (García-Herrero et al., 2024), educational expenditure and the distribution of funding in accordance with the needs of students.

However, socio-economic status at the individual level is not the only factor influencing achievement, with contextual or school socio-economic status also being of high importance (Lizasoain et al., 2007). The unequal distribution of students in schools according to their socio-economic status not only has a direct effect on their performance, but also helps to exacerbate a country's social and economic inequalities (Murillo & Martínez-Garrido, 2018). In relation to this contextual effect, there are two main perspectives of statistical analysis: school segregation and the isolation index (OECD, 2019). School segregation is understood as the unequal distribution of students in schools according to personal characteristics or social and cultural backgrounds “la distribución

desigual de los estudiantes en las escuelas en función de sus características personales o procedencia social y cultural" (Murillo & Duk, 2016). The isolation index, on the other hand, refers to the likelihood of students with a set of characteristics to be enrolled in schools with different characteristics. A thorough study of these two issues can provide valuable conclusions for an equitable system to achieve higher levels of equity, limiting the negative influence that aspects of this type can have on the teaching-learning process.

6.1.2. Grade repetition, grade and age of the students

The second variable that consistently appears as one of the most influential on academic performance is grade repetition. This variable has been associated in all cases with academic disadvantage, implying that having repeated a year leads to negative consequences for students in terms of educational achievement. On many occasions, students who repeat a year may see their level of self-esteem reduced, observing greater difficulties for their normal academic development, even leading to academic dropout.

Of the three countries studied, repetition proves to be the variable with the strongest negative influence on performance in two of the three competences in Spain, science and reading comprehension, and Ireland, mathematics and reading comprehension (although there is not much difference with the most negative variable) and in all three competences in Portugal.

Repeating students in Spain and Portugal represent similar values in the respective samples (close to 20%), but the weight of this variable in the statistical models differs considerably depending on the country, with a greater importance in the Spanish case. In Ireland, around 6% of students have repeated a year, showing that this variable has a significant weight in the reduction of academic performance.

The fact that in Spain and Portugal there are similar values in terms of the number of repeating students, but the influence of this variable depends on the place we are studying, raises an important debate about the usefulness or not of this measure to improve student learning. Educational research is of the opinion that this is not an effective measure, since students who repeat grades do not show an improvement in their performance (Cappella & Weinstein, 2001; Gómez, 2013; Jimerson, 2001). Therefore, the question is whether this measure is beneficial in terms of promoting student achievement or whether it is detrimental to their education. Some research alludes to the way in which grade repetition is used, since, taken as the only measure to indicate the need to improve a student's

performance (or that he/she has not reached the required level of competence), and without being complemented by other types of educational reinforcement, both prior and subsequent, its positive implications are diminished. Such pedagogical practices should be rethought because, while it is true that grade repetition is a key factor in the decline in academic performance, the direction of the relationship between predictor and criterion variable is not clear. That is, repeating students do not perform worse because of their previous performance, but also because of the fact that they have repeated a grade (Choi de Mendizábal, 2016; Jacob & Lefgren, 2009; Manacorda, 2012). In view of the above, we recommend further study of the relationship between this variable and academic performance based on the pedagogical practices carried out in each country.

The importance of checking the correlations between the variables prior to the elaboration of the hierarchical models when interpreting the results was indicated earlier. In the case of grade repetition, both in Spain and Portugal, this variable and the students' grade itself show a medium/high correlation (0.68 in Spain and 0.668 in Portugal), having opted for their inclusion in the statistical modelling as the values are not too high. Even so, the correlation between both variables must be considered, so the results of both variables are interpreted jointly and with caution.

Although these two variables refer to different characteristics of the students, they are based on some common concepts for their construction: the grade of the students themselves and their age, variables also considered independently in this analysis. Due to the close relationship between these three variables, it has been decided to include them in the same section and thus interpret them jointly.

In contrast to the directly negative relationship between grade repetition and academic performance, age and grade level have a positive influence on academic performance. This makes sense, as this is interpreted to mean that the higher these values, the higher the performance.

PISA measures the performance of students aged between 15 years and 3 months and 16 years and 2 months at the time of the assessment. The age of students may be associated with their level of cognitive development, which would imply a greater maturity to cope with this assessment test. Hence, the existence of differences in performance according to the month of birth has been studied and demonstrated (Valdés, 2023). Similarly, the fact that a student of this age is in higher grades seems to indicate that he/she has already achieved the expected competencies for his/her corresponding

academic level according to his/her age, which implies a positive relationship with performance.

Student age was found to significantly influence performance in all three competences in Spain and Ireland. While grade was found to be the most important variable in Portugal (in all competences) and one of the most important in the case of Spain.

6.1.3. Student Gender

The gender of students has been shown to be another vitally important variable in explaining academic performance in different disciplines. Educational research has shown that boys outperform girls in mathematics and science, while the opposite is true for reading literacy.

This is a variable which again features consistently in the models, except for reading literacy in Portugal and science literacy in Ireland.

Historically, it has been suggested that there is a gender gap in science, technology, engineering, and mathematics (STEM) subjects in favour of male students. The results presented in this research support this assertion, although a much more detailed and focused study of this gender gap in scores would be needed to confirm this unequivocally. Results from the most recent PISA test (PISA 2022) indicate that, at least in the case of science proficiency, the gender gap has narrowed (OECD, 2023). Similarly, other research over the last decade has found similar results (Zhu et al., 2018).

In recent decades, several countries have made progress in reducing the gender gap through initiatives that encourage girls' and women's participation in these subjects or by ensuring equal access to education. One example of such initiatives is Girls who code, an organisation dedicated to providing STEM-focused educational programmes for girls aged 13-17, thus trying to reduce this gender gap (Girls Who Code, 2024). This suggests that this gap is not innate but can be avoidable if it is addressed (OECD, 2019).

6.1.4. Number of school changes

This variable indicates the number of changes of school experienced by the student during his/her academic history. In both Spain and Ireland, with the exception of science proficiency in the latter country, the number of school changes has been shown to

negatively affect student performance, although to a lesser extent than the variables analysed so far.

The fact that a student changes school may be due to a variety of reasons and, in turn, may be associated with a few consequences that hinder the learning process. Some of these consequences that may result from school changes are measured in PISA, for example: the sense of belonging to the school, student satisfaction or student self-efficacy (Ministerio de Educación y Formación Profesional, 2019), all of which are factors related to school climate.

Most of the studies linked to changes of school are related to changes of educational stage (Azorín Abellán, 2019; Flores Macías & Gómez Bastida, 2010; Guevara Molina, 2022; Lorenzo, 2010), with this variable referring to the number of changes of school being one of the least studied to date and generating a new field where educational research can make headway. With a view to establishing educational policies and practices that reduce the negative impact of this variable, and bearing in mind that the variable itself is difficult to intervene directly, it would be advisable to carry out exhaustive studies to check what other factors may be moderating the relationship between the number of school changes and performance, such as the feeling of belonging to the school or the school climate, in order to be able to intervene on them.

6.1.5. Language spoken at home and immigration status

The language spoken at home or mother tongue variable was found to significantly influence performance in three of the nine cases presented. In PISA, its measurement is specifically portrayed in a question asking students about the language most commonly used at home, with a choice between the test language or another.

Research has shown that mother tongue is a significant factor in academic performance at different educational levels (Mateos Claros et al., 2019; Özdemir, 2016; Riederer & Verwiebe, 2015). This study corroborates those results, but only in the case of Ireland when working on performance in mathematical competence and in Spain when studying its relationship with reading comprehension and performance in science.

In this case, the results are not similar, since in Ireland this variable has a positive influence, while in Spain the relationship is negative. In this sense, Spanish students whose mother tongue is different from the one in which they took the test obtained worse results in science and reading comprehension, coinciding with studies carried out in other

contexts (Entorf & Minoiu, 2005). In contrast, Irish students with a different mother tongue than that of the test perform better in mathematical competence.

A particular feature of the Spanish case is the coexistence of co-official languages in the territory which, depending on the Autonomous Community, have a greater or lesser weight in the educational curriculum. Perhaps it would be of interest to study the relationship between mother tongue and performance in the Autonomous Communities with co-official languages to be able to make informed decisions on the subject and to determine what influence this has in Spain. One approach to the research proposed was that carried out by Doncel and Cabrera (2020), where the authors analysed the different styles of bilingualism used in Spanish education on performance in the three basic PISA competences based on PISA 2015 data.

In Ireland, on the other hand, the language of instruction is English, although Gaelic is also used in some areas.

This variable is also closely associated with migratory status, since many students who have had to emigrate from their countries of origin have, therefore, a different mother tongue to that of their new country of residence. UNESCO advocates mother tongue education as far as possible, since the fact that pupils are taught in a language other than their mother tongue hinders effective learning (UNESCO, 2022b), an argument that would support the results obtained in Spain but differs from the Irish results.

Analysing this related variable, migration status, we observe that it has a significant relationship with performance in all three skills in Ireland, in all cases being associated with a reduction in performance. These results should be interpreted carefully, as these two variables were found to have a medium-high correlation (0.548). Some studies have analysed the influence of these two variables on academic performance in Ireland, and the authors may be providing clues as to the unequal influence of these two variables based on the type of immigration in this country (Darmody & Smyth, 2018). Highly skilled immigration that finds it difficult to obtain jobs commensurate with their training. This is why we cannot make unequivocal statements about the influence of immigration status or mother tongue on the performance of Irish students, and why more in-depth studies should be carried out to understand the reasons for the results we have obtained.

6.1.6. Teachers with doctoral studies

The only variable related to teacher characteristics that has been shown to have a significant relationship with academic performance is the rate of teachers with level 6 studies in the International Standard Classification of Education (ISCED), i.e. with doctoral studies. This variable influences academic performance in mathematics and science in Spain and in science in Portugal.

While in Spain this variable has a positive relationship with performance, the opposite is true in Portugal. It is interesting that in contexts as similar as these two territories of the Iberian Peninsula, such differences are generated. Portugal has an index with an average value of 0.098, whereas in Spain this value is 0.033. The number of teaching staff with doctoral studies in Portugal far exceeds the same group in Spain, almost tripling the Spanish rate. These results seem to indicate that the higher the rate of teachers with an academic background equivalent to ISCED level 6, the more likely it is that academic performance will decrease. But this is not a hypothesis that we can confirm without a doubt, among other aspects, because the Spanish results must be taken carefully due to the high percentage of imputed values referring to this variable (approximately 20%). In this respect, this makes it difficult for us to really establish the relationship between the educational level of teachers and the performance of 15-year-old students adequately.

6.1.7. Duration on Early Childhood Education and Care

Some international agencies, such as UNESCO, suggest the importance of early childhood education (2022a), in part because of their malleability and their capacity to learn and absorb new knowledge, attitudes, etc. As has long been asserted, children are sponges (Montessori, 1986). This fact has also been studied in educational research indicating, or rather confirming, this relationship (Cordero Ferrera et al., 2013; Martínez-Abad & León, 2023).

The information obtained from the study presented here is contradictory, since in Portugal this variable is shown to have a positive influence on students' academic performance in mathematics and reading comprehension, with a negative influence in the case of Irish students' mathematical competence. In all cases, the effects are not very high.

The global trend is to promote education during this stage by applying measures aimed at the institutions themselves, at families or at the community; measures such as

increasing funding for early education, favoring equity and social justice. This has not been one of the objectives of this research, but it could be of great interest to study the relationship between the financing of education in different countries and the academic performance of their students, which would make it possible to verify whether increased spending on education helps to improve performance.

6.2. Factors influencing school effectiveness

Following the structure outlined in the previous section in which the factors related to academic performance were presented, the conclusions derived from the study of the variables that have been shown to influence the effectiveness of schools are commented on and discussed below. Special attention is given to those factors that have appeared repeatedly in at least two of the three countries studied, in accordance with the results presented in section 5.6 of this document (Analysis of the effect of procedural variables on school effectiveness).

In order to be able to make a correct interpretation of the results, it is necessary to take, first of all, the values of the initial ICC calculated prior to the elaboration of the multilevel models, which indicates the percentage of variance in the scores due to the characteristics of the second level of analysis, the school level. The following table shows the values of this index (table 6.1).

Table 6.1

Initial CCI values for each competence in each country

	Spain	Ireland	Portugal
Maths	13,14%	16,29%	30,94%
Science	11,39%	14,48%	28,29%
Reading	14,21%	15,85%	29,18%

The ICC values indicate, in addition to the appropriateness of the application of the multilevel models for the characteristics that are related to academic performance, the existence of differences in performance that could be due to differences between schools in each country. Thus, in order to obtain information on these variables in which schools

differ and in which modifications of some kind could be made (unlike in the case of contextual variables), a binary logistic regression study was carried out to analyse the influence of process variables on the school effectiveness of the schools.

In Spain, nine variables out of the 76 initially incorporated in the procedure were shown to influence school effectiveness. In Ireland, the number of variables is reduced to three. In Portugal, these factors are further reduced, with only two variables making up the model. In all cases, the models show a more than acceptable fit according to *Nagelkerke's R²* statistic, with more than 55% explanatory power. Similarly, the predictive accuracy of the models is very good, with an overall accuracy rate of between 83.6% (Portugal) and 88.9% (Spain).

Three variables have coincided in at least two models: metacognition understood as students' ability to summarise (Spain and Portugal), self-concept of reading as perceived competence (Spain and Ireland) and the use of ICT outside school for homework (Spain and Ireland).

6.2.1. Student's Metacognition - summarising

Research on school effectiveness has indicated that not only cognitive competences, but also metacognitive competences (Creemers, 1994) or the socio-affective development of individuals (Creemers & Kyriakides, 2008), among others, as part of their comprehensive education (as indicated in figure 1.2), are expected results of the educational process.

For this reason, it is not surprising that there is a relationship between metacognition, understood as students' ability to summarise, and school effectiveness. In both Portugal and Spain, this is the variable with the strongest positive relationship with the likelihood of a school being characterised as highly effective. In absolute terms, the interpretation that can be made of these results is that students who are more able to summarise a text adequately are more likely to belong to high-performing schools. However, this goes further, since this variable is also measuring the way in which students apply previously acquired cognitive processes to a new task to tackle it successfully.

The analysis presented here demonstrates how this is a factor of great importance for school effectiveness that could be promoted and worked on by schools themselves, aiming to increase their effectiveness without the need for large capital investments or structural changes in education systems.

6.2.2. Self-concept of reading: Perception of competence

Since the 1970s Brookover et al. (1979) expanded the understanding of school effectiveness by including new factors such as “el autoconcepto de la capacidad académica o la autoconfianza de los alumnos” (Murillo, 2003, p. 6). Despite this, there is not much research that has taken such factors into account in the study of school effectiveness.

The results of the analysis indicate that students' self-concept about reading positively affects school effectiveness in Spain and Ireland, although to different extents, with a much greater influence in the English-speaking country. A student's self-concept about himself or herself is a fundamental part of his or her overall development, which is why it should be one of the objectives to be achieved by schools, that their students consider themselves capable of facing the different challenges that society and, in short, life may present them with.

Commonly, variables of this type have been considered more as outcome variables of education than as process variables (Inglés et al., 2009), limiting the interpretations that can be derived from their study to objectives to be achieved with the teaching-learning process, rather than malleable procedures to foster student success.

6.2.3. Use of ICT outside of school for schoolwork activities

There is currently a wide-ranging debate on the use of ICT inside and outside the classroom by students at different educational stages, largely linked to the capacity and potential of these devices, particularly mobile devices. Initially, the inclusion of technology in education was seen as an opportunity for progress, although it has subsequently come to be seen as an instrument that fosters educational inequality among students, among other aspects, due to the difficulty of access to them by the most disadvantaged sectors (Alarcó Estévez, 2022).

The variable with an influence on school effectiveness in Ireland and Spain in terms of ICT use focuses on the use of technology to carry out school tasks outside the educational institution itself. Contrary to what might be expected, although in line with other research (Gamazo, 2019), the relationship between the two variables is inverse. In other words, the greater the use of ICT in this way, the greater the likelihood that the

school is considered to have low effectiveness. This conclusion has been extrapolated to the use of ICT in the classroom, considering that their use, in this case the use of mobile devices, has negative consequences on academic performance. The effect of this extrapolation is being, in different European regions such as Spain, the prohibition of the use of mobile devices in schools, which means that the possibilities and limitations that this has on student learning must be rethought.

This fact is closely related to another of the current educational debates, that of homework outside school, and therefore, to understand in depth, the effects of ICT use on school effectiveness, it would be interesting to study these two factors together, together with many others that could be directly intervened in school practices.

6.3. Limitations

The conduct of this research has highlighted the existence of numerous limitations linked to the study of factors related to academic performance and school effectiveness. Firstly, the limitations derived from the use of large-scale assessment tests in general and PISA in particular are addressed. Then, we propose those limitations found during the development of this doctoral thesis linked both to the methodology used and to the difficulties encountered in its development.

The literature has called for the need for longitudinal studies to understand existing educational trends in order to guide educational decision-making (Bokhove et al., 2019; Li, 2016; Wu, Gao, et al., 2020). Clearly, in this field, conducting studies such as these poses a particular difficulty associated with the identification of students and other hard-to-control variables such as dropout. A large-scale assessment test that aims to alleviate this limitation as far as possible is TIMSS, since it proposes a cohort study of two groups of students in which, theoretically, the group with the lowest educational level analysed in one edition (4th year of Primary Education in the case of Spain) should correspond to the group with the highest educational level in the immediately following edition (2nd year of Compulsory Secondary Education in Spain). The way in which this limitation linked to the use of large-scale assessment tests could then be alleviated would be to conduct trend studies in which, despite the fact that populations are not stable over time, we try to describe educational change by measuring or comparing measures at two or more different points in time (Bisquerra, 2004; Cohen et al., 2017; Sullivan & Calderwood, 2017). In any case, as mentioned above, the results derived from conducting

research of this type cannot be taken as causal statements but rather as explanatory approximations of the phenomena studied.

Another inherent criticism and limitation of large-scale assessment tests is cultural bias, i.e. that their conceptualisation and implementation do not fully take into account the differential characteristics and idiosyncrasies of the participating countries (Lenkeit, 2012). Partly because of this criticism, experts from the different countries that intend to participate in these assessment tests actively collaborate in the conceptualisation of the different assessment instruments, trying to minimise the effect that this may have on the educational outcomes that are being measured through the rigorous standardisation process that ILSAs follow (Ramos Zincke, 2021).

To these two limitations can be added the need for data at the classroom level to verify and understand the good practices developed by teachers that influence the academic performance of students and, at a macro level, the effectiveness of schools (Burns et al., 2020; He & Fischer, 2020; Wu, Gao, et al., 2020). Although an approach to constructing three large-scale assessment tests is presented, the stated limitation refers specifically to the PISA tests, since in TIMSS, for example, the information necessary to properly group students into classrooms is available. As mentioned, not having data at the classroom level makes it impossible to study variables that could be of some importance for school effectiveness, variables such as teaching methodologies, the conditions on which student grouping is based, etc.

We must take a certain perspective on the results obtained through secondary analysis of ILSAs, not only when analysing the results of the competency tests themselves, but also when considering those variables obtained through self-reporting by students, teachers, and management teams. This extraction of data through self-reports may result in a loss of objectivity (Danhier, 2018; Wu, Gao, et al., 2020; Wu, Shen, et al., 2020) and in another limitation linked to social desirability when answering these questionnaires (Lopes et al., 2022). These two limitations are difficult to solve, as the loss of objectivity that could be alleviated by obtaining information from external observers, for example, would infinitely limit the quality and, above all, the quantity of information that could be obtained.

A limitation could be considered to be that raised by some authors who indicate how ILSAs encourage competition between countries and between schools themselves (Murillo, 2004) by ranking them on the basis of the performance of their students. Basically, it is understood that tests such as PISA are tools for the commercialisation of

education that have more negative than positive consequences. In the case of Spain, the Programme for International Student Assessment has been used as an unequivocal basis on which to build educational legislation seeking to improve the quality of the education system, resulting in the Ley Orgánica 8/2013, de 9 de diciembre, para la mejora de la calidad educativa (2013), where in its Preamble it was stated that the system does not allow progress towards improving the quality of education as shown by the results of Spanish students in the PISA test. Similarly, it proposed external assessment of students at the end of the school year as a guarantee of direct improvement in quality, given that in some twenty OECD countries participating in PISA, measures of this type are being developed that have a positive effect on the performance of their students. This is a clear example of the commercialisation of education, the trivialisation of students' results in PISA and the decontextualisation of measures aimed at improving quality in the name of PISA.

School effectiveness is influenced by several factors that are not directly related to the contextual characteristics of the academic environment. These factors may be related to aspects such as the socio-economic characteristics of the student body (which has already been confirmed as a factor influencing academic performance), the family environment, motivational or cultural conditions of students, and so on. This non-contextual information is obtained in PISA through self-reporting by students, schools, or teachers, and may therefore be biased and limiting in that it does not represent reliably what it is intended to measure. Nevertheless, we have studied the process variables that may be affecting the effectiveness of schools to provide a more comprehensive approach that takes these characteristics into account.

Finally, it is worth mentioning the major limitations encountered in this doctoral thesis: the number of missing values for certain variables, the impossibility of applying mixed methods and the lack of collaboration on the part of the schools.

In relation to the high percentage of missing values detected in some of the variables incorporated into the multilevel analysis, the case of the index of teachers with doctoral studies in Spain is presented. The existence of 20% missing values in this variable means that, although it has been shown to have a positive influence on performance in mathematics and science, this relationship cannot be confirmed. Above all, considering that this same variable appears to be significant in the case of scientific competence in Portugal, showing an opposite influence.

At the beginning of this doctoral thesis, the objective was to establish a phase of qualitative data collection in specific schools selected from the quantitative analyses carried out in order to complement them and provide a broader view of the phenomenon studied, in accordance with the suggestion of the use of mixed methods made by various authors (Burns et al., 2020; Ertem, 2021; Wu, Shen, et al., 2020).

