



Universidad Autónoma de Baja California

Instituto de Investigación y Desarrollo Educativo

“Factores de enseñanza que explican la Eficacia Escolar en México: Un estudio a partir del análisis multinivel de PISA 2018”

TESIS

Que para obtener el grado de

MAESTRO EN CIENCIAS EDUCATIVAS

Presenta

César Gómez Monarrez

Ensenada, B. C., México, 01 de diciembre de 2020



Universidad Autónoma de Baja California
Instituto de Investigación y Desarrollo Educativo
Maestría en Ciencias Educativas



“Factores de enseñanza que explican la Eficacia Escolar en México: Un estudio a partir del análisis multinivel de PISA 2018”

TESIS

Que para obtener el grado de

MAESTRO EN CIENCIAS EDUCATIVAS

Presenta

César Gómez Monarrez

APROBADO POR:

Dra. Alicia Aleli Chaparro Caso López
Directora de tesis

Dr. Luis Horacio Pedroza Zúñiga
Sinodal

Dra. Karla María Díaz López
Sinodal





Ensenada, B.C., a 09 de noviembre de 2020

ASUNTO: Voto aprobatorio sobre trabajo de tesis de grado de Maestría.

“Dr. Sergio Gerardo Malaga Villegas”
Coordinador de la Maestría en Ciencias Educativas
Presente.

Después de haber efectuado una revisión minuciosa sobre el trabajo de tesis presentado por el **C. CÉSAR GÓMEZ MONARREZ** para poder presentar la defensa de su examen y obtener el grado de Maestría en Ciencias Educativas, me permito comunicarle que he dado mi VOTO APROBATORIO, sobre su trabajo intitulado:

“FACTORES DE ENSEÑANZA QUE EXPLICAN LA EFICACIA ESCOLAR EN MÉXICO: UN ESTUDIO A PARTIR DEL ANÁLISIS MULTINIVEL DE PISA 2018”.

Esperando reciba el presente de conformidad, quedo de Usted.

Atentamente



DRA. ALICIA ALELÍ CHAPARRO CASO LÓPEZ



Ensenada, B.C., a 09 de noviembre de 2020

ASUNTO: Voto aprobatorio sobre trabajo de tesis de grado de Maestría.

“Dr. Sergio Gerardo Malaga Villegas”
Coordinador de la Maestría en Ciencias Educativas
Presente.

Después de haber efectuado una revisión minuciosa sobre el trabajo de tesis presentado por el **C. CÉSAR GÓMEZ MONARREZ** para poder presentar la defensa de su examen y obtener el grado de Maestría en Ciencias Educativas, me permito comunicarle que he dado mi VOTO APROBATORIO, sobre su trabajo intitulado:

“FACTORES DE ENSEÑANZA QUE EXPLICAN LA EFICACIA ESCOLAR EN MÉXICO: UN ESTUDIO A PARTIR DEL ANÁLISIS MULTINIVEL DE PISA 2018”.

Esperando reciba el presente de conformidad, quedo de Usted.

Atentamente

A handwritten signature in black ink, appearing to read "L. H. P. Z.", written over a horizontal line.

DR. LUIS HORACIO PEDROZA ZÚÑIGA



Ensenada, B.C., a 09 de noviembre de 2020

ASUNTO: Voto aprobatorio sobre trabajo de tesis de grado de Maestría.

“Dr. Sergio Gerardo Malaga Villegas”
Coordinador de la Maestría en Ciencias Educativas
Presente.

Después de haber efectuado una revisión minuciosa sobre el trabajo de tesis presentado por el **C. CÉSAR GÓMEZ MONARREZ** para poder presentar la defensa de su examen y obtener el grado de Maestría en Ciencias Educativas, me permito comunicarle que he dado mi VOTO APROBATORIO, sobre su trabajo intitulado:

“FACTORES DE ENSEÑANZA QUE EXPLICAN LA EFICACIA ESCOLAR EN MÉXICO: UN ESTUDIO A PARTIR DEL ANÁLISIS MULTINIVEL DE PISA 2018”.

Esperando reciba el presente de conformidad, quedo de Usted.

Atentamente

DRA. KARLA MARÍA DÍAZ LÓPEZ

Dedicatória

*A mi padre César Gómez Acosta, quien a es mi ejemplo de vida y me ha enseñado con lo
qué significa ser un buen hombre.*

*A mi madre Imelda Concepción Monarrez Paredes, quien siempre me ha brindado
todo su amor y apoyo incondicional.*

*A mi hermana Margarita Gómez Monarrez, pues por ella decidí
dedicarme a la enseñanza.*

*A mi hermana Liseth Rivas Monarrez, por siempre
cuidar, escuchar y apoyar a su hermano menor.*

Agradecimientos

A Alicia Alelí Chaparro Caso López, por confiar en mí y brindarme todo su apoyo en cada paso del camino; sin duda, la mejor mamá académica que pude haber tenido.

A Fernando Martínez Abad, por todas sus enseñanzas y mostrarme que no soy el único loco que disfruta y se divierte con la estadística.

A Luis Horacio Pedroza Zúñiga, por ser gran amigo y siempre tener un buen consejo bajo la manga.

A Karla María Díaz López, por impulsarme a continuar formándome como investigador.

A Adriana Gamazo, por su paciencia y guía durante mi estancia en Salamanca, España.

A Iván Contreras, por su siempre escucharme en mis momentos de desesperación.

A Estrella Velasco, por ser nuestra tía académica y velar por todos nosotros.

A Sharon Solis del Moral, por todo su cariño y apoyo en todo momento.

Al equipo administrativo del IIDE por todo su apoyo.

A los docentes que me formaron en la maestría.

A la UABC por ser mi alma mater.

Al CONACyT por el financiamiento.

Tabla de Contenido

Resumen.....	1
Introducción	2
1 Planteamiento del Problema	4
1.1 Objetivos de Investigación.....	9
1.1.1 Objetivo general.....	9
1.1.2 Objetivos específicos	9
1.2 Justificación.....	10
1.2.1 Relevancia social	10
1.2.2 Relevancia teórica.....	11
1.2.3 Relevancia metodológica	12
2 Revisión de Literatura	14
2.1 Programa para la Evaluación Internacional de Alumnos (PISA).....	14
2.1.1 Fundamentos	14
2.1.2 Prueba	16
2.1.3 Cuestionarios de contexto	19
2.1.4 Limitaciones de PISA	21
2.2 Revisión de los Cuestionarios de Contexto.....	26
2.2.1 Resultados de la revisión de los cuestionarios de contexto	26
2.3 Eficacia Educativa.....	32
2.3.1 Eficacia escolar	41
2.3.2 Enseñanza eficaz.....	49
2.4 Revisión Sistemática de Literatura de las Variables de Enseñanza Asociadas a los Resultados de PISA.....	56
2.4.1 Metodología del mapeo y SRL	56

2.4.2	Extracción de datos del mapeo y SRL	60
2.4.3	Informe del mapeo y SRL.....	62
3	Método.....	78
3.1	Hipótesis.....	78
3.2	Diseño.....	78
3.3	Participantes	79
3.4	Instrumentos	80
3.5	Variables del Estudio	81
3.5.1	Limpieza de datos	81
3.5.2	Variables elegidas	82
3.6	Análisis de Datos.....	89
3.6.1	Fase 1: Estimación de residuos	90
3.6.2	Fase 2: Clasificación de la eficacia escolar.....	92
3.6.3	Fase 3: Análisis del peso explicativo	94
4	Resultados.....	97
4.1	Fase 1: Estimación de Residuos	97
4.1.1	Diagnósticos de colinealidad	97
4.1.2	Modelos multinivel en competencia lectora	99
4.1.3	Modelos multinivel en competencia matemática.....	104
4.1.4	Modelos multinivel en competencia científica	109
4.2	Fase 2: Clasificación de la Eficacia Escolar.....	115
4.2.1	Transformación de los residuos escolares.....	115
4.2.2	Conglomeración.....	124
4.2.3	Caracterización de las agrupaciones	126
4.2.4	Evaluación de la calidad de las agrupaciones.....	128

4.3	Fase 3: Análisis del Peso Explicativo	130
4.3.1	Diagnóstico de colinealidad entre variables explicativas	130
4.3.2	Regresiones logísticas multinomiales	131
4.3.3	Precisión y exactitud del modelo de regresión final	140
5	Discusión	142
5.1	Fase 1: Estimación de Residuos	143
5.2	Fase 2: Clasificación de la Eficacia Escolar.....	150
5.3	Fase 3: Análisis del Peso Explicativo	152
6	Conclusiones, Limitaciones y Prospectiva para Futuros Estudios	157
	Referencias.....	163

Índice de Tablas

Tabla 1. Histórico de los resultados de México en PISA	6
Tabla 2. Histórico de las variables contextuales recogidas entre PISA 2000 y PISA 2018	27
Tabla 3. Histórico de las variables de enseñanza recogidas entre PISA 2000 y PISA 2018	29
Tabla 4. Principales elementos de los factores de enseñanza del Modelo Dinámico	55
Tabla 5. Criterios de calidad del mapeo y de la Revisión Sistemática de Literatura.....	59
Tabla 6. Recuento del número de estudios multinivel por autor	64
Tabla 7. Recuento del número de estudios multinivel por revista.....	65
Tabla 8. Recuento de los estudios multinivel por competencia y edición PISA	67
Tabla 9. Variables de enseñanza asociadas con los resultados PISA	69
Tabla 10. Variables contextuales asociadas a los resultados PISA	71
Tabla 11. Coeficientes de correlación intraclase reportados en los análisis multinivel de PISA .	76
Tabla 12. Características del muestreo final del estudio.....	80
Tabla 13. Variables de criterio introducidas en los modelos multinivel.....	83
Tabla 14. Variables predictoras introducidas en los modelos multinivel	84
Tabla 15. Variables explicativas introducidas en las regresiones logísticas multinomiales.....	85
Tabla 16. Resumen de los análisis y softwares empleados en cada fase del estudio.....	89
Tabla 17. Síntesis de los 30 diagnósticos de colinealidad entre las variables predictoras.....	98
Tabla 18. Componentes de varianza del modelo nulo en competencia lectora	99
Tabla 19. Modelo condicional en competencia lectora con estimación robusta del error	101
Tabla 20. Componentes de varianza del modelo condicional en competencia lectora.....	101
Tabla 21. Descriptivos de los residuos escolares en competencia lectora	102
Tabla 22. Componentes de varianza del modelo nulo en competencia matemática.....	105

Tabla 23. Modelo condicional en competencia matemática con estimación robusta del error ..	106
Tabla 24. Componentes de varianza del modelo condicional en competencia matemática	107
Tabla 25. Descriptivos de los residuos escolares en competencia matemática	107
Tabla 26. Componentes de varianza del modelo nulo en competencia científica	110
Tabla 27. Modelo condicional en competencia científica con estimación robusta del error	111
Tabla 28. Componentes de varianza del modelo condicional en competencia científica.....	112
Tabla 29. Descriptivos de los residuos escolares en competencia científica.....	112
Tabla 30. Resumen de las variables que integraron cada modelo multinivel.....	115
Tabla 31. Descriptivos de los residuos escolares en competencia lectora transformados	116
Tabla 32. Descriptivos de los residuos escolares en competencia matemática transformados ..	119
Tabla 33. Descriptivos de los residuos escolares en competencia científica transformados	122
Tabla 34. Clasificaciones de la eficacia de las escuelas	124
Tabla 35. Distancia entre centroides de clústers finales	124
Tabla 36. Comparaciones múltiples entre agrupaciones por competencia: Bonferroni	125
Tabla 37. Descriptivos de las escuelas de baja eficacia, residuo centrado y alta eficacia	127
Tabla 38. Rendimiento de los estudiantes por clúster.....	128
Tabla 39. Diagnóstico de colinealidad entre variables explicativas	130
Tabla 40. Correlación entre residuos y variables explicativas.....	131
Tabla 41. Modelos de regresión logística multinomial - explicativas escolares.....	132
Tabla 42. Modelos de regresión logística multinomial - explicativas de enseñanza	133
Tabla 43. Modelo final de regresión logística multinomial - escolares y de enseñanza.....	134
Tabla 44. Matriz de confusión del modelo de regresión logística multinomial final	141

Índice de Figuras

Figura 1. Representación de los niveles de anidamiento en educación	8
Figura 2. Gráfico de líneas del número de indicadores contextuales entre PISA 2000 y PISA 2018	28
Figura 3. Gráfico de líneas del número de indicadores de enseñanza entre PISA 2000 y PISA 2018	31
Figura 4. Marco conceptual para la Eficacia Educativa de Creemers (1991).....	37
Figura 5. Modelo CIPO de Scheerens (1991).....	38
Figura 6. Modelo Dinámico de Eficacia Educativa	40
Figura 7. Factores a nivel escuela del Modelo Dinámico de Eficacia Educativa	46
Figura 8. Factores a nivel aula del Modelo Dinámico de Eficacia Educativa	52
Figura 9. Proceso de extracción de datos para el mapeo y revisión sistemática de literatura	61
Figura 10. Recuento de los análisis multinivel de datos PISA	63
Figura 11. Recuento de los análisis multinivel de datos PISA por país	66
Figura 12. Recuento de los análisis multinivel por edición PISA	68
Figura 13. Histograma de los residuos escolares en competencia lectora	103
Figura 14. Gráfico Q-Q de los residuos escolares en competencia lectora	103
Figura 15. Nube de puntos de los residuos escolares en competencia lectora.....	104
Figura 16. Histograma de los residuos escolares en competencia matemática.....	108
Figura 17. Gráfico Q-Q de los residuos escolares en competencia matemática.....	108
Figura 18. Nube de puntos de los residuos escolares en competencia matemática	109
Figura 19. Histograma de los residuos escolares en competencia científica	113
Figura 20. Gráfico Q-Q de los residuos escolares en competencia científica	113

Figura 21. Nube de puntos de los residuos escolares en competencia científica.....	114
Figura 22. Histograma de los residuos escolares en competencia lectora transformados	117
Figura 23. Gráfico Q-Q de los residuos escolares en competencia lectora transformados.....	117
Figura 24. Nube de puntos de los residuos escolares en competencia lectora transformados....	118
Figura 25. Histograma de los residuos escolares en competencia matemática transformados...	119
Figura 26. Gráfico Q-Q de los residuos escolares en competencia matemática transformados .	120
Figura 27. Nube de puntos de los residuos escolares en competencia matemática transformados	121
Figura 28. Histograma de los residuos escolares en competencia científica transformados	122
Figura 29. Gráfico Q-Q de los residuos escolares en competencia científica transformados.....	123
Figura 30. Nube de puntos de los residuos escolares en competencia científica transformados	123
Figura 31. Gráfico 3D de los clústers de los residuos escolares transformados	126
Figura 32. Contrastes marginales en escuelas de alta eficacia.....	129
Figura 33. Contrastes marginales en escuelas de residuo centrado	129
Figura 34. Contrastes marginales en escuelas de baja eficacia.....	129
Figura 35. Gráfico de líneas de la variable adaptación de la instrucción por clasificación de eficacia	135
Figura 36. Gráfico de líneas de la variable retroalimentación percibida por clasificación de eficacia	136
Figura 37. Gráfico de líneas de la variable tiempo promedio en el aula por clasificación de eficacia	137
Figura 38. Gráfico de líneas de la variable clima disciplinario en las lecciones por clasificación de eficacia	138

Figura 39. Gráfico de líneas de la variable uso de las TIC durante la clase por clasificación de eficacia	139
Figura 40. Gráfico de líneas de la variable clima escolar discriminatorio por clasificación de eficacia	140

Índice de Ecuaciones

(1). Distribución de Johnson S_B	93
(2). Distribución de Johnson S_L	93
(3). Distribución de Johnson S_U	93
(4). Modelo nulo en competencia lectora	99
(5). ICC del modelo nulo en competencia lectora	100
(6). Modelo condicional en competencia lectora.....	100
(7). ICC del modelo condicional en competencia lectora.....	101
(8). Modelo nulo en competencia matemática.....	104
(9). ICC del modelo nulo en competencia matemática.....	105
(10). Modelo condicional en competencia matemática	105
(11). ICC del modelo condicional en competencia matemática	107
(12). Modelo nulo en competencia científica	109
(13). ICC del modelo nulo en competencia científica	110
(14). Modelo condicional en competencia científica.....	110
(15). ICC del modelo condicional en competencia científica.....	112

Resumen

El objetivo de esta investigación fue determinar los factores de enseñanza que explican la eficacia escolar de los centros educativos mexicanos evaluados en PISA 2018. La muestra estuvo compuesta por 7 267 estudiantes y 268 escuelas. El estudio precisó de tres fases de análisis, la estimación de residuos escolares en competencia lectora, matemática y científica mediante la construcción de modelos multinivel, la clasificación de la eficacia escolar a partir de la técnica de conglomeración k-medias, y el análisis del valor explicativo de los procesos educativos sobre la eficacia escolar con regresiones logísticas multinomiales. Cinco variables de enseñanza y una escolar explicaron significativamente la probabilidad de ser un centro de alta o baja eficacia: la adaptación de la instrucción, la retroalimentación dada por el docente, el tiempo promedio en clase, el clima disciplinario en las lecciones, el uso de las TIC en la instrucción y el clima escolar discriminatorio. Los hallazgos sugieren que la enseñanza eficaz de los docentes es un elemento clave para la mejora escolar y el incremento del logro de los estudiantes mexicanos.

Palabras clave: eficacia docente, eficacia escolar, modelos jerárquicos lineales, evaluación educativa.

Introducción

El estudio se enmarca dentro del cuerpo de conocimiento sobre Eficacia Educativa, pues aporta información tanto a la línea de investigación sobre Eficacia Escolar, como a la de Enseñanza Eficaz. Dos líneas que se desarrollaron de forma paralela pero que, como se aprecia en este documento, actualmente se conciben como interdependientes para la mejora educativa de las escuelas, las prácticas docentes, y los resultados estudiantiles.

El principal objetivo de este estudio fue determinar los factores de enseñanza que explican la eficacia escolar de los centros educativos mexicanos evaluados en PISA 2018, a partir de la información recogida con los cuestionarios de contexto estudiantiles y escolares. Con ello, se pretende generar información aplicable al contexto mexicano sobre las prácticas de enseñanza que impactan sobre los resultados en PISA y que pueden servir como insumo para la mejora educativa de este país.

La investigación precisó de tres enfoques analíticos distintivos: (a) la estimación de residuos en competencia lectora, matemática y científica de los centros educativos mexicanos a partir de la construcción de modelos multinivel; (b) la clasificación de la eficacia escolar de los centros educativos mexicanos mediante la aplicación de la técnica de conglomeración K-medias; y, (c) el análisis del peso explicativo de los procesos educativos sobre la eficacia escolar de los centros educativos mexicanos con regresiones logísticas multinomiales.

Para abordar lo anterior, el presente documento está integrado por cinco capítulos. El primero que delimita el planteamiento del problema, define los objetivos (general y específicos), y justifica social, teórica y metodológicamente el estudio.

En el segundo que integra la revisión de literatura, constituida por seis apartados. El Programa Internacional para la Evaluación de los Alumnos (PISA), en el que se describen las

características del proyecto PISA, sus fundamentos, la prueba, los cuestionarios de contexto y principales limitaciones. Una revisión documental de los cuestionarios de contexto, en la que se registran la recurrencia de las variables contextuales y de enseñanza a través del tiempo. La Eficacia Educativa, en el que se detalla en qué consiste este cuerpo de conocimiento, las disciplinas que la integran y la teoría que le atañe. La Eficacia Escolar, en la que se define qué es una escuela eficaz y los factores relativos a estas. La Enseñanza Eficaz, en la que se expone qué es un enseñante eficaz y las prácticas que le integran. Y, por último, una Revisión Sistemática de Literatura que expone cómo se han realizado los análisis multinivel con datos PISA y cuáles han sido las variables de enseñanza de mayor interés del año 2000 al año 2020.

En el tercero se presenta el método propuesto en la presente investigación. Se describe el diseño, los participantes e instrumentos incluidos en PISA 2018, las variables utilizadas y los análisis empleados para el cumplimiento a cada objetivo.

Finalmente, en el cuarto y quinto se presentan y discuten los resultados obtenidos; y en el sexto y séptimo se argumentan las conclusiones, se declaran las limitaciones y se emiten recomendaciones.

1 Planteamiento del Problema

El Programa para la Evaluación Internacional de Alumnos (PISA, por sus siglas en inglés) es un proyecto de carácter comparativo que ha sido empleado desde el año 2000 por múltiples países, 32 en el primer ciclo PISA y 79 en el último. Evalúa a cada uno de los participantes en su labor en el desarrollo de competencia lectora, matemática y científica (Organización para la Cooperación y el Desarrollo Económico [OCDE], 2006). Los datos que publica cada tres años han brindado la oportunidad de reconocer la medida en que cada país participante ha preparado a sus estudiantes de 15 años de edad al término de su educación formal para la vida (Anderson et al., 2009; Turner y Adams, 2007).

De acuerdo con Jornet (2016), si bien se han realizado otras pruebas internacionales, como el Estudio Internacional de Tendencias en Matemáticas y Ciencias, y el Estudio Internacional de Progreso en Comprensión Lectora (TIMSS y PIRLS respectivamente, por sus siglas en inglés), PISA ha sido la de mayor impacto político, social y académico. Cordero et al. (2013) sostuvieron que, el impacto de PISA se ha debido principalmente a las inesperadas diferencias en el nivel de competencia de los estudiantes entre naciones alrededor del mundo. Dichas diferencias, a su vez, han propiciado un clima de insatisfacción y un incremento exponencial en el estudio empírico de las bases de datos proporcionadas por PISA en busca de las principales explicaciones (Cordero et al., 2013; Reynolds et al., 2014).

Aunque PISA fue diseñada para satisfacer una visión macro analítica, orientar políticas educativas, este programa ha sido optimizado por investigadores independientes a nivel meso y micro analítico. Pues, según Jornet et al. (2012), se ha vuelto frecuente aprovechar las bases de datos que este proyecto ofrece para la identificación de los factores educativos claves y, con ello, planear rutas de mejora escolar o, bien, dar razón del porqué de determinados resultados.

Transformándola así, en una herramienta de mayor utilidad a nivel escuela y a nivel aula (Bove et al., 2016; Karakolidis et al., 2016).

Naturalmente, han existido diferencias entre lo que funciona en una u otra parte del mundo (Caro et al., 2016). No son los mismos contextos, escuelas, docentes y estudiantes. Por consiguiente, los hallazgos de los análisis secundarios de PISA han diferido entre naciones. Por ejemplo, Barnard-Brak et al. (2018) encontraron, con la muestra estadounidense de PISA 2012, que una menor *oportunidad de aprender* se asocia significativamente con una menor probabilidad de responder correctamente un ítem de conocimientos y procedimientos matemáticos. En cambio, Gamazo et al. (2018) hallaron, con la muestra española de PISA 2015, que un *clima de disciplina* y el *nivel de justicia del profesorado* se relacionan significativamente con el rendimiento global de los estudiantes.

En el caso de México, un país históricamente situado por debajo de la media establecida por la OCDE en las tres áreas evaluadas (ver Tabla 1), se ha aprovechado de forma escasa y parcial la información recabada de este país en PISA (Chaparro y Gamazo, 2020). Escasa porque existe un número muy reducido de estudios al respecto; y parcial porque, de acuerdo con Anderson et al. (2009) e Iñiguez-Berrozpe y Marcaletti (2018), los estudios existentes suelen contar con dos limitaciones: (a) se opta por una definición de constructo u objeto de estudio distante a la propia del programa a gran escala; y, (b) se elige un enfoque analítico que no contempla la naturaleza de las observaciones.

Tabla 1*Histórico de los resultados de México en PISA*

Competencia	Promedio	2000	2003	2006	2009	2012	2015	2018
Lectora	OCDE	500	494	492	493	496	493	487
	México	422	400	410	425	424	423	420
Matemática	OCDE	500	500	498	496	494	490	489
	México	387	385	406	419	413	408	409
Científica	OCDE	500	500	500	501	501	493	489
	México	422	405	410	416	415	416	419

Nota: Elaboración propia con base en Chaparro y Gamazo (2020) y OCDE (2019a).

Este campo de análisis se ha desarrollado a partir de la información recogida por los cuestionarios de contexto que, según la OCDE (2017a), han tenido como base de conocimiento la investigación sobre Eficacia Educativa. Esta línea de investigación ha permitido la identificación e integración de los principales factores asociados a los resultados de los estudiantes: aspectos no cognitivos y metacognitivos, como lo son las actitudes, creencias, motivaciones y aspiraciones; antecedentes de los estudiantes, por ejemplo, el índice social, económico y cultural (ISEC; Coleman et al., 1966); y, prácticas educativas, en las que se incluyen las políticas escolares y la labor del docente (Creemers y Kyriakides, 2008).

En lo que refiere a las prácticas educativas, dentro y fuera de los análisis secundarios de PISA, el estudio de los factores relativos a la escuela cuenta con una extensa trayectoria. Su origen puede remontarse a Coleman et al. (1996), quienes afirmaron que el aporte de la escuela no era significativo en comparación con las características de fondo de los estudiantes, como su grupo racial y sus condiciones socioeconómicas. De acuerdo a Chapman et al. (2016), el informe Coleman et al. (1966) fue el detonante de los primeros estudios empíricos sobre las escuelas denominadas eficaces (Chapman et al., 2016), aquellas que tenían la capacidad de producir efectos

(Weber, 1971; Edmonds, 1979). Al día de hoy, la Eficacia Escolar y los factores que operan a este nivel son una de las líneas de investigación más proliferales en el ámbito educativo.

No obstante, según Reynolds et al. (2014), debido a que gran parte de los estudios se han enmarcado casi de forma exclusiva en las escuelas, el poder explicativo y la puesta en práctica de los hallazgos se ha visto obstaculizada al prestar menor atención a lo que acontece en el aula: la labor docente. En las últimas tres décadas, se ha demostrado consistentemente que la enseñanza, y los factores relativos a esta, son los de mayor importancia al explicar los resultados escolares y estudiantiles (Muijs et al., 2014). No solo se ha evidenciado que a nivel aula se explica mayor variabilidad de los resultados de los estudiantes que a nivel escuela (Hill y Rowe, 1996, 1998; Kyriakides et al., 2000; Scheerens y Bosker, 1997; Teddlie y Reynolds, 2000; Yair, 1997), sino que, también, la varianza a nivel escuela y a nivel aula puede explicarse por lo que hacen los docentes (Creemers, 1994; Muijs y Reynolds, 2011, 2017), es decir, la Eficacia Escolar depende de la eficacia del docente.

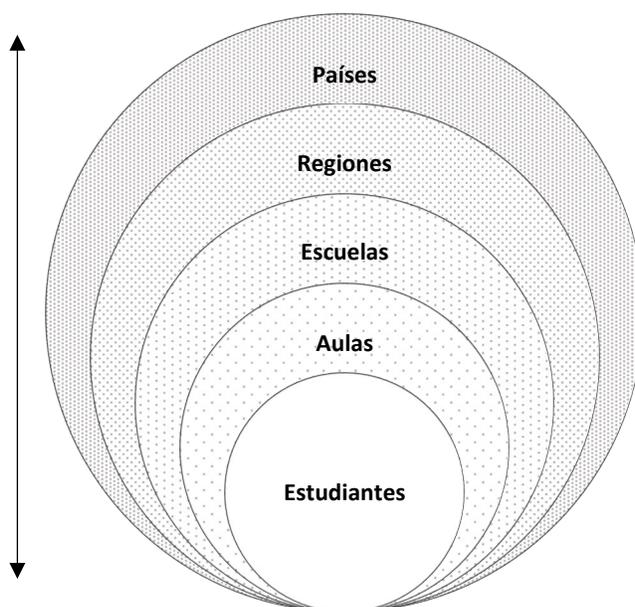
De acuerdo con Chapman et al. (2016), la identificación de la importancia del docente ha sido gracias al desarrollo de los estadísticos multinivel, pues han permitido análisis a gran escala cada vez más complejos; como los análisis secundarios de PISA. En la estadística inferencial clásica, como la regresión lineal y multivariable, se precisa de la independencia de las observaciones (Pituch y Stevens, 2015). En los datos recogidos por PISA las observaciones son dependientes, el método de muestreo es estratificado por dos etapas (OCDE, 2016a), primero se eligen al azar las escuelas y, de las escuelas seleccionadas, se escogen a los estudiantes; por lo tanto, no existe independencia en las observaciones (Iñiguez-Berrozpe y Marcaletti, 2018). Los estudiantes, al compartir contextos sociales, en los que viven, trabajan y aprenden, tienden a ser similares (Hox et al., 2018). Los estadísticos multinivel fueron desarrollados principalmente para

analizar y respetar la dependencia de las observaciones, habituales en educación (Gully y Phillips, 2019).

A partir de algunos estudios desarrollados en la década de 1980 (Aitkin et al., 1981; Aitkin y Longford, 1986; Goldstein, 1987; Raudenbush y Bryk, 1986) las observaciones dependientes se entienden como series de relaciones anidadas, jerárquicas o multinivel. En educación, por ejemplo, los estudiantes se encuentran anidados en aulas, las aulas en escuelas, las escuelas en regiones, y las regiones en países; donde cada nivel de anidamiento tiene un efecto bidireccional sobre el otro (ver Figura 1, Vandenberg y Richardson, 2019). Como lo indicaron Humphrey y LeBreton (2019), el análisis de relaciones entre los diferentes niveles jerárquicos, y las variables definidas en cada nivel, dieron lugar a las investigaciones denominadas multinivel; los cuales, en el caso de PISA, han permitido obtener resultados más apegados a la realidad educativa (Chapman et al., 2016; Reynolds et al., 2014).

Figura 1

Representación de los niveles de anidamiento en educación



Con base en lo anterior, se puede afirmar que: (a) se ha desaprovechado la información de México recabada por PISA (Chaparro y Gamazo, 2020); (b) se ha optado por definiciones de constructo distantes a la base de conocimientos de los cuestionarios de contexto de PISA, la Eficacia Educativa (Anderson et al., 2009); (c) se ha obstaculizado el poder explicativo y la puesta en práctica de los hallazgos al prestar menor atención al aula y la labor docentes (Reynolds et al., 2014); y, (d) se han elegido enfoques analíticos que no contempla la naturaleza de las observaciones, anidadas, jerárquicas o multinivel (Iñiguez-Berrozpe y Marcaletti, 2018).

Por consiguiente, en el presente estudio se aspiró a responder la siguiente pregunta de investigación ¿qué factores de enseñanza explican la eficacia escolar de los centros educativos mexicanos muestreados en PISA 2018?

1.1 Objetivos de Investigación

1.1.1 Objetivo general

Determinar los factores de enseñanza que explican la eficacia escolar de los centros educativos mexicanos muestreados en PISA 2018, a partir de la información recogida con los cuestionarios de contexto estudiantiles y escolares.

1.1.2 Objetivos específicos

1. Estimar los residuos en competencia lectora, matemática y científica de los centros educativos mexicanos.
2. Clasificar la eficacia escolar de los centros educativos mexicanos a partir su residuo.
3. Analizar el peso explicativo de los procesos educativos sobre la eficacia escolar de los centros educativos mexicanos.

1.2 Justificación

1.2.1 Relevancia social

Desde el año 2000 a la actualidad, México ha participado en la evaluación PISA. En respuesta, la OCDE ha emitido una serie de informes cíclicos sobre los avances y retrocesos, y un conjunto de recomendaciones para la elaboración de estrategias político-educativas de este país (Jiménez, 2018). No obstante, los informes y puntajes han parecido no ser suficientes por sí mismos para lograr la mejora educativa y aumentar los resultados de los estudiantes mexicanos (Martínez-Rizo, 2016).

De acuerdo con Jornet (2016) y Rivas (2015), después de cada evaluación PISA quedan muchas interrogantes. ¿Hubo escuelas que aportaron más o menos de lo que podían conseguir? ¿Cómo lo lograron? ¿En qué podríamos enfocarnos para obtener mejores resultados?

La falta de respuestas a estas interrogantes, incluso, ha derivado en diversos cuestionamientos entorno la utilidad práctica de la prueba PISA más allá de la comparabilidad entre países. Alcaraz et al. (2013) destacó la siguiente situación:

La información proporcionada por estas evaluaciones externas es fundamentalmente acumulativa y, en el mejor de los casos, resulta útil para identificar el progreso de cada escuela o del sistema, pero su utilidad es menor cuando se trata de orientar las prácticas de mejora (pp. 581).

Al respecto, este trabajo de investigación ofrece información que podría dar respuesta a este tipo de cuestionamientos. Particularmente, se identifica la proporción de escuelas que rindieron en menor medida y más allá de lo esperado en PISA 2018, se determinan los factores de enseñanza que explican la Eficacia Escolar de México, y se discute el peso y el efecto de ciertas prácticas en el contexto educativo mexicano. De ser retomada, dicha información podría ser

insumo para aumentar las probabilidades de éxito de los diversos esfuerzos educativos (Popham, 2013), pues, como lo indicó Zabalza (2007), la labor educativa tendrá mayor efecto si se poseen el conocimiento teórico sobre la lógica y los factores que inciden en los resultados de las escuelas.

1.2.2 Relevancia teórica

El cuerpo de conocimiento sobre Eficacia Educativa cuenta con una extensa trayectoria (Chapman et al., 2016). Desde el informe Coleman et al. (1966) al día de hoy, se ha caracterizado por un crecimiento exponencial en la investigación empírica sobre los factores educativos que inciden en los resultados de los estudiantes y la mejora de las escuelas (Reynolds et al., 2014). No obstante, gran parte de los estudios se han enmarcado, casi de forma exclusiva, en los factores a nivel escuela, como el liderazgo del directivo, la gestión escolar, entre otras (Reynolds et al., 2014).

De acuerdo con Muijs et al. (2014), en el estudio de la eficacia, la falta de atención en el nivel aula y los factores relativos a la enseñanza ha mermado el poder explicativo y la puesta en práctica de los hallazgos. En las últimas tres décadas, se ha evidenciado constantemente que, no solo a nivel aula se explica mayor variabilidad de los resultados de los estudiantes que a nivel escuela (Hill y Rowe, 1996, 1998; Kyriakides et al., 2000; Scheerens y Bosker, 1997; Teddlie y Reynolds, 2000; Yair, 1997), sino que, también, la variabilidad a nivel aula y a nivel escuela puede explicarse por la labor docente: la enseñanza (Creemers, 1994; Muijs y Reynolds, 2011, 2017).

Aunque en PISA no se cuenta con variables a nivel aula, en este estudio se aprovecha la información recogida por los cuestionarios de contexto estudiantiles y escolares para analizar los factores de enseñanza que explican la eficacia de las escuelas mexicanas en PISA 2018. Con ello, se aporta información a dos líneas de investigación que se desarrollaron en paralelo pero que hoy se conciben como interdependientes, la Eficacia Escolar y la Enseñanza Eficaz. Aunado a ello, se pretende responder a dos elementos fundamentales en el estudio de la Eficacia Educativa

(Campbell et al., 2004; Reynolds et al., 2014): (a) la actualización continua del conocimiento con el propósito de que esta sea pertinente, coherente y de calidad; y, (b) el estudio de la medida en que los resultados son generalizables y aplicables a los diversos contextos, y en qué medida estos contrastan (p. ej. Reynolds et al., 2002; Townsend, 2007).

1.2.3 Relevancia metodológica

Entre los avances metodológicos del cuerpo de conocimiento sobre Eficacia Educativa se encuentra el empleo de los estadísticos multinivel (Creemers et al., 2010). Estos se han posicionado -dada su naturaleza multivariable y multinivel- como una alternativa para la medición e identificación de centros eficaces. Dado que respetan la estructura jerárquica de los datos habituales en educación, facilitan el control de efectos significativos provenientes de variables contextuales; aumentando así, la precisión de las estimaciones residuales (Raudenbush et al., 2016).

En la investigación educativa de la eficacia se han utilizado comúnmente dos indicadores: los residuos y el valor añadido (Joaristi et al., 2014; Lizasoain y Angulo, 2014; Martínez-Abad et al., 2017; Pedroza et al., 2018). El primero es generalmente utilizado en estudios de corte transversal y nos indican la diferencia entre el rendimiento esperado en un momento en el tiempo, de acuerdo a sus características contextuales, y el rendimiento “real” obtenido. El segundo es particularmente utilizado cuando se tiene información longitudinal, aunque este tiene base en los residuos, este nos indica el crecimiento del rendimiento de un sujeto (o escuela) a través del tiempo. Ambos siguen una lógica lineal, es decir, a mayor residuo o valor añadido mayor eficacia y, por el contrario, a mayor residuo o valor añadido negativo menor eficacia. En ese sentido, el papel de los residuos y el valor añadido ha sido fundamental en análisis explicativos (de corte cuantitativos) u

observación *in situ* de centros eficaces (en estudios de método mixto). No obstante, aún no existe un consenso metodológico sobre identificar y seleccionar a los centros eficaces.

En este estudio se realizan dos aportes a la identificación y clasificación de escuelas eficaces. El primero es la aplicación de la Transformación de Johnson (1949) para facilitar la interpretación de los residuos y, con ello, la identificación de centros eficaces. El segundo es el empleo de tres criterios para la clasificación de escuelas eficaces mediante K-medias.

2 Revisión de Literatura

2.1 Programa para la Evaluación Internacional de Alumnos (PISA)

PISA ha sido considerada como una de las evaluaciones a gran escala más relevantes a nivel internacional (Jiménez, 2018). Fue diseñado entre los años 1997 y 1999 por la Organización para la Cooperación y el Desarrollo Económico (OCDE) e implementada por primera vez en el año 2000 a 43 países. Si bien han existido otras pruebas internacionales, como el Estudio Internacional de Tendencias en Matemáticas y Ciencias y el Estudio Internacional de Progreso en Comprensión Lectora (TIMSS y PIRLS, respectivamente), PISA ha sido la de mayor impacto e influencia en materia político-educativa de los países participantes (Jornet, 2016).

2.1.1 Fundamentos

La OCDE se ha caracterizado por una extensa tradición mediática en el desarrollo de líneas de acción en casi todos los ámbitos de la política pública, excepto defensa (Jiménez, 2018). Desde su origen en la década de 1960, el ámbito educativo ha formado parte de sus intereses fundamentales. Evidencia de ello, fueron sus primeras publicaciones y recopilaciones esporádicas sobre estadísticas relativas a los diversos sistemas educativos nacionales. No obstante, a mediados de la década de 1980, representantes de diversos países expusieron su preocupación sobre la calidad y la fiabilidad de la recolección y el análisis datos (Gamazo, 2019).

Esta situación impulsó a la OCDE a la creación del programa de Indicadores de Sistemas Educativos (INES). Este dio respuesta a las principales demandas de los países participantes, entre ellas (Martens, 2007; Schleicher, 2006): (a) reconfigurar su sistema de publicaciones, de esporádicas a periódicas; y, (b) desarrollar indicadores de relevancia política educativa a partir de estimaciones estadísticas documentadas, y ofrecer resúmenes para la presentación y el debate político.

El desarrollo de indicadores educativos dio pie a constantes evoluciones en la colaboración de la OCDE en el ámbito educativo. Este organismo multilateral transitó gradualmente del desarrollo de indicadores sobre elementos básicos -como los recursos humanos, gasto en educación, ingreso y egreso de estudiantes- hacia la medición del rendimiento en competencias y avances educativos. Según Schleicher, “El resultado más significativo de este proceso ha sido el desarrollo de un ciclo de estudios del rendimiento estudiantil, dirigidos por la OCDE y realizados por el Programa para la Evaluación Internacional de Alumnos (PISA)” (2006, p. 23,).

PISA ha sido el proyecto emblemático de la OCDE para la evaluación de los sistemas educativos nacionales. Este ha asumido la responsabilidad de su estudio y comparación mediante indicadores comunes con el propósito de favorecer la extracción de conclusiones sobre su equidad y eficacia y, con ello, la informada toma de decisiones político-educativas (Bonnet, 2006; Vélaz, 2006). Esta labor implicó el desarrollo de dos instrumentos: una prueba objetiva y cuestionarios de contexto. Su diseño, pilotaje e implementación estuvo a cargo de Ray Adams como director internacional del proyecto, un comité de países participantes, y un Consorcio compuesto por expertos en áreas específicas relevantes (Turner y Adams, 2007): el *Australian Council for Educational Research* (ACER), el *National Institute for Educational Measurement* (Cito Group) en Países Bajos, el *German Institute for International Educational Research* (DIPF), el *Service de Pédagogie Expérimentale* en la *Université de Liège* y el *CAPSTAN Linguistic Quality Control Company* en Bélgica, el *Westat* y el *Educational Testing Service* (ETS) en los Estados Unidos y el *National Institute for Educational Policy Research* (NIER) en Japón.

De acuerdo con Harlen (2002), el equipo -más allá de encargarse de detalles de gestión- influyó en gran medida en el objeto de evaluación de PISA. Acordaron aquello que se considera como resultados deseables, aunque estos estuvieran o no en los currículos de cada país. Ello

significó una de las principales diferencias con TIMSS y PIRLS: las unidades de evaluación, que se exponen a continuación.

2.1.2 Prueba

Desde el año 2000, con la prueba PISA se ha medido el nivel de competencia lectora, matemática y científica de los estudiantes de 15 años de edad. A diferencia de TIMSS y PIRLS, más allá de la reproducción de conocimientos, se ha propuesto específicamente estimar la capacidad de extrapolar y aplicar lo aprendido en entornos conocidos y desconocidos. En cada implementación se ha establecido y rotado la competencia principal -que constituye aproximadamente el 66% de la prueba- y las competencias secundarias -que constituyen casi el 34% de la prueba-. De acuerdo con la OCDE, las competencias se evalúan con diferentes cuadernillos en bloques de pares para evitar el exceso de tiempo de medición. Por ello los estudiantes responden diferentes ítems, primero responden un cuadernillo de la competencia principal y, después, un cuadernillo sobre una de las competencias secundarias (OCDE, 2017b).

2.1.2.1 Competencia lectora

En PISA 2018, la competencia lectora fue el dominio principal a evaluar. Ello significó que dos tercios de los ítems de la prueba correspondieron a esta competencia. Para ello, se reunieron esfuerzos internacionales como la guía de John de Jong y Peter Foltz de *Pearson*, la coordinación de Jean-François Rouet de la *University of Poitiers* en Francia, y la experiencia de Dominique Lafontaine de la *Université de Liège* en Bélgica, Kevin Chung de la *University of Hong Kong* en China, Paul van den Broek de la *Universiteit Leiden* en Países Bajos, Sari Sulkunen de la *University of Jyväskylä* en Finlandia y Sascha Schroeder del *Max Planck Institute for Human Development* en Alemania (OCDE, 2019b).

La trayectoria acumulada de la OCDE y el consenso entre expertos dieron como resultado la redefinición de la competencia lectora. Esta ha variado a lo largo del tiempo con la intención de reflejar las cambiantes circunstancias sociales, económicas y tecnológicas. En PISA 2018 se conceptualizó como “la capacidad del individuo para comprender, usar, evaluar, reflexionar y abordar textos con el fin de lograr sus objetivos, desarrollar su conocimiento y potencial, y participar en la sociedad” (OCDE, 2019b, p. 14).

En la vida cotidiana la lectura es una actividad amplia. Evaluarla en términos de la OCDE, la extrapolación y aplicación del conocimiento, implicó reconocer e incluir en la medida de lo posible la generalidad y diversidad de este dominio. En PISA 2018, con base en el marco de Snow y el RAND Reading Group (2002) sobre las tres fuentes que influyen en la comprensión lectora, se operacionalizaron tres características determinantes para el diseño de su prueba y los resultados obtenidos sobre el nivel de competencia de los estudiantes (OCDE, 2019b, p. 15):

- Procesos (aspectos): la búsqueda y selección de textos en función de su relevancia; localizar y recuperar información; comprender y representar información; evaluar y reflexionar contenido en relación con su calidad y credibilidad.
- Formatos de texto: de fuente única o múltiple; estático o dinámico; discontinuo -como listas-, continuo -como oraciones o párrafos- o mixto.
- Situaciones: libros, apuntes, novelas, cartas personales, documentos oficiales, informes académicos, entre otros.

2.1.2.2 Competencia matemática

En PISA 2018 la competencia matemática fue un dominio secundario. Ello implicó que únicamente le correspondiera una sexta parte de la prueba. De acuerdo con la OCDE (2019b) el

marco para su evaluación fue una continuación del trabajo hecho para PISA 2012, ciclo en el que fue la competencia principal. Por lo tanto, esta fue entendida como

la capacidad de un individuo para formular, emplear e interpretar las matemáticas en una variedad de contextos. Incluye razonamiento matemático y uso de conceptos, procedimientos, hechos y herramientas matemáticos para describir, explicar y predecir fenómenos. Ayuda a las personas a reconocer el papel que desempeñan las matemáticas en el mundo y a emitir juicios y decisiones fundamentados que necesitan ciudadanos constructivos, comprometidos y reflexivos (OCDE, 2013, p. 25).

Su evaluación, al igual que en PISA 2012, integró tres aspectos interrelacionados (OCDE, 2019b, p. 15-16):

- **Procesos:** la OCDE ha definido tres categorías, formular situaciones matemáticamente; emplear conceptos, procedimientos y razonamientos matemáticos; interpretar, aplicar y evaluar resultados matemáticos. En estas categorías incluyen siete aptitudes esenciales para abordar temas específicos, matematizar, representar, razonar, discutir, comunicar, idear, utilizar fórmulas, técnicas, lenguaje y operaciones simbólicas, y emplear herramientas matemáticas.
- **Contenidos:** cantidad, espacio y forma, cambio y relaciones, e incertidumbre en la resolución de problemas matemáticos.
- **Contextos:** los ámbitos en la vida cotidiana del estudiante en el que se encuentran problemas matemáticos, por ejemplo, el social, el educativo y el personal.

2.1.2.3 Competencia científica

En PISA 2018 la competencia científica, al igual que la matemática, fue un dominio secundario. Ello conllevó a que aproximadamente el 16.67% de los ítems pertenecieran a esta competencia. El

marco para su evaluación y definición se recuperaron de PISA 2015, ciclo en el que fue la competencia principal; refinado y ampliado el trabajo realizado para PISA 2006. Esta se precisó como

la capacidad de comprometerse con cuestiones relacionadas con la ciencia y con las ideas de la ciencia como ciudadano reflexivo. Una persona alfabetizada científicamente está dispuesta a participar en un discurso razonado sobre ciencia y tecnología, que requiere las competencias para: explicar fenómenos científicamente [...], evaluar y diseñar investigaciones científicas [...] e interpretar datos y pruebas científicamente (OCDE, 2017c, p. 22).

Su evaluación, a diferencia de PISA 2015, en PISA 2018 envolvió tres aspectos (OCDE, 2019b, pp. 16):

- Contexto: en materia macro (asuntos globales o nacionales) o micro científica (asuntos personales) que requieren el uso de conocimientos científicos.
- Conocimiento: el discernir de los fundamentos de la ciencia, conceptos, hechos y teorías. Donde se ve implicado el conocimiento sobre lo que nos rodea (conocimiento de contenido); el cómo se producen tales saberes (conocimiento de proceso) y qué subyace a su origen (conocimiento epistemológico).
- Competencias: la aptitud para diseñar, explicar y evaluar estudios científicos, así como interpretar datos y evidencias.

2.1.3 Cuestionarios de contexto

Desde el primer ciclo, la prueba de competencias se ha acompañado de un segundo instrumento, los cuestionarios de contexto. Estos se han implementado con el propósito de recoger información complementaria al nivel de competencia de los estudiantes. Las bases de datos que aportan los

cuestionarios, en paralelo con el nivel de competencia de los estudiantes, han resultado de utilidad para la explicación de las prácticas eficaces y los resultados obtenidos (Jornet et al., 2012).

El marco para los cuestionarios de PISA 2012 ha sido la base para los instrumentos de los últimos dos ciclos, PISA 2015 y PISA 2018, con la intención de obtener insumos equivalentes y lograr establecer tendencias. Por consiguiente, el Marco de Referencia para el Cuestionario de Contexto PISA 2018, al igual que en las dos ocasiones anteriores, incluyó tres dimensiones: constructos no cognitivos y metacognitivos, constructos relacionados con los antecedentes del estudiante, y procesos educativos (OCDE, 2019b).

