

revista de **e**EDUCACIÓN

Nº 412 ABRIL-JUNIO 2026



GOBIERNO
DE ESPAÑA

MINISTERIO
DE EDUCACIÓN, FORMACIÓN PROFESIONAL
Y DEPORTES

2025
CALIDAD
REVISTAS
CIENTÍFICAS
ESPAÑOLAS

Ministerio de Ciencia, Innovación y Universidades

AGENCIAS
NACIONALES
DE EVALUACIÓN

RECETES

Ministerio de Educación, Formación Profesional y Deportes

Ministerio de Ciencia, Innovación y Universidades

Ministerio de Educación, Formación Profesional y Deportes

Ministerio de Ciencia, Innovación y Universidades

Ministerio de Educación, Formación Profesional y Deportes

Ministerio de Ciencia, Innovación y Universidades

revista de
eEDUCACIÓN



Nº 412 ABRIL-JUNIO 2026

revista de EDUCACIÓN

Nº 412 Abril-Junio 2026

Revista trimestral

Fecha de inicio: 1952

MINISTERIO DE EDUCACIÓN Y FORMACIÓN PROFESIONAL

SECRETARÍA DE ESTADO DE EDUCACIÓN Y FORMACIÓN PROFESIONAL

Instituto Nacional de Evaluación Educativa
Paseo del Prado, 28, 4ª planta
28014 Madrid
España

Edita
© SECRETARÍA GENERAL TÉCNICA
Subdirección General de Atención al Ciudadano, Documentación y Publicaciones

Catálogo de publicaciones del Ministerio: sede.educacion.gob.es
Catálogo general de publicaciones oficiales: publicacionesoficiales.boe.es

Edición: 2026
NIPO pdf: 164-24-063-9
NIPO ibd: 164-24-061-8
NIPO html: 164-24-062-3
ISSN línea: 1988-592X
ISSN papel: 0034-8082
Depósito Legal: M.57/1958

Diseño de portada: Iris Bermejo Rodríguez
Fotografía: Ron Lach (Pexels)
Fuente: <https://www.pexels.com/es-es/foto/sentado-joven-nino-adolescente-9783350/>
Maqueta: Ibersaf Industrial S.L.

CONSEJO DE DIRECCIÓN

PRESIDENCIA

Abelardo de la Rosa
Secretario de Estado de Educación

Esther Monterrubio Ariznabarreta
Secretaria General de Formación Profesional

VOCALES

Santiago Antonio Roura Gómez
Subsecretario de Educación, Formación Profesional y Deportes

Mónica Domínguez García
Directora General de Evaluación y Cooperación Territorial

Carmen Tovar Sánchez
Directora del Instituto Nacional de Evaluación Educativa

Lucio Calleja Bachiller
Subdirector General de Cooperación Territorial
e Innovación Educativa

Helena Ramos García
Subdirectora General de Ordenación Académica

CONSEJO EDITORIAL

DIRECTORA

Carmen Tovar Sánchez

EDITOR JEFE

José Luis Gaviria Soto

EDITOR ADJUNTO

David Reyero García

VOCALES

Antonio Cabrales Goitia (University College London); Caterina Casalmiglia (Universitat Autònoma de Barcelona); Antonio Lafuente García (Consejo Superior de Investigaciones Científicas); Leoncio López-Ocón Cabrera, (Consejo Superior de Investigaciones Científicas); Carlos Marcelo García (Universidad de Sevilla); Miquel Martínez Martín (Universitat de Barcelona); Francisco Michavila Pitarch; (Universidad Politécnica de Madrid); Juan Manuel Moreno Olmedilla (Banco Mundial); Clara Eugenia Núñez (Universidad Nacional de Educación a Distancia); Lucrecia Santibáñez (Claremont Graduate University); Denise Vaillant (Universidad ORT, Uruguay); Pablo Zoido (Banco Interamericano de Desarrollo).

REDACCIÓN

Jefe de Redacción: Julio López Sevilla

Colaboradores: Ruth Martín Escanilla, Alberto Díaz-Ron-cero Canales, Jasmine Blasco Yurevich y Marta Carrera González

ASESORES CIENTÍFICOS

Internacional

Aaron Benavot (State University of New York, SUNY-Al-bany); Abdeljalil Akkari (Universidad de Ginebra); Mark Bray (University of Hong Kong); José Joaquín Brunner (Universidad Diego Portales, Chile); Dirk Hastedt (Executive Director; International Association for the Evaluation of Educational Achievement, IEA); Andreas Schleicher (Director; Directorate for Education and Skills, OCDE).

Nacional

Teresa Aguado Odina (Universidad Nacional de Educación a Distancia); Margarita Bartolomé (Universitat de Barcelona); Antonio Bolívar (Universidad de Granada); Josefina Cambra Giné (Colegio de Doctores y Licenciados); Anna Camps i Mundó (Universitat Autònoma de Barcelona); César Coll Salvador (Universitat de Barcelona); Agustín Dosil Maceira (Universidad LiberQuaré); Gerardo Echeita Sarrionandia (Universidad Autónoma de Madrid); Juan Manuel Escudero Muñoz (Universidad de Murcia); Mariano Fernández Enguita (Universidad Complutense de Madrid); Joaquín Gairín Sallán (Universitat Autònoma de Barcelona); José Luis García Garrido (Universidad Nacional de Educación a Distancia); Daniel Gil Pérez (Universitat de València); Fuensanta Hernández Pina (Universidad de Murcia); Carmen Labrador Herraiz (Universidad Complutense de Madrid); Miguel López Meler (Universidad de Málaga); Elena Martín Ortega (Universidad Autónoma de Madrid); Rosario Martínez Arias (Universidad Complutense de Madrid); Inés Miret (Neturity S.L., Madrid); Juan Ignacio Pozo (Universidad Autónoma de Madrid); Joaquim Prats Cuevas (Universitat de Barcelona); Manuel de Puelles (Universidad Nacional de Educación a Distancia); Tomás Recio Muñiz (Universidad de Cantabria); Luis Rico Romero (Universidad de Granada); Juana M.ª Sancho Gil (Universitat de Barcelona); Mercedes Vico Monteoliva (Universidad de Málaga); Antonio Viñao Frago (Universidad de Murcia); Jesús García Laborda (Universidad de Alcalá).

Investigación

FERNANDO GIL CANTERO: La educación no es una actividad transhumanista.....	3
LAURA ORTEGA TORRES Y CÉSAR ANTONIO CHÁVEZ ÁLVAREZ: Análisis de funcionamiento diferencial en evaluaciones a gran escala	21
ELENA CASTRO-RODRÍGUEZ, JUAN LUIS PIÑEIRO Y ENRIQUE CASTRO: Resolución de problemas matemáticos en la formación del profesorado de educación primaria española	45
MARIO ALARCÓN BRAVO Y JOSÉ JOAQUÍN BRUNNER RIED: Autonomía universitaria: un modelo analítico multinivel desde una perspectiva latinoamericana	71
MICAELA SÁNCHEZ-MARTÍN, MARTA GUTIÉRREZ-SÁNCHEZ, EVA MARÍA OLMEDO-MORENO Y FERNANDO NAVARRO-MATEU: Adherencia a la declaración PRISMA en los meta-análisis de intervenciones experimentales publicados en Educación: una meta-revisión sistemática.....	95
MERY ESTEFANÍA BUESTÁN-JÁTIMA, ELENA ORTEGA-CAMPOS, LETICIA DE LA FUENTE-SÁNCHEZ Y JUAN GARCÍA-GARCÍA: Eficacia de los Programas Psico-socio-educativos aplicados en Justicia Juvenil en España: Revisión Sistemática.....	131
GARAZI ORMAZABAL-ARIZKORRETA, ANDONI ARGUIÑANO MADRAZO E INAKI KARRERA XUARROS: Sentbrando el proyecto vital desde la infancia: Aprendizaje Sociocultural en <i>Eskola Txikiak</i>	159
ANDREA BETTI, PABLO BIDERBOST Y ESTHER VAQUERO: Comparación entre el pensamiento visual y el aula invertida en la enseñanza de las relaciones internacionales	183
Reseñas	213

INVESTIGACIÓN

La educación no es una actividad transhumanista

Education is not a transhumanist activity

<https://doi.org/10.4438/1988-592X-RE-2025-412-740>

Fernando Gil Cantero

<https://orcid.org/0000-0002-6636-7601>

Universidad Complutense de Madrid

Resumen

Introducción. El debate contemporáneo sobre el transhumanismo ha generado interpretaciones que lo acercan a la educación como vía de mejora del ser humano. Este artículo cuestiona esa equivalencia, argumentando que educación y transhumanismo responden a lógicas y fines radicalmente distintos. **Método.** A través de un enfoque filosófico y pedagógico, se analizan críticamente las premisas del transhumanismo desde la antropología, la ética y la teoría educativa. El artículo contrapone el ideal transhumanista de ampliación funcional de capacidades con la tradición teleológica de la formación humana. **Resultados.** Se sostiene que, mientras el transhumanismo apuesta por la optimización técnica del individuo -especialmente en sus capacidades cognitivas, físicas o sensoriales-, la educación se orienta a la maduración ética del sujeto, mediante relaciones humanas, cultura y esfuerzo. No toda ampliación de capacidades constituye una mejora educativa: esta solo ocurre cuando la autonomía se orienta éticamente hacia fines valiosos, reconociendo la vulnerabilidad y dependencia humanas. Asimismo, se rechaza la reducción de la ética a parámetros biologicistas y se reivindica una comprensión del ser humano como agente moral teleológicamente orientado. **Discusión.** Frente a la hegemonía cultural de la eficiencia y la fascinación tecnológica, el artículo defiende el papel insustituible de la educación como proceso humanizador. La pedagogía no debe subordinarse a la lógica del rendimiento ni a modas intelectuales, sino custodiar valores esenciales como el esfuerzo, la prudencia y la aceptación de los límites. En este marco, se propone una crítica a la ilusión tecnocrática de la perfectibilidad ilimitada, subrayando que el fin de la educación no es programar sujetos funcionales, sino formar personas capaces de obrar con sentido y responsabilidad en un mundo compartido.

Palabras clave: transhumanismo; posthumanismo; pedagogía; mejoramiento humano; naturaleza humana; educación; autonomía.

Abstract

Introduction. The contemporary debate on transhumanism has led to interpretations that bring it closer to education as a means of human enhancement. This article challenges that equivalence, arguing that education and transhumanism follow radically different logics and pursue fundamentally distinct ends. **Method.** Using a philosophical and pedagogical approach, the article critically examines the premises of transhumanism from the perspectives of anthropology, ethics, and educational theory. It contrasts the transhumanist ideal of functional enhancement of capacities with the teleological tradition of human formation. **Results.** The article argues that while transhumanism aims at the technical optimization of the individual -especially regarding cognitive, physical, or sensory abilities- education is oriented toward the ethical maturation of the person through human relationships, culture, and personal effort. Not every enhancement of capacity constitutes educational improvement: true improvement occurs only when autonomy is ethically oriented toward meaningful ends, acknowledging human vulnerability and interdependence. The article also rejects the biologicist reduction of ethics to evolutionary utility and affirms a view of the human being as a moral agent teleologically directed. **Discussion.** In the face of a cultural hegemony of efficiency and technological fascination, the article defends the irreplaceable role of education as a humanizing process. Pedagogy must not be subordinated to the logic of performance or to intellectual trends, but must safeguard essential formative values such as effort, prudence, and the acceptance of limits. Within this framework, the article offers a critique of the technocratic illusion of unlimited perfectibility, emphasizing that the ultimate goal of education is not to program functional subjects, but to form persons capable of acting with meaning and responsibility in a shared world.

Keywords: transhumanism; posthumanism; pedagogy; human enhancement; human nature; education; autonomy.

Introducción

En este artículo vamos a mantener tres tesis. En primer lugar, que la educación no es una actividad transhumanista porque la noción de mejora *educativa* es diferente a la noción de mejora transhumanista. En segundo lugar, que la realidad es mucho más diversa, diferente y amplia que lo que podemos captar con nuestros sentidos y capacidades, por lo que, para comprender el mundo -también la educación-, necesitamos una metafísica realista o sustancialista, no atomista o analítica. Y, por último, que para entender bien la educación necesitamos más visiones de sentido que de significado, más teorías comprensivas que descriptivas, más argumentos razonables que racionales o empíricos.

En el pensamiento pedagógico hay una tendencia no solo a adoptar las modas intelectuales del momento, sino incluso a identificarse completamente con ellas. Así, por ejemplo, cuando en la década de los 90 -especialmente en el ámbito académico español- comenzó a desarrollarse la idea de una intensa relación entre educación y tecnología, se llegó a afirmar que la educación *es* tecnología. Cuando se empezó a reconocer la importancia de la formación en competencias, se concluyó que la educación *es* una competencia. Al subrayar la dimensión ética de los procesos educativos, se acabó diciendo que la educación *es* ética. Hay así una inclinación a confundir las partes con el todo, especialmente en un campo como la Pedagogía donde, además de trabajar con conceptos imprecisos, se modifican sus significados según convenga. Pues bien, lo mismo está ocurriendo ahora, como vamos a mostrar, con el transhumanismo.

El título y contenido de este ensayo es, en parte, una réplica a la reciente afirmación del profesor José Luis Gaviria quien sostiene que “la educación es una actividad transhumanista en esencia” porque “(...) hay algo en común entre las técnicas transhumanistas y la educación. En los dos casos se trata de hacer algo al hombre, en transformar al propio hombre (2024, p. 14). Por esa misma lógica, podríamos afirmar también que ‘la educación es una actividad agrícola porque hay algo en común entre las técnicas agrícolas y la educación. En los dos casos se trata de implantar, crecer, desarrollar, etc.’. La educación, en efecto, participa de la lógica transhumanista, pero no es una actividad transhumanista. Del mismo modo, la educación participa de la metáfora agrícola, pero no es una actividad agrícola. Gaviria confunde las partes con el todo porque parte del presupuesto erróneo de que nuestra obligación como pedagogos es pensar “(...) en el papel de la educación en el transhumanismo” (*op. cit.*, p. 14) cuando en realidad de lo que se trata es de pensar en el papel del transhumanismo *en* la educación.

Sí que hay que agradecer al profesor Gaviria que reconozca que la discusión en torno al transhumanismo y la educación nos lleva inevitablemente a las preguntas más radicales que se pueden plantear. Por ejemplo, cuando advierte: “(p)arece lógico que en esa misma dinámica de duda total nos planteemos algo todavía más fundamental: qué significa ser un ser humano.” (*op. cit.*, p. 2); o cuando al seguir su propia argumentación tiene que parar y preguntarse ni más ni menos: “¿Cuándo surgen los valores, y por tanto la vida moral?” (*op. cit.*, p. 10). En cualquier caso, parte ya de cierta consideración poética al sostener que “(e)s posible que en nuestra naturaleza esté inscrita la determinación de violar todas las reglas que impidan realizar cualquier posibilidad que esté a nuestro alcance, sin que el miedo a perder el paraíso nos detenga.” (*op. cit.*, p. 2).

Podemos discutir si conviene o no poner un semáforo frente a un cruce de calles, pero lo que está claro es que hoy la Pedagogía se encuentra ante una encrucijada jamás planteada en la educación (Arana, 2024; Bostrom, 2024; Goering, Klein, Specker Sullivan, *et al.*, 2021; Kurzweil, 2025; Lacalle, 2022). Hace años, ya Fukuyama (2009) consideraba que el transhumanismo era la idea más peligrosa del mundo. Y G. Steiner tenía las mismas preocupaciones cuando además desconocía los actuales desconcertantes¹ avances tecnológicos:

¿Debe seguir adelante la investigación genética, sea cual sea la consecuencia social, humana? ¿Hay órdenes de la verdad abstracta irreconciliables con la justicia humana? (...). Cabe pensar que el ansia de conocimiento, el grito del cazador, que infectó la mente occidental hace casi 3000 años, nos está llevando a una emboscada (Steiner, 2022, p. 39 y p. 43).

En lo que sigue voy a tratar de argumentar el título del artículo yendo desde las cuestiones más metafísicas y epistemológicas a las más concretas. Como se trata de un ensayo y, sobre todo, de respuesta a otro organizaré mi argumentación en torno a una serie de puntos donde discrepo de las propuestas de Gaviria, por eso algunos apartados están redactados en un sintagma negativo para acentuar la oposición a la tesis contraria mantenida por él. Una última advertencia: no voy a perder de vista la perspectiva pedagógica que subyace a todo mi planteamiento de tal modo que si, en algunos puntos, sobre todo los primeros, no se observa directamente no es porque no esté presente sino porque estoy abordando una *condición de posibilidad* previa, de tal modo que, sin ella, desde mi perspectiva, por lo menos, no habría educación ni Pedagogía.

La realidad excede lo cognoscible

Esta tesis es relevante para nuestro análisis, no tanto por sus implicaciones en el conocimiento de la realidad física, sino porque establece un fundamento ontológico indispensable: la afirmación de que la realidad no se agota en lo que podemos conocer. Esta premisa abre la posibilidad

¹ Sé que este adjetivo puede sonar pueril pero hoy inquieta saber que, por ejemplo, el Instituto de zoología de Kunming, en China, inyectó en 2019 un gen humano relacionado con la inteligencia en macacos que mostraron capacidades cognitivas superiores a las de sus pares; que la extensión del examen de embriones con diagnóstico genético preimplantacional se acerca poco a poco a prácticas eugenésicas: en países como Islandia y Dinamarca la población de síndrome de Down es prácticamente inexistente; (Lukács de Perény, 2023, p. 155), o que en 2023 el equipo de científicos liderados por Jacob Hanna del Instituto Weizmann de Ciencias en Israel lograron crear el primer modelo completo de embrión humano sin utilizar óvulos, espermatozoides ni útero (Oldak, Wildschutz, Bondarenko, *et al.*, 2023)

de reconocer los valores como una dimensión objetiva del ser, accesible al conocimiento, pero no reducible a él, cuestión que se abordará con mayor detalle en el siguiente apartado.

En cambio, autores como Gaviria -y con él, quienes conciben el transhumanismo como un paso evolutivo natural- parten de forma acrítica de una premisa de raíz kantiana: que lo real es tal como lo conocemos. Es decir, incurren en una confusión entre epistemología y ontología, al identificar las condiciones de posibilidad del conocimiento humano con la estructura misma del ser. De este modo, el conocimiento empírico, construido desde nuestras facultades cognitivas (naturales o artificiales), es erróneamente tomado como una representación fiel y acabada de la esencia de las cosas. Esto se refleja en afirmaciones como las siguientes: "(l) os juicios morales pueden ser objeto de estudio científico (...). Por eso los juicios morales pueden ser apoyados por las máquinas". Y más adelante:

(...) un transhumano híbrido o simbiótico con una inteligencia artificial no sólo podría tener vida moral, sino que podría ser moralmente más eficaz, ya que mientras mantiene el impulso consciente de perdurabilidad, podría hacer juicios morales más acertados, pues el componente artificial aumentaría sus capacidades de juicio, y por tanto de *valoración de las circunstancias concretas en las que la disyuntiva moral se plantease* (*op. cit.*, p. 11; cursiva añadida).

Esta afirmación es falaz porque aparenta lógica, pero oculta premisas incorrectas. La realidad no se reduce a la capacidad de conocerla, ni siquiera cuando dicha capacidad se ve ampliada por tecnologías artificiales. No existe una correspondencia fija, directa o automática entre el conocimiento (siempre parcial, situado y revisable) de la estructura del mundo y la acción moral. De hecho, ningún comité de ética clínica asumiría que un juicio moral puede ser generado automáticamente mediante el cruce de datos y principios teóricos. Entre los principios morales más claros y la acción concreta que debe realizarse en una situación particular no media un proceso causal ni deductivo. No es una operación que pueda resolverse por algoritmos, porque la acción moral exige deliberación, interpretación, responsabilidad y, sobre todo, *experiencia*, no solo cálculo. Un medio puede ser eficaz para alcanzar un fin, pero esto no garantiza que ese fin tenga valor intrínseco o sea humanamente deseable. La eficacia técnica puede ser condición necesaria para una acción, pero nunca suficiente para justificar su legitimidad ética.

En esta misma línea conviene añadir dos advertencias importantes. La primera tiene que ver con los riesgos de delegar en exceso la toma de decisiones éticas en sistemas automatizados, lo que puede llevarnos a

una progresiva “deshabilitación moral” o incluso a un “empobrecimiento moral humano” que infantilice y dificulte nuestro crecimiento como sujetos responsables (Green, 2018, p. 22). Otra dificultad es lo que se ha denominado “el problema del alineamiento de valores”. Como señala Eberl (2018) si bien es cierto que algunos métodos de biomejora pueden servir para aumentar la capacidad de razonar con prudencia y ayudar a los agentes morales a alinear su voluntad con sus *deseos racionales de orden superior*, la cuestión decisiva queda sin resolver: ¿quién determina esos deseos? ¿Cómo se forman y legitiman?

Por otra parte -y este punto es ontológicamente más relevante- si aceptamos que la realidad es solo aquello que nuestras capacidades, sean naturales o artificiales, pueden transformar o dominar, entonces caeremos en la trampa de pensar que *lo que podemos hacer define lo que somos*. Así, Gaviria concluye erróneamente que lo transformable o ampliable constituye evidencia suficiente de la naturaleza humana y de sus fines. Pero la maleabilidad no agota el conocimiento de la naturaleza humana, del mismo modo que no comprendemos el valor intrínseco de la naturaleza o de los seres vivos simplemente a partir de su utilidad o disponibilidad para nosotros. La verdad no siempre reside en lo que podemos hacer, ni el sentido humano se reduce a la capacidad de intervenir con eficacia. Confundir lo posible con lo legítimo no es una superación de los límites humanos, sino una negación de su profundidad constitutiva.

Lo natural desde una perspectiva teleológica de fines

La noción de «lo natural» puede abordarse desde dos perspectivas clásicas. Por un lado, desde una mirada genetista, centrada en el origen biológico: lo natural como aquello que está determinado por la génesis o el principio vital. En este sentido, es natural el crecimiento de una semilla de roble según la información genética que contiene desde el inicio. Esta perspectiva pone el énfasis en el punto de partida. Por otro lado, cabe una comprensión teleológica de la naturaleza: lo natural es aquello que realiza su fin propio, que despliega su forma plena. En este segundo sentido, un roble es plenamente natural no sólo porque nace como tal, sino porque alcanza su forma acabada, con ramas, hojas y frutos, realizando así su *potencia*. Es natural no sólo por su origen, sino por su *plenitud*.

Pues bien, el transhumanismo se enreda con estas distinciones al eliminar, como hace Gaviria, la perspectiva teleológica de la naturaleza humana. Al negarla nos quedamos sin la posibilidad de responder adecuadamente a ciertos interrogantes fundamentales. En efecto, si no existe

una finalidad propia en los seres humanos, ¿con qué criterio podemos afirmar que hacer X es mejor que hacer Y?

Si la naturaleza humana carece de fines propios, si eliminamos la posibilidad de hallar un orden normativo en perspectiva teleológica –‘llega a ser el que eres’-, es radicalmente imposible y aún inútil discutir qué propuestas del transhumanismo son aceptables y cuáles no. La consecuencia no se hace esperar: una “(...) naturaleza humana naturalizada hasta el extremo es al mismo tiempo una naturaleza humana disponible, apta para su transformación técnica y presunta mejora» (Marcos y Pérez Marcos, 2018, p. 16). Si no hay un fin constitutivo de la naturaleza humana al que dirigirse tampoco hay un fin constitutivo de la naturaleza humana que nos aleje. Si no hay una forma humana que cumplir, tampoco hay una deformación que evitar. Por tanto, todo es posible y todo está permitido.

La fuerza objetiva de una norma moral no se fundamenta en su utilidad biológica

Uno de los supuestos centrales del pensamiento transhumanista es la reducción del ser humano a una entidad exclusivamente natural o físico-biológica, comparable a cualquier otro ser vivo, cuyo deseo principal es la perdurabilidad, entendida como supervivencia y proyección futura. Desde esta perspectiva, el transhumanismo no presenta problema alguno, sino que se muestra como una extensión lógica del deseo de conservar y amplificar nuestras capacidades adaptativas. Si, además, se sostiene que los valores surgen simplemente del deseo consciente de un organismo que interactúa con su entorno -como afirma Gaviria cuando señala que: “(...) los valores surgen (...) en el momento en que un ser vivo consciente enfrenta al mundo con un deseo propio” (*op. cit.*, p. 10)- entonces, los valores no remiten a una realidad objetiva, sino que se reducen a expresiones funcionales del deseo biológico. En ese marco, el transhumanismo aparece no solo como compatible con la ética, sino como su *potenciador*.

El problema filosófico decisivo surge precisamente aquí. Si los valores tienen su origen exclusivo en el deseo -individual o colectivamente consensuado-, ¿cómo explicar su fuerza normativa?, ¿en virtud de qué podemos hablar de verdad o falsedad moral, de acierto o error ético? Si los valores no son más que proyecciones de preferencias subjetivas o construcciones sociales contingentes, resulta difícil justificar por qué consideramos moralmente significativos conceptos como el error moral,

la mentira, la tergiversación o la injusticia. La propia idea de una búsqueda de mayor precisión moral presupone la existencia de una realidad que trasciende nuestras perspectivas particulares. Solo si reconocemos un horizonte objetivo es posible hablar con sentido de acierto y error, de verdad o falsedad, también en el ámbito moral. Como escribe Thomas Nagel:

El intento serio de identificar lo que es subjetivo y particular de nuestra perspectiva, o relativo y propio a la comunidad, lleva inevitablemente a lo objetivo y universal. Esto vale tanto para la ética como para la ciencia o la lógica (Nagel, 2000, p. 28).

Esto implica que la realidad no es solo físico-química, sino también valorativa, estimativa: está atravesada por significados y fines que no se reducen a su utilidad biológica. ¿El respeto al medio ambiente o a los animales responde únicamente a un cambio de perspectiva cultural, o más bien a un reconocimiento de valores objetivos que estaban velados y que ahora se *desvelan*? ¿El derecho al voto de las mujeres responde solo a una concesión de los hombres o hay *algo más*? Si múltiples culturas, a lo largo del tiempo, convergen en ciertos principios morales, ¿no se debe a la percepción compartida de una misma realidad moral, vinculada a la naturaleza humana? (Lewis, 1990, pp. 81-96).

Esta necesidad de referencia ontológica no es exclusiva de posiciones iusnaturalistas clásicas. Incluso autores poco afines a ese enfoque reconocen que la comprensión moral remite a algo más que a meros consensos. Así lo expresa Savater:

Esta capacidad humana de alcanzar la comprensión de lo esencial es inseparable de la pretensión de un derecho natural (...). *Y poder conocer esa verdad fundamental supone también que tal verdad existe, que hay una verdad natural o, si se prefiere, una verdadera naturaleza, distinta radicalmente de las opiniones que los hombres tienen en común* (Savater, 1988, p. 167; cursiva añadida).

Hay que volver, pues, a reconocer la naturaleza de las cosas y, en particular, las valoraciones éticas que se desvelan desde la misma naturaleza humana. Sin ontología no hay ética. Sin ontología los asuntos humanos, todos, se relativizan. Como ha argumentado el profesor Ibáñez-Martín: "(...) si no hubiera verdad alguna sobre el ser humano podríamos hacer con él lo que nos diera la gana, en la medida en que fuéramos poderosos" (Ibáñez-Martín, 2017, p. 21). Y es que

(e)n último término, el deber es relativo al ser. Y como el deber atañe al hombre, entonces es relativo al ser del hombre, a la naturaleza humana. (...) ¿Cómo puede ser necesario para un ser algo que con ese

ser no tiene nada que ver? Es un dislate. *Lo que es bueno o malo para un ser depende de lo que ese ser es, no 'parece', o 'quiero yo que sea'...* (Millán-Puelles, 2014, p. 516; cursiva añadida).

Por eso la fuerza objetiva de una norma moral no se fundamenta en su utilidad biológica para la supervivencia del individuo o del grupo, sino en una propiedad intrínseca del objeto mismo al que dicha norma se refiere. Esta distinción se manifiesta con especial claridad en situaciones límite. Pensemos en el gesto de un policía que, ante un secuestro, se ofrece a intercambiarse por uno de los rehenes porque este tiene hijos pequeños. Desde una lógica estrictamente adaptativa o algorítmica, resulta difícil justificar moralmente tal acción: ¿qué tipo de máquina la recomendaría?, ¿qué cálculo de maximización de la supervivencia la haría razonable? Sin embargo, desde el punto de vista humano, ese juicio moral resulta plenamente inteligible. El policía actúa movido por el reconocimiento de un valor superior, por la convicción -vvida y encarnada- de que existe algo más importante que la propia supervivencia. Su decisión no es el resultado de un cálculo utilitarista, sino la respuesta a un orden de sentido inscrito en la realidad misma, que se le impone como exigencia moral objetiva y que le permite saber, y también sentir, lo que debe hacer.

La educación no es una mejora programada

Es cierto que educarse implica perfeccionarse, pero no de forma inmediata ni mecánica. La educación es, en parte, actualización de las potencialidades inscritas en nuestras capacidades, pero lo es mediante un proceso indirecto, pausado, lento y, sobre todo, profundamente mediado: por las actividades culturales, por las relaciones con los demás (Bárcena, 2025) y por el esfuerzo personal (Woo, 2024). Se trata de un perfeccionamiento que requiere tiempo, trabajo y participación del sujeto. Lejos de tratarse de una optimización externa o de un resultado garantizado, la educación es una experiencia interna de transformación: no lineal, no plenamente controlable ni asegurada en sus efectos, pero intencional y orientada a fines que pueden ser anticipados de manera razonable, aunque nunca producidos de forma mecánica².

² Agradezco a uno de los revisores del artículo la sugerencia de interpretar la educación no solo como transformación, sino también -en un sentido más fuerte- como transfiguración, entendida como el proceso mediante el cual la persona alcanza y revela su plenitud o aquello a lo que está llamada a ser. Sin desconocer el interés y la fecundidad pedagógica de esta perspectiva (Mortari, Valbusa, y Bombieri, 2026; Gil Cantero, 2025; García Gutiérrez, 2023; López Quintás, 2014), en este trabajo se ha optado deliberadamente por el concepto de transformación, con el fin de subrayar el carácter interno, no técnico y no garantizable del cambio educativo, así como su apertura constitutiva, sin introducir de entrada una teleología plena que podría desplazar el eje argumentativo del análisis propuesto.

Por ello, la lógica educativa se aleja radicalmente de la lógica transhumanista. Esta última concibe la mejora como un proceso rápido, técnico, sin mediaciones culturales ni vínculos sociales, sin esfuerzo del sujeto, sin narratividad ni trayectoria personal. Una mejora, en suma, concebida como producto final, no como camino.

En la educación lo crucial no es solamente el resultado alcanzado, sino el proceso mismo mediante el cual se alcanza. Por ejemplo, no es lo mismo llegar a la cima de una montaña tras una dura escalada que ser colocado directamente allí. Del mismo modo, no es igual mejorar un texto tras una lectura crítica exigente que aceptar sin más las correcciones automáticas de un programa de inteligencia artificial.

En definitiva, la educación no se reduce a una simple mejora de funciones o capacidades como resultado programado y garantizado, sino que es un proceso, un camino, una intención sostenida en el tiempo con esfuerzo personal. “Y esto es lo verdaderamente grandioso de la educación: su poder autoestructurante no solo como fin perfectivo al que llegamos sino como medio de nuestro propio desarrollo humano” (Gil Cantero, 2022, p. 23). Este desarrollo se singulariza, además, en una forma de ser particular, en una historia biográfica, narrativa, propia, sobre nuestro *ser educando*, a través del surco que marca la entraña antropológica del propio esfuerzo.

La educación es una mejora con mundo

Hay un aspecto particularmente revelador en el modo en que la educación promueve la mejora humana, y que la diferencia radicalmente del enfoque transhumanista. Mientras que éste, como proyecto técnico y cerrado sobre sí mismo, permite establecer una correspondencia anticipada entre intervención y resultado -cuánta memoria va a ganar, qué nivel de coeficiente de inteligencia va a alcanzar, cuánta vista va a ganar, qué nivel de atención va a mejorar, etc.-, la educación no funciona ni puede funcionar así. No hay ni puede haber una garantía técnica de resultados, ni una linealidad causa-efecto que determine su éxito. Y no solo porque no podamos prever los resultados con precisión, sino -más profundamente- porque no queremos reducir la mejora humana a una cuestión de eficiencia funcional.

El transhumanismo concibe la mejora como una potenciación abstracta de funciones, capacitante, desligada de cualquier contenido o contexto. En cambio, la educación es siempre un *proceso de mejora inserto en el mundo*: acontece en un entramado de mediaciones culturales, vínculos personales y experiencias vividas. Podríamos decir, en breve, que

la tecnología transhumanista se desarrolla sin mundo mientras que la educación se desarrolla con mundo, en el mundo y para el mundo. *Con mundo* porque es a través de mediaciones culturales de contenido; *en el mundo* porque se enraíza en las experiencias concretas, en las relaciones humanas (Vila y Álvarez, 2025) y en los contextos sociales; y *para el mundo* porque persigue la realidad compartida, no solo potenciar al individuo.³

Gaviria sostiene -en sintonía con el pensamiento transhumanista- que el ideal de mejora del ser humano “(...) no (es) muy distinto de la naturaleza optimizante de la educación. Lo que en un caso se consigue por medios tecnológicos en el otro se hace por medio de elementos simbólicos, contenidos culturales y relaciones personales.” (*op. cit.*, p. 12). Pero este paralelismo pierde de vista lo esencial: no es posible equiparar una mejora simbólica, encarnada, relacional, con una optimización funcional, abstracta y técnica. Por ejemplo: ¿qué efecto tendrían las intervenciones transhumanistas sobre un niño salvaje, privado del mundo educativo? Ninguno significativo. El «buen salvaje» no puede realizar su humanidad sin ese mundo compartido que solo la educación proporciona.

Por tanto, transhumanismo y educación no son simplemente medios distintos para un mismo fin, sino procesos esencialmente diferentes, no intercambiables. La educación responde a una estructura antropológica que el transhumanismo no puede replicar, porque la naturaleza humana tiene exigencias propias —culturales, simbólicas, relacionales— que no pueden ser suplidas solo por mejoras técnicas. En consecuencia, no es correcto afirmar que la educación sea una forma de transhumanismo.