The pandemic caused by the Covid-19 virus during the 2019-2020 academic year prevented the collaboration of secondary schools due to restrictions on access to them. For this reason, it was not possible to apply the questionnaire designed to find out the self-perceived school effectiveness of active secondary education teachers, nor were we able to obtain information on the educational practices developed in high- and low-residue schools. Consequently, we were unable to achieve the specific objective related to the development of a catalogue of good practices aimed at improving school effectiveness.

Nevertheless, taking as a reference the variables that have been shown to influence school effectiveness, some suggestions for improvement are made in the three countries studied:

- In Spain: to foster students' ability to synthesis and evaluate the credibility of the information they receive, to encourage students to improve their reading skills by adapting their reading to their individual needs, to provide tools and training to teachers to help them improve their classroom management and to relate appropriately to students and, ultimately, to improve the classroom climate by fostering a good teacher-student relationship.
- In Ireland: improving students' digital competence by using ICT more effectively, since the use of ICT outside school has a negative influence on school effectiveness, but students' perceived autonomy in its use has a positive effect on this variable. Similarly, it would be advisable to encourage reading by adapting it to students' needs, seeking to awaken in them an interest that helps them to feel more competent in this aspect.
- In Portugal: improving students' ability to synthesis and maintain or foster positive teacher-student relationships in language classes as a guarantee of increased school effectiveness.

Ultimately, it should be clear which uses can and should be made of large-scale assessment tests and which should not. ILSAs are intended to monitor the state of education systems, but if important decisions are to be made based on them, they should

not be regarded as absolute truths, but only as tools to assist in obtaining information and formulating deeper research objectives to guide such decision-making, for example, in the development of education legislation.

6.4. Future research lines

Throughout the presentation, some possible future lines of research have been hinted at, although they are specifically specified in this section.

This research sought to gain a comprehensive perspective on the factors associated with academic performance, opting for the development of multilevel models with random intercept and fixed slopes. This view has provided information on the average effects that the different variables could have on performance. To further study the effects of these variables on the performance of a specific group of schools, such as those with high or low effectiveness, a completely randomised design could be proposed in which the magnitude of the estimated variation for each variable in each specific school is specifically identified (Tapia Blásquez, 2016), which could provide valuable information for educational research.

Another possible future line of research is to carry out studies of educational trends that could be developed in order to find out how the influence of different factors has changed over time and the inclusion of new factors due to political, social or cultural changes. An example of these changes can be seen by comparing the analysis of the PISA 2015 data (Gamazo, 2019), where migration status was relevant, in contrast to this study constructed on the PISA 2018 data where this variable does not show any importance. It is interesting to compare whether there have been structural changes in terms of the characteristics of the population group we are referring to, since if so, it would be necessary to compare the relationship between these changes in terms of the idiosyncrasies of the group itself and their impact on students' educational outcomes.

In relation to the above, the possibility of applying mixed methods, already indicated as a limitation of this doctoral thesis, is derived. The mixed methods approach would imply the contribution of qualitative information that could not be considered in this study. The possibility of applying, for example, the questionnaire elaborated as part of this research, would provide that real vision on the educational practices developed in effective and ineffective schools, obtaining, therefore, other information that would allow us to understand the causes of this difference in terms of effectiveness.

The different influence of the same variable on the different competences has been verified. Evidently, the three main competencies studied in PISA are closely related, so that their individualised study sometimes involves the appearance of certain incongruities or contradictions that provide evidence on the aspects to be worked on. This is the case, for example, of the gender variable, which indicates the differences in performance in favour of males in mathematics and science and in favour of females in reading comprehension. One of the current initiatives that seek to alleviate the differences in performance based on the gender of students has been described above, but, despite this type of proposal, it has been observed that instead of increasing the performance of female students in STEM subjects or that of male students in reading comprehension, these differences have been reducing at the same time as the performance itself. In other words, the scores of the two groups have moved closer together, but towards balance rather than towards excellence, with the performance of the higher-achieving group decreasing and that of the lower-achieving group increasing.

Finally, it could be affirmed that the doctoral thesis developed has adequately achieved the objectives set. It has achieved an approach to the good practices carried out in high and low effectiveness schools, although highlighting the need to study them in greater depth with the combination of obtaining more qualitative data that complement this research and help to understand to a greater extent how school improvement can be achieved.

6.5. Dissemination and publications

The development of this research work has resulted in publications of different types to disseminate the findings and some of the results of the research. These publications have been:

Frade-Martínez, C., Gamazo, A. & Olmos-Migueláñez, S. (2024, in press). Evolución de los predictores del nivel competencial de los estudiantes españoles: un estudio comparativo entre PISA 2015 y 2018. *Revista de Investigación Educativa*.

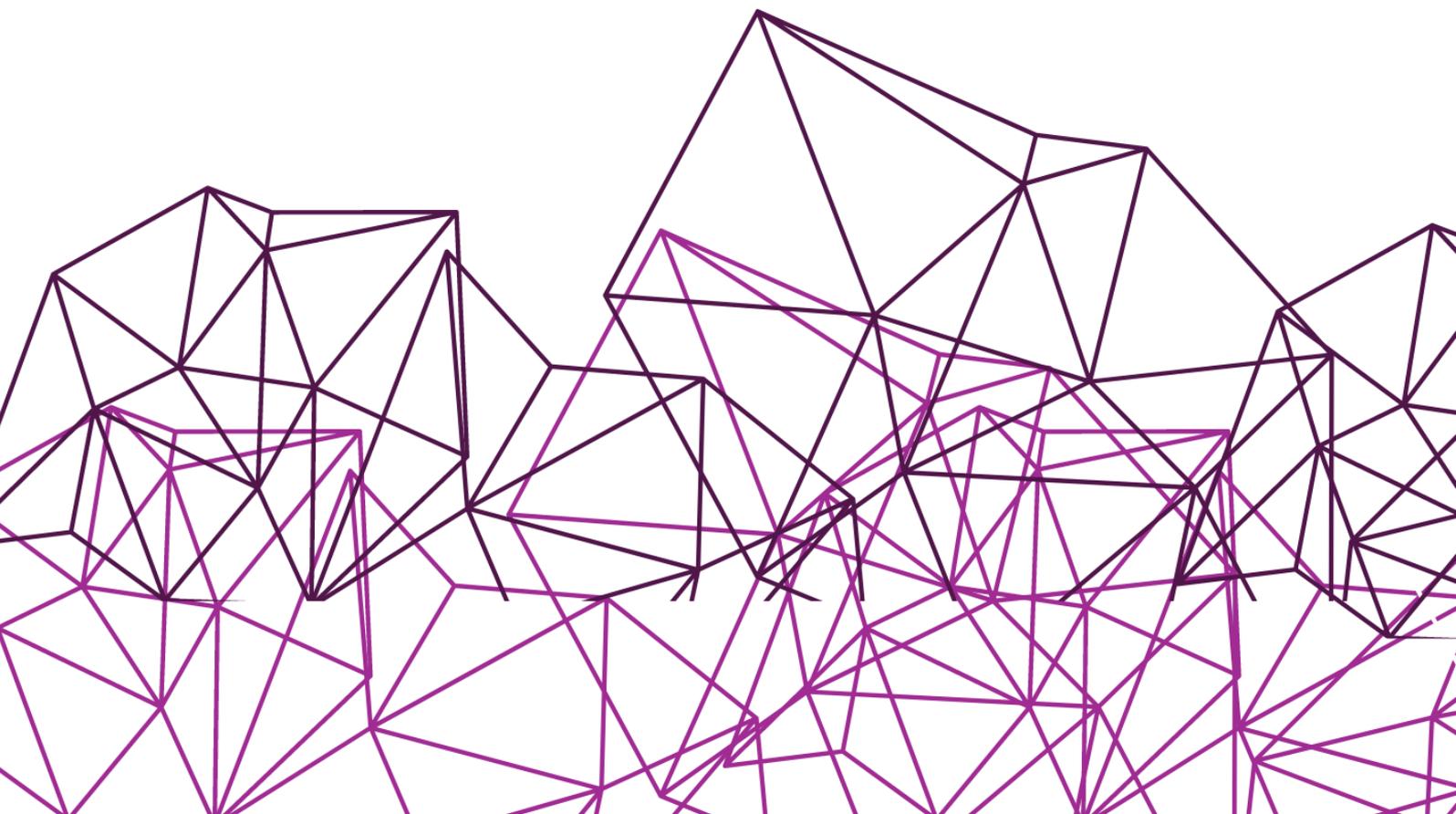
Frade-Martínez, C., Gamazo, A. & Olmos-Migueláñez, S. (2024, in press). Pruebas de evaluación a gran escala y eficacia escolar: una revisión sistemática de la literatura. *Education in the Knowledge Society*.

Frade-Martínez, C., O'Hara, J., Gamazo, A., Olmos-Migueláñez, S. & Brown, M. (2024). A multilevel investigation of factors related to achievement in Ireland and Spain

using PISA data. *Frontiers in Education*, 9.
<https://doi.org/10.3389/feduc.2024.1306197>

- Frade-Martínez, C., Olmos-Migueláñez, S. & Gamazo, A. (2020). *School effectiveness and PISA tests. Factors of school success*. Communication presented at VIII International Conference on Technological Ecosystems for Enhancing Multiculturality (TEEM 2020), held in Salamanca.
- Frade-Martínez, C. & Olmos-Migueláñez, S. (2021). *Detección de escuelas eficaces para la mejora del sistema*. Communication presented at II Congreso Internacional de Innovación y Tendencias Educativas (INNTEd 2021), held online.
- Frade-Martínez, C., Olmos-Migueláñez, S. & Gamazo, A. (2021). *Factors associated with the school performance of Spanish students. A study based on PISA 2018 data*. Communication presented at IX International Conference on Technological Ecosystems for Enhancing Multiculturality (TEEM 2021), held in Barcelona.
- Frade-Martínez, C., O'Hara, J., Olmos-Migueláñez, S., Gamazo, A. & Brown, M. (2022). *Comparing school performance in Ireland and Spain: an analysis of factors emerging from PISA 2018*. Communication presented at ESAI Conference 2022 held in Dublín.
- Frade-Martínez, C., Gamazo, A. & Olmos-Migueláñez, S. (2022). *¿Escuelas de alta y baja eficacia? Caracterización y clasificación de las escuelas eficaces en Castilla y León y Andalucía*. Communication presented at XX Congreso Internacional de Investigación Educativa, held in Santiago de Compostela.
- Frade-Martínez, C., Olmos-Migueláñez, S., Gamazo, A. & O'Hara, J. (2022). *Researching school effectiveness: a comparative study of Spain, Portugal and Ireland from PISA 2018*. Communication presented at ECER 2022 – The European Conference on Educational Research, held online.
- Frade-Martínez, C., Gamazo, A. & Olmos-Migueláñez, S. (2023). *Factors influencing school effectiveness in Spain. A secondary analysis of PISA 2018*. Communication presented at XI International Conference on Technological Ecosystems for Enhancing Multiculturality (TEEM 2023), held in Barcelona.

REFERENCIAS



Referencias

- Agasisti, T., Avvisati, F., Borgonovi, F., & Longobardi, S. (2021). What School Factors are Associated with the Success of Socio-Economically Disadvantaged Students? An Empirical Investigation Using PISA Data. *Social Indicators Research*, 157(2), 749-781. <https://doi.org/10.1007/s11205-021-02668-w>
- 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. <https://doi.org/10.2307/2981882>
- Alarcó Estévez, G. (2022). Mediación de las TIC en el contexto de una educación inclusiva. *Serie Científica de la Universidad de las Ciencias Informáticas*, 15(7), 31-54.
- Ayala Franco, E., López Martínez, R. E., & Menéndez Domínguez, V. H. (2021). Modelos predictivos de riesgo académico en carreras de computación con minería de datos educativos. *RED: revista de educación a distancia*, 21(66), 1.
- Azorín Abellán, C. M. (2019). Las transiciones educativas y su influencia en el alumnado. *Edetania: estudios y propuestas socio-educativas*, 55, 223-248. https://doi.org/10.46583/edetania_2019.55.444
- Bellod Redondo, J. F. (2016). PIGS: Austeridad fiscal, reformas estructurales y crecimiento potencial. *Revista de economía mundial*, 43, 161-177. <https://doi.org/10.33776/rem.v0i43.3855>
- Bisquerra, R. (2004). *Metodología de la investigación educativa*. La Muralla.
- Blanco Blanco, A., 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. <https://doi.org/10.4438/1988-592X-RE-2014-365->

- Bokhove, C., Miyazaki, M., Komatsu, K., Chino, K., Leung, A., & Mok, I. A. C. (2019). The Role of “Opportunity to Learn” in the Geometry Curriculum: A Multilevel Comparison of Six Countries. *Frontiers in Education, 4*.
<https://doi.org/10.3389/feduc.2019.00063>
- Brookover, W. B., Beady, C., Flood, P., Schweitzer, J., & Wisenbaker, J. (1979). *School Social Systems and Student Achievement: Schools Can Make a Difference*. Praeger.
- Brookover, W. B., & Lezotte, L. W. (1979). *Changes in School Characteristics Coincident with Changes in Student Achievement: Executive Summary*. Institute for Research on Teaching, Michigan State University.
- Brookover, W. B., & Schneider, J. M. (1975). Academic Environments and Elementary School Achievement. *Journal of Research and Development in Education, 9*(1), 82-91.
- Burin, D. I., Coccimiglio, Y., González, F. M., & Bulla Afanador, J. J. (2016). Habilidades digitales y lectura en entornos digitales. Desarrollos recientes sobre comprensión lectora digital. *Psicología, Conocimiento y Sociedad, 6*(1), 191-206.
- Burns, E. C., Martin, A. J., & Collie, R. J. (2020). Supporting and thwarting interpersonal dynamics and student achievement: A multi-level examination of PISA 2015. *International Journal of Research & Method in Education, 43*(4), 364-378.
<https://doi.org/10.1080/1743727X.2020.1757639>
- Caponera, E., & Losito, B. (2016). Context factors and student achievement in the IEA studies: Evidence from TIMSS. *Large-scale Assessments in Education, 4*(1).
<https://doi.org/10.1186/s40536-016-0030-6>
- Cappella, E., & Weinstein, R. S. (2001). Turning around reading achievement: Predictors

- of high school students' academic resilience. *Journal of Educational Psychology*, 93(4), 758-771. <https://doi.org/10.1037/0022-0663.93.4.758>
- Catena Martínez, A., Trujillo Mendoza, H. M., & Ramos Álvarez, M. M. (2003). *Análisis multivariado: Un manual para investigadores*. Biblioteca Nueva.
- Chitty, C. (1997). The school effectiveness movement: Origins, shortcomings and future possibilities. *Curriculum Journal*, 8(1), 45-62. Scopus. <https://doi.org/10.1080/09585176.1997.11070761>
- Choi de Mendizábal, A. (2016). Explorando los límites de PISA. *Revista de Sociología de la Educación-RASE*, 9(1), 163-165.
- Cohen, L., Manion, L., & Morrison, K. (2017). Surveys, longitudinal, cross-sectional and trend studies. En *Research Methods in Education* (8.^a ed.). Routledge.
- Coleman, J. S. (1966). *Equality of Educational Opportunity*.
- Cordero Ferrera, J. M., Crespo Cebada, E., & Pedraja Chaparro, F. M. (2013). Rendimiento educativo y determinantes según PISA: Una revisión de la literatura en España. *Revista de educación*, 362, 273-297. <https://doi.org/10.4438/1988-592X-RE-2011-362-161>
- Costa, P., & Araújo, L. (2018). Skilled Students and Effective Schools: Reading Achievement in Denmark, Sweden, and France. *Scandinavian Journal of Educational Research*, 62(6), 850-864. <https://doi.org/10.1080/00313831.2017.1307274>
- Creemers, B. P. M. (1994). *The Effective Classroom*. Cassell.
- Creemers, B. P. M., & Kyriakides, L. (2008). *The Dynamics of Educational Effectiveness: A contribution to policy, practice and theory in contemporary schools*. Routledge; Scopus. <https://doi.org/10.4324/9780203939185>
- Cvetkovic-Vega, A., Maguiña, J. L., Soto, A., Lama-Valdivia, J., & López, L. E. C.

- (2021). Estudios transversales. *Revista de la Facultad de Medicina Humana*, 21(1), 179-185. <https://doi.org/10.25176/rfmh.v21i1.3069>
- Danhier, J. (2018). How Big is the Handicap for Disadvantaged Pupils in Segregated Schooling Systems? *British Journal of Educational Studies*, 66(3), 341-364. <https://doi.org/10.1080/00071005.2017.1322682>
- Darmawan, I., & Keeves, J. (2006). *Accountability of teachers and schools: A value-added approach*.
- Darmody, M., & Smyth, E. (2018). Immigrant Student Achievement and Educational Policy in Ireland. En L. Volante, D. Klinger, & O. Bilgili (Eds.), *Immigrant Student Achievement and Education Policy: Cross-Cultural Approaches* (pp. 119-135). https://doi.org/10.1007/978-3-319-74063-8_8
- Ding, H., & Homer, M. (2020). Interpreting mathematics performance in PISA: Taking account of reading performance. *International Journal of Educational Research*, 102, 101566. <https://doi.org/10.1016/j.ijer.2020.101566>
- Doncel Abad, D., & Cabrera Álvarez, P. (2020). Comunidades Autónomas bilingües, identidades y desempeño educativo según PISA 2015. *Revista de educación*, 387, 163-188. <https://doi.org/10.4438/1988-592X-RE-2020-387-443>
- Draper, D. (1995). Assessment and Propagation of Model Uncertainty. *Journal of the Royal Statistical Society*, 57(1), 45-97. <https://doi.org/10.2307/2346087>
- Edel Navarro, R. (2003). El rendimiento académico: Concepto, investigación y desarrollo. *REICE: Revista Iberoamericana sobre Calidad, Eficacia y Cambio en Educación*, 1(2), 9.
- Edmonds, R. (1979). Effective Schools for the Urban Poor. *Educational Leadership*, 37(1), 15-24.
- Entorf, H., & Minoiu, N. (2005). What a Difference Immigration Policy Makes: A

- Comparison of PISA Scores in Europe and Traditional Countries of Immigration. *German Economic Review*, 6(3), 355-376. <https://doi.org/10.1111/j.1468-0475.2005.00137.x>
- Ertem, H. Y. (2021). Examination of Turkey's PISA 2018 reading literacy scores within student-level and school-level variables. *Participatory Educational Research*, 1, 248-264. <https://doi.org/10.17275/per.21.14.8.1>
- Fenstermacher, G., & Soltis, J. (1998). *Enfoques de la enseñanza*. Amorrortu.
- Fernández García, T., & Andrés Cabello, S. (2019). Crisis y educación en España: Sobre la equidad y la igualdad de oportunidades. *Sistema: revista de ciencias sociales*, 254, 3-22.
- Flores Macías, R. del C., & Gómez Bastida, J. (2010). Un estudio sobre la motivación hacia la escuela secundaria en estudiantes mexicanos. *REDIE: Revista Electrónica de Investigación Educativa*, 12(1), 5.
- Gamazo, A. (2019). *Factores asociados al rendimiento y a la eficacia escolar: Un estudio basado en métodos mixtos a partir de PISA 2015*. <https://doi.org/10.14201/gredos.140406>
- Gamazo, A., & Martínez-Abad, F. (2020). An Exploration of Factors Linked to Academic Performance in PISA 2018 Through Data Mining Techniques. *Frontiers in Psychology*, 11, 1-17. <https://doi.org/10.3389/fpsyg.2020.575167>
- Gamazo, A., Martínez-Abad, F., Olmos-Migueláñez, S., & Rodríguez-Conde, M. J. (2018). Evaluación de factores relacionados con la eficacia escolar en PISA 2015. Un análisis multinivel. *Revista de educación*, 379, 56-84. <https://doi.org/10.4438/1988-592X-RE-2017-379-369>
- García-Herrero, M., Rodríguez-Conde, M.-J., & Martínez-Abad, F. (2024). Factores de calidad docente asociados a la equidad educativa: Formación del profesorado y

- estrategias docentes. *Revista Electrónica Interuniversitaria de Formación del Profesorado*, 27(1), Article 1. <https://doi.org/10.6018/reifop.595181>
- Gaviria Soto, J. L., & Castro Morera, M. (2004). A Multilevel Study on Factors of School Effectiveness in Developing Countries: The Case of Brazilian Resources. *Education Policy Analysis Archives*, 12, 20-20. <https://doi.org/10.14507/epaa.v12n20.2004>
- Gaviria Soto, J. L., & Castro Morera, M. (2005). *Modelos jerárquicos lineales*. La Muralla.
- Gil-Flores, J. G., & García-Gómez, S. (2017). Importancia de la actuación docente frente a la política educativa regional en la explicación del rendimiento en PISA. *Revista de educación*, 378, 52-77. <https://doi.org/10.4438/1988-592X-RE-2017-378-361>
- Girls Who Code. (2024). *Girls Who Code | Home*. <https://girlswhocode.com/>
- Goldstein, H. (1987). *Multilevel Models in Educational and Social Research*. Charles Griffin and Company Limited.
- Gómez, G. (2013). Los efectos de la repitencia en tanto que política pública en cuatro países del Cono Sur: Argentina, Brasil, Chile y Uruguay: un análisis en base a PISA 2009. *Revista Latinoamericana de Educación Comparada: RELEC*, 4(4), 59-70.
- Gómez-Fernández, N., & Mediavilla, M. (2021). Exploring the relationship between Information and Communication Technologies (ICT) and academic performance: A multilevel analysis for Spain. *Socio-Economic Planning Sciences*, 77, 101009. <https://doi.org/10.1016/j.seps.2021.101009>
- Gortázar, L. (2019). ¿Favorece el sistema educativo español la igualdad de oportunidades? *Información Comercial Española, ICE: Revista de economía*, 910, 15-29. <https://doi.org/10.32796/ice.2019.910.6917>

- Grilli, L., Pennoni, F., Rampichini, C., & Romeo, I. (2016). Exploiting TIMSS and PIRLS Combined Data: Multivariate Multilevel Modelling of Student Achievement. *The Annals of Applied Statistics*, 10(4), 2405-2426. <https://doi.org/10.1214/16-AOAS988>
- Guevara Molina, J. (2022). *Efecto del cambio de establecimiento en el rendimiento académico de los estudiantes*. Universidad de Chile.
- Guo, S., Li, L., & Zhang, D. (2018). A multilevel analysis of the effects of disciplinary climate strength on student reading performance. *Asia Pacific Education Review*, 19, 1-15. <https://doi.org/10.1007/s12564-018-9516-y>
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2014). *Multivariate data analysis* (7^a). Pearson Education Limited.
- Hayes, A. F. (2006). A Primer on Multilevel Modeling. *Human Communication Research*, 32(4), 385-410. <https://doi.org/10.1111/j.1468-2958.2006.00281.x>
- He, J., & Fischer, J. (2020). Differential associations of school practices with achievement and sense of belonging of immigrant and non-immigrant students. *Journal of Applied Developmental Psychology*, 66, 101089. <https://doi.org/10.1016/j.appdev.2019.101089>
- Heyneman, S. P., & Lee, B. (2014). The Impact of International Studies of Academic Achievement on Policy and Research. En L. Rutkowski, M. von Davier, & D. Rutkowski, *Handbook of International Large-Scale Assessment*. Chapman and Hall/CRC.
- Hox, J. J., & Kreft, I. G. G. (1994). Multilevel Analysis Methods. *Sociological Methods & Research*, 22(3), 283-299. <https://doi.org/10.1177/0049124194022003001>
- Hox, J. J., Moerbeek, M., & van de Schoot, R. (2010). *Multilevel Analysis: Techniques and Applications, Second Edition* (2.^a ed.). Routledge.