Las dos primeras dimensiones, resultados no cognitivos y antecedentes, recogen datos sobre el entorno y la educación recibida por el estudiante. Usualmente, esta información se obtiene mediante cuestionarios dirigidos al alumno o a los padres de familia, en los que se incluyen preguntas sobre (OCDE, 2017a):

- Variables de resultados no cognitivos: compromiso, motivación, interés y creencias hacia el aprendizaje, compartir y respetar valores y creencias, etc.
- Variables de antecedentes: estatus económico, social y cultural (ISEC), ambiente familiar y de hogar, origen étnico, estatus migratorio e historial educativo.

La tercera dimensión, sobre procesos educativos, recoge datos sobre las prácticas de enseñanza, y la política y la gobernabilidad de las escuelas. Este es el elemento central del cuestionario de contexto, pues es el insumo para la emisión de recomendaciones político-educativas y el análisis estadístico de investigadores independientes en busca de las principales explicaciones. Su base de conocimiento fue la línea de investigación sobre Eficacia Educativa que, de acuerdo con la OCDE (2017a), ha permitido la identificación de los factores con influencia esperada sobre los resultados de los estudiantes:

- Variables de enseñanza: organización y gestión del aula, clima de aula, desafío cognitivo, manejo del tiempo y oportunidades de aprendizaje (dentro y fuera de clase), apoyo y credenciales docentes, etc.
- Variables escolares: liderazgo directivo, gestión, currículo organizado, ambiente escolar, participación de los padres y evaluación estudiantil e institucional, etc.

Además de lo mencionado, en PISA 2018 se ofrecieron adicionalmente cinco cuestionarios de carácter voluntario para determinados sujetos (OCDE, 2019b):

- Cuestionario de familiaridad con la informática: dirigido a estudiantes e informó su empleo, aptitud y disposición hacia las tecnologías de la información y la comunicación (TIC) de los estudiantes.
- Cuestionario de bienestar: enfocado a estudiantes para estar al tanto de su salud, percepción de la vida, vínculos sociales, y actividades escolares y extraescolares.
- Cuestionario de carrera educativa: un sondeo para estudiantes referente a su visión a futuro en el ámbito académico y profesional.
- Cuestionario para familia: dirigido a padres y madres de familia e incluyó preguntas acerca de cómo se involucran en la formación de su hijo (en el centro educativo y en el hogar), lo que esperan de sus hijos, entre otras cosas.
- Cuestionario para docentes: con el que se indagaron aspectos del docente relativos a su desarrollo y competencias profesionales iniciales, convicciones, actitudes y su labor en el aula. En el caso de México este cuestionario no fue aplicado.

2.1.4 Limitaciones de PISA

Al día de hoy PISA ha proveído siete ciclos de evidencias sobre estudiantes, escuelas y sistemas educativos nacionales. Desde su primera aplicación, los datos proporcionados por este estudio han

dado oportunidad a análisis secundarios sobre diversos temas relacionados con la educación, las características de los estudiantes, los procesos escolares y las prácticas de enseñanza, por mencionar algunos. De acuerdo con Caro y Kyriakides (2019), reunir la cantidad de información que este proyecto ofrece conlleva un diseño complejo.

En este apartado se abordan las principales limitaciones de PISA reportadas en la literatura. La subestimación, la sobreestimación, la posición de los ítems, la inferencia causal con datos transversales, y la ausencia de datos a nivel aula.

2.1.4.1 Subestimación

Como en otras evaluaciones internacionales, PISA ha establecido criterios de exclusión en el muestro para prevenir la subestimación. No participan estudiantes que contaban con discapacidad funcional; discapacidad cognitiva, conductual o emocional; o quienes presentaran insuficiente dominio del idioma en el que se aplicaron los instrumentos (OCDE, 2016a). No obstante, LeRoy et al. (2018) y Gamazo (2019) han dado evidencia del incumplimiento de los criterios de exclusión de PISA. En la muestra española, por ejemplo, se han incluido a los estudiantes con necesidades educativas especiales (NEE) por no contar con un diagnóstico oficial. Esta situación se agrava en contextos vulnerables, donde regularmente no se cuenta con mecanismos para la detección y diagnóstico con NEE. Ello ha resultado en subestimaciones de los resultados obtenidos a nivel escuela, comunidad y país.

2.1.4.2 Sobreestimación

Spaull (2018) y Marchionni y Vázquez (2018) analizaron la sobreestimación producto de contrariedades en la elegibilidad de algunos estudiantes en el muestreo de PISA. Por un lado, Spaull (2018) abordó las consecuencias de elegir únicamente a los jóvenes de 15 años que se encuentran en el grado 7 o superior. En países desarrollados la mayoría de los jóvenes de 15 años

cumplen con este criterio, sin embargo, en los países en desarrollo una parte importante de esta población se encuentra fuera de las escuelas o con un atraso educativo severo. Este autor abordó particularmente el caso de Turquía, un país en el que un tercio de su población de 15 años no asiste a la escuela y, por ende, no es representada. Tras el análisis de los resultados e informes PISA concluyó que existe una sobreestimación del rendimiento en competencias de esta población, y emite una advertencia en relación con la generalización y comparación entre países desarrollados y en desarrollo.

Por otro lado, Marchionni y Vázquez (2018) exploraron el efecto de un año de escolaridad adicional sobre el rendimiento de la competencia matemática en siete países latinoamericanos en PISA 2012. Dado que cada país establece las normas para el acceso a primer grado de primaria, al elegir los estudiantes para la muestra, en algunos contextos los jóvenes de 15 años pueden contar con un año adicional de escolaridad. Los autores emplearon regresión aguda y difusa y encontraron que contar con un año de escolaridad adicional tiene un fuerte efecto sobre el logro en competencia matemática. Ello podría implicar un desequilibrio en las comparaciones internacionales, y un sesgo en los resultados de algunas escuelas que cuentan con un alto grado de estudiantes de 15 años con un año adicional de escolaridad y que fueron incluidos en la muestra.

2.1.4.3 Posición de los ítems

Nagy et al. (2018) examinaron la validez de los puntajes en función de la posición de los ítems. Tras el análisis de patrones de respuesta de la muestra alemana en PISA 2018, identificaron que la probabilidad de responder acertadamente disminuyó a medida que los ítems se encuentran al final de la prueba por causa de la fatiga. Esta situación afecta especialmente a los dominios secundarios que comprenden la segunda etapa evaluativa, pues en lugar de reflejar la capacidad del estudiante de emplear su conocimiento en diversas situaciones podrían reflejar los efectos de su fatiga para

contestar. Finalmente, los autores recomendaron tomar en cuenta la posición de los ítems al emitir conclusiones sobre el rendimiento en competencias.

2.1.4.4 Inferencia causal con datos transversales

Scheerens et al. (2015) indicaron que, entre los inconvenientes de PISA, la naturaleza transversal de sus datos limita la emisión de causalidades. Cada implementación de este proyecto mantiene un conjunto de elementos comunes para alinear las puntuaciones a lo largo del tiempo. Sin embargo, la configuración técnica de PISA y la marcada desvinculación entre los ítems de cada ciclo inhiben la obtención de datos longitudinales (Robitzsch y Lüdtke, 2018) e imposibilitan la emisión de modelos de valor añadido (Gamazo, 2019).

De acuerdo con la OCDE (2011), los modelos de valor agregado han adquirido particular importancia en la investigación educativa. Se han convertido en una base cuantitativa más precisa (en comparación con las puntuaciones brutas) al considerar el rendimiento previo de los estudiantes, separar las contribuciones ajenas a la escuela y al aula, y disminuir el grado de error en la identificación de centros eficaces. No obstante, los modelos de valor agregado requieren estrictamente de datos longitudinales, deben ser las mismas escuelas y los mismos estudiantes evaluados en al menos tres puntos en el tiempo. Generalmente, en cada edición PISA los estudiantes y las escuelas no son los mismos, por lo que se ha optado por otros modelos con menor precisión que el valor agregado: el valor agregado contextualizado y la estimación de residuos escolares.

2.1.4.5 Ausencia de datos a nivel aula

La comunidad científica internacional ha denunciado la ausencia de datos a nivel aula en PISA (Gamazo, 2019; Lafontaine et al., 2015; Sandilands et al., 2014; Scheerens et al., 2015). Desde principios del siglo XX se ha reconocido que los procesos de aula no pueden ser ignorados en los

análisis sobre la eficacia educativa (Chapman et al., 2016). Aunque en PISA 2015 y PISA 2018 se manifestó un mayor interés en la información relacionada con el profesorado y se aumentó el número de ítems sobre la enseñanza en los cuestionarios estudiantiles y escolares (OCDE, 2017a), “esto no supone una equivalencia con el nivel de aula, ya que los alumnos no se encuentran identificados en función del docente que les atiende” (Gamazo, 2019, pp. 246).

2.1.4.6 Comparabilidad intercultural

Por último, diversos autores han tratado el tema de la comparabilidad intercultural del estudio PISA (He et al., 2018; Marksteiner et al., 2018). Con un proyecto de carácter comparativo como el de la OCDE, se debe estar seguro de que se mide lo mismo. La falta de equivalencia de los instrumentos podría resultar en la incomparabilidad de las mediciones y, con ello, la incomparabilidad entre los países. Por un lado, Kankaraš y Moors (2014) encontraron que en PISA 2009 existieron moderadas faltas de equivalencia entre las mediciones efectuadas en regiones de una misma nación y entre países alrededor del mundo. Por otro lado, Arffman (2010) analizó tres textos en inglés y uno en finlandés utilizados en la prueba PISA 2000 y encontró que ninguno era totalmente equivalente en dificultad.

Entre los factores que podrían propiciar la incomparabilidad intercultural, la traducción de los instrumentos es uno de los más reconocidos. Tal como lo mencionó el Instituto Nacional para la Evaluación de la Educación (2011), “cuando un instrumento se traduce, es muy probable que se alteren los constructos en medición” (p. 9). Ante tal dilema, suelen existir diversos esfuerzos con el propósito de controlar el impacto de la traducción. No obstante, es complejo obtener una correspondencia perfecta entre las diversas lenguas en las que se traducen algunas pruebas. A este fenómeno se le denomina sesgo por traducción y puede favorecer a algunas de las poblaciones a las que se les aplique el instrumento.

2.2 Revisión de los Cuestionarios de Contexto

Los cuestionarios de contexto han permitido hacer análisis específicos respecto a los factores explicativos de las puntuaciones PISA. No obstante, en algunos casos los estudios pueden resultar limitados debido a la cantidad de variables recogidas por esta evaluación. En este apartado se pretende conocer y describir las variables que históricamente han sido de relevancia para PISA y que han sido incluidas en los cuestionarios de contexto.

Metodológicamente, se realizó un análisis documental de dos tipos de variables históricamente recogidas por los cuestionarios de contexto entre PISA 2000 y PISA 2018: factores contextuales estudiantiles y escolares, y factores relacionados con la enseñanza de los docentes. Para realizar la presente revisión, se analizaron documentos oficiales publicados por la OCDE: informes técnicos y cuestionarios de contexto. Las fases y exigencias de esta técnica se determinaron con base en Massot et al. (2009, pp. 351-352):

1. El rastreo e inventario de los documentos existentes y disponibles.
2. La clasificación de los documentos identificados.
3. La selección de los documentos más pertinentes para los propósitos de la investigación.
4. Una lectura en profundidad del contenido de los documentos seleccionados [...].
5. Una lectura cruzada y comparativa de los documentos en cuestión.

2.2.1 Resultados de la revisión de los cuestionarios de contexto

Como se puede observar en la Tabla 2, en lo que respecta a las variables contextuales estudiantiles y escolares, la mayoría de los indicadores son simples, es decir, provienen de información recabada por un único ítem. No obstante, el *índice social, económico y cultural* de los estudiantes, y los *recursos materiales* y *recursos humanos* de la escuela son indicadores compuestos contruidos a partir de información proveniente de diversos ítems.

Tabla 2

Histórico de las variables contextuales recogidas entre PISA 2000 y PISA 2018

Descripción de la variable	Año						
	2000	2003	2006	2009	2012	2015	2018
Estudiantes	Género	x	x	x	x	x	x
	Estatus inmigratorio	x	x	x	x	x	x
	Índice social, económico y cultural ⁱ	x	x	x	x	x	x
	Lenguaje hablado en el hogar	x	x	x	x	x	x
	Grado escolar	x	x	x	x	x	x
	Condición de repetidor		x		x	x	x
	Cambio escolar		x		x		x
Total	5	7	6	7	6	7	7
Directivos	Tipo de escuela	x	x	x	x	x	x
	Recursos materiales ⁱ	x	x	x	x	x	x
	Recursos humanos ⁱ	x	x	x	x	x	x
	Tamaño de la escuela	x	x	x	x	x	x
	Ratio estudiantes-docentes	x	x	x	x	x	x
	Tamaño de las aulas			x		x	x
Total	5	5	6	5	6	6	6

Nota: Elaboración propia con base en OCDE (2002a; 2002b; 2005a; 2005b; 2009a; 2009b; 2012; 2014; 2017b).

ⁱ = Índice compuesto con base en información procedente de diversos ítems.

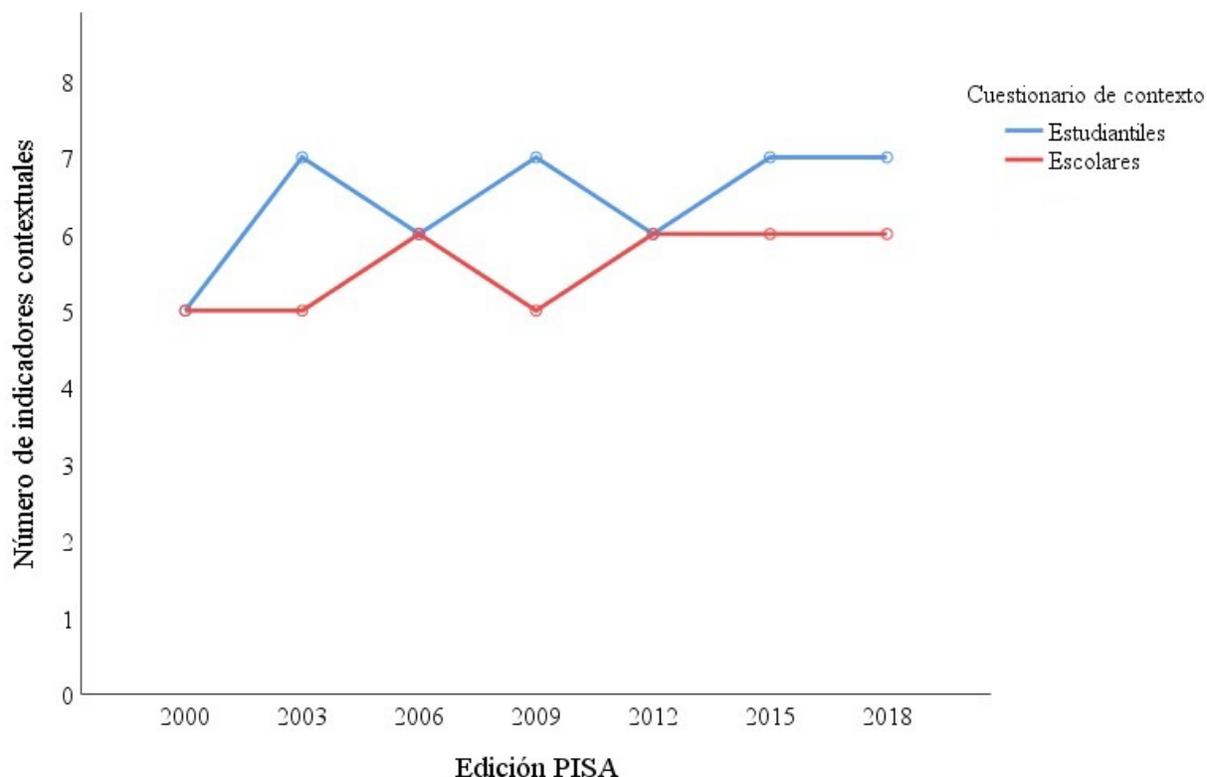
A su vez, por un lado, se puede identificar que existe un conjunto de indicadores entre PISA 2000 y PISA 2018 que permiten la comparabilidad de los fenómenos educativos a través del tiempo. En los cuestionarios aplicados a estudiantes se encuentra el *género*, el *estatus migratorio*, el *índice social, económico y cultural*, y el *lenguaje hablado en el hogar*. En cambio, en los cuestionarios dirigidos a directivos se halla el *tipo de escuela*, los *recursos materiales*, los *recursos humanos*, el *tamaño de la escuela* y el *ratio-estudiantes docentes*.

Por otro lado, se puede concluir que ha sido constante el interés de PISA por recoger información contextual -individual y grupal- a partir de los cuestionarios estudiantiles y escolares (ver Figura 2). Si bien en algunas implementaciones ha habido un menor número de indicadores

(en lo que refiere a los indicadores simples), en las últimas dos aplicaciones se ha brindado el número máximo de indicadores contextuales en el histórico de datos PISA.

Figura 2

Gráfico de líneas del número de indicadores contextuales entre PISA 2000 y PISA 2018



Nota: Elaboración propia con base en OCDE (2002a; 2002b; 2005a; 2005b; 2009a; 2009b; 2012; 2014; 2017b).

Respecta a las variables relacionadas con la enseñanza, como se aprecia en la Tabla 3 gran parte de los indicadores fueron compuestos y, por ende, contruidos a partir de más de una respuesta obtenida con los cuestionarios. Únicamente el *tiempo de aprendizaje por semana* y el *uso de las TIC* -ambos incluidos en los cuestionarios estudiantiles- fueron indicadores simples.

Tabla 3*Histórico de las variables de enseñanza recogidas entre PISA 2000 y PISA 2018*

Descripción de la variable	Año						
	2000	2003	2006	2009	2012	2015	2018
Estudiantes							
Clima disciplinario en la lección ^í	x	x		x	x	x	x
Relación docente-estudiante ^í	x	x	x	x			
Apoyo docente ^í	x	x			x	x	x
Presión docente en el logro académico ^í	x						
Tiempo de aprendizaje por semana		x				x	x
Enseñanza interactiva ^í			x				
Enfoque en la aplicación del conocimiento ^í			x				
Actividades prácticas en la lección ^í			x		x		
Aprendizaje basado en la investigación ^í			x			x	
Estimulación del docente ^í				x			x
Estructuración y andamiaje ^í				x			
Activación cognitiva ^í					x		
Manejo del aula ^í					x		
Orientación estudiantil ^í					x		
Uso de las TIC					x		x
Retroalimentación ^í					x	x	x
Instrucción directa ^í					x	x	x
Adaptación de la instrucción ^í						x	x
Interés del docente ^í							x
Total	4	4	5	4	9	7	9
Directivos							
Comportamiento docente ^í	x	x		x		x	x
Moral y compromiso docente ^í	x	x			x		
Consenso docente sobre la enseñanza ^í		x					
Enfoque del docente ^í					x		
Total	2	3	0	1	2	1	1

Nota: Elaboración propia con base en OCDE (2002a; 2002b; 2005a; 2005b; 2009a; 2009b; 2012; 2014; 2017b).

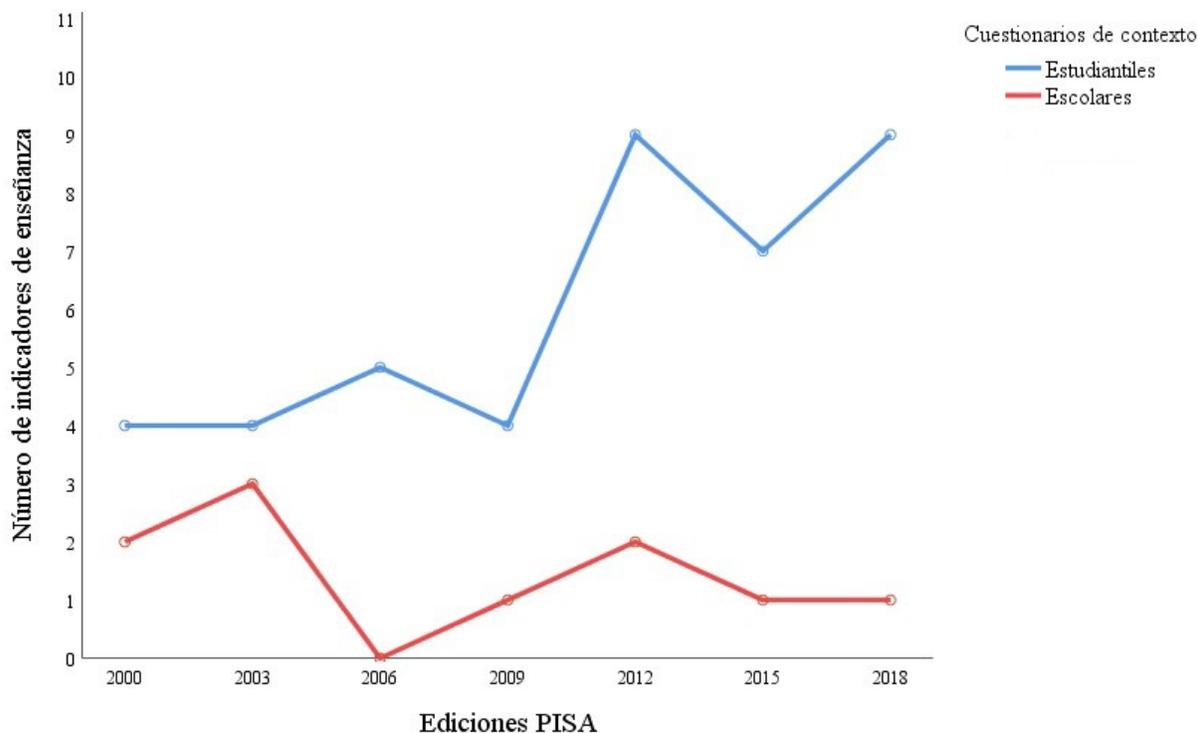
TIC = Tecnologías de la Información y Comunicación; ^í = Índice compuesto con base en información procedente de diversos ítems.

Además, se puede observar que la mayoría de los indicadores han cambiado de forma recurrente a lo largo de las diversas aplicaciones PISA. Únicamente el *clima disciplinario* -del cuestionario dirigido a estudiantes- y el *comportamiento docente* -del cuestionario dirigido a directivos- se han mantenido relativamente constantes, a excepción de la edición 2006 en el primer indicador y a las ediciones 2006 y 2012 en el segundo. De tal forma que podría mermarse la comparabilidad de los indicadores de la enseñanza a lo largo del tiempo.

Pese a la falta de comparabilidad entre ediciones PISA, se puede observar en el Figura 3 que existe un aumento en la recolección de información relativa a la enseñanza mediante la aplicación de cuestionarios dirigidos a los estudiantes. A partir de este incremento se pueden emitir dos inferencias: la primera es que existe un creciente interés de la OCDE por la recolección de información sobre las prácticas de enseñanza de los docentes a través de la opinión de los estudiantes; y la segunda es que los indicadores sobre la enseñanza han sido útiles para complementar los puntajes y comprender el fenómeno educativo de los países participantes en la evaluación PISA.

Figura 3

Gráfico de líneas del número de indicadores de enseñanza entre PISA 2000 y PISA 2018



Nota: Elaboración propia con base en OCDE (2002a; 2002b; 2005a; 2005b; 2009a; 2009b; 2012; 2014; 2017b).

Caso contrario en los indicadores sobre la enseñanza obtenidos a partir de las respuestas de los directivos, donde se disminuyó de entre dos y tres indicadores en los primeros dos ciclos (2000 y 2003) a un indicador en los últimos dos (2015 y 2018). Aunque esto podría interpretarse como una disminución del interés de la OCDE por recolectar datos sobre la enseñanza mediante la opinión de los directivos, es preciso considerar que en PISA 2015 y PISA 2018 se aplicó un nuevo cuestionario dirigido a docentes con el que se obtuvo la información que anteriormente se recolectaba a través el director.

De forma general, del análisis documental se pudo concluir lo siguiente:

1. En relación con el total de aplicaciones PISA a la fecha, en la última edición 2018 se cuenta con el número máximo de indicadores contextuales (estudiantiles y escolares), ello

permite realizar análisis más completos sobre el efecto de las características contextuales sobre el rendimiento de los estudiantes y robustece la estimación de residuos escolares al utilizar las variables contextuales como factores de ajuste.

2. Existe un conjunto de indicadores que permiten la comparabilidad contextual -estudiantil y escolar- a través del tiempo de las evaluaciones PISA. Ello permite el estudio del impacto de políticas educativas en busca de disminuir las segregaciones educativas en relación con el contexto de los estudiantes; si bien los avances en este tema son complejos de observar, la comparabilidad de las características contextuales a través del tiempo podría dar fe de ello.
3. Los indicadores de enseñanza han cambiado constantemente entre ediciones PISA. Aunque es natural que las evaluaciones internacionales se encuentren en constante evolución para su mejora, en el caso de PISA ha habido avances y retrocesos respecto a la cantidad de información recogida con los cuestionarios de contexto; a pesar de que en la edición 2018 se ha aumentado considerable el número de ítems en comparación con las primeras cuatro ediciones (2000, 2003, 2006 y 2009), se han eliminado indicadores que podrían resultar útiles para la explicación de los resultados obtenidos en PISA 2018.

2.3 Eficacia Educativa

De acuerdo con Burušić, Babarović y Velić (2016), no es tarea sencilla proporcionar una definición única sobre Eficacia Educativa. Esta debe ser adecuada a los diversos entornos, circunstancias y aspiraciones. Al retomar datos del proyecto PISA, en este trabajo la Eficacia Educativa refiere al grado de logro educativo sobre el nivel de competencia matemática, lectora y científica de los estudiantes, una vez consideradas las características contextuales de los estudiantes y las escuelas.

La investigación sobre Eficacia Educativa y Mejora (*Educational Effectiveness and Improvement, EEI*) se ha desarrollado con rapidez desde la década de 1970. Según Chapman et al. (2016), este cuerpo de conocimiento ha evidenciado ser el más reconocido y proliferante en el ámbito de la investigación educativa. Sus primeras revisiones de literatura dan certeza sobre su progreso exponencial: la revisión *The search for effective schools* de Reynolds (1982) tuvo únicamente 100 referencias; en cambio, el *International Handbook of School Effectiveness* de Teddlie y Reynolds (2000) refirió a 1,200 publicaciones.

La EEI dio forma a tres disciplinas distintivas e interrelacionadas, la Eficacia Escolar (*School Effectiveness, SE*), la Enseñanza Eficaz (*Teacher Effectiveness, TE*) y la Mejora Escolar (*School Improvement, SI*) (Chapman et al., 2016):

- El estudio de la Eficacia Escolar ha buscado determinar qué factores dentro de los centros educativos tienen un efecto positivo en el aprendizaje y en el desarrollo de los estudiantes.
- El estudio de la Enseñanza Eficaz ha indagado qué características, comportamientos y prácticas tienen los “buenos” maestros que propician el aprendizaje y el desarrollo de los escolares.
- El estudio de la Mejora Escolar ha pretendido determinar cómo las bases de conocimiento sobre la Eficacia Escolar y la Enseñanza Eficaz pueden transferirse a las escuelas y aulas para mejorar los resultados de los estudiantes.

Históricamente las tres líneas de investigación han divergido en metodología, valores base y propósitos generales. En el estudio de la Eficacia Escolar han predominado los métodos cuantitativos, con la recolección de datos a gran escala para determinar el residuo, el valor agregado (*value added*) y las principales características de las escuelas eficaces (Stoll y Sammons, 2007). Los datos sobre la escuela han focalizado en las características organizativas y

administrativas de los centros escolares, más que en la cultura escolar, entendida como “los patrones de significado transmitidos históricamente y que incluyen las normas, los valores, las creencias, las ceremonias, los rituales, las tradiciones, y los mitos comprendidos, quizás en distinto grado, por las personas miembros de la comunidad escolar” (Elías, 2015, p. 288). Esta disciplina asumió que los resultados educativos están dados y centró su atención en resultados académicos, los cuales eran responsabilidad de los sistemas educativos estatales y su personal (Chapman et al., 2016).

Por el contrario, en la Mejora Escolar preponderaron los estudios cualitativos, con la recolección natural de los datos al explorar la perspectiva de docentes y estudiantes sobre la escolarización. Esta disciplina dio lugar a una perspectiva crítica de las decisiones de los sistemas educativos estatales. No es de sorprender que las diferencias epistemológicas entre los estudios de Eficacia Escolar y Mejora Escolar propiciaran diversas confrontaciones entre investigadores, las cuales comenzaron a disiparse con la fusión de ambas perspectivas y la adopción de métodos mixtos en los estudios de la Mejora Escolar (Hopkins y Reynolds, 2001; Reynolds, 1988).

La Enseñanza Eficaz se desarrolló de forma independiente y general antes que las otras dos disciplinas: la Eficacia Escolar y la Mejora Escolar (Muijs et al., 2014). De acuerdo con Creemers (1994), aunque los estudios precursores sobre Enseñanza Eficaz mostraron semejanza con las propiedades intelectuales de la Eficacia Escolar y la Mejora Escolar, las tres disciplinas cristalizaron su vínculo entre finales década de los ochenta y el año 2000 debido a un gran número de estudios sobre la eficacia docente a cargo de estudiosos de la Eficacia Escolar asociados en el *Invisible College for Research on Teaching* (Brophy y Good, 1984); en este trabajo colaborativo se moldearon dos eruditos de la Enseñanza Eficaz, Bert Creemers y Sam Stringfield.

Con el tiempo, la Eficacia Escolar, la Enseñanza Eficaz y la Mejora Escolar comenzaron a transitar de una posición particular y diferencial a una perspectiva común y abarcativa sobre metodología, orientación, y propósito científico. A grandes rasgos, las tres disciplinas se encontraron influidas por:

- un énfasis en el empleo de un enfoque científico que asegure la validez de los datos sobre los diversos entornos educativos, escuelas y aulas;
- un propósito social con el uso del conocimiento científico obtenido para mejorar las condiciones educativas de los estudiantes, en especial aquellos más desfavorecidos;
- un compromiso por ofrecer a los actores educativos el conocimiento preciso para la mejora de sus prácticas; y
- un sentido de equidad al buscar abordar las necesidades de todos y cada uno de los niños para reducir las diferencias en el logro académico.

Según Chapman et al. (2016), la sinergia de estas tres disciplinas significó el surgimiento de un nuevo paradigma educativo, el paradigma en Eficacia Educativa y Mejora. La generación de este, a su vez, se vio favorecido por la creación de nuevas estructuras organizativas: en especial el *International Congress for School Effectiveness and Improvement* (ICSEI) en 1988, acompañado con la revista *School Effectiveness and School Improvement* (SESI) en 1990; la *Society for Research on Educational Effectiveness* (SREE) y su revista *Journal of Research on Educational Effectiveness* (JREE), y la evolución de diversos *Special Interest Groups* (SIGs) dentro de la *American Educational Research Association* (AERA) en los Estados Unidos, los SIGs en la *Educational Effectiveness of the British Educational Research Association* (BERA) y la *European Association for Research on Learning and Instruction* (EARLI), así como la red *Educational Effectiveness and Quality Assurance* de la *European Conference for Educational*

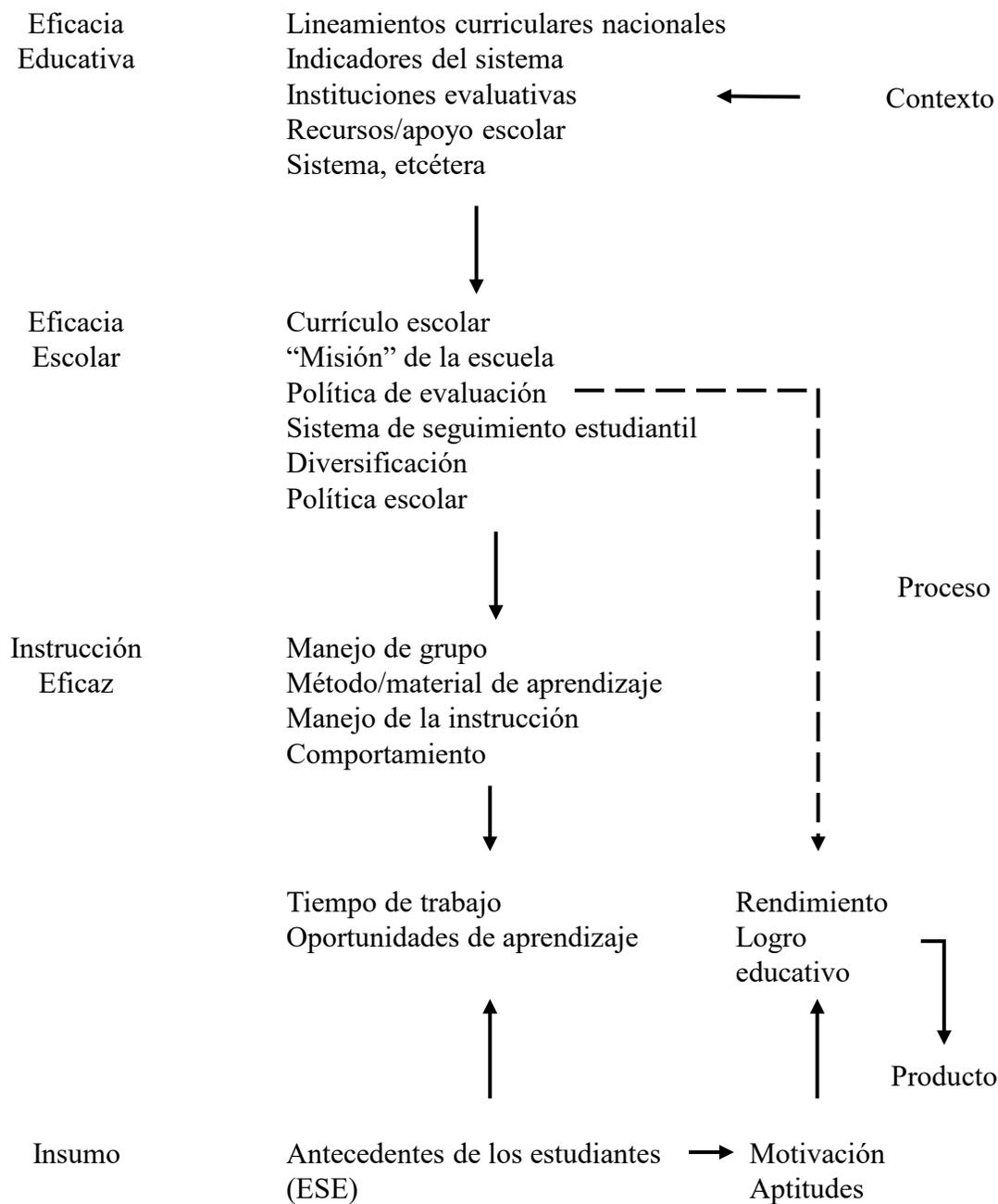
Research (ECER). De acuerdo con Reynolds et al. (2014), estas estructuras organizativas llevaron a la EEI a su internacionalización. Con ellas, se establecieron redes mundiales de colaboración, que dieron pie a la investigación conjunta en múltiples países. Así como la unión de iniciativas y enfoques de investigación entre la Eficacia Escolar y la Mejora Escolar. Tal es el caso del *International School Effectiveness Research Project* (ISERP), un estudio sobre los efectos diferenciales de la escuela en nueve países de cuatro continentes con el propósito de identificar los factores eficaces específicos -que su función depende del contexto- y universales -que su función es independiente al contexto- (Reynolds et al., 2002).

Históricamente, los estudiosos de la EEI se han comprometido con el desarrollo de teorías que expliquen los fenómenos educativos (p. ej. Creemers, 1991; Mortimore et al., 1988; Scheerens, 1991). Según Reynolds et al. (2014) más allá de la asociación de variables, tener una teoría explicativa es trascendental para contar con una estructurar base para comprender los fenómenos educativos, de tal forma que facilite el conocimiento a los nuevos interesados en este campo: inmersos y ajenos a la investigación, profesionales de la educación y encargados de la toma de decisiones político-educativas.

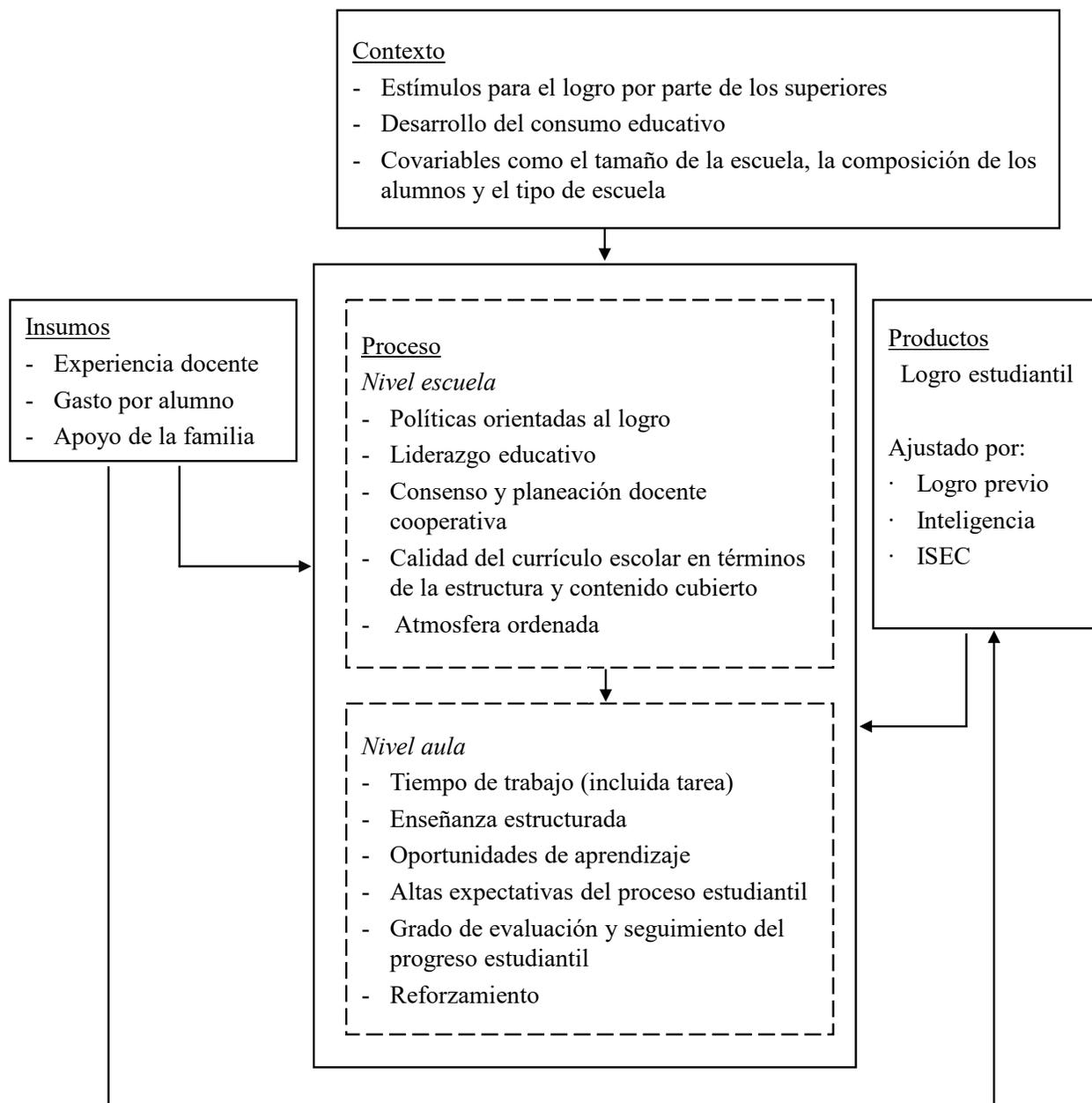
Inicialmente, el cuerpo de conocimiento sobre EEI se apoyó en teorías parciales para el desarrollo de una teoría global: (a) la teoría de contingencia empleada para la discusión de las variaciones contextuales; (b) la construcción de coaliciones de la ciencia política para discutir el liderazgo en las escuelas; (c) la psicología cognitiva y la teoría motivacional para las discusiones de la psicología de la instrucción; y (d) algunos acercamientos al orden causal (Bosker y Scheerens, 1994). Con base en el uso de estas, surgieron diversas aproximaciones teóricas sobre la Eficacia Educativa, entre ellas: el *Marco Conceptual para la Eficacia Educativa* de Creemers (1991) y el *Modelo CIPO* de Scheerens (1991), la Figura 4 y la Figura 5 respectivamente.

Figura 4

Marco conceptual para la eficacia educativa de Creemers (1991)



Nota: Adaptación al español de "Towards a theory on educational effectiveness" por B. P. M. Creemers, 1993, *International Congress for School Effectiveness and Improvement*. Norrköping, Suecia.
ESE = Estatus socioeconómico.

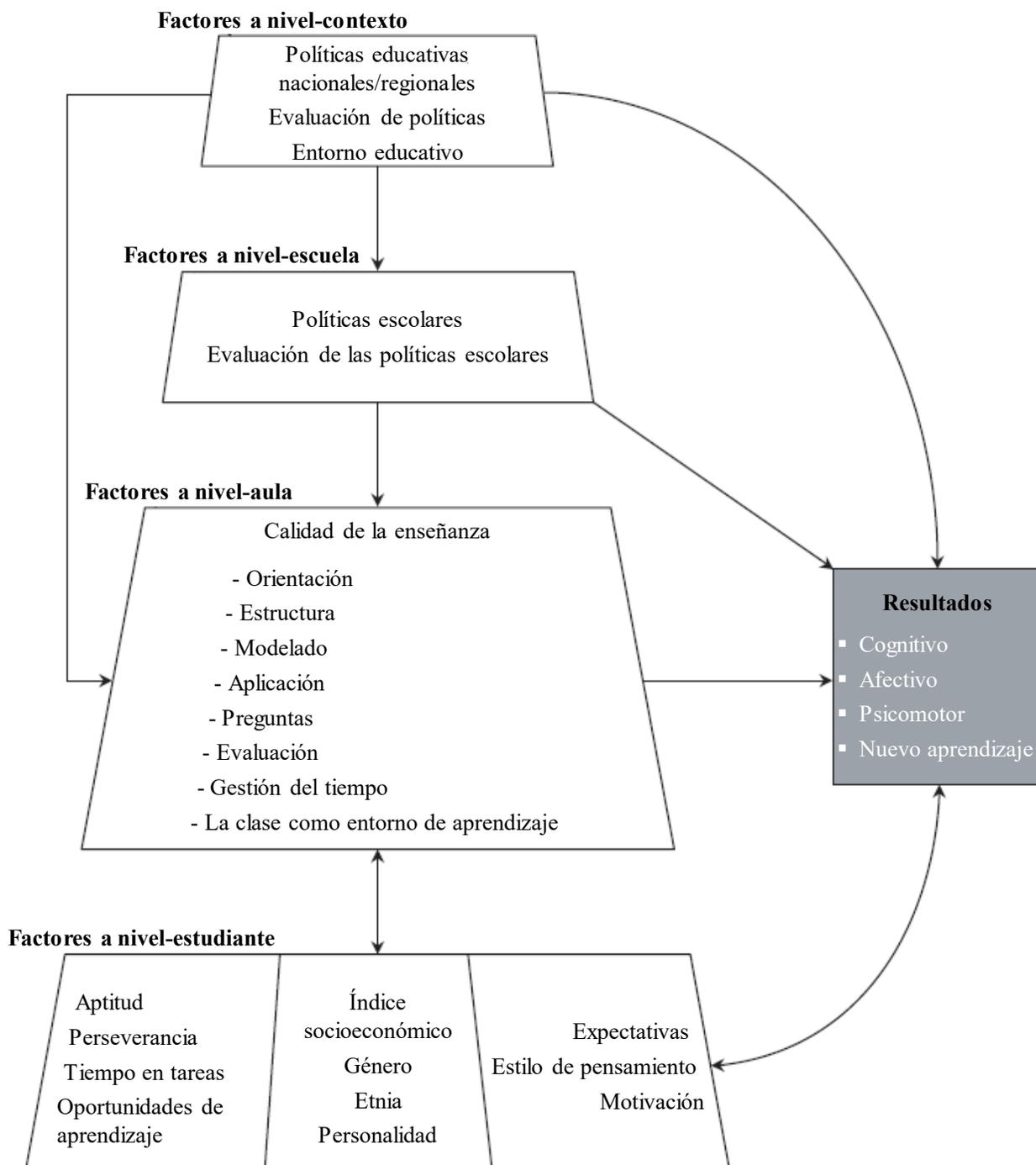
Figura 5*Modelo CIPO de Scheerens (1991)*

Nota: Adaptación al español de "Towards a theory on educational effectiveness" por B. P. M. Creemers, 1993, *International Congress for School Effectiveness and Improvement*. Norrköping, Suecia.

De acuerdo con Creemers (1993), pese a los avances, los aportes de Creemers (1991) y Scheerens (1991) no eran teóricas completas, sino marcos conceptuales como los mismos autores los denominan. La característica atórica de las primeras décadas de la investigación sobre EEI desató múltiples críticas, particularmente en los estudios relativos a la Eficacia Escolar y la Mejora Escolar (Creemers y Kyriakides, 2008). Según Scheerens y Bosker (1997), esta situación se agravaba con el infrecuente uso o apropiación de los modelos existentes en otros estudios empíricos similares de la época.

Recientemente, Creemers y Kyriakides (2008) establecieron la teoría dinámica sobre Eficacia Educativa (*The Dynamics of Educational Effectiveness*) con el propósito de explicar de forma sistemática e integrada la asociación entre variables. Esta incorpora y analiza simultáneamente los diversos niveles educativos: el contexto, la escuela, el aula y el estudiante. Se considera que los elementos que comprenden a los diversos niveles educativos tienen efectos directos e indirectos sobre los resultados de los estudiantes. Esta teoría ha estado a prueba en múltiples estudios y ha presentado resultados prometedores (Antoniou y Kyriakides, 2011, 2013; Azkiyah et al., 2016; Christoforidou et al., 2014; Christoforidou y Xirafidou, 2014; Creemers y Kyriakides, 2015; Kyriakides, 2008; Kyriakides et al., 2013; Kyriakides et al., 2013; Kyriakides y Creemers, 2008, 2009; Kyriakides et al., 2009; Panayiotou et al., 2015; Panayiotou et al., 2014).

Figura 6

Modelo Dinámico de Eficacia Educativa

Nota: Adaptación al español de "The Dynamics of Educational Effectiveness: A Contribution to Policy Practice and Theory in Contemporary Schools" de B. P. M. Creemers & L. Kyriakides, 2008, Routledge.

La Figura 6 representa la principal estructura del modelo dinámico de eficacia educativa. Como se puede apreciar, esta teoría se refiere a múltiples factores de eficacia que operan a diferentes niveles; por lo tanto, el modelo propuesto pertenece a un enfoque multinivel, dado que enfatiza la relación anidada entre los distintos niveles educativos (Creemers y Kyriakides, 2008). En esa secuencia, a continuación, se presenta el nivel escuela, y los factores que influyen en este nivel sobre las situaciones de aula. Y consecutivamente, lo que refiere a la enseñanza y sus factores asociados a los resultados de los estudiantes.

2.3.1 Eficacia escolar

De acuerdo a Scheerens (2016), la Eficacia Escolar se refiere al grado de logro de las metas en una escuela. Naturalmente, las metas escolares están sujetas a diversas circunstancias y pueden ser de diversa índole; los puntajes en pruebas objetivas han sido recurrentemente reconocidos como criterios de eficacia (Scheerens, 2013). En este estudio, al retomar datos proporcionados por el proyecto PISA, la Eficacia Escolar se concibe en el marco de la OCDE (2017c), el grado de éxito de las escuelas sobre el nivel de competencia lectora, matemática y científica de los estudiantes, una vez consideradas sus circunstancias contextuales.