No es lo mismo una actividad productiva que una acción perfectiva

Conviene ahora plantear una cuestión clave: si, como hemos argumentado, el transhumanismo y la educación no son procesos de mejora intercambiables, ¿por qué rechazar que la educación se inicie desde un “punto de partida” tecnológicamente mejorado? Si existen medios artificiales para potenciar la memoria, la atención, la inteligencia, la resistencia física, o la agudeza sensorial, ¿por qué no emplearlos antes de educar?

Algunos transhumanistas con sensibilidad humanista intentan así una conciliación temporal: primero vendría la mejora tecnológica de ciertas

³ Argumento inspirado en Gomá (2023, p. 25 y ss.) que él utiliza para desarrollar su idea de una filosofía mundana.

capacidades, y luego la educación como proceso integral. En este sentido, Gaviria sostiene que “la mejora artificial transhumanista no es más que la definición de un nuevo punto de partida, uno que sitúa el comienzo de la educación en un estadio superior” (*op. cit.*, p. 17). Este planteamiento tiene un acierto importante pues supone reconocer la necesidad ineludible de la educación, esto es, que ningún estadio -artificial o no- superior de capacidades cognitivas, sensoriales o físicas nos garantiza un sujeto plenamente educado.

Ahora bien, siendo esto verdad, me gustaría mostrar que las mejoras artificiales del transhumanismo no son un punto de partida *neutro* desde la perspectiva de la educación. Como advierte Žižek:

(...) ¿y si nuestra vida interior, incluidos nuestros más destacables logros espirituales, estuviera anclada en nuestra finita existencia física y sus limitaciones, de manera que con la transición a la Singularidad nos viéramos privados de los aspectos básicos de nuestra vida interior? (2023, p. 157).

En efecto, tales intervenciones pueden afectar negativamente a la disposición del sujeto a implicarse en su propio desarrollo: ¿Por qué habría de esforzarme en desarrollar mi capacidad de concentración si ya presto atención sin dificultad gracias a una intervención externa? ¿Qué sentido tiene intentar superarme físicamente si mis capacidades han sido modificadas para rendir sin fatiga? De hecho, esta lógica ya opera con la inteligencia artificial: el esfuerzo que yo dejo de hacer por mí mismo para redactar un resumen, entender un párrafo de un libro o elaborar una ruta de senderismo es *una oportunidad perdida de educación*, de mejora personal, que ya no se queda en mí, que ya no me forma o configura, aunque, eso sí, el resultado del resumen, de la comprensión del párrafo y de la ruta campestre sea mejor de lo que yo hubiese hecho. En términos aristotélicos he realizado una actividad productiva pero no una acción perfectiva. El resultado ha mejorado, sí, pero desde fuera de mí, sin configurarme, sin surco, sin haberlo vivido y experimentado interiormente.

Autonomía: ¿criterio pedagógico suficiente?

Para los defensores del transhumanismo una vez aceptada la secuencia temporal propuesta -primero las intervenciones transhumanistas, después la educación-, lo único que cabe debatir ya es qué tipos de mejoras son aceptables y cuáles no. Con este mismo planteamiento concluye Gaviria su artículo, por cierto, de forma brillante y en plena sintonía

con el texto de Bostrom (2005, p. 9 y ss.). No vamos a discutir cada uno de los criterios propuestos que, como puede comprobarse, se sostienen razonablemente sobre la base de los logros éticos asociados al respeto de los derechos humanos. Lo que sí vamos a cuestionar es el principio rector -o metacriterio- que articula y justifica el resto: la autonomía del sujeto como medida educativa última. Gaviria lo expresa así:

(...) es perfectamente válido establecer algunos criterios para determinar qué intervenciones biotécnicas son aceptables y cuáles no.

Para mí, el criterio principal tiene que ver con la autonomía de los sujetos. Se trata de distinguir entre lo que aumenta nuestra autonomía y lo que la disminuye. El aumento de nuestras capacidades incrementa nuestra autonomía. La implantación de contenidos inmunes a la crítica o conductas automáticas incontrolables disminuye nuestra autonomía (*op. cit.*, p. 20).

Este planteamiento, sin embargo, incurre en un error frecuente cuando se reflexiona sobre la relación entre educación y transhumanismo desde una perspectiva exclusivamente científica y naturalista: suponer que la autonomía es una realidad mensurable, ampliable o reducible de forma lineal y técnica, como si fuera una función independiente del sujeto y de su inserción vital en el mundo.

Parta empezar no está nada claro en qué consiste “aumentar nuestra autonomía”. ¿Quién es más autónomo en una sociedad hiperconsumista: quien dispone de recursos ilimitados para consumir o quien opta por una vida austera? ¿Quién es más autónomo frente a la tradición religiosa: quien conserva su fe con convicción o quien la rechaza en nombre de la autonomía individual? ¿Quién es más autónomo: el hijo que sigue siendo socio del Atlético de Madrid al igual que sus padres o el que cambia de equipo y se hace del Real Madrid?

Por otra parte, tampoco está claro qué tipo de capacidades *concretas* incrementan nuestra autonomía. ¿La inteligencia, la voluntad, las emociones, la atención, la libertad...? ¿La capacidad de enamorarse aumenta nuestra autonomía? ¿La capacidad de criar un hijo incrementa nuestra autonomía? Para los liberales, por ejemplo, lo que nos hace más libres-autónomos- es la propiedad, para otros, nos esclaviza. Más aún: No hay una línea ascendente de incremento de autonomía en el ser humano que garantice en paralelo su felicidad.

Pasando a una reflexión más de fondo, conviene señalar que la “verdadera educación” -Gaviria *dixit*- no consiste en maximizar la autonomía como si fuera un fin en sí mismo o un indicador acumulativo de mejora. Esta propuesta no solo incurre en una reducción funcionalista

de la educación, sino que también confunde, una vez más, las partes con el todo. Educar no consiste solo en ampliar las posibilidades de acción del individuo, sino en *formar su disposición interior a sustentar libremente su obrar en bienes superiores, desde un juicio prudente y responsable*. Desde esta perspectiva, la autonomía no debe entenderse como una capacidad técnica que se adquiere, sino como una virtud ética que se cultiva. No se gana, se forma; no se mide por la cantidad de opciones disponibles ni por la eficacia en la ejecución, sino por la capacidad de discernir lo valioso y obrar en consecuencia. El sujeto verdaderamente educado no es quien más autonomía posee, sino quien sabe usar su autonomía de forma humanizadora, al servicio del bien y de los otros. Por eso, los grandes referentes morales de la humanidad -héroes, santos, sabios- no son admirados por la magnitud de su autonomía, sino por el uso noble, justo y compasivo que hicieron de ella.

La tarea de la Pedagogía frente al transhumanismo

El transhumanismo no es ya una hipótesis futurista, sino un proceso cultural y tecnológico en marcha, cada vez más asumido por la sociedad como parte del horizonte de lo posible. La oferta de mejoras artificiales -algunas ya disponibles- se irá integrando con creciente naturalidad en nuestras vidas. Aunque se establezcan límites éticos o jurídicos desde instituciones internacionales, es previsible que surjan iniciativas paralelas que los desborden, como los biohackers o “biólogos de garaje”, que operan fuera del sistema.

Ante este panorama, no parece que la Pedagogía deba implicarse directamente en el análisis técnico, jurídico o sociológico de cada nueva mejora artificial. Esas tareas, aunque necesarias, pertenecen a otros campos. Lo que sí le corresponde a una Pedagogía ambiciosa es custodiar un legado y mantener una convicción. El *legado* de salvaguardar la riqueza humanizadora de la educación: ayudar a alcanzar una madurez que resalte los deseos racionales de orden superior con respecto a lo que enriquece realmente una vida humana. La *convicción* de que la educación no se reduce a la mejora funcional del ser humano sino a la configuración de una biografía singular, propia, que pretende ser mejor cada día con el esfuerzo personal, más preocupada por el cambio interior como tarea permanente que como mejora artificial.

En tiempos de fascinación por lo posible, debemos recordar lo valioso. No nos toca competir con el ritmo de la innovación biotecnológica, sino preservar aquellos valores formativos que, aunque lleguen a considerarse marginales por antiguos, constituyen el núcleo humanizador de

toda experiencia educativa: el esfuerzo, la aceptación de sí mismo, la prudencia en el uso de la libertad, la relación con los otros en un mundo compartido, la conciencia de nuestra vulnerabilidad y dependencia.

Por supuesto, soy consciente de que muchos preferirán “(...) conformarse con el sentimiento de victoria sin haber puesto un pie en la cancha de baloncesto o sentirse amado sin un amante” (Güell, Echarte-Alonso y Murillo, 2019, p. 207), pero lo que la Pedagogía debe decirles es claro: ni es lo mismo, ni eres el mismo al elegir una cosa o la otra.

Referencias bibliográficas

- Arana, J. (Dir.). (2024). *El futuro de la identidad humana a debate. Protagonistas de la polémica sobre el transhumanismo*. Tecnos.
- Bárcena, F. (2025). Una filosofía de la relación educativa: mediación existencial, transmisión y testimonio [A Philosophy of the Educational Relationship: Existential Mediation, Transmission and Testimony]. *Teoría de la Educación. Revista Interuniversitaria*, 37(2), 1-18. <https://doi.org/10.14201/teri.32244>
- Bostrom, N. (2005). Transhumanist values. *Review of Contemporary Philosophy*, 4(1-2), 87-101. <https://nickbostrom.com/ethics/values.pdf>
- Bostrom, N. (2024). *Deep Utopia. Life and Meaning in a Solved World*. Ideapress.
- Eberl, J. T. (2018). Can Prudence Be Enhanced? *Journal of Medicine and Philosophy*, 43(5), 506-526. <https://doi.org/10.1093/jmp/jhy021>
- Fukuyama, F. (2009, 23 de octubre). Transhumanism, *Foreign Policy*. <https://foreignpolicy.com/2009/10/23/transhumanism/>
- García Gutiérrez, J. (2023). Pensar la antropología educativa desde la transfiguración: Apuntes críticos al transhumanismo desde una pedagogía de la vida del espíritu. *Temps d'Educació*, 65, 93-111. <https://doi.org/10.1344/TempsEducacio2024.65.7>
- Gaviria, J. L. (2024). ¿Transhumanismo ‘contra’ educación? [Transhumanism ‘against’ Education?]. *Teoría de la Educación. Revista Interuniversitaria*, 36(2), 1-23. <https://doi.org/10.14201/teri.31762>
- Gil Cantero, F. (2022). La Pedagogía ante el desfase prometeico del transhumanismo. *Revista de Educación*, 396, abril-junio, 11-33. <https://doi.org/10.4438/1988-592X-RE-2022-396-528>
- Gil Cantero, F. (2025). Controversias en torno a la educación transformadora. En C. Naval y A. Bernal (Eds.), *La educación transformadora, luces y sombras* (pp. 31-64). Eunsa.

- Goering, S., Klein, E., Specker Sullivan, L. *et al.* (2021) Recommendations for Responsible Development and Application of Neurotechnologies. *Neuroethics*, 14, 365-386. <https://doi.org/10.1007/s12152-021-09468-6>
- Gomá, J. (2023). *Universal concreto. Método, ontología, pragmática y poética de la ejemplaridad*. Taurus.
- Green, B. P. (2018). Ethical Reflections on Artificial Intelligence. *Scientia et Fides* 6(2), 9-31. <http://doi.org/10.12775/SetF.2018.015>
- Güell, F., Echarte-Alonso, L. E. & Murillo, J. I. (2019). Objetividad y bio-mejora humana. La paradoja de lo natural. *Scientia et Fides*, 7(2), 195-214. <https://doi.org/10.12775/SetF.2019.023>
- Ibáñez-Martín, J. A. (2017). *Horizontes por los educadores. Las profesiones educativas y la promoción de la plenitud humana*. Dykinson.
- Kurzweil, R. (2025). *La singularidad está más cerca: cuando nos fusionamos con la IA*. Deusto.
- Lacalle, M. (Dir.).(2022). *Transhumanismo: ¿homo sapiens o ciborg?* Universidad Francisco de Vitoria.
- Lewis, C. S. (1990). *La abolición del hombre*. Encuentro.
- López Quintás, A. (2014). Transfiguraciones que deciden el proceso de crecimiento personal. En Autor, *La ética o es transfiguración o no es nada* (357-396). BAC
- Lukács de Perény, M. (2023). *Neo entes. Tecnología y cambio antropológico en el siglo XXI*. Homo Legens.
- Marcos, A. y Pérez Marcos, M. (2018). *Meditación de la naturaleza humana*. BAC.
- Millán-Puelles, A. (2014). *Obras completas*. Vol. V. *Economía y libertad* (1974). Rialp.
- Mortari, L., Valbusa, F., y Bombieri, R. (2026). Aprendizaje socioemocional y florecimiento humano. *Teoría de la Educación. Revista Interuniversitaria*, 38(1), 123-145. <https://doi.org/10.14201/teri.32149>
- Nagel, Th. (2000). *La última palabra*. Gedisa.
- Oldak, B., Wildschutz, E., Bondarenko, V. et al. (2023). Complete human day 14 post-implantation embryo models from naive ES cells. *Nature* 622, 562-573. <https://doi.org/10.1038/s41586-023-06604-5>
- Savater, F. (1988). *Ética como amor propio*. Mondadori.
- Steiner, G. (2022). *¿Tiene futuro la verdad? El hombre al norte del mañana*. Almuzara.
- Vila, E. S. y Álvarez, V. E. (2025). La relación educativa en un mundo digitalizado: paradojas de la hiperconexión [The Educational Relationship in a Digitalized World: Paradoxes of Hyperconnection]. *Teoría de*

- la Educación. Revista Interuniversitaria*, 37(2), 19-36. <https://doi.org/10.14201/teri.32283>
- Woo, J. G. (2024). Education amid the deluge of enhancement discourses. *Educational Philosophy and Theory*, 56(12), 1198–1209. <https://doi.org/10.1080/00131857.2024.2364036>
- Žižek, S. (2023). *Hegel y el cerebro conectado*. Paidós.

Análisis de funcionamiento diferencial en evaluaciones a gran escala

Differential item functioning analysis in large-scale assessments

<https://doi.org/10.4438/1988-592X-RE-2025-412-733>

Laura Ortega Torres

<https://orcid.org/0009-0009-5776-0581>

Centro Nacional de Evaluación para la Educación Superior, CENEVAL (México)

César Antonio Chávez Álvarez

<https://orcid.org/0009-0000-4197-8281>

Centro Nacional de Evaluación para la Educación Superior, CENEVAL (México)

Resumen

Los análisis de funcionamiento diferencial de los ítems (DIF, por sus siglas en inglés), son esenciales para sustentar la calidad y equidad de una evaluación, ya que señalan ítems que pudieran estar sesgados. Se han desarrollado diferentes métodos para realizar estos análisis y se ha comparado su eficacia por medio de la manipulación de diversas variables, incluido el número de participantes. Generalmente se estudian grupos con hasta algunos miles de integrantes, sin embargo, existen evaluaciones a gran escala que se aplican a un número considerablemente mayor de personas. Este estudio exploró la eficacia de cuatro métodos tradicionales para analizar DIF en muestras de personas con un tamaño similar al de los grupos de evaluados que presentan los Exámenes Nacionales de Ingreso (EXANI) en una misma aplicación en México y que pudieran usarse para analizar otras pruebas de ingreso que se apliquen a nivel nacional o aplicaciones completas de Evaluaciones Internacionales a Gran Escala (ILSA, por sus siglas en inglés). Se realizaron simulaciones en las que se generaron muestras de tamaños similares a las de las aplicaciones de los EXANI y se manipularon los valores de los parámetros de dificultad del 10% o el 20% de los ítems de una de sus áreas de contenido para que presentaran DIF Uniforme moderado y alto. Se verificó la eficacia de los métodos de Mantel-Haenszel, Regresión Logística, Lord y Raju, comparando el porcentaje de ítems detectados con su grado de DIF de forma correcta (ítems manipulados), así como el porcentaje de falsas detec-

ciones en los reactivos sin DIF. Los métodos de Mantel-Haenszel, Lord y Raju fueron los más eficaces para detectar funcionamiento diferencial uniforme en las muestras simuladas, por lo que se confirma su posible implementación en evaluaciones a gran escala con ítems dicotómicos con muestras de hasta 200,000 evaluados.

Palabras clave: Análisis de Funcionamiento Diferencial Uniforme, Ítems Dicotómicos, Evaluaciones a Gran Escala, Método Mantel-Haenszel, Método de Regresión Logística, Método de Lord, Método de Raju

Abstract

Differential Item Functioning (DIF) analyses are essential to support the quality and fairness of an assessment, as they identify items that may be biased. Various methods have been developed to perform these analyses, and their effectiveness has been compared by manipulating different variables, including the number of examinees. Typically, studies focus on groups with some thousands of participants; however large-scale assessments are administered to a significantly larger number of individuals. This study explored the effectiveness of four traditional methods for analyzing DIF in samples similar in size to the groups of examinees taking the Entrance exams (EXANI by its acronym in Spanish) in a single administration in Mexico and that could be applied in other entrance tests administered nationwide or complete administrations of International Large-Scale Assessments (ILSA). Simulations were conducted, in which samples of sizes comparable to those in the EXANI's administrations were generated, and the difficulty parameter values of 10% or 20% of the items of one of their content areas were manipulated to exhibit moderate and high Uniform DIF. The effectiveness of the Mantel-Haenszel, Logistic regression, Lord and Raju methods was verified by comparing the percentage of items correctly detected with their degree of DIF (manipulated items) as well as the percentage of false detections in non-DIF items. The Mantel-Haenszel, Lord and Raju methods were the most effective in detecting uniform DIF in the simulated samples, confirming their potential implementation in large-scale assessments with dichotomous items for samples of up to 200,000 examinees.

Keywords: Uniform differential item functioning, Dichotomic items, Large-scale assessments, Mantel-Haenszel Method, Logistic Regression Method, Lord Method, Raju Method

Introducción

Una propiedad fundamental de cualquier evaluación es que no favorezca a alguna persona o grupo con características específicas. Los análisis de funcionamiento diferencial de los ítems (DIF, por sus siglas en inglés:

Differential Item Functioning) ayudan a determinar que las mediciones no estén sesgadas y reflejan si se está evaluando el mismo constructo en todos los examinados; por lo que son fundamentales en el desarrollo de pruebas y sus procesos de validación (Walker, 2011). Específicamente, el estándar 4.10 de los Standards for Educational and Psychological Testing sugiere la implementación de estos análisis para determinar parte de las propiedades psicométricas de los ítems (AERA et al., 2014).

El DIF se refiere a cuando dos grupos de personas con la misma habilidad que se pretende medir tienen distinta probabilidad de éxito al responder un reactivo, es decir, después de controlar su habilidad se observa que esa probabilidad difiere entre ambos grupos. Un reactivo puede comportarse estadísticamente diferente al conjunto de ítems con los que fue ensamblado sin presentar un sesgo en el contenido (Wu et al., 2016), por lo tanto, el estar marcado con DIF es una condición necesaria, pero no suficiente para determinar sesgo. Para ello, la causa del funcionamiento diferencial debe atribuirse a un factor irrelevante al constructo y es necesario realizar análisis de contenido posteriores para poder incluir tanto el soporte estadístico como la interpretación cualitativa que indiquen que el DIF se debe a una característica del ítem, irrelevante al objetivo de la evaluación, dictaminarlo como sesgado y decidir si se envía a revisión o se elimina del banco (Sireci & Rios, 2013; Zwick, 2012).

Los análisis de DIF buscan identificar ítems que son particularmente difíciles para miembros de grupos específicos, comparando a participantes de diferentes géneros, culturas, razas o que hablan idiomas distintos. Estas variables siguen siendo centrales en los análisis de DIF en distintos exámenes o instrumentos en particular o como parte de estudios longitudinales alrededor del mundo (Cassiani-Miranda et al., 2025; Shamsaldeen et al., 2024; Suk & Han, 2024; Villalonga-Olives et al., 2024) para descartar que las diferencias observadas en estos grupos se deben a características intrínsecas de los ítems.

Existen varios métodos para realizar estos análisis, los cuales tienen sustento en estudios de simulación como parte de la investigación para su uso. En estos estudios se manipulan distintas variables debido a que algunas pueden influir en la detección del funcionamiento diferencial. Estas manipulaciones pueden ser procedimientos en común para los distintos métodos de DIF, como el de purificación, que se refiere a la eliminación iterativa de reactivos con funcionamiento diferencial para mejorar la precisión de los resultados (French & Maller, 2007; Özdemir, 2015; Wang et al., 2009). También se pueden basar en condiciones generales de las evaluaciones para analizar sus escenarios más comunes e incluyen el número de ítems que forman una prueba o área, el rango de valores que

toman los parámetros psicométricos de los ítems (como su dificultad y discriminación), el grado de DIF, el número de participantes por grupo y sus distribuciones de puntajes (Fidalgo et al., 2000; Gierl et al., 2004; Liu et al., 2019; Rogers & Swaminathan, 1993; Wang et al., 2009). Otras manipulaciones tienen fundamento empírico, por ejemplo, se usa un porcentaje del 10% de ítems con DIF, ya que es el porcentaje encontrado en pruebas estandarizadas de logro o relacionado con la raza y el género; o uno del 20%, que es el relacionado a oportunidades de aprendizaje diferentes (French & Maller, 2007; Narayanan & Swaminathan, 1994).

Sin embargo, la manipulación de la cantidad de participantes en estudios de simulación previos no es generalizable a pruebas a gran escala, como las usadas en las Evaluaciones Internacionales a Gran Escala (ILSA, por sus siglas en inglés). Los estudios de DIF que manipulan el tamaño de los grupos a contrastar reportan cantidades entre 100 y 2,000 participantes (French & Maller, 2007; Gierl et al., 2004; Halpin, 2024; Narayanan & Swaminathan, 1994; Park et al., 2021; Wang et al., 2009). Sin embargo, instrumentos que forman parte de las ILSAs, son contestados por un mayor número de sustentantes, por ejemplo, en el Programa para la Evaluación Internacional de los Estudiantes (PISA) pueden participar casi 700,000 estudiantes de 81 países (OECD, 2023), alrededor de 300,000 estudiantes de 50 países por grado en el Estudio Internacional de Tendencias en Matemáticas y Ciencias (TIMSS) (von Davier et al., 2024), en el Estudio Internacional de Progreso en Comprensión Lectora (PIRLS) casi 400,000 estudiantes de más de 50 países (Fishbein et al., 2024), más de 80,000 estudiantes de 24 países en el Estudio Internacional sobre Educación Cívica y Ciudadana (ICCS) (Schulz et al., 2025) y en el Estudio Internacional sobre Competencia Digital (ICILS), cerca de 133,000 estudiantes de 34 países (Fraillon, 2024).

Alrededor del mundo también existen pruebas aplicadas como herramientas en procesos de ingreso a un nivel educativo de un país (Velázquez Ramírez, 2022) y estas pueden ser aplicadas a un gran número de personas. En Madrid, España en 2024, la Evaluación para el Acceso a la Universidad (EvAU) fue presentada por poco más de 30,000 aspirantes (Universidad Autónoma de Madrid, 2024). En México, el Centro Nacional de Evaluación para la Educación Superior elabora los Exámenes Nacionales de Ingreso (EXANI), que son usados para el ingreso a los niveles de Educación Media Superior y Superior y respondidos por cientos de miles de aspirantes.

Los resultados de todos estos instrumentos tienen un alto impacto a nivel nacional o personal, por lo que es indispensable garantizar que sus ítems estén libres de sesgo. En estudios con algunas ILSAs se han

usado métodos de DIF tradicionales aunque con muestras conformadas por entre 500 y 7,000 estudiantes aproximadamente (Abulela & Rios, 2022; Demirus & Pektaş, 2022; Kucam & Gülleroğlu, 2023; Liu & Bradley, 2021; Özdemir, 2015; Uyar, 2020), pero no a poblaciones mayores. Con tamaños de muestras similares también se han comparado modalidades de aplicación del ingreso a la Educación Superior con el EXANI-II a través de análisis de DIF (Vidal González et al., 2022). Sin embargo, para comparar grupos con un mayor número de participantes, es necesario probar que estos métodos son robustos en estas condiciones, ya que ni los estudios de simulación ni los que usan datos empíricos de exámenes cuentan con decenas o cientos de miles de observaciones. El presente estudio aborda esa problemática llevando a cabo simulaciones con tamaños de muestra y otras condiciones similares a las de las evaluaciones de los EXANI, probando la efectividad de cuatro métodos tradicionales. Para realizar cualquier análisis de DIF se requiere definir una variable de comparación que usualmente son variables demográficas de los participantes. Es posible realizar comparaciones entre varios grupos, sin embargo, generalmente se compara un par solamente. El siguiente paso es identificar a los grupos a comparar, donde a uno se le denomina de referencia y a otro focal. El primero es el estándar contra el cual el focal es comparado y tiende a ser un grupo mayoritario; mientras que el focal es un grupo de especial interés y generalmente comprende a la minoría de los examinados (Zumbo, 2007). Finalmente, la variable condicional para ligar a los evaluados usualmente es el puntaje en una prueba. Si el examen está compuesto por áreas, es preferible utilizar sus puntajes por separado (Sireci & Rios, 2013).

Existen dos tipos de DIF de acuerdo con la existencia de una interacción entre el nivel de habilidad y el grupo de pertenencia y ambos derivan de rasgos adicionales a la habilidad principal medida por el ítem, por ejemplo, habilidades verbales en un reactivo de matemáticas (Mellenbergh, 1982). Así los integrantes de dos grupos con la misma habilidad principal pueden diferir en el rasgo secundario alterando su ejecución en el ítem. Si esa diferencia es constante en la habilidad principal, entonces se obtiene DIF Uniforme; pero si depende de la habilidad principal, entonces se obtiene DIF No Uniforme. Al no haber una interacción entre el nivel de habilidad y el grupo de pertenencia en el DIF Uniforme, un ítem es más fácil para todos los miembros de un grupo que para los integrantes del otro, dando una ventaja constante al primero (Walker, 2011). Este DIF ocurre con mayor frecuencia en pruebas estandarizadas (Narayanan & Swaminathan, 1996) y se ha sugerido que es apropiado enfocarse en detectarlo, especialmente cuando las pruebas se conforman por ítems dicotómicos,

debido a que el No Uniforme requiere de interacciones específicas que son más comunes en los ítems politómicos (Jodoin & Gierl, 2001).

Existen varios métodos para analizar DIF (en Magis y colaboradores, 2010, y Ohiri y colaboradores, 2024; se encuentran revisiones de varios de ellos), sin embargo, existen cuatro que aunque fueron desarrollados hace varias décadas, siguen siendo ampliamente usados y punto de comparación para nuevas propuestas. A diferencia de otros métodos para los que no existen criterios para clasificar diferentes niveles de DIF, estos cuentan con criterios para cuantificar tamaños del efecto con poblaciones grandes (Sireci & Rios, 2013) y por lo tanto pueden brindar información específica y estandarizada sobre el funcionamiento de los ítems en evaluaciones a gran escala.

El método de Mantel-Haenszel al no basarse en la Teoría de Respuesta al Ítem (TRI) es de fácil implementación y no requiere de muestras grandes de sujetos. Prueba si existe una asociación entre la pertenencia a un grupo y la respuesta a los ítems de manera condicional sobre el puntaje total de la prueba (Magis et al., 2010). Compara la razón de momios de respuesta correcta de los dos grupos para diferentes intervalos de habilidad, a partir de una prueba de χ^2 (Dorans & Holland, 1992). Se obtiene el Alpha Mantel-Haenszel (MH), un estadístico que toma valores de cero a infinito positivo, donde el valor de uno indica que no hay diferencia en el rendimiento de los grupos; si es mayor a uno, el rendimiento del grupo de referencia es mejor al del grupo focal y lo contrario sucede si es menor a uno.

El Alpha MH suele transformarse a una escala simétrica conocida como Delta, cuyo estadístico se obtiene con la siguiente fórmula: $-2.35 \cdot \ln(\text{Alpha MH})$ (CENEVAL, 2009). Este estadístico se interpreta como una diferencia entre la dificultad de los ítems para los grupos de referencia y focal en la escala Delta (Roussos et al., 1999). Valores positivos indicarán que el ítem funciona a favor del grupo focal (es más difícil para el grupo de referencia); valores negativos que favorece al grupo de referencia (es más difícil para el grupo focal); y el valor de cero que no existe DIF. También se usan sus valores absolutos para clasificar el nivel de DIF:

Tabla I. Interpretación del estadístico Delta de Mantel-Haenszel

Nivel de DIF	Valor de Delta
Sin importancia	$ \delta < 1$
Leve a moderado	$1 \leq \delta < 1.5$
Moderado a alto	$1.5 \leq \delta $

Fuente: elaboración propia a partir de Zwick, 2012

Un valor absoluto de Delta menor a uno indica que la diferencia entre los grupos de referencia y focal no es significativa a un nivel de 5%, por lo que su grado de DIF es Sin importancia. Un valor absoluto de Delta mayor o igual a 1.5, indica que la diferencia entre ambos grupos es significativamente mayor que uno al nivel de 5% y se clasifica como Moderada a alta. Los ítems que no cumplen con los criterios mencionados se clasifican como con DIF Leve a moderado, debido a que una diferencia Delta de un punto, aunque indeseable, puede ser tolerada; mientras que una diferencia de Delta de dos o más debe ser evitada y el valor de 1.5 representa su límite inferior (Zwick, 2012). El método de Mantel-Haenszel es un referente para otros por su fácil y rápida implementación, porque cuenta con criterios de clasificación del nivel de DIF, detecta satisfactoriamente DIF Uniforme y puede ayudar a detectar DIF No Uniforme.

Swaminathan y Rogers (1990) indicaron que el método de Mantel-Haenszel no tiene poder estadístico para detectar DIF No Uniforme y propusieron un método basado en una regresión logística para detectar ambos tipos de funcionamiento diferencial. Un modelo logístico se ajusta con la probabilidad de contestar correctamente un ítem como variable dependiente, y el puntaje total de la prueba, el grupo de pertenencia y la interacción entre ambos como variables independientes. Se prueba si existe DIF Uniforme por medio del efecto del grupo y DIF No Uniforme por medio de la interacción de las variables. Zumbo y Thomas propusieron criterios para detectar DIF a partir del cambio en el coeficiente de determinación de las regresiones logísticas (ΔR^2), como medida del tamaño del efecto (en Magis et al., 2010). Posteriormente, Jodoin y Gierl (2001) propusieron una escala menos conservadora (ver la Tabla II), reportando que con esta nueva escala el porcentaje de ítems detectados con DIF Leve a moderado aumentaba del 6.8% al 68.2%. En contraste, se ha mostrado que el porcentaje de ítems dictaminados con DIF es menor en comparación con los detectados con la Delta de Mantel-Haenszel (Hidalgo & López-Pina, 2004).

Tabla II. Interpretación del estadístico ΔR^2 de la regresión logística

Nivel de DIF	Criterio de Zumbo y Thomas	Criterio de Jodoin y Gierl
Sin importancia	$\Delta R^2 \leq 0.13$	$\Delta R^2 \leq 0.035$
Leve a moderado	$0.13 \leq \Delta R^2 < 0.26$	$0.035 \leq \Delta R^2 < 0.07$
Moderado a alto	$0.26 \leq \Delta R^2$	$0.07 \leq \Delta R^2$

Fuente: elaboración propia a partir de Magis y colaboradores, 2010

El método de regresión logística es considerado una extensión del método de Mantel-Haenszel, diseñado para detectar ambos tipos de fun-

cionamiento diferencial, por medio de un modelo reconocido, flexible y de fácil implementación que se ajusta a muestras pequeñas (Jodoin & Gierl, 2001) y con tasas menores de falsos positivos que el método de Mantel-Haenszel (Abulela & Rios, 2022).

El método de Lord se basa en la hipótesis nula de tener para ambos grupos los mismos parámetros estimados del ítem a partir de la TRI, así como en un estadístico para su diferencia con una distribución χ^2 (Lord, 1980 en Magis et al., 2010). Puede ajustarse cualquier modelo de la TRI, por lo que puede detectar DIF Uniforme y No Uniforme, siempre y cuando los parámetros estén escalados en una métrica común antes de realizar la prueba estadística. Se ha probado su utilidad para detectar DIF porque controla la detección de falsos positivos (Uyar, 2020), aunque no siempre se encuentra congruencia con el método de Mantel-Haenszel, ya que este puede detectar ítems con DIF que no se detectan con el método de Lord (Raju et al., 1993).

Este método requiere tamaños de muestra de moderados a altos para estimaciones de parámetros estables, un conocimiento más avanzado en psicometría, así como programaciones y cálculos más complejos. Por lo que, si los exámenes se diseñan bajo algún modelo de la TRI, este método es muy sencillo de implementar ya que es una extensión de los análisis que comúnmente se realizan. Únicamente se han establecido criterios para determinar la magnitud del funcionamiento diferencial para el modelo de un parámetro, con el que se mide DIF Uniforme, porque solo calcula la discrepancia entre los parámetros de dificultad de los grupos focal y de referencia. Si la diferencia es negativa, el parámetro de dificultad del grupo focal es menor que el del grupo de referencia y se favorece al primero. Si la diferencia es positiva, el parámetro de dificultad del grupo focal es mayor que para el de referencia y este último se ve favorecido. El tamaño del efecto se estima como 2.35 veces la diferencia entre las dificultades de los ítems estimadas con los grupos bajo análisis (Penfield & Camilli, 2007). Debido a que únicamente se consideran las dificultades de los ítems, este estadístico es similar al de Mantel-Haenszel por lo que se usan los mismos criterios de la escala Delta para interpretarlo (ver la Tabla D).

El método de Raju se desarrolló porque uno de los principales supuestos del método de Lord es que la habilidad del participante es conocida. Sin embargo, se ha demostrado que cuando tanto la habilidad como los parámetros de los ítems son desconocidos, se puede caer en un Error de Tipo I (Cohen & Kim, 1993). Es por ello que se desarrolló otra aproximación basada en la TRI para detectar DIF, a partir de estimar el área entre las Curvas Características del ítem de ambos grupos bajo estudio

y del estadístico Z con la hipótesis nula de que la diferencia verdadera entre las áreas es cero (Raju, 1990). Nuevamente, cualquier modelo de la TRI puede ser implementado por lo que puede usarse para detectar ambos tipos de DIF. Dados los cálculos más complejos que requiere este método suele considerarse al de Lord más eficiente. Sin embargo, cuando el tamaño de la muestra es pequeño y existe poco DIF, el método de Raju es más efectivo (Hidalgo-Montesinos & Lopez-Pina, 2002).