<https://doi.org/10.4324/9780203852279>

Huang, H., & Sebastian, J. (2015). The role of schools in bridging within-school achievement gaps based on socioeconomic status: A cross-national comparative study. *Compare: A Journal of Comparative and International Education*, 45(4), 501-525. <https://doi.org/10.1080/03057925.2014.905103>

IEA. (2023). *TIMSS*. TIMSS - Trends in International Mathematics and Science Study. <https://www.iea.nl/studies/iea/timss>

IEA. (2024). *PIRLS*. PIRLS - Progress in International Reading Literacy Study. <https://www.iea.nl/studies/iea/pirls>

INEE. (2022). Propuestas para garantizar la igualdad de oportunidades en educación | Blog de INEE [INEE - Instituto Nacional de Evaluación Educativa]. *INEE*. <http://blog.intef.es/inee/2022/07/20/pif117/>

Inglés, C. J., Pastor, Y., Torregrosa, M. S., Redondo, J., & García, J. M. (2009). Diferencias en función del género y el curso académico en dimensiones del autoconcepto: Estudio con una muestra de adolescentes españoles. *Anuario de psicología*, 40(2), 271-288.

Instituto Nacional de Evaluación Educativa [INEE], Ministerio de Educación y Formación Profesional. (2023). *TIMSS - Trends in International Mathematics and Science Study*. Instituto Nacional de Evaluación Educativa. <https://www.educacionyfp.gob.es/inee/evaluaciones-internacionales/timss.html>

International Congress for School Effectiveness and Improvement [ICSEI]. (2024). *The ICSEI Story*. <https://www.icsei.net/about-icsei-membership/>

Jacob, B. A., & Lefgren, L. (2009). The Effect of Grade Retention on High School Completion. *American Economic Journal: Applied Economics*, 1(3), 33-58. <https://doi.org/10.1257/app.1.3.33>

- Jiménez Hernández, M. (1994). Competencia social: Intervención preventiva en la escuela. *Infancia y Sociedad: Revista de estudios*, 24, 21-48.
- Jiménez Moreno, J. A. (2014). *Evaluación a gran escala: Su desarrollo e inconvenientes de su modelo teórico*. Congreso Internacional de Educación, México.
- Jiménez Moreno, J. A. (2016). El Papel de la Evaluación a Gran Escala como Política de Rendición de Cuentas en el Sistema Educativo Mexicano. *Revista Iberoamericana de Evaluación Educativa*, 9(1).
<https://doi.org/10.15366/riee2016.9.1.007>
- Jimerson, S. R. (2001). Meta-analysis of grade retention research: Implications for practice in the 21st century. *School Psychology Review*, 30(3), 420-437.
- 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 Lineales. *Estudios sobre educación*, 27, 37-61.
<https://doi.org/10.15581/004.27.37-61>
- Jornet Meliá, J. M., & Suárez Rodríguez, J. M. (1996). Pruebas estandarizadas y evaluación del rendimiento: Usos y características métricas. *Revista de investigación educativa, RIE*, 14(2), 141-163.
- Kameshwara, K. K., Sandoval-Hernandez, A., Shields, R., & Dhanda, K. R. (2020). A false promise? Decentralization in education systems across the globe. *International Journal of Educational Research*, 104, 101669.
<https://doi.org/10.1016/j.ijer.2020.101669>
- Keeves, J. P., Hungi, N., & Afrassa, T. (2005). Measuring value added effects across schools: Should schools be compared in performance? *Studies in Educational Evaluation*, 31(2), 247-266. <https://doi.org/10.1016/j.stueduc.2005.05.012>

- Kitchenham, B. (2004). *Procedures for Performing Systematic Reviews*. Keele University.
- Kitchenham, B., Budgen, D., & Brereton, P. (2011). Using mapping studies as the basis for further research—A participant-observer case study. *Information & Software Technology, 53*, 638-651. <https://doi.org/10.1016/j.infsof.2010.12.011>
- Kleinbaum, D. G., & Klein, M. (2010). *Logistic Regression: A Self-Learning Text*. Springer New York. <https://doi.org/10.1007/978-1-4419-1742-3>
- Kline, R. B. (2011). *Principles and Practice of Structural Equation Modeling*. Guilford Publications.
- Krüger, N. S. (2013). Segregación social y desigualdad de logros educativos en Argentina. *Education Policy Analysis Archives, 21*(86), 1-30. <https://doi.org/10.14507/epaa.v21n86.2013>
- Lacave Rodero, C., Molina Díaz, A. I., Fernández Guerrero, M. M., & Redondo Duque, M. Á. (2015). Análisis de la fiabilidad y validez de un cuestionario docente. *Actas del simposio-taller sobre estrategias y herramientas para el aprendizaje y la evaluación: JENUI 2015, 23*.
- Lamas, H. A. (2015). Sobre el rendimiento escolar. *Propósitos y representaciones, 3*(1), 313-386. <http://dx.doi.org/10.20511/pyr2015.v3n1.74>
- Laukaityte, I., & Rolfsman, E. (2020). Low, medium, and high-performing schools in the Nordic countries. Student performance at PISA Mathematics 2003-2012. *Education Inquiry, 11*(3), 276-295. <https://doi.org/10.1080/20004508.2020.1721256>
- Lee, M., Ryoo, J. H., & Walker, A. (2021). School Principals' Time Use for Interaction with Individual Students: Macro Contexts, Organizational Conditions, and Student Outcomes. *American Journal of Education, 127*(2), 303-344.

<https://doi.org/10.1086/712174>

- Lee, V. E. (2000). Using Hierarchical Linear Modeling to Study Social Contexts: The Case of School Effects. *Educational Psychologist*, 35(2), 125-141. https://doi.org/10.1207/S15326985EP3502_6
- Lenkeit, J. (2012). How effective are educational systems? A value-added approach to study trends in PIRLS. *Journal of Educational Research Online*, 4(2), 143-173.
- Levine, D. U., & Lezotte, L. W. (1990). Unusually Effective Schools: A Review and Analysis of Research and Practice. *School Effectiveness and School Improvement*, 1(3), 221-224. <https://doi.org/10.1080/0924345900010305>
- Ley Orgánica 8/2013, de 9 de diciembre, para la mejora de la calidad educativa, Jefatura del Estado, 295 Boletín Oficial del Estado 97858 (2013). <https://www.boe.es/eli/es/lo/2013/12/09/8>
- Lezotte, L. W., Edmonds, R., & Ratner, G. (1974). *A final report: Remedy for school failure to equitable basic school skills*. Michigan State University, Department of Urban and Metropolitan Studies.
- Li, H. (2016). How is formative assessment related to students' reading achievement? Findings from PISA 2009. *Assessment in Education: Principles, Policy & Practice*, 23(4), 473-494. <https://doi.org/10.1080/0969594X.2016.1139543>
- Liou, P.-Y., & Hung, Y.-C. (2015). Statistical techniques utilized in analyzing PISA and TIMSS data in science education from 1996 to 2013: A methodological review. *International Journal of Science and Mathematics Education*, 13(6), 1449-1468. <https://doi.org/10.1007/s10763-014-9558-5>
- Liu, H., Damme, J., Gielen, S., & Van den Noortgate, W. (2014). School processes mediate school compositional effects: Model specification and estimation. *British Educational Research Journal*, 41. <https://doi.org/10.1002/berj.3147>

- Lizasoain, L., & Angulo, A. (2014). Buenas prácticas de escuelas eficaces del País Vasco. Metodología y primeros resultados. *Participación educativa. Revista del Consejo Escolar del Estado*, 3(4), 17-28. <https://doi.org/10.4438/1886-5097-PE>
- Lizasoain, L., Angulo, A., Azpillaga, V., Bartau, I., Danborenea, M. D., del Frago, R., Etxeberria, F., Intxausti, N., Joaristi, L., Méndez Usillos, Y., Núñez Fernández, C., & Valadez, C. (2015). *La eficacia escolar en los centros del País Vasco* (p. 89). Universidad del País Vasco (UPV-EHU), Instituto de Evaluación e Investigación Educativa (ISEI-IVEI). <https://redined.educacion.gob.es/xmlui/handle/11162/203206>
- Lizasoain, L., Joaristi, L. M., Lukas, J. F., & Santiago, K. (2007). El efecto contextual del nivel socioeconómico sobre el rendimiento académico en la educación secundaria obligatoria en la Comunidad Autónoma Vasca (España). Estudio diferencial del nivel socioeconómico familiar y el del centro escolar. *Archivos Analíticos de Políticas Educativas*, 15(8).
- Lopes, J., Oliveira, C., & Costa, P. (2022). Determinantes escolares y de los estudiantes en el rendimiento lector: Un análisis multinivel con estudiantes portugueses. *Revista de Psicodidáctica*, 27(1), 29-37. <https://doi.org/10.1016/j.psicod.2021.05.001>
- Lorenzo, F. (2010). Enfoque orientativo en la transición de primaria a secundaria. *Autodidacta. Revista de la Educación en Extremadura*, 9, 121-127.
- Maas, C. J. M., & Hox, J. J. (2005). Sufficient Sample Sizes for Multilevel Modeling. *Methodology*, 1(3), 86-92. <https://doi.org/10.1027/1614-2241.1.3.86>
- Manacorda, M. (2012). The Cost of Grade Retention. *The Review of Economics and Statistics*, 94(2), 596-606. https://doi.org/10.1162/rest_a_00165
- Marteletto, L., & Andrade, F. (2014). The Educational Achievement of Brazilian

- Adolescents: Cultural Capital and the Interaction between Families and Schools. *Sociology of Education*, 87(1), 16-35.
<https://doi.org/10.1177/0038040713494223>
- Martín Carrasquilla, O., Santaolalla Pascual, E., & Muñoz San Roque, I. (2022). La brecha de género en la Educación STEM. *Revista de Educación*, 396.
<https://doi.org/10.4438/1988-592X-RE-2022-396-533>
- Martínez Rizo, F. (2009). Evaluación formativa en aula y evaluación a gran escala: Hacia un sistema más equilibrado. *Revista electrónica de investigación educativa*, 11(2), 1-18.
- Martínez-Abad, F., Gamazo, A., & Rodríguez-Conde, M.-J. (2020). Educational Data Mining: Identification of factors associated with school effectiveness in PISA assessment. *Studies in Educational Evaluation*, 66, 100875.
<https://doi.org/10.1016/j.stueduc.2020.100875>
- Martínez-Abad, F., & León, J. (2023). Inferencia causal en investigación educativa: Análisis de la causalidad en estudios observacionales de carácter transversal. *RELIEVE. Revista Electrónica de Investigación y Evaluación Educativa*, 29(2), 1-24. <https://doi.org/10.30827/relieve.v29i2.26843>
- Martins, L., & Veiga, P. (2010). Do inequalities in parents' education play an important role in PISA students' mathematics achievement test score disparities? *Economics of Education Review*, 29(6), 1016-1033.
<https://doi.org/10.1016/j.econedurev.2010.05.001>
- Mateos Claros, F., Olmedo Ruiz, F. J., Esteban Ibáñez, M., & Amador Muñoz, L. V. (2019). Lengua materna, cultura y rendimiento en un contexto multicultural de Educación Infantil. *Ocnos: Revista de Estudios sobre Lectura*, 18(2), 44-54.
https://doi.org/10.18239/ocnos_2019.18.2.1966

- Ministerio de Educación y Formación Profesional. (2019). *PISA 2018. Programa para la Evaluación Internacional de los Estudiantes. Informe español (PISA)*. INEE. <https://sede.educacion.gob.es/publiventa/pisa-2018-programa-para-la-evaluacion-internacional-de-los-estudiantes-informe-espanol/evaluacion-examenes/23505>
- Ministerio de Educación y Formación Profesional. (2023). *PISA*. PISA. <https://www.educacionyfp.gob.es/inee/evaluaciones-internacionales/pisa.html>
- Ministerio de Educación y Formación Profesional, & Subdirección General de Atención al Ciudadano, Documentación y Publicaciones. (2018). *TIMSS 2019. Marcos de la evaluación*. Instituto Nacional de Evaluación Educativa. https://www.libreria.educacion.gob.es/libro/timss-2019-marcos-de-la-evaluacion_181302/
- Mora-Ruano, J. G., Schurig, M., & Wittmann, E. (2021). Instructional Leadership as a Vehicle for Teacher Collaboration and Student Achievement. What the German PISA 2015 Sample Tells Us. *Frontiers in Education*, 6. <https://doi.org/10.3389/feduc.2021.582773>
- Mortimore, P., Sammons, P., Stoll, L., Lewis, D., & Ecob, R. (1989). A study of effective junior schools. *International Journal of Educational Research*, 13(7), 753-768. [https://doi.org/10.1016/0883-0355\(89\)90026-8](https://doi.org/10.1016/0883-0355(89)90026-8)
- Muñoz Martín, M. (2006). Investigación: «El valor añadido en Educación y la función de producción educativa». *Avances en Supervisión Educativa*, 2, 1-21.
- Murillo, F. J. (2003). El movimiento de investigación de Eficacia Escolar. En F. J. Murillo Torrecilla, *La investigación sobre Eficacia Escolar en Iberoamérica. Revisión internacional del estado del arte* (p. 446). Convenio Andrés Bello. <https://sede.educacion.gob.es/publiventa/d/14324/19/0>

- Murillo, F. J. (2004). La investigación sobre «eficacia escolar» a debate: Análisis de las críticas y aportaciones. *Tendencias pedagógicas*, 9, 111-130.
- Murillo, F. J. (2005). *La investigación sobre eficacia escolar*. Octaedro.
- Murillo, F. J. (2007). *Investigación Iberoamericana sobre Eficacia Escolar* (Educación para la Integración). Convenio Andrés Bello.
[https://repositorio.minedu.gob.pe/bitstream/handle/20.500.12799/284/155.%20Investigación%20Iberoamericana%20sobre%20eficacia%20escolar.pdf?sequence=1&isAllowed=y](https://repositorio.minedu.gob.pe/bitstream/handle/20.500.12799/284/155.%20Investigaci%20Iberoamericana%20sobre%20eficacia%20escolar.pdf?sequence=1&isAllowed=y)
- Murillo, F. J. (2008). Los modelos multinivel como herramienta para la investigación educativa. *Magis: Revista Internacional de Investigación en Educación*, 1, 45-62.
- Murillo, F. J., & Duk, C. (2016). Segregación Escolar e Inclusión. *Revista latinoamericana de educación inclusiva*, 10(2), 11-13.
<https://doi.org/10.4067/S0718-73782016000200001>
- Murillo, F. J., & Hernández-Castilla, R. (2011). Efectos escolares de factores socioafectivos. Un estudio Multinivel para Iberoamérica. *Revista de investigación educativa, RIE*, 29(2), 407-428.
- Murillo, F. J., & Martínez-Garrido, C. (2018). Magnitud de la Segregación escolar por nivel socioeconómico en España y sus Comunidades Autónomas y comparación con los países de la Unión Europea. *Revista de Sociología de la Educación-RASE*, 11(1), 37-58. <https://doi.org/10.7203/RASE.11.1.10129>
- Myrberg, E., Johansson, S., & Rosén, M. (2019). The Relation between Teacher Specialization and Student Reading Achievement. *Scandinavian Journal of Educational Research*, 63(5), 744-758.
<https://doi.org/10.1080/00313831.2018.1434826>
- Ning, B., Van Damme, J., Van Den Noortgate, W., Gielen, S., Bellens, K., Dupriez, V.,

- & Dumay, X. (2016). Regional inequality in reading performance: An exploration in Belgium. *School Effectiveness and School Improvement*, 27(4), 642-668. <https://doi.org/10.1080/09243453.2016.1202842>
- Ning, B., Van Damme, J., Van den Noortgate, W., Yang, X., & Gielen, S. (2015). The influence of classroom disciplinary climate of schools on reading achievement: A cross-country comparative study. *School Effectiveness and School Improvement*, 26, 1-26. <https://doi.org/10.1080/09243453.2015.1025796>
- Nunez Moscoso, J. (2017). Los métodos mixtos en la investigación en educación: Hacia un uso reflexivo. *Cadernos de Pesquisa*, 47(164), 632-649. <http://dx.doi.org/10.1590/198053143763>
- OECD. (2006). *El programa PISA de la OCDE: Qué es y para qué sirve* (p. 31). OECD. <https://www.oecd.org/pisa/39730818.pdf>
- OECD. (2009). *PISA Data Analysis Manual: SPSS, Second Edition*. OECD Publishing. https://www.oecd-ilibrary.org/education/pisa-data-analysis-manual-spss-second-edition_9789264056275-en
- OECD. (2011). *La Medición del Aprendizaje de los Alumnos: Mejores Prácticas para Evaluar el Valor Agregado de las Escuelas*. OECD Publishing. <https://doi.org/10.1787/9789264090170-es>
- OECD. (2014). *PISA 2012 Technical Report* (PISA). OECD Publishing. <https://www.oecd.org/pisa/pisaproducts/PISA-2012-technical-report-final.pdf>
- OECD. (2019a). *PISA 2018 Assessment and Analytical Framework*. OECD Publishing. https://www.oecd-ilibrary.org/education/pisa-2018-assessment-and-analytical-framework_b25efab8-en
- OECD. (2019b). *PISA 2018 Technical Report*. OECD. <https://www.oecd.org/pisa/data/pisa2018technicalreport/>

- OECD. (2023a). *PISA 2022 Assessment and Analytical Framework (PISA)*. OECD Publishing. <https://www.oecd.org/pisa/publications/pisa-2021-assessment-and-analytical-framework.htm>
- OECD. (2023b). *PISA 2022. Programa para la Evaluación Internacional de los Estudiantes. Informe español* (Instituto Nacional de Evaluación Educativa). OECD Publishing. https://www.oecd-ilibrary.org/education/pisa-2022-results-volume-i_53f23881-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. <https://doi.org/10.1177/1474904115627159>
- Ozili, P. K. (2022). *The Acceptable R-Square in Empirical Modelling for Social Science Research* (SSRN Scholarly Paper 4128165). <https://doi.org/10.2139/ssrn.4128165>
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., ... Alonso-Fernández, S. (2021). Declaración PRISMA 2020: Una guía actualizada para la publicación de revisiones sistemáticas. *Revista Española de Cardiología*, 74(9), 790-799. <https://doi.org/10.1016/j.recesp.2021.06.016>
- Pajares, R., Rico, L., & Sanz, Á. (2004). *Aproximación a un modelo de evaluación: El proyecto PISA 2000*. Ministerio de Educación, Cultura y Deporte. https://www.libreria.educacion.gob.es/libro/aproximacion-a-un-modelo-de-evaluacion-el-proyecto-pisa-2000_174544/