La investigación sobre Eficacia Escolar cuenta con una extensa trayectoria. Sus orígenes pueden remontarse a los estudios de Coleman et al. (1966), quienes afirmaron que la contribución de la escuela al rendimiento de los estudiantes era insignificante. En estos estudios se afirmaba que poco importaba si la escuela contaba o no con suficientes recursos materiales, financieros o humanos. Lo que realmente incidía en el logro del estudiante era su grupo racial, su rendimiento previo, su condición socioeconómica y el nivel educativo alcanzado por sus padres.

A partir de estos primeros hallazgos, empezaron a surgir estudios diversos a profundidad sobre el proceder de las escuelas que tienen un aporte sobre el rendimiento de los estudiantes con

cualquier tipo de desventaja (racial, económica o familiar). Después de observar cuatro escuelas con esta descripción, Weber (1971) concluyó que en estas coexisten elementos comunes para su éxito: un fuerte liderazgo, altas expectativas, un buen ambiente, atención individualizada y evaluación minuciosa.

Años más tarde, Edmonds (1978) se interesó en el trabajo de Weber (1971) y se dedicó a describir a las escuelas ubicadas en zonas urbanas pobres con un aporte sobre el rendimiento de sus estudiantes. A este tipo de escuelas se les denominó Escuelas Eficaces (*Effective Schools*), definidas como aquellas que tenían “el poder o capacidad de producir efectos” sobre los resultados de los estudiantes (Compact Oxford English Dictionary, 1991, como se citó en Collen, 2003, p. 137).

Tras el trabajo realizado, Edmonds (1979) se propuso crear y propiciar que las escuelas con rendimiento bajo y rendimiento medio se convirtiesen en eficaces. Para ello, con base en la investigación generada a la fecha (especialmente la de Weber, 1971) elaboró un modelo de cinco factores: fuerte liderazgo administrativo, clima de altas expectativas, atmósfera ordenada y segura, énfasis en el aprendizaje, y evaluación constante. Este trabajo fue considerado pionero en la Investigación sobre Mejora Escolar (*School Improvement Research*, Teddlie y Reynolds, 2003).

Como se mencionó, estas investigaciones se realizaron principalmente en escuelas con estatus socioeconómico bajo. Lo cual, según Teddlie y Reynolds (2003), propició que las conclusiones derivadas de estos estudios tuviesen un sesgo en relación con el estatus socioeconómico de las escuelas y los estudiantes. Esta situación trajo consigo diversas críticas de la comunidad científica que apuntaban a la necesidad de reformar el método de estudio de la Eficacia Escolar y los enfoques analíticos utilizados hasta el momento.

Una década después se desarrollaron los modelos multinivel, un nuevo analítico más sofisticado que comprende la diversidad de los contextos. Estos estadísticos permitieron explorar y contrastar la eficacia en escuelas de los diversos estratos socioeconómicos, bajo, medio y alto; permitiendo así, la obtención de resultados más ajustados a la realidad (Amador y López-González, 2007). A partir de estos estadísticos se identificó la relevancia de las prácticas docentes en la eficacia de las escuelas; Stringfield et al. (1985), Teddlie et al. (1989), y Virgilio et al. (1991) evidenciaron que la eficacia del docente es inherente a la Eficacia Escolar, es decir, no puede haber eficacia en la escuela sin un docente eficaz.

Teddlie y Reynolds (2003) resumieron el desarrollo de la investigación en Eficacia Escolar en cuatro etapas cronológicas:

1. Insumo-Producto (entre mediados de la década los 60 y principios de los 70) en los que se estudiaron las relaciones entre los insumos escolares, como los recursos financieros, físicos y humanos, y el producto obtenido en el estudiante, la nota final del curso.
2. Insumo-Proceso-Producto (a partir de la década de los 70) en los que se examinaron los elementos clave en el proceso de las escuelas eficaces. A su vez, el producto se diversificó, es decir, se amplió la gama de posibles resultados (actitudinales, emocionales y académicos).
3. Mejora Escolar (entre finales de la década de los 70 y mediados de la década de los 80) en la que foco de atención estuvo en torno a la mejora de las escuelas mediante la incorporación del conocimiento científico generado a la fecha -sobre las escuelas eficaces- a través de programas de mejora.

4. Contexto-Proceso-Producto (entre finales de la década de los 80 a la actualidad) en la que se incorporaron factores contextuales al estudio de la eficacia y se desarrollaron enfoques analíticos más sofisticados que han abonado en la mejora de la producción de conocimiento científico.

2.3.1.1 Factores a nivel escuela

En la teoría dinámica, de Creemers y Kyriakides (2008), se espera que una escuela eficaz tenga efectos indirectos sobre los resultados de los estudiantes. Esta idea se originó a partir de diversos estudios que dieron evidencia de la elevada significancia del nivel aula en comparación con el nivel escuela; y que, además, la eficacia del centro dependía de la eficacia docente (Creemers, 1994; Kyriakides et al., 2000; Teddlie y Reynolds, 2000; Yair, 1997; entre otros). Por consiguiente, en esta teoría los factores a nivel escuela tienen la finalidad de incidir sobre el acontecer en el aula: la enseñanza; y, con ello, influir sobre el logro de los estudiantes.

El modelo dinámico integra cuatro factores globales a nivel escuela (Creemers y Kyriakides, 2008):

- Política escolar sobre la enseñanza y medidas adoptadas para mejorarla. Acciones escolares que tienen por objetivo facilitar la gestión del tiempo en el aula; mejorar la práctica docente; y abordar el absentismo estudiantil y docente.
- Evaluación de la política escolar respecto a la enseñanza y acciones tomadas para mejorarla. Acciones que tienen por objetivo la recolección información sobre la enseñanza durante el año escolar.
- Política escolar para crear un ambiente de aprendizaje y acciones para mejorarlo. Acciones escolares que regulan el comportamiento del estudiante fuera del aula; facilitan la colaboración e interacción entre docentes; propician el vínculo con la comunidad, padres

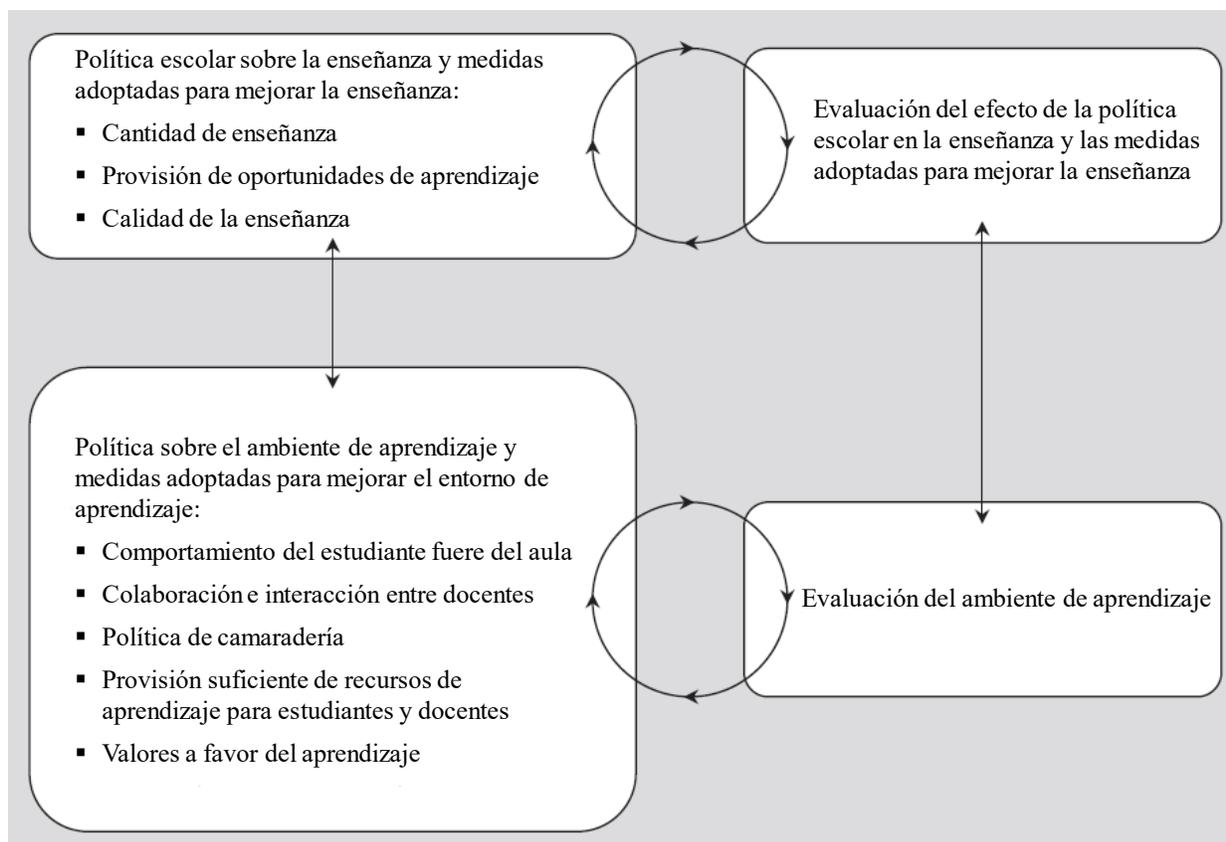
o tutores; aumentan la suficiencia de recursos de aprendizaje para estudiantes y docentes; y promueven los valores a favor del aprendizaje.

- Evaluación de la política escolar respecto al ambiente de aprendizaje. Al igual que en la evaluación de la política en enseñanza, configura las acciones escolares que tienen por finalidad la recolección de información sobre el ambiente de aprendizaje durante el año escolar.

Como se puede apreciar en la Figura 7, a diferencia de otros modelos, estos factores tienen un fuerte énfasis en el impacto sobre los resultados de los estudiantes a través de facilitar o mejorar la práctica de los docentes en el aula. Según Creemers y Kyriakides (2008), esto se debe a que la función principal de una escuela es el aprendizaje de los estudiantes que, usualmente, tiene lugar en el aula. Es así que en este modelo el componente central es eficacia de la enseñanza y la profesionalización de la labor docente.

Figura 7

Factores a nivel escuela del Modelo Dinámico de Eficacia Educativa



Nota: Adaptación al español de “*The Dynamics of Educational Effectiveness: A Contribution to Policy Practice and Theory in Contemporary Schools*” de B. P. M. Creemers & L. Kyriakides, 2008, Routledge.

2.3.1.2 Identificación de escuelas eficaces

De acuerdo con Fernández (2003), en el estudio de la Eficacia Escolar, la identificación empírica de un conjunto de escuelas denominadas eficaces ha tenido un papel fundamental. Pese a ser una práctica recurrente, desde Weber (1971) y Edmonds (1978) a la actualidad, aún no existe un consenso metodológico en términos y medidas base del supuesto de eficacia de las escuelas. Sammons et al. (2016), por ejemplo, delineó al menos nueve formas de concebir y medir la eficacia:

- La eficacia de la escuela en forma de un “puntaje medio bruto”, expresada como una desviación del gran promedio o de la eficacia media de las escuelas. Si bien, esta vía es sencilla, este tipo de eficacia cuanta con diversas limitaciones al no reconoce la diversidad que envuelve a los centros en términos de características contextuales escolares y estudiantiles y, por lo tanto, las comparaciones entre escuelas son poco justas.
- El progreso medio de los estudiantes en un tiempo determinado en comparación con un rendimiento previo. Esta eficacia se estima a partir de un análisis de regresión en el que se controla el rendimiento previo y los antecedentes de los estudiantes (como su estatus socioeconómico, género, origen étnico, etc.).
- Efecto absoluto. Este tipo de eficacia se mide a partir de grupos de control que no reciben escolaridad.
- El impacto de la escolarización en el rendimiento promedio de todos los estudiantes en una escuela, ajustados por logros previos y el estatus socioeconómico.
- Eficacia Tipo A y Tipo B. La primera consiste en la diferencia entre el rendimiento real de un estudiante y el rendimiento que se hubiese esperado si el niño hubiera asistido a una escuela típica, entendida como aquella que cuenta con estudiantes con los mismos antecedentes y aptitudes. La segunda refiere a la diferencia entre el desempeño real de un estudiante en una escuela en específico y el desempeño esperado si el estudiante hubiese asistido a una escuela con características contextuales idénticas, pero con eficacia promedio.
- El efecto de diferentes escuelas en el rendimiento del estudiante a lo largo del tiempo. En esta eficacia, por ejemplo, se compara el crecimiento del aprendizaje entre escuelas primarias, secundarias y preparatorias.

- Tamaño de la eficacia escolar medido mediante el coeficiente de correlación intraclase (ICC, por sus siglas en inglés). Este tipo de eficacia implica la aplicación de modelos multinivel y, por ejemplo, la comparación entre el ICC inicial obtenido con el modelo nulo y el ICC final obtenido con el modelo agregado contextualizado, en el que se controla el logro previo y características contextuales de los estudiantes.
- Efectos diferenciales o variación dentro de la escuela. Estos se identifican a partir la variación en los efectos escolares individuales para diferentes grupos de estudiantes (por ejemplo, hombres y mujeres, e índice socioeconómico bajo, medio y alto), obtenidos a partir de modelos multinivel por pendientes aleatorias ajustadas.
- Eficacia escolar individual basada en estimaciones residuales a través de modelos multinivel. Se identifica a las escuelas y/o estudiantes que logran resultados por encima o por debajo de lo que sería previsto una vez controladas diversas características escolares y estudiantiles.

Si bien la elección de una tipología u otra puede depender de los datos disponibles y el propósito de la investigación, los estudiosos de la eficacia han propuesto diversos criterios sustantivos para disminuir el grado de error (o probabilidad de elegir falsos positivos) en la identificación de escuelas eficaces: (a) la ubicación del centro por encima (o por debajo) de un corte en la nube de puntos (Bartau et al., 2017; Gamazo et al., 2018; Intxausti et al., 2016; Joaristi et al., 2014; Martínez-Abad, 2019); (b) la posición educativa del centro en comparación con los pares (Rutledge et al., 2015; Timmermans et al., 2013; Walker et al., 2014); y (c) la diferencia significativa en comparación con la media (Sulis y Porcu, 2015).

2.3.2 Enseñanza eficaz

La Enseñanza Eficaz ha sido referida de diferentes formas a través del tiempo: los comportamientos docentes como su calidez y cortesía (p. ej. Ornstein y Levine, 1981); los conocimientos de los maestros respecto a su materia y sus estudiantes (p. ej. Hill et al., 2005; Mayer y Marland, 1997); y las creencias de los enseñantes, por ejemplo, sobre la inclusión educativa (p. ej. Stanovich y Jordan, 1998). En este estudio, la Enseñanza Eficaz refiere a la capacidad de lograr la meta escolar planificada y, por ende, mejorar los resultados de los estudiantes (Campbell et al., 2004; Good et al., 2009).

Esta línea de investigación cuenta con una larga historia. Germinó y evolucionó de forma independiente antes que las otras dos disciplinas: la eficacia y la mejora de las escuelas (Chapman et al., 2016). Robinson (2004) encontró que el estudio de la Enseñanza Eficaz puede remontarse a las primeras décadas del siglo pasado en Reino Unido, donde intentaron desarrollar los primeros modelos para transitar de una práctica de enseñanza artesanal a una con base en el conocimiento científico de la época. Los debates encabezados por expertos y los estudios embrionarios del trabajo docente fueron de suma importancia en la búsqueda del consenso profesional sobre los principios racionales de una enseñanza eficaz ubicados en el nexo entre la teoría y la práctica. Producto de ello, en ese periodo surgieron cinco amplias propuestas:

- a) planeación y preparación minuciosa con base en saberes sólidos de la materia;
- b) comprensión de las diferentes formas en que interactúan la enseñanza y lo enseñado;
- c) construcción lógica, sistemática, manejable y ordenada de la lección;
- d) habilidades básicas de enseñanza, como exposición, narración, e ilustración; y,
- e) el poder o presencia personal del docente al enseñar.

No obstante, estas y otras propuestas no fueron retomadas sino hasta finales del siglo XX. Según Reynolds et al. (2014), pese a ser tradicionalmente paralelas, el estudio de la Enseñanza Eficaz se vio eclipsado por la investigación sobre Eficacia Escolar. Esto debido principalmente a que gran parte de las primeras investigaciones en educación se enmarcaron casi de forma exclusiva en la eficacia de las escuelas, prestando así menor atención a lo que acontece en las aulas: la enseñanza.

Como lo sostuvieron Creemers (1994), Teddlie y Stringfield (1993) a la enseñanza le tomó un par de décadas más recibir la misma atención que la dada a la escuela. De acuerdo a Muijs et al. (2014), esto sucedió hasta la década de los ochenta, en la que se ha demostrado consistentemente que, no solo a nivel aula se puede explicar más variabilidad de los resultados estudiantiles que a nivel escuela (Kyriakides et al., 2000; Scheerens y Bosker, 1997; Teddlie y Reynolds, 2000; Yair, 1997), sino que también la varianza a nivel escuela y a nivel aula puede explicarse por lo que hacen los docentes: enseñar (Creemers, 1994; Muijs y Reynolds, 2011, 2017).

Con base en Campbell et al. (2003) y Kyriacou (2009), es posible identificar cinco etapas cronológicas en la investigación sobre Enseñanza Eficaz:

1. Presagio-Producto (1930-1940), en la que se estudiaron las características individuales de los docentes sobre el aprendizaje de los estudiantes, como su personalidad y actitudes (Schofield y Start, 1979; Silverman, 1988).
2. Estudios experimentales (1940-1960), en los que se investigaron los efectos de los distintos estilos de enseñanza sobre el aprendizaje (Anderson et al., 1979; Hatcher, 1941).
3. Proceso-producto (1960-1980), en los que se examinó la influencia de los comportamientos de los docentes en las aulas sobre el logro y progreso de los alumnos, por ejemplo, la

comunicación y el ritmo al instruir (Brophy y Evertson, 1974; Medley, 1977; Spada, 1987; Waxman y Fash, 1983).

4. Conocimiento y creencias de los docentes (1990-2000), en los que se exploraron las relaciones del conocimiento de la materia, el conocimiento pedagógico, y las creencias como la autoeficacia y expectativas sobre el logro de los estudiantes (Calderhead, 1996; Gallagher, 1991; Hativa, 2000).
5. Contexto-proceso-producto (2000 a la actualidad), en la que se analizan los factores a nivel aula que tienen efecto sobre el grado de logro de los objetivos socialmente acordados para la labor docente, de forma particular, pero no exclusiva, los asociados al aprendizaje del estudiante; por ejemplo, el clima del aula y la adaptación de la instrucción (Azigwe et al., 2016; Obilor, 2019; Prasertcharoensuk et al., 2018).

Basados en los principales hallazgos de cada etapa cronológica, Creemers y Kyriakides (2008) integraron un modelo teórico/empírico que analiza los diferentes niveles de anidamiento en el ámbito educativo (estudiante, aula, escuela y sistema) y tiene por núcleo la práctica docente: el Modelo Dinámico sobre Eficacia Educativa. De acuerdo a Muijs et al. (2014), este modelo comprende la importancia del nivel aula como predictor de los resultados de los estudiantes y de la eficacia educativa, otorgándole así, un fuerte énfasis al rol del docente y, por ende, a su enseñanza. Al tener en cuenta esto, a continuación, se describen los factores a nivel aula propuestos por estos autores.

2.3.2.1 Factores a nivel aula

De acuerdo con Kyriakides et al. (2018), en la teoría dinámica los elementos a nivel aula refieren a ocho factores que han evidenciado su asociación consistente con los resultados escolares y estudiantiles (ver Figura 8): orientar, estructurar, modelar, aplicar, cuestionar, evaluar, gestionar

el tiempo y propiciar un entorno de aprendizaje. Esta teoría (como la conciben Chapman et al., 2016) circunscribe factores que aluden a los modelos de enseñanza directa y dominio del aprendizaje (incluidos en Joyce et al., 2000), y perspectivas relativas tanto al constructivismo, como el aprendizaje activo (incluidas en Simons et al., 2002).

Figura 8

Factores a nivel aula del Modelo Dinámico de Eficacia Educativa



Nota: Adaptación al español de “*The Dynamics of Educational Effectiveness: A Contribution to Policy Practice and Theory in Contemporary Schools*” de B. P. M. Creemers & L. Kyriakides, 2008, Routledge.

Orientar. Se refiere a los comportamientos del docente que propician que los estudiantes identifiquen las razones por las cuales se presentan, realizan o producen, ciertas lecciones y actividades: (a) proveer los objetivos de una tarea/lección específica; y, (b) desafiar a los estudiantes a identificar la razón por la cual se lleva a cabo una actividad. De ello, se espera que los estudiantes perciban como significativo lo que acontece en el aula y, por ende, aumenten su participación y motivación en su aprendizaje. Este factor ha evidenciado estar significativamente relacionado con los resultados de los estudiantes (Azigwe et al., 2016) y ser un fuerte predictor de

su autorregulación, como la realización de actividades planificadas y el seguimiento cuidadoso de la clase (Yerdelen y Sungur, 2018).

Estructurar. Se refiere a los vínculos que determina el docente entre las diferentes partes de la lección: (a) comenzar con resúmenes y/o revisión de objetivos; (b) esbozar el contenido que se cubrirá y señalar las transiciones entre las partes de la lección; (c) recalcar las ideas clave; y, (d) revisar las ideas principales al final de la lección. De acuerdo con los hallazgos de Peculea y Peculea (2018), los estudiantes perciben este factor como un facilitador en el alcance de los objetivos educativos establecidos.

Modelar. Se refiere a que el docente debería promover en los estudiantes habilidades de orden superior que: (a) ayuden a la resolución de problemas; y, (b) faciliten la organización de su propio aprendizaje (por ejemplo, su autorregulación y aprendizaje activo). En ese sentido, la investigación ha demostrado que los docentes eficaces propician el uso o, en su defecto, el desarrollo de estrategias para resolución de diferentes tipos de situaciones (Sumantri et al., 2018; Yew y Goh, 2016).

Aplicar. Se refiere a proporcionar a los estudiantes: (a) la oportunidad de practicar o aplicar lo aprendido; y (b) utilizar la práctica o aplicación como punto de partida del siguiente aprendizaje. De acuerdo con Kirschner (2002), aprender no debería ser un proceso constante e ininterrumpido de únicamente la adquisición de información; los docentes eficaces utilizan el trabajo de asiento, o tareas en pequeños grupos, para proporcionar la práctica necesaria y ofrecer retroalimentación directa.

Cuestionar. Se refiere a la capacidad del docente para: (a) plantear preguntas de diferente nivel de dificultad (es decir, proceso y producto), contemplando sus características personales y contextuales; (b) ofrecer el tiempo suficiente para dar respuesta y/o discutir en clase; y, (c) tratar

con las múltiples respuestas de los estudiantes. Según Hill (2016) y Yang (2017), realizar diferentes preguntas durante la lección aumenta la participación, promueve la verbalización y desarrollo de ideas, ayuda a aclarar conceptos y facilita en la evaluación de la comprensión de los estudiantes.

Evaluar. Se refiere a partes, que se conciben, fundamentales en el inicio, transcurso y final de la enseñanza, en la que se: (a) emplean técnicas apropiadas para recopilación de la información sobre los conocimientos, habilidades y necesidades de los estudiantes; (b) retroalimenta a los estudiantes en sus avances y áreas de oportunidad; y, (c) valora la propia práctica docente. De acuerdo con de Jong et al. (2004), Kyriakides (2005) y Shepard (1989), este factor se ha asociado significativamente con la eficacia en todos los niveles, aula, escuela y sistema.

Gestionar el tiempo. Se refiere a la cantidad de tiempo utilizado para el cumplimiento de la meta educativa. En se sentido, los docentes eficaces: (a) priorizan las actividades; (b) organizan y administran el entorno; y, (c) maximizar las tasas de participación. Este factor, opera a diferentes niveles, y es considerado uno de los más importantes dentro de la teoría dinámica, pues comprende las oportunidades de aprender y el tiempo en la tarea de los estudiantes. Este, a su vez, se ha asociado significativamente con la participación de los estudiantes y el rendimiento de los estudiantes en todas las áreas.

Propiciar un entorno de aprendizaje. Se refiere a la calidad del ambiente para el cumplimiento de la meta educativa. Este factor consta de cinco componentes, recurrentemente declarados como los más importantes en estudios empíricos y metaanálisis: (a) interacción docente-estudiante; (b) interacción estudiante-estudiante; (c) el trato del docente hacia los estudiantes; (d) la competencia entre estudiantes; y, (e) el desorden en el aula. En la literatura, los

dos primeros componentes han sido considerados indicadores de clima de aula, mientras que los otros tres han aludido a los esfuerzos del docente por gestionar el clima del aula.

Los factores y principales elementos del modelo dinámico se resumen en la Tabla 4.

Tabla 4

Principales elementos de los factores de enseñanza del Modelo Dinámico

Factor	Principales elementos
Orientar	Proveer los objetivos de una tarea/lección específica Desafiar a los estudiantes a identificar la razón por la cual se lleva a cabo una actividad
Estructurar	Comenzar con resúmenes y/o revisión de objetivos Esbozar el contenido que se cubrirá y señalar las transiciones entre las partes de la lección Recalcar las ideas clave Revisar las ideas principales al final de la lección
Cuestionar	Plantear preguntas de diferente nivel de dificultad (es decir, proceso y producto), contemplando sus características personales y contextuales Ofrecer el tiempo suficiente para dar respuesta y/o discutir en clase Tratar con las múltiples respuestas de los estudiantes
Modelar	Ayudar a los estudiantes a usar o, en su defecto, desarrollar estrategias de resolución de problemas Promover habilidades que faciliten organizar el propio aprendizaje (p. ej., la autorregulación y el aprendizaje activo)
Aplicar	Ofrecer oportunidades de práctica o aplicación del aprendizaje Emplear la práctica o aplicación como punto de partida para el siguiente aprendizaje
Ambiente de aprendizaje	Establecer reglas para tratar con el desorden y la competencia estudiantil Ejercitar activamente el comportamiento de los estudiantes, al interactuar con y entre ellos
Manejar el tiempo	Organizar y administrar el entorno Maximizar las tasas de participación
Evaluar	Emplear técnicas apropiadas para recopilar datos sobre los conocimientos y habilidades de los estudiantes. Utilizar los insumos de la evaluación para evaluar la práctica docente. Utilizar los insumos de la evaluación para retroalimentar a los estudiantes.

2.4 Revisión Sistemática de Literatura de las Variables de Enseñanza Asociadas a los Resultados de PISA

Desde la aplicación 2000 a la actualidad, en PISA ha existido un creciente interés por recoger información relativa a la enseñanza a través de diferentes cuestionarios de contexto: estudiantiles, escolares y, en las últimas dos ediciones, docentes. A partir de dicha información, investigadores independientes se han dado a la tarea de realizar análisis multinivel de PISA en busca de las principales explicaciones de los resultados. En este apartado se presenta un mapeo y una Revisión Sistemática De Literatura (SRL, por sus siglas en inglés) con el objetivo de conocer cómo se han realizado los análisis multinivel de PISA e identificar qué variables de enseñanza han generado mayor interés a lo largo del tiempo y cuáles han sido las más asociadas con los resultados estudiantiles y escolares de PISA.

De acuerdo con Petersen et al. (2015), aunque los mapeos y las Revisiones Sistemáticas de Literatura comparten puntos en común (como, el proceso de búsqueda y selección), difieren en objetivos y enfoques analíticos. Los mapeos se diseñan para ofrecer una visión amplia de un tema estudiado a través de la clasificación y recuento de las contribuciones. Mientras que las Revisiones Sistemáticas de Literatura se realizan con el propósito de profundidad en la evidencia y la solidez de la misma (por ejemplo, cómo han depurado los datos, cómo se han analizado y qué han encontrado).

2.4.1 Metodología del mapeo y SRL

A continuación, se describe la metodología empleada para la realización del mapeo y la Revisión Sistemática de Literatura: preguntas de investigación, criterios de inclusión, criterios de exclusión, criterios de calidad, fuentes y cadenas de búsqueda utilizadas.

2.4.1.1 Preguntas de investigación

Las preguntas de investigación orientaron y dieron forma al mapeo y a la Revisión Sistemática de Literatura, de tal forma que se generaron dos grupos de preguntas. El primer grupo, correspondió al mapeo y atendió a aspectos globales para proporcionar una visión general de los estudios multinivel de PISA:

- PM1: ¿Cuántos estudios se han publicado entre el año 2000 y 2020?
- PM2: ¿Cuáles autores han sido los más activos en el tema?
- PM3: ¿Cuáles son las revistas en las que más se han publicado?
- PM4: ¿Qué datos de países han sido los más analizados?
- PM5: ¿Qué competencia (lectora, matemática y científica) ha sido la de mayor interés?
- PM6: ¿Qué ciclo PISA ha sido el más estudiado?

El segundo grupo de preguntas concernió a la Revisión Sistemática de Literatura con el propósito de profundizar en las variables, decisiones metodológicas y limitaciones declaradas en los estudios multinivel de PISA:

- PS1: ¿Cuáles han sido variables de enseñanza más asociadas con los resultados PISA?
- PS2: ¿En qué tanto por ciento se aplican los modelos multinivel por pendientes fijas y por pendientes aleatorias?
- PS3: ¿Con qué frecuencia se opta por todos los valores plausibles proporcionados en PISA?
- PS4: ¿Con qué frecuencia utilizan máxima verosimilitud restringida y máxima restringida?
- PS5: ¿Con qué frecuencia utilizan pesos muestrales escolares y/o estudiantiles?
- PS6: ¿Cuáles han sido los coeficientes de correlación intraclase reportados?
- PS7: ¿Cuáles son las limitaciones de PISA más reportadas?

2.4.1.2 Criterios de inclusión y criterios de exclusión

Una vez construidas las cuestiones de investigación, se definieron los criterios de inclusión y exclusión. En los criterios de inclusión (CI) se enunciaron las características que debían cumplir los estudios para ser seleccionados:

- CI1: La publicación es un artículo publicado en una revista indexada con revisión por pares.
- CI2: La versión completa de la publicación está disponible a través de la suscripción de nuestra institución.
- CI3: La publicación está escrita en inglés.
- CI4: La publicación presenta como variable criterio el rendimiento estudiantil o escolar en el *Programme for International Student Assessment*.
- CI5: La publicación presenta como variable predictora alguna de las variables de enseñanza proporcionadas en el *Programme for International Student Assessment*.
- CI6: La publicación describe la aplicación de estadísticos multinivel.
- CI7: La publicación es la más reciente y completa de las publicaciones relacionadas sobre el mismo estudio.

En cambio, en los criterios de exclusión (CE) se expusieron las características que limitaron la participación de algunos estudios:

- CI1: La publicación no es un artículo publicado en una revista indexada con revisión por pares.
- CI2: La versión completa de la publicación no está disponible a través de la suscripción de nuestra institución.
- CI3: La publicación no está escrita en inglés.

- CI4: La publicación no presenta como variable criterio el rendimiento estudiantil o escolar en el *Programme for International Student Assessment*.
- CI5: La publicación no presenta como variable predictora alguna de las variables de enseñanza proporcionadas en el *Programme for International Student Assessment*.
- CI6: La publicación no describe la aplicación de estadísticos multinivel.
- CI7: La publicación no es la más reciente y ni completa de las publicaciones relacionadas sobre el mismo estudio.

2.4.1.3 Criterios de calidad (CC)

Según Beltrán (2005), es necesario determinar criterios de calidad para evitar la integración de estudios que aumenten el sesgo e inhiban la comparabilidad de los estudios. En este trabajo, se construyó una rubrica con siete criterios de calidad con la intención de valorar de la forma más objetiva posible la calidad de los estudios (ver Tabla 5).

Tabla 5

Criterios de Calidad del Mapeo y de la Revisión Sistemática de Literatura

	Criterio de calidad	Puntuación
CC1	¿Los objetivos o preguntas del estudio están descritos de forma clara?	S/N/Parcial
CC2	¿Los datos del <i>Programme for International Student Assessment</i> permiten abordar el objetivo del estudio?	S/N/Parcial
CC3	¿Emplean los nombres originales de las variables del <i>Programme for International Student Assessment</i> ?	S/N/Parcial
CC4	¿El método del estudio está claramente escrito?	S/N/Parcial
CC5	¿El método es adecuado para responder a los objetivos/preguntas del estudio?	S/N/Parcial
CC6	¿Se responde a los objetivos/preguntas del estudio?	S/N/Parcial
CC7	¿Las conclusiones se basan en los resultados expuestos?	S/N/Parcial

Nota: S = Sí; N = No.

2.4.1.4 Bases de datos y cadenas de búsqueda

Por último, se seleccionaron las bases de datos y las cadenas de búsqueda. Semejante a Gamazo (2019), las bases de datos se eligieron con base en tres criterios: acceso a través de los medios institucionales disponibles; admisión del uso de cadenas lógicas de búsqueda; y permisividad de búsqueda en texto completo y/o campos clave (título, resumen, palabras clave, etc.) de la publicación. Considerando estos criterios, se optó por dos bases de datos: EBSCO y Scopus.

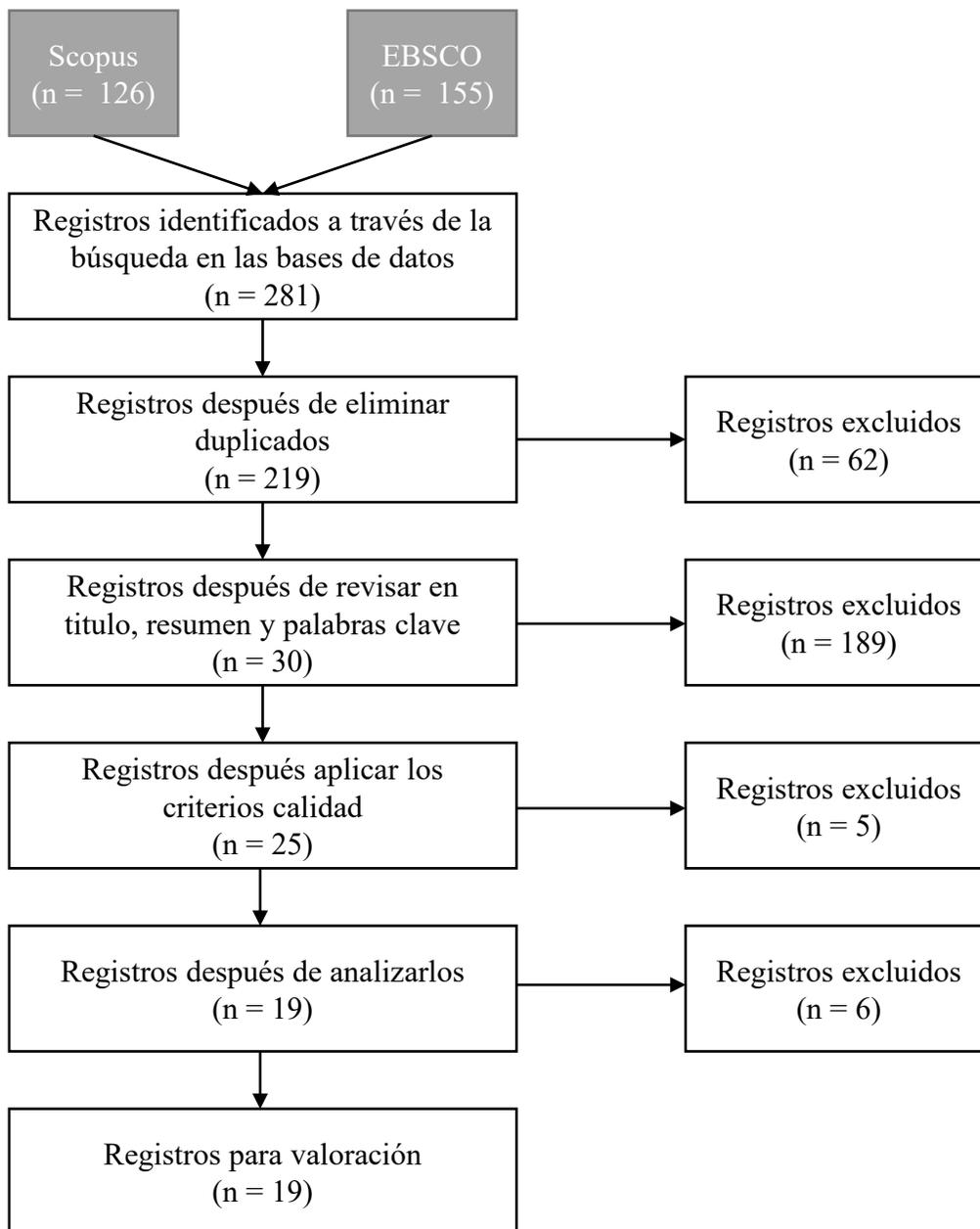
Las cadenas de búsqueda se determinaron con base en revisiones previas de estudios multinivel de PISA y la adaptación a las características de cada base de datos. De tal manera que en EBSCO se solicitaron únicamente publicaciones académicas y se empleó la cadena: ((“hierarchical linear models” OR “multilevel models” OR “multilevel analysis”) AND (pisa)). Mientras que en Scopus se utilizó la cadena: TITLE-ABS-KEY ((“hierarchical linear models” OR “multilevel models” OR “multilevel analysis”) AND (pisa)).

2.4.2 Extracción de datos del mapeo y SRL

Una vez definidas las especificaciones metodológicas, se ejecutaron las consultas en las bases de datos seleccionadas, se registraron las publicaciones identificadas, se eliminaron los registros duplicados, se aplicaron los criterios de inclusión y exclusión, se evaluó la calidad de los estudios, se analizaron los documentos y se determinó el número final de artículos a valorar. Los resultados de este proceso se pueden ver en la Figura 9.

Figura 9

Proceso de extracción de datos para el mapeo y Revisión Sistemática de Literatura



2.4.3 Informe del mapeo y SRL

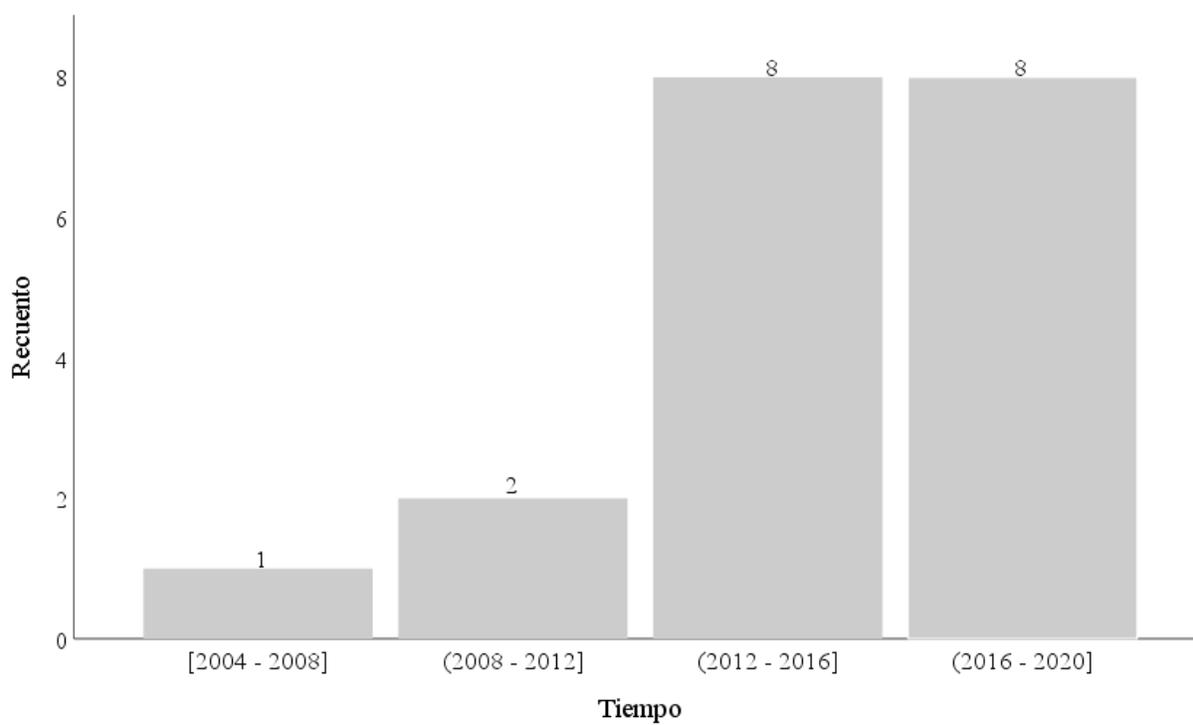
A continuación, se da respuesta a las preguntas de investigación a través del análisis de las 19 publicaciones. Los resultados se presentan en función del orden de las interrogantes, primero las correspondientes al mapeo y después las referentes a la Revisión Sistemática de Literatura.

2.4.3.1 Resultados del mapeo

Para dar respuesta a la primera pregunta (PM1), sobre los estudios publicados a lo largo del tiempo, se analizó el número artículos publicados. Como se observa en la Figura 10, los primeros análisis multinivel de PISA se encontraron entre el año 2004 y 2012 (Areepattamannil y Kaur, 2012; Ma y Crocker, 2007; Yildirim, 2012); no fue sino hasta después del segundo año mencionado (2012) que aumentó de forma exponencial la aplicación de estadísticos multinivel en datos PISA (Areepattamannil, 2014; Bove et al., 2016; Cairns y Areepattamannil, 2017; Chen y Cui, 2019; Cordero et al., 2015; Eickelmann et al., 2016; Gamazo et al., 2017; Grabau y Ma, 2017; Guo et al., 2018; Lau y Ho, 2015; Lau y Lam, 2017; Martínez-Abad et al., 2020; Meggiolaro, 2017; Mikk et al., 2015; Ning et al., 2015; Shera, 2014).

Figura 10

Recuento de los análisis multinivel de datos PISA



Respecto a los autores más activos en el tema (PM2), se identificaron 39 autores en las 19 publicaciones. Como se plasma en la Tabla 6, Areepattamannil fue el autor de mayor referencia con tres participaciones en investigaciones multinivel de datos PISA (Areepattamannil, 2014; Areepattamannil y Kaur, 2012; Cairns y Areepattamannil, 2017), seguido por Gamazo, Ma, Martínez-Abad y Rodríguez-Conde con dos colaboraciones cada uno (Gamazo et al., 2017; Grabau y Ma, 2017; Ma y Crocker, 2007; Martínez-Abad et al., 2020).

Tabla 6*Recuento del número de estudios multinivel por autor*

Nombre del autor/a	Nº. de estudios
Areepattamannil, S.	3
Gamazo, A.; Ma, X.; Martínez-Abad, F.; Rodríguez-Conde, M. J.	2
Bove, G.; Cairns, D.; Chen, F.; Cordero, J. M.; Crocker, R.; Cui, Y.; Eickelmann, B.; Gerick, J.; Gielen, S.; Grabau, L. J.; Guo, S.; Ho, E. S. C.; Kalk, K.; Kaur, B.; Koop, C.; Krips, H.; Lam, T. Y.; Lau, K.; Lau, K. L.; Marella, D.; Meggiolaro, S.; Mikk, J.; Ning, B.; Olmos-Migueláñez, S.; Pedraja, F.; Säälík, Ü.; Shera, P.; Simancas, R.; Van Damme, J.; Van den Noortgate, W.; Vanlaar, G.; Vitale, V.; Yıldırım, S.; Zhang, D.	1

Al igual que con los autores, se registraron las revistas en las que se han publicado mayormente los estudios multinivel (PM3). Como se observa en la Tabla 7, se encontraron 16 diferentes revistas, de las cuales se destaca el *International Journal of Science and Mathematics Education*, el *International Journal of Science Education* y la Revista de Educación, con dos publicaciones cada una (Areepattamannil y Kaur, 2012; Cordero et al., 2015; Gamazo et al., 2017; Grabau y Ma, 2017; Lau y Lam, 2017; Mikk et al., 2015). De estas, según SCIMAGO, la revista con mayor impacto fue la Revista Educación con un *h*-índice de 20.

Tabla 7*Recuerdo del número de estudios multinivel por revista*

Nombre de la revista	<i>h</i> -índice	No. de estudios
<i>International Journal of Science and Mathematics Education</i>	31	2
<i>International Journal of Science Education</i>	93	2
Revista de Educación	20	2
<i>Alberta Journal of Educational Research</i>	18	1
<i>Asia Pacific Education Review</i>	23	1
<i>Asia-Pacific Education Researcher</i>	20	1
<i>Education and Information Technologies</i>	31	1
<i>Educational Psychology</i>	52	1
<i>International Education Studies</i>	15	1
<i>Journal of Adolescence</i>	101	1
<i>Journal of Educational, Cultural and Psychological Studies</i>	4	1
<i>Journal of Experimental Education</i>	50	1
<i>Research in Science Education</i>	45	1
<i>Scandinavian Journal of Educational Research</i>	29	1
<i>Social Psychology of Education</i>	39	1
<i>Studies in Educational Evaluation</i>	33	1

Posteriormente, se reconoció a los países más analizados con datos PISA (PM4). Como se aprecia en la Figura 11, los países más explorados son China (6), España (5), Canadá (4), Estonia (4), Finlandia (4), Japón (4), Australia (3), Brasil (3), Corea (3), Estados Unidos (3), Indonesia (3), Italia (3), México (3), Países Bajos (3), Rusia (3), Singapur (3), Suecia (3) y Turquía (3). Cabe mencionar que, gran parte de los países no fueron investigados de forma individual, sino agrupados en estudios comparativos sobre una o más variables de interés (Cairns y Areepattamannil, 2017; Chen y Cui, 2019; Eickelmann et al., 2016; Lau y Lam, 2017; Mikk et al., 2015).

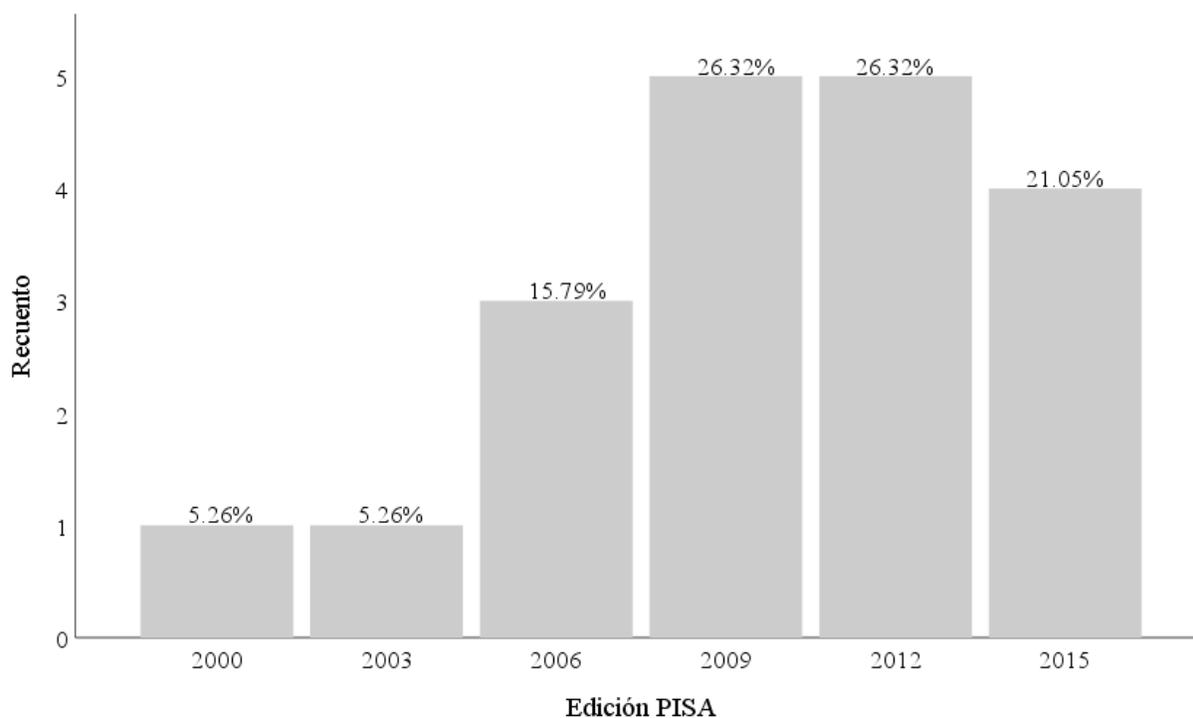
Tabla 8*Recuento de los estudios multinivel por competencia y edición PISA*

Edición PISA	Competencia			Dominio principal
	Lectora	Matemática	Científica	
2000	1	-	-	Lectora
2003	-	1	-	Matemática
2006	-	-	3	Científica
2009	4	2	2	Lectora
2012	1	4	-	Matemática
2015	2	2	4	Científica
2018	-	-	-	Lectora
Total	8	9	9	

Por último, se da respuesta a la pregunta relativa al ciclo PISA más analizado (PM6). Como se puede observar en la Figura 12, se presenta un crecimiento exponencial en los análisis multinivel de los datos proporcionados por las ediciones PISA, siendo la 2009 y la 2012 las más estudiadas, seguidos por la 2015. Es importante aclarar que, aunque los datos PISA 2018 están disponibles desde diciembre del año 2019, en esta revisión no se identificó publicación alguna sobre esta última edición; ello puede deberse al tiempo promedio al que se somete un trabajo de investigación por revisión por pares y/o que aún hay interés en explorar los datos de ediciones anteriores (p. ej. Martínez-Abad et al., 2020), entre otras cosas.