Este método solo cuenta con criterios establecidos para el modelo de un parámetro y usa la diferencia entre los parámetros de dificultad y el procedimiento de Penfield y Camilli (2007) para estimar la dirección del sesgo y el tamaño del efecto, así como los criterios de la escala Delta (ver la Tabla I). Por lo que, aunque los estadísticos para estimar las diferencias para los métodos de Lord y Raju difieren (Özdemir, 2015), la estimación del tamaño del sesgo para el modelo de un parámetro en ambos métodos es idéntica.

Método

Se llevaron a cabo simulaciones donde se replicaron grupos con cantidades de participantes similares a las de las poblaciones de los EXANI y otras características del diseño de estas pruebas, como su número de reactivos o distribución de puntajes. Tomando como base los reactivos de uno de estos exámenes, algunos se manipularon para que presentaran DIF Uniforme Leve a moderado o Moderado a alto. Posteriormente, se realizaron análisis de DIF con los métodos de Mantel-Haenszel, Regresión logística, Lord y Raju y se compararon las detecciones correctas e incorrectas de funcionamiento diferencial, así como las no detecciones.

Muestra

Las poblaciones de los EXANI son grandes, aunque estas difieren de un examen a otro. En 2019¹, una versión del EXANI-III (ingreso al posgrado), fue presentada por 14,650 participantes, 1,698 para ingresar a la misma institución; mientras que una del EXANI-I (ingreso a la educación media superior), por 114,757, con 77,449 evaluados en un solo proceso de admisión y una del EXANI-II (ingreso a la educación superior), fue contestada por 202,264 participantes, 42,966 en la misma institución. A partir de estas cantidades y para dar continuidad a los estudios previos, se determinó que cada grupo de referencia y focal se compondría por 400, 1,000, 2,000, 10,000, 50,000 y 100,000 integrantes, con un total de

¹ Se usó este año como referente para no tener datos sesgados por la pandemia.

800, 2,000, 4,000, 20,000, 100,000 y 200,000 participantes simulados por estudio.

Tomando en cuenta que los EXANI se diseñan para distribuir los puntajes de los participantes de manera normal con una media de 50% de aciertos, para este estudio se simularon las habilidades de todos los grupos a contrastar para que tuvieran una media de cero y una desviación estándar de uno. De esta manera se controló que las diferencias encontradas no se debieran a un factor distinto al número de examinados en una evaluación a gran escala.

Instrumentos

Los EXANI son pruebas de alto impacto, de alta demanda, con referencia a la norma y dirigidas a poblaciones que participan en procesos de admisión a la educación media superior (EXANI-I), licenciatura (EXANI-II) y posgrado (EXANI-III) en México. Estos exámenes, hasta mediados del 2021, se conformaban por uno de admisión y uno de diagnóstico, donde los primeros contaban con un número similar de áreas e ítems: los EXANI-I y II se componían de cuatro áreas; en el EXANI-I cada área presentaba 20 ítems y en el EXANI-II, se presentaban 25 ítems por área. El EXANI-III contaba con siete áreas, cada una con 20 ítems; por lo que 20 fue el número de ítems simulado para esta investigación. Todos los ítems de esta generación de exámenes eran dicotómicos y se analizaban con el modelo de dos parámetros de la TRI para determinar su calidad estadística.

Procedimiento

Se usó el modelo de dos parámetros de la TRI para estimar los parámetros de los reactivos, se obtuvieron los valores de dificultad y discriminación de los reactivos que evalúan el área de Comprensión Lectora de una versión del EXANI-I Admisión presentada por 76,772 participantes. El valor mínimo de discriminación fue 0.520 y la media de dificultad 0.237. De los 20 ítems, se seleccionaron cuatro, dos con valores de dificultad media y dos con valores extremos (uno fácil y uno difícil). Estos valores fueron manipulados para que la diferencia absoluta entre ellos reflejara DIF Leve a moderado o Moderado a alto, modificando sus parámetros de dificultad para que las diferencias absolutas entre ellos fueran 0.6 o 1.1. La siguiente tabla muestra los valores de dificultad originales, los modificados y sus diferencias absolutas.

Tabla III. Valores del parámetro de discriminación (a) y dificultad (b) de los ítems simulados con DIF

Ítem	Discriminación	Dificultad			Diferencia absoluta (bf -br)	Tamaño DIF	Grupo favorecido
		Original	Grupo de referencia (br)	Grupo focal (bf)			
2	1.317	-1.255	-0.755	-1.855	1.1	Moderado a alto	Focal
8	1.073	-0.329	-0.029	-0.629	0.6	Leve a moderado	Focal
13	1.131	0.494	-0.006	1.094	1.1	Moderado a alto	Referencia
19	1.112	1.598	1.298	1.898	0.6	Leve a moderado	Referencia

Fuente: elaboración propia

La Tabla III muestra que se obtuvieron ítems con DIF de magnitudes Leves a moderadas y Moderadas a altas, con dificultades medias, altas y bajas, y que favorecen a ambos grupos. En cuanto a la discriminación, los cuatro valores eran cercanos a 1. Se manipuló un 10% o un 20% de los ítems con DIF, usando los dos primeros valores de la Tabla III o los cuatro.

Las simulaciones se llevaron a cabo en el Programa R versión 3.5.2 (R Core Team, 2018) y los análisis de DIF con la versión 5 de la paquetería difR (Magis et al., 2010).

Se simularon las habilidades y las cadenas de respuestas calificadas para cada grupo de referencia y focal (de tamaño 400 a 100,000), a partir tanto de los valores de los parámetros de los ítems correspondientes al estudio (10% o 20% con DIF), como de la media y desviación estándar de habilidad especificados a partir de la muestra de los EXANI (cero y uno, respectivamente). Las diferencias absolutas entre las medias de habilidad de los grupos focal y de referencia fueron menores a 0.0052 y de sus desviaciones estándar menores a 0.0026.

Una vez obtenidas las cadenas de respuestas calificadas se les agregó la etiqueta del grupo correspondiente (focal o referencia) y se integraron para realizar los análisis de DIF con los métodos de Mantel-Haenszel, Regresión Logística, Lord y Raju. En los métodos de Lord y Raju únicamente se implementó el modelo de Rasch (el parámetro de discriminación tuvo un valor fijo de 1). En todas las simulaciones se obtuvo, para cada uno de los 20 ítems, el estadístico del tamaño del efecto, así como la clasificación del grado de DIF de acuerdo con los criterios establecidos para cada método.

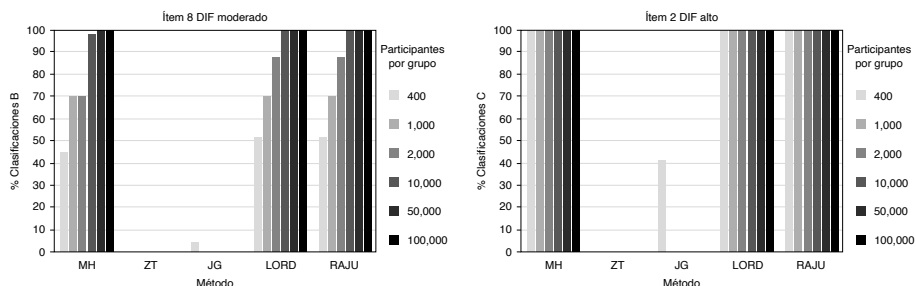
Se llevaron a cabo 100 iteraciones para cada muestra de participantes estudiada y cada porcentaje de ítems con DIF, iniciando con la simulación de las habilidades y cadenas de respuestas para cada grupo y finalizando con los análisis de DIF. Con los resultados de las 100 iteraciones, para cada ítem se obtuvo el porcentaje de reactivos clasificados con DIF Sin importancia (A), Moderado (B) y Alto (C) por cada método implementado, con lo que se obtuvo el porcentaje de detecciones o de no detecciones. Cuando el ítem fue manipulado para tener un grado de DIF Moderado se obtuvo el porcentaje de detecciones correctas (clasificaciones B), el de detecciones incorrectas (clasificaciones C) y el porcentaje de no detecciones (clasificaciones A). Cuando el ítem fue manipulado para tener un grado de DIF Alto, se obtuvo el porcentaje de detecciones correctas (clasificaciones C), el de detecciones incorrectas (clasificaciones B) y el porcentaje de no detecciones (clasificaciones A).

Para el resto de los ítems cuyos parámetros no fueron manipulados y eran idénticos en ambos grupos, se obtuvieron, de manera conjunta, las proporciones de detecciones incorrectas de DIF al ser clasificados como con DIF (Moderado o Alto).

Resultados

Primero se verificó la efectividad de los métodos en la prueba con el 10% de sus ítems con funcionamiento diferencial. El Gráfico I muestra el porcentaje de detecciones correctas en los dos reactivos con DIF. Los métodos de Mantel-Haenszel, Lord y Raju detectan tanto el DIF Moderado como el DIF Alto. En la detección del primero existe un efecto del tamaño de la muestra, donde a mayor cantidad de participantes, mayor el número de detecciones correctas de DIF, siendo con 400 participantes alrededor del 50% para los tres métodos, porcentaje que se incrementa y a partir de 10,000 evaluados a prácticamente el 100%. Los métodos de Lord y Raju son más precisos para detectar el DIF Moderado, ya que sus porcentajes de detección son mayores que los del Mantel-Haenszel. El método de regresión logística no detecta DIF Moderado con los criterios de Zumbo y Thomas ni con los de Jodoin y Gierl. También se observa que los métodos de Mantel-Haenszel, Lord y Raju detectan de manera exacta el DIF Alto con cualquier tamaño de muestra. El método de regresión logística únicamente lo detectó en el 42% de las simulaciones con el criterio de Jodoin y Gierl y 400 participantes.

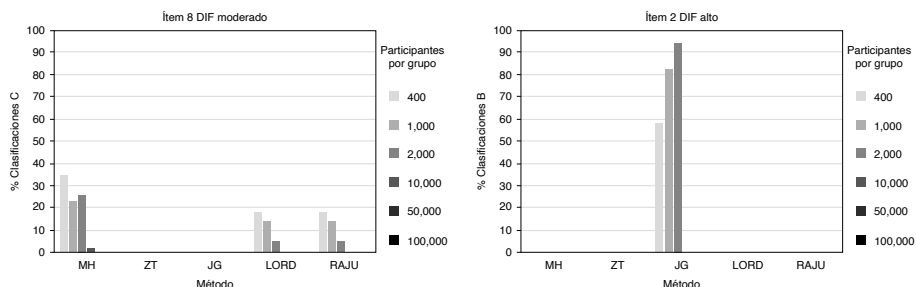
Gráfico I. Porcentaje de detecciones correctas de DIF para cada método² con un 10% de DIF



Fuente: elaboración propia

El siguiente gráfico muestra las detecciones incorrectas de DIF (cuando los métodos clasificaron a los reactivos con DIF en una categoría distinta a la que fue generada). Los métodos de Mantel-Haenszel, Lord y Raju pueden clasificar el DIF Moderado como Alto con 10,000 participantes o menos, siendo mayor esta sobreestimación para el primero que para los basados en la TRI y para muestras de 400 examinados. El método de regresión logística clasifica el DIF Alto como Moderado con el criterio de Jodoin y Gierl y con tamaños de muestra de hasta 2,000 participantes.

Gráfico II. Porcentaje de detecciones incorrectas de DIF con un 10% de DIF



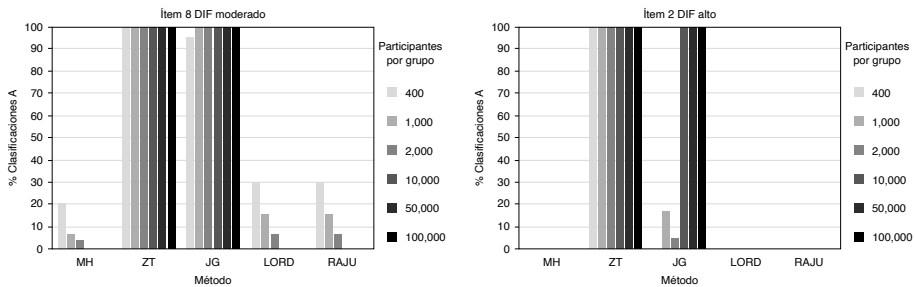
Fuente: elaboración propia

En el Gráfico III se presentan las no detecciones de DIF, donde se observa que el porcentaje de no detección de DIF Moderado del método de Mantel-Haenszel es menor que el de los métodos de Lord y Raju, el

² MH: Mantel-Haenszel; ZT: Regresión Logística, criterio Zumbo y Thomas; JG: Regresión Logística, criterio Jodoin y Gierl.

cual llega máximo al 30% en muestras pequeñas, y que este porcentaje disminuye conforme aumenta el número de participantes, siendo cero a partir de 10,000. El método logístico no detecta este DIF con ningún criterio. El mismo gráfico muestra que no existen no detecciones de DIF Alto con los métodos de Mantel-Haenszel, Lord y Raju y confirma que el método de regresión logística nunca lo detecta con el criterio de Zumbo y Thomas, y que bajo el criterio de Jodoin y Gierl no lo detecta a partir de 10,000 participantes.

Gráfico III. Porcentaje de no detecciones de DIF con un 10% de DIF



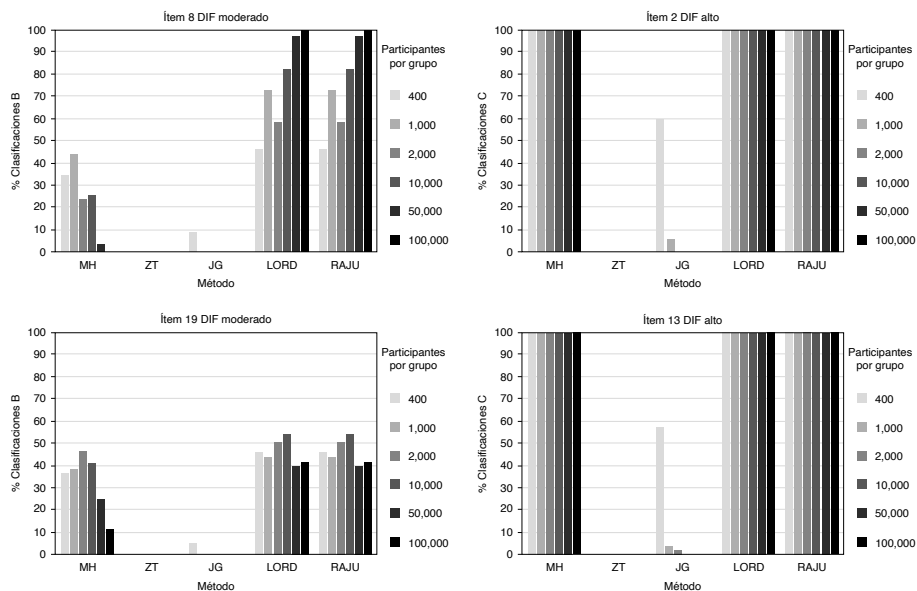
Fuente: elaboración propia

Para los 18 ítems cuyos parámetros fueron idénticos entre los grupos a contrastar, se obtuvo el porcentaje de falsas detecciones de DIF. Únicamente hubo detecciones erróneas con los métodos de Mantel-Haenszel, Lord y Raju en menos del 1.90% de estos ítems en las simulaciones con 400 participantes y en menos del 0.07% con los métodos de la TRI en aquellas con 1,000 participantes.

Posteriormente se verificó la efectividad de los mismos métodos con el 20% de los ítems con DIF: dos con DIF Moderado y dos con Alto. El siguiente gráfico muestra el porcentaje de detecciones correctas. Se observa que el tener más ítems con DIF alteró la eficacia de los métodos para detectar DIF Moderado. Al comparar las figuras de la izquierda de los gráficos I y IV (primera fila), que corresponden al mismo ítem, se observa que las proporciones de detecciones correctas de los métodos Mantel-Haenszel, Lord y Raju, disminuyeron, aunque los métodos basados en la TRI siguen siendo más efectivos que el de Mantel-Haenszel. Para este último se observa un efecto del tamaño de la población inverso al obtenido anteriormente: sus detecciones correctas disminuyen conforme aumenta el número de participantes. Con los métodos de Lord y Raju se mantiene el efecto anterior: su precisión aumenta conforme crece el número de participantes, aunque ambos métodos son precisos

únicamente con 100,000 participantes. El método de regresión logística no detecta DIF Moderado con ningún criterio.

Gráfico IV. Porcentaje de detecciones correctas de DIF con un 20% de DIF



Fuente: elaboración propia

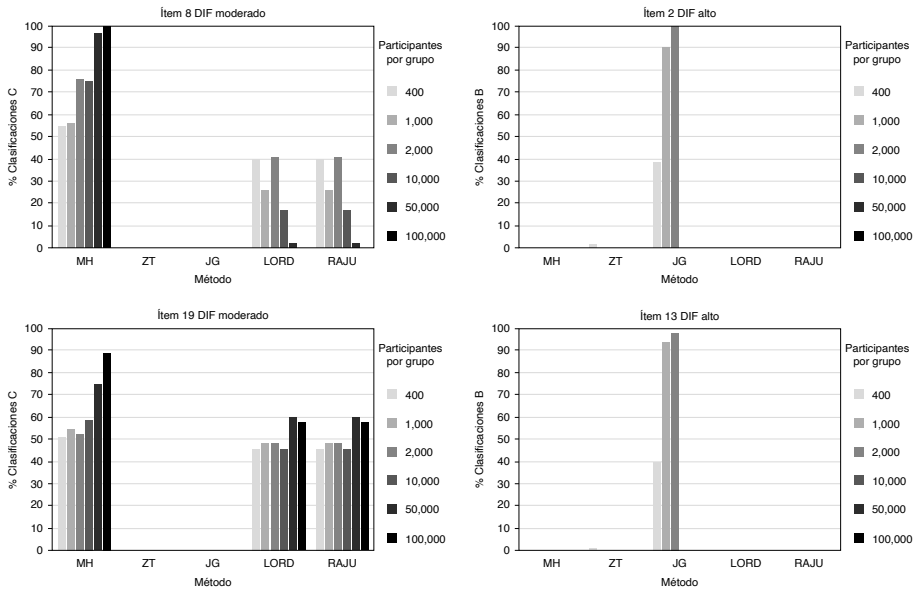
Las figuras de la derecha en los gráficos I y IV (primera fila), corresponden al mismo ítem, y muestran que no se alteró la eficacia para detectar DIF Alto al añadir ítems con DIF: los mismos tres métodos siguen detectándolo en todas las ocasiones y el de regresión logística lo detectó en un máximo de 61% bajo el criterio de Jodoin y Gierl con 400 participantes.

Las figuras de la segunda fila del Gráfico IV muestran las detecciones correctas de DIF en los ítems que se manipularon adicionalmente para esta condición. En la izquierda se muestra que el porcentaje de detección de DIF Moderado es menor al 50% con los métodos de Mantel-Haenszel, Lord y Raju y en ninguno se observa un efecto del tamaño de la población. Cabe resaltar que este ítem (el 19 en la Tabla III), originalmente era uno difícil que con la manipulación para generar el DIF se volvió el más difícil con DIF y el segundo más difícil de la versión, por lo que el tener una dificultad extrema pudo afectar la eficacia de la detección de este DIF. La figura de la derecha muestra que en el DIF Alto se replican los

resultados previos: los métodos de Mantel-Haenszel, Lord y Raju tienen un porcentaje de 100% de detección; mientras que el método de regresión logística lo detectó cerca del 60% de las simulaciones bajo el criterio de Jodoin y Gierl y con 400 participantes.

En el gráfico V se presentan las detecciones incorrectas de DIF donde se observa nuevamente, del lado izquierdo, que cuando los métodos de Mantel-Haenszel, Lord y Raju no detectan correctamente el DIF Moderado lo clasifican como Alto. Para el primero, esta sobreestimación aumenta conforme aumenta el número de participantes. El método de regresión logística nunca clasifica el DIF Moderado como Alto. Las gráficas de la derecha muestran que este método clasifica a ambos ítems con DIF Alto como con Moderado únicamente bajo el criterio de Jodoin y Gierl, y con poblaciones de 400 a 2,000 participantes.

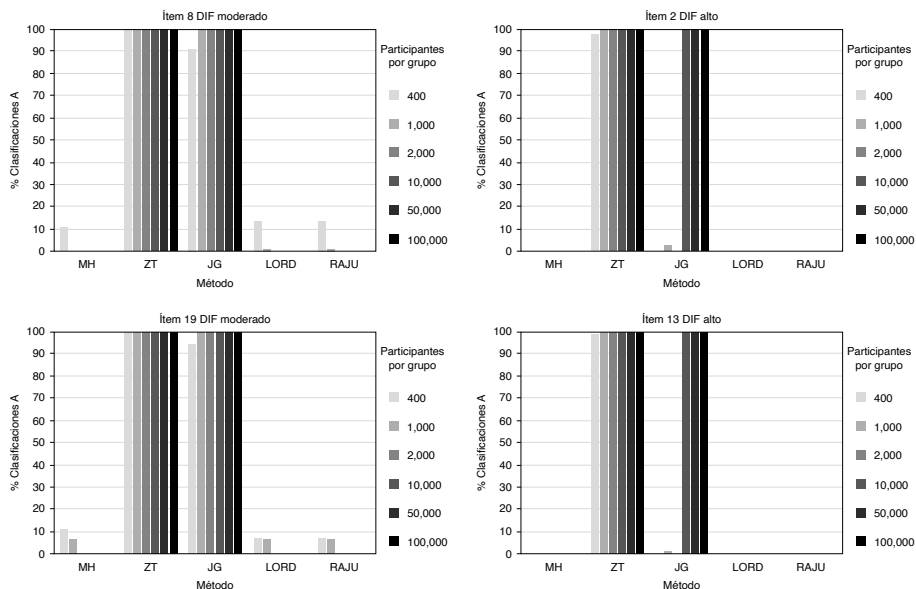
Gráfico V. Porcentaje de detecciones incorrectas de DIF con un 20% de DIF



Fuente: elaboración propia

En cuanto al porcentaje de no detecciones, a la derecha del siguiente gráfico se observa que los métodos de Mantel-Haenszel, Lord y Raju llegan a no detectar el DIF Moderado con 1,000 o 400 participantes y que el de regresión logística clasifica a estos ítems como sin DIF con ambos criterios, al igual que a los ítems con DIF Alto, con excepción de las condiciones donde los clasificó como con DIF Moderado.

Gráfico VI. Porcentaje de no detecciones de DIF con un 20% de DIF



Fuente: elaboración propia

Finalmente, el porcentaje de detecciones incorrectas de DIF en los ítems cuyos parámetros fueron iguales entre los grupos contrastados disminuyó con más ítems con DIF en el área, siendo de 1.63% para los métodos de Mantel-Haenzel, Lord y Raju en las simulaciones con 400 participantes y de cero en las condiciones restantes y con el método de regresión logística.

Conclusiones

Este estudio verificó la efectividad de cuatro métodos para detectar funcionamiento diferencial uniforme en poblaciones cinco a 50 veces mayores a las analizadas en estudios previos de simulación y con datos empíricos para verificar que pudieran ser usados en estudios con poblaciones completas de algunas ILSAs u otros exámenes que se apliquen a gran escala.

Los resultados muestran que el método de regresión logística no es apropiado para analizar DIF en ítems de evaluaciones a gran escala, debido a que únicamente lo detectó con el criterio de Jodoin y Gierl con hasta 2,000 participantes. Estos resultados son consistentes con reportes anteriores que indican que este método es efectivo para grupos de hasta 1,000 sustentantes (French & Maller, 2007; Jodoin & Gierl, 2001).

Adicionalmente se obtuvo que con 20 ítems y el 10% de ellos con DIF (un ítem con DIF Moderado y uno con Alto), a mayor cantidad de integrantes por grupo se mejora la detección de DIF Moderado con los métodos de Mantel-Haenszel, Lord y Raju, siendo más exactos los dos últimos con hasta 10,000 participantes por grupo, ya que hasta este tamaño de muestra, el de Mantel-Haenszel lo sobreestima o clasifica como Alto en mayor medida. También se encontró que estos tres métodos son 100% precisos para detectar DIF Moderado a partir de 50,000 evaluados por grupo, aunque esta precisión disminuyó considerablemente al aumentar el porcentaje de ítems con DIF al 20%. Cabe resaltar que ninguno de los tres métodos coincidió en su clasificación en todas las simulaciones de DIF Moderado. Por ejemplo, en 70 simulaciones de grupos con 1,000 participantes, los tres métodos lo detectaron correctamente, aunque en 18 de estas corridas difirieron en la clasificación.

Los métodos de Mantel-Haenszel, Lord y Raju fueron precisos para detectar DIF Alto a partir de 400 participantes por grupo, sin importar el porcentaje de DIF en el área, el número de evaluados, la dificultad del ítem o si beneficiaba al grupo focal o al de referencia. También se confirma que los resultados obtenidos con los métodos de Lord y Raju son idénticos con grandes cantidades de participantes, tanto para el DIF Moderado como para el Alto.

Finalmente, se observó que con grupos de contraste con 400 y hasta 100,000 participantes no hubo un mayor Error Tipo I con el método Mantel-Haenszel, es decir, este método no detectó DIF cuando no lo hubo, en cambio en los métodos basados en la TRI, se identificaron detecciones incorrectas mínimas y solo se presentaron en grupos con 1,000 participantes o menos. Por lo que los falsos positivos de estos tres métodos en ítems sin DIF son mínimos e inexistentes con muestras de participantes de gran tamaño.

Aunque ningún método fue completamente exacto para detectar DIF Uniforme, los de Mantel-Haenszel, Lord y Raju fueron los más eficaces ya que, aunque su clasificación no fue siempre precisa, el funcionamiento diferencial no pasó desapercibido y cuando no lo clasificaron correctamente, lo sobreestimaron. Por ello, se sugiere que en estudios enfocados en los sesgos de evaluaciones a gran escala que busquen comparar poblaciones completas y no solo muestras, se implementen los métodos de Mantel-Haenszel, Lord o Raju para analizar el DIF en ítems dicotómicos en pruebas con aproximadamente 20 ítems, presentados por 800 hasta 200,000 participantes o con grupos a contrastar entre 400 y 100,000 personas cada uno. Para obtener la mayor información posible, se pueden realizar análisis con varios métodos (Özdemir, 2015), aunque debido a que los métodos de Lord y Raju arrojan los mismos resultados

con estos tamaños de muestra al usar los mismos criterios, y a que el segundo implica cálculos más avanzados, puede ser suficiente aplicar conjuntamente los métodos de Mantel-Haenszel y Lord. Con la clasificación obtenida, se podrían detectar aquellos ítems que requieran un análisis cualitativo, especialmente aquellos que ambos métodos marquen con DIF Alto, para determinar si están sesgados y así establecer las acciones a seguir para esos reactivos.

El usar estos métodos para los análisis de DIF con poblaciones de estas magnitudes podría permitir realizar investigaciones usando toda la información posible, sin la necesidad de obtener muestras de los grupos evaluados de las ILSAs o realizar comparaciones entre los participantes de países, regiones geográficas o económicas que hubieran contestado el examen en el mismo idioma. De igual manera se podrían realizar estudios con poblaciones completas de aplicaciones de los EXANI u otros exámenes similares que se apliquen a nivel nacional. En todos estos casos, sería el primer paso para asegurar que estas evaluaciones con alto impacto estén libres de sesgo.

Este estudio aporta información para la implementación de tres métodos tradicionales de DIF en muestras grandes de participantes que responden instrumentos de evaluaciones a gran escala. Sin embargo, al ser el tamaño de la muestra solo una de las variables estudiadas, se requiere indagar si otras como diferentes distribuciones, tamaños diferentes de los grupos contrastados, mayores cantidades de reactivos en la prueba, mayores porcentajes de reactivos con DIF y métodos de purificación, afectan la efectividad de estos métodos. De igual manera, existen y siguen surgiendo nuevos métodos para detectar funcionamiento diferencial (Halpin, 2024; Liu et al., 2019; Suk & Han, 2024; Villalonga-Olives et al., 2024) y es necesario investigar su efectividad en muestras de participantes similares a los del presente estudio. Finalmente, es necesario corroborar la efectividad de estos métodos con datos empíricos, por lo que se podrían retomar estudios en los que se haya detectado DIF en reactivos de alguna ILSA y realizar los análisis con la población completa.

Referencias bibliográficas

- Abulela, M. A. A., & Rios, J. A. (2022). Comparing the robustness of three nonparametric DIF procedures to differential rapid guessing. *Applied Measurement in Education*, 35(2), 81–94. <https://doi.org/10.1080/08957347.2022.2067542>
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (Eds.).

- (2014). *Standards for educational and psychological testing*. American Educational Research Association.
- Cassiani-Miranda, C. A., Scopetta, O., Barrios-Villadiego, M. A., Tirado-Otálvaro, A. F., & Duran-Bedoya, A. C. (2025). Construct validity of the PHQ-9 in university students in Colombia: A Rasch analysis approach. *Medicina Clínica y Social*, 9(1), e521. <https://doi.org/10.52379/mcs.v9.521>
- Centro Nacional de Evaluación para la Educación Superior. (2009). *Procedimientos Básicos para el Análisis de Reactivos*.
- Cohen, A. S., & Kim, S. H. (1993). A comparison of Lord's X^2 and Raju's area measures in detection of DIF. *Applied Psychological Measurement*, 17(1), 39-52. <https://doi.org/10.1177/014662169301700109>
- Demirus, K. B., & Pektaş, S. (2022). Investigation of TIMSS 2015 science test items in terms of differential item functioning according to language and culture. *International Journal of Education Technology and Scientific Researches*, 7(18), 1166-1178. <http://dx.doi.org/10.35826/ijetsar.499>
- Dorans, N. J., & Holland, P. W. (1992). DIF detection and description: Mantel-Haenszel and standardization. *ETS Research Report Series*, 1992(1), i-40. <https://doi.org/10.1002/j.2333-8504.1992.tb01440.x>
- Fidalgo, A. M., Mellenbergh, G. J., & Muñiz, J. (2000). Effects of amount of DIF, test length, and purification type on robustness and power of Mantel-Haenszel procedures. *Methods of Psychological Research Online*, 5(3), 43-53.
- Fishbein, B., Yin, L., & Foy, P. (2024). *PIRLS 2021 User Guide for the International Database* (2nd ed.). Boston College, TIMSS & PIRLS International Study Center. <https://pirls2021.org/data>
- Fraillon, J. (2024). *An international perspective on digital literacy: Results from ICILS 2023*. International Association for the Evaluation of Educational Achievement. https://www.iea.nl/sites/default/files/2024-11/ICILS_2023_International_Report_0.pdf
- French, B. F., & Maller, S. J. (2007). Iterative purification and effect size use with logistic regression for differential item functioning detection. *Educational and Psychological Measurement*, 67(3), 373-393. <https://doi.org/10.1177/0013164406294781>
- Gierl, M. J., Gotzmann, A., & Boughton, K. A. (2004). Performance of SIBTEST when the percentage of DIF items is large. *Applied Measurement in Education*, 17(3), 241-264. https://doi.org/10.1207/s15324818ame1703_2
- Halpin, P. F. (2024). Differential item functioning via robust scaling. *Psychometrika*, 89(3), 796-821. <https://doi.org/10.1007/s11336-024-09957-6>

- Hidalgo-Montesinos, M. D., & Lopez-Pina, J. A. (2002). Two-stage equating in differential item functioning detection under the graded response model with the Raju area measures and the Lord statistic. *Educational and Psychological Measurement*, 62(1), 32-44. <https://doi.org/10.1177/0013164402062001003>
- Hidalgo, M. D., & López-Pina, J. A. (2004). Differential item functioning detection and effect size: A comparison between logistic regression and Mantel-Haenszel procedures. *Educational and Psychological Measurement*, 64(6), 903-915. <https://doi.org/10.1177/0013164403261769>
- Jodoin, M. G., & Gierl, M. J. (2001). Evaluating type I error and power rates using an effect size measure with the logistic regression procedure for DIF detection. *Applied Measurement in Education*, 14(4), 329-349. https://doi.org/10.1207/S15324818AME1404_2
- Kucam, E., & Gülleroğlu, H. D. (2023). Examination of differential item functioning in PISA 2018 Mathematics Literacy test with different methods. *Journal of Measurement and Evaluation in Education and Psychology*, 14(2), 128-153. <https://doi.org/10.21031/epod.1122857>
- Liu, R., & Bradley, K. D. (2021). Differential item functioning among English language learners on a large-scale mathematics assessment. *Frontiers in Psychology*, 12, 657335. <https://doi.org/10.3389/fpsyg.2021.657335>
- Liu, Y., Yin, H., Xin, T., Shao, L., & Yuan, L. (2019). A comparison of differential item functioning detection methods in cognitive diagnostic models. *Frontiers in Psychology*, 10, 1137. <https://doi.org/10.3389/fpsyg.2019.01137>
- Magis, D., Béland, S., Tuerlinckx, F., & De Boeck, P. (2010). A general framework and an R package for the detection of dichotomous differential item functioning. *Behavior Research Methods*, 42(3), 847-862. <https://doi.org/10.3758/BRM.42.3.847>
- Mellenbergh, G. J. (1982). Contingency table models for assessing item bias. *Journal of Educational Statistics*, 7(2), 105-118. <https://doi.org/10.2307/1164960>
- Narayanan, P., & Swaminathan, H. (1994). Performance of the Mantel-Haenszel and simultaneous item bias procedures for detecting differential item functioning. *Applied Psychological Measurement*, 18(4), 315-328. <https://doi.org/10.1177/014662169401800403>
- Narayanan, P., & Swaminathan, H. (1996). Identification of items that show nonuniform DIF. *Applied Psychological Measurement*, 20(3), 257-274. <https://doi.org/10.1177/014662169602000306>
- Organisation for Economic Co-operation and Development. (2023). *PISA 2022 results (Volume I): The state of learning and equity in education*. <https://doi.org/10.1787/53f23881-en>