- Paterson, L., & Goldstein, H. (1991). New Statistical Methods for Analysing Social Structures: An introduction to multilevel models. *British Educational Research Journal*, 17(4), 387-393. <https://doi.org/10.1080/0141192910170408>
- Pedrosa, I., Suárez-Álvarez, J., & García-Cueto, E. (2013). Evidencias sobre la validez de contenido: Avances teóricos y métodos para su estimación. *Acción Psicológica*, 10(2), 3-18. <https://doi.org/10.5944/ap.10.2.11820>
- Pomianowicz, K. (2021). Educational achievement disparities between second-generation and non-immigrant students: Do school characteristics account for tracking effects? *European Educational Research Journal*, 14749041211039929. <https://doi.org/10.1177/14749041211039929>
- Purkey, S. C., & Smith, M. S. (1983). Effective Schools: A Review. *The Elementary School Journal*, 83(4), 427-452. <https://doi.org/10.1086/461325>
- Ramos Zincke, C. (2021). Las evaluaciones internacionales a gran escala y la regulación global de los sistemas educacionales: Un análisis integrativo. *CTS: Revista iberoamericana de ciencia, tecnología y sociedad*, 16(48), 159-191.
- Raudenbush, S. W., & Bryk, A. S. (1992). Hierarchical linear models: Applications and data analysis methods. En *Hierarchical linear models: Applications and data analysis methods*. Sage.
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods*. SAGE.
- Raudenbush, S. W., & Willms, J. D. (1995). The Estimation of School Effects. *Journal of Educational and Behavioral Statistics*, 20(4), 307-335. <https://doi.org/10.2307/1165304>
- 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.
<https://doi.org/10.1093/esr/jcv063>
- Rodríguez Gómez, G. (1991). *Investigación evaluativa en torno a los factores de eficacia escolar de los centros públicos de E.G.B* [Http://purl.org/dc/dcmitype/Text, UNED. Universidad Nacional de Educación a Distancia].
<https://dialnet.unirioja.es/servlet/tesis?codigo=40974>
- Ruiz de Miguel, C. (2009). Las escuelas eficaces: Un estudio multinivel de factores explicativos del rendimiento escolar en el área de matemáticas. *Revista de educación*, 348, 355-376.
- Ruiz de Miguel, C., & Castro Morera, M. (2006). Un Estudio Multinivel Basado en PISA 2003: Factores de Eficacia Escolar en el área de Matemáticas. *Education Policy Analysis Archives/Archivos Analíticos de Políticas Educativas*, 14, 1-26.
- Rutter, M., Maughan, B., Mortimore, P., & Ouston, J. (1979). *Fifteen thousand hours: Secondary schools and their effects on children*. Harvard University Press.
- Sabariego Puig, M. (2004). La investigación educativa: Génesis, evolución y características. En R. Bisquerra, *Metodología de la investigación educativa* (pp. 52-87). La Muralla. <https://dialnet.unirioja.es/servlet/articulo?codigo=1090802>
- Sammons, P., Hillman, J., & Mortimore, P. (1995). *Key Characteristics of Effective Schools: A Review of School Effectiveness Research*. Institute of Education, University of London. <https://eric.ed.gov/?id=ED389826>
- Samperio Pacheco, V. M. (2019). Ecuaciones estructurales en los modelos educativos: Características y fases en su construcción. *Apertura (Guadalajara, Jal.)*, 11(1), 90-103. <https://doi.org/10.32870/ap.v11n1.1402>
- Sanders, W. L., & Wright, S. P. (2008). A response to Amrein-Beardsley (2008): "Methodological concerns about the Education Value-Added Assessment

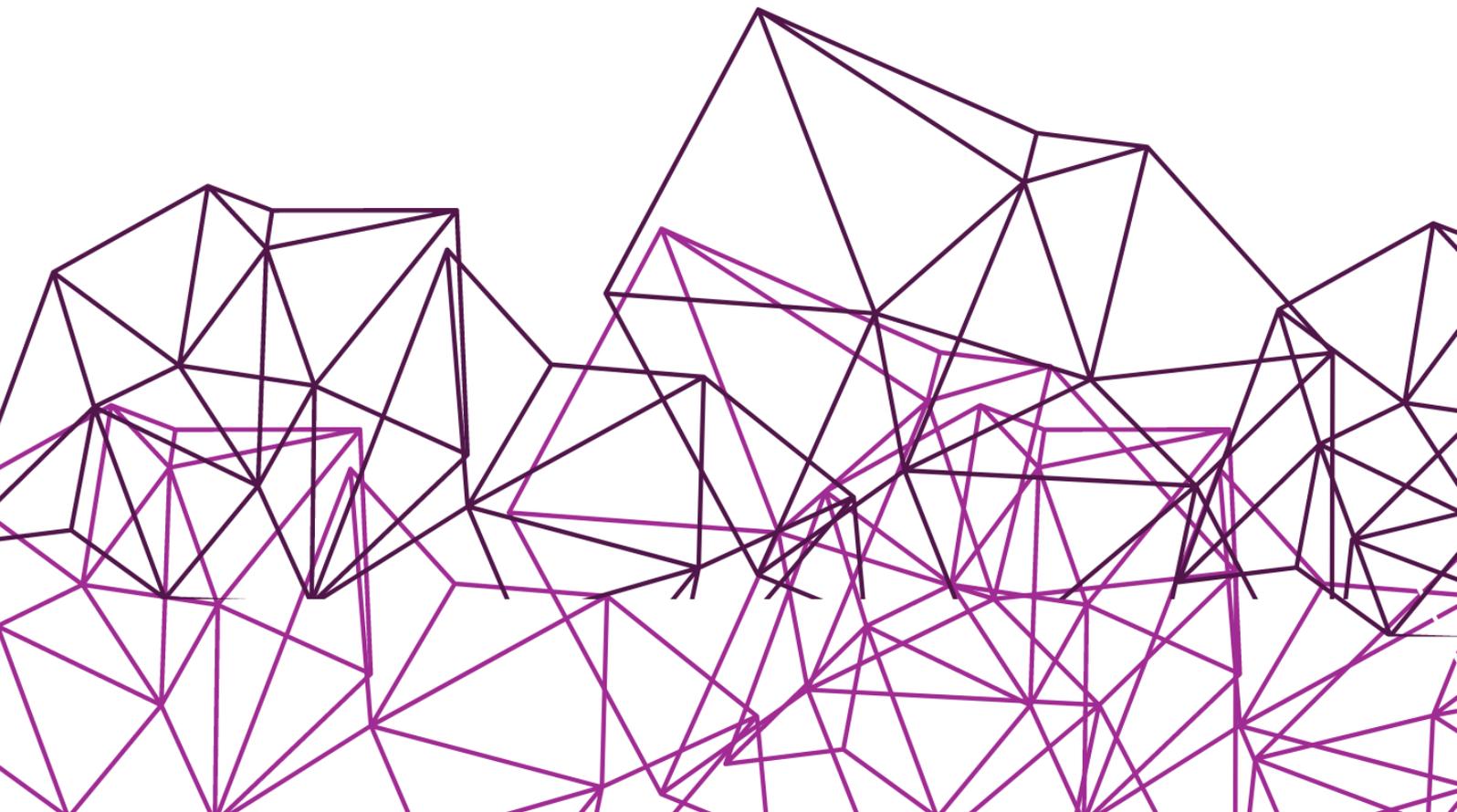
- System.” *A SAS White paper*.
- Scheerens, J. (1990). School Effectiveness Research and the Development of Process Indicators of School Functioning. *School Effectiveness and School Improvement*, *1*(1), 61-80. <https://doi.org/10.1080/0924345900010106>
- Scheerens, J., & Bosker, R. J. (1997). *The foundations of educational effectiveness* (1st ed). Pergamon.
- Scheerens, J., Luyten, H., van den Berg, S. M., & Glas, C. A. W. (2015). Exploration of direct and indirect associations of system-level policy-amenable variables with reading literacy performance. *Educational Research and Evaluation*, *21*(1), 15-39. <https://doi.org/10.1080/13803611.2015.1008520>
- Schleicher, A. (2019). *PISA 2018: Insights and Interpretations*. OECD Publishing.
- Shin, J., Lee, H., & Kim, Y. (2009). Student and School Factors Affecting Mathematics Achievement: International Comparisons Between Korea, Japan and the USA. *School Psychology International*, *30*(5), 520-537. <https://doi.org/10.1177/0143034309107070>
- Sirin, S. (2005). Socioeconomic Status and Academic Achievement: A Meta-Analytic Review of Research. *Review of Educational Research*, *75*. <https://doi.org/10.3102/00346543075003417>
- Snijders, T. A. B., & Bosker, R. J. (2011). *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*. SAGE.
- Sortkær, B., & Reimer, D. (2018). Classroom disciplinary climate of schools and gender – evidence from the Nordic countries. *School Effectiveness and School Improvement*, *29*(4), 511-528. <https://doi.org/10.1080/09243453.2018.1460382>
- Sousa, S., Park, E. J., & Armor, D. J. (2012). Comparing Effects of Family and School Factors on Cross-national Academic Achievement using the 2009 and 2006 PISA

- Surveys. *Journal of Comparative Policy Analysis: Research and Practice*, 14(5), 449-468. <https://doi.org/10.1080/13876988.2012.726535>
- Stevens, J. P. (2009). *Applied Multivariate Statistics for the Social Sciences* (5.^a ed.). Routledge. <https://doi.org/10.4324/9780203843130>
- Stringfield, S. C., & Slavin, R. E. (1992). A hierarchical longitudinal model for elementary school effects. En B. P. M. Creemers & G. J. Reezigt, *Evaluation of educational effectiveness* (pp. 35-69). ICO.
- Sullivan, A., & Calderwood, L. (2017). Surveys: Longitudinal, Cross-sectional and Trend Studies. En *The BERA/SAGE Handbook of Educational Research: Two Volume Set* (1-2, pp. 395-415). SAGE Publications Ltd. <https://doi.org/10.4135/9781473983953>
- Tan, C. Y., & Dimmock, C. (2022). The relationships among between-class ability grouping, teaching practices, and mathematics achievement: A large-scale empirical analysis. *Educational Studies*, 48(4), 471-489. <https://doi.org/10.1080/03055698.2020.1780416>
- Tan, C. Y., & Liu, D. (2018). What is the influence of cultural capital on student reading achievement in Confucian as compared to non-Confucian heritage societies? *Compare: A Journal of Comparative and International Education*, 48(6), 896-914. <https://doi.org/10.1080/03057925.2017.1369392>
- Tapia Blásquez, P. A. (2016). *Algunas contribuciones a la modelación multinivel en la Investigación Escolar en España* [[Http://purl.org/dc/dcmitype/Text](http://purl.org/dc/dcmitype/Text), Universitat Politècnica de València]. 10.4995/Thesis/10251/67388
- Teddlie, C., & Stringfield, S. (1993). *Schools Make a Difference: Lessons Learned from a 10-Year Study of School Effects*. Teachers College Press, 1234 Amsterdam Avenue, New York, NY 10027.

- Torres-Carrion, P., González González, C., Aciar, S., & Rodriguez, G. (2018). *Methodology for Systematic Literature Review applied to Engineering and Education*. <https://doi.org/10.1109/EDUCON.2018.8363388>
- 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. <https://doi.org/10.1177/0038040716683779>
- UNESCO. (2022a). *Por qué es importante la atención y educación de la primera infancia* | UNESCO. UNESCO. <https://www.unesco.org/es/articulos/por-que-es-importante-la-atencion-y-educacion-de-la-primera-infancia>
- UNESCO. (2022b). *Por qué la educación en la lengua materna es esencial* | UNESCO. UNESCO. <https://www.unesco.org/es/articulos/por-que-la-educacion-en-la-lengua-materna-es-esencial>
- Valdés, M. T. (2023). The effect of the month of birth on academic achievement: Heterogeneity by social origin and gender. *European Societies*, 1-27. <https://doi.org/10.1080/14616696.2023.2289652>
- van Hek, M., Kraaykamp, G., & Pelzer, B. (2018). Do schools affect girls' and boys' reading performance differently? A multilevel study on the gendered effects of school resources and school practices. *School Effectiveness and School Improvement*, 29(1), 1-21. <https://doi.org/10.1080/09243453.2017.1382540>
- Vargas Halabi, T., & Mora Esquivel, R. (2017). Tamaño de la muestra en modelos de ecuaciones estructurales con constructos latentes: Un método práctico. *Actualidades Investigativas en Educación*, 17(1), 1-34. <http://dx.doi.org/10.15517/aie.v17i1.27294>
- Weber, G. (1971). Inner-City Children Can Be Taught to Read: Four Successful Schools. *CBE Occasional Papers*, 18, 41.

- Woltman, H., Feldstain, A., MacKay, J. C., & Rocchi, M. (2012). An introduction to hierarchical linear modeling. *Tutorials in Quantitative Methods for Psychology*, 8(1), 52-69. <https://doi.org/10.20982/tqmp.08.1.p052>
- Wu, H., Gao, X., & Shen, J. (2020). Principal leadership effects on student achievement: A multilevel analysis using Programme for International Student Assessment 2015 data. *Educational Studies*, 46(3), 316-336. <https://doi.org/10.1080/03055698.2019.1584853>
- Wu, H., Shen, J., Zhang, Y., & Zheng, Y. (2020). Examining the effect of principal leadership on student science achievement. *International Journal of Science Education*, 42(6), 1017-1039. <https://doi.org/10.1080/09500693.2020.1747664>
- Wu, M. (2005). The role of plausible values in large-scale surveys. *Studies in Educational Evaluation*, 31(2), 114-128. <https://doi.org/10.1016/j.stueduc.2005.05.005>
- Yetişir, M. İ., & Bati, K. (2021). The Effect of School and Student-Related Factors on PISA 2015 Science Performance in Turkey. *International Journal of Psychology and Educational Studies*, 8(2), 170-186. <https://doi.org/10.52380/ijpes.2021.8.2.433>
- Zhu, Y., Kaiser, G., & Cai, J. (2018). Gender equity in mathematical achievement: The case of China. *Educational Studies in Mathematics*, 99(3), 245-260. <https://doi.org/10.1007/s10649-018-9846-z>

ANEXOS





VNiVERSiDAD
D SALAMANCA

School effectiveness and PISA tests. Factors associated with academic performance

EXTENDED ABSTRACT

Author: Cristina Frade Martínez

Directors:

Dr. Susana Olmos Migueláñez

Dr. Adriana Gamazo García

Salamanca, febrero 2024



Abstract

This doctoral thesis has been developed within the Doctoral Programme Training in the Knowledge Society at the University Institute of Education Sciences of the University of Salamanca (IUCE). It is part of the project awarded under the 2018 call for Knowledge Generation Projects under the funding of the Ministerio de Ciencia e Innovación of the Gobierno de España entitled *School Effectiveness for the Improvement of the System - EFI-6* (PGC2018-099174-B-I00). In addition, it has been funded by a grant for pre-doctoral contracts with reference PRE2019-087412 by MCIN/AEI/10.13039/501100011033 and by *FSE invierte en tu futuro* associated with the previous project.

As a complementary part of the pre-doctoral training developed in this period (2019 – present), three research stays have been carried out to provide an international perspective on the topic studied. The first one took place at the *Centre for Evaluation, Quality & Inspection* (EQI) at Dublin City University (DCU) in Ireland, with the aim of selecting high- and low-effectiveness schools in Ireland and obtaining information on the practices developed in them. The second visit was to the *Escola Superior de Educação e Ciências Sociais* of the Politécnico de Leiria (Portugal), achieving the same objective as in the Irish case, but applied to the Portuguese context. Finally, a one-month stay at the *Universidade da Coruña* (UDC) was carried out at national level, where high- and low-effectiveness schools in the Galician community were selected and their contextual characteristics were compared with those of the schools in Castilla y León.

The main objective of this thesis is to generate a framework of good educational practices related to school effectiveness at an international level, specifically studying the cases of three European countries: Spain, Ireland, and Portugal. This general objective is to be achieved through the following specific objectives:

- To select school-level and student-level context variables with significant relationships with student performance based on the data provided by the PISA 2018 tests using multilevel modelling methodology.
- To detect schools that perform above and below expectations according to their contextual characteristics in these three countries. In other words, to determine which schools have high and low school effectiveness.

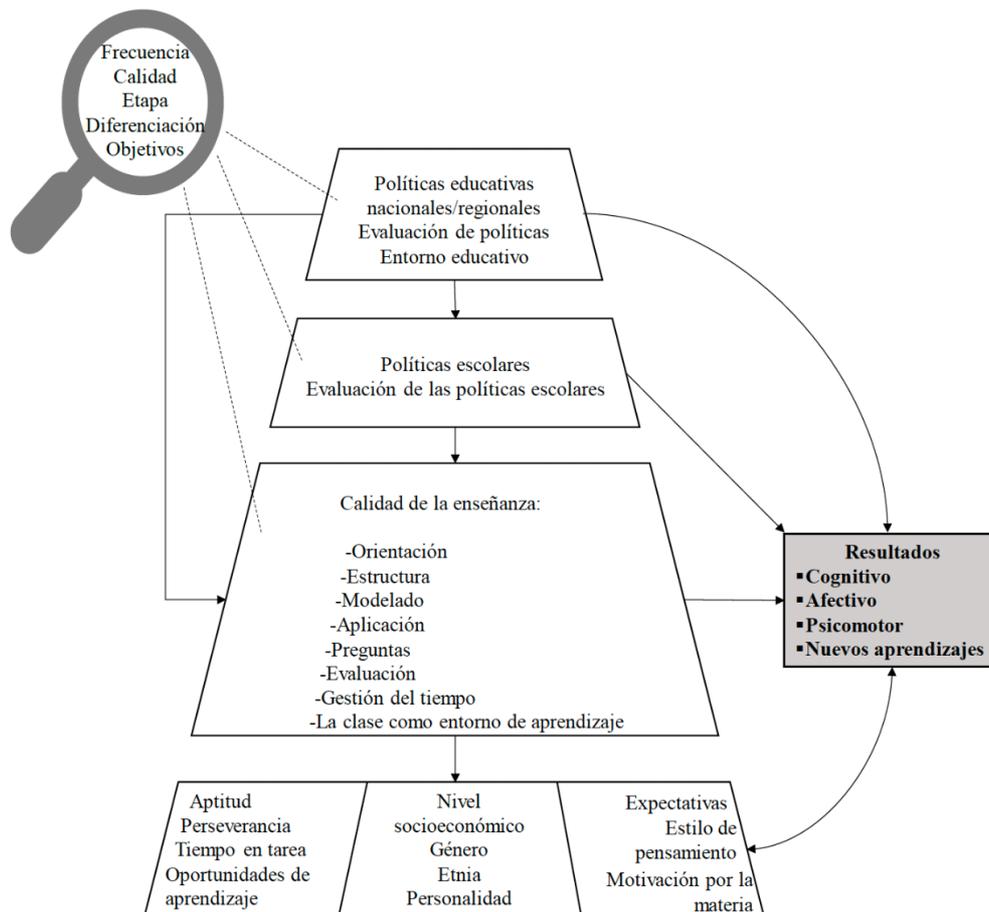
- To analyse the variables related to school effectiveness using binary logistic regression to determine the relationships between different process variables and this criterion variable.
- To find out the good practices implemented by schools with high residuals (high effectiveness) to try to replicate them in schools with lower residual values.

At the end of the 1960s, the *Equality of Educational Opportunity Study*, also known as the Coleman Report (Coleman, 1966), was published, a document in which reference was made to the importance of socioeconomic and demographic factors of students as a reason for inequalities in their academic performance, as opposed to the influence of the characteristics of the school; the main source of inequality up to that time. From this moment on, a new stream of research began in the field of education related to school effectiveness (Research in School Effectiveness - IEE) to try to demonstrate that schools do produce effects on educational performance and to understand which factors can contribute to it, to a greater or lesser extent.

This research line experienced a great growth in the 1990s, with the emergence of different models for the development of this type of research, such as the dynamic model proposed by Creemers and Kyriakides (2008), according to which there are factors at different levels that can help to explain educational performance. These range from characteristics related to the context at the country level, to the level of the educational institution or the student him/herself, as can be seen in Figure 1:

Figure 1

Dynamic model of school effectiveness



Note. Taken from *Factors associated with school performance and effectiveness: a mixed methods study from PISA 2015* (p. 95), by Gamazo, (2019), adapted from Creemers y Kyriakides, (2008).

Alongside these developments in EEI, there has been a growing interest in measuring educational performance through standardised assessment tests, also known as International Large-Scale Assessments (ILSA). The nature of these assessments makes it possible to characterise the variables extracted in terms of the level to which they refer according to the dynamic model (Creemers & Kyriakides, 2008). Standardised tests are generally developed by an international organisation and are intended to present a comparative view of education systems at the international level.

Three of these ILSAs are presented in this doctoral thesis:

5. The Programme for International Student Assessment (PISA): proposed by the Organisation for Economic Co-operation and Development (OECD) and aimed at assessing the skills of students at the end of compulsory schooling (15 years of age) in three basic skills: reading comprehension, mathematics, and science. It is held every three years.
6. The Trends in International Mathematics and Science Study (TIMSS): developed by the International Association for the Evaluation of Educational Achievement (IEA) for students in the 4th year of primary education and 2nd year of compulsory secondary education for the assessment of specified cognitive skills (mathematics and science). Implemented every four years.
7. The Progress in International Reading Literacy Study (PIRLS): also conducted by the IEA and aimed at 4th grade students to assess their reading literacy proficiency. It takes place every 5 years.

Within the framework of the proposed research and with the aim of finding out about the studies, especially of a secondary nature, carried out on the international large-scale assessment tests (ILSA) and their relationship with school effectiveness and academic performance, a systematic review of the literature was carried out. The main objective of the review focused on the evolution and characteristics of the secondary analyses of the three ILSAs mentioned, as well as their relationship with school effectiveness and academic achievement, to assess possible improvements in conducting similar studies and to define new lines of research.

To this end, we have opted for the combination of two review methodologies that provide a certain added value to this research and allow the extraction of information in a precise and planned manner, facilitating the replication of the proposed review. These two methodologies are mapping (Kitchenham et al., 2011; Torres-Carrion et al., 2018) and the systematic literature review (Systematic Literature Review or SLR) (Kitchenham, 2004).

The first methodology (mapping) aims to provide an overview of the subject matter (Kitchenham et al., 2011). While SLR is used to collect and synthesise relevant information in a rigorous manner to provide an in-depth understanding of different aspects related to ILSA-based secondary studies. The guidelines set out in the PRISMA 2020 Statement (Page et al., 2021) were followed in conducting this systematic literature review, thus giving the process a systematised character.