Figura 12

Recuento de los análisis multinivel por edición PISA



2.4.3.2 Resultados de la SRL

La primera pregunta de la Revisión Sistemática de Literatura se centró en las variables de enseñanza más asociadas de forma significativa con los resultados estudiantiles y escolares de PISA (PS1). Como se observa en la Tabla 9, se identificaron 19 variables de enseñanza que han tenido un impacto significativo sobre el rendimiento educativo en PISA, en al menos una ocasión.

Tabla 9*Variables de enseñanza asociadas con los resultados PISA*

Nombre la variable	Nº. de publicaciones
Clima disciplinario en el aula	9
Relación maestro-estudiante	5
Enseñanza basada en la investigación de los estudiantes	4
Adaptación de la instrucción	2
Comportamiento del docente	2
Enseñanza con enfoque en la aplicación del conocimiento	2
Instrucción docente directa	2
Injusticia por parte del profesorado	2
Tiempo de enseñanza efectivo	2
Retroalimentación percibida	2
Uso de las TIC en el aula	2
Actividades prácticas	1
Apoyo docente	1
Estrategias de activación cognitiva	1
Estimulación docente	1
Evaluación formativa	1
Moral docente	1
Orientación dada a los estudiantes	1
Presión académica	1

Fuente: Elaboración propia.

El *clima disciplinario en el aula* fue la variable más asociada con los resultados PISA. La evidencia de nueve estudios coincidió respecto la relación positiva de este factor con la variable criterio, es decir, en la medida que los estudiantes percibieron una mayor disciplina en el aula se obtuvieron también mayores resultados estudiantiles y escolares (Areepattamannil, 2014; Bove et

al., 2016; Cordero et al., 2015; Gamazo et al., 2017; Guo et al., 2018; Lau y Ho, 2015; Ma y Crocker, 2007; Ning et al., 2015; Shera, 2014).

La siguiente variable más asociada fue la *relación maestro-estudiante*. Donde, al igual que el *clima disciplinario*, mayoritariamente se asoció de forma positiva con el nivel de competencia lectora, matemática y científica (Areepattamannil, 2014; Cordero et al., 2015; Mikk et al., 2015; Ning et al., 2015), a excepción del estudio de Bove et al. (2016), quienes encontraron que este factor tuvo un impacto negativo sobre el rendimiento matemático. En general, en cuatro de cinco estudios, una percepción más favorable de los estudiantes hacia las relaciones con sus maestros coincidió con puntajes más altos.

Caso contrario ocurrió con la *enseñanza basada en la investigación de los estudiantes*. En los cuatro análisis en que esta práctica se asoció de forma significativa, su relación con el rendimiento fue negativa (Areepattamannil, 2012; Cairns y Areepattamannil, 2017; Grabau y Ma, 2017; Lau y Lam, 2017). Los hallazgos coincidieron en que una instrucción basada en la investigación de los estudiantes no concurre con los mejores resultados en PISA.

Diferentes variables se asociaron de forma significativa en dos ocasiones: la *adaptación de la instrucción* (Gamazo et al., 2017; Lau y Lam, 2017); el *comportamiento del docente* (Bove et al., 2016; Guo et al., 2018), la *enseñanza con enfoque en la aplicación del conocimiento* (Areepattamannil y Kaur, 2012; Lau y Lam, 2017); la *instrucción directa* (Bove et al., 2016; Lau y Lam, 2017); la *injusticia por parte del profesorado* (Gamazo et al., 2017; Chen y Cui, 2019); el *tiempo efectivo de enseñanza* (Ma y Crocker, 2007; Martínez-Abad et al., 2020); la *retroalimentación percibida* (Gamazo et al., 2017; Lau y Lam, 2017); y, el *uso de las TIC en el aula* (Eickelmann et al., 2016; Meggiolaro, 2017).

La segunda pregunta se enfocó en las variables contextuales más asociadas de forma significativa con los resultados estantilles y escolares de PISA (PS2). En este caso, se identificaron 15 variables con un impacto significativo sobre el rendimiento educativo en PISA (ver Tabla 10).

Tabla 10

Variables contextuales asociadas a los resultados PISA

Nombre la variable	Nº. de publicaciones
ISEC del estudiante	12
Género del estudiante	11
ISEC promedio escolar	10
Estatus migratorio del estudiante	5
Tasa hombres y mujeres en la escuela	5
Grado escolar del estudiante	3
Lenguaje del estudiante	3
Tamaño de la escuela	3
Escasez de docentes	2
Número de cambios de centro	1
Condición de repetidor	1
Proporción de docentes certificados	1
Tamaño del aula	1
Tasa de repetidores	1
Tipo de escuela	1

Nota: ISEC = Índice social, económico y cultural.

La variable contextual más asociada con los resultados PISA fue el *índice social, económico y cultural del estudiante*. Esta variable tuvo un impacto positivo en todos los estudios que resultó significativa, es decir, que en 12 investigaciones, a mayor *índice social, económico y cultural del estudiante* se obtuvieron, también, mayores puntajes (Areepattamannil y Kaur, 2012; Bove et al., 2016; Cairns y Areepattamannil, 2017; Chen y Cui, 2019; Cordero et al., 2015;

Gamazo et al., 2017; Lau y Lam, 2017; Ma y Crocker, 2007; Meggiolaro, 2017; Ning et al., 2015; Shera, 2014; Yıldırım, 2012). Este hallazgo era de esperarse, pues desde mediados del siglo XX Coleman et al. (1966) dieron certeza de la importancia de este factor en el logro académico de los estudiantes y la eficacia de las escuelas.

La siguiente variable más asociada fue el *género del estudiante*. Un total de 11 estudios dieron evidencia de que el ser hombre o mujer tiene un efecto sobre los resultados obtenidos en la prueba PISA (Bove et al., 2016; Chen y Cui, 2019; Cordero et al., 2015; Gamazo et al., 2017; Lau y Ho, 2015; Lau y Lam, 2017; Ma y Crocker, 2007; Meggiolaro, 2017; Ning et al., 2015; Shera, 2014; Yıldırım, 2012). Dado que esta variable es dicotómica y depende del cómo los autores etiquetaron los valores (por ejemplo: 0= hombre y 1=mujer o 0=mujer y 1=hombre) no se puede ser concluyente sobre la linealidad de las relaciones, no obstante, este hallazgo da luz sobre las aún existentes brechas de género en el desarrollo de las compendias lectora, matemática y científica.

Otra variable que se refirió de manera recurrente fue el *índice social, económico y cultural medio escolar*. Al igual que con el *ISEC del estudiante*, esta variable tuvo un efecto positivo en todos los análisis secundarios de PISA en los que resultó significativa. Los hallazgos de diez diferentes estudios coincidieron en que a mayor *ISEC medio escolar* se obtienen, a su vez, mayores niveles de competencia (Bove et al., 2016; Chen y Cui, 2019; Gamazo et al., 2017; Guo et al., 2018; Lau y Ho, 2015; Lau y Lam, 2017; Meggiolaro, 2017; Ning et al., 2015; Shera, 2014; Yıldırım, 2012).

Dos variables se asociaron significativamente en cinco estudios: el *estatus migratorio del estudiante* y la *tasa de hombres y mujeres en la escuela*. De forma semejante al *género del estudiante*, al ser variables de naturaleza nominal, la interpretación de los hallazgos depende de la

codificación de los valores efectuada por los autores y, por lo tanto, es recomendable no ser determinante respecto a la linealidad de la asociación de los factores. Pese a ello, es posible concluir que estos estudios dieron evidencia de las existentes disparidades entre estudiantes migrantes y no migrantes (Cordero et al., 2015; Gamazo et al., 2017; Lau y Lam, 2017; Ma y Crocker, 2007; Meggiolaro, 2017) y el impacto del género de los estudiantes a nivel escuela (Gamazo et al., 2017; Lau y Ho, 2015; Lau y Lam, 2017; Meggiolaro, 2017; Ning et al., 2015).

Otras variables tuvieron un efecto significativo en tres investigaciones: el *grado escolar* y el *lenguaje del estudiante* y el *tamaño de la escuela*. En lo que respecta al primer factor, los tres estudios coincidieron en que a mayor *grado escolar del estudiante* se obtienen, igualmente, mayores puntajes en PISA (Gamazo et al., 2017; Lau y Lam, 2017; Ning et al., 2015). En lo que refiere a la segunda variable, no es posible establecer la linealidad de la relación debido a su naturaleza polinomial, sin embargo, los tres estudios dan certeza de la incidencia del lenguaje del estudiante sobre el logro en la prueba PISA (Areepattamannil y Kaur, 2012; Gamazo et al., 2017; Ma y Crocker, 2007). En el caso del tercer factor hubo hallazgos divergentes: dos positivos (Areepattamannil, 2014; Areepattamannil y Kaur, 2012) y un negativo (Gamazo et al., 2017) en cuanto a la relación del *tamaño de la escuela* con los resultados en PISA.

La *escasez de docentes* se asoció significativamente en dos estudios: Areepattamannil y Kaur encontraron que esta variable predijo negativamente el rendimiento en ciencias de los estudiantes migrantes de Canadá; mientras que Gamazo et al. (2017) hallaron una asociación positiva sobre el logro en ciencias de los estudiantes españoles.

Múltiples variables se asociaron significativamente en una sola ocasión: el *número de cambios de centro* (Gamazo et al., 2017); la *condición de repetidor* (Gamazo et al., 2017); la

proporción de docentes certificados (Guo et al., 2018); el *tamaño de las aulas* (Cordero et al., 2015); la *tasa de repetidores* (Gamazo et al., 2017); y el *tipo de escuela* (Guo et al., 2018).

La tercera pregunta (PS3) refiere a qué tanto por ciento se aplicaron modelos multinivel por pendientes fijas y por pendientes aleatorias. En el 15.8% se optó por pendientes fijas. Mientras que en el 26.3% se prefirió pendientes aleatorias. No obstante, sorprendió que en el 57.9% no se reportó cuál de las dos vías analíticas utilizaron. Si bien en algunas estimaciones la utilización de una u otra vía no tiene un efecto significativo (como el cálculo de residuos), esta decisión informa si se pretendió generalizar a todas las unidades j que componen la población objetivo (Raudenbush et al., 2016) o si la intención fue la emisión de inferencias sobre las unidades j que comprenden la muestra (Pedroza et al., 2018).

La cuarta pregunta (PS4) hizo referencia a la frecuencia con la que se utilizaron todos los valores plausibles proporcionados por PISA. De las investigaciones que especificaron su utilización (el 79% del total), el 93.33% utilizaron todos los valores plausibles y solo el 6.66% optó por emplear solo un valor plausible. Ningún estudio indicó haber optado por la media de los valores plausibles, lo cual pudo haber llevado a estimaciones más sesgadas (Wu, 2005).

La quinta pregunta (PS5) fue respecto al uso de máxima verosimilitud y máxima verosimilitud restringida. De forma semejante a la pregunta PS2, el 73.7% no indicó haber utilizado uno de los dos estimadores. De los estudios que sí especificaron, el 80% optó por máxima verosimilitud, mientras que el otro 20% eligió máxima verosimilitud restringida. Existen diferentes opiniones respecto al uso de una u otra, por un lado, Scherbaum y Pesner (2019) planteó que ambas son igualmente funcionales y, por otro lado, Boedeker (2017) ha comentado que la segunda es la menos limitada en lo que refiere al tamaño de la muestra.

La sexta pregunta (PS6) fue en relación con el uso de los pesos muestrales escolares y/o estudiantiles. Más de la mitad de los estudios (63.2%) indicaron haber empleado los pesos muestrales proporcionados por PISA. Mientras que el 36.8% no indicó haber incorporado esta información en la aplicación de modelos multinivel. De acuerdo con la OCDE (2015), de no utilizar los pesos muestrales no se podría lograr una representación completa de cada escuela y estudiante, no se equilibrarían las participaciones (p. ej. el sobre o submuestreo), no se ajustaría la subrepresentación debido a la no respuesta de las escuelas o de los estudiantes, y no se recortaría la influencia indebida de un subconjunto de escuelas o estudiantes, entre otros.

La séptima pregunta (PS7) se orientó en los coeficientes de correlación intraclase (CCI) reportados. Del total de los estudios seleccionados, únicamente el 63.16% informó el CCI obtenido con el modelo nulo. En la Tabla 11 se observa que los países con mayor CCI fueron Turquía (57%), Italia (50%), Japón (50%), China (44%), India (40%) y Albania (30%); en cambio, los países con menor CCI fueron Canadá (21%; 19%), Estados Unidos (21%; 18%), España (12%) y Finlandia (7%). [(2017) aseveró que la magnitud del CCI puede deberse a las fuertes desigualdades existentes en un país, por ejemplo, las diferencias entre el índice social, económico y cultural de los estudiantes y escuelas; otros autores han compartido este planteamiento: Chaparro y Gamazo (2020) en el caso de México y Quiroz et al. (2018) en el caso de Argentina.

Tabla 11

Coefficientes de correlación intraclase reportados en los análisis multinivel de PISA

CCI	País(es)	Autor(es)
57%	Turquía	Yıldırım (2012)
50%	Italia	Bove et al. (2016)
50%	Japón	Guo et al. (2018) ^a
44%	Shanghái	Guo et al. (2018) ^a
44%	Shanghái	Ning et al. (2015) ^b
40%	India	Areepattamannil (2014)
30%	Albania	Shera (2014)
21%	Canadá	Areepattamannil y Kaur (2012) ^c
21%	Estados Unidos	Guo et al. (2018) ^a
21%	Todos	Chen y Cui (2019)
19%	Canadá	Areepattamannil y Kaur (2012) ^c
18%	Estados Unidos	Grabau y Ma (2017)
17%	Todos	Cairns y Areepattamannil (2017)
12%	España	Gamazo et al. (2017)
12%	España	Martínez-Abad et al. (2020)
7%	Finlandia	Ning et al. (2015) ^b

Nota: CCI = Coeficiente de correlación intraclase. Se le colocó un superíndice (^a, ^b, ^c) a los autores pertenecientes a un mismo estudio. El 12% de CCI de Gamazo et al. (2017) y el 12% de Martínez-Abad et al. (2020) fue el resultado de la media de los tres CCI obtenidos, respectivamente.

La última pregunta se centró en las limitaciones de PISA más reportadas (PS8). Se encontró que el 63.2% de los estudios seleccionados informó al menos una de las limitaciones de haber realizado un análisis secundario de datos PISA. De las cuales, se destacaron:

- El uso de datos transversales (6). Los autores coinciden que, debido a la naturaleza transversal de los datos, los resultados solo pueden ser correlacionales (Cairns y Areepattamannil, 2017; Guo et al, 2018; Lau y Ho, 2015; Shera, 2014). Un diseño longitudinal podría fortalecer la causalidad de las relaciones encontradas (Yıldırım, 2012);

sin embargo, esto no es posible con datos PISA debido a sus inconsistencias y falta de replicabilidad a lo largo del tiempo (Martínez-Abad et al., 2020).

- La falta de evidencia empírica directa (6). Las variables utilizadas en los análisis secundarios -principalmente- se basan en las percepciones de estudiantes y directivos (Bove et al., 2016; Cairns y Areepattamannil, 2017; 2016; Yıldırım, 2012), es decir, PISA no mide de forma directa la información que proporciona sobre los procesos educativos (Shera, 2014). Ello podría implicar una captura incompleta sobre la realidad educativa y, en consecuencia, una falta de claridad sobre las características que verdaderamente podrían incidir sobre los resultados estudiantiles (Mikk et al., 2015; Lau y Ho, 2015).
- Diferencias culturales (1). De acuerdo con Ning et al. (2015), más que una percepción sobre los procesos educativos, las respuestas de estudiantes y directivos podrían reflejar diferencias en la forma que las personas responden a cuestionarios como los de la OCDE; por ejemplo, hay evidencia de que “los encuestados en China tienden a tener respuestas positivas extremas similares, menos respuestas negativas extremas y un mayor número de respuestas neutrales en comparación con sus contrapartes finlandesas” (p. 16).
- La operacionalización de los índices construidos (1). Según Ericlmann (2016), otra limitación es que los datos proporcionados por PISA no se recopilieron para atender los intereses de los análisis secundarios de PISA. Se debe tener precaución en el cómo se abordan los constructos y no alejarse del cómo están definidos por la OCDE.

3 Método

3.1 Hipótesis

Las hipótesis que guía este estudio es que la eficacia escolar de los centros educativos mexicanos es explicada por las siguientes variables de enseñanza:

- *el comportamiento de los docentes,*
- *el uso de las tecnologías de la información y comunicación (TIC),*
- *el tiempo promedio en clase,*
- *el clima disciplinario en las lecciones,*
- *el apoyo docente,*
- *la instrucción directa,*
- *la retroalimentación,*
- *la estimulación,*
- *la adaptación de la instrucción, y*
- *el interés del profesor.*

3.2 Diseño

El diseño del estudio fue de carácter no experimental, también denominado ex-post-facto (Ary et al., 2010), ya que se trata de un análisis secundario los datos proporcionados por PISA. A su vez, forma parte de las investigaciones de tipo transversal debido a que la medición y recolección de los datos fue en un solo momento en el tiempo (Ary et al., 2019).

3.3 Participantes

El trabajo de investigación se realizó a partir del conjunto de datos correspondientes a la aplicación de PISA 2018 en México y, por ende, en este estudio no se implementó un proceso de selección de participantes sino que se utilizó la base de datos de esta aplicación. No obstante, es preciso informar la población objetivo, el método de muestreo y la muestra de la que se derivó el mismo. Desde su primer ciclo, PISA ha estado dirigida a estudiantes “entre 15 años y 3 meses (completos) y los 16 años y 2 meses (completos) al comienzo del periodo de prueba, que asisten a instituciones educativas ubicadas dentro del país” (OECD, 2016a, p. 2), sin embargo, tradicionalmente solo se les enuncia como ‘estudiantes con 15 años de edad’.

El método de muestreo es estratificado en dos etapas (OECD, 2016a). Primero eligen al azar a las escuelas, cuya probabilidad de ser elegidas es proporcional al número de estudiantes inscritos de la población objetivo; y, segundo, escogen aleatoriamente a los estudiantes (con el rango de edad mencionado) inscritos en las escuelas previamente seleccionadas.

En PISA 2018 la muestra mexicana estuvo compuesta por 286 escuelas y 7299 estudiantes. No obstante, en este estudio fueron excluidos los datos de un conjunto de escuelas con el propósito de asegurar la precisión de los análisis (Chaparro y Gamazo, 2019; Gamazo, 2019; Joaristi et al., 2014; y, Martínez-Abad et al., 2017). Para de garantizar la menor pérdida de información se empleó el criterio de exclusión de Hayes (2006). De las bases de datos, se eliminaron a los centros escolares en los que participaron menos de cuatro estudiantes, de tal manera que se aseguraron el 76% de los pesos muestrales, es decir, la representatividad poblacional. Finalmente, para efectos de este trabajo, la muestra estuvo integrada por 268 escuelas y 7267 estudiantes mexicanos.

En la Tabla 12 se caracteriza la muestra utilizada, según las variables proporcionadas por PISA 2018. Como puede observarse, aproximadamente, los estudiantes se distribuyeron

equitativamente en cuanto al género (52.4% de mujeres y 47.6% de hombres). Más de tres cuartas partes se encontraban en el curso indicado en relación con su edad (85%). Hubo mayoría de estudiantes nativos (98.7%), no repetidores de curso (90.1%) y hablantes del mismo idioma en el que fue la prueba (97.8%).

Tabla 12

Características del muestreo final del estudio

Variable	Categorías	Porcentaje válido
Género	Mujer	52.4%
	Hombre	47.6%
Grado escolar en relación con su edad	Adecuado	85%
	No adecuado	15%
Estatus migratorio	Nativos	98.7%
	Migrantes	1.3%
Estatus de repetidor	Ha repetido	9.9%
	No ha repetido	90.1%
Idioma hablado en casa	Idioma del test	97.8%
	Otro lenguaje	2.2%

Nota: La tabla se generó una vez excluidas las escuelas con menos de cuatro estudiantes. El *Grado escolar en relación con su edad* era una variable politómica y fue transformada a dicotómica.

3.4 Instrumentos

Como se ha mencionado, dado que el presente estudio es un análisis secundario de la base de datos de PISA 2018, no se preparó o implementó instrumento alguno. Pese a ello, es necesario describir los instrumentos de los que proceden los datos e integran la evaluación PISA: las pruebas objetivas y los cuestionarios de contexto.

De acuerdo con la OECD (2019), las pruebas objetivas evaluaron la competencia lectora, matemática y científica de la muestra mencionada (ver 3.3 Participantes). Las pruebas estuvieron constituidas por un aproximado de 400 ítems sobre el dominio principal en PISA 2018, la

competencia lectora, y 200 ítems sobre cada dominio secundario, la competencia matemática y científica. Los ítems fueron de uno de tres tipos de respuesta: abierta (respuesta desarrollada), cerrada (respuesta exacta) u opción múltiple.

Los cuestionarios de contexto recogieron información adicional, complementaria a la competencia de los estudiantes (OECD, 2017c). Al igual que en las aplicaciones PISA 2012 y 2015, con estos cuestionarios se capturaron variables relacionadas con tres dimensiones: constructos no cognitivos y metacognitivos, constructos relacionados con los antecedentes de los estudiantes, y procesos educativos (OECD, 2019). Aunque en PISA 2018 se aplicaron otros cuestionarios, para efectos de este estudio únicamente se recuperaron los datos de los cuestionarios estudiantiles y escolares.

3.5 Variables del Estudio

Como se mencionó al final del apartado anterior (3.4 Instrumentos), las bases de datos estudiantiles y escolares fueron utilizadas para cumplir con el propósito de este trabajo. Como es usual en análisis secundarios, a continuación, se describe el proceso seguido en la limpieza de datos y las variables utilizadas en cada base.

3.5.1 Limpieza de datos

La limpieza de datos consistió en la búsqueda y corrección (o eliminación) de registros incorrectos en las bases de datos. Dado que son distintas variables, se describe el proceso seguido en la Fase 1 y en la Fase 3.

3.5.1.1 Limpieza de datos de la Fase 1

1. Las variables fueron elegidas con base en estudios empíricos similares.
2. Las variables politómicas se recodificaron en variables *dummy*, donde (0) fue la ausencia y (1) la presencia de una condición.

3. Las variables dicotómicas se recodificaron en las mismas variables, de valores (1) y (2) a valores (0) y (1).
4. Los valores perdidos de las variables a nivel escuela fueron imputados
5. El nombre de todas las variables elegidas se redujo a ocho caracteres.

3.5.1.2 Limpieza de datos de la Fase 3

1. Las variables fueron elegidas con base en el Modelo Dinámico de Eficacia Educativa de Creemers y Kyriakides (2008).
2. Las variables a nivel estudiante fueron agregadas a nivel escuela.
3. Las variables politómicas se recodificaron en variables *dummy*, donde (0) fue la ausencia y (1) la presencia de una condición.
4. Se revisó que las variables dicotómicas tuviesen valores (0) y (1).
5. Se revisó que no hubiese valores perdidos en las variables.

3.5.2 Variables elegidas

Se hizo uso de diferentes tipos de variables en función del objetivo de cada fase de análisis. A las variables se les denominó con base en la naturaleza del estadístico (ver 3.6 Análisis de datos): (a) de *criterio* y *predictoras* en la Fase 1; y, (b) *explicativas* y de *resultado* en la Fase 3.

3.5.2.1 Fase 1

En esta fase las variables de criterio fueron los 30 valores plausibles, diez por cada competencia (ver Tabla 13). Dado que en PISA cada estudiante evaluado representa a un conjunto de estudiantes de la población objetivo, informada con el peso muestral de su participación, las puntuaciones proporcionadas no corresponden de forma directa a un resultado concreto. En lugar de esto, PISA estima un rango de posibles puntuaciones obtenidas por la población representada (OECD, 2017c).

Tabla 13*Variables de criterio introducidas en los modelos multinivel*

Variable	Etiqueta	Tipo	Rango
VP1Lec			
...	Diez valores plausibles en competencia lectora	Continua	0 – 749.586
VP10Lec			
VP1Mat			
...	Diez valores plausibles en competencia matemática	Continua	0 – 757.313
VP10Mat			
VP1Cie			
...	Diez valores plausibles en competencia científica	Continua	0 – 704.589
VP10Cie			

Nota: VP = Valor Plausible.

Las variables predictoras fueron los factores contextuales a nivel estudiante y nivel escuela. Como se aprecia en la Tabla 14, por factores contextuales se entienden aquellas características individuales o grupales en las que los actores educativos no tienen capacidad de incidir (Pedroza et al., 2018). Por ejemplo, el *género*, el *estatus migratorio* o el *índice social, económico y cultural* del estudiante.

Tabla 14*Variables predictoras introducidas en los modelos multinivel*

	Descripción de la variable	Tipo	Rango
Nivel estudiante	Género del estudiante	Nominal	0: Masculino 1: Femenino
	Estatus migratorio del estudiante ^d	Nominal	0: Nativo 1: Migrante
	Índice social, económico y cultural del estudiante ⁱ	Continua	-4.90 – 2.74
	Estatus de repetidor del estudiante ^d	Nominal	0: No repetidor 1: Repetidor
	Lengua hablada por el estudiante	Nominal	0: Lenguaje de la prueba 1: Otra lengua
	Grado escolar del estudiante	Ordinal	1° secundaria – 2° bachillerato
Nivel escuela	Porcentaje de mujeres en la escuela ^a	Continua	0 – 1
	Tasa de migrantes en la escuela ^a	Continua	0 – 1
	Índice social, económico y cultural medio en la escuela ^a	Continua	-3.53 – 1.25
	Tasa de repetidores en la escuela ^a	Continua	0 – 1
	Tipo de escuela ^d	Nominal	0: Pública 1: Privada
	Escasez de recursos en la escuela ⁱ	Continua	-1.42 – 2.96
	Escasez de personal en la escuela ⁱ	Continua	-1.45 – 4.04
	Tamaño de la escuela	Discreta	29 – 7564
	Tamaño de las aulas en la escuela	Discreta	13 – 53
	Ratio de estudiantes por docente en la escuela	Continua	2.33 – 100

Nota: ⁱ = Índice compuesto por información procedente de diversos ítems; ^d Variable *dummy*; ^a Variable agregada.

3.5.2.2 Fase 3

En la última fase la variable de resultado fue la Eficacia Escolar, producto de la aplicación de técnicas de conglomeración a los residuos transformados obtenidos en la Fase 1. La información sobre esta variable se encuentra en apartado 4.3.

Las variables explicativas fueron los factores de enseñanza y factores escolares incluidos en los cuestionarios de contexto estudiantil y escolares. En este caso, la mayor parte de las variables fueron índices compuestos por información procedente de diversos ítems (ver Tabla 15), a excepción de las variables *actividades extracurriculares creativas*, *oferta de clases adicionales* y *tiempo promedio en clase*.

Tabla 15

Variables explicativas introducidas en las regresiones logísticas multinomiales

	Descripción de la variable	Tipo	Rango
De enseñanza	Adaptación de la instrucción ^í	Continua	-2.2652 – 2.0073
	Apoyo docente ^í	Continua	-2.7106 – 1.3411
	Clima disciplinario en las lecciones ^í	Continua	-2.7124 – 2.0345
	Comportamientos docentes que dificultan el aprendizaje ^í	Continua	-2.0409 – 3.7879
	Estimulación del profesor ^í	Continua	-2.3003 – 2.0871
	Instrucción directa ^í	Continua	-2.9425 – 1.8202
	Interés docente ^í	Continua	-2.2177 – 1.8245
	Retroalimentación ^í	Continua	-1.6391 – 2.0165
	Tiempo promedio en clase	Continua	10 – 120
	Uso de las TIC durante la clase	Continua	-1.2188 – 2.4394
Escolares	Actividades extracurriculares creativas ^d	Nominal	0 – 1
	Calidad escolar ^í	Continua	-0.7055 – 1.4373
	Clima escolar discriminatorio	Continua	-1.1549 – 2.4591
	Oferta de clases adicionales	Nominal	0 – 1
	Políticas escolares para la participación de los padres ^í	Continua	-1.2358 – 1.3784
	Uso general de las TIC en la escuela	Continua	-1.7161 – 1.5264

Nota: TIC = Tecnologías de la Información y Comunicación. ^í = Índice compuesto por información procedente de diversos ítems; ^d Variable *dummy*.

Con base en la OCDE (2019b), a continuación, se presentan los ítems que conformaron a cada uno de los índices. Ello con el propósito de precisar los múltiples elementos tomados en cuenta para su construcción.

- Adaptación de la instrucción
 - El profesor adapta la lección a las necesidades y conocimientos de mi clase.
 - El maestro proporciona ayuda individual cuando un estudiante tiene dificultades para comprender un tema o tarea.
 - El maestro cambia la estructura de la lección sobre un tema que a la mayoría de los estudiantes les resulta difícil de entender.
- Apoyo docente
 - El maestro muestra interés en el aprendizaje de cada alumno
 - El maestro brinda ayuda adicional cuando los estudiantes la necesitan
 - El profesor ayuda a los estudiantes con su aprendizaje
 - El maestro continúa enseñando hasta que los alumnos entiendan
- Clima disciplinario en las lecciones
 - Los estudiantes no escuchan lo que dice el maestro
 - Hay ruido y desorden
 - El profesor tiene que esperar mucho tiempo para que los estudiantes se callen
 - Los estudiantes no pueden trabajar bien
- Comportamiento del docente que dificulta el aprendizaje
 - Los maestros no satisfacen las necesidades individuales de los estudiantes
 - Ausentismo docente
 - El personal de la escuela se resiste al cambio

- Los profesores son demasiado estrictos con los estudiantes
- Los maestros no están bien preparados para las clases
- Estimulación del docente
 - El maestro anima a los estudiantes a expresar su opinión sobre un texto
 - El maestro ayuda a los estudiantes a relatar las historias que leen a sus vidas
 - El profesor muestra a los alumnos cómo la información de los textos se basa en lo que ya saben
 - El maestro plantea preguntas que motivan a los estudiantes a participar activamente
- Instrucción directa
 - El maestro establece objetivos claros para el aprendizaje
 - El maestro hace preguntas para verificar si han entendido lo que se enseñó
 - Al comienzo de una lección, el maestro presenta un breve resumen de la lección anterior
 - El maestro dice lo que se tiene que aprender
- Interés del docente
 - El docente deja claro que le gusta enseñarles
 - El entusiasmo del profesor inspira a los estudiantes
 - El profesor de muestra que le gusta tratar el tema de la lección
 - El maestro mostró placer en la enseñanza
 - Los estudiantes no comienzan a trabajar por mucho tiempo después de que comienza la lección
- Retroalimentación
 - El maestro da retroalimentación sobre los puntos fuertes en la materia

- El profesor dice en qué áreas aún se puede mejorar
- El profesor dice cómo se puede mejorar el rendimiento
- Calidad escolar
 - La mayoría de los maestros de escuela parecen competentes y dedicados
 - Los estándares de rendimiento son altos en la escuela
 - Satisfacción con el contenido enseñado y los métodos de instrucción utilizados en la escuela
 - Satisfacción con el ambiente disciplinario en la escuela
 - La escuela supervisa cuidadosamente el progreso de los estudiantes
 - La escuela proporciona información periódica y útil sobre el progreso de los estudiantes
 - La escuela hace un buen trabajo educando a los estudiantes
- Políticas escolares para la participación de los padres
 - La escuela ofrece un ambiente acogedor para que los padres se involucren
 - La escuela proporciona una comunicación efectiva entre la escuela y las familias
 - La escuela involucra a los padres en el proceso de toma de decisiones de la escuela
 - La escuela ofrece educación para padres
 - La escuela informa a las familias sobre cómo ayudar a los estudiantes con la tarea y otras actividades relacionadas con la escuela
 - La escuela coopera con servicios comunitarios para fortalecer los programas escolares y el desarrollo estudiantil

3.6 Análisis de Datos

Como se introdujo en el apartado anterior, el estudio estuvo compuesto de tres fases de análisis: (a) el cálculo de residuos escolares en competencia lectora, matemática y científica; (b) la clasificación de la eficacia de las escuelas; y, (c) la explicación de la eficacia de las escuelas. En la Tabla 16 se resumen los análisis realizados y software utilizados en cada fase.

Tabla 16

Resumen de los análisis y softwares empleados en cada fase del estudio

Fase	Análisis	Software
1. Estimación de residuos	Diagnóstico de colinealidad entre variables contextuales	SPSS v.26
	Modelos multinivel en competencia lectora	HLM7
	Modelos multinivel en competencia matemática	HLM7
	Modelos multinivel en competencia científica	HLM7
	Prueba de normalidad, homocedasticidad y linealidad de los residuos	SPSS v.26
2. Clasificación de la eficacia escolar	Estandarización y normalización de los residuos	R Project
	Prueba de normalidad, homocedasticidad y linealidad de los residuos transformados	SPSS v.26
	Clasificación de los residuos	SPSS v.26
	Graficación 3D	R Project
	Contrastes marginales	R Project
3. Análisis del peso explicativo	Diagnóstico de colinealidad entre variables explicativas	SPSS v.26
	Regresión logística multinomial – explicativas de enseñanza	SPSS v.26
	Regresión logística multinomial – explicativas escolares	SPSS v.26
	Regresión logística multinomial – explicativas de enseñanza y escolares	SPSS v.26
	Matriz de confusión	SPSS v.26
	Precisión, exactitud y valor Kappa	Calculadora virtual

Nota: El enlace de la calculadora virtual es <http://www.marcovanetti.com/pages/cfmatrix/?noc=3>

3.6.1 Fase 1: Estimación de residuos

La estimación de los residuos en competencia lectora, matemática y científica de los centros educativos mexicanos se efectuó mediante la aplicación de modelos multinivel. Por cada competencia, se construyeron 10 modelos independientes (uno por cada variable criterio) y un modelo final que comprende los parámetros estimados en los 10 modelos independientes; es decir, que el *output* final del modelado es un modelo que integra la información proveniente de 10 modelos multinivel.

Previo al modelado, se realizaron diagnósticos de colinealidad entre variables predictoras de nivel estudiante y de nivel escuela. Ello con el propósito de evitar la inclusión de variables altamente correlacionadas que, de acuerdo con Darlington y Hayes (2017), podrían afectar el poder de las pruebas sobre los coeficientes de regresión; pues aumentan los errores estándar de las pendientes y vuelven más complicado el encontrar valores estadísticamente significativos. Dado que en esta fase se cuenta con 30 variables criterio (ver 3.5.2 Variables de cada fase), fue necesario realizar 30 diagnósticos de colinealidad. Para cada diagnóstico, el criterio estadístico utilizado fue el factor de inflación de la varianza (VIF, por sus siglas en inglés). Como lo indica su nombre, el VIF cuantifica la intensidad de influencia de un predictor sobre los otros en un análisis de regresión (Roni et al., 2020). Con base en Garson (2014) y Gordon (2014), se consideró que no hubo problemas de colinealidad si el $VIF \leq 4$; cuando este estuvo presente, el modelo fue simplificado al excluir las variables con un VIF alto (Chatterjee y Simonoff, 2013).

Una vez eliminadas las variables con un alto índice de colinealidad, se procedió a la construcción de los modelos multinivel. Las decisiones metodológicas tuvieron base en recomendaciones y evidencia empírica de estudios similares:

- Modelos multinivel con pendientes fijas. Gamazo (2019), Martínez-Garrido (2015), y Özdemir (2016) probaron que no existen diferencias significativas entre residuos obtenidos a partir de pendientes fijas y pendientes aleatorias. De tal forma, que es recomendable elegir en función del principio de parsimonia. Al incluir un menor número de variables, se optó por pendientes fijas.
- Diez valores plausibles por competencia. De acuerdo con Von Davier et al. (2009), debido a que lleva a resultados con mayor sesgo, el promedio de los valores plausibles no debe emplearse como variable criterio. Dado que es posible en HLM7, en este trabajo se integraron en la variable criterio los diez valores plausibles correspondientes en cada modelo multinivel.
- Máxima verosimilitud restringida. Aunque Scherbaum y Pesner (2019) plantearon que la máxima verosimilitud y la máxima verosimilitud restringida son igualmente funcionales (ML y REML, por sus siglas en inglés), Boedeker (2017) comentó que REML se utiliza en mayor medida al no limitarse (como ML) a grandes muestras de nivel-1 y nivel-2. Por tal situación, se procedió a realizar estimaciones con REML.
- Pesos muestrales estudiantiles y escolares. De acuerdo con Tat et al. (2019), los pesos muestrales son necesarios para la estimación de parámetros poblacionales, errores de muestreo, y emisión de inferencias poblacionales válidas. Al igual que Martínez-Abad et al. (2020), dado que no fue posible la incorporación de las réplicas de los pesos, únicamente se utilizaron los pesos muestrales estudiantiles y escolares.

Inicialmente, se generó el modelo nulo. Este integró únicamente a la variable criterio, es decir, no incluyó variable predictora alguna. Particularmente, su utilidad radicó en el cálculo de la varianza total que se encuentra sistemáticamente dentro de las escuelas y entre las escuelas (Shera,

2014), el coeficiente de correlación intraclase (CCI). De acuerdo con Lee (2000), es recomendable no realizar modelos multinivel si el CCI es menor al 10% de la varianza total en la variable criterio.

Posteriormente, se construyó el modelo multinivel condicional. Este incorporó a ambos tipos de variables, de criterio y predictoras, a nivel estudiante y a nivel escuela. Se siguió un paso a paso en el que se eliminó una a una las variables predictoras sin efecto significativo sobre la variable criterio ($p > .05$). Una vez obtenido el modelo final, se calculó el CCI y se informó el total de la varianza explicada por parte de las variables predictoras incluidas en los modelos y qué porcentaje de ella aún quedaba por explicar.

Por último, con las tres versiones finales de los modelos multinivel, se estimaron los residuos de los centros educativos mexicanos. Definidos, en este estudio, como la diferencia entre el resultado esperado, en una escuela con estudiantes y características contextuales idénticas, y el resultado “verdadero” obtenida por la escuela. Como procedimiento de rutina, se presentaron estadísticos descriptivos y contrastes de:

1. Normalidad con histograma, gráfico Q-Q e índice Kolmogorov-Smirnov.
2. Homocedasticidad y linealidad (simultáneamente) con una nube de puntos, donde en el eje Y se ubicaron los residuos y en el eje X las puntuaciones pronosticadas.

3.6.2 Fase 2: Clasificación de la eficacia escolar

La clasificación de la eficacia escolar de los centros educativos mexicanos, a partir de su residuo, se realizó mediante la técnica de conglomeración K-medias. Sin embargo, al igual que Martínez-Garrido (2015), previo a la aplicación de esta técnica se optó por la transformación de los residuos a una escala estandarizada y normal con el propósito de facilitar las interpretaciones. Este procedimiento se realizó con las Familias de Distribuciones de Johnson, que, en términos

generales, consiste en la identificación y transformación de la distribución dentro una de tres familias de distribuciones Johnson (1949):

- S_B cuando x es acotada. La transformación aplicada es

$$Z = \gamma + \eta \ln \left(\frac{x - \varepsilon}{\lambda + \varepsilon - x} \right) \quad (1)$$

Sujeta a:

Condiciones de los parámetros: $\eta, \lambda > 0, -\infty < \gamma < \infty, -\infty < \varepsilon < \infty$

Condiciones de la variable x : $\varepsilon < x < \varepsilon + \lambda$

- S_L cuando x acotada por debajo o log-normal. La transformación aplicada es

$$Z = \gamma + \eta \ln(x - \varepsilon) \quad (2)$$

Sujeta a:

Condiciones de los parámetros: $\eta > 0, -\infty < \gamma < \infty, -\infty < \varepsilon < \infty$

Condiciones de la variable x : $x > \varepsilon$

- S_U cuando x no-acotada. La transformación aplicada es

$$Z = \gamma + \eta \operatorname{senh}^{-1} \left(\frac{x - \varepsilon}{\lambda} \right) \quad (3)$$

Sujeta a:

Condiciones de los parámetros: $\eta, \lambda > 0, -\infty < \gamma < \infty, -\infty < \varepsilon < \infty$

Condiciones de la variable x : $-\infty < x < \infty$

Una vez efectuada las transformaciones, se presentaron estadísticos descriptivos y contrastes de normalidad, homocedasticidad y linealidad de los residuos.

A los residuos transformados se le aplicaron técnicas de *clustering* para la identificación y clasificación de la eficacia de las escuelas. Entre las técnicas existes, en este estudio se optó por K-medias. Según Owsiniński (2020), K-medias cuenta con un algoritmo robusto que permite

distinguir y nombrar de forma óptima lo similar y lo diferente. No obstante, dado que es desconocido número el ideal de conglomerados, esta técnica -al igual que otras- requiere una configuración *a priori* del número de conglomerados. Como solución a esto, Ben HajKacem et al. (2019) propusieron utilizar diferentes modelos heurísticos y tomar las agrupaciones que proporcionen mayor equilibrio.

Por consiguiente, para valorar y determinar el número adecuado de clústers, se aplicaron tres criterios, dentro de las tipologías presentadas por Vilà-Bañós et al. (2014): (a) la menor distancia promedio entre la escuela y su centroide, es decir, la menor media entre las distancias euclidianas de cada grupo; (b) el menor tamaño de conglomerados, en otras palabras, la menor desviación estándar promedio entre cada grupo; y, (c) un mínimo de 20 escuelas por grupo.

Por último, se realizó una evaluación de la calidad de las agrupaciones. Pues, como lo indicaron Almeida et al. (2011), además de medir la efectividad del algoritmo de agrupación, es preciso informar sobre la dinámica de las relaciones en los conjuntos de datos. Para ello, se aplicaron nueve contrastes marginales que informaron la densidad entre pares de variables, los cuales, permiten valorar la calidad de las agrupaciones a través de la concentración de los puntos en la dispersión (Görke et al., 2015; Puschmann et al., 2016). Una agrupación de calidad se caracteriza, por ejemplo, por único núcleo “denso” (Kriegel et al., 2011).

3.6.3 Fase 3: Análisis del peso explicativo

El análisis del peso explicativo de los procesos educativos sobre la eficacia escolar de los centros educativos mexicanos realizó mediante regresiones logística multinomiales. Se optó por este tipo de regresión dado que el objetivo de investigación es la explicación de la eficacia escolar de los centros educativos, independientemente de si exista o no una linealidad en la naturaleza de la

eficacia; diferente a que lo que se podría obtener con otro tipo de regresiones, como la ordinal, que se enfoca particularmente la explicación de relaciones lineales.

Previo a la regresión, se efectuaron dos análisis: (a) diagnósticos de colinealidad entre variables explicativas, cuyo criterio aplicado fue el de Garson (2014) y Gordon (2014), aquellas variables con un $VIF > 4$ tuvieron problemas de colinealidad y fueron excluidas; y, (b) análisis de correlación entre variables explicativas y residuos en competencia lectora, matemática y científica para dar mayor sustento a las interpretaciones, donde se tomó en consideración, a su vez, los criterios del tamaño del efecto de Cohen (1988): un $r = .2$ es una correlación pequeña, un $r = .5$ es una correlación media y un $r = .8$ es una correlación alta.

Tras descartar problemas de colinealidad, se procedió a las regresiones logística multinomiales. En este caso, el “Residuo centrado” se estableció como categoría de referencia y se procedió a un paso a paso en el que se fueron eliminando una a una las variables con ausencia de significancia sobre una de las categorías a explicar, la “Baja eficacia” o la “Alta eficacia”. El modelo de regresión final incluyó únicamente aquellas variables explicativas con un efecto significativo en alguna de las dos categorías ($p < .05$).

Cabe mencionar que, para este análisis, fue necesaria la realización de tres modelos de regresiones logísticas multinomiales. Ello con el fin de evitar la sobredivisión de la varianza, lo cual, inhibe la identificación de variables con efectos significativos. La primera regresión, incluyó únicamente las variables explicativas escolares; la segunda, solo las variables explicativas de enseñanza; y, la tercera, exclusivamente las variables que resultaron significativas en la primera y segunda regresión (escolares y de enseñanza). Con ello, se identificaron las variables escolares y de enseñanza que más explicaron la eficacia escolar de los centros educativos mexicanos.

Además, con el propósito de complementar la información proporcionada por los coeficientes de regresión, se presentaron estadísticos descriptivos e inferenciales de las escuelas de baja y alta eficacia. Los estadísticos descriptivos e inferenciales fueron útiles para: (a) conocer la composición de las escuelas, (b) identificar tipos de segregación educativa dentro y entre escuelas (por ejemplo, índices socioeconómicos altos en escuela de alta, y bajos en escuela de baja eficacia); y, (c) complementar coeficientes contraintuitivos (por ejemplo, que el clima del aula aumente la baja eficacia y aumente la alta eficacia).

Por último, se generó la tabla de clasificación final y se procedió a la construcción de una matriz de confusión. Como lo indicó Twain (2018), esta matriz fue útil para evaluar el modelo de regresión final en términos de precisión, exactitud y acuerdo. Los cálculos se realizaron a través de la calculadora virtual proporcionada por el Dr. Venetti (2007).

4 Resultados

4.1 Fase 1: Estimación de Residuos

La Fase 1 tuvo por objetivo la estimación de residuos en competencia lectora, matemática y científica de los centros educativos mexicanos. De tal forma que, a continuación, se presentan los resultados obtenidos de los diagnósticos de colinealidad entre variables predictoras, los Modelos Multinivel construidos por competencia evaluada (lectora, matemática y científica) y los residuos escolares estimados por cada dominio PISA.

4.1.1 *Diagnósticos de colinealidad*

Tal como se observa en la Tabla 17, ninguna variable a nivel estudiante y a nivel escuela tuvo problemas de colinealidad. El VIF más alto fue de 2.251, correspondiente a la variable *Tipo de escuela*; no obstante, con base en Garson (2014) y Gordon (2014), este no se consideró lo suficientemente elevado como para justificar la retirada de la variable.

Tabla 17*Síntesis de los 30 diagnósticos de colinealidad entre las variables predictoras*

Variables	VIF			
	Lectora	Matemática	Científica	
Nivel estudiante	Género del estudiante	1.012	1.012	1.012
	Estatus migratorio del estudiante	1.028	1.028	1.028
	ISEC del estudiante	1.028	1.028	1.028
	Estatus de repetidor del estudiante	1.844	1.844	1.844
	Lengua hablada por el estudiante	1.032	1.032	1.032
	Grado escolar del estudiante	1.827	1.827	1.827
Nivel escuela	Tasa de mujeres en la escuela	1.154	1.154	1.154
	Tasa de migrantes en la escuela	1.107	1.107	1.107
	ISEC medio en la escuela	2.153	2.153	2.153
	Tasa de repetidores en la escuela	1.409	1.409	1.409
	Tipo de escuela	2.251	2.251	2.251
	Escasez de recursos en la escuela	1.979	1.979	1.979
	Escasez de personal en la escuela	1.552	1.552	1.552
	Tamaño de la escuela	1.545	1.545	1.545
	Tamaño de las aulas en la escuela	1.366	1.366	1.366
	Ratio de estudiantes por docente en la escuela	1.192	1.192	1.192

Nota: En esta tabla únicamente se reportan los VIF más altos obtenidos en cada variable. Un VIF>4 indicaría problemas de colinealidad. ISEC = Índice social, económico y cultural.