- Ohiri, S. C., Momoh, O. C., & Ikeanumba, C. B. (2024). Differential item functioning detection methods: An overview. *International Journal of Research Publication and Reviews*, 5(2), 1555-1564. <https://doi.org/10.55248/gengpi.5.0224.0505>
- Özdemir, B. (2015). A comparison of IRT-based methods for examining differential item functioning in TIMSS 2011 mathematics subtest. *Procedia-Social and Behavioral Sciences*, 174, 2075-2083. <https://doi.org/10.1016/j.sbspro.2015.02.004>
- Park, S. E., Ahn, S., & Zopluoglu, C. (2021). Differential item functioning effect size from the multigroup confirmatory factor analysis for a meta-analysis: A simulation study. *Educational and Psychological Measurement*, 81(1), 182-199. <https://doi.org/10.1177/0013164420925885>
- Penfield, R. D., & Camilli, G. (2007). *Differential item functioning and item bias*. En C. R. Rao y S. Sinharay (Eds.), *Handbook of Statistics: Vol. 26. Psychometrics* (pp. 125-167). Elsevier. [https://doi.org/10.1016/S0169-7161\(06\)26005-X](https://doi.org/10.1016/S0169-7161(06)26005-X)
- R Core Team. (2018). *R: A language and environment for statistical computing. R foundation for Statistical Computing*.
- Raju, N. S. (1990). Determining the significance of estimated signed and unsigned areas between two item response functions. *Applied Psychological Measurement*, 14(2), 197-207. <https://doi.org/10.1177/014662169001400208>
- Raju, N. S., Drasgow, F., & Slinde, J. A. (1993). An empirical comparison of the area methods, Lord's chi-square test, and the Mantel-Haenszel technique for assessing differential item functioning. *Educational and Psychological Measurement*, 53(2), 301-314. <https://doi.org/10.1177/0013164493053002001>
- Rogers, H. J., & Swaminathan, H. (1993). A comparison of logistic regression and Mantel-Haenszel procedures for detecting differential item functioning. *Applied Psychological Measurement*, 17(2), 105-116. <https://doi.org/10.1177/014662169301700201>
- Roussos, L. A., Schnipke, D. L., & Pashley, P. J. (1999). A generalized formula for the Mantel-Haenszel differential item functioning parameter. *Journal of Educational and Behavioral Statistics*, 24(3), 293-322. <https://doi.org/10.3102/10769986024003293>
- Shamsaldeen, F., Wang, J., & Ahn, S. (2024). Evaluating the fairness of a high-stakes college entrance exam in Kuwait. *Language Testing in Asia*, 14(1), 27. <https://doi.org/10.1186/s40468-024-00301-4>
- Schulz, W., Ainley, J., Fraillon, J., Losito, B., Agrusti, G., Damiani, V., & Friedman, T. (2025). *Education for citizenship in times of global cha-*

- llenge: IEA international civic and citizenship education study 2022 international report. Springer. <https://doi.org/10.1007/978-3-031-65603-3>
- Sireci, S. G., & Rios, J. A. (2013). Decisions that make a difference in detecting differential item functioning. *Educational Research and Evaluation, 19*(2-3), 170-187. <https://doi.org/10.1080/13803611.2013.767621>
- Suk, Y., & Han, K. T. (2024). A psychometric framework for evaluating fairness in algorithmic decision making: Differential algorithmic functioning. *Journal of Educational and Behavioral Statistics, 49*(2), 151-172. <https://doi.org/10.3102/10769986231171711>
- Swaminathan, H., & Rogers, H. J. (1990). Detecting differential item functioning using logistic regression procedures. *Journal of Educational Measurement, 27*(4), 361-370. <https://doi.org/10.1111/j.1745-3984.1990.tb00754.x>
- Universidad Autónoma de Madrid. (2024). *EvAU 2024: Presentación de resultados* [Presentación de PowerPoint]. <https://www.uam.es/uam/media/doc/1606958184663/rueda-de-prensa-uam.-datos-evau-2024-compressed.pdf>
- Uyar, S. (2020). Latent class approach to detect differential item functioning PISA 2015 science sample. *Eurasian Journal of Educational Research, 20*(88), 179-198. <https://ejer.com.tr/latent-class-approach-to-detect-differential-item-functioning-pisa-2015/>
- Velázquez Ramírez, S. (2022). Exámenes que contribuyen a los procesos de admisión. *Revista Ceneval Investiga, 4*, 22-35. <https://online.flippingbook.com/view/556549215/>
- Vidal Gonzáles, F. L., Pacheco Leyva, Y., & Galindo Cortés, G. (2022). Comparabilidad en las aplicaciones remotas y presenciales del EXANI-II. *Revista Ceneval Investiga, 6*, 28-37. <https://online.flippingbook.com/view/917214422/>
- Villalonga-Olives, E., Khademi, A., Pan, Y. Y., & Ransome, Y. (2024). Unveiling disparities: examining differential item functioning's impact on racial health equity among white and black populations. *Public Health, 231*, 80-87. <https://doi.org/10.1016/j.puhe.2024.03.006>
- von Davier, M., Kennedy, A., Reynolds, K., Fishbein, B., Khorramdel, L., Aldrich, C., Bookbinder, A., Bezirhan, U., & Yin, L. (2024). *TIMSS 2023 international results in Mathematics and Science*. Boston College, TIMSS & PIRLS International Study Center. <https://doi.org/10.6017/lse.tpisc.timss.rs6460>
- Walker, C. M. (2011). What's the DIF? Why differential item functioning analyses are an important part of instrument development and valida-

- tion. *Journal of Psychoeducational Assessment*, 29(4), 364-376. <https://doi.org/10.1177/0734282911406666>
- Webster, N. S., & Bowe, A. G. (2021). Examining measurement equivalency within the Latin American module of the International Civic Competence Study 2009. *Journal of Community Psychology*, 49(7), 2972-2982. <https://doi.org/10.1002/jcop.22515>
- Wang, W. C., Shih, C. L., & Yang, C. C. (2009). The MIMIC method with scale purification for detecting differential item functioning. *Educational and Psychological Measurement*, 69(5), 713-731. <https://doi.org/10.1177/0013164409332228>
- Wu, M., Tam, H. P., & Jen, T.-H. (2016). *Educational measurement for applied researchers*. Springer.
- Zieky, M. (2006). Fairness review in assessment. En S. M. Downing y T. M. Haladyna (Eds.), *Handbook of test development* (pp. 359-376). Lawrence Erlbaum Associates.
- Zumbo, B. D. (2007). Three generations of DIF analyses: Considering where it has been, where it is now, and where it is going. *Language Assessment Quarterly*, 4(2), 223-233. <https://doi.org/10.1080/15434300701375832>
- Zwick, R. (2012). A review of ETS differential item functioning assessment procedures: Flagging rules, minimum sample size requirements, and criterion refinement. *ETS Research Report Series*, 2012(1), i-30. <https://doi.org/10.1002/j.2333-8504.2012.tb02290.x>

Resolución de problemas matemáticos en la formación del profesorado de educación primaria española¹

Mathematical problem-solving in primary education teacher training programmes in Spain

<https://doi.org/10.4438/1988-592X-RE-2025-412-734>

Elena Castro-Rodríguez

<https://orcid.org/0000-0002-2560-8982>

Universidad de Granada

Juan Luis Piñeiro

<https://orcid.org/0000-0002-9616-3925>

Universidad Metropolitana de Ciencias de la Educación, Chile

Enrique Castro

<https://orcid.org/0000-0002-2786-500X>

Profesor catedrático jubilado. Universidad de Granada

Resumen

Introducción: La relevancia de la resolución de problemas matemáticos queda manifestada en los documentos curriculares de Educación Primaria. Sin embargo, para que esto pueda llevarse a cabo es necesario formar a los docentes y conocer sus planes de estudio. Por ello, este trabajo describe la presencia de la resolución de problemas y su enseñanza en las guías docentes de las asignaturas obligatorias y optativas del área de Didáctica de la Matemática pertenecientes a los grados de Maestro en Educación Primaria en España. **Método:** Realizamos un análisis descriptivo y un análisis clúster de 151 guías docentes de asignaturas obligatorias y optativas del área de Didáctica de la Matemáticas de 39 universi-

¹ Realizado con el apoyo de los proyectos PID2020-113601 GB-I00, PID2024-157106NB-I00 financiados por MCIN/AEI/10.13039/501100011033, y FEDER, UE y del proyecto FOVI240238 para el fomento a la vinculación internacional para instituciones de investigación de la Agencia Nacional de Investigación y Desarrollo (ANID) del Ministerio de Ciencias de Chile.

dades públicas. **Resultados:** Los resultados muestran un énfasis en la competencia para resolver problemas sobre elementos de conocimiento especializado. Asimismo, en relación al conocimiento sobre la resolución de problemas, aparece con mayor frecuencia las menciones a tipos de problemas, mientras que, respecto al conocimiento didáctico de la resolución de problemas, los tópicos relativos a la invención de problemas son los más frecuentes. **Conclusiones:** Concluimos que cada universidad ha interpretado la enseñanza y aprendizaje de la resolución de problemas desde su propio entendimiento por lo que la formación entre maestros españoles es muy heterogénea. Además, destacamos la necesidad de incluir aspectos como la consideración del resolutor o factores no cognitivos para mejorar la formación sobre resolución de problemas de los futuros maestros.

Palabras clave: estudios universitarios, formación inicial, investigación curricular, matemáticas, profesor, solución de problemas

Abstract

Introduction: *The relevance of mathematical problem solving is reflected in the curricular documents of Primary Education. However, in order for this to be carried out, it is necessary to train teachers and know their study plans. Therefore, this work describes the presence of problem solving and its teaching in the syllabuses of the compulsory and elective subjects in the area of Didactics of Mathematics belonging to the university degrees of Teacher in Primary Education in Spain.* **Method:** *We have carried out a content analysis and cluster analysis of 151 syllabuses from 39 Spanish public universities.* **Results:** *The results show an emphasis on problem-solving competence over elements of specialized knowledge. Likewise, regarding knowledge about problem characterization appear more frequently, while regarding pedagogical problem solving knowledge, topics related to the problem posing are more frequent.* **Conclusions:** *We conclude that each university has interpreted the teaching and learning of problem solving from its own understanding, so the training among Spanish teachers is very heterogeneous. In addition, we highlight the need to include aspects such as the consideration of the solver or non-cognitive factors to improve teacher training on problem solving.*

Keywords: curriculum research, initial training, mathematics, problem solving, teacher, university studies

Introducción

La resolución de problemas (RP) —entendida como una tarea para la que el resolutor no dispone de un procedimiento inmediato y directo a la solución, pero sí la asunción de que existe un camino posible— es un

medio importante para desarrollar el conocimiento matemático en los niños de educación primaria (Britz y Richard, 1992; Schoenfeld, 2013). Resolver problemas significativos contribuye a desarrollar habilidades de pensamiento de orden superior y descubrir un repertorio de estrategias que preparan al escolar para resolver nuevos problemas (Lester y Cai, 2016). En este sentido, los niños adquieren sentido de las ideas matemáticas mediante la participación activa en la resolución de una variedad de problemas matemáticos (Britz y Richard, 1992).

La relevancia que ha adquirido la RP en las primeras edades se ha recogido en el documento curricular español actual para la etapa de educación primaria (RDL 157/2022, de 1 de marzo). En él destaca el papel fundamental que ocupa la RP, ya que se promueve como principio pedagógico fomentar la integración de las competencias dedicando un tiempo del horario lectivo a la RP (RDL 157/2022, de 1 de marzo). En este currículo, la resolución de problemas es la competencia a la que se dedica especial atención, “considerándola más como principio metodológico que como contenido en sí mismo” (Contreras, p. 66, 2022). Sin embargo, para que esta propuesta curricular pueda llevarse a cabo, es necesario preparar a los futuros maestros. En este sentido, este trabajo pretende contribuir a ello centrándose en dar respuesta a dos objetivos. En primer lugar, describir la presencia de las asignaturas del área de Didáctica de la Matemática en los títulos universitarios oficiales que habilitan para el ejercicio de la profesión de Maestro en Educación Primaria en España; y en segundo lugar, describir la presencia de la RP y su enseñanza en las guías docentes de las asignaturas del área de Didáctica de la Matemática pertenecientes a dichos títulos.

Conocer los planes de estudio es importante pues si estas ofertas y experiencias curriculares son internamente coherentes tienden a tener un mayor impacto en el conocimiento de los futuros docentes (Tatto, 2018). En esta línea, Hiebert et al. (2019) indagan si los maestros noveles retenían y utilizaban las matemáticas estudiadas durante su formación inicial. Si bien sus resultados muestran diferencias entre los participantes, estos siempre obtuvieron mejores resultados en los contenidos trabajados en la formación más que en los no abordados. Asimismo, Morris y Hiebert (2017) estudian si los contenidos de matemáticas estudiados en un programa de formación de maestros de primaria tienen algún efecto en sus prácticas una vez que los futuros profesores están en servicio. Los resultados mostraron que cuando los profesores habían puesto atención al contenido estudiado durante su formación, estos tendían a utilizarlos en sus planificaciones de enseñanza. Por tanto, la evidencia sugiere que existe influencia entre los contenidos estudiados en la formación con el

conocimiento especializado utilizado en la enseñanza. En este sentido, conocer y comprender los planes de formación de maestros de las universidades españolas puede ser de gran utilidad para los formadores y diseñadores de currículos.

La resolución de problemas en la formación del profesor de matemáticas

A pesar de que la investigación en educación matemática se ha centrado en la RP durante más de 50 años, su implementación continúa siendo un desafío para los docentes (Liljedahl y Cai, 2021). Entre las posibles causas, algunos autores destacan las creencias de los docentes, considerándola como un medio para alcanzar una solución centrándose en aspectos procedimentales (Son y Lee, 2021). Independientemente de los motivos, muchos docentes necesitan ayuda para desarrollar en su aula la RP, y una fuente de esta ayuda es el desarrollo profesional.

Diversos expertos en el área se han centrado en el desarrollo profesional relativo a la RP (e.g., Felmer y Perdomo-Díaz, 2017; Leung, 2016; Wake et al., 2016). Por ejemplo, en Chile, se llevó a cabo la iniciativa denominada Activación de la Resolución de Problemas en Aula con el objetivo de introducir la resolución de problemas en las clases de los docentes de primaria y secundaria (Felmer y Perdomo-Díaz, 2017). Para ello, se llevaron a cabo tres talleres de carácter práctico en varias regiones del país, en donde se fomentaba la enseñanza de la resolución de problemas de manera colaborativa y no rutinaria. En los talleres, hubo oportunidades para que los docentes resolvieran problemas y también para reflexionar sobre la noción de problema. Los resultados mostraron que hubo una mejora estadísticamente significativa en el desempeño al resolver problemas entre los escolares cuyos maestros participaron en los talleres.

En otras ocasiones, el desarrollo profesional se llevó a cabo a través de proyectos en donde los investigadores tienen reuniones con los docentes para reflexionar sobre la planificación de la enseñanza y posteriormente, sobre su puesta en práctica (e.g., Hähkiöniemi y Francisco, 2019; Leung, 2016; Wake et al., 2016). Por ejemplo, el proyecto *Informal Mathematical Learning* en Estados Unidos o el proyecto *The Inquiry-Based Mathematics Teaching* en Finlandia, incluyen un componente de desarrollo profesional destinado a ayudar a los profesores de secundaria a guiar a los escolares en la resolución de problemas. Para ello, grabaron las lecciones impartidas y realizaron reuniones con los docentes donde reflexionaron

conjuntamente sobre su actuación en el aula, concluyendo que la implementación de sesiones de resolución de problemas puede verse afectada por la limitación de tiempo. De manera similar, en el proyecto de Wake et al. (2016) llevado a cabo en Inglaterra, los investigadores planificaron en colaboración con los profesores de secundaria las sesiones de resolución de problemas. Inmediatamente tras su puesta en práctica, realizaron una discusión sobre las intervenciones específicas que hizo el docente y cómo afectó la planificación y la comunicación entre los estudiantes en su desempeño en resolución de problemas.

A diferencia de los anteriores, otros proyectos incluyen la invención de problemas como parte esencial. Por ejemplo, el proyecto desarrollado por Leung (2016) en Taiwán centrado en formar a maestros de primaria en la enseñanza de la invención de problemas, concluyendo que la forma en que los maestros organizan las tareas, cómo interactúan con los estudiantes y la forma en que los niños trabajan con sus compañeros en grupo contribuye a mejorar el aprendizaje (Leung, 2016).

Dejando a un lado la formación continua, otros trabajos se centran en el desarrollo profesional relativo a la resolución de problemas que se lleva a cabo en la formación inicial de docentes secundaria (e.g., Gourdeau, 2019; Mellone et al., 2021). En esta línea, Mellone et al. (2021) analizaron los cambios en las concepciones y creencias de los futuros profesores de secundaria sobre la resolución de problemas después de un curso de Matemáticas. A través de distintos instrumentos como encuestas y entrevistas, mostraron que las concepciones y creencias que tienen sobre resolución de problemas están condicionadas por sus experiencias y por el contexto cultural en el que viven.

También integrado en un entorno de formación inicial de profesores, Gourdeau (2019) describe un curso para futuros profesores de secundaria en Canadá. En él, la resolución de problemas ocupa un papel central ya que su objetivo es que los futuros profesores experimenten la resolución de problemas en un nivel profundo y comprendan los procesos involucrados. Para ello, además de cuidar la motivación y el deseo personal por resolver el problema, considera cinco aspectos: comprender o tratar de comprender procesos y objetos, extraer características y definirlas, encontrar patrones y estructuras, representar, y comunicar sobre nuestra comprensión.

Particularmente en España, el diseño del plan de formación inicial de maestros de Educación Primaria queda recogido en diversos documentos oficiales de nivel nacional, institucional y formador, los cuales establecen distintos niveles de concreción (Rico et al., 2014). A nivel nacional, se establecen de manera vinculante el número de créditos del

título y su reparto entre los distintos módulos que lo conforman, como el módulo de formación didáctico disciplinar con 100 ECTS que incluye la enseñanza y aprendizaje de la Matemática, Ciencias Sociales, Ciencias Experimentales, Lenguas, Educación física y Educación musical y plástica y visual (O. ECI/3857/2007, de 27 de diciembre).

A nivel institucional, cada universidad concreta las directrices nacionales a través del diseño y estructura del plan de estudios, así como los descriptores y créditos propios de cada materia por lo se establecen tanto el número de asignaturas como su diseño.

En un último nivel de concreción, los correspondientes departamentos universitarios desarrollan los programas de las asignaturas (denominados como guías docentes) ateniéndose a lo dispuesto en los documentos nacionales e institucionales. Las guías docentes contienen información general (número de créditos ECTS, semestre, departamento, carácter, profesorado responsable o bibliografía) y aspectos relativos a la planificación de la materia, los cuales pueden ser organizados en los niveles curriculares de expectativas de aprendizaje (finalidades que establecen prioridades en el aprendizaje), contenidos (menciones que organizan el conocimiento a través de conceptos, procedimientos o actitudes), metodología (estrategias, acciones, técnicas de enseñanza, tareas o materiales necesarios que en su conjunto indican cómo se llevará a cabo la enseñanza) y evaluación (pautas, sistemas o técnicas cuyo fin es identificar aquello que han aprendido los estudiantes), coincidiendo con los niveles básicos del currículo propuestos por Tyler (1986). En este estudio nos situamos en este último nivel de concreción.

Algunas universidades incluyen asignaturas en la formación inicial de maestros de primaria que desarrollan el conocimiento profesional sobre la RP, en las que se contempla la enseñanza de estrategias o heurísticos, tipos de problemas y aspectos metacognitivos entre otros. Sin embargo, hasta el momento no hemos encontrado indicios de estudios que se centren en la enseñanza de la RP que se desarrolla en el actual plan de formación inicial de maestros en España.

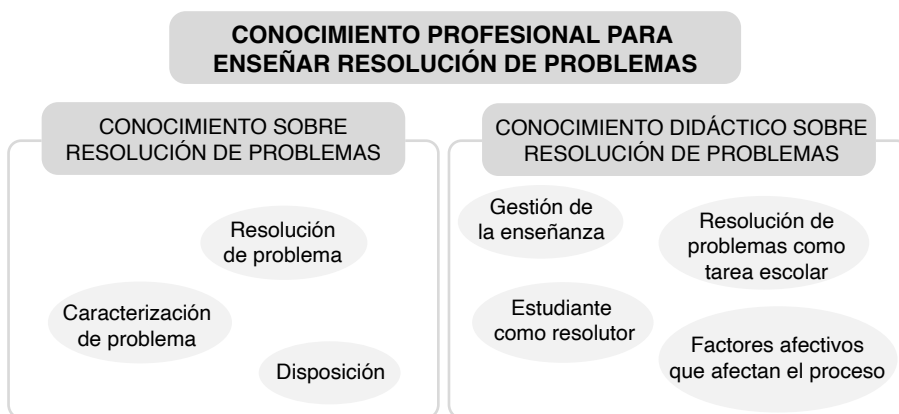
Conocimiento profesional para enseñar resolución de problemas

En este trabajo tomamos una perspectiva del conocimiento basada en categorías. En este contexto, utilizamos dos constructos teóricos para describir el conocimiento necesario para enseñar a resolver problemas en educación primaria: la competencia para resolver problemas y el

triángulo didáctico. Este último es empleado por su potencialidad para desentrañar aspectos del acto de enseñar (Schoenfeld, 2013). Además, recurrimos a dos perspectivas que permiten delimitar las categorías. Un primer elemento es el marco propuesto por Chapman (2015) para enseñar a resolver problemas, el *Mathematical problem-solving knowledge for teaching*. Un segundo aspecto son las directrices curriculares de diversos países (Piñeiro et al., 2022) pues delimitan el conocimiento de los profesores (Ball et al., 2008). Específicamente, presentamos una propuesta relativa a la enseñanza y aprendizaje de la RP² que destaca elementos del conocimiento del contenido y el conocimiento didáctico o pedagógico vinculados a la RP tomando en consideración lo que los profesores deben enseñar en las escuelas sobre RP (gráfico 1).

En este modelo, se diferencia entre competencia para resolver problemas y el conocimiento profesional (o teórico) sobre resolución de problemas. La competencia para resolver problemas “implica la capacidad de formular, representar y resolver problemas matemáticos; la disposición productiva incluye creencias; y el razonamiento adaptativo incluye la capacidad de pensamiento lógico y reflexión” (Chapman 2015, p. 21). Por su parte, los conocimientos profesionales corresponden al conocimiento necesario para desempeñar la tarea de enseñar y que en los modelos de conocimiento tradicionales son usualmente diferenciados entre conocimiento del contenido y didáctico del contenido.

Gráfico 1. Componentes del conocimiento profesional para enseñar RP



Fuente: Piñeiro et al. (2019)

² Esta propuesta de conocimiento profesional para enseñar resolución de problemas se encuentra detallada en Piñeiro et al. (2019). Dicho trabajo está dedicado a la descripción de este modelo.

El primer componente de este modelo, el conocimiento sobre la RP, se configura en torno a los conocimientos sobre: caracterización de problema, resolución y disposición.

La categoría caracterización o noción de problema alude a los conocimientos necesarios para reconocer si una tarea constituye o no un problema, i.e., que una tarea sea considerada problema, o no, depende del resolutor (consideración del resolutor); que los resolutores no posean un procedimiento predeterminado ni directo para resolver la tarea (tarea sin procedimiento de resolución conocido); y finalmente, que una tarea sea considerada un problema depende de su estructura, i.e., tipos de tareas que se presentan como problemas.

Los conocimientos relativos a la categoría de resolución de problemas engloban aspectos teóricos específicos como: a) el conocimiento de las etapas o fases de RP, esto es, comprender el problema, establecer un plan, ejecutarlo y evaluarlo (Polya, 1981); b) estrategias de RP (estrategias heurísticas); c) el conocimiento de las formas en las que los resolutores de problemas se autorregulan, monitorean y controlan heurísticos, estrategias específicas y conocimientos matemáticos, permitiendo tomar decisiones acertadas sobre lo que se hace (metacognición); d) conocimiento sobre la influencia del dominio afectivo en la RP, pues en función de la idoneidad del reto planteado, se ponen en juego emociones, lo que a su vez moviliza el intelecto (aspectos no cognitivos).

En relación a la disposición, para resolver un problema se debe aceptar el desafío de resolverlo. Particularmente, nos referimos al hecho de que no es necesario estar motivado para resolver un problema con éxito, pero si se debe aceptar el compromiso de realizar la tarea.

El segundo componente, el conocimiento didáctico sobre la RP, tiene como eje articulador el triángulo didáctico y sus interacciones, por lo que en él distinguimos cuatro elementos:

Conocimiento del estudiante como resolutor de problemas. La interacción entre estudiante y RP se corresponde con un conocimiento del maestro que permite poner atención en lo que pueden hacer, qué puede interferir y hasta dónde se puede desarrollar la habilidad de RP de los estudiantes.

Conocimiento sobre la RP como tarea escolar. La interacción entre el maestro y la RP se corresponde con conocimientos sobre la tarea etiquetada como problema y su proceso de resolución. Esto se traduce en sostener conocimiento para seleccionar problemas, estrategias y modelos de RP, y los beneficios de la invención de problemas (dónde y cuándo plantearlos).

Conocimiento sobre los factores no cognitivos que afectan a la RP. El tercer componente tiene relación con la interacción entre el profesor y el

estudiante en términos de los factores no cognitivos (emociones, actitudes y creencias) de cada uno en torno a la RP. En este sentido, es necesario proponer situaciones que favorezcan creencias relacionadas con el tiempo necesario para resolver un problema, las opciones de resolución múltiple o la posibilidad de varias respuestas, así como ser conscientes de que la ansiedad y frustración son emociones normales en el proceso de resolución de un problema, al igual que el disfrute y la satisfacción al resolverlos.

Conocimiento sobre la gestión de la RP. Un cuarto componente supone la relación triple entre el profesor, el estudiante y la RP. En él se incluye el conocimiento de los enfoques de enseñanza de la RP, la gestión del discurso, la gestión de los bloqueos de los estudiantes, la gestión de la evaluación de la RP y la gestión de los recursos tanto manipulables como intangibles.

Método

El estudio presenta un enfoque cualitativo, específicamente realizamos un análisis de documentos (Bowen, 2009) de todas las guías docentes de las asignaturas del área de Didáctica de la Matemática pertenecientes a los títulos de grado universitarios que habilitan para el ejercicio de la profesión de Maestro en Educación Primaria de las 39 universidades públicas españolas. Identificamos 151 asignaturas del área de Didáctica de la Matemática, de las cuales 114 son de carácter básico u obligatorio y 37 de carácter optativo. Particularmente, las guías pertenecen al curso académico 2022-2023, y dado que son documentos públicos todas estaban accesibles. Las unidades de análisis fueron aquellos fragmentos de las guías con alguna mención a la RP pertenecientes a los apartados de objetivos, competencias, contenidos, metodología y evaluación.

Categorías

Las categorías para el análisis de los documentos en este estudio atienden a las dimensiones: conocimiento para enseñar RP y niveles curriculares, obtenidas del marco teórico y presentadas anteriormente. En la primera dimensión, se incluyeron dos categorías extra. La primera de ellas denominada Tema, que incluye aquellos fragmentos que sólo contienen los términos Resolución de problemas sin mayor especificación. La segunda, Competencia de RP, en donde se codificaron los fragmentos que hacían mención a la habilidad o competencia para resolver problemas de los futuros maestros.

La segunda de las dimensiones, denominada niveles curriculares, hace referencia a la estructura de las guías docentes, e incluye las categorías

consideradas como fundamentales en la planificación de la enseñanza y en el currículo (Rico, 2013; Tyler, 1986): expectativas de aprendizaje (en las guías docentes son expresadas a través de competencias, objetivos de aprendizaje o resultados de aprendizaje), contenidos (en los documentos son presentados a través del contenido del programa o el temario de la asignatura), metodología y evaluación. Todas estas categorías pueden verse en la tabla I.

Tabla I. Categorías de análisis

Dimensión conocimiento para enseñar RP		
Competencia para resolver problemas		
Conocimiento sobre la RP	Noción de problema	Tarea sin procedimiento de resolución conocido Consideración del resolutor Tipos de tareas que se presentan como problemas
	RP	Fases de resolución y su caracterización Estrategias Metacognición Factores no cognitivos
	Disposición	Aceptación del desafío de resolver el problema
Conocimiento didáctico sobre la RP	Estudiante como resolutor	Características de resolutores exitosos Dificultades y errores
	RP como tarea escolar	Selección de problemas Modelos de resolución y estrategias Invencción de problemas
	Factores no cognitivos	Influencia mutua de creencias y concepciones
	Gestión de la enseñanza de la RP	Enfoques o vías de acceso Gestión del discurso Gestión de los atascos Gestión de la evaluación Gestión de los recursos
Tema		
Dimensión niveles curriculares		
Expectativas de aprendizaje		
Contenidos		
Metodología		
Evaluación		

Fuente: Elaboración propia

Dichas categorías, de tipo deductivo ya que están formuladas desde la teoría (Fraenkel et al., 2011), fueron sometidas a una evaluación mediante el juicio de expertos. Particularmente, tres investigadores del área con una larga trayectoria profesional en el tema. Todos coincidieron en que no era necesaria ninguna modificación, estableciéndose por tanto una total concordancia entre ellos.

Análisis de datos

El análisis de los datos obtenidos se llevó a cabo en dos fases. En la primera fase los documentos se codificaron en base a las categorías de análisis utilizando la herramienta MAXQDA. Por ejemplo, el objetivo “Dominar y aplicar criterios para inventar problemas de matemáticas dirigidos a Educación Primaria” fue codificado dentro de la categoría expectativas de aprendizaje en la dimensión de niveles curriculares y en la dimensión del conocimiento profesional para enseñar la RP en la categoría conocimiento didáctico sobre la RP-RP como tarea escolar-Invencción de problemas. Asimismo, este proceso no fue excluyente debido a que en una misma frase se aludían a diferentes conocimientos de acuerdo a nuestras categorías.

En primer lugar, dos investigadores codificaron el 5% de los documentos de manera independiente, alcanzando una índice kappa de 0,944 y siendo el porcentaje de acuerdo entre codificadores del 95%. Los desacuerdos fueron revisados por los investigadores hasta llegar a consensos para la codificación del resto de guías.

En la segunda fase, en primer lugar, calculamos estadísticos descriptivos del número de asignaturas obligatorias y optativas, así como sus créditos. En segundo lugar, calculamos tablas de frecuencias y estadísticos descriptivos de las categorías por cada área. En tercer lugar, tablas de concurrencias y matrices de relaciones entre los componentes del conocimiento profesional y los niveles curriculares. Una matriz de relaciones entre códigos es una representación visual de las intersecciones de dos códigos en un mismo enunciado (gráfico II). Finalmente, se realizó un análisis clúster considerando como variables los componentes del conocimiento profesional con el fin de obtener grupos de documentos. Se utilizó el método de similitud basado en el coeficiente de coincidencia simple. Este método considera la ocurrencia de códigos en el documento, pues solo se tiene en cuenta si se ha asignado un código a un documento, y define la similitud entre los atributos considerados como el número de coincidencias dividido por el número total de atributos.

Resultados

Los resultados son presentados en tres apartados. En el primero, describimos la situación general de las asignaturas del área de didáctica de la matemática en la titulación. En el segundo apartado, nos centramos en la presencia de los componentes del conocimiento profesional para enseñar RP. Finalmente, en el último apartado, mostramos los resultados del análisis clúster y describimos las características de los grupos de documentos formados.

Asignaturas del área de Didáctica de la Matemática en el grado de Educación Primaria

Como se aprecia en la tabla II, la mediana del número de asignaturas de carácter básico u obligatorio es 3, siendo en 6 instituciones sólo 2 asignaturas (ej., Universidad de Murcia o la Universidad de Vigo), mientras que 3 universidades presentan el máximo de 4 asignaturas (ej., Universidad de Huelva). La mediana de créditos totales por institución es 18, el mínimo es 12 (ej., Universidad de León o la Universidad de Vigo) y el máximo 27 (ej., Universidad de Cádiz).

Tabla II. Estadísticos descriptivos del número de asignaturas obligatorias y optativas y créditos del grado de Maestro en Educación Primaria del área de Didáctica de la Matemática por universidad

	Asignaturas						Créditos					
	N	Mín	Máx	Me	\bar{x}	σ	N	Mín	Máx	Me	\bar{x}	σ
Obligatorias	114	2	4	3	2,9	0,6	724	12	27	18	18,5	2,368
Optativas	37	0	4	1	0,9	1,17	197	0	24	3	4,9	6,3

Nota: N= total; Mín=mínimo; Máx=máximo; Me= mediana; \bar{x} =media; σ =desviación estándar.
Fuente: Elaboración propia

El número de créditos de las asignaturas obligatorias varía desde 3 hasta 12, constando en su mayoría (82,5%) de 6 créditos (ver tabla III).

Tabla III. Frecuencia de créditos en las asignaturas obligatorias y optativas del grado de Maestro en Educación Primaria pertenecientes al área de Didáctica de la Matemática

Obligatorias			Optativas		
Nº ECTS	Frecuencia	Porcentaje	Nº ECTS	Frecuencia	Porcentaje
3	1	0,9	3	7	18,9
5	3	2,6	4	2	5,4
6	94	82,5	4,5	2	5,4
7	2	1,8	6	25	67,6
8	1	0,9	9	1	2,7
9	12	10,5			
12	1	0,9			
TOTAL	114	100,0	TOTAL	37	100

Fuente: Elaboración propia

Como se observa en la tabla IV, estas asignaturas se imparten desde primero hasta cuarto, predominando en el segundo y tercer curso.