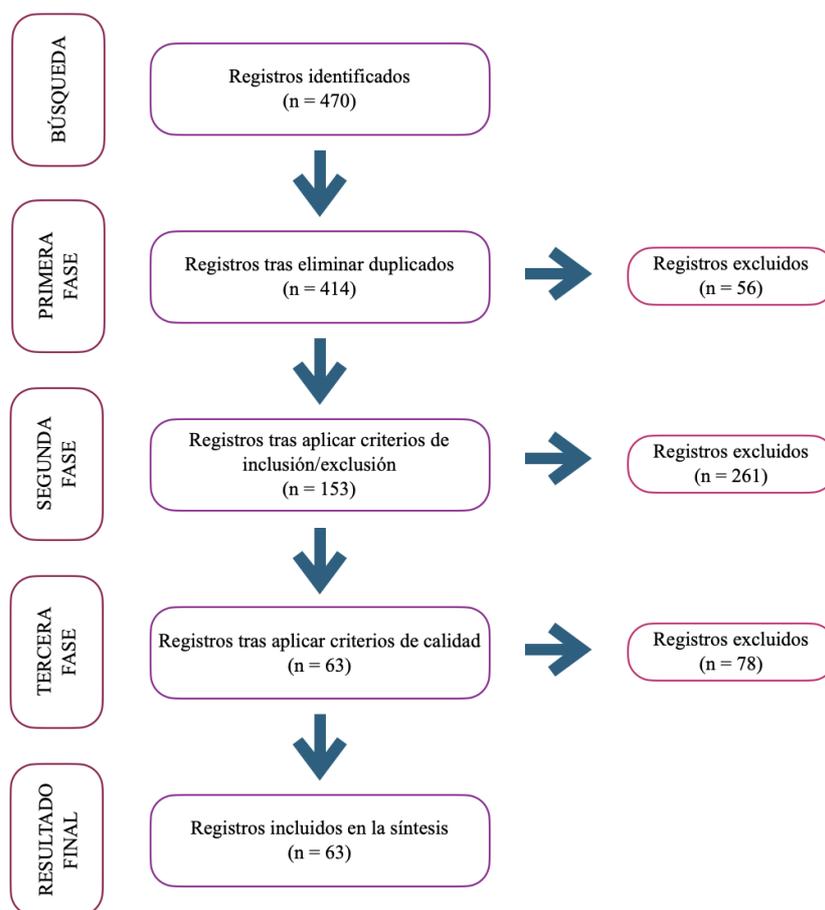
Different research questions were posed depending on the methodology used, differentiating between mapping (M) and SLR (PI) research questions:

- M1: How many studies have been published in scientific journals linking any standardised test (PISA, TIMMS or PIRLS) to school effectiveness?
- M2: What has been the evolution of publications on this topic over time, and when has research on this subject been at its peak?
- M3: Who are the leading authors in this field?
- M4: Which countries conduct the most such studies?
- M5: In which scientific journals are most studies related to large-scale testing and school effectiveness published?
- M6: What are the most used methodologies for secondary data analysis of international large-scale assessment tests?
- PI1: What are the most used predictor variables in research on school performance and effectiveness?
- PI2: Has research based on large-scale test results focused on academic performance or school effectiveness?
- PI3: Do the authors consider the hierarchical structure of large-scale test data (student, classroom, school, region, country, etc.) in their analyses?
- PI4: Do you work with plausible values and sampling weights of different hierarchies (pupils and schools) in parameter estimation?
- PI5: Are there many limitations to this type of study, and what are the most common?
- PI6: Generally, what are the societal implications of large-scale evaluation evidence-based studies?

The procedure involved different phases in which the articles from which to extract the information necessary to answer the questions posed were delimited, as shown in the following flow chart (figure 2):

Figure 2

Flowchart of the resource search and selection process



Analysing the 63 articles that made up the final sample, the following results were obtained (table 1):

Table 1

Responding to the research questions raised in the systematic literature review

Question	Answer
M1	Since 2000 (the first PISA report), 63 studies constructed with data from one of the selected ILSAs have been published.
M2	There is an increasing trend in the number of secondary studies based on ILSA data, with the year 2020 standing out with 14 publications.
M3	The 63 articles selected were signed by a total of 133 authors, with Martínez-Abad standing out with 4 publications.
M4	Spain is the country in which most studies of this type have been carried out, both comparing their results with those of other countries and analysing their specific case.
M5	Two scientific journals account for more than 20% of the selected research: Revista de Educación and School Effectiveness and School Improvement.

Question	Answer
M6	The most used methodology for data analysis has been multilevel modelling.
PI1	The most used predictor variables are, at student level: gender, socio-economic and cultural level, language spoken at home, immigration rate and grade repetition; and, at school level: average socio-economic level of the school, location (urban/rural), ownership, shortage of teachers at the school and shortage of educational resources.
PI2	The criterion variable is mostly academic performance (62 studies).
PI3	In 60 of the 63 investigations, the hierarchical structure of the data is considered, usually differentiating between two levels: student and school, although sometimes a third level referring to national characteristics was also included.
PI4	Methodological considerations regarding the treatment and use of plausible values are considered in more than 95% of the studies. This percentage is reduced in relation to the use of sampling weights, which are used correctly in 42 of the selected articles (66.67%).
PI5	Not all studies report limitations, the most common being: the cross-sectional nature of large-scale assessment tests, the need for longitudinal studies, the lack of information on students' prior performance, differences between scales and indicators for different editions of the same test, possible bias due to self-reporting, lack of classroom-level data, and students perceived social desirability.
PI6	The implications of these studies have been the formulation of recommendations to alleviate the deficiencies detected or the proposal of new lines of research.

Just like the evolution experienced in School Effectiveness Research, the techniques and instruments used for the analysis of the results of these studies have been changing. More than three decades ago, educational research underwent a revolution due to the contributions of different authors who showed that, in this field, observations cannot be considered independent (Aitkin & Longford, 1986; Raudenbush & Bryk, 2002). Therefore, it became necessary to apply other statistical techniques for the treatment of these data, proposing the use of multilevel models or hierarchical-linear models, among others.

Multilevel methodology is a statistical regression technique that builds on the nested organisation of data into higher-order units (students in classrooms, classrooms in schools, schools in countries, etc.) and serves to try to explain the variance between different levels of analysis simultaneously. They are therefore extensions of classical linear regression models in which several regression models are constructed for each level of analysis (Murillo, 2008). Multilevel models present a linear combination of different predictor variables with a criterion variable, each associated with a coefficient indicating

how much variation in the predictor variable affects the criterion variable (Gamazo, 2019; Hayes, 2006).

It has been proven in educational research that the use of multilevel models has increased over the years, using this methodology to conduct research with two (Gamazo et al., 2018; Yetişir & Bati, 2021), three (Gaviria Soto & Castro Morera, 2004; M. Lee et al., 2021; Tan & Liu, 2018) or even more levels of analysis (Murillo & Hernández-Castilla, 2011). In the case at hand, we present how the correct construction of two-level multilevel models (student and school) should be carried out using the data provided by PISA 2018.

The appropriate use of multilevel models for data analysis is associated with the need to comply with a few basic assumptions about the distribution of the elements that compose them and the relationships that are established between them:

Observation dependence

Inferential statistical techniques are constructed based on the independence between the different observations in a sample, but in the case of multilevel models, the existence of dependence between the observations of the different levels is necessary. This degree of dependence is measured by calculating the Intraclass Correlation Coefficient (ICC), which indicates the percentage of variability between students' scores that is attributable to the school level.

Linearity of the relationship between predictor variables and criterion variable

In the case of this assumption, multilevel regression techniques coincide with traditional regression assuming the linear nature of the relationship between the predictor variables and the criterion variable.

Distribution and variance of residuals

When we refer to residuals, we mean the difference between the actual score and the estimated score for a given unit according to the variables that make up the statistical model (Joaristi et al., 2014).

The residuals are assumed to be normally distributed with mean 0 and constant variance. Similarly, the homoscedasticity assumption of equality of the variance of the residual is assumed.

This research is based on a quantitative analysis approach in order to identify the variables that predict the two phenomena under study: academic performance and school effectiveness.

Education is a complex phenomenon, and therefore so is educational research. This complexity makes research design difficult and has meant that in order to achieve the different objectives set out in the same study it is necessary to use different methodologies within the same paradigm or even a combination of methodologies from the quantitative and qualitative paradigms, thus constituting a methodological approach known as Mixed Methods (Nunez Moscoso, 2017).

The research carried out proposes the study of different aspects related to the educational reality using different methodologies of quantitative analysis.

The research is divided into three main phases: first, a secondary analysis of the PISA tests is presented, focusing on the determination of the contextual factors associated with students' academic performance in the three main competencies studied in this test: reading comprehension, mathematics and science, using a multilevel analysis methodology (Gaviria Soto & Castro Morera, 2005; Raudenbush & Bryk, 1992). Then, based on the results obtained, the schools with high and low effectiveness were selected following the procedure used in previous research (Gamazo, 2019; Lizasoain & Angulo, 2014); to finally proceed with the study of those procedural variables significantly related to school effectiveness according to the data obtained in PISA using binary logistic regression.

Once these first three phases had been completed, and with the aim of going deeper into the factors associated with school effectiveness and obtaining different information from the very agents involved in the educational process, a questionnaire was drawn up and subsequently validated for active teachers at schools with high and low effectiveness on self-perceived effectiveness at their work centre.

This research presents a non-experimental *ex post facto* design (Bisquerra, 2004) as it is a secondary analysis of the data provided by the 2018 edition of the PISA tests. As it is a measurement carried out at a specific moment in time, this study could also be characterised as cross-sectional (Bisquerra, 2004; Cvetkovic-Vega et al., 2021).

The study sample presented here is made up of students who participated in PISA in the 2018 edition and who were enrolled in schools in Spain, Portugal, or Ireland. The selection of these countries was determined by their similar socio-economic characteristics and because they belong to the group of southern European countries known in financial and economic jargon as PIIGS (Portugal, Italy, Ireland, Greece and Spain) due to the difficulties they had in economic recovery after the 2008 crisis (Bellod Redondo, 2016).

Following the procedure used by several authors in similar studies, schools with fewer than 20 students assessed were eliminated from the analysis (Liu et al., 2014; Martínez-Abad et al., 2020) in order to ensure a correct analysis of the school-level variables (Gamazo, 2019). This meant that in the end the study sample consisted of 44650 students (34411 Spanish, 5551 Irish and 4688 Portuguese), 1309 schools (976 Spanish, 155 Irish and 178 Portuguese) and 24052 teachers (19991 Spanish and 4061 Portuguese).

To ensure the representativeness of the data obtained, it is necessary to use sampling weights in the analyses derived from the results of this test. Sampling weights are, in essence, a weight assigned to each student and school to represent the eligible cohort in the corresponding PISA edition. As stated in the PISA Data Analysis Manual (OECD, 2009), weighting is the recognition that some sample units are more important than others and should contribute more than others to the calculation of any population estimate.

Plausible values reflect the range of skills that a student could reasonably have based on his or her responses to the items (Wu, 2005), because students do not answer all of the questions asked in the competency tests. They are used to facilitate addressing problems related to biases arising from the estimation of population parameters. In this edition, a total of 10 plausible values were generated for each of the competences studied, values which have been considered in the statistical analyses carried out.

In order to detect the contextual factors related to the academic performance of students from the three selected European countries at the end of secondary education, the multilevel methodology (Raudenbush & Bryk, 1992; Snijders & Bosker, 2011) was used, thanks to which a statistical model was designed for each of the three competences studied in each of the three countries by combining the variables with a significant relationship with academic performance in that particular competence ($\alpha = .05$). The choice of this technique for the analysis of the data made it possible to consider the effect of contextual variables (included in Annex I) belonging to different levels, respecting the hierarchical structure of the data and facilitating the elimination of biases arising from the use of other types of data aggregation or disaggregation techniques (Gaviria Soto & Castro Morera, 2005; Raudenbush & Bryk, 1992).

The equations representing the final multilevel models for each competence in each of the three countries are presented below. The data referring to the influence of each of these variables on the respective criterion variable can be seen in the tables included in Annex II.

Final models for Spain

$$\begin{aligned} Maths_{ij} = & \gamma_{00} + \gamma_{01} * SCHSIZE_j + \gamma_{02} * STAFFSHORT_j + \gamma_{03} * PROAT6_j + \gamma_{10} \\ & * ESCS_{ij} + \gamma_{20} * GENDER_{ij} + \gamma_{30} * REPEAT_{ij} + \gamma_{40} * GRADE_{ij} \\ & + \gamma_{50} * AGE_{ij} + \gamma_{60} * SCCHANGE_{ij} + u_{0j} + r_{ij} \end{aligned}$$

(Equation 1)

$$\begin{aligned} Science_{ij} = & \gamma_{00} + \gamma_{01} * OTT1_j + \gamma_{02} * SCHSIZE_j + \gamma_{03} * STAFFSHORT_j + \gamma_{04} \\ & * PROAT6_j + \gamma_{10} * ESCS_{ij} + \gamma_{20} * GENDER_{ij} + \gamma_{30} * REPEAT_{ij} \\ & + \gamma_{40} * GRADE_{ij} + \gamma_{50} * AGE_{ij} + \gamma_{60} * LANG_{ij} + \gamma_{70} * SCCHANGE_{ij} \\ & + u_{0j} + r_{ij} \end{aligned}$$

(Equation 2)

$$\begin{aligned} Reading_{ij} = & \gamma_{00} + \gamma_{01} * OTT2_j + \gamma_{02} * SCHSIZE_j + \gamma_{03} * STAFFSHORT_j + \gamma_{04} \\ & * LOCAT_j + \gamma_{10} * ESCS_{ij} + \gamma_{20} * GENDER_{ij} + \gamma_{30} * REPEAT_{ij} + \gamma_{40} \\ & * GRADE_{ij} + \gamma_{50} * AGE_{ij} + \gamma_{60} * LANG_{ij} + \gamma_{70} * SCCHANGE_{ij} + u_{0j} \\ & + r_{ij} \end{aligned}$$

(Equation 3)

Final models for Ireland

$$\begin{aligned} Maths_{ij} = & \gamma_{00} + \gamma_{01} * N2ESCS_j + \gamma_{10} * ESCS_{ij} + \gamma_{20} * GENDER_{ij} + \gamma_{30} \\ & * DURECEC_{ij} + \gamma_{40} * REPEAT_{ij} + \gamma_{50} * AGE_{ij} + \gamma_{60} * LANG_{ij} \\ & + \gamma_{70} * INM2_{ij} + \gamma_{80} * SCCHANGE_{ij} + u_{0j} + r_{ij} \end{aligned}$$

(Equation 4)

$$\begin{aligned} Science_{ij} = & \gamma_{00} + \gamma_{01} * N2ESCS_j + \gamma_{10} * ESCS_{ij} + \gamma_{20} * REPEAT_{ij} + \gamma_{30} * AGE \\ & + \gamma_{40} * INM_2_{ij} + u_{0j} + r_{ij} \end{aligned}$$

(Equation 5)

$$\begin{aligned} Reading_{ij} = & \gamma_{00} + \gamma_{01} * N2ESCS_j + \gamma_{10} * ESCS_{ij} + \gamma_{20} * GENDER_{ij} + \gamma_{30} \\ & * REPEAT_{ij} + \gamma_{40} * AGE_{ij} + \gamma_{50} * INM_2_{ij} + \gamma_{60} * SCCHANGE_{ij} \\ & + u_{0j} + r_{ij} \end{aligned}$$

(Equation 6)

Final models for Portugal

$$\begin{aligned}
\text{Maths}_{ij} = & \gamma_{00} + \gamma_{01} * \text{CLSIZE}_j + \gamma_{02} * \text{N2ESCS}_j + \gamma_{10} * \text{ESCS}_{ij} + \gamma_{20} \\
& * \text{GENDER}_{ij} + \gamma_{30} * \text{DURECEC}_{ij} + \gamma_{40} * \text{REPEAT}_{ij} + \gamma_{50} * \text{GRADE}_{ij} \\
& + u_{0j} + r_{ij}
\end{aligned}$$

(Equation 7)

$$\begin{aligned}
\text{Science}_{ij} = & \gamma_{00} + \gamma_{01} * \text{CLSIZE}_j + \gamma_{02} * \text{PROAT6}_j + \gamma_{03} * \text{N2ESCS}_j + \gamma_{10} * \text{ESCS}_{ij} \\
& + \gamma_{20} * \text{GENDER}_{ij} + \gamma_{30} * \text{REPEAT}_{ij} + \gamma_{40} * \text{GRADE}_{ij} + u_{0j} + r_{ij}
\end{aligned}$$

(Equation 8)

$$\begin{aligned}
\text{Reading}_{ij} = & \gamma_{00} + \gamma_{01} * \text{CLSIZE}_j + \gamma_{02} * \text{N2ESCS}_j + \gamma_{10} * \text{ESCS}_{ij} + \gamma_{20} \\
& * \text{DURECEC}_{ij} + \gamma_{30} * \text{REPEAT}_{ij} + \gamma_{40} * \text{GRADE}_{ij} + u_{0j} + r_{ij}
\end{aligned}$$

(Equation 9)

In these equations:

Maths/Science/Reading_{ij} refers to the average score in that competence in each school.

y_{00} is the average of the schools in the specified competence.

$y_{01} - y_{04}$ are the level 2 covariates.

$y_{10} - y_{80}$ are the level 1 covariates.

u_{0j} is the difference between the school's reading literacy score minus the average of all schools.

r_{ij} is the level 1 residue.

These models served as the basis for calculating the difference between the actual score obtained in each school and the score estimated according to its contextual characteristics (residual). Based on the result obtained, the high and low performing schools in each country were defined, generating the criterion variable used in the next phase: school effectiveness. A total of 184 high and 202 low performing schools were detected in Spain, 29 high and 29 low performing schools in Ireland, and 29 high and 32 low performing schools in Portugal.

To test the relationship between the different non-contextual variables chosen (defined in Annex III) and school effectiveness, the Pearson correlation was calculated. If the result of this coefficient indicated the existence of significant correlations ($\alpha = .05$), this correlation was analysed using binary logistic regression techniques to determine the relationships between the different process variables and school effectiveness. This statistical technique is the most widely recommended for the study of dichotomous

criterion variables (Kleinbaum & Klein, 2010). Three different levels of study were considered based on the characteristics of students, teachers, and schools.

In the case of Spain, the results of the binary logistic regression indicate the relationship between 9 variables and school effectiveness (2 at teacher level and 7 at student level). These variables are expressed in table 2:

Table 2

Binary Logistic regression results in Spain

	B	Odds ratio	Converted Odds ratio	p-value
Metacognition: summarising	4.581	97.619		<0.001
Metacognition: assessing credibility	4.538	93.488		<0.001
Perception of cooperation at school	2.534	12.604		<0.001
Self-concept of reading: perceptions of competence	2.186	8.9		0.038
Teacher self-efficacy in classroom management	1.645	5.181		0.023
Teacher self-efficacy in building positive relationships with students	-2.018	0.133	7.519	0.003
Self-concept of reading: perception of difficulty	-2.711	0.066	15.152	0.009
Use of ICTs outside school for homework assignments	-3.376	0.034	29.412	<0.001
Subjective well-being: positive affect	-3.438	0.032	31.25	<0.001

Note: The *converted Odds ratio* column includes the probability of variables with a negative OR to facilitate comparison (1/OR).

In Ireland, binary logistic regression indicated the influence of 3 variables on school effectiveness (Table 3):

Table 3

Binary Logistic regression results in Ireland

	B	Odds ratio	Converted Odds ratio	p-value
Self-concept of reading: perceptions of competence	7.278	1447.561		0.003
Perceived autonomy in relation to ICT use	5.127	168.9497		0.031
Use of ICTs outside school for homework assignments	-5.929	0.003	333.333	0.005

In Portugal, 2 variables were found to be significantly related to school effectiveness (Table 4):

Table 4

Binary Logistic regression results in Portugal

	B	Odds ratio	p-value
Metacognition: summarising	4.74	114.487	0.000
Disciplinary climate in language lessons	2.392	10.94	0.025

After calculating the corresponding regressions, the goodness of fit of the different models was checked by calculating the *Nagelkerke's R²* statistic, all exceeding 55%, which indicates adequate values. Similarly, their predictive accuracy was checked, obtaining results of between 83% and 88.9%.

This study has been able to determine the existence of two variables with a consistent influence on students' academic performance regardless of the context analysed: their socio-economic and cultural level and grade repetition. The first of these was always positively related to performance, while the other indicated a negative influence on performance in the three countries studied. Socio-economic status appears as a significant variable in all three countries studied, although with a different contribution depending on the context. In the case of Spain, it is of great positive importance if it refers to the socio-economic status of individual students, being the variable with the greatest influence on performance in the three competences. Comparing this result with those of other studies (Gamazo, 2019), it is striking that the average socio-economic status of schools is not a significant factor in the PISA 2018 data. In contrast, this is in direct contrast to the results obtained in both Ireland and Portugal, where the socio-economic status variable is shown to have a strong relationship with academic performance at both levels (student and school). In the case of Ireland, as in the case of Spain, the socio-economic level of the students is the variable with the strongest influence on performance in all three competences. In Portugal, on the other hand, although both variables have a positive influence on performance, the variable that proves to have the greatest weight on all three competences is the students' grade.

Another variable that consistently appears as one of the most influential on academic performance is grade repetition. This variable has been associated in all cases with academic disadvantage, which implies that having repeated a year leads to negative

consequences for students in terms of educational achievement. On many occasions, students who repeat a year may see their level of self-esteem reduced, observing greater difficulties for their normal academic development, even leading to academic dropout.

Of the three countries studied, repetition proves to be the variable with the strongest negative influence on performance in two of the three competences in Spain, science and reading comprehension, and Ireland, mathematics and reading comprehension (although there is not much difference with the most negative variable) and in all three competences in Portugal.

The fact that in Spain and Portugal there are similar values in terms of the number of repeating students, but the influence of this variable depends on the place we are studying, raises an important debate about the usefulness or not of this measure to improve student learning. Educational research is of the opinion that this is not an effective measure, since students who repeat grades do not show an improvement in their performance (Cappella & Weinstein, 2001; Gómez, 2013; Jimerson, 2001). Therefore, the question is whether this measure is beneficial in terms of promoting student achievement or whether it is detrimental to their education.