4.1.2 Modelos multinivel en competencia lectora

El modelo nulo en competencia lectora se expresó a través de la Ecuación 4.

$$CLectora_{ij} = \gamma_{00} + u_{0j} + r_{ij} \quad (4)$$

Donde:

$CLectora_{ij}$ estuvo compuesta por los 10 valores plausibles en competencia lectora de cada estudiante i en una escuela j ;

γ_{00} fue el intercepto (o gran media) en competencia lectora de todos los estudiantes i incluidos en escuelas j ;

u_{0j} fue el residuo en competencia lectora de cada escuela j , es decir, la diferencia entre la gran media γ_{00} y la puntuación obtenida en la escuela j en competencia lectora; y

r_{ij} es el residuo en competencia lectora del estudiante i de una escuela j , es decir, cuánto dista cada estudiante de lo esperado en su escuela.

Una vez obtenido el modelo nulo, se extrajo la estimación final de los componentes de varianza (ver Tabla 18) para calcular el coeficiente de correlación intraclase (CCI).

Tabla 18

Componentes de varianza del modelo nulo en competencia lectora

Efecto aleatorio	Desviación estándar	Componentes de varianza	g.l.	x^2	p-valor
Nivel-2	51.292	2630.908	267	4811.379	<.001
Nivel-1	64.842	4204.542			

A partir de las unidades incluidas en la tercera columna de la Tabla 18, se computó el CCI inicial en competencia lectora. Como se aprecia en la Ecuación 5, este fue superior al 10%

recomendable (Lee, 2000) para efectuar Modelos Multinivel. El 38.5% de la variabilidad en competencia lectora de los estudiantes mexicanos fue atribuible a nivel escuela.

$$ICC_{CLectora} = \frac{2630.908}{2630.908 + 4204.542} = 0.385 \quad (5)$$

El modelo condicional, para el nivel de competencia lectora de los estudiantes, quedó definido por la Ecuación 6.

$$CLectora_{ij} = \gamma_{00} + \gamma_{01}SchESCS + \gamma_{02}SchRep + \gamma_{10}StuIMMIG + \gamma_{20}StuESCS + \gamma_{30}StuRep + \gamma_{40}StuGra + u_{0j} + r_{ij} \quad (6)$$

Donde:

$CLectora_{ij}$ estuvo compuesta por los 10 valores plausibles en la competencia lectora de cada estudiante i en una escuela j ;

γ_{00} fue el intercepto (o gran media) en competencia lectora de todos los estudiantes i incluidos en escuelas j ;

$\gamma_{10} - \gamma_{40}$ fueron covariables a nivel estudiante;

$\gamma_{01} - \gamma_{04}$ fueron covariables a nivel escuela;

u_{0j} fue el residuo en competencia lectora de cada escuela j ; y

r_{ij} fue el residuo en competencia lectora del estudiante i en una escuela j .

En la Tabla 19 se presentan los coeficientes, el error estándar, el t -ratio y el p -valor de las variables predictoras que integraron el modelo condicional final en competencia lectora. Como se observa en la columna correspondiente al t -ratio, en orden de importancia, las variables que más influyeron en el nivel de competencia lectora de los estudiantes mexicanos “en PISA 2018” fueron el *estatus migratorio del estudiante*, el *ISEC del estudiante*, el *ISEC medio de la escuela* y el *grado escolar del estudiante*, el *estatus de repetidor del estudiante* y la *tasa de repetidores en la escuela*.

Tabla 19*Modelo condicional en competencia lectora con estimación robusta del error*

Efecto fijo	Coefficiente	Error estándar	t-ratio	p-valor
Intercepto γ_{00}	458.085	6.858	66.799	<.001
ISEC medio en la escuela γ_{01}	19.352	5.270	3.672	.001
Tasa de repetidores en la escuela γ_{02}	-31.253	13.149	-2.377	.018
Estatus migratorio del estudiante γ_{10}	-55.908	14.208	-3.935	<.001
ISEC del estudiante γ_{20}	6.311	1.730	3.647	.001
Estatus de repetidor del estudiante γ_{30}	-18.363	6.897	-2.662	.009
Grado escolar del estudiante γ_{40}	17.126	5.050	3.391	.001

Nota: ISEC = Índice social, económico y cultural.

Tras construir el modelo condicional en competencia lectora, se obtuvo la estimación final de los componentes de varianza (ver Tabla 20).

Tabla 20*Componentes de varianza del modelo condicional en competencia lectora*

Efecto aleatorio	Desviación estándar	Componentes de varianza	g.l.	x^2	p-valor
Nivel-2	29.852	891.135	265	1746.561	<.001
Nivel-1	63.809	4071.541			

Con base en los elementos incluidos en la tercera columna de la Tabla 20, se calculó el CCI final (Ecuación 7). Con este, se pudo llegar a dos conclusiones: (a) los factores contextuales explicaron 53.25% de la variabilidad dentro y entre las escuelas; y (b) aún era posible explicar el 46.75% de la varianza.

$$ICC_{C\text{Lectora}} = \frac{891.135}{891.135 + 4071.541} = 0.18 \quad (7)$$

4.1.2.1 Residuos en competencia lectora

Finalmente, se estimaron los residuos escolares en competencia lectora, la diferencia entre el rendimiento esperado en una escuela con características contextuales idénticas y el resultado “verdadero” obtenido por la escuela. Los estadísticos descriptivos de los valores obtenidos se encuentran en la Tabla 21.

Tabla 21

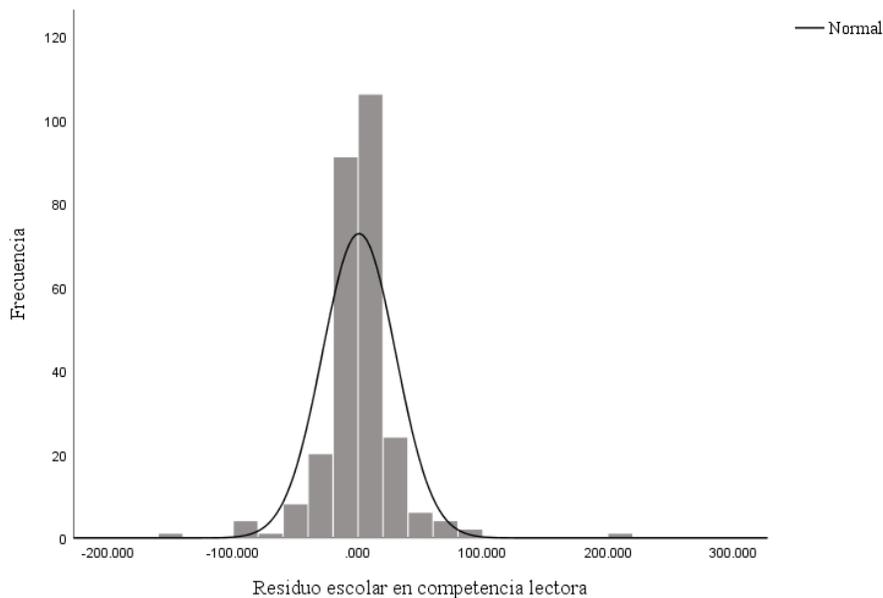
Descriptivos de los residuos escolares en competencia lectora

	N	Mínimo	Máximo	Media	Desv. Est.	Asimetría		Curtosis	
						Valor	Desv. Error	Valor	Desv. Error
Residuos	268	-157.492	216.791	.779	29.372	.762	.149	14.389	.297

El contraste de Kolmogorov-Smirnov presentó un índice menor a .001, es decir, la distribución de los residuos en competencia lectora se desvió ligeramente de la curva normal. En la Figura 13, se aprecia que ello pudo deberse a una alta concentración de valores en torno a la media, una característica de las distribuciones leptocúrticas.

Figura 13

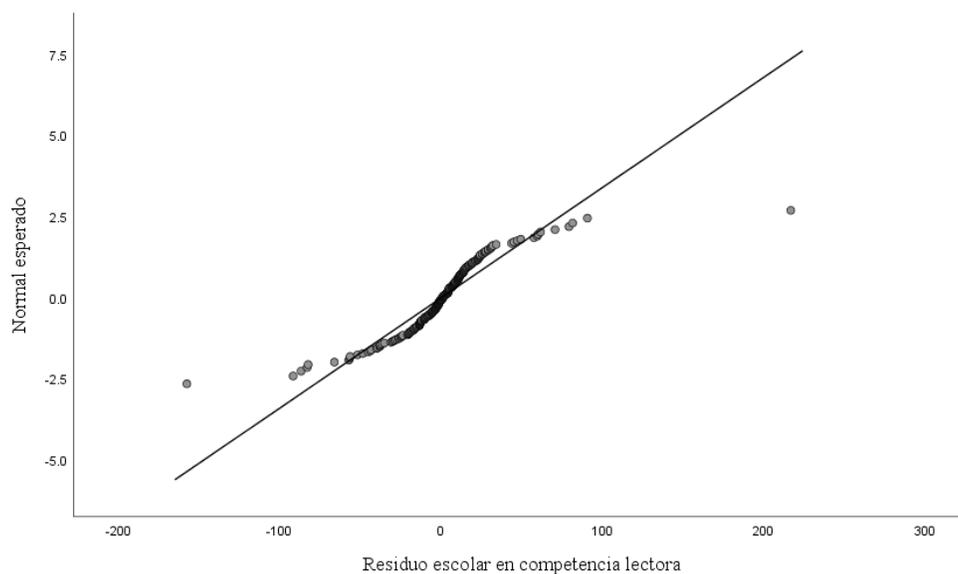
Histograma de los residuos escolares en competencia lectora



En la Figura 14, se observa que gran parte de los valores circulan entorno a la normal, encontrando algunos valores atípicos moderados y extremos que pudieron haber producido la desviación mencionada.

Figura 14

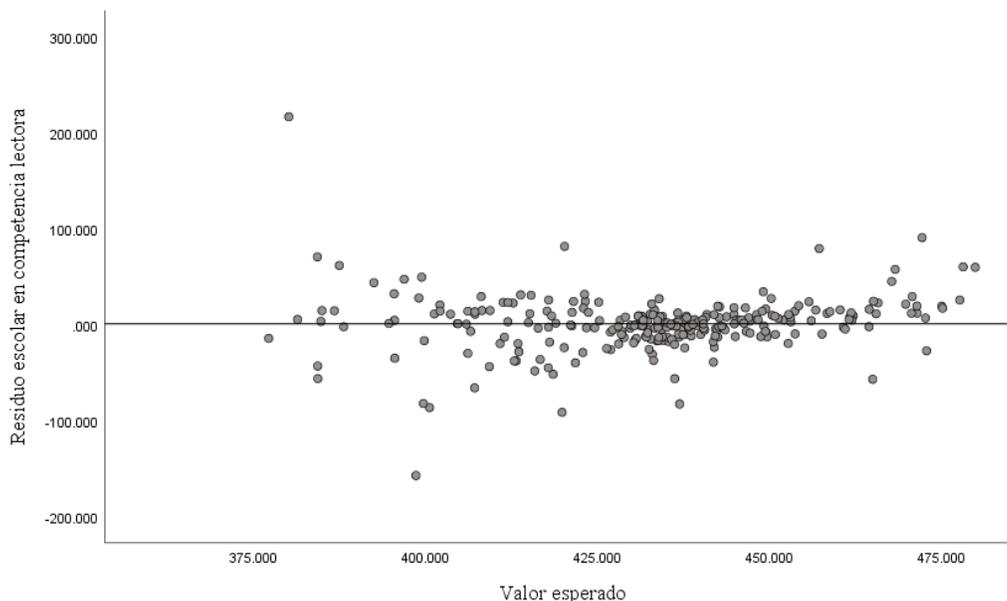
Gráfico Q-Q de los residuos escolares en competencia lectora



En la Figura 15, se aprecia una leve aproximación a la heterocedasticidad. El recorrido de la varianza fue relativamente constante, aunque algunos valores se alejan de forma moderada del “cilindro” generado por cada variable (x, y).

Figura 15

Nube de puntos de los residuos escolares en competencia lectora



4.1.3 Modelos multinivel en competencia matemática

El modelo nulo en competencia matemática se expresó por medio de la Ecuación 8.

$$CMatemática_{ij} = \gamma_{00} + u_{0j} + r_{ij} \quad (8)$$

Donde:

$CMatemática_{ij}$ estuvo compuesta por los 10 valores plausibles en la competencia matemática de cada estudiante i en una escuela j ;

γ_{00} es el intercepto (o gran media) en competencia matemática de todos los estudiantes i incluidos en escuelas j ;

u_{0j} es el residuo en competencia matemática de cada escuela j ; y

r_{ij} es el residuo en competencia matemática del estudiante i en una escuela j .

Tras haber generado el modelo nulo, se obtuvo la estimación final de los componentes de varianza (ver Tabla 22) para el cálculo del CCI inicial.

Tabla 22

Componentes de varianza del modelo nulo en competencia matemática

Efecto aleatorio	Desviación estándar	Componentes de varianza	g.l.	x^2	p-valor
Nivel-2	45.389	2060.206	267	3973.683	<.001
Nivel-1	62.852	3950.409			

En función de los valores ubicados en la tercera columna de la Tabla 22, se cuantificó el CCI inicial en competencia matemática (Ecuación 9). El CCI sobrepasó el mínimo sugerido (10%0 para la aplicación de Modelos Multinivel (Lee, 2000): el 34.3% de la variabilidad en el rendimiento matemático de los estudiantes mexicanos fue atribuible a nivel centro.

$$ICC_{CMatemática} = \frac{2060.206}{2060.206 + 3950.409} = 0.343 \quad (9)$$

El modelo condicional relativo al logro en la competencia matemática quedó definido por la Ecuación 10.

$$CMatemática_{ij} = \gamma_{00} + \gamma_{01}SchIMMIG + \gamma_{02}SchESCS + \gamma_{10}StuGen + \gamma_{20}StuIMMIG + \gamma_{30}StuESCS + \gamma_{40}StuRep + \gamma_{50}StuGra + u_{0j} + r_{ij} \quad (10)$$

Donde:

$CMatemática_{ij}$ estuvo compuesta por los 10 valores plausibles en la competencia matemática de cada estudiante i en una escuela j ;

γ_{00} es el intercepto (o gran media) en competencia matemática de todos los estudiantes i incluidos en escuelas j ;

$\gamma_{10} - \gamma_{q0}$ son covariables a nivel estudiante;

$\gamma_{01} - \gamma_{0q}$ son covariables a nivel escuela;

u_{0j} es el residuo en competencia matemática de cada escuela j ; y

r_{ij} es el residuo en competencia matemática del estudiante i en una escuela j .

En la Tabla 23, se muestran los coeficientes, el error estándar, el t -ratio y el p -valor de las variables predictoras que incluyeron el modelo condicional final en competencia matemática. En función de su peso sobre la variable criterio (t -ratio), las variables con mayor influencia fueron el *estatus migratorio del estudiante*, el *estatus de repetidor del estudiante*, el *género del estudiante*, el *ISEC medio en la escuela*, el *grado escolar del estudiante*, el *ISEC del estudiante* y la *tasa de migrantes en la escuela*.

Tabla 23

Modelo condicional en competencia matemática con estimación robusta del error

Efecto fijo	Coefficiente	Error estándar	t -ratio	p -valor
Intercepto γ_{00}	441.154	7.198	61.289	<.001
Tasa de migrantes en la escuela γ_{01}	-109.104	52.679	-2.071	.039
ISEC medio en la escuela γ_{02}	15.625	5.011	3.118	.002
Género del estudiante γ_{10}	-13.694	4.271	-3.206	.003
Estatus migratorio del estudiante γ_{20}	-44.670	13.690	-3.263	.002
ISEC del estudiante γ_{30}	4.824	2.027	2.380	.02
Estatus de repetidor del estudiante γ_{40}	-26.866	8.337	-3.222	.003
Grado escolar del estudiante γ_{50}	15.892	5.600	2.838	.005

Nota: ISEC = Índice social, económico y cultural.

La estimación final de los componentes de varianza del modelo condicional final en competencia matemática se encuentra en la Tabla 24.

Tabla 24

Componentes de varianza del modelo condicional en competencia matemática

Efecto aleatorio	Desviación estándar	Componentes de varianza	g.l.	X ²	p-valor
Nivel-2	27.387	750.031	265	1562.722	<.001
Nivel-1	62.437	3898.339			

El CCI final se calculó con base en los componentes de la tercera columna de la Tabla 24. Con este, se llegó a la conclusión de que: (a) los factores contextuales explicaron 53.06% de la variabilidad dentro y entre las escuelas; y (b) aún era posible explicar el 46.95% de la varianza.

$$ICC_{CMatemática} = \frac{750.031}{750.031 + 3898.339} = 0.161 \quad (11)$$

4.1.3.1 Residuos en competencia matemática

Por último, se estimaron los residuos escolares en competencia matemática, la diferencia entre el resultado esperado en una escuela con características contextuales idénticas y el resultado “real” obtenido por la escuela. Los descriptivos de estos valores se encuentran en la Tabla 25.

Tabla 25

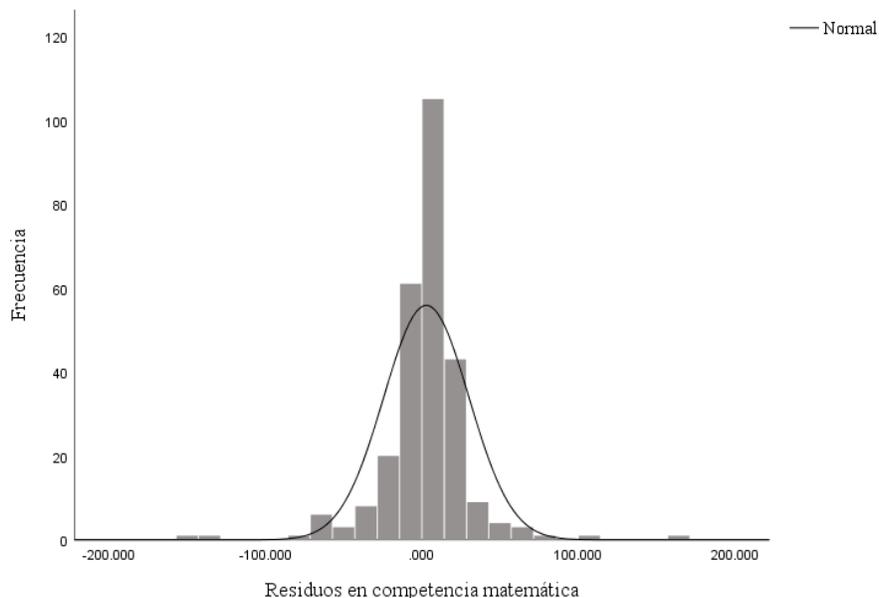
Descriptivos de los residuos escolares en competencia matemática

	N	Mínimo	Máximo	Media	Desv. Est.	Asimetría		Curtosis	
						Valor	Desv. Error	Valor	Desv. Error
Residuos	268	-148.514	165.681	-2.831	27.323	-0.366	.149	10.77	.297

Con la prueba Kolmogorov-Smirnov se obtuvo un valor menor a .001, por lo tanto, la distribución de los residuos se alejaba levemente de la curva normal. En la Figura 16, se muestra que ello pudo ser consecuencia de un gran número de valores alrededor de la media, similar a las distribuciones leptocúrticas.

Figura 16

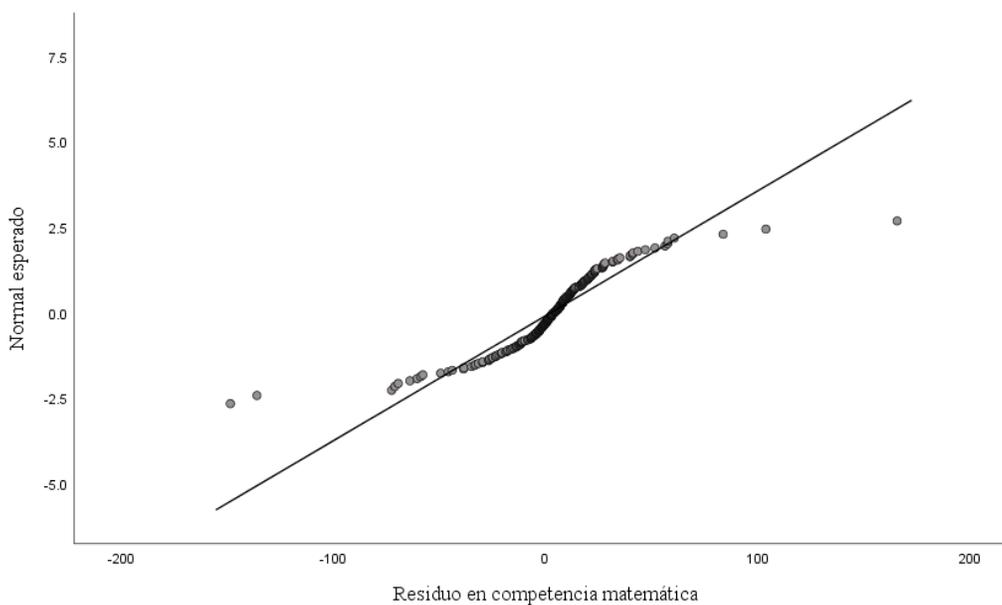
Histograma de los residuos escolares en competencia matemática



Como se aprecia en la Figura 17, la mayor parte de los valores se aproximan a la normal esperada, habiendo algunos que se alejan de forma moderada y extrema de esta recta.

Figura 17

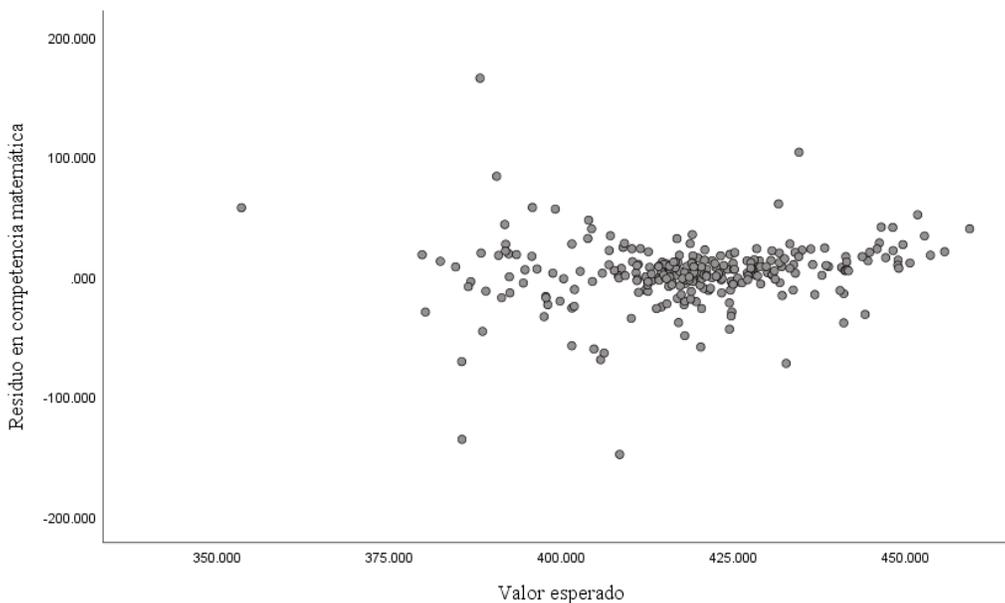
Gráfico Q-Q de los residuos escolares en competencia matemática



La nube de puntos de la Figura 18, muestra una distribución con una ligera aproximación a la heterocedasticidad. El recorrido de la varianza fue relativamente constante, pese a la presencia de algunos *outliers*.

Figura 18

Nube de puntos de los residuos escolares en competencia matemática



4.1.4 Modelos multinivel en competencia científica

El modelo nulo en competencia científica se expresó con la Ecuación 12.

$$CCientífica_{ij} = \gamma_{00} + u_{0j} + r_{ij} \quad (12)$$

Donde:

$CCientífica_{ij}$ estuvo compuesta por los 10 valores plausibles en la competencia científica de cada estudiante i en una escuela j ;

γ_{00} es el intercepto (o gran media) en competencia científica de todos los estudiantes i incluidos en escuelas j ;

u_{0j} es el residuo en competencia científica de cada escuela j ; y

r_{ij} es el residuo en competencia científica del estudiante i en una escuela.

Tras efectuar el modelo nulo, se obtuvo la estimación final de los componentes de varianza (ver Tabla 26).

Tabla 26

Componentes de varianza del modelo nulo en competencia científica

Efecto aleatorio	Desviación estándar	Componentes de varianza	g.l.	x^2	p -valor
Nivel-2	39.546	1563.906	267	3804.510	<.001
Nivel-1	58.633	3437.831			

En función de los valores incluidos en la tercera columna de la Tabla 26, se cuantificó el CCI inicial. Como se observa en la Ecuación 13, este rebasó el 10% recomendado para la aplicación de Modelos Multinivel (Lee, 2000): el 31.3% de la variabilidad en competencia científica fue atribuible a nivel centro.

$$ICC_{CCientífica} = \frac{1563.906}{1563.906 + 3437.831} = 0.313 \quad (13)$$

El modelo condicional, para el rendimiento en competencia científica de los estudiantes, quedó definido por la Ecuación 14.

$$CCientífica_{ij} = \gamma_{00} + \gamma_{01}SchESCS + \gamma_{02}SchRep + \gamma_{10}StuGen + \gamma_{20}StuIMMIG + \gamma_{30}StuESCS + \gamma_{40}StuRep + u_{0j} + r_{ij} \quad (14)$$

Donde:

$CCientífica_{ij}$ estuvo compuesta por los 10 valores plausibles en la competencia científica de cada estudiante i en una escuela j ;

γ_{00} es el intercepto (o gran media) en competencia científica de todos los estudiantes i incluidos en escuelas j ;

$\gamma_{10} - \gamma_{q0}$ son covariables a nivel estudiante;

$\gamma_{01} - \gamma_{0q}$ son covariables a nivel escuela;

u_{0j} es el residuo en competencia científica de cada escuela j ; y

r_{ij} es el residuo en competencia científica del estudiante i en una escuela j .

En la Tabla 27, se presentan los coeficientes, el error estándar, el t -ratio y el p -valor de las variables predictoras que integraron el modelo condicional final en competencia científica. En orden de importancia, los factores contextuales que incidieron en la variable criterio fueron la *tasa de repetidores en la escuela*, el *ISEC medio en la escuela*, el *ISEC del estudiante*, el *género del estudiante*, el *estatus migratorio del estudiante* y el *estatus de repetidor del estudiante*.

Tabla 27

Modelo condicional en competencia científica con estimación robusta del error

Efecto fijo	Coefficiente	Error estándar	t -ratio	p -valor
Intercepto γ_{00}	448.650	6.334	70.828	<.001
ISEC medio en la escuela γ_{01}	14.054	4.181	3.362	.001
Tasa de repetidores en la escuela γ_{02}	-41.187	10.610	-3.882	<.001
Género del estudiante γ_{10}	-11.044	3.513	-3.144	.002
Estatus migratorio del estudiante γ_{20}	-39.601	14.196	-2.790	.006
ISEC del estudiante γ_{30}	5.364	1.634	3.282	.002
Estatus de repetidor del estudiante γ_{40}	-12.839	6.438	-1.994	.047

Nota: ISEC = Índice social, económico y cultural.

Tras construir el modelo condicional en competencia científica, se obtuvo la estimación final de los componentes de varianza (ver Tabla 28).

Tabla 28

Componentes de varianza del modelo condicional en competencia científica

Efecto aleatorio	Desviación estándar	Componentes de varianza	g.l.	x^2	p -valor
Nivel-2	23.836	568.151	265	1496.443	<.001
Nivel-1	58.612	3435.419			

A partir de los elementos incluidos en la tercera columna de la Tabla 28, se calculó el CCI final (Ecuación 15). Con este, se pudo llegar a dos conclusiones: (a) los factores contextuales explicaron 54.63% de la variabilidad dentro y entre las escuelas; y (b) aún era posible explicar el 45.37% de la varianza.

$$ICC_{CCcientífica} = \frac{568.151}{568.151 + 3435.419} = 0.142 \quad (15)$$

4.1.4.1 Residuos en competencia científica

Finalmente, se estimaron los residuos escolares en competencia científica, la diferencia entre el rendimiento esperado en una escuela con características contextuales idénticas y el resultado “verdadero” obtenido por la escuela. Los estadísticos descriptivos de los valores obtenidos se encuentran en la Tabla 29.

Tabla 29

Descriptivos de los residuos escolares en competencia científica

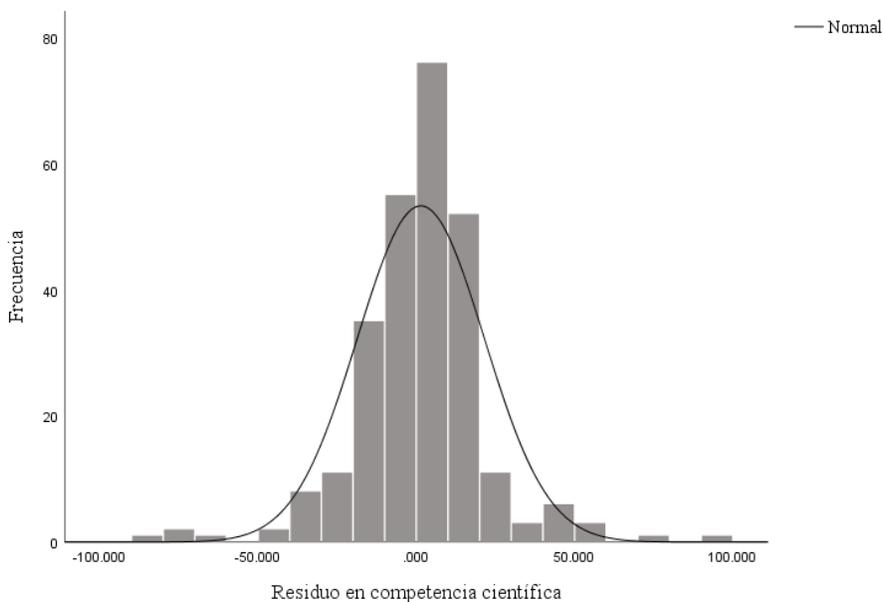
	N	Mínimo	Máximo	Media	Desv. Est.	Asimetría		Curtosis	
						Valor	Desv. Error	Valor	Desv. Error
Residuos	268	-81.804	93.978	1.417	20.053	-0.115	.149	4.244	.297

El índice de Kolmogorov-Smirnov fue menor a .001, es decir, la distribución de los residuos se desvió ligeramente de la curva normal. En la Figura 19 se aprecia que ello pudo deberse

a una alta concentración de valores en torno a la media, una característica de las distribuciones leptocúrticas.

Figura 19

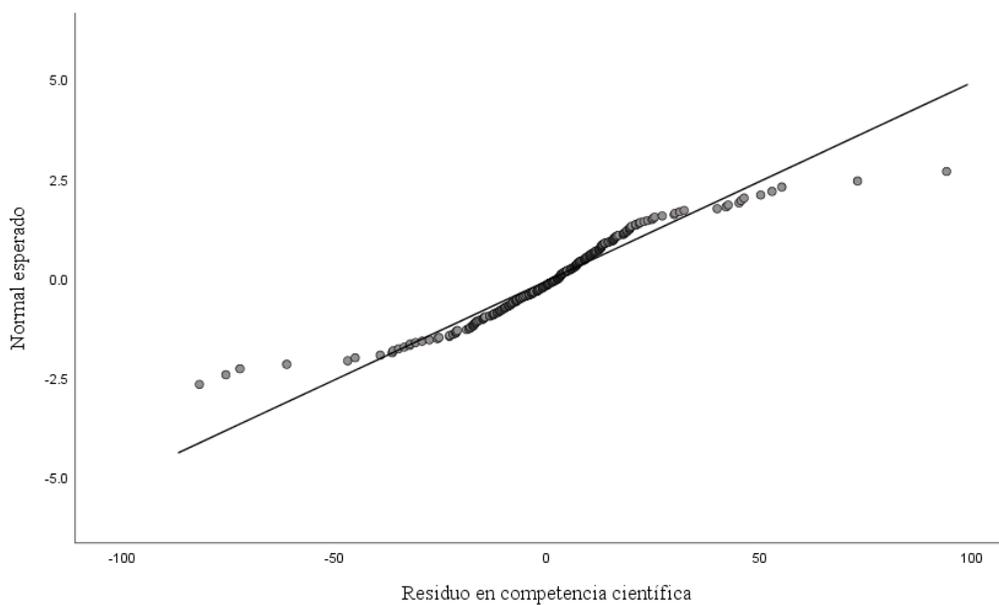
Histograma de los residuos escolares en competencia científica



En la Figura 20 se observa que gran parte de los valores fluctúan en torno a la normal, encontrando algunos valores atípicos moderados y extremos que la pudieron haber producido.

Figura 20

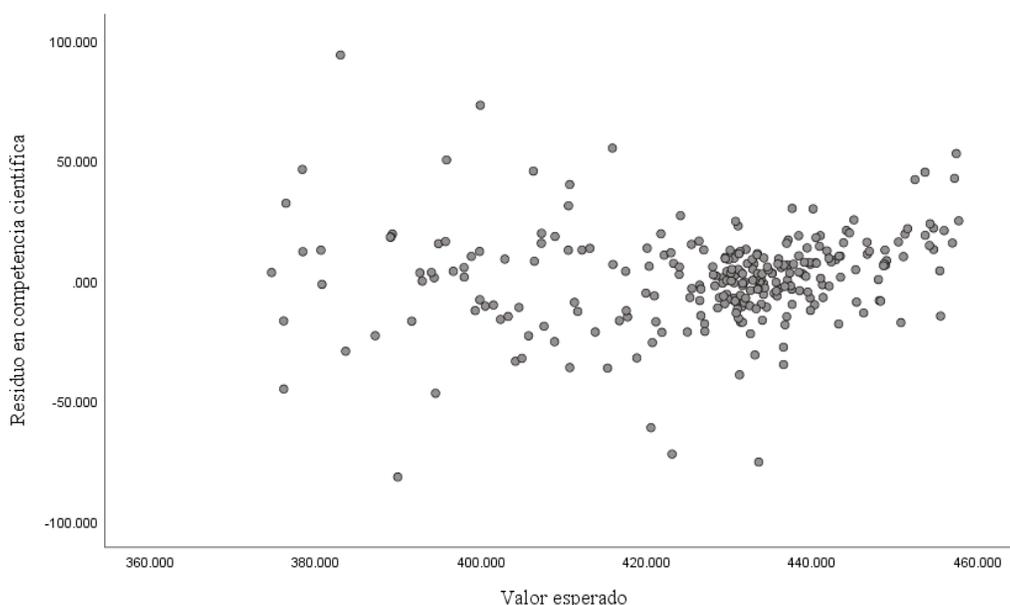
Gráfico Q-Q de los residuos escolares en competencia científica



En la Figura 21 se aprecia una leve aproximación a la heterocedasticidad. El recorrido de la varianza fue relativamente constante, aunque algunos valores se alejan de forma moderada.

Figura 21

Nube de puntos de los residuos escolares en competencia científica



Como se puede ver en la Tabla 30, hubo coincidencias y diferencias entre las variables que integraron los tres modelos multinivel correspondientes a cada competencia evaluada en PISA. Por un lado, hubo variables que coincidieron en los tres modelos: el *estatus de repetidor*, el *estatus migratorio*, el *ISEC del estudiante* y el *ISEC medio escolar*. Por otro lado, algunas de ellas solo coincidieron en dos modelos, por ejemplo: el *género del estudiante*, en el caso de la competencia matemática y científica, el *grado escolar del estudiante*, en el caso de la competencia lectora y matemática, y la *tasa de repetidores*, en el caso de la competencia lectora y científica. Únicamente la *tasa de migrantes* se asoció significativamente en un modelo multinivel, el correspondiente a la competencia matemática.

Tabla 30

Resumen de las variables que integraron cada modelo multinivel

Variables	Modelo por competencia		
	Lectora	Matemática	Científica
Estatus de repetidor del estudiante	*	*	*
Estatus migratorio del estudiante	*	*	*
Género del estudiante		*	*
Grado escolar del estudiante	*	*	
ISEC del estudiante	*	*	*
ISEC medio en la escuela	*	*	*
Tasa de migrantes en la escuela		*	
Tasa de repetidores en la escuela	*		*

4.2 Fase 2: Clasificación de la Eficacia Escolar

La Fase 2 tuvo por objetivo la clasificación de la eficacia de las escuelas en función de su residuo en el nivel de competencia lectora, matemática y científica de sus estudiantes en PISA 2018. De manera que, a continuación, se presenta la estandarización y normalización de los residuos, la aplicación del algoritmo K-medias, la evaluación y caracterización de las agrupaciones.

4.2.1 Transformación de los residuos escolares

Se realizaron tres transformaciones de Johnson, una por cada grupo de residuos obtenidos en la Fase 1: en competencia lectora, matemática y científica.

4.2.1.1 En competencia lectora

La distribución de los residuos en competencia lectora se identificó en la familia SU y, por ende, la transformación aplicada correspondió a la Ecuación 3.

$$Z = \gamma + \eta \operatorname{senh}^{-1} \left(\frac{X - \varepsilon}{\lambda} \right) \quad (3)$$

Donde:

$$\gamma = .102727$$

$$\lambda = 13.54047$$

$$\varepsilon = 3.632437$$

$$\eta = .9362725$$

Como se muestra en la Tabla 31, los residuos en competencia lectora transformados oscilaron entre -2.867 y 3.333, contaron con media cercana a cero y desviación estándar próxima a uno.

Tabla 31

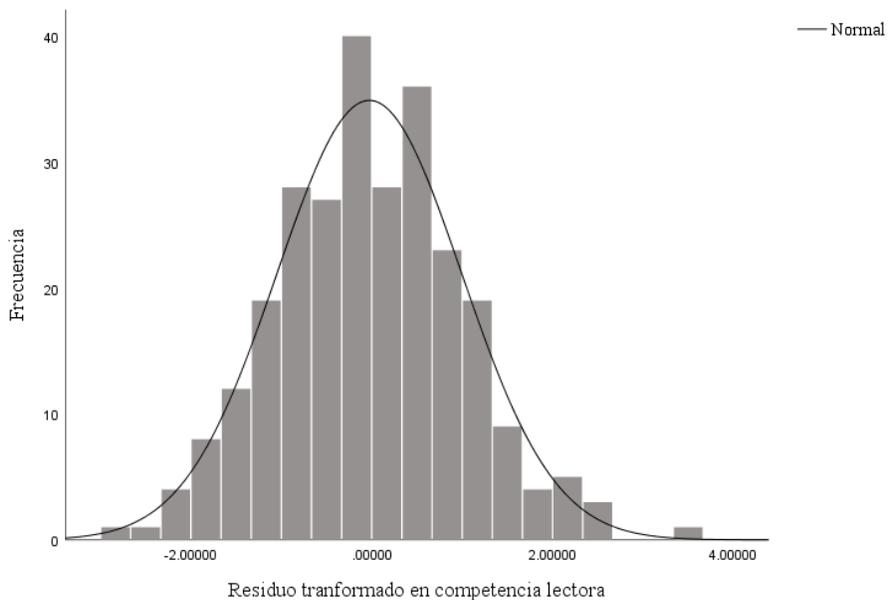
Descriptivos de los residuos escolares en competencia lectora transformados

	N	Mínimo	Máximo	Media	Desv. Est.	Asimetría		Curtosis	
						Valor	Desv. Error	Valor	Desv. Error
Residuos	268	-2.867	3.333	-0.027	1.021	.089	.149	.085	.297

El índice Kolmogorov-Smirnov fue de .2, es decir, los residuos se distribuyeron de forma semejante a la normal. Ello se corroboró a través de la Figura 22, donde el histograma se apegó fielmente a la campana de Gauss.

Figura 22

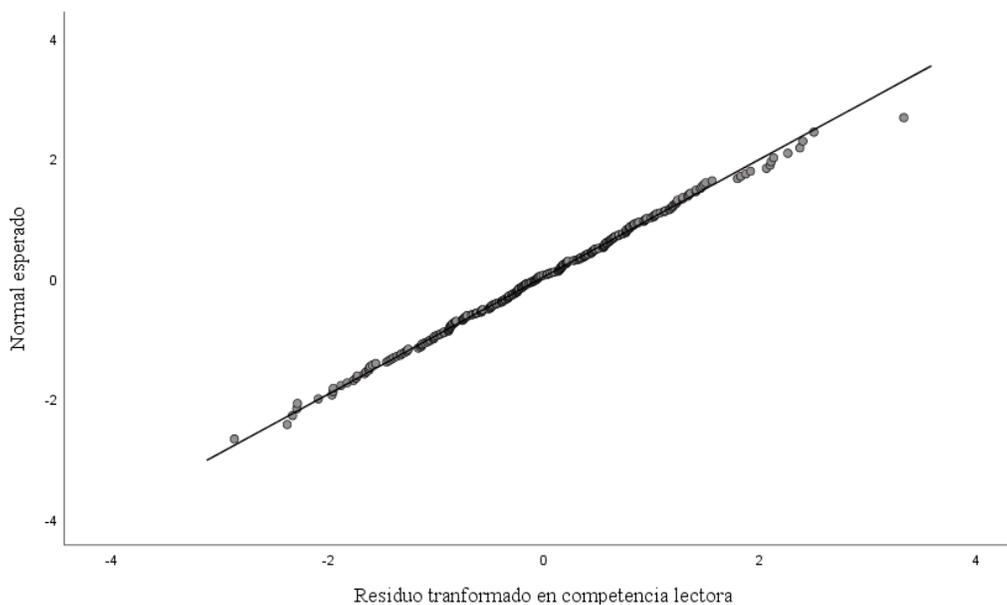
Histograma de los residuos escolares en competencia lectora transformados



La Figura 23 refleja que la mayoría de los valores se alinean sobre la diagonal que representa la normal, encontrándose mínimos *outliers*.

Figura 23

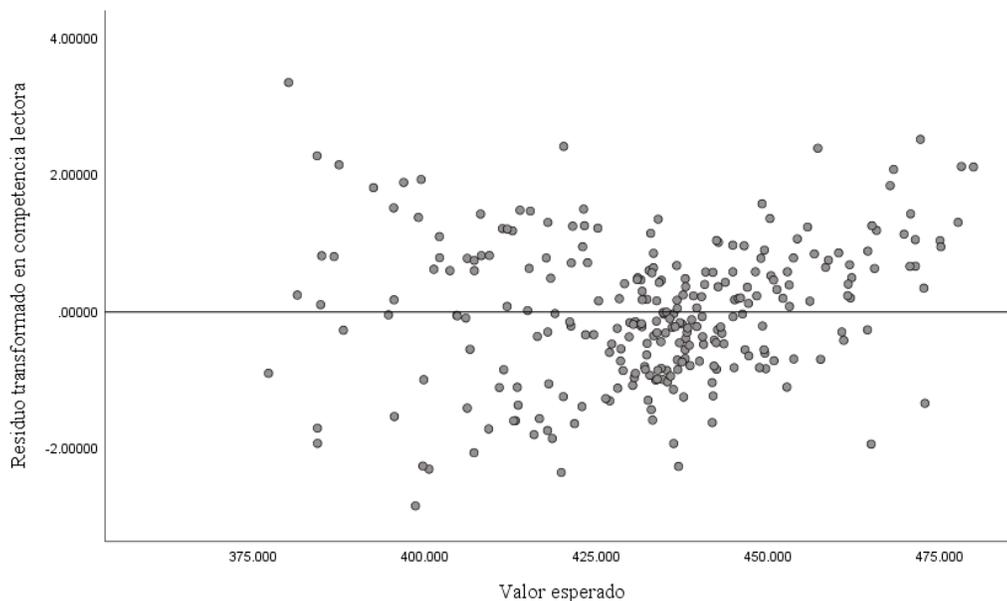
Gráfico Q-Q de los residuos escolares en competencia lectora transformados



En la Figura 24 se distingue que los residuos escolares transformados cumplen con la homocedasticidad y linealidad. La varianza es relativamente constante en el recorrido de los valores (X, Y).

Figura 24

Nube de puntos de los residuos escolares en competencia lectora transformados



4.2.1.2 En competencia matemática

La distribución de los residuos en competencia matemática se identificó en la familia SU y, en consecuencia, la transformación aplicada correspondió a la Ecuación 3.

$$Z = \gamma + \eta \operatorname{senh}^{-1} \left(\frac{X - \varepsilon}{\lambda} \right) \quad (3)$$

Donde:

$$\gamma = .09420382$$

$$\lambda = 11.40906$$

$$\varepsilon = 6.087694$$

$$\eta = .8552218$$

Como se distingue en la Tabla 32, los residuos en competencia matemática transformados fluctuaron entre -2.729 y 2.944, poseyeron media colindante a cero y desviación estándar cercana a uno.

Tabla 32

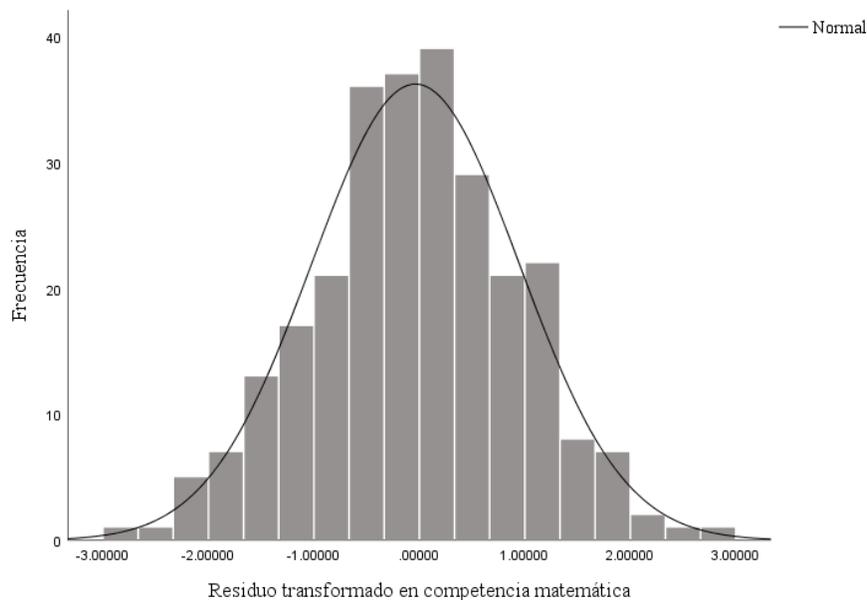
Descriptivos de los residuos escolares en competencia matemática transformados

	N	Mínimo	Máximo	Media	Desv. Est.	Asimetría		Curtosis	
						Valor	Desv. Error	Valor	Desv. Error
Residuos	268	-2.729	2.944	-0.373	.984	.012	.149	-0.029	.297

La prueba Kolmogorov-Smirnov presentó un valor de .2, esto es que la distribución de los residuos en competencia matemática correspondió a la curva normal. Ello se ve representado en la Figura 25.