Tabla IV. Frecuencia del curso en donde se ofertan las asignaturas obligatorias y optativas del grado de Maestro en Educación Primaria pertenecientes al área de Didáctica de la Matemática

	Obligatorias		Optativas	
	Frecuencia	Porcentaje	Frecuencia	Porcentaje
1	20	17,5	4	10,8
2	41	36	2	5,4
3	39	34,2	9	24,3
4	14	12,3	22	59,5
TOTAL	114	100	37	100

Fuente: Elaboración propia

En el caso de las asignaturas optativas, aproximadamente la mitad de las instituciones, 20 de las 39 universidades, presentan alguna asignatura optativa del área de Didáctica de la Matemática, siendo la mediana de asignaturas 1 (tabla II). Encontramos 3 instituciones que presentan el máximo 4 (ej., Universidad de Valencia) y 19 que no presentan ninguna asignatura optativa en su plan de formación (ej., Universidad de Alcalá o la Universidad de Burgos). El número total de créditos por institución

presenta gran variación, oscilando desde 0 hasta 24, siendo 3 ECTS la mediana (tabla II). En relación a los créditos de cada asignatura (tabla III), encontramos que el 67,6% suelen presentar 6 ECTS, y suelen ofertarse en cuarto curso (tabla IV).

Por otra parte, observamos dos formas o modelos de organizar la materia. Un primer modelo que consta de asignaturas centradas fundamentalmente en el desarrollo del conocimiento del contenido matemático y asignaturas que se focalizan en el desarrollo del conocimiento didáctico de los contenidos. Un ejemplo de ello es la Universidad de Sevilla que presenta dos asignaturas de carácter obligatorio: Matemáticas Específicas para Maestros, y Didáctica de Matemáticas para Maestros. Un segundo modelo combina ambos tipos de conocimiento en todas las asignaturas, las cuales se separan en función del contenido matemático que trabajan. Por ejemplo, la Universidad de las Palmas de Gran Canaria que contiene tres asignaturas de carácter obligatorio: Matemáticas y su didáctica I (Números y el álgebra), Matemáticas y su didáctica II (Medida y la geometría), y Matemáticas y su didáctica III (Estadística y la probabilidad).

Componentes del conocimiento para enseñar RP en las guías docentes

Respecto al número de veces que aparece cada una de las categorías del conocimiento (ver tabla V), encontramos que la mayoría de los documentos contienen alguna mención a la RP, siendo sólo un 5% de las asignaturas obligatorias y un 16% de las optativas las que no presentan ninguna mención. Cabe destacar que hemos encontrado 1 asignatura obligatoria de la Universidad de Almería y 3 optativas centradas en la RP de las Universidades Autónoma de Madrid, da Coruña y de Girona, aunque esta última dejó de ofertarse.

Al distinguir entre tipos de conocimiento, no existen grandes diferencias en su presencia en las guías de las asignaturas obligatorias, siendo algo superior las menciones a la competencia RP con un 37% de enunciados codificados, seguido del conocimiento sobre la RP con un 33,6% y el conocimiento didáctico sobre RP con un 28,2%. En las asignaturas optativas, la distribución varía, ya que estas contienen un mayor porcentaje de enunciados codificados relativos al conocimiento didáctico y la competencia de RP (tabla V).

Tabla V. Frecuencia de componentes del conocimiento para enseñar RP en las guías docentes de las asignaturas obligatorias y optativas

	Obligatorias		Optativas	
	Documentos N=114	Enunciados codificados N=1092	Documentos N=37	Enunciados codificados N=160
Competencia RP	103 (90,4%)	412 (37,7%)	26 (70,3%)	60 (37,5%)
Conocimiento sobre la RP	100 (87,8%)	367 (33,6%)	17 (45,9%)	31 (19,4%)
Conocimiento didáctico sobre la RP	103 (90,4%)	308 (28,2%)	23 (62,2%)	66 (41,3%)
Tema	5 (4,4%)	5 (0,5%)	3 (8%,1)	3 (1,9%)
Documentos con código(s)	108 (95%)		31 (83,8%)	
Documentos sin código(s)	6 (5%)		6 (16,2%)	

Fuente: Elaboración propia

Sobre la presencia de los componentes en los niveles curriculares, encontramos que las categorías (competencia para resolver problemas y conocimiento profesional –conocimiento sobre la RP y conocimiento didáctico sobre la RP³) fundamentalmente en los dos primeros niveles (expectativas de aprendizaje y contenidos), mientras que la competencia destaca en los niveles de expectativas de aprendizaje y metodología (ver gráfico II). Las directrices nacionales (O. ECI/3857/2007, de 27 de diciembre) establecen como competencia general del área de Matemáticas «Plantear y resolver problemas vinculados con la vida cotidiana» (p. 53759), por lo que la competencia de RP aparece mayormente como expectativas de aprendizaje, aunque en algunos casos la encontramos particularizada a un campo concreto (números, medida, geometría,..) o sin la mención a la vida cotidiana (ej., “Ser capaz de resolver problemas que impliquen la conexión entre diferentes bloques de contenidos”).

Los contenidos detallan distintos conocimientos relativos a la RP a través del temario de la asignatura, tanto teórico como práctico. No obstante, en algunas guías docentes aparece un tema denominado “Resolución de problemas” pero no manifiesta mayor especificación.

En general, los niveles de metodología y evaluación suelen estar centrados en describir aspectos organizativos de la asignatura lo que explica la baja presencia de los componentes del conocimiento en ellos. Específicamente, cuando la RP se incluye dentro de la metodología de las guías, se suele incluir como sesiones prácticas destinadas a la RP, como parte

³ Para efectos del gráfico hemos agrupado las subcategorías.

del desarrollo de sesiones teóricas, o en algunos casos se menciona el uso de la metodología de aprendizaje basada en problemas en combinación con otras metodologías.

Por último, los documentos que mencionan la RP en el apartado de evaluación, la incluyen como parte de las pruebas escritas donde se afirma que los estudiantes deberán resolver problemas, por ejemplo «Examen de preguntas conceptuales, ejercicios y problemas a final del curso» presente en la asignatura Matemáticas de la Educación Primaria de la Universidad de las Islas Baleares. Además, de formar parte de las actividades evaluables que los estudiantes realizan durante el curso.

Gráfico II. Matriz de relaciones entre los componentes y los niveles curriculares.

Sistema de códigos	Expectativas de aprendizaje	Contenidos	Metodología	Evaluación
Competencia RP	220	55	126	63
Conocimiento sobre la RP	156	177	49	16
Conocimiento didáctico sobre la RP	209	102	39	14

Fuente: Elaboración propia

Conocimiento sobre la RP

Centrándonos en las categorías de conocimiento sobre RP, observamos en la tabla VI una tendencia similar entre las asignaturas obligatorias y optativas, pues la categoría con mayor presencia en las guías analizadas es tipos de tareas (por ejemplo, los contenidos de la guía docente de la asignatura Bases Matemática para la Educación Primaria de la Universidad de Granada incluyen “Tipos de problemas aritméticos”), y en menor medida encontramos las categorías tarea sin procedimiento de resolución conocido (por ejemplo, el tema RP de la asignatura Matemáticas y su Didáctica I, de la Universidad del País Vasco, incluye “Distinción entre problema y ejercicio”), fases de resolución (por ejemplo, “Fases en la resolución de un PAEV” asignatura RP y Conexiones Matemáticas de la Universidad de Almería) y estrategias (por ejemplo “Estrategias heurísticas” de la Universidad de Barcelona dentro del bloque temático de RP de la asignatura Razonamiento y Actividad Matemática en Primaria). Específicamente, la mayoría de las guías docentes de las asignaturas de carácter obligatorio incluyen distintos tipos de tareas que se pueden presentar como problemas (95%). Las menciones sobre la noción de problema como una tarea sin procedimiento de resolución conocido, las fases de resolución y las estrategias se encuentran aproximadamente en un 26% de los documentos, mientras que la presencia del resto de categorías en estas asignaturas es casi nula con un porcentaje menor al 3%.

En relación con las asignaturas optativas, el panorama es similar, pero con una presencia significativamente menor en relación a este conocimiento. La categoría mayoritaria continúa siendo tipos de tareas presente en un 35% de los documentos, seguido de la tarea sin un procedimiento conocido (13,5%), las fases de resolución (10,8%) y las estrategias (8,1%). El resto de categorías no están presentes en ninguna guía docente de las asignaturas optativas.

Tabla VI. Frecuencia de componentes de conocimiento sobre la RP

		Obligatorias		Optativas	
		Documentos N=114	Enunciados codificados N=1092	Documentos N=37	Enunciados codificados N=160
Noción de problema	Tarea sin procedimiento de resolución conocido	31 (27,2%)	48	5 (13,5%)	5
	Consideración del resolutor	2 (1,8%)	3	0	0
	Tipos de tareas que se presentan como problemas	95 (83,3%)	258	13 (35,1%)	16
RP	Fases de resolución y su caracterización	29 (25,4%)	46	4 (10,8%)	6
	Estrategias	30 (26,3%)	57	3 (8,1%)	4
	Metacognición	3 (2,6%)	6	0	0
	Factores no cognitivos	3 (2,6%)	5	0	0
Disposición	Aceptación del desafío de resolver el problema	2 (1,8%)	2	0	0

Fuente: Elaboración propia

Conocimiento didáctico sobre RP

La invención de problemas es la categoría del conocimiento didáctico sobre RP con mayor presencia en las guías (tabla VII), hecho razonable ya que, como comentamos anteriormente, atendiendo a las directrices nacionales, una competencia general del área de Matemáticas es “Plantear y resolver problemas vinculados con la vida cotidiana” (O. ECI/3857/2007, de 27 de diciembre, p. 53759). No obstante, sólo la encontramos en un 66,7% de las guías de las asignaturas obligatorias.

La segunda categoría con mayor presencia en ambos grupos de asignaturas es la gestión del discurso, un ejemplo es la competencia básica “Que los estudiantes puedan transmitir información, ideas, problemas y soluciones a un público tanto especializado como no especializado” de la asignatura Matemáticas y su Didáctica III de la Universidad de Salamanca.

En los documentos de las asignaturas obligatorias, también encontramos mención a la selección de problemas (por ejemplo, el resultado de aprendizaje “Extraer problemas de la vida cotidiana para plantear situaciones de aprendizaje” de la asignatura El Conocimiento Matemático en la Educación Primaria II, de la Universidad de Cádiz), los modelos de resolución y estrategias (por ejemplo, el contenido “análisis de estrategias de resolución” Matemáticas y su Didáctica I, de la Universidad de La Rioja) y los enfoques o vías de acceso (como el contenido “pautas para la aplicación de conocimientos y RP en el aula de Primaria” de la asignatura Matemáticas y su Didáctica I de la Universidad Autónoma de Madrid).

Tabla VII. Frecuencia de componentes de conocimiento didáctico sobre RP

		Obligatorias		Optativas	
		Documentos	Enunciados codificados	Documentos	Enunciados codificados
Estudiante como resolutor	Características de resolutores exitosos	5 (4,4%)	5	4 (10,8%)	5
	Dificultades y errores	10 (8,8%)	15	3 (8,1%)	4
RP como tarea escolar	Selección de problemas	28 (24,6%)	42	7 (18,9%)	8
	Modelos de resolución y estrategias	22 (19,3%)	54	3 (8,1%)	6
	Inventación de problemas	76 (66,7%)	125	9 (24,3%)	18
Factores no cognitivos	Influencia mutua de creencias y concepciones	6 (5,3%)	6	0	0

	Obligatorias		Optativas	
	Documentos	Enunciados codificados	Documentos	Enunciados codificados
Enfoques o vías de acceso	19 (16,7%)	34	8 (21,6%)	15
Gestión de la enseñanza de la RP	48 (42,1%)	71	9 (24,3%)	15
Gestión de los atascos	0	0	2 (5,4%)	3
Gestión de la evaluación	1 (0,9%)	2	1 (2,7%)	1
Gestión de los recursos	11 (9,6%)	14	8 (21,6%)	9

Fuente: Elaboración propia

Guías docentes y componentes del conocimiento

Como se ha visto en el apartado anterior, la mayoría de los documentos contienen alguna mención a la RP. Sin embargo, en ellas no hay menciones a cada uno de los componentes. Con el fin de obtener agrupaciones de guías docentes que presentan similitudes hemos realizado un análisis clúster. Dado que en la herramienta MAXQDA requiere precisar el método y número de clúster a priori, es decir, utiliza métodos no jerárquicos, el análisis se hizo aumentando el número de clúster hasta comprobar que los nuevos agrupamientos se producían dividiendo los grupos minoritarios. Dicho procedimiento nos llevó a identificar los clústeres presentados en las tablas VIII y IX. La primera de ellas muestra los 4 grupos de asignaturas obligatorias formados tras el análisis.

El primero clúster agrupa a un único programa, el de la asignatura Introducción a la Enseñanza y el Aprendizaje de las Matemáticas en Educación Primaria de la Universidad de Huelva, que incluye en su documento todas las categorías a excepción de la disposición del resolutor y la gestión de la enseñanza. El segundo grupo (clúster 2) se formó con los siete programas que solo tienen enunciados referidos a la competencia para resolver problemas, no hay mención a otros conocimientos como ocurre en el resto. El tercer clúster agrupa siete documentos, y es el único que incluye la categoría de disposición del resolutor, además incluye la gestión de la enseñanza sobre RP en todos sus programas. Finalmente, el clúster 4, y mayoritario, se formó con los 99 programas restantes que no contienen mención a la categoría de disposición del resolutor ni a los factores no cognitivos.

Tabla VIII. Clúster asignaturas obligatorias

		Clúster 1 N=1	Clúster 2 N=7	Clúster 3 N=7	Clúster 4 N=99
Competencia resolución problemas		1 (100%)	1 (14%)	7 (100%)	94 (95%)
Conocimiento sobre la RP	Noción de problema	1 (100%)	0	7 (100%)	91 (92%)
	RP	1 (100%)	0	7 (100%)	33 (33%)
	Disposición	0	0	2 (28.5%)	0
Conocimiento didáctico sobre la RP	Estudiante como resolutor	1 (100%)	0	4 (57%)	6 (6%)
	RP como tarea escolar	1 (100%)	0	7 (100%)	79 (80%)
	Factores no cognitivos	1 (100%)	0	5 (71.5%)	0
	Gestión de la enseñanza de la RP	0	0	7 (100%)	58 (59%)

Fuente: Elaboración propia

La tabla IX muestra los 3 clúster formados a partir de las asignaturas optativas. Como podemos observar, no hay ninguna asignatura optativa que incluya la categoría disposición del resolutor ni factores no cognitivos. El primero de ellos (clúster 1) agrupa cinco programas, y se caracteriza por no incluir la categoría gestión de la enseñanza. El segundo grupo (clúster 2) se formó a partir de los diez programas, y se caracteriza por tener menciones a la categoría RP. Finalmente, el clúster 3, y mayoritario, se formó con los programas que solo incluyen la competencia para resolver problemas, la noción de problema, la RP como tarea escolar y la gestión de la enseñanza.

Tabla IX. Clúster asignaturas optativas

		Clúster 1 N=5	Clúster 2 N=10	Clúster 3 N=22
Competencia resolución problemas		5 (100%)	10 (100%)	11 (50%)
Conocimiento sobre la RP	Noción de problema	5 (100%)	9 (90%)	2 (9%)
	RP	0	5 (50%)	0
	Disposición	0	0	0

	Clúster 1 N=5	Clúster 2 N=10	Clúster 3 N=22	
Conocimiento didáctico sobre la RP	Estudiante como resolutor	1 (20%)	3 (30%)	0
	RP como tarea escolar	2 (40%)	9 (90%)	5 (23%)
	Factores no cognitivos	0	0	0
	Gestión de la enseñanza de la RP	0	10 (100%)	9 (41%)

Fuente: Elaboración propia.

Discusión y conclusiones

En este trabajo damos respuesta a dos objetivos: describir la presencia de las asignaturas del área de Didáctica de la Matemática pertenecientes los títulos universitarios oficiales que habilitan para el ejercicio de la profesión de Maestro en Educación Primaria en España; y describir la presencia de la RP y su enseñanza en las guías docentes de las asignaturas del área de Didáctica de la Matemática pertenecientes a dichos títulos.

En relación al primer objetivo, las directrices nacionales (O. ECI/3857/2007, de 27 de diciembre) contemplan dentro del módulo Didáctico Disciplinar, el área de Matemáticas. Estas directrices no especifican el número de asignaturas y sus créditos, sólo el número de créditos totales (100 ECTS) de dicho módulo, por lo que, a nivel global la mediana de asignaturas obligatorias y optativas es 3 y 1, y de créditos es 18 y 3 respectivamente. En este sentido, a pesar de que el número de créditos ha aumentado respecto al plan de formación de maestros de primaria previo, consideramos que sigue siendo insuficiente debido a la alta presencia que tiene la materia de matemáticas en el nivel de educación primaria (RDL 157/2022, de 1 de marzo), junto con los bajos resultados que obtienen los maestros en formación en pruebas e investigaciones como TEDS-M (INEE, 2012). Además, destaca que sólo la mitad de las universidades ofertan alguna asignatura optativa relativa al área, por lo que durante sus estudios universitarios los futuros maestros no tienen la posibilidad de aumentar su formación relativa a la Didáctica de la Matemática.

En relación a la forma de organizar la materia en las asignaturas obligatorias hemos identificado que las universidades tienden diseñar el itinerario de dos formas diferentes. Un primer modelo en que las asignaturas están centradas fundamentalmente en el desarrollo del conocimiento del contenido matemático y asignaturas que se focalizan en el desarrollo del conocimiento didáctico de los contenidos. En este primer modelo,

podemos ver reflejados modelos del conocimiento del profesor de matemáticas como el de Ball et al. (2008) o el de Carrillo et al. (2018), que separan en dos dominios el conocimiento del contenido y el conocimiento didáctico del contenido. Un segundo modelo que combina ambos tipos de conocimiento en todas las asignaturas, las cuales se separan en función del contenido matemático que trabajan, aunque cabe la posibilidad de centrarse únicamente en aspectos del contenido dejando de lado la didáctica.

En cuanto al segundo objetivo, encontramos que la mayoría de las guías docentes contienen alguna mención a la RP, especialmente en las asignaturas obligatorias y dentro de los apartados de competencias, objetivos y/o contenidos. En este sentido encontramos diferencias con los planes de formación de otros países como Chile (Felmer y Perdomo-Díaz, 2017), donde la RP está presente mayormente en los objetivos y metodología.

Los documentos incluyen casi siempre los distintos tipos de tareas que se presentan como problemas y fomentan fundamentalmente la competencia de resolver problemas ya que es una competencia propuesta a nivel nacional (O. ECI/3857/2007, de 27 de diciembre): “Plantear y resolver problemas vinculados con la vida cotidiana”. Sin embargo, destaca que el planteamiento de problemas es muchas veces suprimido, dejando las menciones al conocimiento didáctico sobre RP en las asignaturas optativas. Además, encontramos que aspectos considerados relevantes por la literatura en el conocimiento sobre RP (Chapman, 2015), como la consideración del resolutor, los factores no cognitivos o la disposición del resolutor, y aspectos del conocimiento didáctico sobre RP como gestión de los atascos, evaluación y recursos se encuentran ausentes de los documentos de ambos tipos de asignaturas.

Los resultados obtenidos en el análisis clúster de las asignaturas obligatorias muestran que existen programas focalizados en aspectos diferentes. Específicamente, sólo el tercer clúster considera al resolutor, lo cual es decisivo en la enseñanza de la resolución de problemas pues si no considera eso, se corre el riesgo de que las tareas que se planteen no sean verdaderos problemas. El clúster dos solo considera la competencia, aspecto a destacar pues ser buen resolutor de problemas no asegura ser un buen profesor. En relación con los clústeres obtenidos en las asignaturas optativas, el primer clúster se enfoca en las tareas que se usan como problemas, sus dificultades y errores y criterios para seleccionarlas. El segundo considera un mayor número de conocimientos, pero deja de lado los factores no cognitivos aspecto crítico en la enseñanza de la RP pues determinarán la forma en que el resolutor afronte los problemas,

por ejemplo, los estudiantes cuya experiencia con la resolución de problemas constó de ejercicios que resolvían en pocos minutos llegaron a creer que todos los problemas se resuelven en ese tiempo, y dejaron de trabajar en problemas que podrían haber resuelto (Schoenfeld, 2013).

En base a nuestros análisis observamos que, dado que a nivel nacional solo se establecen competencias que deben adquirirse, las guías docentes son muy diversas. Cada universidad ha interpretado la enseñanza y aprendizaje de la RP desde su propio entendimiento por lo que la formación relativa a este aspecto entre maestros españoles es muy heterogénea. Estos contrastes pueden provocar diferencias en el rendimiento de los futuros docentes en el aula, ya que como muestran estudios previos (Son y Lee, 2021) sus concepciones sobre RP están relacionadas con sus prácticas de enseñanza. Además, podrían continuar transmitiendo visiones tradicionales de la RP, ya que en gran parte de las guías docentes sólo se abordan aspectos procedimentales de la RP.

Dado que las nuevas directrices de educación primaria incluyen dedicar un tiempo del horario lectivo a la RP (RDL 157/2022, de 1 de marzo), y que la investigación muestra influencia entre los contenidos estudiados en la formación y el conocimiento especializado utilizado en la práctica (e.g., Morris y Hiebert, 2017) es necesario preparar a los futuros maestros en lo que respecta a su conocimiento profesional para la enseñanza de la RP. Esta preparación ha de ir más allá de una competencia a desarrollar (Contreras, 2022) y para ello, los nuevos programas de formación pueden tomar el modelo propuesto por Piñeiro et al., (2019) que detalla qué conocimiento profesional es necesario para la enseñanza de la RP, o tomar como ejemplo algunas de las guías docentes que contemplan gran parte de los conocimientos requeridos en dicho modelo como el programa de la asignatura Introducción a la Enseñanza y el Aprendizaje de las Matemáticas en Educación Primaria de la Universidad de Huelva o la asignatura RP y Conexiones Matemáticas de la Universidad de Almería.

Finalmente, este estudio aporta información sobre la formación inicial de los maestros de primaria relativa a la RP en España que puede ser útil para el diseño de futuros planes de formación. Sin embargo, somos conscientes de que en muchos casos las guías docentes son demasiado breves y desarrollan fundamentalmente aspectos organizativos, por lo que en investigaciones futuras es necesario ampliar la información obtenida con entrevistas a formadores, con el temario o apuntes y observaciones dentro del aula.

Referencias bibliográficas

- Ball, D. L., Thames, M. H. y Phelps, G. (2008). Content knowledge for teaching: What makes it special? *Journal of Teacher Education*, 59(5), 389-407. <http://dx.doi.org/10.1177/0022487108324554>
- Bowen, G. A. (2009). Document analysis as a qualitative research method. *Qualitative Research Journal*, 9(2), 27-40. <https://doi.org/10.3316/QRJ0902027>
- Britz, J. y Richard, N. (1992). *Problem solving in the early childhood classroom*. National Education Assn.
- Carrillo, J., Climent, N., Montes, M., Contreras, L. C., Flores-Medrano, E., Escudero-Ávila, D., Vasco, D., Rojas, N., Flores, P., Aguilar- González, Á., Ribeiro, M. y Muñoz-Catalán, M. C. (2018). The mathematics teacher's specialised knowledge (MTSK) model. *Research in Mathematics Education*, 20(3), 236-253. <https://doi.org/10.1080/14794802.2018.1479981>
- Contreras, L. C. (2022). La nueva propuesta curricular y la formación del profesor. En T. F. Blanco, C. Núñez-García, M. C. Cañadas y J. A. González-Calero (Eds.), *Investigación en Educación Matemática XXV* (pp. 63-79). SEIEM.
- Chapman, O. (2015). Mathematics teachers' knowledge for teaching problem solving. *LUMAT*, 3(1), 19-36. <https://doi.org/10.31129/lumat.v3i1.1049>
- Felmer, P. y Perdomo-Díaz, J. (2017). Un programa de desarrollo profesional docente para el nuevo currículo de matemática: la resolución de problemas como eje articulador. *Educación Matemática*, 29(1), 201-217. <https://doi.org/10.24844/em2901.08>
- Fraenkel, J. R., Wallen, N. E. y Hyun, H. H. (2011). *How to design and evaluate research in education*. McGraw-Hill.
- Gourdeau, F. (2019). Problem solving as a subject and as a pedagogical approach, and the ongoing dialogue between mathematics and mathematics education. En P. Felmer, P. Liljedahl y B. Koichu (Eds.), *Problem solving in mathematics instruction and teacher professional development* (pp. 23-42). Springer. https://doi.org/10.1007/978-3-030-29215-7_21
- Hähkiöniemi, M. y Francisco, J. (2019). Teacher guidance in mathematical problem-solving lessons: Insights from two professional development programs. En P. Felmer, P. Liljedahl y B. Koichu (Eds.), *Problem solving in mathematics instruction and teacher professional development* (pp. 279-296). Springer.

- Hiebert, J., Berk, D., Miller, E., Gallivan, H. y Meikle, E. (2019). Relationships between opportunity to learn mathematics in teacher preparation and graduates' knowledge for teaching mathematics. *Journal for Research in Mathematics Education*, 50(1), 23-50. <https://doi.org/10.5951/jresmetheduc.50.1.0023>
- Instituto Nacional de Evaluación Educativa. (2012). *TEDS-M Informe Español. Estudio Internacional sobre la formación en matemáticas de los maestros. IEA. Informe español*. Secretaría General Técnica. <https://www.educacionyfp.gob.es/dctm/inee/internacional/teds-mlinea.pdf?documentId=0901e72b8143866e>
- Lester, F. K. y Cai, J. (2016). Can mathematical problem solving be taught? preliminary answers from 30 years of research. En Felmer et al. (Eds.), *Posing and solving mathematical problems, research in mathematics education* (pp. 117-126). Springer. https://doi.org/10.1007/978-3-319-28023-3_8
- Leung, S. S. (2016). Mathematical problem posing: A case of elementary school teachers developing tasks and designing instructions in Taiwan. En P. Felmer et al. (Eds.), *Posing and solving mathematical problems, research in mathematics education* (pp. 327-344). Springer.
- Liljedahl, P. y Cai, J. (2021). Empirical research on problem solving and problem posing: A look at the state of the art. *ZDM*, 53(4), 723-735. <https://doi.org/10.1007/s11858-021-01291-w>
- Mellone, M., Pacelli, T. y Liljedahl, P. (2021). Cultural transposition of a thinking classroom: To conceive possible unthoughts in mathematical problem solving activity. *ZDM*, 53, 785-798. <https://doi.org/10.1007/s11858-021-01256-z>
- Morris, A. K. y Hiebert, J. (2017). Effects of teacher preparation courses: Do graduates use what they learned to plan mathematics lessons? *American Educational Research Journal*, 54(3), 524-567. <https://doi.org/10.3102/0002831217695217>
- Orden ECI/3857/2007, de 27 de diciembre, por la que se establecen los requisitos para la verificación de los títulos universitarios oficiales que habiliten para el ejercicio de la profesión de Maestro en Educación Primaria. *Boletín Oficial del Estado*, 312, de 29 de diciembre de 2007. <https://www.boe.es/eli/es/o/2007/12/27/eci3857>
- Piñeiro, J. L., Castro-Rodríguez, E. y Castro, E. (2019). Componentes de conocimiento del profesor para la enseñanza de la resolución de problemas en educación primaria. *PNA*, 13(2), 124-139. <https://doi.org/10.30827/pna.v13i2.7876>
- Piñeiro, J. L., Castro-Rodríguez, E. y Castro, E. (2022). What problem-solving knowledge is required in mathematical teaching? *A curricular*

- approach. Curriculum Perspective*, 42, 1-12. <https://doi.org/10.1007/s41297-021-00152-6>
- Pólya, G. (1981). *Cómo plantear y resolver problemas* (2a ed.). Trillas.
- Real Decreto 157/2022, de 1 de marzo, por el que se establecen la ordenación y las enseñanzas mínimas de la Educación Primaria. *Boletín Oficial del Estado*, 52, de 2 de marzo de 2022. <https://www.boe.es/eli/es/rd/2022/03/01/157/con>
- Rico, L. (2013). *Análisis Didáctico en Educación Matemática. Metodología de investigación, formación de profesores e innovación curricular* (pp. 23-58). Comares.
- Rico, L., Gómez, P. y Cañadas, M. (2014). Formación Inicial en educación matemática de los maestros de primaria en España, 1991-2010. *Revista de Educación*, 363, 35- 59. <https://doi.org/10.4438/1988-592X-RE-2011-362-152>
- Schoenfeld, A. H. (2013). Reflections on problem solving theory and practice. *The Mathematics Enthusiast*, 10(1-2), 9-34.
- Son, J. W. y Lee, M. Y. (2021). Exploring the relationship between preservice teachers' conceptions of problem solving and their problem-solving performances. *International Journal of Science and Mathematics Education*, 19, 129-150. <https://doi.org/10.1007/s10763-019-10045-w>
- Tatto, M. T. (2018). The mathematical education of primary teachers. En M. T. Tatto, M. C. Rodriguez, W. M. Smith, M. D. Reckase y K. Bankow (Eds.), *Exploring the mathematical education of teachers using TEDS-M data* (pp. 205-256). Springer. https://doi.org/10.1007/978-3-319-92144-0_8
- Tyler, R. (1986). *Principios básicos del currículo*. Troquel.
- Wake, G., Swan, M. y Foster, C. (2016). Professional learning through the collaborative design of problem-solving lessons. *Journal of Mathematics Teacher Education*, 19(2), 243-260. <https://doi.org/10.1007/s10857-015-9332-9>

Autonomía universitaria: un modelo analítico multinivel desde una perspectiva latinoamericana

University autonomy: a multilevel analytical model from a Latin American perspective

<https://doi.org/10.4438/1988-592X-RE-2025-412-735>

Mario Alarcón Bravo

<https://orcid.org/0000-0002-3939-724X>

Universidad Diego Portales

José Joaquín Brunner Ried

<https://orcid.org/0000-0002-7832-0237>

Universidad Diego Portales y Universidad de Tarapacá

Resumen

Este ensayo examina la autonomía universitaria como un concepto dinámico y multifacético, influenciado por factores políticos, económicos y sociales. Propone un modelo analítico que aborda este fenómeno desde tres niveles: macro (político-sistémico), meso (institucional) y micro (individual). A diferencia de enfoques tradicionales centrados en definiciones generales o estáticas, este modelo adopta una perspectiva relacional y dinámica, examinando las interacciones entre el Estado, el mercado, las organizaciones universitarias y los actores académicos individuales. Su principal contribución radica en identificar variables analíticas específicas para cada nivel, lo que facilita la comparación empírica de las dinámicas de autonomía en distintos contextos. El análisis a nivel macro se centra en la relación entre las universidades, el Estado y el mercado, destacando cómo las reformas basadas en la Nueva Gestión Pública han promovido un modelo de autonomía regulada que tensiona el equilibrio entre independencia institucional y rendición de cuentas. A nivel meso, se exploran las transformaciones en la gobernanza universitaria, marcadas por el desplazamiento del modelo colegiado hacia estructuras gerenciales, lo que ha generado tensiones en la distribución del poder interno. Por último, a nivel micro, se analiza cómo las presiones externas afectan la autonomía individual de los académicos, especialmente en sus actividades de investigación, docencia y participación profesional. El ensayo subraya que la autonomía universitaria en América Latina ha evolucionado his-

tóricamente, desde su consolidación con la Reforma de Córdoba en 1918 hasta los desafíos contemporáneos impuestos por la globalización, las políticas neoliberales y la mercantilización de la educación superior. Asimismo, destaca la necesidad de enfoques analíticos integrales que reconozcan las tensiones y dinámicas internas de las universidades en el contexto actual. Las reflexiones finales destacan la defensa de la autonomía universitaria como principio esencial para sociedades democráticas y equitativas, concebida como un constructo relacional en constante negociación con el entorno.

Palabras clave: nueva gerencia pública, autonomía universitaria, gerencialismo, profesión académica, libertad académica, modelo multinivel, Latinoamérica.

Abstract

This essay examines university autonomy as a dynamic and multifaceted concept influenced by political, economic and social factors. It proposes an analytical model that approaches the phenomenon from three levels: macro (political-systemic), meso (institutional) and micro (individual). Unlike traditional approaches that focus on general or static definitions, this model takes a relational and dynamic approach, examining the interactions between the state, the market, university organisations, and individual academic actors. The model's primary contribution is the identification of specific analytical variables for each level, facilitating the empirical comparison of autonomy dynamics in different contexts. The macro-level analysis focuses on the relationship between universities, the state and the market, highlighting how reforms based on New Public Management have promoted a model of regulated autonomy that strains the balance between institutional independence and accountability. At the meso level, it examines the changes in university governance marked by the shift from the collegiate model to managerial structures, which has created tensions in the distribution of internal power. Finally, at the micro level, it analyses how external pressures affect the individual autonomy of academics, particularly in their research, teaching and professional activities. The essay emphasises that university autonomy in Latin America has evolved historically, from its consolidation with the Cordoba Reform in 1918 to the contemporary challenges posed by globalisation, neoliberal policies and the commodification of higher education. It also highlights the need for comprehensive analytical approaches that recognise the internal tensions and dynamics of universities in the current context. The concluding reflections emphasise the importance of defending university autonomy as an essential principle for the development of democratic and just societies, stressing that it must be understood as a relational construct, subject to constant negotiation between universities and their environments.

Keywords: new public management, university autonomy, managerialism, academic profession, academic freedom, multilevel model, Latin America.

Introducción

Desde la década de 1980, la Nueva Gestión Pública (NGP) ha transformado la organización universitaria en todo el mundo (Leišytė, Marquina y Jones, 2025; Shattock, 2014; Veiga et al., 2018). Dichas transformaciones se plasman en rasgos convergentes: universidades más integradas y orientadas a la eficiencia; reequilibrio del poder entre unidades académicas y administrativas; creciente profesionalización de la gestión; designación—y no elección—de autoridades; incorporación de actores externos en los órganos de gobierno; centralización de las decisiones estratégicas; énfasis en la responsabilidad individual por sobre el trabajo colectivo y, finalmente, fortalecimiento de los cargos intermedios. Estos procesos han sido ampliamente documentados en Europa (Bruckmann y Carvalho, 2018; Krüger et al., 2018; Veiga et al., 2018), Asia (Da Wan et al., 2019), Norteamérica (Lavigne, 2018; Stephenson et al., 2025) y América Latina (Acosta, 2020; Brunner y Alarcón, 2024; Marquina et al., 2022).