The importance of checking the correlations between the variables prior to the elaboration of the hierarchical models when interpreting the results was indicated earlier. In the case of grade repetition, both in Spain and Portugal, this variable and the students' grade itself show a medium/high correlation (0.68 in Spain and 0.668 in Portugal), having opted for their inclusion in the statistical modelling as the values are not too high. Even so, the correlation between both variables must be considered, so the results of both variables are interpreted jointly and with caution.

Although these two variables refer to different characteristics of the students, they are based on some common concepts for their construction: the grade of the students themselves and their age, variables also considered independently in this analysis. Due to the close relationship between these three variables, it has been decided to include them in the same section and thus interpret them jointly.

In contrast to the directly negative relationship between grade repetition and academic performance, age and grade level have a positive influence on academic performance. This makes sense, as this is interpreted to mean that the higher these values, the higher the performance.

The age of the students may be associated with their level of cognitive development, which would imply a greater maturity to face this assessment test. Hence, the existence

of differences in performance according to the month of birth has been studied and demonstrated (Valdés, 2023). Similarly, the fact that a student of this age is in higher grades seems to indicate that he/she has already achieved the expected competences for his/her corresponding academic level according to his/her age, which implies a positive relationship with performance. Student age was found to have a significant influence on performance in all three competences in Spain and Ireland. While grade was found to be the most important variable in Portugal (in all competences) and one of the most important in the case of Spain.

Student gender has been shown to be another vitally important variable in explaining academic performance in different disciplines. It is a variable that again features consistently in the models, except for reading literacy in Portugal and science literacy in Ireland.

Historically, the existence of a gender gap in science, technology, engineering, and mathematics (STEM) subjects in favour of male students has been indicated. The results presented in this research support this assertion, although a much more detailed and focused study of this gender gap in scores would be needed to confirm this unequivocally. Other research over the last decade has found similar results (Zhu et al., 2018). In recent decades, several countries have made progress in reducing the gender gap thanks to initiatives that encourage the participation of girls and women in these subjects, such as *Girls who code*, an organisation dedicated to offering educational programmes focused on STEM subjects and aimed at girls between 13 and 17 years of age, thus trying to reduce this gender gap (Girls Who Code, 2024).

In both Spain and Ireland, except for science proficiency in Ireland, the number of school changes has been shown to negatively affect student performance, although to a lesser extent than the variables discussed above. The fact that a student changes school may be due to a variety of reasons and may in turn be associated with several consequences that hinder the learning process. Some of these consequences that may result from school changes are measured in PISA, such as: the sense of belonging to the school, student satisfaction or student self-efficacy (Ministerio de Educación y Formación Profesional, 2019), all of which are factors related to school climate. In order to establish educational policies and practices that reduce the negative impact of this variable, and taking into account that the variable itself is difficult to intervene directly, it would be advisable to carry out exhaustive studies to check what other factors may be moderating the relationship between the number of school changes and performance, such as the

feeling of belonging to the school or school climate, in order to be able to intervene on them.

Research has shown that mother tongue is a significant factor in academic performance at different educational levels (Mateos Claros et al., 2019; Özdemir, 2016; Riederer & Verwiebe, 2015). This study corroborates these results, but only in the case of Ireland when working on performance in mathematical competence and in Spain when studying its relationship with reading comprehension and performance in science. In Ireland, this variable has a positive influence, while in Spain the relationship is negative.

This variable is also closely associated with migration status, as many students who have had to emigrate from their country of origin have a different mother tongue to that of their new country of residence. The migration status variable showed a significant relationship with performance in all three competences in Ireland, in all cases being associated with a reduction in performance. These results should be interpreted carefully, as these two variables were found to have a medium-high correlation (0.548). Some studies have analysed the influence of these two variables on student performance in Ireland, and the authors may be providing clues as to the unequal influence of these two variables based on the type of immigration in this country (Darmody & Smyth, 2018). Highly skilled immigration that finds it difficult to obtain jobs commensurate with their training. This is why we cannot make unequivocal statements about the influence of immigration status or mother tongue on the performance of Irish students, and why more in-depth studies should be carried out to understand the reasons for the results we have obtained.

The only variable related to teacher characteristics that has been shown to have a significant relationship with educational performance is the rate of teachers with doctoral studies. Specifically in mathematics and science performance in Spain and science in Portugal. While in Spain this variable shows a positive relationship with performance, the opposite is true in Portugal. It is interesting that in contexts as similar as these two territories of the Iberian Peninsula, such differences are generated. Portugal has an index with an average value of 0.098, whereas in Spain this value is 0.033. The number of teaching staff with doctoral studies in Portugal far exceeds the same group in Spain, almost tripling the Spanish rate. These results seem to indicate that the higher the rate of teachers with an academic background equivalent to ISCED level 6, the more likely it is that student performance will decrease. But this is not a hypothesis that we can confirm without a doubt, among other aspects, because the Spanish results must be taken carefully

due to the high percentage of imputed values referring to this variable (approximately 20%). In this respect, this makes it difficult for us to really establish the relationship between the educational level of teachers and the performance of 15-year-old students adequately.

Some international bodies, such as UNESCO, suggest the importance of early childhood education (2022), in part because of their malleability and their capacity to learn and absorb new knowledge, attitudes, etc. As has long been asserted, children are sponges (Montessori, 1986). This fact has also been studied in educational research indicating, or rather confirming, this relationship (Cordero Ferrera et al., 2013; Martínez-Abad & León, 2023).

The information obtained from the study presented here is contradictory, since in Portugal this variable is shown to have a positive influence on students' academic performance in mathematics and reading comprehension, with a negative influence in the case of Irish students' mathematical competence. In all cases, the effects are not very high.

In terms of variables influencing school effectiveness, three variables were found to be consistent in at least two models: metacognition understood as students' ability to summarise (Spain and Portugal), reading self-concept as perceived competence (Spain and Ireland) and the use of ICT outside school for homework (Spain and Ireland).

Metacognition, understood as students' ability to summarise, was found to be the variable with the greatest positive influence on the effectiveness of schools in Spain and Portugal. In absolute terms, the interpretation that can be made of these results is that students who are more able to summarise a text adequately are more likely to belong to highly effective schools. However, this goes further, since this variable is also measuring the way in which students apply previously acquired cognitive processes to a new task in order to tackle it successfully. The analysis presented here demonstrates how this factor is of great importance for school effectiveness and could be promoted and worked on by schools themselves, aiming to increase their effectiveness without the need for large capital investments or structural changes in education systems.

The results of the analysis indicate that pupils' self-concept about reading positively affects school effectiveness in Spain and Ireland, although to different extents; its influence being much greater in the English-speaking country. Commonly, variables of this type have been considered more as educational outcome variables than as process variables (Inglés et al., 2009), limiting the interpretations that can be derived from their

study to objectives to be achieved with the teaching-learning process, rather than malleable procedures to foster student success.

There is currently a wide-ranging debate on the use of ICT inside and outside the classroom by students at different educational stages, largely linked to the capacity and potential of these devices, particularly mobile devices. Initially, the inclusion of technology in education was seen as an opportunity for progress, although it has subsequently come to be seen as an instrument that fosters educational inequality among students, among other aspects, due to the difficulty of access to them by the most disadvantaged sectors (Alarcó Estévez, 2022).

The variable with an influence on school effectiveness in Ireland and Spain in terms of ICT use focuses on the use of technology to carry out school tasks outside the educational institution itself. Contrary to what might be expected, although in line with other research (Gamazo, 2019), the relationship between the two variables is inverse. In other words, the greater the use of ICT in this way, the greater the likelihood that the educational institution is considered to have low effectiveness. This conclusion has been extrapolated to the use of ICT in the classroom, considering that their use, in this case the use of mobile devices, has negative consequences on student performance.

This fact is closely related to another of the current educational debates, that of homework outside school, and therefore, in order to understand in depth the effects of ICT use on school effectiveness, it would be interesting to study these two factors together, together with many others that could be directly intervened in school practices.

In the development of this research, a series of limitations have been detected, which are set out below:

- The need for longitudinal studies to understand educational trends in order to guide educational decision-making (Bokhove et al., 2019; Li, 2016; H. Wu, Gao, et al., 2020). Conducting such studies in the field of education involves a particular difficulty associated with the identification of students and other hard-to-control variables such as dropout.
- Cultural bias, or that the conceptualisation and application of ILSAs do not fully take into account the differential characteristics and idiosyncrasies of the participating countries (Lenkeit, 2012). To alleviate this limitation, experts from the different participating countries collaborate in their conceptualisation, trying to minimise the effect this may have on the educational outcomes being measured.

- The need for data at the classroom level that makes it possible to verify and understand the good practices developed by teachers that influence the academic performance of students and, at a macro level, the effectiveness of schools (Burns et al., 2020; He & Fischer, 2020; H. Wu, Gao, et al., 2020). Not having data at the classroom level makes it impossible to study variables that could have a certain importance in school effectiveness, such as teaching methodologies, the conditions underpinning the grouping of students, etc.
- The potential loss of objectivity resulting from obtaining contextual information through self-reports from students, teachers and schools (Danhier, 2018; H. Wu, Gao, et al., 2020; H. Wu, Shen, et al., 2020).
- The promotion of competitiveness between countries and schools based on the results of large-scale assessment tests (Murillo, 2004).
- The number of missing values in certain variables prevents confirmation of the relationships established according to the analyses carried out, as for example in the case of doctoral teaching staff in Spain (20% of missing values).
- The lack of collaboration of the schools in obtaining qualitative information on the educational practices carried out in them, specifically on those related to the variables that were shown to influence school effectiveness in some way. There was a particular reason for this lack of collaboration: the consequences of the pandemic caused by the Covid-19 virus.

Ultimately, it should be clear which uses can and should be made of large-scale assessment tests and which should not. ILSAs are intended to monitor the state of education systems, but if important decisions are to be made based on them, they should not be considered as absolute truths but only as tools that help to obtain information and to formulate more in-depth research objectives to guide such decision-making.

As future lines of research, it is proposed to carry out trend studies that could be developed in order to find out how the influence of different factors has changed over time and the inclusion of new factors due to political, social or cultural changes. An example of these changes can be seen by comparing the analysis of the PISA 2015 data (Gamazo, 2019), where migration status was relevant, in contrast to this study constructed on the PISA 2018 data where this variable does not show any importance. It is interesting to compare whether there have been structural changes in terms of the characteristics of the population group we are referring to, since if so, it would be necessary to compare the

relationship between these changes in terms of the idiosyncrasies of the group itself and their impact on students' educational outcomes.

Similarly, it is proposed that the application of mixed methods would provide qualitative information that could not be considered in this study. As part of this research, a questionnaire has been developed and validated for in-service secondary school teachers on self-perceived effectiveness in their schools. The application of this questionnaire could provide this real insight into the educational practices developed in effective and ineffective schools, thus obtaining other information that would allow us to understand the causes of this difference in terms of effectiveness.

Finally, it could be affirmed that this doctoral thesis has adequately achieved the objectives set. It has achieved an approach to the good practices carried out in schools with high and low effectiveness, although highlighting the need for further study with the combination of obtaining qualitative data that complement this research and help to understand to a greater extent how school improvement can be achieved.

REFERENCES

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.
<https://doi.org/10.2307/2981882>
- Alarcó Estévez, G. (2022). Mediación de las TIC en el contexto de una educación inclusiva. *Serie Científica de la Universidad de las Ciencias Informáticas*, 15(7), 31-54.
- Bellod Redondo, J. F. (2016). PIGS: Austeridad fiscal, reformas estructurales y crecimiento potencial. *Revista de economía mundial*, 43, 161-177.
<https://doi.org/10.33776/rem.v0i43.3855>
- Bisquerra, R. (2004). *Metodología de la investigación educativa*. La Muralla.
- Bokhove, C., Miyazaki, M., Komatsu, K., Chino, K., Leung, A., & Mok, I. A. C. (2019). The Role of “Opportunity to Learn” in the Geometry Curriculum: A Multilevel Comparison of Six Countries. *Frontiers in Education*, 4.
<https://doi.org/10.3389/feduc.2019.00063>
- Burns, E. C., Martin, A. J., & Collie, R. J. (2020). Supporting and thwarting interpersonal dynamics and student achievement: A multi-level examination of PISA 2015. *International Journal of Research & Method in Education*, 43(4), 364-378.
<https://doi.org/10.1080/1743727X.2020.1757639>
- Cappella, E., & Weinstein, R. S. (2001). Turning around reading achievement: Predictors of high school students’ academic resilience. *Journal of Educational Psychology*, 93(4), 758-771. <https://doi.org/10.1037/0022-0663.93.4.758>
- Coleman, J. S. (1966). *Equality of Educational Opportunity*.
- Cordero Ferrera, J. M., Crespo Cebada, E., & Pedraja Chaparro, F. M. (2013). Rendimiento educativo y determinantes según PISA: Una revisión de la literatura

- en España. *Revista de educación*, 362, 273-297. <https://doi.org/10.4438/1988-592X-RE-2011-362-161>
- Creemers, B. P. M., & Kyriakides, L. (2008). *The Dynamics of Educational Effectiveness: A contribution to policy, practice and theory in contemporary schools*. Routledge; Scopus. <https://doi.org/10.4324/9780203939185>
- Cvetkovic-Vega, A., Maguiña, J. L., Soto, A., Lama-Valdivia, J., & López, L. E. C. (2021). Estudios transversales. *Revista de la Facultad de Medicina Humana*, 21(1), 179-185. <https://doi.org/10.25176/rfmh.v21i1.3069>
- Danhier, J. (2018). How Big is the Handicap for Disadvantaged Pupils in Segregated Schooling Systems? *British Journal of Educational Studies*, 66(3), 341-364. <https://doi.org/10.1080/00071005.2017.1322682>
- Darmody, M., & Smyth, E. (2018). Immigrant Student Achievement and Educational Policy in Ireland. En L. Volante, D. Klinger, & O. Bilgili (Eds.), *Immigrant Student Achievement and Education Policy: Cross-Cultural Approaches* (pp. 119-135). https://doi.org/10.1007/978-3-319-74063-8_8
- Gamazo, A. (2019). *Factores asociados al rendimiento y a la eficacia escolar: Un estudio basado en métodos mixtos a partir de PISA 2015*. <https://doi.org/10.14201/gredos.140406>
- Gamazo, A., Martínez-Abad, F., Olmos-Migueláñez, S., & Rodríguez-Conde, M. J. (2018). Evaluación de factores relacionados con la eficacia escolar en PISA 2015. Un análisis multinivel. *Revista de educación*, 379, 56-84. <https://doi.org/10.4438/1988-592X-RE-2017-379-369>
- Gaviria Soto, J. L., & Castro Morera, M. (2004). A Multilevel Study on Factors of School Effectiveness in Developing Countries: The Case of Brazilian Resources. *Education Policy Analysis Archives*, 12, 20-20.

<https://doi.org/10.14507/epaa.v12n20.2004>

Gaviria Soto, J. L., & Castro Morera, M. (2005). *Modelos jerárquicos lineales*. La Muralla.

Girls Who Code. (2024). *Girls Who Code | Home*. <https://girlswhocode.com/>

Gómez, G. (2013). Los efectos de la repitencia en tanto que política pública en cuatro países del Cono Sur: Argentina, Brasil, Chile y Uruguay: un análisis en base a PISA 2009. *Revista Latinoamericana de Educación Comparada: RELEC*, 4(4), 59-70.

Hayes, A. F. (2006). A Primer on Multilevel Modeling. *Human Communication Research*, 32(4), 385-410. <https://doi.org/10.1111/j.1468-2958.2006.00281.x>

He, J., & Fischer, J. (2020). Differential associations of school practices with achievement and sense of belonging of immigrant and non-immigrant students. *Journal of Applied Developmental Psychology*, 66, 101089. <https://doi.org/10.1016/j.appdev.2019.101089>

Inglés, C. J., Pastor, Y., Torregrosa, M. S., Redondo, J., & García, J. M. (2009). Diferencias en función del género y el curso académico en dimensiones del autoconcepto: Estudio con una muestra de adolescentes españoles. *Anuario de psicología*, 40(2), 271-288.

Jimerson, S. R. (2001). Meta-analysis of grade retention research: Implications for practice in the 21st century. *School Psychology Review*, 30(3), 420-437.

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 Lineales. *Estudios sobre educación*, 27, 37-61. <https://doi.org/10.15581/004.27.37-61>

- Kitchenham, B. (2004). *Procedures for Performing Systematic Reviews*. Keele University.
- Kitchenham, B., Budgen, D., & Brereton, P. (2011). Using mapping studies as the basis for further research—A participant-observer case study. *Information & Software Technology, 53*, 638-651. <https://doi.org/10.1016/j.infsof.2010.12.011>
- Kleinbaum, D. G., & Klein, M. (2010). *Logistic Regression: A Self-Learning Text*. Springer New York. <https://doi.org/10.1007/978-1-4419-1742-3>
- Lee, M., Ryoo, J. H., & Walker, A. (2021). School Principals' Time Use for Interaction with Individual Students: Macro Contexts, Organizational Conditions, and Student Outcomes. *American Journal of Education, 127*(2), 303-344. <https://doi.org/10.1086/712174>
- Lenkeit, J. (2012). How effective are educational systems? A value-added approach to study trends in PIRLS. *Journal of Educational Research Online, 4*(2), 143-173.
- Li, H. (2016). How is formative assessment related to students' reading achievement? Findings from PISA 2009. *Assessment in Education: Principles, Policy & Practice, 23*(4), 473-494. <https://doi.org/10.1080/0969594X.2016.1139543>
- Liu, H., Damme, J., Gielen, S., & Van den Noortgate, W. (2014). School processes mediate school compositional effects: Model specification and estimation. *British Educational Research Journal, 41*. <https://doi.org/10.1002/berj.3147>
- Lizasoain, L., & Angulo, A. (2014). Buenas prácticas de escuelas eficaces del País Vasco. Metodología y primeros resultados. *Participación educativa. Revista del Consejo Escolar del Estado, 3*(4), 17-28. <https://doi.org/10.4438/1886-5097-PE>
- Martínez-Abad, F., Gamazo, A., & Rodríguez-Conde, M.-J. (2020). Educational Data Mining: Identification of factors associated with school effectiveness in PISA assessment. *Studies in Educational Evaluation, 66*, 100875.

<https://doi.org/10.1016/j.stueduc.2020.100875>

- Martínez-Abad, F., & León, J. (2023). Inferencia causal en investigación educativa: Análisis de la causalidad en estudios observacionales de carácter transversal. *RELIEVE. Revista Electrónica de Investigación y Evaluación Educativa*, 29(2), 1-24. <https://doi.org/10.30827/relieve.v29i2.26843>
- Mateos Claros, F., Olmedo Ruiz, F. J., Esteban Ibáñez, M., & Amador Muñoz, L. V. (2019). Lengua materna, cultura y rendimiento en un contexto multicultural de Educación Infantil. *Ocnos: Revista de Estudios sobre Lectura*, 18(2), 44-54. https://doi.org/10.18239/ocnos_2019.18.2.1966
- Ministerio de Educación y Formación Profesional. (2019). *PISA 2018. Programa para la Evaluación Internacional de los Estudiantes. Informe español (PISA)*. INEE. <https://sede.educacion.gob.es/publiventa/pisa-2018-programa-para-la-evaluacion-internacional-de-los-estudiantes-informe-espanol/evaluacion-examenes/23505>
- Murillo, F. J. (2004). La investigación sobre «eficacia escolar» a debate: Análisis de las críticas y aportaciones. *Tendencias pedagógicas*, 9, 111-130.
- Murillo, F. J. (2008). Los modelos multinivel como herramienta para la investigación educativa. *Magis: Revista Internacional de Investigación en Educación*, 1, 45-62.
- Murillo, F. J., & Hernández-Castilla, R. (2011). Efectos escolares de factores socioafectivos. Un estudio Multinivel para Iberoamérica. *Revista de investigación educativa, RIE*, 29(2), 407-428.
- Nunez Moscoso, J. (2017). Los métodos mixtos en la investigación en educación: Hacia un uso reflexivo. *Cadernos de Pesquisa*, 47(164), 632-649. <http://dx.doi.org/10.1590/198053143763>
- OECD. (2009). *PISA Data Analysis Manual: SPSS, Second Edition*. OECD Publishing.

https://www.oecd-ilibrary.org/education/pisa-data-analysis-manual-spss-second-edition_9789264056275-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. <https://doi.org/10.1177/1474904115627159>
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., ... Alonso-Fernández, S. (2021). Declaración PRISMA 2020: Una guía actualizada para la publicación de revisiones sistemáticas. *Revista Española de Cardiología*, 74(9), 790-799. <https://doi.org/10.1016/j.recesp.2021.06.016>
- Raudenbush, S. W., & Bryk, A. S. (1992). Hierarchical linear models: Applications and data analysis methods. En *Hierarchical linear models: Applications and data analysis methods*. Sage.
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods*. SAGE.
- 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. <https://doi.org/10.1093/esr/jcv063>
- Snijders, T. A. B., & Bosker, R. J. (2011). *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*. SAGE.
- Tan, C. Y., & Liu, D. (2018). What is the influence of cultural capital on student reading