Figura 25

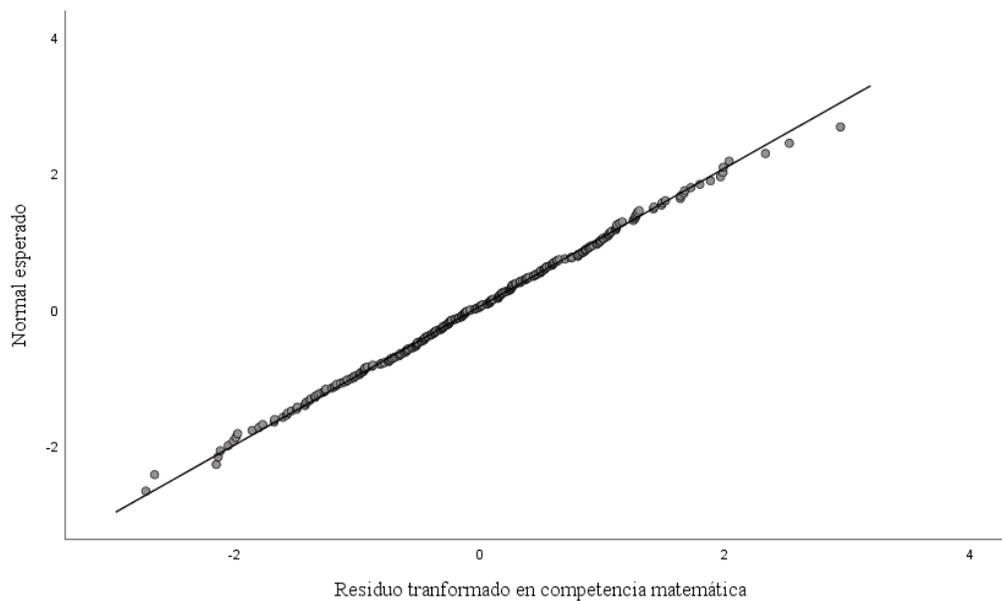
Histograma de los residuos escolares en competencia matemática transformados



Como se muestra en el gráfico Q-Q de la Figura 26, la mayoría de los valores coinciden con la diagonal que representa la curva normal. Una cantidad mínima de valores fueron relativamente atípicos.

Figura 26

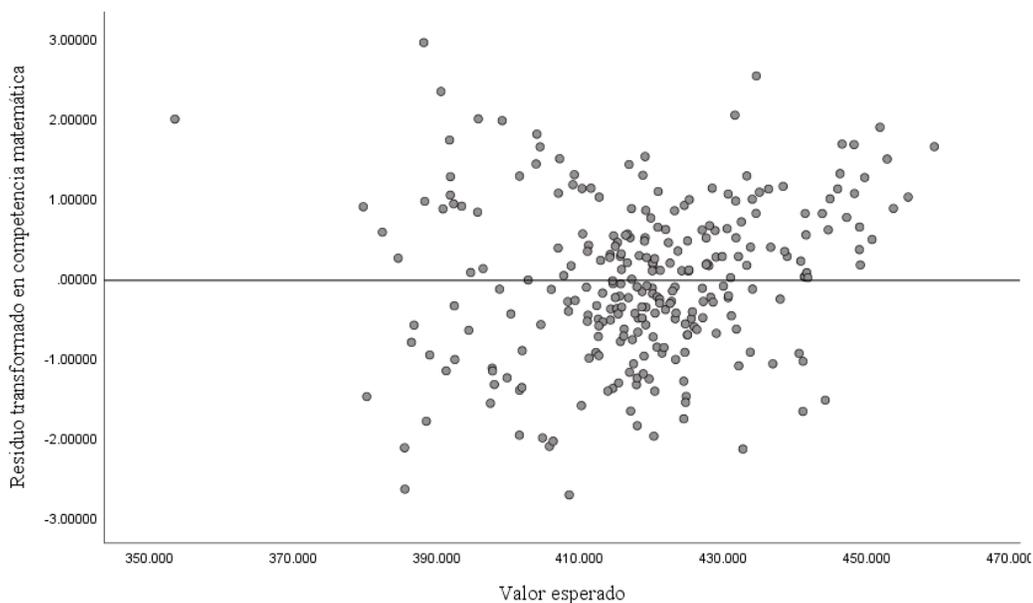
Gráfico Q-Q de los residuos escolares en competencia matemática transformados



En la Figura 27, se observa que la homocedasticidad y linealidad de los residuos se cumple. La varianza se mantiene parcialmente constante en el recorrido de los valores X.

Figura 27

Nube de puntos de los residuos escolares en competencia matemática transformados



4.2.1.3 En competencia científica

La distribución de los residuos en competencia científica se identificó en la familia SU y, por lo tanto, la transformación aplicada correspondió a la Ecuación 3.

$$Z = \gamma + \eta \operatorname{senh}^{-1} \left(\frac{X - \varepsilon}{\lambda} \right) \quad (3)$$

Donde:

$$\gamma = .1712988$$

$$\lambda = 14.94085$$

$$\varepsilon = 4.286547$$

$$\eta = 1.090435$$

Como se identifica en la Tabla 33, los residuos en competencia matemática transformados variaron entre -2.502 y 2.889, tuvieron una media contigua a cero y una desviación estándar próxima a uno.

Tabla 33

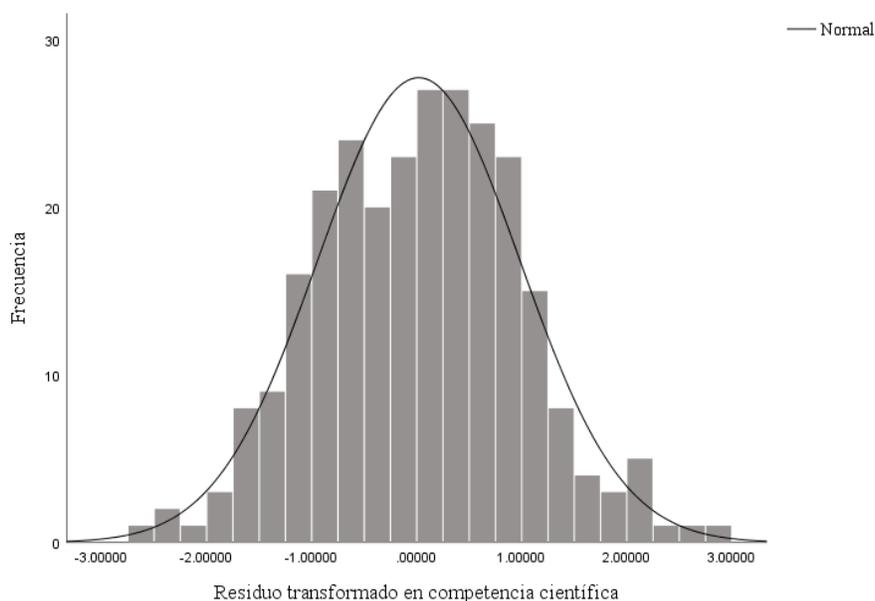
Descriptivos de los residuos escolares en competencia científica transformados

	N	Mínimo	Máximo	Media	Desv. Est.	Asimetría		Curtosis	
						Valor	Desv. Error	Valor	Desv. Error
Residuos	268	-2.502	2.889	.019	.964	.08	.149	-0.014	.297

La prueba Kolmogorov-Smirnov presentó un valor de .2, es decir, la distribución de los residuos en competencia científica se aproxima a la normalidad Gaussiana. Ello puede corroborarse en la Figura 28.

Figura 28

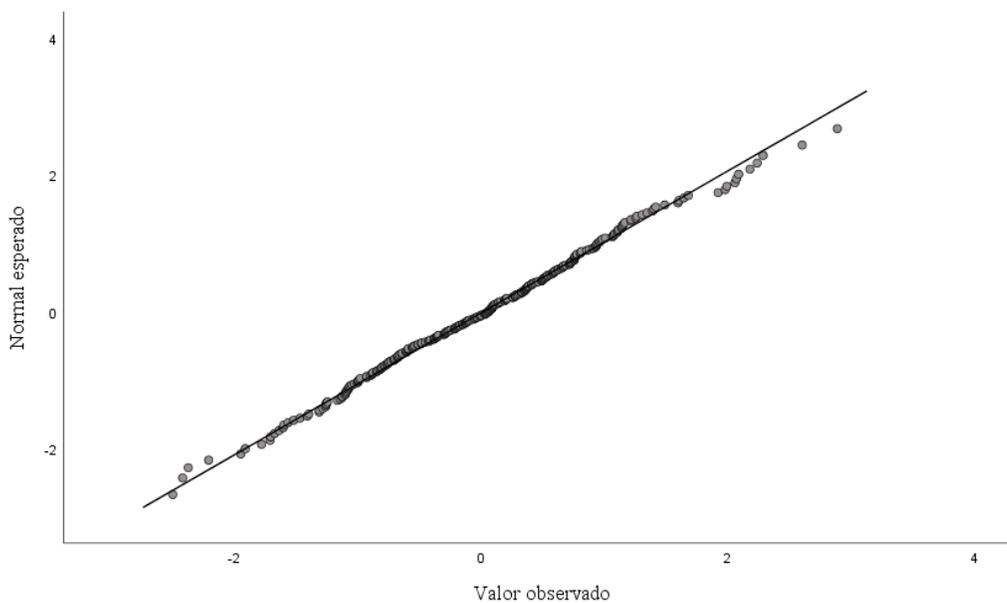
Histograma de los residuos escolares en competencia científica transformados



Como se observa en la Figura 29, la mayoría de los valores se ajustan a la diagonal que representa la normal, habiendo una cantidad mínima de residuos que se alejan de la misma.

Figura 29

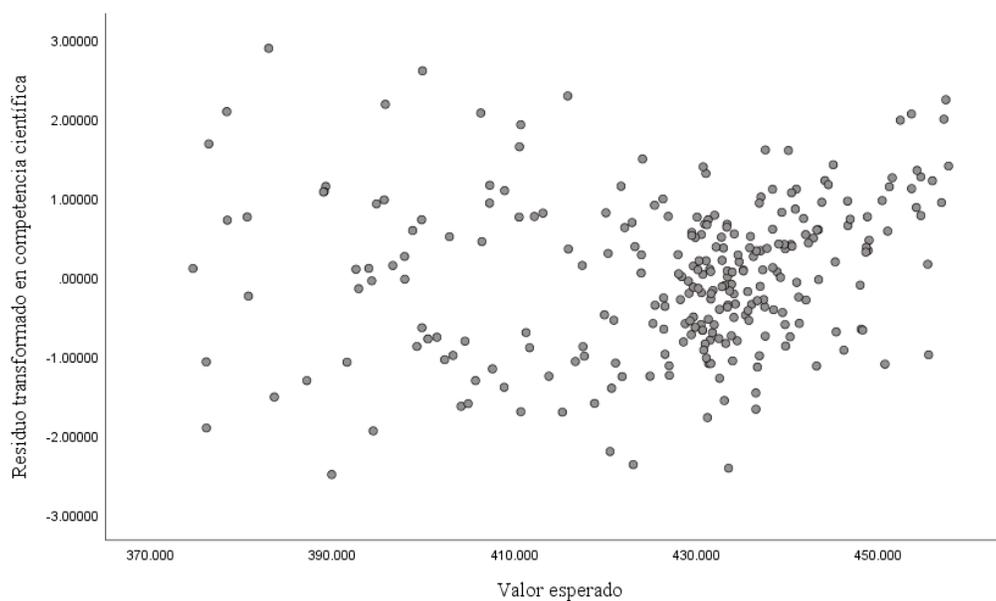
Gráfico Q-Q de los residuos escolares en competencia científica transformados



En la Figura 30 se aprecia que se cumple con la homocedasticidad y linealidad de los residuos en competencia científica. La varianza es relativamente constante a lo largo del eje X.

Figura 30

Nube de puntos de los residuos escolares en competencia científica transformados



4.2.2 Conglomeración

Tras aplicar el algoritmo de K-medias sobre los residuos escolares transformados (en competencia lectora, matemática y científica) y poner a prueba los tres criterios propuestos en el método, tres agrupaciones se determinaron como el número adecuado de conglomerados (ver Tabla 34).

Tabla 34

Clasificaciones de la eficacia de las escuelas

	Media de la distancia euclidiana de cada clúster	Desviación estándar de cada clúster	n
	-1.091388254	0.051530122	76
	-0.008283629	0.059652192	120
	1.110118504	0.017883776	72
Total	0.003482207	0.04302203	268

Con base en la media de las distancias euclidianas se nombró a cada agrupación (ver Tabla 35). “Baja Eficacia” a las escuelas con media cercana a (-1), “Residuo centrado” a aquellas con media aproximada a (0), y “Alta eficacia” a los centros escolares con una media próxima a (1).

Tabla 35

Distancia entre centroides de clústers finales

Residuo transformado \ Clúster	-1 Baja eficacia	0 Residuo centrado	1 Alta eficacia
Competencia lectora	-1.14984	-0.01007	1.13020
Competencia matemática	-1.07182	-0.06702	1.10426
Competencia científica	-1.05251	0.05224	1.09590

A través de la prueba Post hoc Bonferroni se corroboró que existen diferencias significativas entre todas las agrupaciones obtenidas (ver Tabla 36).

Tabla 36

Comparaciones múltiples entre agrupaciones por competencia: post hoc bonferroni

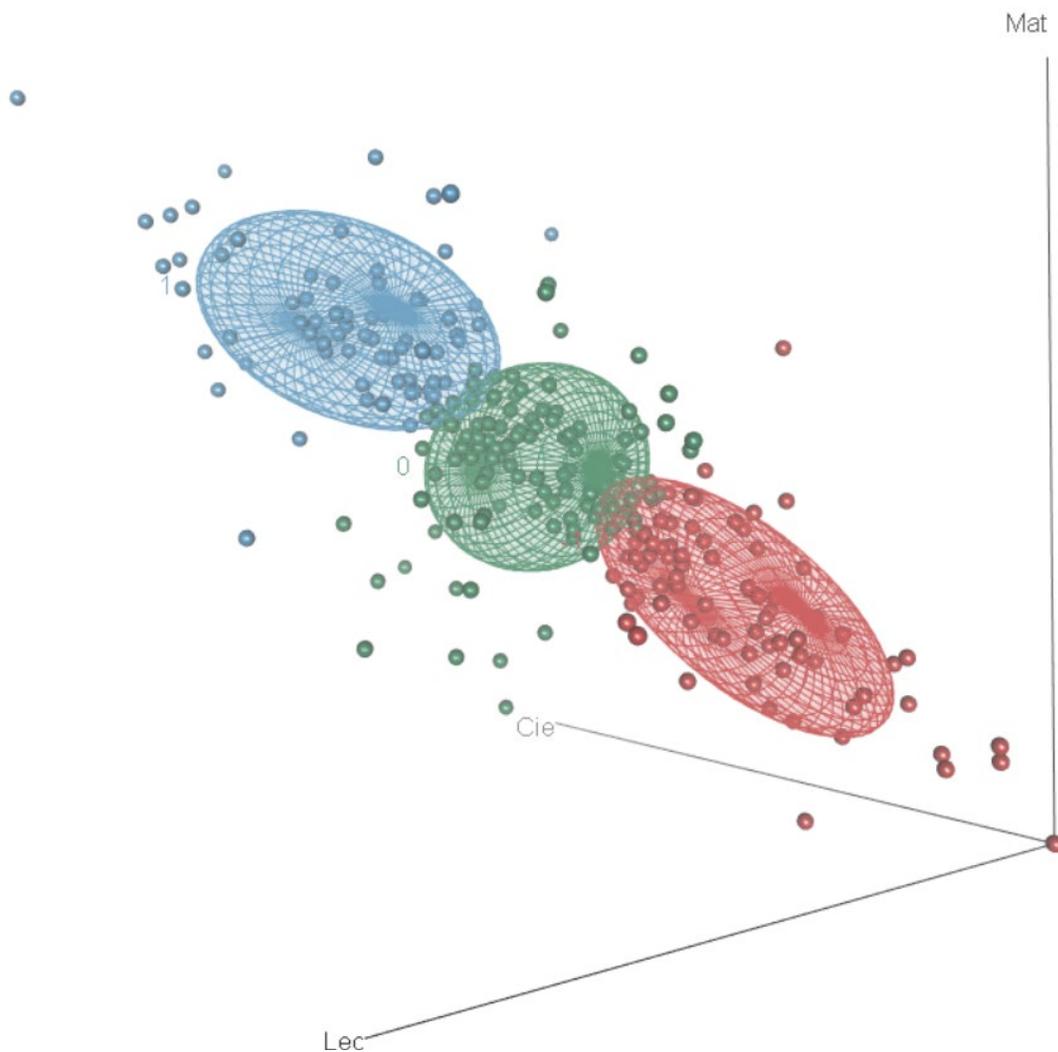
Residuo transformado	(I) clúster	(J) clúster	Diferencia de medias (I-J)	Desv. Estándar	Sig.
Competencia lectora	-1	0	-1.1398*	.0835	<.001
	0	1	-1.1403*	.0849	<.001
	1	-1	2.2800*	.0937	<.001
Competencia matemática	-1	0	-1.0048*	.0821	<.001
	0	1	-1.1713*	.0835	<.001
	1	-1	2.1761*	.0921	<.001
Competencia científica	-1	0	-1.1047*	.0790	<.001
	0	1	-1.0437*	.0804	<.001
	1	-1	2.1484*	.0887	<.001

Nota: * la diferencia de medias es significativa.

En la nube de puntos de tres dimensiones (Figura 31) puede apreciarse cómo se distribuyeron las agrupaciones. Las escuelas de “Baja eficacia” están en color rojo, las de “Residuo centrado” en verde, y las de “Alta eficacia” en azul.

Figura 31

Gráfico 3D de los clústers de los residuos escolares transformados



Nota: Alta eficacia (1; azul), Eficacia esperada (0; verde), y Baja eficacia (-1; rojo).

4.2.3 Caracterización de las agrupaciones

En la Tabla 37 se encuentran los descriptivos de las agrupaciones generadas. Los tres conglomerados son semejantes en relación con el *porcentaje de mujeres* (52.3%, 51% y 53.9%), *tasa de migrantes* (2.2%, 1% y 1.7%), *tasa de repetidores* (16.3%, 15.6% y 13.6%), *escasez de recursos* (.714, .442 y .29), *escasez de personal* (-0.133, .083 y -0.084) y *ratio de estudiantes por docente* (aprox. 33, 34 y 31). No obstante, al menos dos agrupaciones difirieron significativamente

en el *índice social, económico y cultural medio* (baja y alta eficacia), *tipo de escuela* (residuo centrado con baja y alta eficacia), *tamaño de la escuela* (residuo centrado con baja y alta eficacia) y *tamaño de las aulas en la escuela* (residuo centrado con baja y alta eficacia).

Tabla 37

Descriptivos de las escuelas de baja eficacia, residuo centrado y alta eficacia

	Eficacia Escolar						Dif. sig.
	Baja eficacia ^b		Residuo centrado ^c		Alta eficacia ^a		
	Media	Recuento	Media	Recuento	Media	Recuento	
Porcentaje de mujeres en la escuela	.523		.51		.54		
Tasa de migrantes en la escuela	.02		.01		.02		
Índice social, económico y cultural medio en la escuela	-1.32		-1.16		-.94		.034 ^{ba}
Tasa de repetidores en la escuela	.16		.16		.14		
Tipo de escuela	Pública	60		113		56	.009 ^{bc}
	Privada	16		7		16	.005 ^{ca}
Escasez de recursos en la escuela	.71		.44		.29		
Escasez de personal en la escuela	-.13		.08		-.08		
Tamaño de la escuela	483		1427		815		<.001 ^{bc}
							.001 ^{ca}
Tamaño de las aulas en la escuela	39		44		38		.006 ^{bc}
							.001 ^{ca}
Ratio de estudiantes por docente en la escuela	32.69		33.96		31.01		

Nota: La media y el recuento se realizó con una tabla cruzada. La diferencia significativa se estimó con el Post hoc Bonferroni. Las variables originalmente cualitativas se interpretan como porcentajes: porcentaje de mujeres en la escuela, tasa de migrantes en la escuela y tasa de repetidores en la escuela.

Las agrupaciones, a su vez, se caracterizaron por diferentes niveles de rendimiento estudiantil en las tres competencias evaluadas en la prueba PISA. Como se puede observar en la Tabla 38, en las escuelas de baja eficacia los estudiantes de 15 años de edad obtuvieron puntajes menores, ubicándose -los mismos- entre los 380 y 382 puntos en la competencia lectora, 371 y 377 en la matemática, y 383 y 386 en la científica. En cambio los pupilos de matriculados en escuelas de alta eficacia obtuvieron los mayores puntajes, registrándose entre los 456 y 460 puntos en

competencia la lectora, 439 y 445 en la matemática, y 451 y 455 en la científica. En las de residuo centrado el alumnado se ubicó en un punto intermedio entre los estudiantes inscritos en las escuelas de baja y alta eficacia, habiendo obtenido entre 420 y 422 puntos en competencia la lectora, 406 y 411 en la matemática, y 419 y 421 en la científica.

Tabla 38

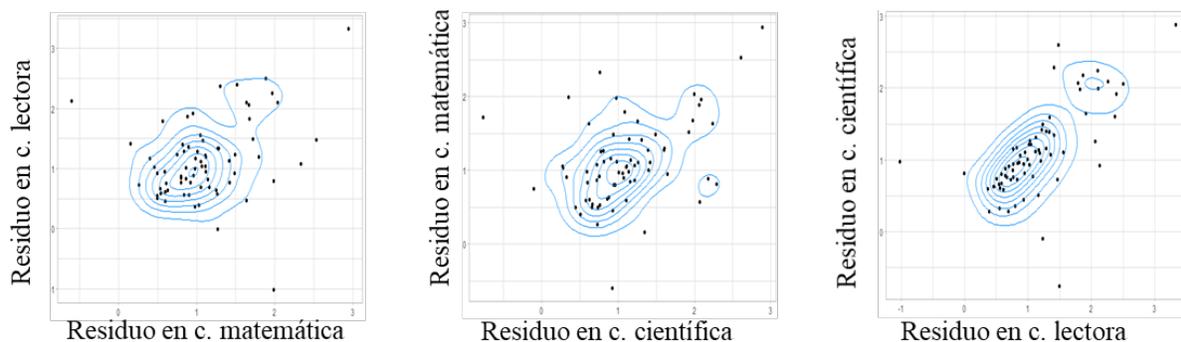
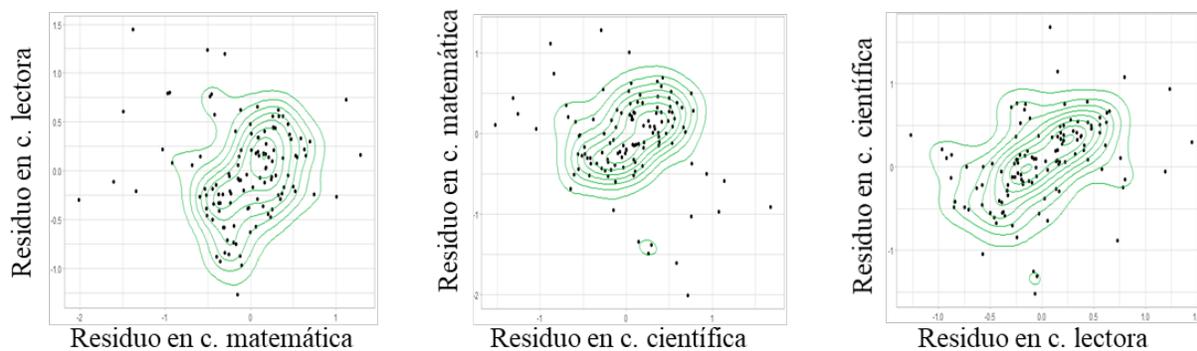
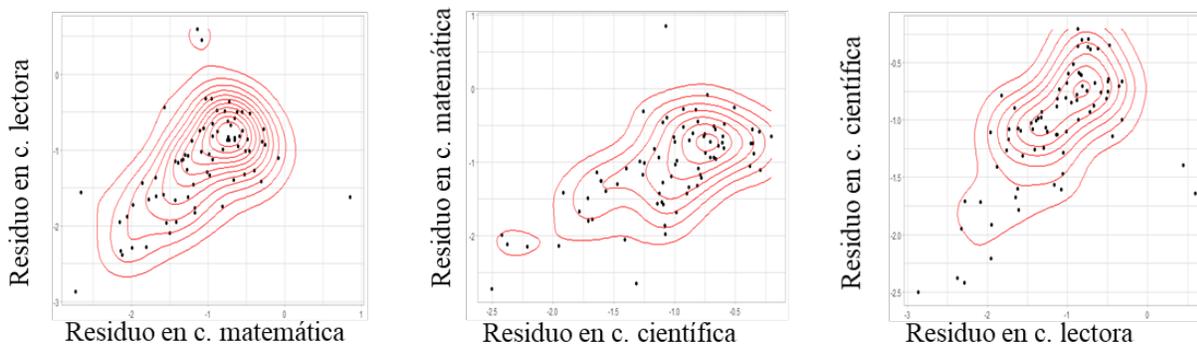
Rendimiento de los estudiantes por clúster

	Lectora		Matemática		Científica	
	Media mínima	Media máxima	Media mínima	Media máxima	Media mínima	Media máxima
Baja eficacia	380	382	371	377	383	386
Residuo centrado	420	422	406	411	419	421
Alta eficacia	456	460	439	445	451	455

Nota: la media mínima y la media máxima se obtuvo a partir del cálculo de medias en cada uno de los valores plausibles por competencia.

4.2.4 Evaluación de la calidad de las agrupaciones

Como se apreciar en las Figuras 32, 33, y 34 la calidad de las agrupaciones es relativamente aceptable. Únicamente en el tercer contraste marginal de la Figura 33 se observa un doble núcleo denso, lo cual, tiene un impacto en la calidad del clúster “Residuo centrado”. En lo que respecta a los clústers “Alta eficacia” y “Baja eficacia”, los contrastes marginales presentaron algunos núcleos externos, ello debido a algunas agrupaciones de valores atípicos.

Figura 32*Contrastes marginales en escuelas de alta eficacia***Figura 33***Contrastes marginales en escuelas de residuo centrado***Figura 34***Contrastes marginales en escuelas de baja eficacia*

4.3 Fase 3: Análisis del Peso Explicativo

La Fase 3 tuvo por objetivo analizar el valor explicativo de las variables de enseñanza sobre la eficacia de las escuelas.

4.3.1 Diagnóstico de colinealidad entre variables explicativas

Tal como se observa en la Tabla 39, únicamente la variable *estimulación docente* presentó problemas de colinealidad. Su VIF superó por al menos tres décimas el criterio propuesto por Garson (2014) y Gordon (2014). En consecuencia, esta variable fue excluida de análisis posteriores.

Tabla 39

Diagnóstico de colinealidad entre variables explicativas

	VARIABLES EXPLICATIVAS	TOLERANCIA	VIF
De enseñanza	Adaptación de la instrucción	0.299	3.349
	Apoyo docente	0.271	3.686
	Clima disciplinario en las lecciones	0.621	1.610
	Comportamientos del docente que dificultan el aprendizaje	0.844	1.185
	Estimulación docente	0.236	4.229
	Instrucción directa	0.290	3.452
	Interés docente	0.390	2.563
	Tiempo promedio en clase	0.856	1.168
	Retroalimentación	0.334	2.994
	Uso de las TIC en clase	0.696	1.437
Escolares	Actividades extracurriculares	0.949	1.053
	Calidad escolar	0.367	2.724
	Clima escolar discriminatorio	0.823	1.215
	Oferta de clases adicionales	0.912	1.096
	Políticas escolares para el involucramiento de la familia	0.387	2.584
	Uso de las TIC en la escuela	0.685	1.460

Nota: TIC = Tecnologías de la Información y Comunicación. Las variables con un VIF > 4 presentaron problemas de colinealidad.

4.3.2 Regresiones logísticas multinomiales

A continuación, se presentan los resultados obtenidos a través de las tres regresiones logísticas multinomiales. La primera incluyó únicamente explicativas escolares; la segunda sólo explicativas de enseñanza; y la tercera explicativas escolares y de enseñanza que resultaron significativas en la primera y segunda regresión. Sin embargo, previo a los análisis de regresión, se estimaron correlaciones entre los residuos en competencia lectora, matemática y científica y el conjunto de variables para dar mayor sustento a las interpretaciones (ver Tabla 40).

Tabla 40

Correlación entre residuos y variables explicativas

Variables explicativas	Lectora		Matemática		Científica	
	<i>r</i>	Sig.	<i>r</i>	Sig.	<i>r</i>	Sig.
Adaptación de la instrucción	.126	.039	.185	.002	.156	.011
Apoyo docente	.076	.213	.129	.034	.119	.051
Clima disciplinario en las lecciones	.174	.004	.22	>.001	.2	.001
Comportamientos del docente que dificultan el aprendizaje	-.044	.469	-.049	.43	-.03	.622
Instrucción directa	-.097	.113	-.062	.309	-.017	.784
Interés docente	.15	.014	.184	.002	.172	.005
Tiempo promedio en clase	-.09	.16	-.013	.835	-.011	.863
Retroalimentación	-.17	.005	-.084	.172	-.117	.056
Uso de las TIC en clase	-.068	.267	.039	.525	.025	.688
Actividades extracurriculares	.028	.648	-.014	.821	-.023	.706
Calidad escolar	.104	.089	.144	.018	.141	.021
Clima escolar discriminatorio	-.196	.002	-.218	>.001	-.239	>.001
Oferta de clases adicionales	.058	.341	.034	.577	.117	.056
Políticas escolares para el involucramiento de la familia	.034	.576	.089	.145	.091	.136
Uso de las TIC en la escuela	-.057	.353	.009	.888	.046	.458

Nota: Se utilizó Pearson con pares de variables cuantitativas y Biserial-Puntual con pares de variables dicotómico-cuantitativas.

Como se puede observar en la Tabla 40, nueve variables presentaron correlaciones significativas sobre al menos uno de los residuos: la *adaptación de la instrucción*, el *apoyo docente*, el *clima disciplinario*, el *interés docente*, la *retroalimentación*, la *calidad escolar* y el *clima escolar discriminatorio*. No obstante, según Cohen (1998), todas estas fueron correlaciones “pequeñas”.

4.3.2.1 Explicativas escolares

Como se puede apreciar en la Tabla 41, de las cinco variables escolares que se incluyeron en la regresión, únicamente el *clima escolar discriminatorio* explicó significativamente el pertenecer a una escuela de “Baja eficacia” y “Alta eficacia”.

Tabla 41

Modelos de regresión logística multinomial de las variables explicativas escolares

		Reg. 1	Reg. 2	Reg. 3	Reg. 4	Reg. 5	Final
Eficacia Escolar	Descripción de la variable	Sig.	Sig.	Sig.	Sig.	Sig.	Sig.
Baja Eficacia	Clima escolar discriminatorio	.352	.538	.443	.446	.598	.603
	Políticas escolares para el involucramiento de la familia	.306	.336	.860	.938	.825 ^x	
	Actividades extracurriculares creativas	.129	.186	.212	.188 ^x		
	Oferta de clases adicionales	.223	.248	.283 ^x			
	Calidad escolar	.324	.259 ^x				
	Uso de las TIC en la escuela	.631 ^x					
Alta Eficacia	Clima escolar discriminatorio	>.001	>.001	>.001	>.001	>.001	>.001
	Políticas escolares para el involucramiento de la familia	.411	.252	.035	.040	.074 ^x	
	Actividades extracurriculares creativas	.171	.239	.271	.246 ^x		
	Oferta de clases adicionales	.383	.503	.489 ^x			
	Calidad escolar	.366	.645 ^x				
	Uso de las TIC en la escuela	.489 ^x					

Nota: Se marcaron con negritas aquellas variables con un efecto significativo. ^x indicó la variable que se eliminó por regresión.

4.3.2.2 Explicativas de enseñanza

Como se presenta en la Tabla 42, de las variables de enseñanza: la adaptación de la instrucción, el clima disciplinario en las lecciones, el tiempo promedio en clase, la retroalimentación docente y el uso de las TIC durante la clase explicaron la alta y baja eficacia de las escuelas.

Tabla 42

Modelos de regresión logística multinomial de las variables explicativas de enseñanza

		Reg. 1	Reg. 2	Reg. 3	Reg. 4	Final
Eficacia Escolar	Descripción de la variable	Sig.	Sig.	Sig.	Sig.	Sig.
Baja Eficacia	Adaptación de la instrucción	.011	.010	.004	.006	.001
	Retroalimentación	.001	.001	.001	>.001	>.001
	Tiempo promedio en clase	.344	.313	.274	.273	.27
	Clima disciplinario en las lecciones	.296	.249	.268	.243	.338
	Uso de las TIC durante la clase	.84	.842	.807	.816	.665
	Interés docente	.105	.106	.094	.196 ^x	
	Instrucción directa	.086	.09	.137 ^x		
	Apoyo docente	.394	.394 ^x			
	Comportamiento docente	.613 ^x				
Alta Eficacia	Adaptación de la instrucción	.864	.897	.709	.815	.383
	Retroalimentación	.249	.281	.275	.079	.15
	Tiempo promedio en clase	.009	.008	.008	.007	.007
	Clima disciplinario en las lecciones	.009	.006	.005	.005	.002
	Uso de las TIC durante la clase	.011	.011	.012	.006	.007
	Interés docente	.080	.077	.052	.116 ^x	
	Instrucción directa	.097	.089	.133 ^x		
	Apoyo docente	.398	.389 ^x			
Comportamiento docente	.568 ^x					

Nota: Se marcaron con negritas aquellas variables con un efecto significativo. ^x indicó la variable que se eliminó por regresión.

4.3.2.3 Explicativas escolares y de enseñanza

En la tercera regresión, las variables escolares y de enseñanza se mantuvieron como explicativas de la eficacia de las escuelas (Tabla 43). La *adaptación de la instrucción* y la *retroalimentación* se asociaron de forma significativas en la baja eficacia; en cambio, el *tiempo promedio de clase*, el *clima disciplinario*, el *uso de las TIC* y el *clima escolar discriminatorio* fueron significativas en relación con la alta eficacia de las escuelas.

Tabla 43

Modelo final de regresión logística multinomial de las variables escolares y de enseñanza

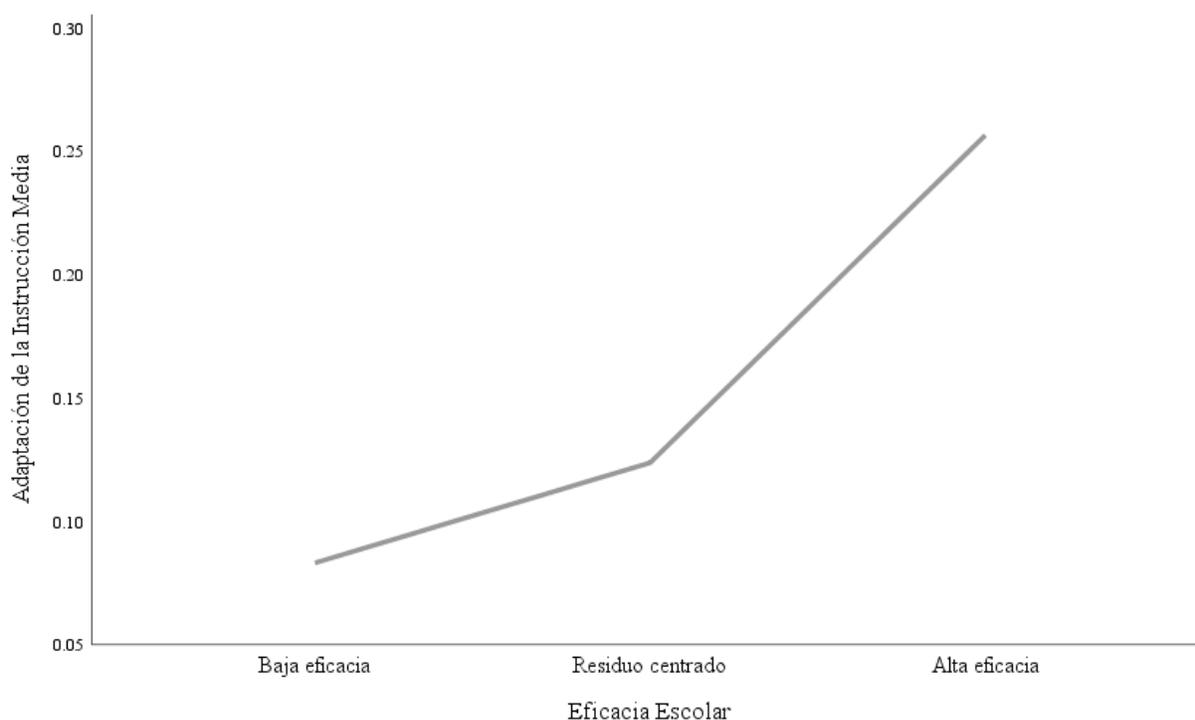
		Reg. 1			
Eficacia Escolar	Descripción de la variable	<i>p</i> -valor	β	<i>Odds Ratio</i>	<i>Odds Ratio inversa</i>
Baja Eficacia	Adaptación de la instrucción	.001	-2.941	.053	18.937
	Retroalimentación	<.001	2.955	19.199	-
	Tiempo promedio en clase	.262	-0.017	.983	1.018
	Clima disciplinario en las lecciones	.338	.512	1.668	-
	Uso de las TIC durante la clase	.598	.290	1.336	-
Alta Eficacia	Clima escolar discriminatorio	.943	.029	1.030	-
	Adaptación de la instrucción	.737	.290	1.337	-
	Retroalimentación	.327	-0.757	.469	2.133
	Tiempo promedio en clase	.010	-0.043	.958	1.044
	Clima disciplinario en las lecciones	.018	1.315	3.725	-
	Uso de las TIC durante la clase	.018	1.302	3.678	-
	Clima escolar discriminatorio	.001	-1.605	.201	4.977

Nota: Se marcaron con negritas aquellas variables con un efecto significativo. Cuando es mayor a 1, Odds Ratio indica el aumento probabilístico por unidad; sin embargo, cuando el Odds Ratio es menor a 1, Odds Ratio inversa indica el aumento probabilístico por unidad.

Por cada aumento en la *adaptación de la instrucción*, la probabilidad de ser una escuela de baja eficacia disminuyó en 18.937 unidades; en cambio, aumentó en 1.337 unidades la probabilidad de ser una escuela de alta eficacia. Como puede verse en la Figura 35, en comparación con la categoría residuo centrado, este factor cuenta con mayor presencia en las escuelas de alta eficacia que en las de baja. Desde una perspectiva ordinal, pareciese que mayores valores de la variable explicativa se relacionan de forma lineal con la variable a explicar.

Figura 35

Gráfico de líneas de la variable adaptación de la instrucción por clasificación de eficacia

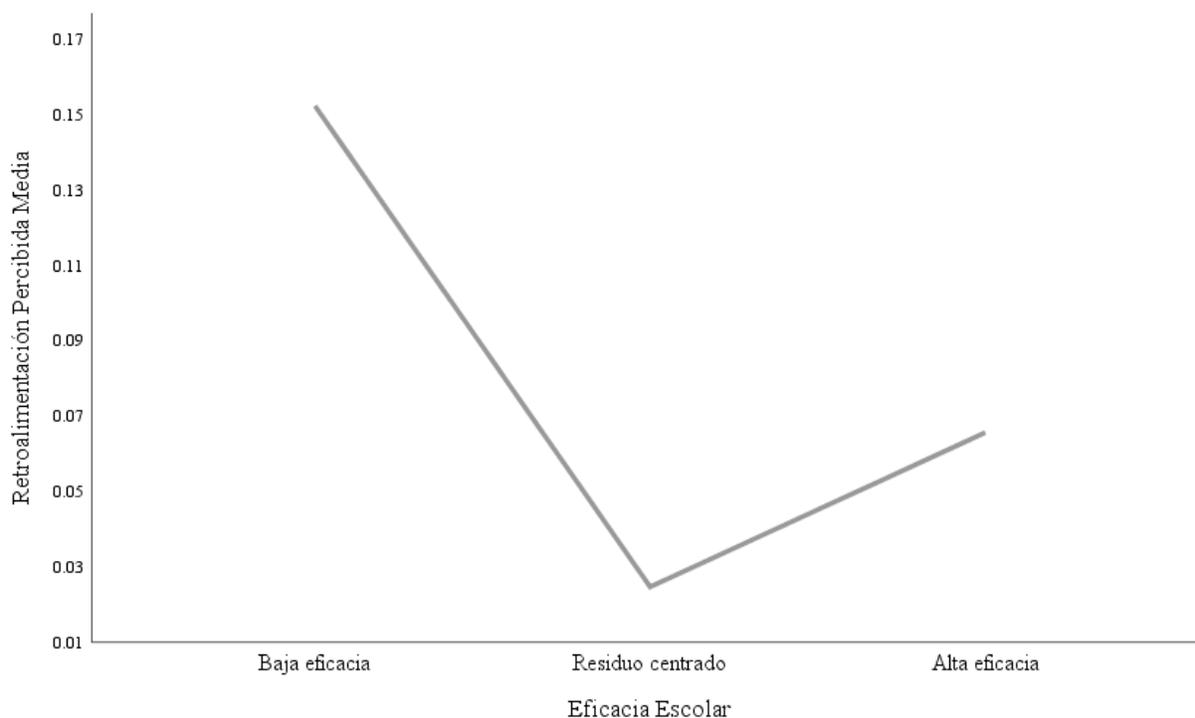


Por cada incremento en la *retroalimentación percibida*, la probabilidad de ser un centro de baja eficacia aumentó en 19.199 unidades y respecto a ser de alta eficacia disminuyó en 2.133 unidades. Tal como se observa en la Figura 36, en comparación a las escuelas consideradas de residuo centrado y alta eficacia, este factor fue notablemente superior en las de baja eficacia.

Contrario a la *adaptación de la instrucción*, no se aprecia una linealidad respecto a la variable a explicar.

Figura 36

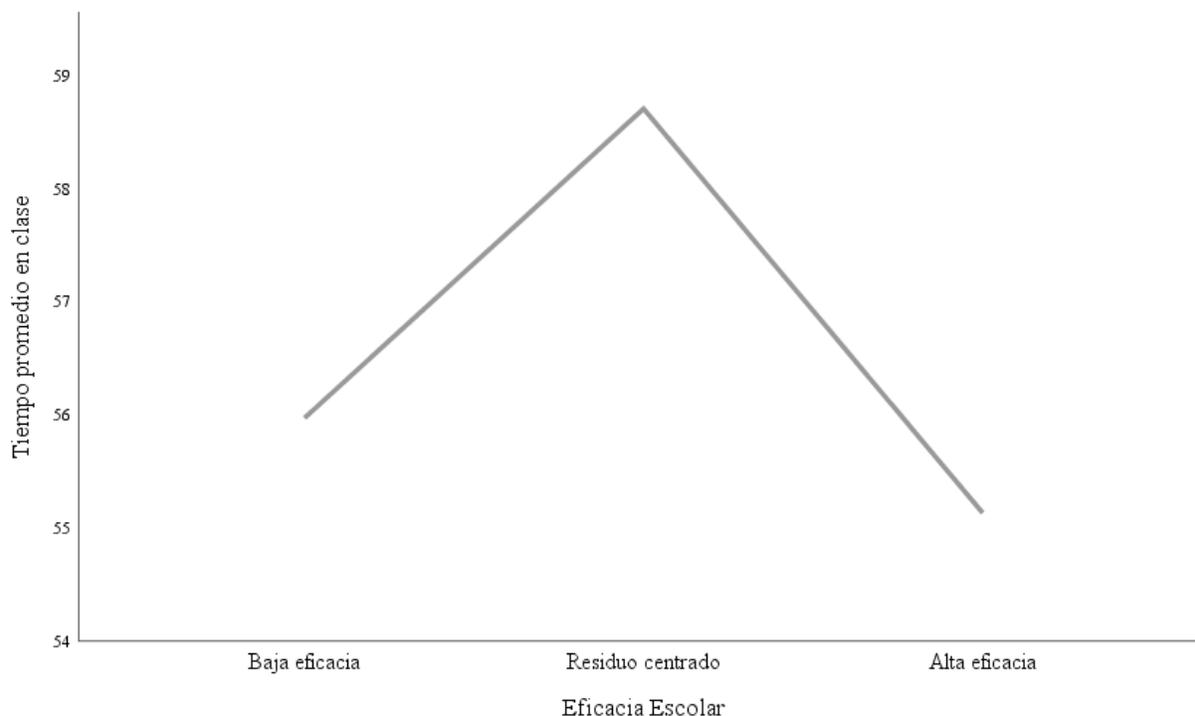
Gráfico de líneas de la variable retroalimentación percibida por clasificación de eficacia



El *tiempo promedio en la clase* disminuyó la probabilidad de ser de baja y de alta eficacia en 1.018 unidades y 1.044 unidades, respectivamente. Este resultado fue contraintuitivo, dado que esta variable explicó negativamente ambos supuestos de eficacia. Ello se vio reflejado en la Figura 37, donde efectivamente las escuelas de residuo centrado tenían mayor *tiempo promedio en aula* que las escuelas de baja y alta eficacia.

Figura 37

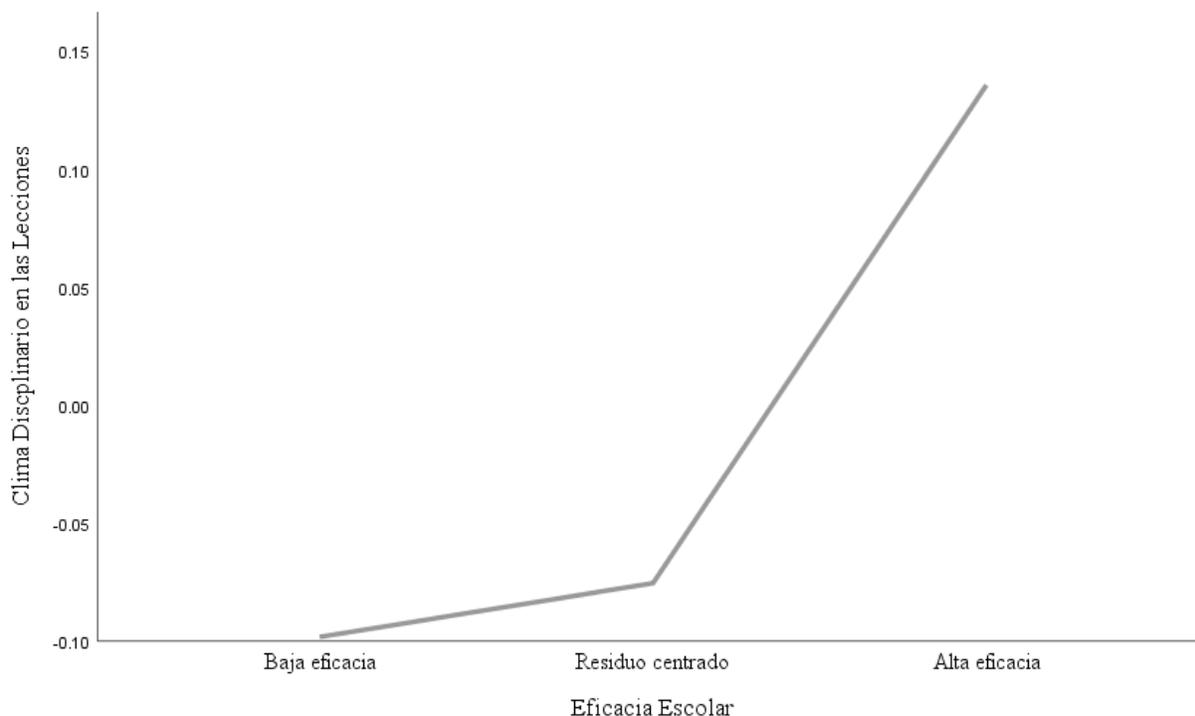
Gráfico de líneas de la variable tiempo promedio en el aula por clasificación de eficacia



El *clima disciplinario en las lecciones* aumentó la probabilidad de ser de baja y de alta eficacia en 1.668 y 3.725 unidades, respectivamente. Al igual que *el tiempo de aprendizaje*, este resultado fue contraintuitivo pese a su leve aproximación a la linealidad, capturada en la Figura 38. Con esta Figura, a su vez, se percibe que (en promedio) hubo un mayor *clima disciplinario en las lecciones* de las escuelas identificadas como de alta eficacia, en contraste con las de baja y residuo centrado.