El debate académico resultante se ha concentrado en la participación de los profesores en la gobernanza, las tareas de docencia, investigación, extensión y gestión, la internacionalización, la adopción de herramientas gerenciales y la satisfacción laboral (Teichler et al., 2022). Tales cuestiones se han abordado, sobre todo, desde los marcos del capitalismo académico (Brunner et al., 2022; Slaughter y Leslie, 1997) y de la teoría de las profesiones (Abbott, 1988). Los estudios comparados revelan variaciones en las condiciones laborales y el estatus de la profesión tanto en países de la OCDE (Kezar et al., 2019; Musselin, 2007; Teichler et al., 2022) como, de forma creciente, en América Latina (Alarcón, Brunner y Labraña, 2025; Brunner y Alarcón, 2024; Santin et al., 2022).

Las grandes encuestas internacionales han impulsado esta agenda: la Carnegie (1991–1993), la Changing Academic Profession (CAP, 2007) y la APIKS (2018–2019), que amplía el foco hacia la carrera y la gobernanza (Teichler et al., 2022) e incluye Argentina (Fuksman y Nosiglia, 2020), México (Estévez Nenninger et al., 2020) y Chile (Brunner y Alarcón, 2025). La profesionalización académica adopta rasgos disímiles en Alemania, Canadá, Estados Unidos, India, Reino Unido, Rusia, China, Japón y Brasil. Finkelstein y Jones (2017) identifican cuatro factores que explican esta diversidad: la centralidad conferida a la investigación, el grado de institucionalización de la carrera, el equilibrio entre gobierno colegial y gerencial, y la heterogeneidad interna de cada sistema.

En contraste, el impacto del gerencialismo sobre la autonomía y la libertad académica ha sido menos explorado. Estudios recientes resaltan la redistribución de poder y las tensiones entre directivos, académicos

y profesionales (Kenny, 2018). Los hallazgos son heterogéneos: persiste la influencia colegial en Países Bajos (Kallenberg, 2015), Portugal (Carvalho y Videira, 2019) y Argentina (Marquina et al., 2022), mientras que se registra pérdida de autonomía en Finlandia (Kallio, Kallio y Blomberg, 2020) y Francia (Mignot Gérard et al., 2023). Persisten, por ello, vacíos analíticos respecto de los factores que condicionan la autonomía universitaria, lo que convierte en imperativa la realización de estudios empíricos comparados (Carvalho y Videira, 2019) y la formulación de marcos conceptuales renovados (Pekkola et al., 2018).

Así, este ensayo propone un modelo analítico multinivel para el estudio de la autonomía universitaria desde una perspectiva latinoamericana. Se argumenta que la autonomía universitaria es un concepto complejo que debe ser analizado a partir de tres niveles: el político-sistémico (macro), el institucional (meso) y el individual (micro). La estructura del ensayo comprende cuatro secciones. En primer lugar, se describe el aparato conceptual utilizado para construir el marco analítico. A continuación, se analiza la evolución de la autonomía académica en América Latina desde la Reforma de Córdoba de 1918. Posteriormente, se presenta la propuesta del modelo analítico multinivel. Finalmente, se ofrecen reflexiones sobre las tensiones y los desafíos que enfrenta la autonomía universitaria en el contexto contemporáneo.

Marco conceptual

Autonomía universitaria

La autonomía universitaria se ha posicionado como principio rector de los sistemas de educación superior al articular la independencia institucional con la responsabilidad social: implica que cada universidad decida sobre su gobernanza, administración, docencia e investigación sin injerencias externas indebidas (Maassen et al., 2017). En Estados Unidos y Europa, esta autonomía se enlaza estrechamente con la libertad académica, concebida como el derecho individual de enseñar e investigar guiado por el rigor intelectual y la búsqueda de la verdad (Bernasconi, 2021; Lerch et al., 2024). En América Latina, por el contrario, la autonomía suele entenderse como un atributo corporativo que, en ciertos casos, dificulta la rendición de cuentas y la respuesta a demandas públicas (Bernasconi, 2024; Mendoza, 2020). Este contraste refleja trayectorias históricas divergentes: los sistemas anglosajones privilegian los derechos individuales y favorecen la flexibilidad e innovación, mientras los latinoamericanos subordinan la libertad académica a la autonomía institucional, reduciendo el margen de acción de docentes e investigadores.

La literatura distingue, por un lado, la autonomía sustantiva —capacidad de fijar fines, planes de estudio y prioridades investigativas— y la procedimental —facultad de definir los medios para alcanzarlos— (Kohtamäki y Balbachevsky, 2018; Maassen et al., 2017; Wen y Marginson, 2023). Por otro lado, Estermann y Steneil (2011) identifican cuatro dimensiones interdependientes: la organizativa, vinculada al diseño de estructuras internas y órganos de gobierno; la académica, relativa a la libertad para estructurar programas, metodologías y líneas de investigación; la financiera, referida a la gestión autónoma de recursos y a la capacidad de generar ingresos; y la de personal, centrada en la potestad de contratar, evaluar y promover al cuerpo académico y administrativo.

En la última década, el concepto ha evolucionado hacia la noción de “autonomía viva”, enfatizando que la autonomía efectiva depende menos de disposiciones legales que de prácticas de autogobierno y de la capacidad estratégica de las universidades para responder a su entorno (Maassen et al., 2017; Maassen, 2024). En Europa, las crecientes exigencias de gobiernos, sector privado y otros actores externos han reforzado los mecanismos de supervisión, mientras que, en América Latina, estas presiones se combinan con dinámicas internas de politización —la “universidad militante” descrita por Brunner (2015)— que pueden restringir tanto la autonomía institucional como la libertad académica (Mendoza, 2020).

Frente a este entramado de tensiones, resulta pertinente adoptar un enfoque analítico multinivel (macro, meso y micro) que permita examinar cómo se reconfigura el poder en la gobernanza universitaria y cuáles son las implicancias de estas transformaciones sobre la autonomía y la libertad académica contemporáneas (Kwiek, 2019).

Cambios a nivel macro: autonomía universitaria y la relación con el Estado y el mercado

En el nivel macro, la autonomía universitaria se entiende como la capacidad de las instituciones para autogobernarse, mantenerse al margen de la intervención estatal directa y actuar como guardianes del conocimiento. Durante los siglos XVII y XVIII, las universidades gozaban de considerable autonomía, operando como entidades independientes en lo que Clark (1983) denominó *torre de marfil*, siguiendo la idea de libertad y aislamiento enunciada por Humboldt. Este concepto sostiene que la búsqueda de la verdad requiere libertad de interferencias externas y cierto ensimismamiento para establecer agendas propias de investigación y enseñanza.

Con el surgimiento del Estado-nación en el siglo XIX y la expansión de las universidades modernas, esta autonomía comenzó a tensionarse.

Las universidades pasaron a formar parte de la estructura estatal, recibiendo financiación pública y participando en la formación de profesionales para el servicio público. Aunque mantenían cierta autonomía, la relación con el Estado se volvió más compleja, equilibrando financiación y libertad académica (Finkelstein y Jones, 2017).

Desde la década de 1980, las reformas inspiradas en la Nueva Gestión Pública (NGP) alteraron radicalmente la relación entre el Estado y las universidades (Veiga et al., 2018). Estas políticas, diseñadas para aumentar eficiencia y reducir costos, introdujeron mecanismos de control externo que limitaron la autonomía tradicional. Se implementaron sistemas de evaluación y rendición de cuentas, fomentando autorregulación pero imponiendo límites a la libertad institucional. Este nuevo régimen, conocido como “autonomía regulada”, ha sido ampliamente documentado tanto en Europa (Krüger et al., 2018) como en América Latina (Acosta-Silva, 2022; Alarcón y Brunner, 2024), y se caracteriza por permitir formas de autogestión circunscritas a marcos regulatorios definidos por el Estado.

En este escenario, la autonomía universitaria se ve tensionada por el control estatal y las crecientes presiones del mercado. La disminución de la financiación pública directa ha impulsado a las universidades a diversificar sus fuentes de ingreso, adoptando estrategias asociadas al capitalismo académico. Este marco conceptual, desarrollado originalmente por Slaughter y Leslie (1997) a partir de estudios en Estados Unidos, Reino Unido y Canadá, ha sido recientemente reinterpretado desde una perspectiva comparada que reconoce las variedades de capitalismo académico presentes en América Latina (Brunner et al., 2022).

Desde esta óptica, las universidades compiten por recursos externos, establecen alianzas con el sector productivo y buscan atraer estudiantes internacionales, lo que ha intensificado la lógica mercantil en la educación superior. Si bien este proceso puede ampliar la autonomía financiera institucional, también limita la capacidad de las universidades para definir de forma autónoma sus prioridades académicas y de investigación, como lo evidencian estudios tanto en Europa (Carvalho, 2017) como en América Latina (Pérez Mora et al., 2022).

Cambios a nivel meso: gobierno universitario y la redistribución del poder

En el nivel meso, la autonomía académica se articula dentro de las propias instituciones de educación superior, donde la distribución interna del poder ha sido alterada por las mismas reformas que afectaron el nivel macro. Tradicionalmente, la gobernanza universitaria se basaba en estructuras colegiadas, donde los académicos desempeñaban un rol cen-

tral en la toma de decisiones a través de senados y comités (Deem y Brehony, 2005; Carvalho y Santiago, 2015). En América Latina, este modelo se ha expresado históricamente en una fuerte valoración del autogobierno universitario y en amplios márgenes de autonomía institucional, concebidos como pilares de la universidad como institución social académica (Alarcón y Brunner, 2024).

Sin embargo, las reformas inspiradas en el gerencialismo introdujeron una estructura más jerárquica y profesionalizada, que trasladó el poder decisional hacia rectores y decanos, muchas veces seleccionados con base en criterios de gestión más que por trayectorias académicas (Carvalho, 2017). A diferencia del caso europeo, en América Latina estos cambios han sido más graduales y conflictivos, resultando en modelos híbridos que combinan prácticas gerenciales emergentes con elementos persistentes de la cultura académica colegial (Alarcón, 2020).

En paralelo, la profesionalización de la gestión ha dado lugar a la emergencia de una capa de administradores no académicos con poder decisional en asuntos clave. Esto ha provocado fricciones entre los cuerpos académicos y los equipos de gestión, así como una percepción generalizada de pérdida de control por parte del profesorado. No obstante, esta percepción tiende a ser menos pronunciada en América Latina que en Europa, como lo muestran estudios comparados recientes (Jones y Weinrib, 2019; Marquina et al., 2022).

Cambios a nivel micro: la fragmentación de la profesión académica y la pérdida de autonomía individual

En el nivel micro, la autonomía académica se refiere a la libertad individual de los académicos para definir sus propias agendas de investigación, métodos de enseñanza y actividades profesionales. Esta libertad ha sido erosionada por la fragmentación y estratificación de la profesión académica (Finkelstein y Jones, 2019). La fragmentación horizontal, producto de la especialización creciente de las actividades académicas, y la vertical, caracterizada por el aumento de categorías laborales precarias, han debilitado aún más la autonomía individual. El empleo precario obliga a los académicos a centrar sus esfuerzos en actividades que aseguren la renovación de contratos, restringiendo su capacidad de seguir sus propias líneas de investigación (Kezar et al., 2019).

En América Latina, el ideal normativo de la profesión académica ha experimentado un desplazamiento progresivo hacia la figura del académico-investigador, consolidado a través de sistemas de incentivos orientados a la productividad científica (Estévez-Nenninger et al., 2020; Bernasconi, Berríos y Véliz, 2021). Si bien esta transformación ha contribuido

al fortalecimiento de capacidades investigativas en diversas instituciones, también ha generado tensiones persistentes en torno al equilibrio entre las funciones de docencia e investigación, y ha intensificado la presión sobre los académicos para alinearse con estándares de desempeño definidos por agendas externas.

Estos cambios no han supuesto una supresión explícita de la autonomía académica individual, sino más bien su reconfiguración bajo parámetros de eficiencia, evaluación y control administrativo. Como resultado, se ha producido una pérdida gradual de control profesional sobre las condiciones y orientaciones del trabajo académico, en un entorno institucional donde las lógicas gerenciales tienden a prevalecer sobre los principios tradicionales de autorregulación disciplinaria (Brunner y Alarcón, 2024).

La autonomía académica en América Latina: a 100 años de la Reforma de Córdoba

La Reforma de Córdoba de 1918 marcó un hito en la educación superior latinoamericana al consagrar la autonomía universitaria como principio rector. Impulsada por estudiantes de la Universidad Nacional de Córdoba en un contexto de dominación oligárquica y clerical, promovió la democratización de instituciones controladas por élites (Balbachevsky, 2014; Bernasconi, 2024). El Manifiesto Liminar proclamó la autonomía como garantía frente al control estatal y eclesiástico, posicionando a las universidades como espacios de pensamiento crítico e innovación (Bergel, 2008; Marsiske, 2017).

El concepto se estructuró en tres dimensiones: la administrativa, que habilitó la elección democrática de autoridades con participación triestamental; la académica, que otorgó autonomía para definir planes de estudio, métodos de enseñanza y líneas de investigación; y la financiera, que permitió gestionar recursos con independencia (Marsiske, 2010; 2017).

La influencia de Córdoba se expandió rápidamente a México, Chile, Perú, Brasil y otros países, adaptándose a distintos contextos. En México, la autonomía fue producto de movilizaciones estudiantiles por la democratización y desvinculación del poder político; en otros casos, fue restringida por dictaduras militares durante las décadas de 1960 y 1970, con represión a estudiantes y académicos, y severas intervenciones estatales (Burbano, 2011; Marsiske, 2017).

Con el retorno democrático en los años 80, se reactivó la defensa de la autonomía, consolidándola como un principio esencial, aunque expuesto

a nuevas presiones. En el siglo XXI, la globalización, el avance tecnológico y las políticas neoliberales han generado tensiones adicionales. Por un lado, la autonomía se ve limitada por exigencias de evaluación externa y estándares de calidad promovidos por organismos internacionales con enfoques mercantilistas (Acosta-Silva, 2021; Sisto, 2020). Por otro, la expansión del sector privado ha intensificado la competencia por recursos, estudiantes y prestigio (Burbano, 2011; Cantard, 2015).

En este escenario, la autonomía se entiende como una construcción relacional, dinámica y situada, resultado de interacciones entre universidades, Estado y actores sociales, condicionadas por contextos históricos y políticos (Ordorika, 2010). Esta perspectiva resalta la autonomía como resguardo ante interferencias externas y, a la vez, como espacio de negociación para asegurar libertad académica e independencia intelectual. Acosta-Silva (2020) la describe como un proceso conflictivo entre demandas universitarias e intereses estatales, mientras Del Bello (2022) enfatiza su rol en la generación y difusión de conocimiento, salvaguardando la integridad institucional frente a presiones externas.

Aunque con formas diversas, Marsiske (2017) identifica elementos comunes en su ejercicio: el autogobierno mediante elección democrática y cogobierno, consolidado en México y Argentina; la libertad académica para definir contenidos, métodos y programas; y la autonomía financiera para administrar presupuestos y generar ingresos. Estos mecanismos han consolidado a las universidades como espacios de debate y transformación social, legitimándolas frente al Estado (Ordorika, 2010).

Sin embargo, el ejercicio de la autonomía ha generado controversias. Originalmente concebida como defensa frente al poder político, ha sido cuestionada por su limitada rendición de cuentas, opacidad y escasa respuesta a demandas sociales (Mendoza, 2020). En algunos casos, su uso como instrumento de confrontación con el Estado ha distorsionado su sentido original.

Las universidades públicas enfrentan hoy el desafío de equilibrar independencia institucional con responsabilidad social, contribuyendo al desarrollo económico y cultural de sus comunidades. Las políticas neoliberales han agudizado esta tensión, al concebirlas como proveedoras de servicios más que como bienes públicos (Marsiske, 2010; Villar, 2015). Si bien la autonomía está consagrada en varias constituciones, lo que ha permitido resistir crisis políticas y económicas, la dependencia del financiamiento público y la presión de mecanismos externos de evaluación siguen limitando su ejercicio pleno (Burbano, 2011; Ordorika, 2010; Villar, 2015).

La autonomía académica en tres niveles: un modelo analítico

El modelo analítico propuesto busca superar las limitaciones de los enfoques tradicionales que han abordado la autonomía universitaria desde perspectivas unidimensionales o normativas, centradas habitualmente en el nivel legal-formal o institucional. A partir de una concepción relacional y dinámica de la autonomía, este modelo permite observar cómo se configuran, negocian y tensionan los márgenes de autonomía universitaria en tres niveles interrelacionados: macro (político-sistémico), meso (institucional) y micro (individual).

Cada uno de estos niveles se operacionaliza a través de variables analíticas específicas, formuladas a partir de la revisión crítica de la literatura y del análisis empírico comparado de universidades latinoamericanas. La siguiente tabla sintetiza la arquitectura del modelo y explicita el foco analítico y las variables clave consideradas en cada nivel.

Tabla 1: Modelo analítico de tres niveles

Nivel	Foco analítico	Variables analíticas clave
Macro	Relación entre Estado, mercado y universidad	Autogobierno académico; bases jurídicas y autonomía académica; financiamiento estatal y autonomía académica
Meso	Gobernanza institucional y redistribución del poder	Control sobre la contratación de académicos; control sobre la carrera académica (promoción y jerarquización); evaluación de pares
Micro	Condiciones del trabajo académico individual	Libertad para definir agendas de investigación; libertad en los métodos de enseñanza; libertad para participar en actividades profesionales

Este modelo no solo contribuye a clarificar analíticamente las dimensiones de la autonomía universitaria, sino que también ofrece una herramienta útil para su estudio empírico comparado. Su enfoque relacional permite comprender cómo las universidades latinoamericanas negocian su autonomía en escenarios de creciente complejidad, tensionadas entre las lógicas de control externo, la reorganización institucional interna y las condiciones cambiantes del trabajo académico.

Con todo, el esquema multinivel exhibe tres límites que conviene explicitar: (i) se apoya principalmente en literatura secundaria y corre el riesgo de homogeneizar realidades nacionales heterogéneas; (ii) la compartimentación rígida de los planos macro, meso y micro atenúa la densidad de las interacciones entre Estado, mercado e instituciones; y (iii) aún carece de corroboración empírica, de modo que su poder expli-

cativo sigue siendo conjetural. Reconocer estas restricciones abre una agenda para refinar y contrastar el modelo mediante estudios comparados y metodologías mixtas.

Nivel macro: político-sistémico

En la educación superior, el nivel macro o político-sistémico es el espacio donde confluyen el Estado, el mercado y las instituciones de educación superior, configurando el marco operativo de las universidades. El triángulo de Clark (1983) ilustra estas relaciones de poder, asociando al Estado con la regulación y las políticas públicas, al mercado con las dinámicas de oferta y demanda, y a las instituciones con la autonomía universitaria.

Este marco conceptual permite analizar cómo las universidades enfrentan de manera simultánea presiones estatales, del mercado y de sus propias estructuras internas. El equilibrio entre estas fuerzas varía según el contexto histórico, político y social. En América Latina, la autonomía universitaria ha sido fundamental para proteger la libertad académica y promover la excelencia investigativa (Bernasconi, 2024), convirtiéndose en un elemento clave para responder a las necesidades de la sociedad.

Autogobierno académico

El autogobierno académico, eje central de la autonomía universitaria en América Latina, permite a las comunidades universitarias tomar decisiones clave sin intervención externa. Esto incluye la elección de autoridades, la definición de planes de estudio, las condiciones del profesorado y la gestión autónoma de recursos, equilibrando libertad académica, independencia política y gestión administrativa (Acosta-Silva, 2008).

Desde los años ochenta, los cambios socioeconómicos y políticos han transformado el significado y la práctica del autogobierno. La crisis económica, el ajuste estructural y las políticas de rendición de cuentas han limitado su alcance, exigiendo eficiencia, calidad y pertinencia social (Acosta-Silva, 2008). Este entorno ha introducido modelos gerenciales basados en gestión por objetivos, indicadores de desempeño e incentivos, tensionando los principios de autonomía (Acosta-Silva, 2022; Labraña y Brunner, 2022).

La incorporación de nuevos actores, como grupos de presión y el sector privado, ha complejizado la gobernanza universitaria, desafiando la cohesión y legitimidad del autogobierno. Preservar la autonomía mientras se responden a demandas sociales y políticas constituye un reto central para las universidades de la región (Acosta-Silva, 2008; Bernasconi, 2015).

Bases jurídicas y autonomía académica

Las bases jurídicas protegen y regulan la autonomía universitaria, brindando un marco que resguarda a las universidades de interferencias externas. En América Latina, la autonomía está consagrada en las constituciones de varios países, otorgando independencia legal frente al Estado (Bernasconi, 2024; Marsiske, 2017). Este marco jurídico permite salvaguardar valores fundamentales de la academia, como la libertad de cátedra, la investigación independiente y el autogobierno institucional.

En México, la autonomía universitaria se institucionalizó en 1929 con la autonomía de la Universidad Nacional Autónoma de México (UNAM), estableciendo un precedente para otras universidades de la región (Acosta-Silva, 2021). Sin embargo, la autonomía jurídica no siempre garantiza una independencia plena, ya que las universidades pueden estar sujetas a regulaciones externas, como las impuestas por agencias de acreditación, que limitan decisiones curriculares o administrativas (Valadés, 2015).

Las reformas educativas impulsadas por organismos internacionales han promovido la estandarización de los sistemas educativos, ajustándolos a un mercado global competitivo (Marquina, 2020). Aunque estas reformas buscan modernizar las instituciones, también pueden restringir su capacidad para adaptar sus políticas y prácticas a los contextos locales, tensionando el equilibrio entre autonomía e intervención externa (Marsiske, 2017).

Financiamiento estatal y autonomía académica

El financiamiento estatal es crucial para la autonomía de las universidades públicas en América Latina (Bernasconi, 2021; Kohtamäki y Balbachevsky, 2018; Mendoza, 2020). Sin embargo, la dependencia de fondos estatales puede limitar el autogobierno, ya que los gobiernos suelen condicionar los recursos al cumplimiento de objetivos políticos o de desarrollo (Marsiske, 2017). Este dilema plantea una tensión fundamental: aunque el financiamiento es esencial, puede restringir la libertad académica si está condicionado a indicadores de eficiencia y productividad.

La autonomía financiera implica equilibrar la independencia institucional con la rendición de cuentas hacia quienes proveen los recursos (Bernasconi, 2020; Marsiske, 2010; Tünnermann, 2008). Si bien el financiamiento estatal ha promovido equidad e inclusión en el acceso a la educación superior, también ha introducido mecanismos de control que desafían la autonomía universitaria.

En la actualidad, las universidades enfrentan presiones para adoptar lógicas gerenciales y responder a indicadores de desempeño impuestos por financiadores externos. Estas exigencias han transformado la

relación entre las universidades y el Estado, generando nuevos desafíos para preservar la autonomía y los valores fundamentales de la educación superior (Sisto, 2020).

Nivel meso: institucional

El análisis de las instituciones de educación superior en el nivel meso aborda las tensiones entre dos modelos de gestión: la colegialidad y el gerencialismo. En este nivel, las universidades deben equilibrar sus funciones tradicionales de enseñanza e investigación con demandas de eficiencia, responsabilidad y competitividad. Estas tensiones inciden directamente en la autonomía académica, influida por tres factores clave: el control sobre la contratación, la carrera académica y la evaluación de pares.

Control sobre la contratación de académicos

El control sobre la contratación es esencial para la autonomía académica. Tradicionalmente, en el modelo colegiado, la selección de académicos se basaba en criterios estrictamente académicos, con decisiones tomadas por la comunidad académica (Bernasconi, 2021; Kohtamäki y Balbachevsky, 2018; Mendoza, 2020). Sin embargo, el gerencialismo introduce criterios estratégicos que buscan alinear la contratación con metas de financiamiento externo y posicionamiento en rankings internacionales, desplazando el control hacia administradores o juntas externas.

En el caso chileno, la adopción de instrumentos de financiamiento por desempeño ha contribuido a una creciente centralización de los procesos de selección en las rectorías, limitando la participación efectiva de los cuerpos académicos. En contraste, en México, programas como el PIFI (Programa Integral de Fortalecimiento Institucional) y el SNI (Sistema Nacional de Investigadores) han introducido incentivos económicos que priorizan perfiles con alta productividad científica y capacidad para captar recursos, subordinando los criterios puramente disciplinarios a consideraciones de eficiencia y competitividad.

Por su parte, la Universidad Nacional de Córdoba continúa operando con comités de selección triestamentales, lo que preserva en parte la lógica colegiada; sin embargo, la reciente institucionalización de una Secretaría de Planeamiento Estratégico ha condicionado la asignación de nuevas plazas al cumplimiento de metas organizacionales, conformando un modelo híbrido en que conviven la deliberación académica y la gestión estratégica (Brunner y Alarcón, 2024). Estos casos ilustran cómo el ejercicio de autonomía en la contratación académica se redefine en función de la interacción entre los incentivos externos y la capacidad de autogobierno institucional.

Control sobre la carrera académica: promoción y jerarquización

El control sobre la carrera académica, particularmente en lo que respecta a la promoción y jerarquización, también impacta la autonomía académica (Bernasconi, 2021; Kohtamäki y Balbachevsky, 2018; Mendoza, 2020). En el modelo colegiado, las decisiones son tomadas por comités académicos que evalúan el desempeño según la calidad de la investigación, la enseñanza y el servicio a la comunidad académica. Este sistema garantiza altos estándares académicos y un proceso justo (Mignot-Gérard et al., 2022).

Sin embargo, con la adopción de prácticas gerenciales, la promoción se ha vuelto más dependiente de indicadores de productividad, como el número de publicaciones y el financiamiento externo (Acosta-Silva, 2021). Aunque estos criterios buscan profesionalizar a los académicos, también pueden distorsionar el proceso al priorizar la cantidad sobre la calidad, generando tensiones entre la lógica gerencial y la autonomía académica (Villanueva, 2022). En América Latina, el control autónomo sobre la promoción es fundamental para evitar interferencias externas en decisiones estratégicas (Marquina, 2020).

Los casos de México, Chile y Argentina muestran una tendencia convergente hacia el uso de indicadores cuantitativos, aunque con intensidades y efectos distintos. En México, los programas de incentivos como el SNI han reforzado una cultura de evaluación orientada por resultados medibles, generando trayectorias académicas más estrechamente alineadas con parámetros de productividad. En Chile, los esquemas de financiamiento competitivo, como FONDECYT (Fondo Nacional de Desarrollo Científico y Tecnológico), han instalado exigencias similares, aunque con una aplicación menos sistemática. En cambio, en Argentina, donde los comités académicos conservan mayor margen de decisión, persiste una tradición evaluativa más integral que permite ponderar una diversidad de méritos.

Estas diferencias reflejan cómo el avance del gerencialismo redefine los criterios y mecanismos de jerarquización académica, modulando el alcance de la autonomía según la fuerza relativa de los incentivos externos y la institucionalización de normas colegiadas (Brunner y Alarcón, 2024).

Evaluación de pares

La evaluación de pares es el mecanismo tradicional para garantizar la calidad del trabajo académico. Este sistema permite que las decisiones sobre investigación y enseñanza sean tomadas por quienes mejor comprenden su valor, reforzando la autonomía académica (Bernasconi,

2021). Sin embargo, la evaluación de pares ha sido complementada, o reemplazada en algunos casos, por sistemas de evaluación basados en métricas de desempeño, como indicadores bibliométricos o rankings (Acosta-Silva, 2021). Si bien estas métricas introducen objetividad, también pueden erosionar la autonomía académica al reducir el control de los académicos sobre la evaluación de su trabajo, generando una cultura de performatividad que prioriza la cantidad sobre la calidad (Bruckmann y Carvalho, 2018). En América Latina, la evaluación de pares sigue siendo crucial para proteger la autonomía académica frente a presiones externas y mantener altos estándares académicos (Marquina, 2020).

No obstante, su aplicación efectiva varía significativamente entre países. En Chile, la evaluación docente y de investigación por parte de colegas es relativamente poco frecuente, lo que refleja una mayor dependencia de sistemas formales y gerenciales de aseguramiento de la calidad. En contraste, Argentina y México mantienen niveles más altos de participación académica en estos procesos, aunque también enfrentan presiones crecientes para incorporar criterios estandarizados. En el caso mexicano, por ejemplo, la evaluación estudiantil se ha institucionalizado de manera extensa, especialmente en la docencia, lo que plantea interrogantes sobre el equilibrio entre legitimidad pedagógica y control profesional (Brunner y Alarcón, 2024).

En definitiva, la evidencia comparada muestra que la tensión entre colegialidad y gerencialismo no desemboca en un reemplazo simple, sino en configuraciones híbridas. Cuando la asignación de recursos y la contratación se vinculan a indicadores externos –Chile y, con mayor intensidad, México– la voz académica se reduce y la autonomía depende de la capacidad de negociación de las unidades de base. Donde persisten mecanismos colegiados robustos (caso argentino) los gestores deben conciliar las lógicas de desempeño con la deliberación disciplinar. Reconocer estas trayectorias diferenciadas es clave para diseñar políticas que fortalezcan la gobernanza participativa sin sacrificar la transparencia ni la responsabilidad institucional.

Nivel micro: individual

El análisis del nivel micro de la autonomía académica se centra en el ámbito personal del académico, donde este ejerce su labor en las áreas de investigación, docencia y actividades profesionales complementarias. En este nivel, la autonomía se define como la capacidad del académico para tomar decisiones informadas, basadas en su juicio experto, libres de coerciones externas, ya sea de la administración universitaria, el Estado

o financiadores privados (Bernasconi, 2021; Kohtamäki y Balbachevsky, 2018; Mendoza, 2020). Esta forma de autonomía es crucial no solo para el desarrollo integral del académico, sino también para que las universidades públicas puedan seguir siendo espacios de pensamiento crítico, creación de conocimiento y responsabilidad social (Marsiske, 2010; Acosta Silva, 2020; Marquina, 2020).

En este contexto, se identifican tres factores principales que condicionan la autonomía académica a nivel individual: la libertad en cuanto a la agenda de investigación, la libertad en los métodos de enseñanza y la libertad para participar en actividades profesionales. Cada uno de estos elementos tiene un impacto significativo en cómo los académicos estructuran su labor y, en última instancia, en la capacidad de las universidades para cumplir con su misión de creación y transferencia de conocimiento.

Autonomía en la agenda de investigación

La libertad para definir temas y preguntas de investigación constituye un componente esencial de la autonomía académica individual. Esta capacidad permite a los académicos desarrollar proyectos relevantes tanto para su disciplina como para la sociedad, sin interferencias externas, y favorece el avance del conocimiento al abrir espacio para la exploración de áreas emergentes e innovadoras que podrían quedar marginadas si las decisiones se rigieran exclusivamente por agendas externas (Kohtamäki y Balbachevsky, 2018).

Sin embargo, la autonomía investigativa enfrenta crecientes restricciones. La dependencia del financiamiento externo impone condiciones que tienden a orientar los proyectos hacia prioridades establecidas por los organismos financiadores, las cuales no siempre coinciden con los intereses académicos ni con los desafíos locales más urgentes (Pérez-Mora et al., 2022). En este contexto, el resguardo de la autonomía requiere mecanismos institucionales que aseguren el acceso a fuentes de financiamiento diversificadas y alineadas con la libertad de investigación.

En América Latina, las reformas recientes, influenciadas por modelos globales de evaluación y financiamiento, han profundizado las tensiones entre prioridades internacionales y agendas locales. Los académicos enfrentan presiones que condicionan la elección de temas, ya sea por exigencias de alineamiento con líneas estratégicas externas o por la necesidad de responder a sistemas de evaluación que valoran ciertos enfoques por sobre otros (Marquina, 2020; Acosta-Silva, 2021).

Aun así, los sistemas nacionales muestran configuraciones distintas (Alarcón et al., 2025). En México predomina una lógica orientada a la productividad y a la vinculación con el mercado; en Chile, las políticas

institucionales promueven una investigación con impacto social y proyección internacional, aunque mediada por una fuerte presión por captar financiamiento competitivo; en Argentina, pese a la adopción parcial de estándares globales, persiste una mayor libertad para definir agendas desde la deliberación académica, con menor interferencia de criterios externos. Estas diferencias sugieren que la autonomía investigativa se negocia en el cruce entre lógicas institucionales, exigencias del financiamiento y marcos normativos nacionales.

Autonomía en los métodos de enseñanza

La libertad académica se expresa también en la esfera de la docencia, donde los académicos deben contar con la capacidad de diseñar y estructurar sus cursos en función de los avances de su disciplina y las características de su estudiantado (Kohtamäki y Balbachevsky, 2018; Mendoza, 2020). Esta autonomía es fundamental para ofrecer una formación pertinente y actualizada, alineada con el conocimiento emergente y con las demandas formativas de cada contexto (Marsiske, 2010). Preservar márgenes de decisión en la enseñanza permite a los docentes experimentar con metodologías, ajustar los ritmos de aprendizaje y aplicar enfoques pedagógicos que favorezcan procesos formativos más significativos (Marquina, 2020).

Sin embargo, en varias instituciones de América Latina esta autonomía enfrenta restricciones derivadas de procesos de estandarización curricular y evaluación impulsados por políticas de aseguramiento de la calidad. Tales mecanismos, si bien buscan garantizar niveles mínimos de desempeño institucional, pueden limitar la capacidad de los docentes para innovar pedagógicamente o para adaptar su enseñanza a realidades específicas (Acosta-Silva, 2021). La expansión de métricas internacionales de evaluación docente ha reforzado estas tendencias, imponiendo criterios de rendimiento que no siempre reflejan las condiciones locales ni las necesidades concretas del estudiantado (Marquina, 2022).

Las regulaciones sobre la docencia varían significativamente entre países. En Argentina y México, la organización de la enseñanza tiende a estar fuertemente normada, tanto en términos de horarios como en las exigencias de evaluación externa. En estos sistemas, la supervisión de la docencia suele recaer en los jefes de unidad académica, y la evaluación por parte de los estudiantes está ampliamente institucionalizada. En contraste, el sistema chileno presenta menores niveles de reglamentación formal respecto a los tiempos lectivos y la estructura de los cursos, lo que abre espacios potenciales para enfoques didácticos más diversos. No obstante, esta mayor autonomía operativa se ve matizada por un alto

grado de escrutinio estudiantil sobre la labor docente (Brunner y Alarcón, 2024; Alarcón et al., 2025).