- achievement in Confucian as compared to non-Confucian heritage societies? *Compare: A Journal of Comparative and International Education*, 48(6), 896-914. <https://doi.org/10.1080/03057925.2017.1369392>
- Torres-Carrion, P., González González, C., Aciar, S., & Rodriguez, G. (2018). *Methodology for Systematic Literature Review applied to Engineering and Education*. <https://doi.org/10.1109/EDUCON.2018.8363388>
- UNESCO. (2022). *Por qué es importante la atención y educación de la primera infancia* | UNESCO. UNESCO. <https://www.unesco.org/es/articles/por-que-es-importante-la-atencion-y-educacion-de-la-primera-infancia>
- Valdés, M. T. (2023). The effect of the month of birth on academic achievement: Heterogeneity by social origin and gender. *European Societies*, 1-27. <https://doi.org/10.1080/14616696.2023.2289652>
- Wu, H., Gao, X., & Shen, J. (2020). Principal leadership effects on student achievement: A multilevel analysis using Programme for International Student Assessment 2015 data. *Educational Studies*, 46(3), 316-336. <https://doi.org/10.1080/03055698.2019.1584853>
- Wu, H., Shen, J., Zhang, Y., & Zheng, Y. (2020). Examining the effect of principal leadership on student science achievement. *International Journal of Science Education*, 42(6), 1017-1039. <https://doi.org/10.1080/09500693.2020.1747664>
- Wu, M. (2005). The role of plausible values in large-scale surveys. *Studies in Educational Evaluation*, 31(2), 114-128. <https://doi.org/10.1016/j.stueduc.2005.05.005>
- Yetişir, M. İ., & Bati, K. (2021). The Effect of School and Student-Related Factors on PISA 2015 Science Performance in Turkey. *International Journal of Psychology and Educational Studies*, 8(2), 170-186. <https://doi.org/10.52380/ijpes.2021.8.2.433>

Zhu, Y., Kaiser, G., & Cai, J. (2018). Gender equity in mathematical achievement: The case of China. *Educational Studies in Mathematics*, 99(3), 245-260.
<https://doi.org/10.1007/s10649-018-9846-z>

ANNEXES

Annex I – Contextual variables

	Variable	Label	Values
Level 1 - Students	Index of socioeconomical and cultural status	ESCS	Continue {-6,07 – 3,72}
	Gender	GENDER	0: Male 1: Female
	Age	AGE	Continue
	Grade	GRADE	1° ESO – 1° Bachillerato
	Index of immigration	IMMIG	0: Native, 1: First generation immigrant 2: Second generation immigrant
	Grade repetition	REPEAT	0: No 1: Yes
	Number of school changes	SCCHANGE	0: No changes 1: One 2: Two or more
	Duration in ECEC	DURECEC	Continue {<1 year – 8 years}
	Language spoken at home	LANG	0: Language of the test 1: Other
Level 2 - Schools	School size	SCHSIZE	Continue {37 – 2698}
	Class size	CLSSIZE	Continue {13 – 53}
	Shortness of Educational resources	EDUSHORT	Continue {-1,42 – 2,96}
	Staff shortness	STAFFSHORT	Continue {-1,45 – 4,04}
	School ownership	SCHOWN	1: Private 2: Charter school 3: Public
	Ratio teacher-students	RATIO	Continue {1 – 51,58}
	Location	LOCAT	1: Rural area (less than 3000 inhabitants) 2: Small village (3000 – 15000 inhabitants) 3: Village (15000 – 100000 inhabitants) 4: Town (100000 – 1000000 inhabitants) 5: City (More than 1000000 inhabitants)
	Teachers with Master Studies	PROAT5AM	Continue {0 - 1}
	Teachers with Doctorate Studies	PROAT6	Continue {0 - 1}
	Teacher gender*	TCH_GEN	Continue {0 – 3}
Teacher age*	TCH_AGE	Continue {33,79 – 55,55}	

Variable	Label	Values
Teacher experience at this school*	TCH_EXPSCH	Continue {1 – 39}
Teacher experience total*	TCH_EXPTOT	Continue {6,78 – 39}
Teacher training*	OTT1	Continue {0 – 1}
Continuous teacher training*	OTT2	Continue {0 – 1}
Average gender of Students*	N2GEN	Continue {1 – 2}
Average ESCS of Students*	N2ESCS	Continue {-1,69 – 1,46}
Average repeat Students*	N2REPEAT	Continue {0 – 1}
Average immigration*	N2INMIG	Continue {1 – 2,43}

Annex II – Final Multilevel models

Final model of mathematical competence in Spain

Fixed effects	Coefficient	Standard error	t-ratio	g. l.	p-value
Intercept, y_{00}	350.194552	43.211894	8.104	72	<0.001
SCHSIZE, y_{01}	0.010999	0.002687	4.094	511	<0.001
STAFFSHORT, y_{02}	-4.614822	1.356288	-3.403	320	<0.001
PROAT6, y_{03}	56.384914	26.729524	2.109	971	0.035
ESCS, y_{10}	13.236273	0.968905	13.661	34	<0.001
GENDER, y_{20}	-17.004681	2.164621	-7.856	24	<0.001
REPEAT, y_{30}	-57.323130	7.347867	-7.801	21	<0.001
GRADE, y_{40}	25.908413	4.743159	5.462	38	<0.001
AGE, y_{50}	10.444725	2.745175	3.805	66	<0.001
SCCHANGE, y_{60}	-12.188733	1.845310	-6.605	19	<0.001

Final model of scientific competence in Spain

Fixed effects	Coefficient	Standard error	t-ratio	g. l.	p-value
Intercept, y_{00}	421.753215	40.088739	10.520	77	<0.001
OTT1, y_{01}	-20.114711	9.332314	-2.155	970	0.031
SCHSIZE, y_{02}	0.007455	0.002636	2.828	605	0.005
STAFFSHORT, y_{03}	-3.984929	1.532136	-2.601	245	0.010
PROAT6, y_{04}	56.897067	27.248689	2.088	970	0.037
ESCS, y_{10}	11.619450	0.964503	12.047	38	<0.001
GENDER, y_{20}	-12.047715	2.067586	-5.827	39	<0.001
REPEAT, y_{30}	-55.750288	4.999875	-11.150	84	<0.001
GRADE, y_{40}	22.022909	3.886366	5.667	93	<0.001
AGE, y_{50}	6.966728	2.513184	2.772	87	0.007
LANG, y_{60}	-9.416640	3.044682	-3.093	39	0.004
SCCHANGE, y_{70}	-10.320797	1.877325	-5.498	19	<0.001

Final model of reading comprehension competence in Spain

Fixed effects	Coefficient	Standard error	t-ratio	g. l.	p-value
Intercept, y_{00}	387.063426	37.030460	10.453	357	<0.001
OTT2, y_{01}	-29.686920	12.372569	-2.399	970	0.017
SCHSIZE, y_{02}	0.007459	0.003463	2.154	970	0.031

Fixed effects	Coefficient	Standard error	t-ratio	g. l.	p-value
STAFFSHORT, y_{03}	-4.225187	1.549762	-2.726	970	0.007
LOCAT, y_{04}	3.458251	1.408653	2.455	970	0.014
ESCS, y_{10}	11.004325	0.861783	12.769	137	<0.001
GENDER, y_{20}	17.296984	1.701709	10.164	110	<0.001
REPEAT, y_{30}	-55.058085	4.418017	-12.462	660	<0.001
GRADE, y_{40}	23.023086	3.372247	6.827	1719	<0.001
AGE, y_{50}	7.870401	2.299009	3.423	302	<0.001
LANG, y_{60}	-9.408444	2.670744	-3.523	135	<0.001
SCCHANGE, y_{70}	-9.043357	1.252330	-7.221	133	<0.001

Final model mathematical competence in Ireland

Fixed effects	Coefficient	Standard error	t-ratio	g. l.	p-value
Intercept, y_{00}	209.192181	71.867914	2.911	153	0.004
N2ESCS, y_{01}	46.316745	4.180229	11.080	153	<0.001
ESCS, y_{10}	20.829723	1.754272	11.874	319	<0.001
GENDER, y_{20}	-11.475024	3.277676	-3.501	84	<0.001
DURECEC, y_{30}	-5.187889	1.543995	-3.360	151	<0.001
REPEAT, y_{40}	-36.039016	5.865523	-6.144	198	<0.001
AGE, y_{50}	19.189583	4.598107	4.173	206	<0.001
LANG, y_{60}	16.994840	6.215276	2.734	68	0.008
INM_2, y_{70}	-4.575120	0.941155	-4.861	76	<0.001
SCCHANGE, y_{80}	-6.716187	2.585708	-2.597	114	0.011

Final model of scientific competence in Ireland

Fixed effects	Coefficient	Standard error	t-ratio	g. l.	p-value
Intercept, y_{00}	271.126877	74.250670	3.652	153	<0.001
N2ESCS, y_{01}	42.978556	5.776254	7.441	153	<0.001
ESCS, y_{10}	23.644826	1.980467	11.939	210	<0.001
REPEAT, y_{20}	-36.821218	5.438212	-6.771	263	<0.001
AGE, y_{30}	14.085468	4.719389	2.985	607	0.003
INM_2, y_{40}	-5.906658	0.863670	-6.839	609	<0.001

Final model of reading comprehension competence in Ireland

Fixed effects	Coefficient	Standard error	t-ratio	g. l.	p-value
Intercept, y_{00}	256.645451	73.776135	3.479	153	<0.001
N2ESCS, y_{01}	47.865185	5.114160	9.359	153	<0.001
ESCS, y_{10}	22.500407	1.929044	11.664	3604	<0.001
GENDER, y_{20}	15.872162	2.984483	5.318	2002	<0.001
REPEAT, y_{30}	-37.556016	5.162031	-7.275	3573	<0.001
AGE, y_{40}	16.049949	4.689439	3.423	801	<0.001
INM_2, y_{50}	-4.723502	0.824796	-5.727	679	<0.001
SCCHANGE, y_{60}	-7.023743	2.422461	-2.899	2265	0.004

Final model mathematical competence in Portugal

Fixed effects	Coefficient	Standard error	t-ratio	g. l.	p-value
Intercept, y_{00}	512.764218	20.516018	24.993	114	<0.001
CLSIZE, y_{01}	1.473761	0.714451	2.063	127	0.041
N2ESCS, y_{02}	14.299145	4.374239	3.269	171	0.001
ESCS, y_{10}	9.143090	1.921933	4.757	117	<0.001
GENDER, y_{20}	-26.166064	3.873004	-6.756	106	<0.001
DURECEC, y_{30}	2.615356	1.302575	2.008	172	0.046
REPEAT, y_{40}	-60.375079	8.777136	-6.879	98	<0.001
GRADE, y_{50}	40.626861	4.721590	8.604	92	<0.001

Final model of scientific competence in Portugal

Fixed effects	Coefficient	Standard error	t-ratio	g. l.	p-value
Intercept, y_{00}	513.437145	15.614328	32.882	170	<0.001
CLSIZE, y_{01}	1.410727	0.569664	2.476	170	0.014
PROAT6, y_{02}	-25.314665	12.600581	-2.009	170	0.046
N2ESCS, y_{03}	15.148369	4.216065	3.593	170	<0.001
ESCS, y_{10}	7.360618	1.691238	4.352	270	<0.001
GENDER, y_{20}	-19.755325	2.868870	-6.886	206	<0.001
REPEAT, y_{30}	-50.110078	6.908927	-7.253	153	<0.001
GRADE, y_{40}	38.726313	3.970577	9.753	104	<0.001

Final model of reading comprehension competence in Portugal

Fixed effects	Coefficient	Standard error	<i>t</i>-ratio	g. l.	<i>p</i>-value
Intercept, y_{00}	470.883977	19.206903	24.516	171	<0.001
CLSIZE, y_{01}	2.318445	0.688131	3.369	171	<0.001
N2ESCS, y_{02}	14.522647	4.283607	3.390	171	<0.001
ESCS, y_{10}	5.094510	2.081919	2.447	943	0.015
DURECEC, y_{20}	2.989061	1.390221	2.150	462	0.032
REPEAT, y_{30}	-50.558995	7.901277	-6.399	317	<0.001
GRADE, y_{40}	43.918494	4.586319	9.576	235	<0.001

Annex III – Non-contextual variables included in Pearson's Correlation

School level:

- Creative extra-curricular activities (Sum) → CREATIV.
- Student behaviour hindering learning (WLE) → STUBEHA.
- Teacher behaviour hindering learning (WLE) → TEACHBEHA.
- School principal's view on teachers' multicultural and egalitarian beliefs (WLE) → SCMCEG.

Teacher level:

- Teacher's view on staff shortage (WLE) → TCSTAFFSHORT.
- Teacher's view on educational material shortage (WLE) → TCEDUSHORT.
- Test language teacher collaboration (WLE) → COLT.
- Exchange and co-ordination for teaching (WLE) → EXCHT.
- Teacher's satisfaction with the current job environment (WLE) → SATJOB.
- Teacher's satisfaction with teaching profession (WLE) → SATTEACH.
- Teacher's self-efficacy in classroom management (WLE) → SEFFCM.
- Teacher's self-efficacy in maintaining positive relations with students (WLE) → SEFFREL.
- Teacher's self-efficacy in instructional settings (WLE) → SEFFINS.
- Opportunity to learn (OTL) aspects of reading comprehension (WLE) → TCOTLCOMP.
- Teacher's stimulation of reading engagement (WLE) → TCSTIMREAD.
- Teacher's initiation of reading strategies (WLE) → TCSTRATREAD.
- Teacher's use of specific ICT applications (WLE) → TCICTUSE.
- Disciplinary climate in test language lessons (WLE) → TCDISCLIMA.
- Direct teacher's instruction (WLE) → TCDIRINS.
- Feedback provided by the teachers (WLE) → FEEDBACK.
- Student assessment/use (adaption of instruction) (WLE) → ADAPTINSTR.
- Feedback provided by the teachers (WLE) → FEEDBINSTR.
- Teacher's attitudes towards immigrants (WLE) → TCATTIMM.
- Teacher's training on global competence (WLE) → GCTRAIN.
- Teachers' multicultural and egalitarian beliefs (WLE) → TCMCEG.
- Teacher's self-efficacy in multicultural environments (WLE) → GCSELF.

Student level:

- Meta-cognition: understanding and remembering → UNDREM.
- Meta-cognition: summarising → METASUM.
- Meta-cognition: assess credibility → METASPAM.
- Disciplinary climate in test language lessons (WLE) → DISCLIMA.
- Teacher support in test language lessons (WLE) → TEACHSUP.
- Teacher-directed instruction (WLE) → DIRINS.
- Perceived feedback (WLE) → PERFEED.
- Parents' emotional support perceived by student (WLE) → EMOSUPS.
- Teacher's stimulation of reading engagement perceived by student (WLE) → STIMREAD.
- Adaptation of instruction (WLE) → ADAPTIVITY.
- Perceived teacher's interest (WLE) → TEACHINT.
- Joy/Like reading (WLE) → JOYREAD.
- Self-concept of reading: Perception of competence (WLE) → SCREADCOMP.
- Self-concept of reading: Perception of difficulty (WLE) → SCREADDIFF.
- Perception of competitiveness at school (WLE) → PERCOMP.
- Perception of cooperation at school (WLE) → PERCOOP.
- Attitude towards school: learning activities (WLE) → ATTLNACT.
- Competitiveness (WLE) → COMPETE.
- Work mastery (WLE) → WORKMAST.
- General fear of failure (WLE) → GFOFAIL.
- Eudaemonia: meaning in life (WLE) → EUDMO.
- Subjective well-being: Positive affect (WLE) → SWBP.
- Resilience (WLE) → RESILIENCE.
- Mastery goal orientation (WLE) → MASTGOAL.
- Student's attitudes towards immigrants (WLE) → ATTIMM.
- Student's interest in learning about other cultures (WLE) → INTCULT.
- Perspective-taking (WLE) → PERSPECT.
- Cognitive flexibility/adaptability (WLE) → COGFLEX.
- Respect for people from other cultures (WLE) → RESPECT.
- Awareness of intercultural communication (WLE) → AWACOM.
- Global-mindedness (WLE) → GLOBMIND.

- Discriminating school climate (WLE) → DISCRIM.
- Subjective well-being: Sense of belonging to school (WLE) → BELONG.
- Student's experience of being bullied (WLE) → BEINGBULLIED.
- ICT use outside of school (leisure) (WLE) → ENTUSE.
- Use of ICT outside of school (for schoolwork activities) (WLE) → HOMESCH.
- Use of ICT at school in general (WLE) → USESCH.
- Interest in ICT (WLE) → INTICT.
- Perceived ICT competence (WLE) → COMPICT.
- Perceived autonomy related to ICT use (WLE) → AUTICT.
- ICT as a topic in social interaction (WLE) → SOIAICT.
- Subject-related ICT use during lessons (WLE) → ICTCLASS.
- Subject-related ICT use outside of lessons (WLE) → ICTOUTSIDE.
- Information about careers (WLE) → INFOCAR.
- Information about the labour market provided by the school (WLE) → INFOJOB1.
- Information about the labour market provided outside of school (WLE) → INFOJOB2.
- Parents' emotional support (WLE) → EMOSUPP.
- Parents' perceived school quality (WLE) → PQSCHOOL.
- School policies for parental involvement (WLE) → PASCHPOL.
- Previous parental support for learning at home (WLE) → PRESUPP.

Anexo 2 – Variables procesuales incluidas en la Correlación de Pearson

A nivel de centro, se busca la relación entre la eficacia escolar y:

- Creative extra-curricular activities (Sum) → CREATIV.
- Student behaviour hindering learning (WLE) → STUBEHA.
- Teacher behaviour hindering learning (WLE) → TEACHBEHA.
- School principal's view on teachers' multicultural and egalitarian beliefs (WLE) → SCMCEG.

A nivel de profesor, se incluyen las siguientes variables:

- Teacher's view on staff shortage (WLE) → TCSTAFFSHORT.
- Teacher's view on educational material shortage (WLE) → TCEDUSHORT.
- Test language teacher collaboration (WLE) → COLT.
- Exchange and co-ordination for teaching (WLE) → EXCHT.
- Teacher's satisfaction with the current job environment (WLE) → SATJOB.
- Teacher's satisfaction with teaching profession (WLE) → SATTEACH.
- Teacher's self-efficacy in classroom management (WLE) → SEFFCM.
- Teacher's self-efficacy in maintaining positive relations with students (WLE) → SEFFREL.
- Teacher's self-efficacy in instructional settings (WLE) → SEFFINS.
- Opportunity to learn (OTL) aspects of reading comprehension (WLE) → TCOTLCOMP.
- Teacher's stimulation of reading engagement (WLE) → TCSTIMREAD.
- Teacher's initiation of reading strategies (WLE) → TCSTRATREAD.
- Teacher's use of specific ICT applications (WLE) → TCICTUSE.
- Disciplinary climate in test language lessons (WLE) → TCDISCLIMA.
- Direct teacher's instruction (WLE) → TCDIRINS.
- Feedback provided by the teachers (WLE) → FEEDBACK.
- Student assessment/use (adaption of instruction) (WLE) → ADAPTINSTR.
- Feedback provided by the teachers (WLE) → FEEDBINSTR.
- Teacher's attitudes towards immigrants (WLE) → TCATTIMM.
- Teacher's training on global competence (WLE) → GCTRAIN.
- Teachers' multicultural and egalitarian beliefs (WLE) → TCMCEG.
- Teacher's self-efficacy in multicultural environments (WLE) → GCSELF.

A nivel de estudiantes, se busca hallar la relación entre eficacia escolar y las siguientes variables:

- Meta-cognition: understanding and remembering → UNDREM.
- Meta-cognition: summarising → METASUM.
- Meta-cognition: assess credibility → METASPAM.
- Disciplinary climate in test language lessons (WLE) → DISCLIMA.
- Teacher support in test language lessons (WLE) → TEACHSUP.
- Teacher-directed instruction (WLE) → DIRINS.
- Perceived feedback (WLE) → PERFEED.
- Parents' emotional support perceived by student (WLE) → EMOSUPS.
- Teacher's stimulation of reading engagement perceived by student (WLE) → STIMREAD.
- Adaptation of instruction (WLE) → ADAPTIVITY.
- Perceived teacher's interest (WLE) → TEACHINT.
- Joy/Like reading (WLE) → JOYREAD.
- Self-concept of reading: Perception of competence (WLE) → SCREADCOMP.
- Self-concept of reading: Perception of difficulty (WLE) → SCREADDIFF.
- Perception of competitiveness at school (WLE) → PERCOMP.
- Perception of cooperation at school (WLE) → PERCOOP.
- Attitude towards school: learning activities (WLE) → ATTLNACT.
- Competitiveness (WLE) → COMPETE.
- Work mastery (WLE) → WORKMAST.
- General fear of failure (WLE) → GFOFAIL.
- Eudaemonia: meaning in life (WLE) → EUDMO.
- Subjective well-being: Positive affect (WLE) → SWBP.
- Resilience (WLE) → RESILIENCE.
- Mastery goal orientation (WLE) → MASTGOAL.
- Student's attitudes towards immigrants (WLE) → ATTIMM.
- Student's interest in learning about other cultures (WLE) → INTCULT.
- Perspective-taking (WLE) → PERSPECT.
- Cognitive flexibility/adaptability (WLE) → COGFLEX.
- Respect for people from other cultures (WLE) → RESPECT.
- Awareness of intercultural communication (WLE) → AWACOM.

- Global-mindedness (WLE) → GLOBMIND.
- Discriminating school climate (WLE) → DISCRIM.
- Subjective well-being: Sense of belonging to school (WLE) → BELONG.
- Student's experience of being bullied (WLE) → BEINGBULLIED.
- ICT use outside of school (leisure) (WLE) → ENTUSE.
- Use of ICT outside of school (for schoolwork activities) (WLE) → HOMESCH.
- Use of ICT at school in general (WLE) → USESCH.
- Interest in ICT (WLE) → INTICT.
- Perceived ICT competence (WLE) → COMPICT.
- Perceived autonomy related to ICT use (WLE) → AUTICT.
- ICT as a topic in social interaction (WLE) → SOIAICT.
- Subject-related ICT use during lessons (WLE) → ICTCLASS.
- Subject-related ICT use outside of lessons (WLE) → ICTOUTSIDE.
- Information about careers (WLE) → INFOCAR.
- Information about the labour market provided by the school (WLE) → INFOJOB1.
- Information about the labour market provided outside of school (WLE) → INFOJOB2.
- Parents' emotional support (WLE) → EMOSUPP.
- Parents' perceived school quality (WLE) → PQSCHOOL.
- School policies for parental involvement (WLE) → PASCHPOL.
- Previous parental support for learning at home (WLE) → PRESUPP.