Figura 38

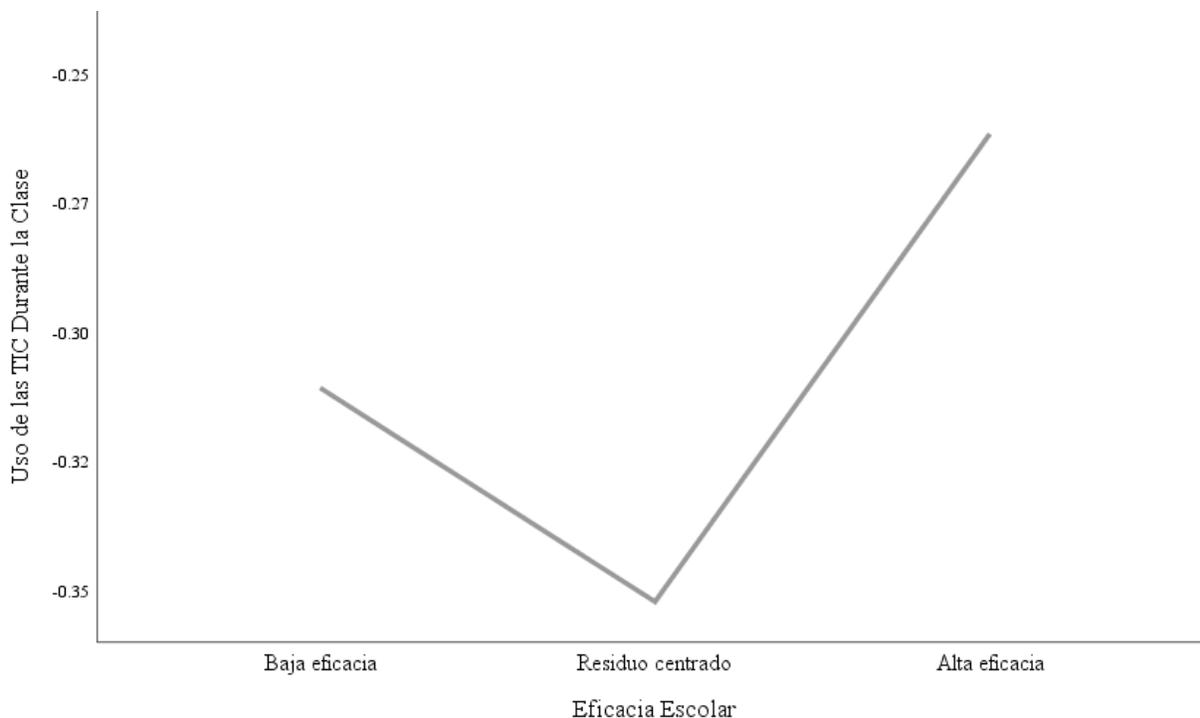
Gráfico de líneas de la variable clima disciplinario en las lecciones por clasificación de eficacia



El uso de las TIC durante la clase incrementó la probabilidad de ser una escuela de baja y de alta eficacia en 1.336 y 3.678 unidades, respectivamente. De forma semejante al tiempo promedio y el clima disciplinario en las lecciones, este resultado fue contraintuitivo. La Figura 39 dio evidencia de ello, donde las escuelas distinguidas como de Residuo centrado tuvieron un menor uso de las TIC durante la clase, siendo en promedio las escuelas de baja y alta eficacia las que dan mayor uso a las TIC.

Figura 39

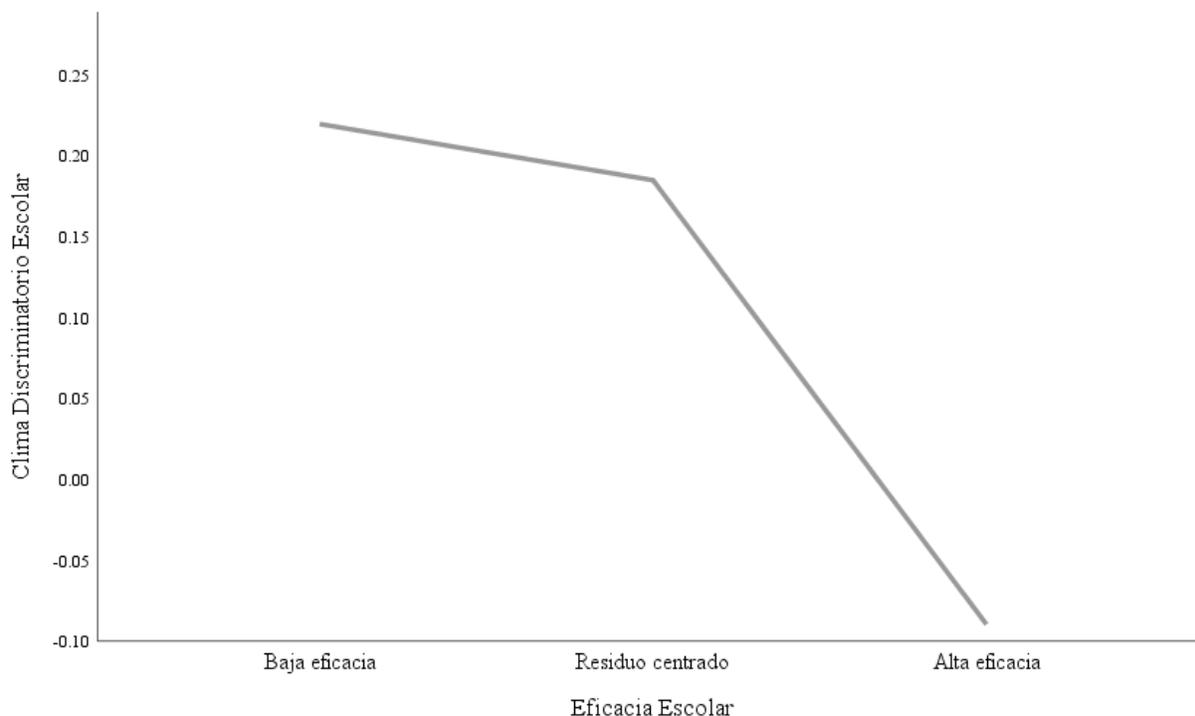
Gráfico de líneas de la variable uso de las TIC durante la clase por clasificación de eficacia



El *clima discriminatorio escolar* aumentó la probabilidad de ser un centro de baja eficacia en 1.030 y disminuyó la probabilidad de ser de alta en 4.977 unidades. Como puede percatarse en la Figura 40, este factor está más presente en las escuelas de baja eficacia y residuo centrado, en comparación con las de alta eficacia. Además, desde la ordinalidad, en esta Figura pareciese que a medida que disminuye el *clima discriminatorio escolar* también aumenta la eficacia de las escuelas.

Figura 40

Gráfico de líneas de la variable clima escolar discriminatorio por clasificación de eficacia



4.3.3 Precisión y exactitud del modelo de regresión final

Por último, con base en Landis y Koch (1977), pudo afirmarse que el modelo de regresión logística final multinomial fue justo, obteniéndose 0.249 en el valor de Kappa. Aunado a ello, el modelo obtuvo un 53.846% de precisión general en clasificaciones correctas de escuelas de baja y alta eficacia, donde, a su vez, se alcanzó una exactitud del 46.154% en los centros de baja eficacia, el 55.705% en los centros de residuo centrado y el 56.522% en los centros de alta eficacia (ver Tabla 44).

Tabla 44*Matriz de confusión del modelo de regresión logística multinomial final*

		Datos verdaderos			Clasificación general	Precisión del producto
		Clas. 1	Clas. 2	Clas. 3		
Resultados del clasificador	Clas. 1	24	18	10	52	46.154%
	Clas. 2	40	83	26	153	55.705%
	Clas. 3	7	13	26	42	56.522%
	Verdad en general	71	114	62	247	
	Precisión del usuario	33.803%	72.807%	41.935%		
Precisión general	53.846%					
Kappa	0.249					

Nota: El valor Kappa se interpretó con base en Landis y Koch (1977).

5 Discusión

El principal objetivo de este estudio fue la identificación de los factores de enseñanza que explican la Eficacia Escolar de los centros educativos mexicanos en PISA 2018. De las 10 hipótesis planteadas, solo cinco fueron aceptadas. Así, las hipótesis correspondientes al efecto de la *adaptación de la instrucción*, la *retroalimentación* dada por el docente, el *tiempo promedio en clase*, el *clima disciplinario en las lecciones* y el *uso de las TIC* fueron aceptadas al explicar de forma significativa el aumento de la probabilidad de ser una escuela de alta o baja eficacia. En cambio, las hipótesis nulas fueron aceptadas en el efecto de los factores relativos al *comportamiento de los docentes*, el *apoyo docente*, la *instrucción directa*, la *estimulación* y el *interés del profesor*.

Para poner a prueba las hipótesis, en esta investigación se precisó de tres fases de análisis: (1) la estimación de los residuos en competencia lectora, matemática y científica de los centros educativos mexicanos mediante la construcción de modelos multinivel; (2) la clasificación de la eficacia escolar de los centros educativos mexicanos de la aplicación de técnicas de conglomeración; y (3) el análisis del peso explicativo de los procesos educativos sobre la eficacia escolar de los centros educativos mexicanos con regresiones logísticas multinomiales. A continuación, se presentan las discusiones correspondientes a cada fase de análisis.

5.1 Fase 1: Estimación de Residuos

En los tres modelos nulos el porcentaje de varianza total que se encuentra sistemáticamente dentro y entre las escuelas (CCI) fue relativamente elevada y superó con creces mínimo recomendado por Lee (2000): 38.5% en la competencia lectora, 34.3% en la matemática y 31.3% en la científica. Estos resultados se contraponen a otros análisis multinivel efectuados con datos de Canadá, España, Estados Unidos y Finlandia, pues su CCI no ha superado el 21%; en cambio, se coincide en mayor medida con estudios efectuados con datos de Albania, China, India, Italia, Japón y Turquía, quienes han obtenido CCI entre el 30% y el 57% (Areepattamannil, 2014; Areepattamannil y Kaur, 2012; Bove et al., 2016; Cairns y Areepattamannil, 2017; Chen y Cui, 2019; Gamazo et al., 2017; Grabau y Ma, 2017; Guo et al., 2018; Ning et al., 2015; Shera, 2014; Yıldırım, 2012). De acuerdo con Krüger (2019), aunque es deseable explicar la mayor cantidad de varianza posible, que exista tanta variabilidad a nivel escuela puede deberse a altos índices de desigualdad social, económica y cultural; este planteamiento ha sido compartido por otros autores como Chaparro y Gamazo (2020), en el caso de México, y Quiroz et al. (2018), en el caso de Argentina.

En lo que refiere a los modelos multinivel finales, en competencia lectora, matemática y científica, las variables a nivel estudiante que resultaron significativas sobre el nivel de competencia estudiantil coincidieron con los hallazgos de estudios previos de otros autores. La variable de mayor peso en las tres competencias fue el *estatus migratorio del estudiante* (Cordero et al., 2015; Gamazo et al., 2017; Lau y Lam, 2017; Ma y Crocker, 2007; Meggiolaro, 2017). Este hallazgo puede alinearse con el planteamiento de Cordero et al. (2015) y Gamazo et al. (2017), quienes afirmaron que las discrepancias entre nativos y migrantes pueden deberse a un aumento en el ingreso de estudiantes provenientes de naciones no hispanohablantes. No obstante, este

argumento podría carecer de validez dado que la variable *lenguaje hablado* no resultó asociarse de forma significativa, por consiguiente, otra explicación podría ser que el “ser migrante” envuelve elementos que trascienden la lengua hablada, por ejemplo, la cultura con la que se formó el estudiante.

La siguiente variable con mayor influencia sobre las tres competencias fue el *índice social, económico y cultural del estudiante*. Este resultado era de esperarse, desde el Informe Coleman (1966), al día de hoy, existen innumerables evidencias sobre el impacto de este factor sobre los diversos resultados educativos. Por ejemplo, en la Revisión Sistemática de Literatura (RSL, incluida en este trabajo) esta variable se ubicó como la más referida entre los diferentes estudios del rendimiento en PISA (Areepattamannil y Kaur, 2012; Bove et al., 2016; Cairns y Areepattamannil, 2017; Chen y Cui, 2019; Cordero et al., 2015; Gamazo et al., 2017; Lau y Lam, 2017; Ma y Crocker, 2007; Meggiolaro, 2017; Ning et al., 2015; Shera, 2014; Yıldırım, 2012). Los autores sostienen que este hallazgo se explica por la aún elevada heterogeneidad económica que existe en los países participantes, tal es el caso de México (Cordero et al., 2015). De acuerdo con Ning et al. (2015), el efecto de este factor se verá disminuido en la medida en que la educación sea más equitativa para todos.

Otra variable estadísticamente significativa fue el *estatus de repetidor del estudiante*. A pesar de que Ikeda y García (2014) afirmaron que aún no existía evidencia concluyente sobre los efectos de la repetición de grado, en este estudio y en el de Gamazo et al. (2017) el ser repetidor fue una desventaja en comparación con aquellos que no han repetido un grado en al menos una ocasión. En las dos investigaciones, este hallazgo fue consistente en las tres competencias evaluadas por la prueba PISA: la lectora, la matemática y la científica. Esto no significa que una potencial solución sea inhibir la posibilidad de repetir el curso mediante una política de “todos

pasan”, sino que existe la necesidad de configurar o adoptar nuevos mecanismos que auxilien el desarrollo de competencias en todos los estudiantes.

El *género del estudiante* solo tuvo impacto en dos modelos, el de competencia matemática y el de competencia científica. Al igual que el *ISEC del estudiante*, era de esperarse que este factor tuviese un efecto significativo, pues en la RSL fue la segunda variable más referida entre los diferentes estudios del rendimiento en PISA (Bove et al., 2016; Chen y Cui, 2019; Cordero et al., 2015; Gamazo et al., 2017; Lau y Ho, 2015; Lau y Lam, 2017; Ma y Crocker, 2007; Meggiolaro, 2017; Ning et al., 2015; Shera, 2014; Yıldırım, 2012). Con este hallazgo se da luz sobre las aún existentes brechas de género en el desarrollo de competencias, particularmente, en el caso de México, la matemática y la científica. Como lo indicaron Ma y Crocker (2007), los esfuerzos por la eliminación de las inequidades por género comienzan por estudiar y dar evidencia de este fenómeno.

El último factor de nivel-1 fue el *grado escolar del estudiante*, encontrándose como relevante en los modelos correspondientes a la competencia lectora y matemática. A pesar de que, en teoría, la delimitación de la población objetivo y los criterios de exclusión establecidos en PISA inhibirían posibles sesgos por edad o curso, al igual que Gamazo et al. (2017), Lau y Lam (2017) y Ning et al. (2018), en esta investigación el haber tenido un grado adicional de estudios significó un mayor rendimiento en comparación con aquellos que no contaron con este. Si bien este resultado no debería tener implicaciones en el marco de la política educativa o escolar, este hallazgo apunta a la consideración del efecto proveniente del *grado escolar del estudiante* en futuros estudios sobre factores asociados.

De las variables a nivel escuela, tres resultaron estar asociadas con el nivel de competencia de los estudiantes. Únicamente el *ISEC medio de la escuela* tuvo un efecto significativo en los tres

modelos construidos. Nuevamente, este hallazgo era predecible, de los factores a nivel centro incluidos en la RSL, el *ISEC medio en la escuela* fue el más referido en los análisis secundarios de datos PISA. Los autores coinciden en que no solo el ISEC del estudiante importa, sino que también el ISEC del contexto escolar en el que se aprende tiene un efecto positivo sobre el rendimiento de los estudiantes (Bove et al., 2016; Chen y Cui, 2019; Gamazo et al., 2017; Guo et al., 2018; Lau y Ho, 2015; Lau y Lam, 2017; Meggiolaro, 2017; Ning et al., 2015; Shera, 2014; Yildirim, 2012). Ante tal situación, Shera (2014) concluyó que, aunque no es posible cambiar el ISEC, sí es posible controlar los factores que influyen en el impacto del ISEC sobre el nivel de competencia de los estudiantes.

El siguiente factor de mayor peso fue *la tasa de repetidores en la escuela*. Además del impacto que significa negativo que tiene el ser repetidor, en este estudio quedó asentado que el rodearse de repetidores se asocia con niveles más bajos de competencia lectora y científica. Este resultado fue contradictorio con Gamazo et al. (2017), quien encontró con la muestra española que *la tasa de repetidores* se asoció positivamente con el rendimiento de los estudiantes. En el caso de México, la asociación negativa (de este factor) podría deberse a que en las escuelas de baja eficacia suelen concentrarse altas tasas de repetidores.

La tercera y última variable significativa a nivel centro fue *la tasa de migrantes en la escuela*. Dado que en la RSL no hubo estudio alguno que encontrase efectos significativos por parte de esta variable, sorprendió que este factor tuviese un impacto negativo sobre el rendimiento matemático de la muestra mexicana. Este hallazgo coincide con el tipo de efecto que tuvo el *estatus migratorio del estudiante*, disminuyendo, también, el nivel de competencia.

La mayoría de los factores introducidos a nivel escuela no resultaron ser estadísticamente significativos. En múltiples naciones, la variable *porcentaje de estudiantes por género* ha tenido

un peso explicativo importante, por ejemplo: en China y España se ha encontrado que, a mayor porcentaje de mujeres en la escuela se obtienen, también, mayores niveles de competencia lectora y científica (Gamazo et al., 2017; Lau y Ho, 2015; Lau y Lam, 2017); mientras que, en Italia, una mayor tasa de estudiantes varones se asoció con puntajes más altos en competencia matemática (Meggiolaro, 2017). En el caso de México, este factor no tuvo relevancia estadística en ninguno de los tres modelos; ello puede deberse a que en este país el porcentaje de hombres y mujeres es relativamente semejante y, por lo tanto, no hay variabilidad estadística (Instituto Nacional de Estadística y Geografía [INEGI], 2015).

El *tipo de escuela* tampoco resultó tener un efecto significativo en ninguna de las tres competencias. Aunque suele creerse que el pertenecer a una escuela pública o privada influye en el desarrollo de competencias y, por ende, en los resultados de la prueba PISA. En la muestra mexicana de la edición 2018 se corroboró que este factor no hizo diferencia. Independientemente del tipo de sostenimiento de la escuela, los estudiantes tuvieron la misma probabilidad de obtener bajos y altos puntajes. En otros países la situación fue distinta, Guo et al. (2018) encontraron que en Japón y Estados Unidos el pertenecer a una escuela pública impacta positivamente sobre la comprensión lectora de los estudiantes.

Otra variable que no resultó significativa fue el *tamaño de la escuela*. En diferentes estudios, esta característica ha tenido relevancia estadística. Por ejemplo, en la muestra canadiense e hindú, el tamaño de la matrícula escolar tuvo un efecto positivo en el nivel de competencia lectora y científica de los estudiantes (Areepattamannil y Kaur, 2012; Areepattamannil, 2014); mientras que en el caso de la muestra española el efecto fue inverso, reduciendo la competencia científica del alumnado (Gamazo et al., 2017). De acuerdo con Ares (2014), aunque en la literatura se encuentran posturas divididas sobre el tema, una matrícula reducida puede afectar sobre la

capacidad de la escuela para implementar clases a un grupo de estudiantes de un solo grado, es decir, de no contar con los suficientes estudiantes se tendría que optar por aulas multigrado, fragmentando así la calidad y la amplitud del currículo. No obstante, en el caso a México este factor careció de relevancia; ello pudo corresponder a que en este país superior no existen aulas multigrado a la educación secundaria o media superior, pues estos niveles educativos se caracterizan por una extensa población estudiantil.

Al igual que el *tamaño de la escuela*, el *tamaño de las aulas* no resultó significativa. En la RSL, únicamente en el estudio de Cordero et al. (2015) se encontraron relaciones estadísticamente relevantes por parte de esta variable. De acuerdo con los hallazgos de estos autores, una clase reducida tuvo un efecto positivo sobre el rendimiento en competencia matemática. Según la OCDE (2012), el tamaño de las clases se considera una de las variables clave a la hora de formular políticas educativas; sin embargo, en el caso de México este factor no fue suficiente por sí mismo para influir en los resultados estudiantiles.

La *escasez de docentes* tampoco hizo diferencia. Areepattamannil y Kaur (2012) evidenciaron que los estudiantes que asisten a escuelas con escasez de docentes obtienen puntajes menores en comparación con sus homólogos sin escasez de docentes. A pesar de que este tema se ha convertido en un desafío en las últimas dos décadas, particularmente en aquellos países en los que se espera que la matrícula aumente (ver Cobbold, 2015), en este trabajo no hubo evidencia de que la variable afectase el rendimiento de los estudiantes. En este caso, no existió claridad sobre cómo exactamente se calculó esta variable y, por lo tanto, no fue posible emitir una razón del por qué en México no tuvo un efecto.

Las últimas dos variables que no resultaron significativas fueron la *escasez de recursos* y la *ratio de estudiantes por docente*. Este resultado era de esperarse, dado que en ninguno de los

estudios analizados en la RSL estos factores tuvieron un peso estadístico sustancial. Aunque la OCDE (2010) ha afirmado, en diversas ocasiones, que la *escasez de recursos* y las altas tasas de alumnos por docente son desafíos que permanecen en el contexto mexicano, estas variables no tuvieron un impacto en variabilidad de los puntajes PISA.

A partir de los CCI finales se dio evidencia de que las variables contextuales explicaron el 53.25% de la variabilidad en competencia lectora, el 53.06% en competencia matemática y el 54.63% en competencia científica; por lo tanto, aún quedaba por explicar la diferencia, el 46.75%, el 46.94% y el 45.37%, respectivamente. Tal como lo planteó Krüger (2017), con estos hallazgos quedó asentado que, efectivamente, hubo marcadas diferencias entre los resultados debido a inequidades sociales, económicas y culturales. Aunque la segregación educativa no se visualiza como un tema prioritario en el marco de la política educativa, Chaparro y Gamazo (2020), comentaron que este tema compete, también, de políticas sociales que figuren como mecanismos de auxilio para extinguir estas desigualdades.

Aunque no es un tema recurrente de discusión en otros estudios, en este trabajo se encontró que los residuos se alejaron considerablemente de la normal, la homocedasticidad y la linealidad. De acuerdo con Raudenbush et al. (2016), en los análisis multinivel se debe asumir que los residuos se distribuyen con normalidad, una media 0 y varianza constante. Dado que, en esta investigación, no fue un objetivo verificar el cumplimiento de estos supuestos, es aventurado establecer alguna conclusión al respecto, no obstante, se considera preciso revisar las posibles implicaciones no cumplir con estos tres supuestos, usuales en los modelos de regresión lineales.

5.2 Fase 2: Clasificación de la Eficacia Escolar

Al igual que Mosquera et al. (2010), en este trabajo se evidenció que, comprado con otras técnicas, Johnson tiene el mejor ajuste en la transformación de datos no-normales a normales.

Existen diferentes enfoques a la hora de valorar la eficacia de las escuelas. Una de las principales disyuntivas en la literatura sobre Eficacia Educativa es a qué se le considera una escuela de alta y de baja eficacia (Lizasoain, 2020). Sammons et al. (2016), por ejemplo, recopilaron al menos nueve formas distintas formas de definir operacionalmente la eficacia, entre ellas el uso de residuos y puntajes promedios brutos. Sin embargo, aún con el uso de estas medidas de eficacia, es posible tener falsos positivos. Por ello, de las nueve formas mencionadas, varios autores han establecido, adicionalmente, criterios de eficacia, por ejemplo: la ubicación de la escuela por encima del percentil 80; la posición del centro por debajo del primer cuartil y por encima del último; el rango pseudocentil en comparación con los pares; la consistencia del rendimiento entre variables de interés o subgrupos de estudiantes, entre otras (Bartau et al., 2017; Castro-Morena et al., 2017; Gamazo et al., 2017; Martínez-Abad, 2019; Murillo y Hernández-Castilla, 2011; Rutledge et al., 2015).

La clasificación realizada difirió parcialmente con otros estudios sobre la eficacia de las escuelas en Iberoamérica. Por ejemplo, Castro-Morera et al. (2015) identificaron escuelas de alto residuo, rendimiento esperado y bajo residuo; Gamazo et al. (2017) y Pedroza et al. (2018) encontraron centros de alta y baja eficacia; y Bartau et al. (2017) solo institutos de alta eficacia. No obstante, estas discrepancias se deben principalmente al objeto de estudio de los investigadores y a las decisiones tomadas, por parte de los mismos, en la identificación y selección de centros eficaces.

En la caracterización de las escuelas clasificadas como alta eficacia, residuo centrado y baja eficacia, se encontraron semejanzas en cuanto *al porcentaje de mujeres, la tasa de migrantes, la tasa de repetidores, la escasez de recursos, la escasez de personal y la ratio de estudiantes por docente*. Sin embargo, en al menos dos categorías hubo diferencias significativas en el *ISEC medio escolar, el tipo de escuela, el tamaño de la escuela y el tamaño de las aulas*. Se identificó un mayor ISEC en las escuelas de alta eficacia, más del doble de escuelas públicas y aulas más grandes en las de residuo centrado y menos estudiantes matriculados en las de baja eficacia. A pesar de que en la Fase 1 se habían controlado los efectos provenientes de las variables contextuales, al clasificar las escuelas sobresalieron las brechas económicas, sociales y culturales existentes en las escuelas mexicanas.

A su vez, se mostró evidencia de que el nivel de eficacia de las escuelas coincide con el nivel de rendimiento de los estudiantes. En las escuelas de alta eficacia están, también, los estudiantes con los mayores puntajes en competencia lectora, matemática y científica; en las de baja los estudiantes con menores puntajes en los tres dominios de la prueba PISA; y, en las de residuo centrado se encuentran los estudiantes que puntuaron en un término intermedio entre los estudiantes matriculados en las escuelas de alta y baja eficacia.

En la evaluación de la calidad de los conglomerados, partir de tres contrastes marginales, fue posible corroborar que las agrupaciones eran de calidad, pues presentaban un único núcleo denso. Como lo afirmaron Almeida et al. (2011), en los estudios que se aplican técnicas de *clustering* es preciso medir la efectividad del algoritmo de agrupación y la dinámica de las relaciones entre conjuntos de datos, en este caso, entre los conjuntos residuales por competencias.

5.3 Fase 3: Análisis del Peso Explicativo

La tercera fase de análisis consistió en el análisis del peso explicativo de los procesos educativos sobre la eficacia escolar de los centros educativos mexicanos. De acuerdo con la OCDE (2017a), en las últimas tres aplicaciones PISA, los cuestionarios de contexto han sido enriquecidos con el cuerpo de conocimiento sobre Eficacia Educativa para aumentar el poder explicativo de los insumos recolectados. Al realizar la revisión documental de los cuestionarios de contexto, se encontró que, efectivamente, en las ediciones 2012, 2015 y 2018 se ha incluido el número máximo de indicadores relativos a la práctica del docente.

El modelo de regresión logística multinomial final proporcionó seis factores explicativos de la eficacia: cinco de enseñanza y uno escolar. Sobre la enseñanza, la *retroalimentación* dada por el docente fue la variable con mayor peso explicativo de la eficacia, al disminuir la probabilidad de que la escuela sea de alta eficacia y aumentar la probabilidad de que sea de baja. Estos resultados coincidieron con los de Gamazo et al. (2017) y Lau y Lam (2017), quienes encontraron, con la muestra española y china, respectivamente, que mayores niveles de *retroalimentación* se asocian negativamente con altos residuos escolares y mayores puntajes en ciencias. Se esperaría que la *retroalimentación* aumentase la alta eficacia; sin embargo, el hecho de que se haya aumentado la baja eficacia de las escuelas puede explicarse porque, usualmente, en estas se concentran altos porcentajes de estudiantes con bajo logro y, por lo tanto, la retroalimentación es una práctica necesaria y constante. Mientras que, en las escuelas de alta eficacia los estudiantes cuentan con mayor logro y requieren en menor medida la retroalimentación de docente. No obstante, esta hipótesis debe ser sometida a comprobación mediante un estudio mayor de corte cualitativo.

La siguiente variable con mayor influencia fue la *adaptación de la instrucción*, la cual, aumentó la probabilidad de que la escuela sea de alta eficacia y disminuyó la probabilidad de que sea de baja. Este hallazgo fue congruente con Gamazo et al. (2017) y Lau y Lam (2017), quienes identificaron que adecuar la estructura de las lecciones tuvo un efecto positivo en la probabilidad de que la escuela sea de alta eficacia y en la obtención de mejores puntajes en ciencias. Con ello, quedó asentado que (al igual que en España y China) en México el considerar las necesidades y conocimientos de los estudiantes, cambiar la organización de la instrucción a favor de la mayoría, y proporcionar ayuda individualizada, resulta beneficioso para la mejora escolar.

Una tercera variable con peso explicativo fue el *clima disciplinario en las lecciones*, el cual aumentó la probabilidad de que la escuela sea de alta y de baja eficacia. Este hallazgo fue considerado contraintuitivo, dado que en ambas categorías hubo un efecto positivo. En todos los estudios analizados en la SRL, este factor tuvo un efecto únicamente en la alta eficacia (Areepattamannil, 2014; Bove et al., 2016; Cordero, 2015; Gamazo et al., 2017; Guo et al., 2018; Lau y Ho, 2015; Ma y Crocker, 2007; Ning et al., 2015; Shera, 2014). Este hallazgo se podría deber a que, en comparación con las escuelas de residuo centrado, en ambos tipos de escuela hay altos niveles de disciplina. Sin embargo, posiblemente, en las de baja eficacia, una alta disciplina tenga el propósito de mejorar los resultados estudiantiles; mientras que en las de alta, apele a mantener los buenos resultados.

El *uso de las TIC durante la clase*, también, explicó el aumento de la probabilidad de ser una escuela de alta y baja eficacia. En los análisis secundarios de datos PISA, Eickelmann et al. (2016) y Meggiolaro (2017) encontraron que el uso de las TIC tuvo un impacto negativo sobre el rendimiento en competencia matemática de los estudiantes europeos. En México, la bifurcación del efecto en el *uso de las TIC en el aula* puede asociarse a las desigualdades socioeconómicas

existentes entre las escuelas de alta y baja eficacia (evidenciadas en la Fase 2). Por lo que es probable que en las de alta eficacia haya un mayor acceso a la tecnología y una mayor cultura digital; mientras que en las de baja eficacia haya un menor acceso a las TIC y una menor cultura digital.

La última variable de enseñanza fue *el tiempo promedio en clase*, cuyo efecto fue negativo en la probabilidad de ser una escuela de alta y baja eficacia. En otros estudios, el tiempo de instrucción ha tenido un efecto positivo sobre el rendimiento en competencia lectora de los estudiantes canadienses (Ma y Crocker, 2007), y se ha asociado con la baja eficacia de las escuelas españolas (Martínez-Abad et al., 2020). Este hallazgo puede deberse a que, en las escuelas de alta eficacia, el exceso de tiempo de instrucción no conlleva un beneficio sobre el rendimiento de los estudiantes porque, probablemente, al tener puntuaciones elevadas, estos se caracterizan por ser más autónomos y no precisar de un mayor tiempo de instrucción; en cambio, en las escuelas de baja eficacia, los estudiantes pueden requerir más tiempo efectivo en clase para mejorar sus resultados.

La única variable escolar con un efecto significativo fue *el clima discriminatorio escolar*, que tuvo un impacto negativo en la probabilidad de ser una escuela de alta eficacia y un efecto positivo en la probabilidad de ser de baja eficacia. Tras la revisión documental de los cuestionarios de contexto, se identificó que esta es la primera ocasión en la que se evalúa *el clima discriminatorio escolar*; por lo tanto, no hay estudios previos a la edición 2018 que hayan analizado esta variable y que permitan la contrastación de los resultados de la presente investigación. No obstante, queda asentado que los niveles de discriminación a nivel centro afectaron estadísticamente la eficacia de las escuelas mexicanas.

Cinco de las diez variables de enseñanza introducidas en los modelos de regresión no tuvieron relevancia estadística al explicar la eficacia de las escuelas. Al igual que en otros estudios, la *estimulación*, los *comportamientos* y el *interés docente* no fueron estadísticamente relevantes. Sin embargo, aunque en otros países (principalmente de Europa y Asia) la *instrucción directa* y el *apoyo docente* han tenido un peso explicativo, ese no fue el caso de México.

La *instrucción directa* tuvo un impacto significativo sobre los resultados PISA de 11 regiones en el mundo. Mientras que los puntajes en ciencias de los estudiantes de Estonia, Canadá, Corea, Finlandia, Japón, Singapur y Taiwán se vieron aumentados (Bove et al., 2016); los puntajes en matemáticas de los estudiantes italianos se vieron disminuidos (Lau y Lam, 2017). Posiblemente, la diferencia, entre esos hallazgos y los de esta investigación, se deba a que en el estudio de la eficacia de las escuelas mexicanas se abarcó las tres competencias evaluadas en PISA; es decir que, en este trabajo, el efecto de la *instrucción directa* se vio diluido al explicar (en conjunto) más de una competencia: la lectora, la matemática y la científica.

En lo que refiere al *apoyo docente*, Gamazo et al. (2017) encontraron que este factor disminuyó la probabilidad de que una escuela sea de alta eficacia. Aunque en ambos estudios la variable de interés fue la eficacia de las escuelas, las decisiones metodológicas correspondientes a la identificación de los centros y los análisis de regresión fueron distintas. En el caso de Gamazo et al. (2017) se utilizaron percentiles y regresión logística binaria, con la categoría baja eficacia como referencia para explicar la alta eficacia; mientras que en este estudio se aplicó k-medias y se usó regresión logística multinomial, con el residuo centrado como categoría de referencia para explicar la alta y la baja eficacia. Por consiguiente, es probable que haya diferencias significativas por parte de este factor al contrastar las escuelas de baja eficacia con las de alta, más no al comparar ambos tipos de escuela con las de residuo centrado.

En lo que refiere a las variables escolares, cinco de las seis variables, incluidas en los modelos de regresión, no resultaron significativas: las *actividades extracurriculares*, la *calidad escolar*, la *oferta de clases adicionales*, las *políticas escolares sobre el involucramiento de la familia* y el *uso de las TIC en la escuela*. Este resultado fue congruente con la teoría dinámica, dado que, como lo indicaron Creemers y Kyriakides (2008), el cumulo de conocimiento sobre Eficacia Educativa ha demostrado recurrentemente que, a nivel centro, únicamente los factores relativos a las políticas escolares sobre la enseñanza y el clima escolar tienen un efecto, indirecto, sobre el rendimiento estudiantil.

Finalmente, se evaluó el modelo final de la regresión logística multinomial mediante una matriz de confusión. La bondad de ajuste del modelo fue justa, de acuerdo Landis y Koch (1977). Aunque hubo un 54% de probabilidad de acertar al explicar la eficacia de las escuelas, la precisión del modelo pudo considerarse como buena, dado que la variable de interés tuvo más de dos categorías. Por regla, el porcentaje máximo de acuerdo que se puede alcanzar se ve reducido a mayor número de categorías, el modelo base, con el que se contrastan los datos observados por el investigador, asume que todas las escuelas corresponden a la categoría con mayor número de casos etiquetados, en este caso, el modelo base asumió que todas las escuelas eran de residuo centrado.

6 Conclusiones, Limitaciones y Prospectiva para Futuros Estudios

A partir de los resultados discutidos, se concluye que los principales aportes de esta tesis son:

- La desigualdad educativa que viven los estudiantes mexicanos. La inequidad por género, estatus migratorio e índice social, económico y cultura, es tal, que aún con el control estadístico, estas brechas reaparecieron tras clasificar las escuelas. Esto hace evidente que el logro educativo no solo concierne al docente, la escuela y/o al sistema, sino que precisan, a su vez, de mecanismos sociales para dar cumplimiento al tercer derecho constitucional de este país: la educación.
- La pertinencia de considerar el contexto al juzgar los resultados PISA. En la edición 2018, en la que hubo 82 países participantes, México se ubicó en la posición 53 en lectura, 61 en matemáticas y 57 en ciencias, lo cual, desde una perspectiva superficial, podría considerarse desalentador. Si bien el rendimiento de las escuelas mexicanas es bajo en comparación con la media de la OCDE, si se ahonda en el tema y se toma en cuenta el contexto, es posible encontrar que la mayoría de las escuelas se sobreponen a sus circunstancias contextuales y tienen un rendimiento mayor al que se esperaría, es decir que, en realidad, gran parte de estas tienen un aporte positivo al rendimiento de los estudiantes. Por consiguiente, la atención no debería estar en la posición de México en los rankings internacionales, sino en las condiciones de fondo (p. ej. el ISEC) que envuelven a sus estudiantes y en las prácticas que tienen lugar en los centros eficaces, aquellos que rinden más allá de lo esperado.
- El nivel de eficacia de las escuelas mexicanas corresponde al nivel de competencia lectora, matemática y científica de los estudiantes. Al realizar estadísticos fundamentales de las agrupaciones fue posible identificar que en las escuelas de alta eficacia se encuentran

también los estudiantes de alto rendimiento en las tres competencias evaluadas en PISA; en las de baja eficacia se ubican los que obtienen menores puntajes; y en las de residuo centrado están los estudiantes con rendimiento medio en los tres dominios, es decir, se encuentran en un punto intermedio entre los estudiantes de baja y alta eficacia.

- La importancia de los factores de enseñanza sobre la eficacia de las escuelas mexicanas. Tras los análisis, quedó en evidencia el significativo efecto de la práctica docente para el alcance de los objetivos y la mejora del rendimiento de las escuelas, pese a las condiciones contextuales que les envuelve a estudiantes y centros. Particularmente, se identificaron cinco variables con relevancia estadística sobre la probabilidad de ser una escuela de baja o alta eficacia:
 - la adaptación que los docentes hacen a su instrucción para responder a las necesidades y conocimientos de la mayoría;
 - la frecuencia en la que dan retroalimentación a sus estudiantes a fin de informar sus puntos fuertes, áreas de oportunidad y estrategias la mejorar su desempeño;
 - el tiempo promedio que el docente dedica a que sus estudiantes aprendan, construyan, comuniquen y apliquen el conocimiento;
 - el clima disciplinario mínimo que el docente mantiene para evitar distractores o pérdida de tiempo lectivo; y,
 - el uso de las TIC como herramienta en el aula para aumentar la accesibilidad a la información de los estudiantes.
- La discriminación presente en los centros educativos mexicanos. Aunque en este trabajo la atención se centró en los factores de enseñanza, se optó, también, por analizar el valor explicativo de un conjunto de factores escolares. La única variable con un efecto

considerable sobre la eficacia de las escuelas fue el clima discriminatorio escolar. Es decir que, en estas escuelas existe un ambiente en el que hay un trato desigual hacia los miembros de la comunidad escolar, particularmente los estudiantes.

- La naturaleza dinámica de la eficacia escolar. Lo que funciona en un centro de alta eficacia, no necesariamente funciona en uno de baja. Si bien hay factores explicativos globales, que funcionan independientemente del contexto, algunos factores se ven afectados por variables moduladoras que provocan que un factor explicativo se comporte de forma distinta en una u otra escuela.
- La pertinencia del uso de regresiones logísticas multinomiales (RLM) al analizar tres categorías de eficacia. Con RLM fue posible encontrar relaciones lineales y no lineales, sin embargo, es probable que de haber optado por modelos de regresión ordinales únicamente se hubiesen encontrado relaciones lineales. Por consiguiente, es aconsejable elegir el algoritmo más comprensivo de la naturaleza cambiante y compleja de los fenómenos educativos.

Las limitaciones y sugerencias que se surgieron durante el desarrollo del estudio fueron las siguientes:

- El tamaño de la muestra mexicana en PISA 2018. A diferencia de las anteriores, en la última edición PISA el conjunto muestral de México fue considerablemente reducido, lo cual, tuvo un impacto negativo en el margen de maniobra de los datos. Dado que en este estudio el descartar aquellos centros con menos de 20 estudiantes mermaba considerablemente la representatividad poblacional, se optó finalmente por considerar un mínimo de cuatro estudiantes por escuela. Tal situación podría traducirse en que se trabajó con centros “pequeños”, lo cual, según Joaristi et al. (2014) tiene un impacto en la

estabilidad (o volatilidad) de los residuos. Por lo tanto, previo a descartar centros, se recomienda investigar el mínimo recomendado de sujetos con el que se puede trabajar, sin que se pierda gran parte de la representatividad poblacional y sin que se volatilizar las puntuaciones residuales.

- La falta de evidencia empírica directa. Los análisis secundarios de datos PISA se basan las opiniones recogidas a través de cuestionarios denominados “de contexto”, es decir, que la información no fue medida de forma directa. Ello podría implicar que los insumos proporcionados por PISA no integren las variables necesarias para el análisis. Por consiguiente, es recomendable complementar los análisis cuantitativos con estudios de corte cualitativo para tener mayor precisión de la realidad educativa que viven los estudiantes, docentes y directivos.
- La calidad de los clústers. Uno de los principales problemas al utilizar k-medias es que no se conoce a priori el número adecuado de conglomerados y, por lo tanto, es necesario realizar diversas pruebas y aplicar diferentes criterios para encontrarlo. En esta búsqueda las decisiones tomadas por el investigador tienen un impacto directo sobre las posibles clasificaciones y el contenido de éstas (Afifi et al., 2020). En este caso, en la calidad de las agrupaciones, la homogeneidad fue relativamente buena, pero se pudieron observar algunos dobles núcleos densos. Por lo tanto, además de los criterios propuestos en este estudio, en futuras investigaciones es recomendable explorar vías para resolver el problema de la clasificación, por ejemplo, complementar con métodos de clúster jerárquicos.
- El poco ajuste del modelo de regresión logística multinomial. Al momento de efectuar los análisis de regresión, los datos habían sido procesados en diferentes etapas del método que restan porcentaje de varianza, por ejemplo, los modelos multinivel y la clasificación de la

eficacia con k-medias. Aunado a ello, la variable dependiente contaba con más de dos categorías, lo cual, disminuyó directamente el grado de acuerdo entre el modelo base y el modelo del investigador, tras la integración de variables explicativas. Por lo tanto, si se pretende aumentar el ajuste el modelo es recomendable transformar la naturaleza de variable a dummy.

- La ausencia de datos a nivel aula. Si bien PISA recopila información sobre el funcionamiento educativo a diferentes niveles de anidamiento, nacionales, regionales, escolares e, incluso, estudiantiles, aún se no recoge información a nivel aula. Esta evaluación ha optado, en cambio, por investigar la práctica docente a través de los estudiantes; sin embargo, por definición, esto vía no significa tener datos a nivel aula. Por lo que, con datos PISA, las inferencias sobre la enseñanza eficaz de los docentes se deben emitir e interpretar con cautela.
- La utilización de variables agregadas. De acuerdo con Bove et al. (2016), una situación problemática en los modelos multinivel es la utilización de variables agregadas (como el índice socioeconómico promedio) para la explicación de variables de criterio de niveles superiores (en este caso, a nivel escuela). Como lo indicaron estos autores, las variables construidas a partir de promedios pueden contener sesgos en la medida de que no se consideren las mediciones individuales de todos los estudiantes. Una posible solución a este dilema, es el excluir las escuelas en las que se muestrearon pocos estudiantes o ajustar el modelo multinivel como lo sugieren Grilli y Rampichini (2011) y Lüdtke et al. (2008).
- La intervención de variables moduladoras. Como se mencionó a lo largo de este trabajo, al identificar los factores explicativos de la eficacia se identificaron algunos factores que tenían un efecto inverso o contraintuitivo, por ejemplo, que el clima disciplinario

aumentara la alta eficacia y que, a su vez, aumentase la baja eficacia. Ello, probablemente, se deba a variables moduladoras inherentes al estudiante, como su género e índice socioeconómico; aunque algunas de ellas fueron controladas mediante modelos multinivel, es posible que haya características de los estudiantes que no hayan sido consideradas dentro de las tradicionalmente analizadas en los modelos multinivel. Por consiguiente, es conveniente poner a prueba otras características que vayan más allá de los antecedentes familiares.

Referencias

- Afifi, A., May, S., Donatello, R. A., & Clark, V. A. (2020). Cluster analysis. In author (eds.), *Practical Multivariate Analysis* (6th ed., pp. 317-338). ISBN: 978-1-138-70222-6
- Aitkin, M., & Longford, N. (1986). Statistical modelling issues in School Effectiveness studies. *Journal of the Royal Statistical Society*, 149(1), 1-43. <https://doi.org/10.2307/2981882>
- Aitkin, M., Anderson, D., & Hinde, J. (1981). Statistical modelling of data on teaching styles. *Journal of the Royal Statistical Society*, 144(4), 419-461. <https://doi.org/10.2307/2981826>
- Alcaraz, N., Caparrós, R., Soto, E., Beltrán, R., Rodríguez, A., y Sánchez, S. (2013). ¿Evalúa PISA la competencia lectora? *Revista de Educación*, 360, 577–599. <https://doi.org/10.4438/1988-592X-RE-2011-360-130>
- Almeida, H., Guedes, D., Meira, W., & Zaki, M. Z. (2011). Is there a best quality metric for graph clusters? *Machine Learning and Knowledge Discovery in Databases*, (pp. 44-59). <https://doi.org/10.1007/978-3-642-23780-5>
- Amador, M., y López-Gonzalez, E. (2007). Una aproximación bibliométrica a los modelos multinivel. *Revista Electrónica de Investigación y Evaluación Educativa*, 13(1), 67-82. https://www.uv.es/RELIEVE/v13n1/RELIEVEv13n1_3.pdf
- Anderson, J. O., Milford, T., & Ross, S. (2009). Multilevel modeling with HLM: Taking a second look at PISA. In M. C. Shelley, L. D. Yore & B. Hand (Eds.), *Quality Research in Literacy and Science Education* (pp. 263–289). ISBN 978-1-4020-8426-3.
- Anderson, L. M., Evertson, C. M., & Brophy, J. E. (1979). An experimental study of effective teaching in first-grade reading groups. *The Elementary School Journal*, 79(4), 193–223. <https://doi.org/10.1086/461151>

- Antoniou, P., & Kyriakides, L. (2011). The impact of a dynamic approach to professional development on teacher instruction and student learning: Results from an experimental study. *School Effectiveness and School Improvement*, 22(3), 291–311. <https://doi.org/10.1080/09243453.2011.577078>
- Antoniou, P., & Kyriakides, L. (2013). A dynamic integrated approach to teacher professional development: Impact and sustainability of the effects on improving teacher behavior and student outcomes. *Teaching and Teacher Education*, 29(1), 1–12. <https://doi.org/10.1016/j.tate.2012.08.001>
- Areepattamannil, S. (2014). International note: what factors are associated with reading, mathematics, and science literacy of indian adolescents? A multilevel examination. *Journal of Adolescence*, 37(4), 367–372. <http://dx.doi.org/10.1016/j.adolescence.2014.02.007>
- Areepattamannil, S., & Kaur, B. (2012). Factors predicting science achievement of immigrant and non-immigrant students: a multilevel analysis. *International Journal of Science and Mathematics Education*, 11(5), 1183–1207. <https://doi.org/10.1007/s10763-012-9369-5>
- Ares, M. (2014). School size policies: a literature review. *OECD Education Working Papers*, No. 106. <https://doi.org/10.1787/5jxt472ddkjl-enb>
- Arffman, I. (2010). Equivalence of Translations in International Reading Literacy Studies. *Scandinavian Journal of Educational Research*, 54(1), 37–59. <https://doi.org/10.1080/00313830903488460>
- Ary, D., Cheser-Jacobs, L., Sorensen-Irvine, C. K. (2010). *Introduction to Research in Education* (8th ed.). Cengage.
- Ary, D., Cheser-Jacobs, L., Sorensen-Irvine, C. K., & Walker, D. (2019). *Introduction to Research in Education* (10th ed.). Cengage.