En todos los casos, la docencia presencial sigue siendo la modalidad predominante, aunque comienzan a emerger experiencias híbridas o virtuales de carácter incipiente. Estas transiciones exigen repensar la autonomía docente no solo como libertad de cátedra, sino también como capacidad para adaptar críticamente las herramientas pedagógicas a nuevos entornos de enseñanza-aprendizaje. Así, el desafío no radica únicamente en resistir la estandarización, sino en encontrar formas institucionales de resguardar y fomentar la autonomía como condición para una docencia reflexiva e innovadora.

Autonomía en las actividades académicas y profesionales

El tercer factor que condiciona la autonomía académica a nivel individual es la libertad para participar en actividades académicas y profesionales complementarias, como conferencias, redes académicas y consultorías (Bernasconi, 2021; Kohtamäki y Balbachevsky, 2018; Mendoza, 2020). Los académicos no solo son docentes e investigadores, sino que también participan activamente en su campo de estudio a través de actividades externas que les permiten mantenerse al día con los últimos desarrollos y aplicar su conocimiento en contextos prácticos (Pérez-Mora et al., 2022).

La participación en consultorías, asesorías técnicas y otras actividades externas constituye también una expresión significativa de la autonomía profesional del académico. En distintos países de América Latina, es común que los docentes universitarios colaboren con organismos públicos, organizaciones sociales, medios de comunicación o el sector productivo, actuando como especialistas en sus respectivas áreas del saber. Estas interacciones no solo les permiten proyectar socialmente el conocimiento universitario, sino que también retroalimentan sus propias prácticas académicas, ampliando sus marcos de análisis y enriqueciendo tanto la docencia como la investigación (Acosta Silva, 2022; Bernasconi, 2024).

No obstante, esta dimensión de la autonomía enfrenta limitaciones estructurales. La escasez de recursos y el insuficiente respaldo institucional dificultan la participación activa en redes internacionales de colaboración o en instancias de proyección profesional. Esta situación se ve agravada por la creciente presión sobre las universidades para cumplir con metas de desempeño institucional, como la acreditación de programas o el mejoramiento de indicadores de eficiencia y empleabilidad. En este escenario, suelen postergarse aquellas actividades que no contribuyen directamente a dichos objetivos (Acosta Silva, 2020).

A ello se suma una intensificación de las tareas de servicio académico, especialmente las orientadas a la atención estudiantil, cuya organización y formalización varía entre países (Alarcón et al., 2025). En México, estas tareas tienden a estar fuertemente normadas y asociadas a metas explícitas, mientras que en Argentina su implementación es más heterogénea. En Chile, en cambio, persiste una menor formalización, lo que refleja diferencias institucionales en la manera en que se regulan estas interacciones. Si bien estas funciones cumplen un rol relevante en la formación estudiantil, también pueden reducir el tiempo disponible para participar en redes externas, consultorías y actividades de divulgación. Sin embargo, su impacto sobre la autonomía profesional parece atenuarse cuando los académicos perciben condiciones laborales que valoran su trabajo y favorecen su desarrollo profesional.

Reflexiones finales

Este ensayo ha revisado críticamente la noción de autonomía universitaria, proponiendo un modelo multinivel que captura la complejidad y las tensiones inherentes al concepto en el contexto de las instituciones de educación superior latinoamericanas. La autonomía, como hemos demostrado, no puede analizarse únicamente desde una perspectiva normativa o prescriptiva, sino que debe entenderse como un constructo dinámico influenciado por factores políticos, económicos y sociales. Al abordar los niveles macro, meso y micro, este análisis permite una visión más matizada de los diversos elementos que condicionan la autonomía universitaria, así como las tensiones emergentes entre la academia y la gestión universitaria.

A nivel macro, la influencia del Estado y el mercado sobre las universidades latinoamericanas plantea preguntas cruciales sobre la capacidad de estas instituciones para resistir la creciente presión hacia la eficiencia y la rendición de cuentas. Si bien es indudable que las universidades deben ser responsables ante la sociedad y sus actores, es imperativo que las políticas públicas respeten los principios de libertad académica y de autogobierno, los cuales han sido fundamentales para el desarrollo del conocimiento.

En el ámbito institucional, el creciente peso de la gestión profesional y la introducción de mecanismos de evaluación estandarizados representan un desafío considerable para el gobierno colegial. Si bien el gerencialismo puede ofrecer soluciones prácticas en términos de eficiencia organizacional, es necesario reflexionar sobre sus efectos a largo plazo en la autonomía académica. La centralización excesiva del poder y la impo-

sición de métricas cuantitativas sobre la producción académica corren el riesgo de debilitar los procesos participativos y el ethos académico, comprometiendo así la capacidad de los académicos para influir en las decisiones clave de las universidades.

Por último, a nivel individual, los académicos se encuentran cada vez más sometidos a presiones externas que limitan su autonomía para investigar, enseñar y contribuir al debate público sin restricciones impuestas por agendas externas. La creciente precarización de la carrera académica, junto con el énfasis en la productividad y la internacionalización, ha generado una fragmentación en la profesión, exacerbando las desigualdades dentro de la comunidad académica y poniendo en cuestión los principios fundamentales de libertad intelectual y profesional.

En conclusión, este ensayo conceptual pone de relieve la importancia de adoptar una perspectiva multinivel para entender la autonomía universitaria en el contexto latinoamericano. Las tensiones entre la autonomía y las demandas externas no son nuevas, pero han adquirido una nueva dimensión bajo las reformas contemporáneas. Aunque las políticas neoliberales y los modelos gerenciales han transformado el gobierno institucional, es esencial que las universidades encuentren formas de preservar su misión académica y social. La autonomía, como principio rector de la educación superior, debe ser defendida no solo como un ideal abstracto, sino como una condición fundamental para el desarrollo de sociedades democráticas, equitativas y basadas en el conocimiento.

Referencias bibliográficas

- Acosta Silva, A. (2008). La autonomía universitaria en América Latina: Problemas, desafíos y temas capitales. *Universidades*, 36, 69–82.
- Acosta Silva, A. (2021). Autonomía universitaria y Estado: La experiencia en Latinoamérica. *Revista de Educación Superior*, 49(193), 1–25.
- Acosta Silva, A. (2022). Autonomía universitaria, gobierno institucional y gobernanza interpretativa en México. *Perfiles Latinoamericanos*, 30(59). <https://doi.org/10.18504/pl3059-016-2022>
- Alarcón, M., Brunner, J. J., and Labraña, J. (2025). Between managerialism and collegiality: The transformation of the academic profession in Ibero-America. *European Journal of Higher Education*, 1–21. <https://doi.org/10.1080/21568235.2025.2491076>
- Alarcón, M. and Brunner, J. J. (2024). Multi-governance in higher education: The case of Chile 2018-2023. *Research in Education*, 119(1), 97-118. <https://doi.org/10.1177/00345237241248535>

- Balbachevsky, E. (2014). The Latin American university model and the challenges posed by the reforms: Perspectives from the academics. In B. Cantwell and I. Kauppinen (Eds.), *The relevance of academic work in comparative perspective* (pp. 239–252). Springer.
- Bergel, M. (2008). Latinoamérica desde abajo. Las redes transnacionales de la Reforma Universitaria (1918–1930). En *La reforma universitaria: Desafíos y perspectivas noventa años después* (pp. 146–184).
- Bernasconi, A. (2021). University autonomy and academic freedom: Contrasting Latin American and US perspectives. *Higher Education Governance and Policy*, 2(1), 56–67.
- Bernasconi, A. (2023). Academic freedom in Latin America and the deceptive comfort of autonomy. *International Higher Education*, 116, 15. <https://doi.org/10.36197/IHE.2021.105.03>
- Bruckmann, S., y Carvalho, T. (2018). Entendiendo el cambio en la educación superior: Un enfoque arquetípico. *Higher Education*, 1–19. <https://doi.org/10.1007/s10734-018-0229-2>
- Brunner, J. J., y Alarcón, M. (2025). More managerialism and less collegiality? The management of Chilean universities from the perspective of academics. In L. Leišytė, M. Marquina, and G. A. Jones (Eds.), *University governance, management and the academic profession* (Vol. 26, pp. 137–152). Springer. https://doi.org/10.1007/978-3-031-86889-4_8
- Brunner, J. J., y Alarcón, M. (2024). Academic management in Chilean universities: An analysis from academics' perspective. *European Journal of Education*, 59(4), e12798.
- Brunner Ried, J. J., y Alarcón-Bravo, M. (2024). El gobierno de las universidades chilenas en perspectiva internacional comparada. *Revista Colombiana de Educación*, (93), 193–217.
- Brunner, J. J., Salmi, J., Labraña, J., Balbachevsky, E., Ganga, F., ... y Schwartzman, S. (2022). *Enfoques de sociología y economía política de la educación superior: Aproximaciones al capitalismo académico en América Latina*. Ediciones Universidad Diego Portales.
- Brunner, J. J. (2015). Ideas y fines de la universidad. *Estudios Públicos*, 139, 155–164.
- Brunner, J. J. (2011). Gobernanza universitaria: Tipología, dinámicas y tendencias. *Revista de Educación*, 355, 137–159.
- Cantard, A. (2014). La autonomía universitaria hoy: Un debate necesario. *Unión de Universidades de América Latina y el Caribe*.
- Cardoso, S., Carvalho, T., and Videira, P. (2019). Is it still worth working in academia? The views from Portuguese academics. *Higher Education Policy*, 32(4), 663–679.

- Carvalho, T. (2017). New Public Management and the academic profession. In J. Huisman and M. Tight (Eds.), *Theory and method in higher education research* (Vol. 3, pp. 59–76). Emerald Publishing.
- Carvalho, T. (2018). New Public Management and the academic profession. In P. Teixeira and J. Shin (Eds.), *Encyclopedia of international higher education systems and institutions*. Springer. https://doi.org/10.1007/978-94-017-9553-1_310-1
- Da Wan, C., Lee, M. N., and Loke, H. Y. (Eds.). (2019). *The governance and management of universities in Asia: Global influences and local responses*. Routledge.
- Del Bello, J. C. (2022). *Los alcances de la autonomía universitaria*. Editorial EUDEBA.
- Estermann, T., and Steinel, M. (2011). University autonomy in Europe. *Beiträge zur Hochschulforschung*, 33(2), 86–105.
- Estévez-Nenninger, E. H., González-Bello, E. O., Valdés-Cuervo, Á., Arcos-Vega, J. L., Ramiro-Marentes, F., and Gutiérrez-Franco, L. E. (2020). Teaching and research of academics in Mexico: Preferences and dedication according to the international survey APIKS. In *Higher Education Forum*, 17, 99–114. Research Institute for Higher Education, Hiroshima University.
- Fuksman, B. U., y Nosiglia, M. C. (2020). El lugar de la investigación en la profesión académica argentina: Hallazgos del estudio internacional APIKS. *Revista Latinoamericana de Políticas y Administración de la Educación*, 12, 61–81.
- Jones, G. A., and Weinrib, J. (2022). The changing context of academic work: Fragmentation, institutional horizontal diversity and vertical stratification. In S. Lysons and T. Burgess (Eds.), *Research handbook on academic careers and managing academics* (pp. 36–46). Edward Elgar Publishing.
- Jones, G. A., and Finkelstein, M. (2019). Looking across systems: Implications for comparative, international studies of academic work. In M. Finkelstein, G. A. Jones y M. Huisman (Eds.), *Professorial pathways: Academic careers in a global perspective* (pp. 265–290). Johns Hopkins University Press.
- Kallenberg, T. (2015). Academic middle managers shaping the landscape between policy and practice. In B. Jongbloed, J. Enders y C. Salerno (Eds.), *Diversity and excellence in higher education* (pp. 201–216). Brill.
- Kallio, T. J., Kallio, K.-M., and Blomberg, A. (2020). From professional bureaucracy to competitive bureaucracy: Redefining universities' organization principles, performance measurement criteria, and reason

- for being. *Qualitative Research in Accounting y Management*, 17(1), 82–108. <https://doi.org/10.1108/QRAM-10-2019-0111>
- Kenny, J. (2018). Re-empowering academics in a corporate culture: An exploration of workload and performativity in a university. *Higher Education*, 75, 365–380.
- Kezar, A., De Paola, T., and Scott, D. T. (2019). *The gig academy: Reforming higher education: Innovation and the public good*. Johns Hopkins University Press.
- Kohtamäki, V., & Balbachevsky, E. (2018). University autonomy from past to present. *Theoretical and methodological perspectives on higher education management and transformation: an advanced reader for PhD students*, 179-194.
- Kwiek, M. (2019). *Changing European academics*. Taylor and Francis.
- Labraña, J., y Brunner, J. J. (2022). La ideología de la nueva gestión pública desde la mirada de los directivos de las universidades chilenas: Un estudio de casos múltiples. *Revista Iberoamericana de Educación Superior*, 13(38), 3–23.
- Lerch, J. C., Frank, D. J., and Schofer, E. (2023). The social foundations of academic freedom: Heterogeneous institutions in world society, 1960 to 2022. *American Sociological Review*, 89(1), 88–125.
- Leišytė, L., Jones, G. A., and Marquina, M. (2025). Trends in higher education governance and the academic profession. In L. Leišytė, M. Marquina and G. A. Jones (Eds.), *University governance, management and the academic profession* (Vol. 26, pp. 3–10). Springer. https://doi.org/10.1007/978-3-031-86889-4_1
- Maassen, P. (2024). Erosion of academic freedom in the European Union. *International Higher Education*, 120.
- Maassen, P., Gornitzka, Å., and Fumasoli, T. (2017). University reform and institutional autonomy: A framework for analysing the living autonomy. *Higher Education Quarterly*, 71(3), 239–257.
- Marquina, M. (2020). Between the global and the local: The study of the academic profession from a Latin American perspective. *Higher Education Governance and Policy*, 1(1), 63–75.
- Marquina, M., Centeno, C. P., and Reznik, N. (2022). Academic power and institutional control of academia in Argentine public universities within the context of a managerial governance model. In S. Lysons and T. Burgess (Eds.), *Research handbook on academic careers and managing academics* (pp. 47–63). Edward Elgar Publishing.
- Marsiske, R. (2017). La autonomía universitaria en América Latina a 100 años del movimiento estudiantil de Córdoba: Una agenda de investigación desde México. *Universidades*, 72, 27–35.

- Mendoza, P. (2020). Autonomy and weak governments: Challenges to university quality in Latin America. *Higher Education*, 80, 719–737.
- Mignot-Gérard, S., Sponem, S., Chatelain-Ponroy, S., and Musselin, C. (2023). Kaleidoscopic collegiality and the use of performance research metrics: The case of French universities. *Higher Education*, 85(4), 887–918.
- Musselin, C. (2007). The transformation of academic work: Facts and analysis. *Research y Occasional Paper Series: CSHE.4.07*. Center for Studies in Higher Education.
- Ordorika Sacristán, I. (2010). La autonomía universitaria: Una perspectiva política. *Perfiles Educativos*, 32(número especial), 79–81.
- Pekkola, E., Carvalho, T., Siekkinen, T., and Johansson, J. E. (2018). The sociology of professions and the study of the academic profession. *ResearchGate*. <https://www.researchgate.net/publication/328107248>
- Pérez Mora, R., Castañeda Bernal, X. Y., y Inguanzo Arias, B. L. (2022). El trabajo científico: Entre la racionalidad económico-burocrática y la racionalidad ético-política. *Analecta Política*, 12(23), 1–26. <https://doi.org/10.18566/apolit.v12n23.a05>
- Sisto, V. (2020). Overwhelmed: Accountability and work intensification in the neo-liberal university. The case of Chile. *Education Policy Analysis Archives*, 28, 7. <https://doi.org/10.14507/epaa.28.4907>
- Slaughter, S., and Leslie, L. L. (1997). *Academic capitalism: Politics, policies, and the entrepreneurial university*. Johns Hopkins University Press.
- Stephenson, G. K., Jones, G. A., Bégin-Caouette, O., and Metcalfe, A. S. (2025). The paradox of increasing managerialism amid historic decentralization: Canadian professors and university governance. In L. Leišytė, M. Marquina and G. A. Jones (Eds.), *University governance, management and the academic profession* (Vol. 26, pp. 71–91). Springer. https://doi.org/10.1007/978-3-031-86889-4_5
- Teichler, U., Aarrevaara, T., and Huang, F. (2022). Conclusion: What we know about the teaching-research nexus in the knowledge-based society. In F. Huang, T. Aarrevaara, and U. Teichler (Eds.), *Teaching and research in the knowledge-based society: Historical and comparative perspectives* (pp. 241–253). Springer.
- Valadés, D. (2015). Autonomía y constitución en América Latina. *Unión de Universidades de América Latina y el Caribe*.
- Villar, A., y Ibarra, A. (2014). La autonomía universitaria, una mirada latinoamericana. *UDUAL*.
- Wen, W., and Marginson, S. (2023). Governance in Chinese universities. In R. W. Watermeyer and D. Knight (Eds.), *University collegiality and the erosion of faculty authority* (pp. 171–197). Emerald Publishing.

Adherencia a la declaración PRISMA en los meta-análisis de intervenciones experimentales publicados en Educación: una meta-revisión sistemática¹

Adherence to the PRISMA Statement in meta-Analyses reporting experimental interventions published in Education: A Systematic Meta-Review

<https://doi.org/10.4438/1988-592X-RE-2025-412-736>

Micaela Sánchez-Martín

<https://orcid.org/0000-0002-9107-574X>

Universidad de Murcia. IMIB*

Marta Gutiérrez-Sánchez

<https://orcid.org/0000-0001-7302-9283>

Universidad de Murcia

Eva María Olmedo-Moreno

<https://orcid.org/0000-0003-0558-1513>

Universidad de Granada

Fernando Navarro-Mateu

<https://orcid.org/0000-0002-3228-623X>

Universidad de Murcia. IMIB*

Resumen

Objetivo: Evaluar la adherencia a la declaración PRISMA de los estudios de meta-análisis sobre intervenciones educativas orientadas a mejorar el rendimiento académico entre 2009 y 2022. **Método:** Revisión sistemática. **Criterios de selección:** metaanálisis de estudios experimentales evaluando intervencio-

¹ Instituto Murciano de Investigación Biosanitaria Pascual Parrilla

nes educativas diseñadas para mejorar el rendimiento académico publicadas en inglés o español entre el 1 de enero de 2009 hasta el 30 de abril de 2022. Criterios de exclusión: otros diseños o resultados y publicaciones no accesibles. Búsqueda bibliográfica en cuatro bases de datos (ERIC, Web of Science, Scopus y PubMed). Extracción de datos mediante protocolo previamente elaborado y registrado. **Datos:** adherencia a las recomendaciones PRISMA, riesgo de sesgos según la AMSTAR 2, y diversas características de los estudios publicados.

Análisis estadísticos: Chi-cuadrado o prueba exacta de Fisher. Como medida de asociación, se calculó las odds ratios y β , con sus intervalos de confianza del 95% (IC95%) y p-valores mediante análisis de regresión logística y lineal. **Resultados:** De 2076 estudios identificados se seleccionaron 69. La puntuación media de la PRISMA fue de 19.7 (SD=4.4) sobre 27. Un 51.8% (n=14) de las recomendaciones tuvo una adherencia superior al 75%. Se encontraron diferencias significativas en el cumplimiento de las diversas recomendaciones en los estudios según los autores hayan declarado su adherencia a la PRISMA en 14 de las 27 recomendaciones (51.8%). El seguimiento de 13 de las 27 recomendaciones (48.1%) se asociaron a un menor riesgo de sesgos. La adherencia a la PRISMA se asoció a un menor riesgo de sesgos, a una publicación más reciente, a la educación en el área sanitaria y a intervenciones con actividades físicas orientadas a la educación superior. **Conclusiones:** La adherencia a la PRISMA en los metaanálisis publicados en educación con intervenciones experimentales para aumentar el rendimiento académico son claramente mejorables.

Financiación: Proyecto de Investigación PID2020-119194RB-I00, financiado por MCIN /AEI/10.13039/501100011033.

Protocolo: registrado en Open Science Framework: <https://doi.org/10.17605/OSF.IO/PSKN6>

Palabras clave: Meta-revisión, Revisión Sistemática, Meta-análisis, PRISMA, intervenciones educativas, rendimiento académico.

Abstract

Objective: To assess adherence to the PRISMA statement in meta-analyses of educational interventions aimed at improving academic performance between 2009 and 2022. **Method:** Systematic review. **Eligibility criteria:** Meta-analyses of experimental studies evaluating educational interventions designed to improve academic performance, published in English or Spanish between January 1, 2009, and April 30, 2022. Exclusion criteria: Other designs or outcomes, and inaccessible publications. Bibliographic search: Conducted in four databases (ERIC, Web of Science, Scopus, and PubMed). Data extraction: Based on a previously developed and registered protocol. **Data:** Adherence to PRISMA recommendations, risk of bias according to AMSTAR 2, and various characteristics of published studies. Statistical analyses: Chi-square test or Fisher's exact test. As measures of association, odds ratios and β were calculated, with 95% confidence intervals (95% CI) and p-values through logistic and linear regression analyses. **Results:** Out of 2,076 identified studies, 69 were included. The mean PRISMA score was 19.7 (SD

= 4.4) out of 27. A total of 51.8% (n = 14) of the recommendations had adherence rates above 75%. Significant differences were found in adherence to the recommendations depending on whether the authors declared compliance with PRISMA in 14 of the 27 items (51.8%). Compliance with 13 of the 27 recommendations (48.1%) was associated with a lower risk of bias. PRISMA adherence was associated with lower risk of bias, more recent publication, health-related education, and interventions involving physical activities in higher education. **Conclusions:** Adherence to PRISMA in published educational meta-analyses of experimental interventions to improve academic performance is clearly improvable.

Funding: Research Project PID2020-119194RB-I00, funded by MCIN / AEI/10.13039/501100011033.

Protocol: Registered on Open Science Framework: <https://doi.org/10.17605/OSF.IO/PSKN6>

Key words: Meta-review, Systematic Review, Meta-analysis, PRISMA, educational interventions, academic performance.

Introducción

En las últimas décadas, el número de publicaciones científicas ha crecido exponencialmente en todas las áreas del conocimiento, incluido el ámbito educativo (Ahn et al., 2012; Ruiz-Corbella et al., 2014). Esta proliferación, junto con la necesidad de identificar y aplicar de manera eficiente las mejores evidencias científicas en la toma de decisiones educativas, ha resaltado la importancia de las revisiones sistemáticas (RS) y metaanálisis (MA) como herramientas fundamentales en el desarrollo de la Educación Basada en la Evidencia (EBE) (Hederich et al., 2014; Sánchez-Martín et al., 2022). A diferencia de las revisiones narrativas tradicionales, las RS y los MA siguen un método científico definido para identificar, evaluar y sintetizar la literatura sobre un tema de manera sistemática y reproducible. Desde que Gene V. Glass en 1976 introdujera el término MA en el ámbito educativo su uso ha aumentado considerablemente en educación (Ahn et al., 2012). Los MA son revisiones sistemáticas que emplean técnicas estadísticas para calcular un efecto global y evaluar aspectos como el sesgo de publicación y la heterogeneidad.

Este incremento en el número de publicaciones se ha acompañado de una creciente preocupación por su calidad metodológica y ha facilitado el desarrollo de estudios centrados en la evaluación metodológica de las publicaciones científicas o meta-investigación (Ioannidis, 2018; Ioannidis et al., 2015). Esta situación también se ha reproducido en el ámbito de la educación (Anguera, 2023; Blanco-Blanco, 2018; Cook et al., 2011), debido a la necesidad de que el conocimiento científico se base en estu-

dios rigurosos, reproducibles y con resultados válidos para su aplicación en la práctica educativa. Estos trabajos de investigación, asimismo denominadas meta-revisiones o R metodológicas, se consideran un tipo especial de revisión en paraguas. Cuyo objetivo principal es evaluar la calidad de la investigación en sí misma, más allá de los resultados analizados. Este tipo de estudios se centran en identificar áreas de mejora en la calidad metodológica y en la forma en la que se presentan los informes publicados.

En la educación médica, diversos estudios han evaluado la calidad metodológica de las publicaciones (Cook et al., 2011; Howley et al., 2008), y, al menos, tres investigaciones han abordado esta cuestión en otros ámbitos educativos (Ahn et al., 2012; Eser & Yurtçu, 2020; Sánchez-Martín et al., 2024a). Los hallazgos de estos estudios sugieren que la calidad metodológica y el riesgo de sesgos (*Risk of Bias - RoB-*) en los MA y las RS en educación no alcanzan los estándares deseables, resaltando la necesidad de mejorar estos diseños para aumentar la confianza en sus resultados. Además, se ha identificado una falta de información clave en muchas publicaciones científicas. Esta deficiencia resulta aún más llamativa considerando el desarrollo de listas de verificación específicamente diseñadas para mejorar la transparencia y reproducibilidad de los MA y RS, como la declaración PRISMA 2020 (*Preferred Reporting Items for Systematic reviews and Meta-Analyses*) (Page et al., 2021a) y la Norma de Presentación de Metaanálisis (*Meta-Analysis Reporting Standards -MARS-*) (*American Psychological Association Publications and Communications Board Working Group on Journal Article Reporting Standards*, 2008; Appelbaum et al., 2018) entre otras (Rubio-Aparicio et al., (2018) y Sánchez-Meca et al., (2021).

Aunque el uso de la guía PRISMA en el ámbito educativo ha aumentado en los últimos años (Holmqvist & Lantz Ekström, 2024; Moreno et al., 2022), el impacto de su utilización ha sido evaluado preferentemente en la educación médica. Diversos estudios sugieren que la adherencia a la PRISMA durante la redacción de una RS se asoció a un menor riesgo de sesgos y, por tanto, a una mayor calidad metodológica (Panic et al., 2013; Sharma & Oremus, 2018). En el ámbito general de la educación, recientemente se han publicado los primeros dos artículos de la primera meta-revisión publicada hasta la fecha utilizando dos de los instrumentos de referencia internacional propuestos para evaluar el RoB y la transparencia y claridad de la publicación, como son, respectivamente, la AMSTAR 2 (*A Measurement Tool to Assess systematic Reviews*) (Shea et al, 2017) y la declaración PRISMA. Esta meta-revisión se centró en los MA publicados basados en estudios experimentales que evaluaban intervenciones educativas orientadas a la mejora del rendimiento académico. Los resultados del primero sugieren que la declaración explícita por parte de los autores de

haber utilizado la guía PRISMA se asoció a un menor RoB (Sánchez-Martín et al., 2024a). En el segundo, la declaración explícita de los autores de haber utilizado la declaración PRISMA aumentaba la probabilidad de incluir un diagrama de flujo. Este flujograma explicita gráficamente todo el proceso de la búsqueda y selección de los artículos y su elaboración e inclusión en el artículo se asoció a un menor RoB (Sánchez-Martín et al., 2025). En ninguno de ellos se evaluó la cumplimentación real de cada una de las recomendaciones individuales de la PRISMA y si la mención explícita por parte de los autores de haberse adherido a la declaración PRISMA se asociaba a una cumplimentación real de sus recomendaciones. Un estudio previo analizando sin la mención explícita de la PRISMA en el texto completo del documento se asoció con un incremento en la calidad del reporte no encontró diferencias en la cumplimentación de la adherencia general a la PRISMA y al comparar dos grupos de RS y MA, unos que hacían una mención explícita de su uso en el manuscrito a texto completo con otro en la que no se mencionaba explícitamente (Panic et al., 2013).

En este tercer análisis, utilizando la misma muestra que los dos trabajos previos, se pretende responder a la pregunta sobre la adherencia real a la declaración PRISMA de los estudios de MA sobre intervenciones educativas orientadas a mejorar el rendimiento académico entre 2009 y 2022 y si había diferencias en el cumplimiento de cada uno de los ítems entre los estudios que mencionan explícitamente el uso de la recomendación PRISMA comparado con aquellos que no la mencionan. En concreto, se plantean los siguientes objetivos específicos: i) Describir el cumplimiento real de cada una de las recomendaciones de la declaración PRISMA de los meta-análisis incluidos en la meta-revisión (objetivo 1); ii) Analizar la asociación del cumplimiento de cada uno de estos ítems en función de la declaración explícita de los autores de haber utilizado la declaración PRISMA (objetivo 2); iii) Evaluar la asociación entre el riesgo de sesgo (RoB) y el cumplimiento de cada una de las recomendaciones de la declaración PRISMA (objetivo 3); y iv) Identificar las características asociadas a la afirmación de los autores de haber utilizado la declaración PRISMA (objetivo 4.a) y al número de recomendaciones cumplidas (objetivo 4.b).

Método

Diseño de investigación

Meta-revisión o RS metodológica diseñada para evaluar la transparencia y calidad de las publicaciones de MA de intervenciones educativas centradas en diferentes resultados académicos (Sánchez-Martín et al., 2024a; Sánchez-Martín et al., 2025).

Protocolo y registro

Esta investigación ha sido redactada siguiendo las pautas de la guía *PRISMA 2020* (Page et al., 2021b). El protocolo se registró en OSF (Open Science Framework: <https://doi.org/10.17605/OSF.IO/PSKN6>).

Criterios de elegibilidad

Los criterios de inclusión fueron: i) MA de diseños experimentales y cuasiexperimentales que evalúan la efectividad de las intervenciones educativas en el desempeño de resultados académicos; ii) escritos en inglés o español; y iii) publicados desde el 1 de enero de 2009 [fecha de publicación de la primera guía PRISMA (Moher et al., 2009)] hasta el 30 de abril de 2022. Los criterios de exclusión fueron: i) MA basados en diseños no experimentales, así como revisiones narrativas tradicionales u otro tipo de publicaciones (conferencia o póster, capítulo de libro, artículo teórico...) ii) MA que evalúan el efecto de las intervenciones educativas sobre otros resultados (p. ej.: mejora cognitiva, habilidades sociales, etc.); y iii) artículos no accesibles al texto completo.

Fuentes de información y estrategias de búsqueda en bases de datos

Las búsquedas se realizaron en las bases de datos: ERIC, WoS, SCOPUS y PubMed. Las estrategias de búsqueda específicas para cada una de las bases de datos utilizadas han sido publicadas previamente en detalle (Sánchez-Martín et al., 2024a). De manera general, la cadena de búsqueda se desarrolló utilizando las siguientes palabras clave, incluyendo sinónimos y otros términos relacionados: “metaanálisis” AND “intervención educativa” AND “resultados académicos” y sus equivalentes en inglés. También se realizó una búsqueda manual en las referencias bibliográficas de los artículos seleccionados.

Selección de los estudios y extracción de información

Las referencias de los MA identificados fueron introducidos en Zotero donde una evaluadora realizó la primera fase de depuración, eliminando duplicados y artículos claramente no relacionados con el objetivo del estudio (screening). Posteriormente, dos revisoras independientes realizaron la selección a título y resumen y a texto completo en función de los criterios de selección, obteniendo un Kappa de Cohen (κ) de .869. Los desacuerdos fueron resueltos por consenso y/o por la intervención de un tercer evaluador.

Para la extracción de los datos se elaboró un protocolo previo al trabajo de campo. En él se incluyó, entre otras, las siguientes variables:

sugerencia de los editores de la revista de adhesión a la PRISMA, adherencia a la PRISMA, evaluación de los ítems individuales de la PRISMA, año de publicación, estrategia de búsqueda (regional o nacional/internacional), país, continente, área de educación (educación, psicología o salud), tipo de intervención, características de la población diana (según nivel educativo), y el RoB mediante la evaluación de la AMSTAR 2. El proceso de extracción de la información fue realizado por dos investigadoras de forma independiente mediante la utilización de un protocolo previamente elaborado. Con un acuerdo inicial entre evaluadores de $\kappa=0.897$, los desacuerdos fueron resueltos mediante consenso y/o por la intervención de un tercer evaluador.

Declaración PRISMA

Se utilizó la versión original de la PRISMA (Moher et al., 2009) en su versión española (Urrútia & Bonfill, 2010). Se eligió utilizar esta versión porque era la que tuvieron disponible la mayoría de los MA publicados en el momento de su realización. Este instrumento se incluye una lista de comprobación con 27 recomendaciones, agrupadas en siete grandes apartados: TÍTULO: 1. Título; II) RESUMEN: 2. Resumen estructurado; III) INTRODUCCIÓN: 3. Justificación; 4. Objetivos; IV) MÉTODO: 5. Protocolo y registro; 6. Criterios de elegibilidad; 7. Fuentes de información; 8. Búsqueda; 9. Selección de los estudios; 10. Proceso de extracción de los datos; 11. Lista de datos; 12. RoB en los estudios individuales; 13. Medidas de resumen; 14. Síntesis de resultados; 15. RoB entre los estudios; 16. Análisis adicionales; V) RESULTADOS: 17. Selección de estudios; 18. Características de los estudios; 19. RoB en los estudios; 20. Resultados de los estudios individuales; 21. Síntesis de los resultados; 22. RoB entre los estudios; 23. Análisis adicionales; VI) DISCUSIÓN: 24. Resumen de la evidencia; 25. Limitaciones; 26. Conclusiones; VII) FINANCIACIÓN: 27. Financiación.

Evaluación del riesgo de sesgo

En el estudio previo se describe en detalle la evaluación del RoB mediante la escala AMSTAR 2. En el primero de los análisis publicados se describe en detalle la escala AMSTAR 2 y los criterios utilizados para la evaluación del RoB (Sánchez-Martín et al., 2024a). Brevemente, se utilizó una evaluación menos restrictiva de los criterios de evaluación propuestos por los autores (Shea et al., 2017). Se definió una evaluación del RoB basada exclusivamente en el número de dominios críticos e independientemente del número de elementos no críticos y dicotomizada como: “riesgo bajo” (hasta dos debilidades críticas) y “riesgo alto” (tres o más debilidades

críticas). Esta adaptación fue decidida siguiendo las indicaciones de los autores originales que sugieren adaptar los criterios de evaluación según los objetivos de los investigadores dado que los resultados del 90% de los 69 estudios incluidos fueron evaluados con una confianza críticamente baja en la evaluación global de la AMSTAR 2 según los criterios originales (Sánchez-Martín et al., 2024a).