Anexo 3 – Documento enviado a los jueces para la validación del cuestionario



**PROPUESTA DE
INSTRUMENTO PARA LA
EVALUACIÓN DE LA EFICACIA
ESCOLAR AUTO PERCIBIDA
POR EL PROFESORADO EN
ACTIVO EN LAS ESCUELAS DE
ALTA Y BAJA EFICACIA
ESCOLAR**

Proyecto I+D+i Escuelas eficaces para la mejora del sistema con referencia PGC2018-099174-B-I00



**VNiVERSIDAD
D SALAMANCA**

Dentro del Proyecto I+D+i “Escuelas eficaces para la mejora del sistema (EFI-6), financiado por el Ministerio de Ciencia e Innovación del Gobierno de España con referencia PGC2018-099174-B-I00, planteamos como objetivo específico el diseño, adaptación y selección de los instrumentos necesarios para realizar la recogida de información, destinados a estudiantes, familias, profesorado y equipos directivos.

Por lo tanto, con el fin de lograr este objetivo, presentamos aquí un cuestionario para la validación, como experto/a, de los ítems propuestos para la evaluación de la eficacia escolar auto percibida por el profesorado en activo en las escuelas de alta y baja eficacia escolar. Se solicita su colaboración para determinar la calidad del cuestionario propuesto, así como de sus dimensiones.

El concepto básico sobre el que pretendemos conseguir información con este cuestionario es el de eficacia escolar, entendida como las buenas prácticas que llevan a cabo las escuelas para conseguir un rendimiento académico superior al esperado en función de sus características contextuales, es decir, el valor añadido que aporta una escuela una vez que se han controlado sus variables contextuales.

Para la evaluación de la calidad del cuestionario (de sus ítems) se han establecido tres criterios: coherencia, relevancia y claridad.

1. **Coherencia:** el ítem presenta relación lógica y apreciable con la dimensión que mide.
 1. El ítem no tiene relación lógica con la dimensión.
 2. El ítem tiene una relación tangencial con la dimensión.
 3. El ítem tiene una relación moderada con la dimensión.
 4. El ítem se encuentra completamente relacionado con la dimensión.
2. **Relevancia:** El ítem es esencial para la dimensión, debe estar presente para medirla.
 1. El ítem puede ser eliminado sin que afecte a la medida de la dimensión.
 2. El ítem tiene cierta relevancia, pero otro ítem está incluyendo en su medida a este.
 3. El ítem es relativamente importante.
 4. El ítem es muy relevante y debe ser incluido.
3. **Claridad:**
 1. El ítem no es claro.
 2. El ítem requiere bastantes modificaciones en el uso de palabras de acuerdo con su significado o a la ordenación de las mismas.

3. Se requiere una modificación muy específica de algunos de los términos del ítem.
4. El ítem se encuentra completamente relacionado con la dimensión que mide.

Tras la valoración de los ítems, se incorpora un cuadro donde se le solicita información sobre la puntuación otorgada a los mismos y si desea indicar algún ítem alternativo.

A continuación, planteamos una evaluación de las dimensiones que componen el cuestionario, para lo cual se valorará la suficiencia de los ítems y la representatividad en el conjunto del cuestionario.

1. Suficiencia de los ítems:

1. Los ítems son suficientes para medir la dimensión.
2. Se debe incrementar el número de ítems para evaluar correctamente la dimensión.
3. Los ítems miden correctamente la dimensión.
4. Los ítems son excesivos para evaluar la dimensión.

2. Relevancia de la dimensión para el cuestionario:

1. La dimensión puede ser eliminada sin que afecte al cuestionario.
2. La dimensión tiene cierta relevancia, pero otra incluye en su medida a ésta.
3. La dimensión es relativamente importante.
4. La dimensión es muy relevante y debe ser incluida.

Las dimensiones que planteamos se definen de la siguiente manera:

1. Proyectos, planes y formación → Encontrar si el centro mantiene programas propios, que es activo en formación y en otros programas...
2. Metodologías y materiales de enseñanza → Se quiere comprobar el tipo de metodología empleada en el centro, así como el tipo de recursos que se emplean en el centro (más tradicionales, más innovadores...).
3. Atención a la diversidad y seguimiento de los estudiantes → Comprobar si se aplican medidas de atención a la diversidad en las aulas y adaptaciones curriculares en línea con el plan de orientación del centro.
4. Evaluación del alumnado y gestión del tiempo → Pretendemos conocer el planteamiento y compromiso de los profesores en aspectos referidos a la evaluación de competencias que se realiza en el centro, así como cómo se lleva a cabo la gestión del tiempo en el centro.

5. Liderazgo y gestión del centro → Conocer el tipo de liderazgo del centro y la implicación del profesorado en su gestión.
6. Coordinación, implicación, clima y convivencia → Conocer las actuaciones planteadas desde el centro y la implicación del profesorado con la organización de estas, así como todo lo referido al buen funcionamiento (o dinámica del centro).
7. Evaluación de los docentes, del centro y de programas → Comprender si en el centro se entiende la evaluación con un carácter formativo y orientado a la mejora.
8. Infraestructura y proyección del centro en el entorno → Percepción del profesorado sobre la relación entre los recursos educativos disponibles en el centro y el rendimiento.

Al final del cuestionario se adjunta un apartado por si quisiera realizar cualquier recomendación o comentario al respecto. Muchas gracias por su participación en este proceso de validación.

Coherencia				Relevancia				Claridad			
1	2	3	4	1	2	3	4	1	2	3	4

DIMENSIÓN 1: PROYECTOS, PLANES Y FORMACIÓN

1. Este centro educativo diseña y desarrolla habitualmente proyectos educativos de interés para toda la comunidad educativa.												
2. Este centro presta especial atención a la formación permanente y al desarrollo profesional del profesorado para la mejora del centro.												
3. Las actividades de formación docente ofertadas en el centro tienen en cuenta las necesidades formativas expresadas por el profesorado.												
4. Los proyectos de innovación educativa iniciados en el centro suelen cumplir con la duración y los objetivos establecidos en su comienzo.												

Comentarios sobre la Dimensión 1:

DIMENSIÓN 2: METODOLOGÍAS Y MATERIALES DE ENSEÑANZA

1. La mayoría de los profesores del centro utilizamos una metodología didáctica basada en la clase magistral y el uso del libro de texto.												
2. El centro dedica especial atención a trabajar la competencia matemática y/o científica.												
3. La mayoría de los profesores utilizamos habitualmente tecnologías digitales para apoyar el aprendizaje de los estudiantes.												
4. La comprensión escrita y expresión oral son las competencias principales a las que atendemos los profesores de este centro.												

Comentarios sobre la Dimensión 2:

DIMENSIÓN 3: ATENCIÓN A LA DIVERSIDAD Y SEGUIMIENTO DE LOS ESTUDIANTES

1. El claustro del centro es sensible y aplica medidas de refuerzo y apoyo a los alumnos con necesidades educativas especiales												
2. En general, el profesorado utiliza metodologías diferenciales para atender a la diversidad en las aulas.												
3. El plan de orientación del centro se elabora conjuntamente entre los distintos implicados (orientador, profesores y familias).												

Comentarios sobre la Dimensión 3:

DIMENSIÓN 4: EVALUACIÓN DEL ALUMNADO Y GESTIÓN DEL TIEMPO

1. En el centro se presta especial atención a la evaluación de los alumnos como proceso educativo, más que sumativo.												
2. En general, los profesores compartimos criterios comunes en el control del rendimiento de los alumnos en un mismo grupo.												
3. El centro participa habitualmente en Evaluaciones Externas (PISA, de nivel, etc.)												
4. En el centro se aplican sistemas de garantía de calidad, reconocidos externamente.												
5. En este centro se insiste en la puntualidad y en el respeto al horario.												

Comentarios sobre la Dimensión 4:

DIMENSIÓN 5: LIDERAZGO Y GESTIÓN DEL CENTRO

1. El modo de actuar de este equipo directivo es más democrático que jerárquico.												
2. Considero que el tipo de liderazgo del equipo directivo tiene una gran influencia en el clima del centro.												
3. Nuestro equipo directivo ha planteado unas metas claras y conoce el funcionamiento del centro.												
4. Generalmente, en este centro el equipo apoya al profesorado implicándole en las decisiones de organización.												
5. En este centro, en general, la plantilla de profesorado es estable.												
6. La organización del centro permite dedicar tiempo a la formación permanente.												

Comentarios sobre la Dimensión 5:

DIMENSIÓN 6: COORDINACIÓN, IMPLICACIÓN, CLIMA Y CONVIVENCIA

1. Es un centro caracterizado por un buen clima de trabajo.												
2. El centro dispone de un plan de convivencia que se aplica y funciona eficazmente.												
3. Habitualmente, no existen conflictos graves de relación entre el personal del centro.												
4. Las reuniones de coordinación son planificadas con anterioridad, planteando los objetivos a conseguir en cada caso.												
5. Como profesor, me identifico con y creo en el proyecto del centro.												
6. La dirección del centro fomenta la participación, la implicación y el compromiso del profesorado y del resto de la comunidad educativa en la gestión total del centro.												

Comentarios sobre la Dimensión 6:

DIMENSIÓN 7: EVALUACIÓN DE LOS DOCENTES, DEL CENTRO Y DE PROGRAMAS

1. En este centro, habitualmente, compartimos experiencias sobre la mejora de la práctica docente en el aula												
2. Además, los profesores hacemos el seguimiento de los progresos y los cambios que introducimos												
3. El equipo de docentes, a partir de los resultados de las evaluaciones de los estudiantes y otras evidencias, elaboramos planes de mejora.												

Comentarios sobre la Dimensión 7:

DIMENSIÓN 8: INFRAESTRUCTURA Y PROYECCIÓN DEL CENTRO EN EL ENTORNO

1. Las instalaciones e infraestructuras del centro posibilitan un buen desarrollo de la labor educativa.												
2. La imagen del centro ante la comunidad próxima es óptima.												
3. Como profesor, percibo que se necesitaría una mayor implicación de las familias del alumnado para conseguir una mejora del centro educativo.												
4. En este centro se ha establecido una comunicación eficaz con toda la comunidad educativa.												

Comentarios sobre la Dimensión 8:

Para finalizar:

1. Como profesor, considero que las condiciones en las que este centro desarrolla su proyecto educativo tienen una influencia significativa en la mejora del aprendizaje de todo su alumnado, contribuyendo así a la equidad educativa.												
2. A modo de RESUMEN u OBSERVACIONES: Por favor, exponga en un breve texto, cómo cree que es considerado su centro por la comunidad en la que está asentado y por qué razones.												

A continuación, deberá proceder a la valoración de las dimensiones en su conjunto:

<u>DIMENSIÓN</u>	Suficiencia				Relevancia para el cuestionario			
	1	2	3	4	1	2	3	4
1. Proyectos, planes y formación.								
2. Metodologías y materiales de enseñanza.								
3. Atención a la diversidad y seguimiento de los estudiantes.								
4. Evaluación del alumnado y gestión del tiempo.								
5. Liderazgo y gestión del centro.								
6. Coordinación, implicación, clima y convivencia.								
7. Evaluación de los docentes, del centro de programas.								
8. Infraestructura y proyección del centro en el entorno.								
Valoración global de las dimensiones y comentarios adicionales del/de la evaluador/a:								

¡Muchas gracias por su colaboración!

Claridad de los ítems

EVALUADOR	CLARIDAD																																					
	D1				D2				D3			D4					D5					D6					D7			D8				OTROS				
	1.1	1.2	1.3	1.4	2.1	2.2	2.3	2.4	3.1	3.2	3.3	4.1	4.2	4.3	4.4	4.5	5.1	5.2	5.3	5.4	5.5	5.6	6.1	6.2	6.3	6.4	6.5	6.6	7.1	7.2	7.3	8.1	8.2	8.3	8.4	O1	O2	
1	4	2	4	2	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	
2	4	4	4	2	2	4	3	4	4	4	4	2	4	4	2	4	2	1	2	3	4	2	2	4	4	3	4	2	4	3	4	4	2	4	4	1	4	4
3	4	4	4	4	4	4	4	2	4	-	3	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	2	4	4	4	4	4	4	4	4	4	
4	3	4	4	3	4	4	4	4	3	3	3	3	4	4	4	4	4	4	3	4	4	4	4	4	3	3	4	3	3	4	4	4	4	4	4	4	4	
5	4	4	4	4	4	2	4	3	4	3	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	
6	4	4	4	3	3	3	4	3	4	4	3	3	4	3	3	2	3	4	4	4	3	1	4	4	4	3	3	4	4	4	4	4	4	4	4	1	3	
7	4	3	4	3	4	-	-	-	4	4	4	3	4	4	3	4	4	4	4	4	4	4	4	3	4	4	4	4	3	4	4	4	4	4	4	4	4	
8	3	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	
9	3	3	4	1	1	4	3	4	4	3	4	1	4	4	4	3	3	4	4	4	4	4	4	1	3	3	2	4	2	3	4	2	3	3	3	3		
10	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	
11	4	1	4	4	4	4	4	4	2	4	4	2	2	4	4	4	4	4	4	4	4	4	2	4	4	4	4	4	4	4	4	4	4	4	4	4	4	
12	3	4	4	4	3	4	3	4	4	4	4	4	4	4	4	4	4	4	2	4	4	4	4	3	4	4	3	3	4	4	4	4	4	4	4	4	4	
13	4	4	4	3	2	4	2	4	4	3	2	4	4	3	3	4	3	4	4	3	3	4	4	3	1	2	-	3	4	3	4	4	4	2	3	4	2	3
14	4	4	4	4	3	3	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	
15	4	4	4	3	1	1	1	1	-	-	3	4	-	-	-	1	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	2	4	4	4	4

Valoración de las dimensiones en su conjunto

EVALUADOR	SUFICIENCIA								RELEVANCIA CUESTIONARIO							
	D1	D2	D3	D4	D5	D6	D7	D8	D1	D2	D3	D4	D5	D6	D7	D8
1	3	3	3	3	3	3	3	2	4	4	4	4	4	4	4	4
2	4	1	3	2	4	3	3	3	4	4	4	4	4	4	4	4
3	4	4	4	4	4	4	3	4	4	4	4	4	4	4	3	4
4	4	3	3	3	3	3	2	4	4	4	4	4	4	4	4	4
5	3	3	3	3	3	3	3	3	4	4	4	4	4	4	4	4
6	4	4	4	3	4	4	4	3	4	4	4	4	4	4	4	3
7	3	3	3	3	4	3	4	4	4	4	4	4	4	4	4	4
8	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
9	4	4	3	3	4	4	4	3	4	4	4	4	4	4	4	3
10	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
11	2	3	2	-	3	2	3	3	4	4	4	4	4	4	4	4
12	4	4	3	4	4	4	3	4	4	4	4	3	4	4	4	4
13	4	3	3	3	4	4	3	4	4	4	4	4	4	4	4	4
14	4	4	4	4	4	4	4	4	4	4	4	4	4	3	4	3
15	2	2	1	1	3	3	2	2	4	4	4	4	4	4	4	4

Anexo 5 – Versión final del cuestionario

Encuesta de opinión del Profesor de Educación Secundaria sobre condiciones de contexto de un centro educativo que favorecen el aprendizaje de los estudiantes

En el proyecto I+D+i “Escuelas eficaces para la mejora del sistema” financiado por el Ministerio de Ciencia e Innovación del Gobierno de España en la Convocatoria del año 2018 y con referencia PGC2018-099174-B-I00 (dirigido por las profesoras de la Universidad de Salamanca, María José Rodríguez Conde y Susana Olmos Migueláñez), nos planteamos la elaboración de un catálogo de buenas prácticas educativas, a partir de un estudio sobre factores asociados a la eficacia y a la mejora de centros escolares. En esta fase del proyecto, necesitamos recoger la opinión directa del profesorado en activo que valore las posibles dimensiones del centro, en cuanto a su importancia como factor condicionador del mejor aprendizaje posible en sus estudiantes.

Se trata de una encuesta anónima y cerrada, que no llevaría mucho tiempo responder. Agradecemos tu participación y si tienes cualquier duda o quisieras obtener más información sobre el proyecto o sus resultados, no dudes en ponerte en contacto con la siguiente investigadora, responsable de esta fase, en el Instituto Universitario de Ciencias de la Educación (IUCE) de la Universidad de Salamanca: Cristina Frade, cristina.frade@usal.es

*** Indica que la pregunta es obligatoria**

IDENTIFICACIÓN DEL CENTRO EDUCATIVO

Localización del centro educativo (Marca solo un óvalo)

- Zona rural
- Zona urbana

Número de ESTUDIANTES aproximado en el CENTRO EDUCATIVO

Número de PROFESORES aproximado en el CENTRO EDUCATIVO

Nivel socioeconómico aproximado medio de las familias del alumnado que acude al CENTRO EDUCATIVO (Marca solo un óvalo)

- Alto
- Medio
- Bajo

IDENTIFICACIÓN DEL PROFESOR

Género (Marca solo un óvalo)

- Femenino
- Masculino
- Prefiero no decirlo

Edad (Marca solo un óvalo)

- Menos de 30 años
- De 31 a 40 años
- De 41 a 50 años
- De 51 a 60 años
- Más de 60 años

Categoría profesional actual (Marca solo un óvalo)

- Profesor Interino
- Funcionario

¿Cuánto tiempo llevas impartiendo clase en este centro?

¿Cuánto tiempo llevas impartiendo clase, en general?

Departamento o materia de Educación Secundaria en el que más docencia impartes:

Nivel más alto de estudios que posees (equivalencia en el sistema actual) (Marca solo un óvalo)

- Grado
- Máster
- Doctorado

LISTADO DE CONDICIONES DEL CENTRO EDUCATIVO QUE FAVORECEN EL APRENDIZAJE DE LOS ESTUDIANTES

* A continuación, se presentan una serie de ítems relacionados con la eficacia del centro educativo para lograr un mejor aprendizaje. Debe valorar de 1 a 5 su grado de acuerdo o desacuerdo con las siguientes afirmaciones; siendo 1 = totalmente en desacuerdo, 2 = en desacuerdo, 3 = Ni en desacuerdo, ni de acuerdo, 4 = de acuerdo y 5 = totalmente de acuerdo (Marca solo un óvalo por fila).

	1	2	3	4	5
El plan de actividades del departamento de orientación se elabora conjuntamente con el equipo directivo y el profesorado.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Como profesor me identifico con la filosofía del centro educativo.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Los profesores compartimos criterios comunes en la evaluación de los alumnos de un mismo grupo.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
El profesorado aplica el DUA (Diseño Universal de Aprendizaje) para aplicar metodologías que faciliten el proceso de enseñanza-aprendizaje de todo el alumnado.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Los proyectos de innovación educativa iniciados en el centro cumplen con los objetivos propuestos.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
El centro utiliza recursos de su entorno (instalaciones deportivas, bibliotecas, etc.).	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
En el centro se presta especial atención a la evaluación de los alumnos como proceso formativo.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
La comunidad educativa del centro está implicada.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
En el centro se aplican sistemas de garantía de calidad, reconocidos externamente.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

	1	2	3	4	5
Este centro presta especial atención a la formación permanente y al desarrollo profesional del profesorado.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Los profesores empleamos metodologías activas (Aprendizaje por proyectos, Aprendizaje basado en problemas, clase invertida, etc.) en nuestras clases.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
En este centro se diseñan proyectos educativos que implican a toda la comunidad educativa.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
En este centro se implica a los estudiantes en la evaluación de su propio aprendizaje (coevaluación, evaluación por pares, autoevaluación, etc.).	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Los profesores utilizamos las tic para apoyar el aprendizaje de los estudiantes.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Los profesores del centro empleamos el libro de texto (analógico o digital) en el desarrollo de nuestras clases.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Las actividades de formación docente ofertadas en el centro tienen en cuenta las necesidades formativas expresadas por el profesorado	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Los profesores del centro utilizamos la clase magistral como metodología didáctica.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
En este centro evaluamos nuestra práctica docente.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
El equipo directivo contempla tiempos para la formación permanente dentro de la jornada laboral.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
El equipo directivo de este centro fomenta la implicación del profesorado en la toma de decisiones en torno a cuestiones organizativas.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
El profesorado utiliza metodologías diversas para atender a las necesidades de los estudiantes en las aulas.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
El modo de actuar de este equipo directivo es más democrático que autoritario.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

	1	2	3	4	5
El centro participa en actividades culturales y de desarrollo de su barrio y/o su localidad.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
No existen conflictos graves de relación entre los miembros de la comunidad educativa del centro.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
El centro dispone de un plan de convivencia que se aplica y funciona eficazmente.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
El claustro del centro aplica medidas de refuerzo y apoyo a todo el alumnado que lo necesite.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
El centro se caracteriza por un buen clima de trabajo entre todos los miembros de la comunidad educativa (profesorado, PAS, alumnado, etc.).	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Considero que el tipo de liderazgo que ejerce el equipo directivo de nuestro centro tiene una gran influencia en el clima del centro.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
El equipo directivo ha definido unos objetivos claros.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
En este centro se ha establecido una comunicación eficaz con toda la comunidad educativa.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Este centro utiliza los resultados de la evaluación para elaborar planes de mejora.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
En este centro se desarrollan proyectos educativos de interés para toda la comunidad educativa.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Como profesor, percibo que se necesitaría una mayor implicación de las familias del alumnado para conseguir una mejora del centro educativo.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Nuestro equipo directivo conoce el funcionamiento del centro.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Como profesor, considero que las condiciones en las que este centro desarrolla su proyecto educativo tienen una influencia significativa en la mejora del aprendizaje de todo su alumnado, contribuyendo así a la equidad educativa.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Para finalizar...

A modo de RESUMEN y OBSERVACIONES: por favor, exponga en un breve texto cómo cree que es considerado su centro por la comunidad en la que está asentado y por qué razones.