- Azigwe, J. B., Kyriakides, L. Panayiotou, A., & Creemers, B. P. M. (2016). The impact of effective teaching characteristics in promoting student achievement in Ghana. *International Journal of Educational Development, 51*, 51–61. <http://dx.doi.org/10.1016/j.ijedudev.2016.07.004>
- Azkiyah, S. N., Doolaard, S., Creemers, B. P. M., & Van Der Werf, M. P. C. (2014). The effects of two intervention programs on teaching quality and student achievement. *Journal of Classroom Interaction, 49*(1), 4–11. <https://www.jstor.org/stable/44735692>
- Barnard-Brak, L., Lan, W. Y., & Yang, Z. (2018). Differences in mathematics achievement according to opportunity to learn: A 4pL item response theory examination. *Studies in Educational Evaluation, 56*, 1–7. <http://dx.doi.org/10.1016/j.stueduc.2017.11.002>
- Bartau, I., Azpillaga, V. y Joaristi, L. M. (2017). Metodología de enseñanza en centros eficaces de la Comunidad Autónoma del País Vasco. *Revista de Investigación Educativa, 35*(1), 93–112. <https://doi.org/10.6018/rie.35.1.225141>
- Beltrán, O. A. (2005). Revisiones sistemáticas de la literatura. *Revista Colombiana de Gastroenterología, 20*(1), 60-69. <http://www.scielo.org.co/pdf/rcg/v20n1/v20n1a09.pdf>
- Ben HajKacem, M. A., Ben N’Cir, C.-E., y Essoussi, N. (2019). Overview of scalable partitional methods for big data clustering. In O. Nasraoui, C.-E. Ben N’Cir, *Clustering Methods for Big Data Analytics: techniques, toolboxes and applications* (pp. 1-24). <https://doi.org/10.1007/978-3-319-97864-2>
- Boedeker, P. (2017). Hierarchical linear modeling with maximum likelihood, restricted maximum likelihood, and fully Bayesian estimation. *Practical Assessment, Research & Evaluation, 22*(2), 1-19. <https://pareonline.net/getvn.asp?v=22&n=2>
- Bonnet, G. (2006). Tener presentes las singularidades lingüísticas y culturales en las evaluaciones internacionales de las competencias de los alumnos: ¿una nueva dimensión para PISA?

Revista de Educación, extraordinario 2006, 91-109.

<http://www.educacionyfp.gob.es/dam/jcr:5799dd27-8859-4a95-ba5f-10880525dddb/articulosre2006-2-.pdf>

Bosker, R. J., & Scheerens, J. (1994). Alternative models of school effectiveness put to the test.

International Journal of Educational Research, 21(2), 159–80.

[https://doi.org/10.1016/0883-0355\(94\)90030-2](https://doi.org/10.1016/0883-0355(94)90030-2)

Bove, G., Marella, D., & Vitale, V. (2016). Influences of school climate and teacher's behavior on

student's competencies in mathematics and the territorial gap between Italian macro-areas in PISA 2012. *Educational, Cultural and Psychological Studies*, (13), 63-95.

<https://doi.org/10.7358/ecps-2016-013-bove>

Bove, G., Marella, D., & Vitale, V. (2016). Influences of school climate and teacher's behavior on

student's competencies' in mathematics and the territorial gap between Italian macro-areas in PISA 2012. *Journal of Educational, Cultural and Psychological Studies*, (13), 63-96.

<https://doi.org/10.7358/ecps-2016-013-bove>

Brophy, J. E., & Evertson, C. M. (1974). *Process-Product Correlations in the Texas Teacher*

Effectiveness Study: Final Report. <https://files.eric.ed.gov/fulltext/ED091394.pdf>

Brophy, J. E., & Good, T. L. (1984). *Teacher behavior and student achievement*. Occasional paper

No. 73. <https://files.eric.ed.gov/fulltext/ED251422.pdf>

Cairns, D., & Areepattamannil, S. (2017). Exploring the relations of inquiry-based teaching to

science achievement and dispositions in 54 countries. *Research in Science Education*.

<https://doi.org/10.1007/s11165-017-9639-x>

Calderhead, J. (1996). Teachers: beliefs and knowledge. In D. C. Berliner & R. C. Calfee (Eds.),

Handbook of educational psychology (p. 709–725). Prentice Hall International.

- Campbell, R. J., Kyriakides, L., Muijs, R. D., & Robinson, W. (2003). Differential Teacher Effectiveness: towards a model for research and teacher appraisal. *Oxford Review of Education*, 29(3), 347-362. <https://doi.org/10.1080/0305498032000120292>
- Campbell, R. J., Kyriakides, L., Muijs, R. D., & Robinson, W. (2004). *Assessing teacher effectiveness: Developing a differentiated model*. Routledge Falmer.
- Caro, D. H., & Kyriakides, L. (2019). Assessment design and quality of inferences in PISA: limitations and recommendations for improvement. *Assessment in Education: Principles, Policy & Practice*, 26(4), 363-368. <https://doi.org/10.1080/0969594X.2019.1645990>
- Caro, D. H., Lenkeit, J., & Kyriakides, L. (2016). Teaching strategies and differential effectiveness across learning contexts: Evidence from PISA 2012. *Studies in Educational Evaluation*, 49, 30–41. <https://doi.org/10.1016/j.stueduc.2016.03.005>
- Castro-Morera, M., García-Medina, A. M., Pedroza-Zúñiga, L. H., y Caso-Niebla, J. (2015). Escuelas de alto y bajo valor añadido. Perfiles diferenciales de las secundarias en baja california. *Archivos Analíticos de Políticas Educativas*, 23(27). <http://dx.doi.org/10.14507/epaa.v23.1917>
- Chaparro, A. A., & Gamazo, A. (2020). Estudio multinivel sobre las variables explicativas de los resultados de México en PISA 2015. *Archivos Analíticos de Políticas Educativas*, 28(26). <https://doi.org/10.14507/epaa.28.4620>
- Chapman, C., Reynolds, D., Muijs, D., Sammons, P., Stringfield, S. & Teddlie, C. (2016). Educational effectiveness and improvement research and practice: The emergence of the discipline. In C. Chapman, D. Muijs, D. Reynolds, P. Sammons, & C. Teddlie, C. (Eds.), *The Routledge International Handbook of Educational Effectiveness and Improvement. Research, policy, and practice* (pp. 1-24). Routledge.

- Chatterjee, S., & Simonoff, J. S. (2013). *Handbook of Regression Analysis*. Wiley.
- Chen, F., & Cui, Y. (2019). Investigating the relation of perceived teacher unfairness to science achievement by hierarchical linear modeling in 52 countries and economies. *Educational Psychology*, 1–23. <https://doi.org/10.1080/01443410.2019.1652248>
- Christoforidou, M., & Xirafidou, E. (2014). Using the dynamic model to identify stages of teacher skills in assessment. *Journal of Classroom Interaction*, 49(1), 12–25. <https://www.jstor.org/stable/44735693>
- Christoforidou, M., Kyriakides, L., Antoniou, P., & Creemers, B. P. M. (2014). Searching for stages of teacher skills in assessment. *Studies in Educational Evaluation*, 40, 1–11. <https://doi.org/10.1016/j.stueduc.2013.11.006>
- Cobbold, C. (2015). Solving the teacher shortage problem in Ghana: critical perspectives for understanding the issues. *Journal of Education and Practice*, 6(9), 1-79. <https://files.eric.ed.gov/fulltext/EJ1082464.pdf>
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Science* (2nd ed.). LEA.
- Coleman, J. S., Campbell, E., Hobson, C., McPartland, J., Mood, A., Weinfeld, F., & York, R. (1966). *Equality of educational opportunity*. Government Printing Office.
- Collen, A. (2003). *Systemic change through praxis and inquiry*. *Praxiology: The International Annual of Practical Philosophy and Methodology* (Volume 11). Transaction Publishers.
- Cordero, J. M., Crespo, E., y Pedraja, F. (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>
- Cordero, J. M., Pedraja, F., y Simancas, R. (2015). Factores del éxito escolar en condiciones socioeconómicas desfavorables. *Revista de Educación*, (370), 172-198.

<https://www.educacionyfp.gob.es/dctm/revista-de-educacion/articulos370/espanol-ingles/06cordero.pdf?documentId=0901e72b81ecb10e>

Creemers, B. P. M. (1993, enero). *Towards a theory on educational effectiveness* [Artículo]. International Congress for School Effectiveness and Improvement, Norrköping, Suecia.
<https://eric.ed.gov/?id=ED361828>

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.

Creemers, B. P. M., & Kyriakides, L. (2015). Developing, testing and using theoretical models of educational effectiveness for promoting quality in education. *School Effectiveness and School Improvement*, 26(1), 102–119. <https://doi.org/10.1080/09243453.2013.869233>

Creemers, B. P. M., Kyriakides, L., & Sammons, P. (2010). *Methodological Advances in Educational Effectiveness Research*. Routledge.

Darlington, R. B. & Hayes, A. F. (2017). *Regression Analysis and Linear Models: Concepts, Applications and Implementation*. ISBN 978-1-4625-2113-5.

de Jong, R., Westerhof, K.J., and Kruiter, J.H. (2004). Empirical evidence of a comprehensive model of school effectiveness: a multilevel study in mathematics in the 1st year of junior general education in the Netherlands. *School Effectiveness and School Improvement*, 15(1), 3–31. <https://doi.org/10.1076/sesi.15.1.3.27490>

De Maeyer, S., van den Bergh, H., Rymenans, R., Van Petegem, P., & Rijlaarsdam, G. (2010). Effectiveness criteria in school effectiveness studies: Further research on the choice for a multivariate model. *Educational Research Review*, 5(1), 81–96.
<https://doi.org/10.1016/j.edurev.2009.09.001>

Edmonds, R. R. (1978). A discussion of the literature and issues related to effective schooling.

CEMREL

Edmonds, R. R. (1979). Effective schools for the urban poor. *Educational Leadership*, 37, 15–27.

Eickelmann, B., Gerick, J., & Koop, C. (2016). ICT use in mathematics lessons and the mathematics achievement of secondary school students by international comparison: Which role do school level factors play? *Education and Information Technologies*, 22(4), 1527–1551. <https://doi.org/10.1007/s10639-016-9498-5>

Elías, M. E. (2015). La cultura escolar: Aproximación a un concepto complejo. *Educare Electronic Journal*, 19(2), 285-301. <http://dx.doi.org/10.15359/ree.19-2.16>

Fernández, T. (2003). Métodos estadísticos de estimación de loa efectos de la escuela y su aplicación al estudio de las escuelas eficaces. *Revista Electrónica Iberoamericana sobre Calidad, Eficacia y Cambio en Educación*, 1(2), 1-28. https://repositorio.uam.es/bitstream/handle/10486/660697/REICE_1_2_9.pdf?sequence=1&isAllowed=y

Gallagher, J. J. (1991). Prospective and practicing secondary school science teachers' knowledge and beliefs about the philosophy of science. *Science Education*, 75(1), 121–133. <https://doi.org/10.1002/sce.3730750111>

Gamazo, A. (2019). *Factores asociados al rendimiento y a la eficacia escolar: un estudio basado en métodos mixtos a partir de PISA 2015* [Tesis doctoral, Universidad de Salamanca]. https://gredos.usal.es/bitstream/handle/10366/140406/DDOMI_GamazoGarc%C3%ADa_A_RendimientoyEficaciaEscolarPISA2015.pdf?sequence=1&isAllowed=y

Gamazo, A., Martínez-Abad, F., Olmos-Migueláñez, S. y Rodríguez-Conde, M. J. (2018). Evaluación de los factores relacionados con la eficacia escolar en PISA 2015. Un análisis

- multinivel. *Revista Educación*, 379, 56-84. <https://doi.org/10.4438/1988-592X-RE-2017-379-369>
- Garson, G. D. (2014). *Logistic Regression: Binary & Multinomial*. Statistical Associates Publishing.
- Goldstein, H. (1987). Multilevel models in educational and social research. Griffin
- Good, T. L., Wiley, C. R. H., & Florez, I. R. (2009). Effective teaching: an emerging synthesis. In L. J. Saha & A. G. Dworkin (Eds.), *International Handbook of Research on Teachers and Teaching* (pp. 803-816). Springer.
- Gordon, R. A. (2015). *Regression Analysis for the Social Science* (2nd ed.). Routledge.
- Görke, R., Kappes, A., & Wagner, D. (2015). Experiments on density-constrained graph clustering. *ACM Journal of Experimental Algorithmics*, 19(1), 6:1-6:31. <http://doi.org/10.1145/2638551>
- Grabau, L. J., & Ma, X. (2017). Science engagement and science achievement in the context of science instruction: a multilevel analysis of U.S. students and schools. *International Journal of Science Education*, 39(8), 1045–1068. <https://doi.org/10.1080/09500693.2017.1313468>
- Grilli, L., & Rampichini, C. (2011). The role of sample cluster means in multilevel models: A view on endogeneity and measurement error issues. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, 7(4), 121–133. <https://doi.org/10.1027/1614-2241/a000030>
- Gully, S. M., & Phillips, J. M. (2019). On finding your level. In S. E. Humphrey & J. M. LeBreton (Eds.), *The Handbook of Multilevel Theory, Measurement, and Analysis* (pp. 11-38). <http://dx.doi.org/10.1037/0000115-000>

- 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), 1–15. <https://doi.org/10.1007/s12564-018-9516-y>
- Harlen, W. (2002). Evaluar la alfabetización científica en el programa de la OECD para la evaluación internacional de estudiantes (PISA). *Enseñanza de las Ciencias*, *20*(2), 209–216. <https://www.raco.cat/index.php/Ensenanza/article/view/21800/21634>
- Hatcher, H. M. (1941). An experimental study to determine the relative effectiveness at the secondary level of two methods of instruction. *The Journal of Experimental Education*, *10*(1), 41–47. <https://doi.org/10.1080/00220973.1941.11010232>
- Hativa, N. (2000). Teacher thinking, beliefs, and knowledge in higher education: an introduction. *Instructional Science*, *28*(5/6), 331–334. <https://www.jstor.org/stable/23371453>
- 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., Barrera-Pedemonte, F., & Buchholz, J. (2018). Cross-cultural comparability of noncognitive constructs in TIMSS and PISA. *Assessment in Education: Principles, Policy & Practice*, 1-17. <https://doi.org/10.1080/0969594X.2018.1469467>
- Hill, H. C., Rowan, B., & Ball, D. L. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American Educational Research Journal*, *42*(2), 371–406. <https://doi.org/10.3102/00028312042002371>
- Hill, J. B. (2016). Questioning techniques: a study of instructional practice. *Peabody Journal of Education*, *91*(5), 660–671. <https://doi.org/10.1080/0161956X.2016.1227190>

- Hill, P. W., & Rowe, K. J. (1996). Multilevel modelling in School Effectiveness research. *School Effectiveness and School Improvement*, 7(1), 1–34. <https://doi.org/10.1080/0924345960070101>
- Hill, P. W., & Rowe, K. J. (1998). Modelling student progress in studies of Educational Effectiveness. *School Effectiveness and School Improvement*, 9(3), 310–333. <https://doi.org/10.1080/0924345980090303>
- Hopkins, D., & Reynolds, D. (2001). The past, present and future of school improvement: towards the third age. *British Educational Research Journal*, 27(4), 459–475. <https://doi.org/10.1080/01411920120071461>
- Hox, J. J., Moerbeek, M., & van de Schoot, R. (2018). Introduction to multilevel analysis. In author (eds.), *Multilevel Analysis, Techniques and Applications. Quantitative Methodology Series* (3rd ed., pp. 1-8). ISBN: 978-131-565-098-2
- Humphrey, S. E., & LeBreton, J. M. (2019). Introduction. In authors (Eds.), *The Handbook of Multilevel Theory, Measurement, and Analysis* (pp. 3-8). <http://dx.doi.org/10.1037/0000115-000>
- Ikeda, M., y García, E. (2014). *Grade repetition: A comparative study of academic and non-academic consequences*. http://dx.doi.org/10.1787/eco_studies-2013-5k3w65mx3hnx
- Instituto Nacional de Estadística y Geografía [INEGI]. (2015). Población. Recuperado en septiembre 22 de 2020, de https://www.inegi.org.mx/contenidos/programas/intercensal/2015/doc/eic_2015_presentacion.pdf
- Intxausti, N., Etxeberria, F., & Bartau, I. (2016). Effective and inclusive schools? Attention to diversity in highly effective schools in the Autonomous Region of the Basque Country.

International Journal of Inclusive Education, 21(1), 14–30.

<https://doi.org/10.1080/13603116.2016.1184324>

Iñiguez-Berrozpe, T., y Marcaletti, F. (2018). Modelos lineales multinivel en SPSS y su aplicación en investigación educativa. *Revista d'Innovació i Recerca en Educació*, 11(1), 26-40.

<http://doi.org/10.1344/reire2018.11.118984>

Jiménez, J. A. (2018). México en PISA 2015. Desaciertos de su aplicación por medios computarizados. *Revista Mexicana de Investigación Educativa*, 23(75), 711-731.

<http://www.comie.org.mx/revista/v2018/rmie/index.php/nrmie/article/view/1173/1156>

Joaristi, L. M., Lizasoain, L., y Azpillaga, V. (2014). Detección y caracterización de los centros escolares de alta eficacia de la comunidad autónoma del país vasco mediante modelos transversales contextualizados y modelos jerárquicos lineales. *Estudios Sobre Educación*,

27, 37–61. <https://doi.org/10.15581/004.27.37-61>

Johnson, N. L. (1949). Systems of frequency curves generated by methods of translation.

Biometrika, 36(1/2), 149. <https://doi.org/10.2307/2332539>

Jornet, J. M. (2016). Análisis metodológico del Proyecto PISA como evaluación internacional.

Revista Electrónica de Investigación y Evaluación Educativa, 22(1), 1-26.

<http://dx.doi.org/07.7203/relieve22.1.8293>

Jornet, J. M., López-González, E., y Tourón, J. (2012). Evaluación de sistemas educativos: teoría y experiencia. *Bordón*, 64(2), 9-11. [https://recyt.fecyt.es/index.php/BORDON/article/view](https://recyt.fecyt.es/index.php/BORDON/article/view/29031/15490)

[/29031/15490](https://recyt.fecyt.es/index.php/BORDON/article/view/29031/15490)

Joyce, B., Weil, M. y Calhoun, E. (2000). *Models of Teaching*. Allyn y Bacon.

- Kankaraš, M., y Moors, G. (2014). Analysis of Cross-Cultural Comparability of PISA 2009 Scores. *Journal of Cross-Cultural Psychology*, 45(3), 381–399. <https://doi.org/10.1177/0022022113511297>
- Karakolidis, A., Pitsia, V., & Emvalotis, A. (2016). Mathematics low achievement in Greece: A multilevel analysis of the Programme for International Student Assessment (PISA) 2012 data. *Themes in Science & Technology Education*, 9(1), 3-24. <http://earthlab.uoi.gr/theste/index.php/theste/article/view/199/118>
- Kriegel, H.-P., Kröger, P., Sander, J., & Zimek, A. (2011). Density-based clustering. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 1(3), 231–240. <https://doi.org/10.1002/widm.30>
- Krüger, N. (2019). La segregación por nivel socioeconómico como dimensión de la exclusión educativa: 15 años de evolución en América Latina. *Archivos Analíticos de Políticas Educativas*, 27(8). <http://dx.doi.org/10.14507/epaa.27.3577>
- Kyriacou, C. (2009). *Effective Teaching in Schools: theory and practice* (3rd ed.). Nelson Thornes.
- Kyriakides, L. (2005). Extending the comprehensive model of educational effectiveness by an empirical investigation. *School Effectiveness and School Improvement*, 16(2), 103–152. <https://doi.org/10.1080/09243450500113936>
- Kyriakides, L. (2008). Testing the validity of the comprehensive model of educational effectiveness: a step towards the development of a dynamic model of effectiveness. *School Effectiveness and School Improvement*, 19(4), 429–446. <https://doi.org/10.1080/09243450802535208>
- Kyriakides, L., & Creemers, B. P. M. (2008). Using a multidimensional approach to measure the impact of classroom-level factors upon student achievement: a study testing the validity of

- the dynamic model. *School Effectiveness and School Improvement*, 19(2), 183–205.
<https://doi.org/10.1080/09243450802047873>
- Kyriakides, L., & Creemers, B. P. M. (2009). The effects of teacher factors on different outcomes: two studies testing the validity of the dynamic model, *Effective Education*, 1(1), 61-85.
Doi: [10.1080/19415530903043680](https://doi.org/10.1080/19415530903043680)
- Kyriakides, L., Archambault, I., & Janosz, M. (2013). Searching for stages of effective teaching: A study testing the validity of the dynamic model in Canada. *Journal of Classroom Interaction*, 48(2), 11–24. <https://www.jstor.org/stable/43858891>
- Kyriakides, L., Campbell, R. J., & Gagatsis, A. (2000). The significance of the classroom effect in primary schools: an application of Creemers' comprehensive model of educational effectiveness. *School Effectiveness and School Improvement*, 11(4), 501–529.
<https://doi.org/10.1076/sesi.11.4.501.3560>
- Kyriakides, L., Christoforou, C., & Charalambous, C. Y. (2013). What matters for student learning outcomes: A meta-analysis of studies exploring factors of effective teaching. *Teaching and Teacher Education*, 36, 143–152. <https://doi.org/10.1016/j.tate.2013.07.010>
- Kyriakides, L., Creemers, B. P. M., & Antoniou, P. (2009). Teacher behavior and student outcomes: Suggestions for research on teacher training and professional development. *Teaching and Teacher Education*, 25(1), 12–23. <https://doi.org/10.1016/j.tate.2008.06.001>
- Kyriakides, L., Creemers, B. P. M., & Panayiotou, A. (2018). Using educational effectiveness research to promote quality of teaching: the contribution of the dynamic model. *ZDM*, 50(3), 381–393. <https://doi.org/10.1007/s11858-018-0919-3>

- Instituto Nacional para la Evaluación de la Educación (2011). *La Teoría del Error de Traducción de Pruebas y las evaluaciones internacionales de TIMSS y PISA*. <https://historico.mejoredu.gob.mx/wp-content/uploads/2019/01/P1C144.pdf>
- Lafontaine, D., Baye, A., Vieluf, S., & Monseur, C. (2015). Equity in opportunity-to-learn and achievement in reading: A secondary analysis of PISA 2009 data. *Studies in Educational Evaluation, 47*, 1–11. <https://doi.org/10.1016/j.stueduc.2015.05.001>
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159. <https://doi.org/10.2307/2529310>
- Lau, K., & Ho, E. S. (2015). Reading performance and self-regulated learning of Hong Kong students: what we learnt from PISA 2009. *The Asia-Pacific Education Researcher, 25*(1), 159–171. <https://doi.org/10.1007/s40299-015-0246-1>
- Lau, K., & Lam, T. Y. (2017). Instructional practices and science performance of 10 top-performing regions in PISA 2015. *International Journal of Science Education, 39*(15), 2128–2149. <http://dx.doi.org/10.1080/09500693.2017.1387947>
- 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
- LeRoy, B.-W., Samuel, P., Deluca, M., & Evans, P. (2018). Students with special educational needs within PISA. *Assessment in Education: Principles, Policy & Practice, 1-11*. <https://doi.org/10.1080/0969594X.2017.1421523>
- Lizasoain, L. (2020). Criterios y modelos estadísticos de eficacia escolar. *Revista de Investigación Educativa, 38*(2), 311-327. <https://doi.org/10.6018/rie.417881>

- Lizasoain, L., y Angulo, A. (2014). Buenas prácticas de escuelas eficaces del país vasco. Metodología y primeros resultados. *Participación Educativa*, 3(4), 17–27. <https://sede.educacion.gob.es/publiventa/d/20026/19/0>
- Lüdtke, O., Marsh, H. W., Robitzsch, A., Trautwein, U., Asparouhov, T., & Muthén, B. (2008). The multilevel latent covariate model: A new, more reliable approach to group-level effects in contextual studies. *Psychological Methods*, 13(3), 203–229. <https://doi.org/10.1037/a0012869>
- Ma, X., & Crocker, R. (2007). Provincial effects on reading achievement. *The Alberta Journal of Educational Research*, 53(1), 87-109. <https://jmss.org/index.php/ajer/article/view/55201>
- Marchionni, M., & Vazquez, E. (2018). The causal effect of an extra year of schooling on skills and knowledge in Latin America. Evidence from PISA. *Assessment in Education: Principles, Policy & Practice*, 1-27. <https://doi.org/10.1080/0969594X.2018.1454401>
- Marksteiner, T., Kuger, S., & Klieme, E. (2018). The potential of anchoring vignettes to increase intercultural comparability of non-cognitive factors. *Assessment in Education: Principles, Policy & Practice*, 1-21. <https://doi.org/10.1080/0969594X.2018.1514367>
- Martens, K. (2007). How to become an influential actor: The ‘comparative turn’ in OECD education policy. En K. Martens, A. Rusconi & K. Leuze (Eds.), *New Arenas of Education Governance* (pp. 40-56). Palgrave Macmillan.
- Martínez-Abad, F. (2019). Identification of factors associated with school effectiveness with data mining techniques: testing a new approach. *Frontiers in Psychology*, 10, <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. *Studies in Educational Evaluation*, 66, 100875. <https://doi.org/10.1016/j.stueduc.2020.100875>
- Martínez-Abad, F., Lizasoain, L., Castro, M. y Joaristi, L. M. (2017). Selección de escuelas de alta y baja eficacia en Baja California (México). *Revista Electrónica de Investigación Educativa*, 19(2), 38-53. <https://doi.org/10.24320/redie.2017.19.2.960>
- Martínez-Garrido, C. (2015). *Investigación sobre Enseñanza Eficaz: Un estudio multinivel para Iberoamérica* [Tesis doctoral, Universidad Autónoma de Madrid]. https://repositorio.uam.es/xmlui/bitstream/handle/10486/665887/martinez_garrido_cynthia.pdf?sequence=1
- Martínez-Rizo, F. (2005). Evitemos lecturas sensacionalistas sin cerrar los ojos. *Revista Mexicana de Investigación Educativa*, 10(24), 231-241. <http://www.comie.org.mx/revista/v2018/rmie/index.php/nrmie/article/view/759/759>
- Martínez-Rizo, F. (2016). Impacto de las pruebas en gran escala en contextos de débil tradición técnica: Experiencia de México y el Grupo Iberoamericano de PISA. *Revista Electrónica de Investigación y Evaluación Educativa*, 22(1), 1-12. <http://dx.doi.org/10.7203/relieve.22.1.8244>
- Massot, I., Dorio, I., y Sabariego, M. (2009). Estrategias de recogida y análisis de la información. En R. Bisquerra (Coord.), *Metodología de la Investigación Educativa* (pp. 329-366). La Muralla, S. A.
- Mayer, D., & Marland, P. (1997). Teachers' Knowledge of Students: a significant domain of practical knowledge? *Asia-Pacific Journal of Teacher Education*, 25(1), 17-34. <https://doi.org/10.1080/1359866970250103>

- Medley, D. M. (1977). *Teacher Competence and Teacher Effectiveness. A Review of Process-Product Research*. <https://files.eric.ed.gov/fulltext/ED143629.pdf>
- Meggiolaro, S. (2017). Information and communication technologies use, gender and mathematics achievement: evidence from Italy. *Social Psychology of Education*, 21(2), 497–516. <https://doi.org/10.1007/s11218-017-9425-7>
- Mikk, J., Krips, H., Säälük, Ü., & Kalk, K. (2015). Relationships between student perception of teacher-student relations and PISA results in mathematics and science. *International Journal of Science and Mathematics Education*, 14(8), 1437–1454. <https://doi.org/10.1007/s10763-015-9669-7>
- Mortimore, P., Sammons, P., Stoll, L., Lewis, D. & Ecob, R. (1988). *School matters: the junior years*. SAGE
- Mosquera, J. C., Mosquera, J. D., y Medina, P. D. (2010). Evaluación del índice de capacidad del proyecto de deserción académica en la Universidad Tecnológica de Pereira (UTP). *Revista Educación en Ingeniería*, (9), 96–103. <https://educacioneningenieria.org/index.php/e-di/article/view/3/2>
- Muijs, D., & Reynolds, D. (2011). *Effective Teaching: research and practice* (3rd ed.). SAGE.
- Muijs, D., & Reynolds, D. (2017). *Effective Teaching: research and practice* (4th ed.). SAGE.
- Muijs, D., Kyriakides, L., van de Werf, G., Creemers, B. P. M., Timperley, H., & Earl, L. (2014). State of the art – teacher effectiveness and professional learning. *School Effectiveness and School Improvement*, 25(2), 231–256. <http://dx.doi.org/10.1080/09243453.2014.885451>
- Murillo, F. J., Hernández-Castilla, R. (2011). Efectos escolares de factores socio-afectivos. Un estudio Multinivel para Iberoamérica. *Revista de Investigación Educativa*, 29(2), 407-427. <https://revistas.um.es/rie/article/view/111811/135341>

- Nagy, G., Nagengast, B., Frey, A., Becker, M., & Rose, N. (2018). A multilevel study of position effects in PISA achievement tests: Student- and school-level predictors in the German tracked school system. *Assessment in Education: Principles, Policy & Practice*, 1-22. <https://doi.org/10.1080/0969594X.2018.1449100>
- Ning, B., Van Damme, J., Gielen, S., Vanlaar, G., & Van den Noortgate, W. (2015). What makes the difference in reading achievement? Comparisons between Finland and Shanghai. *Scandinavian Journal of Educational Research*, 60(5), 515–537. <http://dx.doi.org/10.1080/00313831.2015.1062413>
- Obilor, E. I. (2019). Influence of teaching effectiveness on students' learning outcome. *International Journal of Innovative Social & Science Education Research*, 7(2), 37-48. <https://seahipaj.org/journals-ci/june-2019/IJISSER/full/IJISSER-J-4-2019.pdf>
- Organización para la Cooperación y el Desarrollo Económico [OCDE]. (2006). *El programa PISA de la OCDE. Qué es y para qué sirve*. <https://www.oecd.org/pisa/39730818.pdf>
- Organización para la Cooperación y el Desarrollo Económico [OCDE]. (2010). *Acuerdo de cooperación México-OCDE para mejorar la calidad de la educación de las escuelas mexicanas*. <https://www.oecd.org/education/school/46216786.pdf>
- Organización para la Cooperación y el Desarrollo Económico [OCDE]. (2011). *La medición del aprendizaje de los alumnos: Mejores prácticas para evaluar el valor agregado de las escuelas*. <http://dx.doi.org/10.1787/9789264090170-es>
- Organización para la Cooperación y el Desarrollo Económico [OCDE]. (2017a). *Marco de referencia para el cuestionario de contexto, PISA 2018*. Instituto Colombiano para la Evaluación de la Educación (ICFES). ISBN: 978-958-11-0736-0

- Organization for Economic Co-operation and Development [OCDE en español]. (2002a). *PISA 2000 Technical Report*. <https://www.oecd.org/pisa/data/33688233.pdf>
- Organization for Economic Co-operation and Development [OCDE en español]. (2002b). *Manual for the PISA 2000 Database*. <https://www.oecd.org/pisa/pisaproducts/PISAMANUAL1020517.zip>
- Organization for Economic Co-operation and Development [OCDE en español]. (2005a). *PISA 2003 Technical Report*. <https://www.oecd.org/education/school/programmeforinternationalstudentassessmentpisa/35188570.pdf>
- Organization for Economic Co-operation and Development [OCDE en español]. (2005b). *PISA 2003 Data Analysis Manual*. <https://www.oecd.org/education/school/programmeforinternationalstudentassessmentpisa/35004299.pdf>
- Organization for Economic Co-operation and Development [OCDE en español]. (2009a). *PISA 2006 Technical Report*. <https://www.oecd.org/pisa/data/42025182.pdf>
- Organization for Economic Co-operation and Development [OCDE en español]. (2009b). *PISA Data Analysis Manual*. http://archivos.agenciaeducacion.cl/Manual_de_Analisis_de_datos_SPSS_version_ingles.pdf
- Organization for Economic Co-operation and Development [OCDE en español]. (2012). *How does class size vary around the world?* <https://doi.org/10.1787/22267077>
- Organization for Economic Co-operation and Development [OCDE en español]. (2013). *PISA 2012 Assessment and Analytical Framework: Mathematics, Reading, Science, Problem Solving and Financial Literacy*. <https://doi.org/10.1787/19963777>

- Organization for Economic Co-operation and Development [OCDE en español]. (2014). PISA 2012 Technical Report. <https://www.oecd.org/pisa/pisaproducts/PISA-2012-technical-report-final.pdf>
- Organization for Economic Co-operation and Development [OCDE en español]. (2016a). *Sampling in PISA*. <https://www.oecd.org/pisa/pisaproducts/SAMPLING-IN-PISA.pdf>
- Organization for Economic Co-operation and Development [OCDE en español]. (2017b). *PISA 2015 Technical Report*. https://www.oecd.org/pisa/data/2015-technical-report/PISA2015_TechRep_Final.pdf
- Organization for Economic Co-operation and Development [OCDE en español]. (2017c). *PISA 2015 Assessment and Analytical Framework: Science, Reading, Mathematic, Financial Literacy and Collaborative Problem Solving*. <http://dx.doi.org/10.1787/9789264281820-en>
- Organization for Economic Co-operation and Development [OCDE en español]. (2019a). *PISA 2018 Results. Combined Executive Summaries*. https://www.oecd.org/pisa/Combined_Executive_Summaries_PISA_2018.pdf
- Organization for Economic Co-operation and Development [OCDE en español]. (2019b). *PISA 2018 Assessment and Analytical Framework*. <https://doi.org/10.1787/b25efab8-en>
- Ornstein, A. C., & Levine, D. U. (1981). Teacher behavior research: overview and outlook. *The Phi Delta Kappan*, 62(8), 592-596. <https://www.jstor.org/stable/20386030>
- Owsiński, J. W. (2020). *Data Analysis in Bi-partial Perspective: Clustering and Beyond*. Springer. <https://doi.org/10.1007/978-3-030-13389-4>
- Özdemir, C. (2016). Equity in the Turkish educational system: A multilevel analysis of social background influences on the mathematics performance of 15-years-old students.

- European Educational Research Journal*, 15(2), 193-217.
<https://doi.org/10.1177/1474904115627159>
- Panayiotou, A., Kyriakides, L., & Creemers, B. P. M. (2015). Testing the validity of the dynamic model at school level: a European study. *School Leadership & Management*, 36(1), 1–20.
<https://doi.org/10.1080/13632434.2015.1107537>
- Panayiotou, A., Kyriakides, L., Creemers, B. P. M., McMahon, L., Vanlaar, G., Pfeifer, M., Rekalidou, G., & Bren, M. (2014). Teacher behavior and student outcomes: Results of a European study. *Educational Assessment, Evaluation and Accountability*, 26(1), 73–93.
<https://doi.org/10.1007/s11092-013-9182-x>
- Peculea, L., & Peculea A. (2018). Contemporary perspectives on improving effective teaching at first year students in engineering higher education- a students' perspective. *Journal Plus Education*, XVIII(nov), 70-87. <https://www.uav.ro/jour/index.php/jpe/article/view/852/917>
- Pedroza, L. H., Peniche, R. y Lizasoain, L. (2018). Criterios para la identificación y selección de escuela eficaces de nivel medio superior. *Revista Electrónica de Investigación Educativa*, 20(1), 14-25. <https://doi.org/10.24320/redie.2018.20.1.2170>
- Petersen, K., Vakkalanka, S., & Kuzniarz, L. (2015). Guidelines for conducting systematic mapping studies in software engineering: An update. *Information and Software Technology*, 64, 1–18. <https://doi.org/10.1016/j.infsof.2015.03.007>
- Pituch, K. A., & Stevens, J. P. (2015). Hierarchical linear modeling. In authors, *Applied Multivariate Statistics for The Social Sciences* (6th ed., pp. 537-617). ISBN 13: 978-1-315-81491-9
- Popham, J. (2013). *Evaluación trans-formativa. El poder transformador de la evaluación formativa*. ISBN: 9788427719125

- Prasertcharoensuk, T., Uttarak, A., & Tang, K. N. (2018). The effect on teaching effectiveness: a multi-level analysis. *The Turkish Online Journal of Educational Technology*, 2, 714-721. https://www.ic.kku.ac.th/documents/pdf/2018_12_3.pdf
- Puschmann, D., Barnaghi, P., & Tafazolli, R. (2016). Marginal distribution clustering of multivariate streaming IoT data. *2016 IEEE 3rd World Forum on Internet of Things (WF-IoT)*. <https://doi.org/10.1109/wf-iot.2016.7845451>
- 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. *Revista Iberoamericana sobre Calidad, Eficacia y Cambio en Educación*, 16(4), 79-97. <https://doi.org/10.15366/reice2018.16.4.005>
- Raudenbush, S. W., & Bryk, A. S. (1986). A hierarchical model for studying school effects. *Sociology of Education*, 59(1), 1-17. <https://doi.org/10.2307/2112482>
- Raudenbush, S. W., Bryk, A. S., Cheong, Y. F., & Congdon, R. T. (2016). *HLM7: Hierarchical Linear and Nonlinear Modeling*. Scientific Software International, Inc.
- Reynolds, D. (1982). The search for effective schools. *School Organisation*, 2(3), 215–237. <https://doi.org/10.1080/0260136820020302>
- Reynolds, D. (1988). British school improvement research: the contribution of qualitative studies. *International Journal of Qualitative Studies in Education*, 1(2), 143–154. <https://doi.org/10.1080/0951839880010203>
- Reynolds, D., Creemers, B. P. M., Stringfield, S., Teddlie, C., & Schaffer, E. (2002). *World class schools: International perspectives in school effectiveness*. Routledge Falmer.
- Reynolds, D., Sammons, P., De Fraine, B., Van Damme, J., Townsend, T., Teddlie, C., & Stringfield, S. (2014). Educational effectiveness research (EER): a state-of-the-art review.

- School Effectiveness and School Improvement*, 25(2), 197-230.
<https://doi.org/10.1080/09243453.2014.885450>
- Rivas, A. (2015). *América Latina después de PISA: Lecciones aprendidas de la educación en siete países (2000-2015)*. CIPPEC-Natura-Instituto Natura. ISBN 978-987-1479-46-7
- Robinson, W. (2004). *Power to Teach*. ISBN 0-203-58104-0
- Robitzsch, A., & Lüdtke, O. (2018). Linking errors in international large-scale assessments: Calculation of standard errors for trend estimation. *Assessment in Education: Principles, Policy & Practice*, 1-22. <https://doi.org/10.1080/0969594X.2018.1433633>
- Roni, S. M., Merga, M. K., & Morris, J. E. (2020). *Conducting Quantitative Research in Education*. <https://doi.org/10.1007/978-981-13-9132-3>
- Rutledge, S. A., Cohen-Vogel, L., Osborne-Lampkin, L., & Roberts, R. L. (2015). Understanding effective high schools. *American Educational Research Journal*, 20(10), 1–33.
<https://doi.org/10.3102/0002831215602328>
- Sammons, P., Davis, S., & Gray, J. (2016). Methodological and scientific properties of school effectiveness research. Exploring the underpinnings, evolution, and future directions of the field. In C. Chapman, D. Muijs, D. Reynolds, P. Sammons, & C. Teddlie, C. (Eds.), *The Routledge International Handbook of Educational Effectiveness and Improvement. Research, policy, and practice* (pp. 25-77). Routledge.
- Sandilands, D. (Dallie), Barclay McKeown, S., Lyons-Thomas, J., & Ercikan, K. (2014). An investigation of school-level factors associated with science performance for minority and majority francophone students in Canada. *Canadian Journal of Science, Mathematics and Technology Education*, 14(2), 135–153. <https://doi.org/10.1080/14926156.2014.903316>

- Scheerens, J. (1991). Process indicators of school functioning: a selection based on the research literature on school effectiveness. *Studies in Educational Evaluation*, 17, 371-403. [https://doi.org/10.1016/S0191-491X\(05\)80091-4](https://doi.org/10.1016/S0191-491X(05)80091-4)
- Scheerens, J. (2013). *What is effective schooling? A review of current thought and practice*. <https://www.ibo.org/globalassets/publications/ib-research/continuum/what-is-effective-schooling-report-en.pdf>
- Scheerens, J. (2016). *Educational Effectiveness and Ineffectiveness. A Critical Review of the Knowledge Base*. Springer. <https://doi.org/10.1007/978-94-017-7459-8>
- Scheerens, J., & Bosker, R. (1997). *The foundations of educational effectiveness*. 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. <http://dx.doi.org/10.1080/13803611.2015.1008520>
- Scherbaum, C. A., & Pesner, E. (2019). Power analysis for multilevel research. In S. E. Humphrey & J. M. LeBreton (Eds.), *The Handbook of Multilevel Theory, Measurement, and Analysis* (pp. 329-352). <http://dx.doi.org/10.1037/0000115-000>
- Schleicher, A. (2006). Fundamentos y cuestiones políticas subyacentes al desarrollo de PISA. *Revista de Educación*, extraordinario 2006, 13-18. <http://www.educacionyfp.gob.es/dam/jcr:a44d5e62-0398-4b87-866c-9997490860cd/re200603-pdf.pdf>
- Schofield, H. L., & Start, K. B. (1979). Product variables as criteria of teacher effectiveness. *The Journal of Experimental Education*, 48(2), 130-136. <https://doi.org/10.1080/00220973.1979.11011726>

- Shepard, L.A. (1989). Why we need better assessment. *Educational Leadership*, 46(2), 4–8.
- Shera, P. (2014). School effects, gender and socioeconomic differences in reading performance: a multilevel analysis. *International Education Studies*, 7(11). <https://doi.org/10.5539/ies.v7n11p28>
- Silverman, S. (1988). Relationships of selected presage and context variables to achievement. *Research Quarterly for Exercise and Sport*, 59(1), 35–41. <https://doi.org/10.1080/02701367.1988.10605471>
- Simons, R.-J., van der Linden, J., & Duffy, T. (2002). *New Learning*. Kluwer Academic Publishers.
- Spada, N. M. (1987). Relationships between instructional differences and learning outcomes: a process–product study of communicative language teaching, *Applied Linguistics*, 8(2), 137–161. <https://doi.org/10.1093/applin/8.2.137>
- Spaull, N. (2018). Who makes it into PISA? Understanding the impact of PISA sample eligibility using Turkey as a case study (PISA 2003–PISA 2012). *Assessment in Education: Principles, Policy & Practice*, 1-25. <https://doi.org/10.1080/0969594X.2018.1504742>
- Stanovich, P. J., & Jordan, A. (1998). Canadian teachers' and principals' beliefs about inclusive education as predictors of effective teaching in heterogeneous classrooms. *The Elementary School Journal*, 98(3), 221-238. <https://doi.org/10.1086/461892>
- Stoll, L. & Sammons, P. (2007). Growing together: school effectiveness and school improvement in the UK. In T. Townsend (Ed.), *International Handbook of School Effectiveness and Improvement* (pp. 207-222). Springer.

- Stringfield, S., Teddlie, C., & Suarez, S. (1985). Classroom interaction in effective and ineffective schools: preliminary results from phase III of the Louisiana school effectiveness study. *The Journal of Classroom Interaction*, 20(2), 31-37. <http://www.jstor.org/stable/43997875>
- Sulis, I., & Porcu, M. (2014). Assessing divergences in mathematics and reading achievement in Italian primary schools: a proposal of adjusted indicators of school effectiveness. *Social Indicators Research*, 122(2), 607–634. <https://doi.org/10.1007/s11205-014-0701-z>
- Sumantri, M. S., Prayuningtyas, A. W., Rachmadtullah, R., & Magdalena, I. (2018). The roles of teacher-training programs and student teachers' self-regulation in developing competence in teaching science. *Advanced Science Letters*, 24(10), 7077–7081(5). <https://doi.org/10.1166/asl.2018.12412>
- Teddlie, C., & Reynolds, D. (2000). *The international handbook of school effectiveness research*. Falmer Press.
- Teddlie, C., & Reynolds, D. (2003). *The international handbook of school effectiveness research*. Taylor & Francis Group
- Teddlie, C., & Stringfield, S. (1993). *Schools make a difference: Lessons learned from a 10-year study of school effects*. Teachers College Press.
- Teddlie, C., Kirby, P.C., & Stringfield, S. (1989). Effective versus ineffective schools: Observable differences in the classroom. *American Journal of Education*, 97(3), 221–36. <http://www.jstor.org/stable/1085165>
- Timmermans, A. C., Bosker, R. J., de Wolf, I. F., Doolaard, S., & van der Werf, M. P. C. (2013). Value added based on educational positions in Dutch secondary education. *British Educational Research Journal*, 40(6), 1057–1082. <https://doi.org/10.1002/berj.3126>

- Townsend, T. (2007). *The international handbook of school effectiveness and school improvement*. Springer.
- Turner, R. & Adams, R. J. (2007). The programme for international student assessment: an overview. *Journal of Applied Measurement*, 8(3), 237–248.
- Twain, M. (2018). Prediction. In K. Imai, *Quantitative Social Sciences: An introduction* (pp. 123-189). Princeton University Press.
- Vandenberg, R. J., & Richardson, H. A. (2019). A primer on multilevel structural modeling: user-friendly guidelines. In S. E. Humphrey & J. M. LeBreton (Eds.), *The Handbook of Multilevel Theory, Measurement, and Analysis* (pp. 449-472). <http://dx.doi.org/10.1037/0000115-000>
- Vélaz, C. (2006). Presentación. Una visión integral de las evaluaciones del PISA (OCDE) con especial atención a la participación de España. *Revista de Educación*, extraordinario 2006, 11-12. <http://www.educacionyfp.gob.es/dam/jcr:15969fbc-2f6c-4222-8694-9f694d12eb64/re200602-pdf.pdf>
- Venetti, M. (2007). Confusion matrix online calculator <http://www.marcovanetti.com/pages/cfmatrix/>
- Vilà-Baños, R., Rubio-Hurtado, M.-J., Berlanga-Silvente, V., y Torrado-Fonseca, M. (2014). Cómo aplicar un cluster jerárquico en SPSS. *Revista d'Innovació i Recerca en Educació*, 7(1), 113-127. <http://diposit.ub.edu/dspace/bitstream/2445/65577/1/628893.pdf>
- Virgilio, L, Teddlie, C., & Oescher, J. (1991). Variance and context differences in teaching at differentially effective schools. *School Effectiveness and School Improvement*, 2(2), 152–68. <https://doi.org/10.1080/0924345910020206>

- Von Davier, M., González, E., & Mislevy, R. (2009). What are plausible values and why are they useful. *IERI Monograph Series*, 2, 9-36. http://www.ierinstitute.org/fileadmin/Documents/IERI_Monograph/IERI_Monograph_Volume_02_Chapter_01.pdf
- Walker, A. D., Lee, M., & Bryant, D. A. (2014). How much of a difference do principals make? An analysis of between-schools variation in academic achievement in Hong Kong public secondary schools. *School Effectiveness and School Improvement*, 25(4), 602–628. <https://doi.org/10.1080/09243453.2013.875044>
- Waxman, H. C., & Fash, M. J. (1983). Utilizing students' perceptions and context variables to analyze effective teaching: a process-product investigation. *The Journal of Educational Research*, 76(6), 321–325. <https://doi.org/10.1080/00220671.1983.10885476>
- Weber, G. (1971). Inner-city children can be taught to read: Four successful schools. Council for Basic Education.
- Wu, M. (2005). The role of plausible values in large-scale surveys. *Studies in Educational Evaluation*, 31(2-3), 114-128. <https://doi.org/10.1016/j.stueduc.2005.05.005>
- Yair, G. (1997). When classrooms matter: implications of between-classroom variability for educational policy in Israel. *Assessment in Education*, 4(2), 225–248. <https://doi.org/10.1080/0969594970040202>
- Yang, H. (2017). A research on the effective questioning strategies in class. *Science Journal of Education*, 5(4), 158–163. <https://doi.org/10.11648/j.sjedu.20170504.16>
- Yerdelen, S., & Sungur, S. (2018). Multilevel investigation of students' self-regulation processes in learning science: classroom learning environment and teacher effectiveness. *International Journal of Science and Mathematics Education*, 17, 89–110. <http://dx.doi.org/10.1007/s10763-018-9921-z>

Yew, E. H. J., & Goh, K. (2016). Problem-based learning: an overview of its process and impact on learning. *Health Professions Education*, 2(2), 75–79.

<http://dx.doi.org/10.1016/j.hpe.2016.01.004>

Yıldırım, S. (2012). Teacher support, motivation, learning strategy use, and achievement: a multilevel mediation model. *The Journal of Experimental Education*, 80(2), 150–172.

<https://doi.org/10.1080/00220973.2011.596855>

Zabalza, M. A. (2007). *Competencias del profesorado universitario. Calidad y desarrollo profesional*. ISBN 84-277-1399-1