Síntesis de los resultados

La concordancia entre evaluadores tanto en el proceso de selección como en el de extracción de datos se midió mediante el índice Kappa de Cohen. El diagrama de flujo, según la PRISMA 2020 (Page et al., 2021a), al igual que la descripción de los estudios incluidos y una lista de los estudios excluidos y sus motivos de exclusión, han sido publicados previamente (Sánchez-Martín et al., 2024a). Se realizó un análisis descriptivo inicial de variables continuas (media y desviación estándar) y categóricas (frecuencia y porcentaje), como, por ejemplo, para la descripción de la adherencia a cada una de las recomendaciones de la declaración PRISMA, evaluada como variable dicotómica (Sí/No). Se utilizó la prueba de chi cuadrado o la prueba exacta de Fisher para analizar las diferencias según los autores declaren la utilización de la PRISMA y el RoB de cada uno de los criterios individuales de la PRISMA (objetivos 2 y 3).

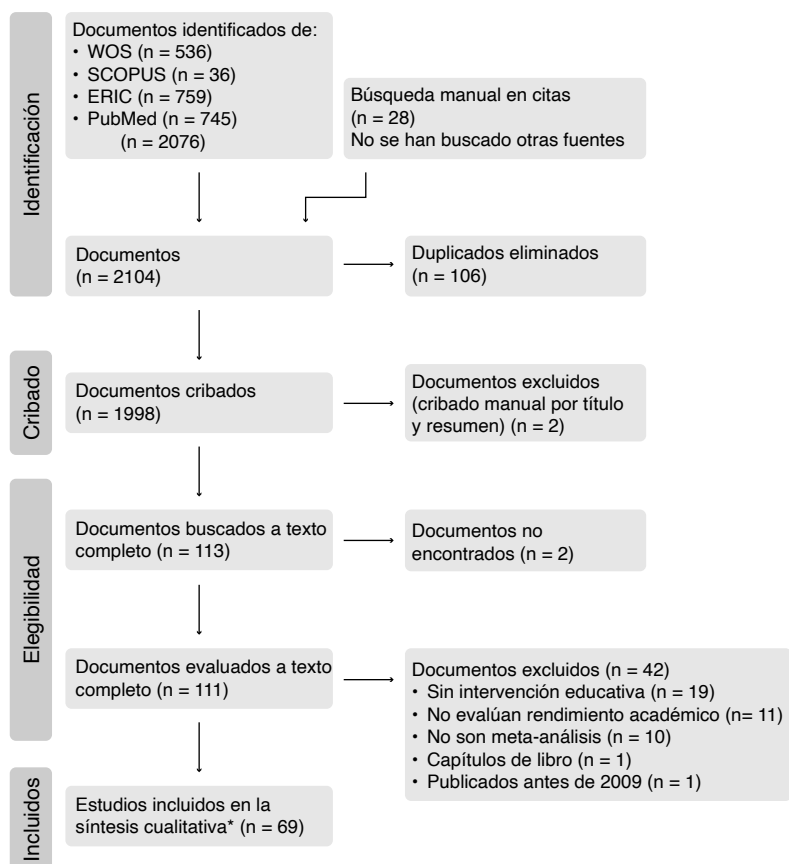
Se asumió como medida de asociación de las características de los estudios (variables independientes) y la declaración explícita de los autores de haber utilizado la PRISMA (evaluado como sí/no) como variable dependiente. Se calculó las odds ratio (OR) y sus intervalos de confianza del 95% (IC95%) con su p-valor mediante diferentes análisis de regresión logística simple (objetivo 4.a). De la misma forma, se evaluó la asociación de las mismas características con la puntuación global de cumplimiento de la PRISMA (variable dependiente) mediante diferentes modelos de regresión lineal simple, calculando la β , su IC95% y su p-valor (objetivo 4.b). Se utilizaron pruebas estadísticas bilaterales y con un nivel alfa de .05. Los análisis estadísticos se realizaron utilizando el software SPSS (V.28.0).

Resultados

La estrategia seguida para seleccionar los estudios se ha descrito con detalle en una publicación anterior, en la que también se puede consultar tanto el listado de artículos excluidos con sus respectivas razones, como las características particulares de los estudios finalmente incorporados (Sánchez-Martín et al., 2024a). El diagrama de flujo correspondiente a la

RS se presenta en el Gráfico I. De forma resumida, inicialmente se identificaron 2.076 estudios, de los cuales 69 cumplieron los criterios y fueron incluidos en los análisis finales. El listado de artículos excluidos y su justificación ha sido descrito previamente (Sánchez-Martín et al., 2024a).

Gráfico I. Diagrama de flujo



* Fuente: Sánchez Martín (2024b)

Las justificaciones generales de los estudios incluidos se presentan en la Tabla suplementaria I. En términos de adherencia a PRISMA, solo 24 estudios (34,8%) mencionaban explícitamente su utilización en el texto, mientras que 33 revistas (47,8%) la sugerían en sus directrices para autores. La puntuación media de adherencia a PRISMA fue de 19,7 (SD=4,4) sobre un máximo de 27. En relación con el RoB, únicamente 21 estudios

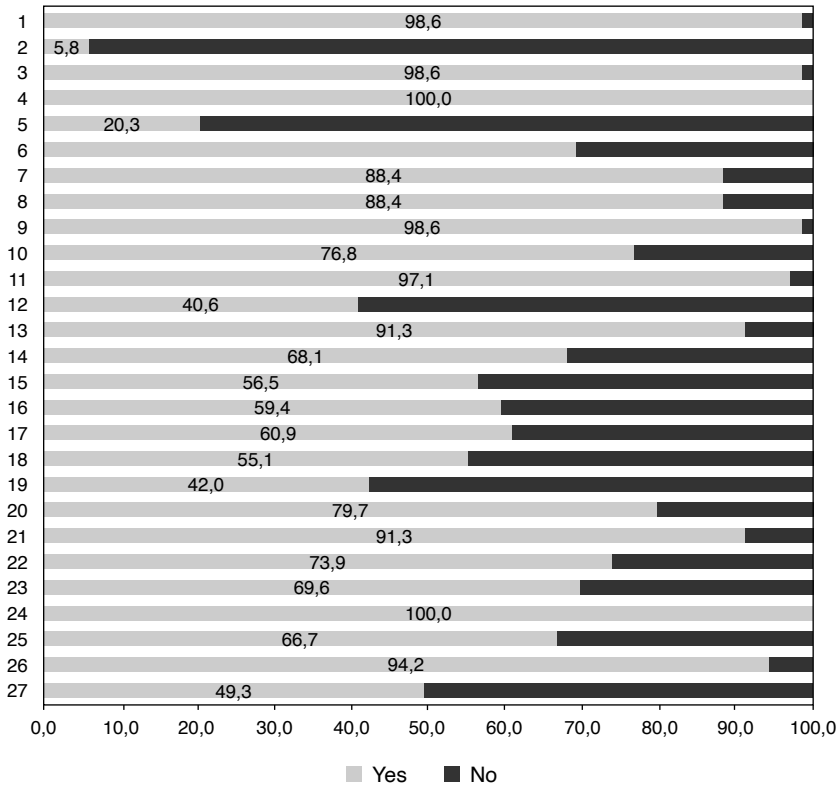
(30,4%) fueron evaluados con un bajo riesgo. La mayoría de los estudios fueron publicados a partir de 2017, y los países con mayor producción fueron, en orden decreciente: Estados Unidos, Turquía y China. Las publicaciones se agruparon en tres grandes áreas del conocimiento: i) Educación general, que concentró la mayoría de los estudios, ii) Intervenciones educativas en salud, e iii) Intervenciones en el ámbito de la psicología. El tipo de intervención más analizado correspondió a estrategias metodológicas. En cuanto a la población de referencia, el 34,8% de los estudios (n=24) abordaron intervenciones dirigidas a una combinación de niveles educativos (primaria, secundaria o educación especial y educación superior), mientras que el 27,5% (n=19) incluyó ambos niveles en sus análisis.

El Gráfico II y la Tabla I presentan los resultados globales de adherencia a las diferentes recomendaciones de la guía PRISMA (objetivo 1) y en la Tabla suplementaria II se presentan la adherencia de a cada una de las recomendaciones de PRISMA, así como la puntuación general obtenida para cada uno de los estudios individuales. En términos generales: un 51,8% de las recomendaciones (n=14) mostraron una adherencia superior al 75%. El rango de adherencia osciló entre un mínimo del 6% y un máximo del 100%. La recomendación con menor adherencia fue el resumen estructurado (6%). Por el contrario, las recomendaciones con un cumplimiento del 100% fueron: INTRODUCCIÓN: Objetivos (ítem 4), MÉTODOS: Criterios de elegibilidad (ítem 6) y DISCUSIÓN: Resumen de la evidencia (ítem 24). Un 18,5% de las recomendaciones (n=5) tuvieron una adherencia inferior al 50%, siendo las de menor cumplimiento: RESUMEN: Resumen estructurado (ítem 2), MÉTODOS: Protocolo y registro (ítem 5), MÉTODOS: Evaluación del RoB en estudios individuales (ítem 12), RESULTADOS: Evaluación del RoB en estudios (ítem 19) y DISCUSIÓN: Información sobre financiación (ítem 27).

Con relación al segundo objetivo “analizar la asociación del cumplimiento de cada uno de estos ítems en función de la declaración explícita de los autores de haber utilizado la declaración PRISMA”, se encontraron diferencias estadísticamente significativas en el cumplimiento de 14 de las 27 recomendaciones de PRISMA (51,8%) según si los autores declararon explícitamente su adherencia a la declaración (Tabla I). Las publicaciones que mencionaron su uso mostraron significativamente un mayor cumplimiento en las siguientes recomendaciones: RESUMEN: 2. Resumen estructurado. MÉTODOS: 5. Protocolo y registro; 7. Fuentes de información; 8. Búsqueda; 12. RoB en los estudios individuales; 15. RoB entre los estudios. RESULTADOS: 17. Selección de estudios; 18. Características de los estudios; 19. RoB en los estudios; 20. Resultados de los estudios individuales; 22. RoB entre los estudios; 23. Análisis adicionales.

DISCUSIÓN: 25. Limitaciones. Y FINANCIACIÓN: 27. Información sobre financiación.

Gráfico II. Porcentajes de cumplimiento de las recomendaciones de la declaración PRISMA de los estudios incluidos (n=69)



Recomendaciones de la declaración PRISMA: TÍTULO: 1. Título; RESUMEN: 2. Resumen estructurado; INTRODUCCIÓN: 3. Justificación; 4. Objetivos; MÉTODO: 5. Protocolo y registro; 6. Criterios de elegibilidad; 7. Fuentes de información; 8. Búsqueda; 9. Selección de los estudios; 10. Proceso de extracción de los datos; 11. Lista de datos; 12. Riesgo de sesgo en los estudios individuales; 13. Medidas de resumen; 14. Síntesis de resultados; 15. Riesgo de sesgos entre los estudios; 16. Análisis adicionales; RESULTADOS: 17. Selección de estudios; 18. Características de los estudios; 19. Riesgos de sesgo en los estudios; 20. Resultados de los estudios individuales; 21. Síntesis de los resultados; 22. Riesgo de sesgo entre los estudios; 23. Análisis adicionales; DISCUSIÓN: 24. Resumen de la evidencia; 25. Limitaciones; 26. Conclusiones; FINANCIACIÓN: 27. Financiación.

El análisis de la asociación entre el riesgo de sesgos (RoB) y el cumplimiento de cada una de las recomendaciones de la PRISMA (objetivo 3) identificó que el cumplimiento de 13 de las 27 recomendaciones (48,1%)

se asoció significativamente con un menor RoB (Tabla I). Estas fueron: MÉTODOS: 5. Protocolo y registro; 10. Proceso de extracción de datos; 12. RoB en los estudios individuales; 14. Síntesis de resultados; 15. RoB entre los estudios; 16. Análisis adicionales. RESULTADOS: 17. Selección de estudios; 18. Características de los estudios; 19. RoB en los estudios; 20. Resultados de los estudios individuales; 22. RoB entre los estudios. DISCUSIÓN: 25. Limitaciones. Y FINANCIACIÓN: 27. Información sobre financiación. Sin embargo, cuatro recomendaciones en las que sí se había encontrado diferencias en su cumplimiento según la declaración de uso de PRISMA por parte de los autores, no se asociaron con diferencias en el RoB: RESUMEN: 2. Resumen estructurado. MÉTODOS: 7. Fuentes de información; 8. Búsqueda. RESULTADOS: 23. Análisis adicionales. En contraste, tres recomendaciones que no mostraron diferencias según la declaración de uso de PRISMA sí se asociaron con un menor RoB. Todas ellas pertenecen a la sección MÉTODOS: 10. Proceso de extracción de datos; 14. Síntesis de resultados; 16. Análisis adicionales.

En la Tabla II se presentan los resultados de los análisis de asociación entre las características de los estudios y la información de cumplimiento de la PRISMA por los autores (objetivo 4.a) y el número de recomendaciones de la PRISMA cumplimentadas realmente (objetivo 4.b). Cuatro características se asociaron con una mayor frecuencia a la declaración de uso de la guía PRISMA en el documento: i) un mayor número de recomendaciones seguidas (puntuación PRISMA), ii) un menor RoB, iii) publicaciones más recientes (entre 2017 y 2022), así como iv) estudios centrados en el área de educación en salud y con intervenciones relacionadas con actividades físicas. Por otro lado, siete características se vincularon significativamente con un mayor número de recomendaciones seguidas: i) que la revista sugiriera el uso de la guía PRISMA en sus directrices para autores, ii) que los autores declararan su adherencia a PRISMA, iii) un menor RoB, iv) publicación en años recientes (2017-2022), v) estudios enmarcados en las áreas de educación en psicología y educación en salud, vi) intervenciones relacionadas con actividades físicas, y vii) población diana compuesta por estudiantes de educación superior, en comparación con los niveles de educación primaria, secundaria y educación especial.

Tabla I. Cumplimiento de las recomendaciones individuales de la declaración PRISMA según la descripción de su utilización y el RoB asociado

	Los autores declaran la utilización de la declaración PRISMA										RoB #		
	Adherencia General a la PRISMA		No Cumplimiento		Sí Cumplimiento		Alto Cumplimiento		Bajo Cumplimiento		p-valor	n (%)	p-valor
	n (%)	n (%)	n (%)	n (%)	n (%)	n (%)	n (%)	n (%)	n (%)	n (%)			
TÍTULO													
1. Título	68 (98.6)	1 (2.2)	0 (0)	44 (97.8)	24 (100.0)	1 (2.1)	47 (97.9)	0 (0.0)	21 (100.0)	1.000*			1.000*
RESUMEN													
2. Resumen estructurado	4 (5.8)	45 (100.0)	0 (0.0)	0 (83.3)	4 (16.7)	47 (97.9)	1 (2.1)	18 (84.7)	3 (14.3)	.081*			.081*
INTRODUCCIÓN													
3. Justificación	68 (98.6)	0 (0.0)	45 (100.0)	1 (4.2)	23 (95.8)	1 (2.1)	47 (97.9)	0 (0.0)	21 (100.0)	.348*			1.000*
4. Objetivos	69 (100.0)	0 (0.0)	45 (100.0)	0 (0.0)	24 (100.0)	0 (0.0)	48 (100)	0 (0.0)	21 (100.0)	1.000*			-
MÉTODOS													
5. Protocolo y registro	14 (20.3)	41 (91.1)	4 (8.9)	14 (58.3)	10 (41.7)	44 (91.7)	4 (8.3)	11 (52.4)	10 (47.6)	.003*			<.001*
6. Criterios de elegibilidad	69 (100.0)	0 (0.0)	45 (100.0)	0 (0.0)	24 (100.0)	0 (0.0)	48 (100.0)	0 (0.0)	21 (100.0)	-			-
7. Fuentes de información	61 (88.4)	8 (17.8)	37 (82.2)	0 (0.0)	24 (100.0)	8 (16.7)	40 (83.3)	0 (0.0)	21 (100.0)	.028*			.095
8. Búsqueda	61 (88.4)	8 (17.8)	37 (82.2)	0 (0.0)	24 (100.0)	7 (14.6)	41 (85.4)	1 (4.8)	20 (95.2)	.044*			.419

	Los autores declaran la utilización de la declaración PRISMA										RoB #	
	Adherencia General a la PRISMA		No Cumplimiento		Sí Cumplimiento		Alto Cumplimiento		Bajo Cumplimiento		p-valor	p-valor
	n (%)	n (%)	Sí n (%)	No n (%)	Sí n (%)	No n (%)	Sí n (%)	No n (%)	Sí n (%)	No n (%)		
9. Selección de estudios	68 (98.6)	1 (2.2)	44 (97.8)	0 (0.0)	24 (100.0)	1 (2.1)	47 (97.9)	0 (0.0)	21 (100.0)	0 (0.0)	1.000*	1.000*
10. Proceso de extracción de datos	53 (76.8)	12 (26.7)	33 (73.3)	4 (16.7)	20 (83.3)	15 (31.2)	33 (68.8)	1 (4.8)	20 (95.2)	1 (4.8)	.349	.027
11. Lista de datos	67 (97.1)	2 (4.4)	43 (95.6)	0 (0.0)	24 (100.0)	2 (4.2)	46 (95.8)	0 (0.0)	21 (100.0)	0 (0.0)	.540	1.000*
12. RoB estudios individuales	28 (40.6)	35 (77.8)	10 (22.2)	6 (25.0)	18 (75.0)	39 (81.2)	9 (18.8)	2 (9.5)	19 (90.5)	2 (9.5)	<.001	<.001
13. Medidas de resumen	63 (91.3)	4 (8.9)	41 (91.1)	2 (8.3)	22 (91.7)	3 (6.2)	45 (93.8)	3 (14.3)	18 (85.7)	3 (14.3)	1.000*	.359
14. Síntesis de resultados	47 (68.1)	17 (37.8)	28 (62.2)	5 (20.8)	19 (79.2)	21 (43.8)	27 (56.2)	1 (4.8)	20 (95.2)	1 (4.8)	.150	.001
15. RoB entre los estudios	39 (56.5)	25 (55.6)	20 (44.4)	5 (20.8)	19 (79.2)	27 (56.2)	21 (43.8)	3 (14.3)	18 (85.7)	3 (14.3)	.006	.001
16. Análisis adicionales	41 (59.4)	22 (48.9)	23 (51.1)	6 (25.0)	18 (75.0)	24 (50.0)	24 (50.0)	4 (19.0)	17 (81.0)	4 (19.0)	.054	.016
RESULTADOS												
17. Selección de los estudios	42 (60.9)	26 (57.8)	19 (42.2)	1 (4.2)	23 (95.8)	26 (54.2)	22 (45.8)	1 (4.8)	20 (95.2)	1 (4.8)	<.001	<.001
18. Características de los estudios	38 (55.1)	27 (60.0)	18 (40.0)	4 (16.7)	20 (83.3)	26 (54.2)	22 (45.8)	5 (23.8)	16 (76.2)	5 (23.8)	.001	.020

Los autores declaran la utilización de la declaración PRISMA										RoB #		
Adherencia General a la PRISMA	No Cumplimiento			Sí Cumplimiento			Alto Cumplimiento			Bajo Cumplimiento		
	n (%)	Sí n (%)	No n (%)	No n (%)	Sí n (%)	No n (%)	No n (%)	Sí n (%)	No n (%)	Sí n (%)	No n (%)	p-valor
19. RoB en los estudios	29 (42.0)	36 (80.0)	9 (20.0)	4 (16.7)	20 (83.3)	38 (79.2)	10 (20.8)	2 (9.5)	19 (90.5)	<.001		
20. Resultados de los estudios individuales	55 (79.7)	14 (31.1)	31 (69.9)	0 (0.0)	24 (100.0)	13 (27.1)	35 (72.9)	1 (4.5)	20 (95.2)	.001*		
21. Síntesis de los resultados	63 (91.3)	6 (13.3)	39 (86.7)	0 (0.0)	24 (100.0)	6 (12.5)	42 (87.5)	0 (0.0)	21 (100.0)	.085*		
22. RoB entre los estudios	51 (73.9)	15 (33.3)	30 (66.7)	3 (12.5)	21 (87.5)	17 (35.4)	31 (64.6)	1 (4.8)	20 (95.2)	.061		
23. Análisis adicionales	48 (69.6)	18 (40.0)	27 (60.0)	3 (12.5)	21 (87.5)	18 (37.5)	30 (62.5)	3 (14.3)	18 (85.7)	.018		
DISCUSIÓN												
24. Resumen de la evidencia	69 (100.0)	0 (0.0)	45 (100.0)	0 (0.0)	24 (100.0)	0 (0.0)	48 (100)	0 (0.0)	21 (100.0)	-		
25. Limitaciones	46 (66.7)	21 (46.7)	24 (53.3)	2 (8.3)	22 (91.7)	20 (41.7)	28 (58.3)	3 (14.3)	18 (85.7)	.001		
26. Conclusiones	65 (94.2)	3 (6.7)	42 (93.3)	1 (4.2)	23 (95.8)	3 (6.2)	45 (93.8)	1 (4.8)	20 (95.2)	1.000*		
FINANCIACIÓN												
27. Financiación	34 (49.3)	31 (68.9)	14 (31.1)	4 (16.7)	20 (83.3)	31 (64.6)	17 (35.4)	4 (19.0)	17 (81.0)	<.001		

* Test exacto de Fisher; # RoB (Risk of Bias o Riesgo de Sesgos): Bajo riesgo (< 2 dimensiones críticas); Alto riesgo (>2 dimensiones críticas). Fuente: Elaboración propia.

Tabla II. Asociación de las características de los estudios incluidos según la afirmación de los autores de haber cumplido las recomendaciones y el número de ítems cumplimentados de la declaración PRISMA

	Los autores afirman informar según PRISMA					N° recomendaciones PRISMA					
	No N (%)	Si N (%)	OR	Inf	Sup	p-valor	N° (SD)	β	Inf	Sup	p-valor
La revista recomienda el uso de la declaración PRISMA											
No	26 (57.8)	10 (41.7)	Ref				18.2 (4.4)	Ref			
Si	19 (42.2)	14 (58.3)	1.9	.7	5.2	.269	21.3 (3.7)	3.1	1.1	5.1	.003
Los autores afirman informar según declaración PRISMA											
No	-	-					17.7 (3.8)	Ref			
Si	-	-	Ref				23.4 (2.4)	5.7	4.0	7.5	<.001
N° recomendaciones PRISMA (SD)	17.7 (3.8)	23.5 (2.4)	1.8	1.3	2.3	<.001	-				
RoB											
Alto	37 (82.2)	11 (45.8)	Ref				17.8 (3.8)	Ref			
Bajo	8 (17.8)	13 (54.2)	5.5	1.8	16.6	.003	23.9 (2.0)	6.8	4.3	7.8	<.001
Año de publicación											
2009-2016	26 (57.8)	5 (20.8)	Ref				18.3 (4.1)	Ref			

		Los autores afirman informar según PRISMA					N° recomendaciones PRISMA				
		95%CI					95%CI				
	No N (%)	Si N (%)	OR	Inf	Sup	p-valor	N° (SD)	β	Inf	Sup	p-valor
2017-2022	19 (42.2)	19 (79.2)	5.2	1.6	16.4	.005	20.9 (4.3)	2.6	.5	4.6	.014
Estrategia de búsqueda con restricciones geográficas											
Regionales o nacionales	11 (24.4)	2 (8.3)	Ref				17.8 (5.1)	Ref			
Internacionales	34 (75.6)	22 (91.7)	3.6	.7	17.6	.120	20.2 (4.1)	2.4	-2	5	.075
Continentes											
Asia y Australia	22 (48.9)	10 (41.7)	Ref				18.7 (5.0)	Ref			
Europa	10 (22.2)	8 (33.3)	1.8	.5	5,8	.353	21.6 (3.4)	2.9	.4	5.4	.023
Norteamérica	13 (28.9)	6 (25.0)	1.0	.3	3.4	.980	19.6 (3.5)	0.9	-1.5	3.4	.447
Área de educación											
Educación	35 (77.8)	8 (33.3)	Ref				17.6 (4.0)	Ref			
Psicología	6 (13.3)	2 (8.3)	1.5	.2	8.6	.677	21.0 (1.9)	3.4	.8	6.0	.012
Salud	4 (8.9)	14 (58.3)	15.3	4.0	59.1	<.001	24.1 (1.8)	3.2	1.3	8.4	<.001

Tipo de intervención	Los autores afirman informar según PRISMA					N° recomendaciones PRISMA					
	95%CI					95%CI					
	No N (%)	Si N (%)	OR	Inf	Sup	p-valor	N° (SD)	β	Inf	Sup	p-valor
Estrategias metodológicas	21 (46.7)	8 (37.5)	Ref				19.1 (4.1)	Ref			
Estrategias organizativas	10 (22.2)	7 (29.2)	1.6	.5	5.7	.439	19.8 (4.9)	0.7	-1.8	3.3	.577
Actividades físicas y otras	3 (6.7)	7 (29.2)	5.4	1.1	25.9	.033	22.8 (2.9)	3.7	.6	6.8	.020
Combinación	11 (24.4)	1 (4.2)	.2	.0	1.9	.165	18.5 (4.5)	-0.6	-3.5	2.3	.681
Población de origen											
Primaria, secundaria y diversidad funcional	6 (13.6)	4 (16.7)	Ref				18.0 (5.2)	Ref			
Educación superior	6 (13.6)	9 (37.5)	2.6	.5	13.1	.239	11.7 (3.8)	3.9	.7	7.1	.019
Todos los niveles incluidos	16 (36.4)	3 (12.5)	.3	.1	1.9	.328	17.6 (4.0)	-0.7	-3.8	2.4	.638
Combinación de niveles	16 (36.4)	8 (33.3)	.9	.2	3.9	.861	20.4 (3.8)	2.0	-1.0	5.0	.182

Fuente: Elaboración propia.

Conclusiones

Hasta donde conocemos, este es el primer estudio en Educación que analiza el nivel de transparencia y calidad de las publicaciones de MA de estudios experimentales, evaluada mediante la adherencia a la declaración PRISMA, y su relación con el RoB, calculado mediante la AMSTAR 2. Ambos instrumentos son ampliamente aceptados en distintas áreas del conocimiento (Moher et al., 2009; Shea et al., 2017). Los resultados obtenidos sugieren que la adherencia a PRISMA en las publicaciones de MA en educación es subóptima (34,8%) y que aquellas que declaran y cumplen sus recomendaciones se asocian, en general, a una mayor calidad metodológica y a un menor RoB.

En el ámbito educativo, apenas se han publicado estudios que analicen la calidad metodológica de los MA. Solo se han identificado dos investigaciones independientes (Ahn et al., 2012; Eser & Yurtçu, 2020) que emplean escalas diseñadas por los propios autores y no instrumentos estandarizados, lo cual limita la comparabilidad de los resultados entre estudios, así como con otras disciplinas. Aunque ninguno de ellos evaluó la adherencia a la declaración PRISMA, ambos trabajos coinciden en señalar deficiencias en la descripción de los estudios incluidos y recomiendan una mayor transparencia en la presentación de los datos. Ninguno de ellos evaluó la adherencia a la declaración PRISMA.

En educación médica, se han publicado estudios sobre la calidad de otros diseños de investigación, como ensayos experimentales, utilizando escalas propias (Cook et al., 2007) o listas de comprobación como la STROBE (*Strengthening the Reporting of Observational Studies in Epidemiology*) (Cook et al., 2011). Estos estudios concluyen que la calidad y transparencia de las publicaciones pueden mejorarse sustancialmente. En este ámbito educativo, la evaluación de la adherencia a PRISMA en RS ha sido más frecuente, con resultados que reflejan un uso inferior al deseable (Cullis et al., 2017; Innocenti et al., 2022; Javidan et al., 2023; Liu et al., 2017; Panic et al., 2013; Sun et al., 2019, 2021; Wasiak et al., 2017; Yang et al., 2024). Un estudio metodológico realizado por los creadores de PRISMA (Page & Moher, 2017) encontró que solo el 57% de los 100 MA analizados en MEDLINE cumplían alguno de sus criterios o extensiones. En casos excepcionales, la adherencia ha alcanzado el 86% (Sharma & Oremus, 2018). Por otra parte, algunos estudios han mostrado mejoras en la calidad de las publicaciones tras la introducción de la declaración PRISMA (Liu et al., 2019). También se han realizado evaluaciones de la adherencia a la MARS (*Meta-Analysis Reporting Standards*), diseñada por la American Psychological Association para mejorar la pre-

sentación de MA en ciencias sociales y educación (American Psychological Association Publications and Communications Board Working Group on Journal Article Reporting Standards, 2008) y Appelbaum et al., (2018). También se han detectado niveles de cumplimiento insuficientes tanto en psicología industrial y organizacional (Kepes et al., 2013; Schalken & Rietbergen, 2017) como en psicología en general (Hohn et al., 2019).

Los resultados de nuestro análisis muestran una variabilidad significativa en la transparencia y calidad de las publicaciones, similar a la observada en estudios previos en educación (Ahn et al., 2012; Eser & Yurtçu, 2020). Sin embargo, no es posible realizar una comparación directa entre estudios debido a las diferencias en los instrumentos utilizados. Además, las diferencias entre revistas educativas y médicas influyen en la forma en que se presentan los resultados (Blanco-Blanco, 2018). Por ejemplo, la recomendación de incluir un resumen estructurado (ítem 2) muestra una baja adherencia en nuestro estudio (5,8%), en contraste con la literatura médica, donde oscila entre el 59% y el 79% (Page & Moher, 2017). Entre las cinco recomendaciones con menor cumplimiento (RESUMEN: 2. Resumen estructurado; MÉTODOS: 5. Protocolo y registro; 12. RoB en los estudios individuales; RESULTADOS: 19. RoB en los estudios; y DISCUSIÓN: 27. Financiación), solo el resumen estructurado no mostró una asociación estadísticamente significativa con el RoB. Una posible explicación es que muchas revistas de educación no exigen resúmenes estructurados en sus directrices para autores.

La segunda recomendación con una menor adherencia (20,3%) es el ítem 5, sobre el protocolo y registro del estudio, un elemento crucial para reducir el sesgo de publicación y mejorar la transparencia. En educación médica, su cumplimiento alcanza el 21% en estudios posteriores a 2009 (Page & Moher, 2017). Iniciativas como PROSPERO (*International Prospective Register of Systematic Reviews*) y OSF (*Open Science Framework*) han sido promovidas para registrar protocolos y aumentar la credibilidad de las RS (Page et al., 2021b) y revistas como *Systematic Reviews* han comenzado a publicar protocolos antes de los resultados finales (<https://systematicreviewsjournal.biomedcentral.com/>). Por último, se ha observado que las RS con protocolos registrados muestran una mayor adherencia a PRISMA (Yang et al., 2024).

La evaluación del RoB en los estudios incluidos, tanto en la descripción del método utilizado (ítem 12), como en la presentación de los resultados (ítem 19), es otro factor clave asociado a la calidad metodológica. En nuestro estudio, la adherencia a estos ítems fue del 40.6% y 42.0%, respectivamente, y su cumplimiento se asoció con un menor RoB. En comparación, en el ámbito sanitario, los niveles de adherencia alcanzan el 65% (ítem 12)

y el 60% (ítem 19) (Page & Moher, 2017). Explicar con detalle cómo se han evaluado, los instrumentos utilizados y cómo se han tenido en cuenta para la síntesis de los resultados es crucial para evaluar la calidad de los resultados del propio MA. En los últimos años se han desarrollado numerosos instrumentos expresamente diseñados para evaluar el RoB de los diferentes diseños de investigación, como la propia AMSTAR 2, ya citada, para la evaluación de RS, la NOS (Newcastle-Ottawa Scale) para los estudios observacionales (Stang, 2010) o la MMAT (Mixed Methods Appraisal Tool) para los estudios con una metodología cuantitativa, cualitativa o mixta (Hong et al., 2018). La evaluación RoB entre los estudios (recomendaciones 15 y 22) se centra en la descripción del método realizado y de los resultados en la evaluación de cualquier riesgo que pudiera poner en peligro la evidencia acumulativa, por ejemplo, el sesgo de publicación (Thornton & Lee, 2000).

Entre las limitaciones del estudio cabe mencionar la restricción de los criterios de inclusión al centrarse en MA de intervenciones experimentales en educación para mejorar el rendimiento académico. Este criterio fue adoptado debido al creciente número de MA publicados en educación y la esperada heterogeneidad entre ellos. La elección del resultado de las intervenciones centradas en la mejora del rendimiento académico respondió a la necesidad de focalizar las intervenciones en un determinado resultado. Futuras meta-revisiones de MA y RS centradas en otros resultados de interés, así como con otros diseños de investigación permitirán un análisis más preciso del problema y una mayor generalización de los resultados a otros ámbitos de la educación. Por otra parte, dado que el objetivo era evaluar la calidad de los estudios publicados, la búsqueda se realizó en cuatro bases de datos y no se exploró literatura gris ni se contactó con autores clave. Además, la categorización del RoB se basó en criterios justificados categorizando los criterios de la AMSTAR 2 según el número de dimensiones críticas.

La declaración PRISMA (Moher et al., 2009) surgió para facilitar una lista de recomendaciones destinadas a mejorar la calidad y transparencia de los informes de RS y MA que sirviera de ayuda tanto a investigadores como a editores y revisores. Esta declaración está en constante evolución. Además de su reciente actualización, la PRISMA 2020 (Page et al., 2021b), han surgido otras extensiones como la PRISMA-P 2015 para protocolos (Moher et al., 2015). La EQUATOR Network (Enhancing the QUality and Transparency of health Research, <https://www.equator-network.org/>) facilita el acceso a estas herramientas y promueve su uso entre investigadores y editores (Sánchez-Martín et al., 2024b; Struthers et al., 2021). Si bien muchas listas de comprobación se han desarrollado en el ámbito sanitario, su implementación en educación puede ser benefi-

ciosa (Pandis & Fedorowicz, 2011; Sánchez-Martín et al., 2023). La replicabilidad y transparencia en investigación educativa requieren esfuerzos conjuntos de investigadores, editores y responsables de políticas científicas. Iniciativas como GoodReports (<https://www.goodreports.org/>) buscan facilitar el uso de estas herramientas desde las primeras etapas de redacción de artículos científicos. Además, revistas internacionales como The Lancet y PLOS han comenzado a exigir listas de comprobación, y en España, publicaciones como ESPIRAL. Cuadernos del Profesorado o Educación XXI han adoptado estrategias similares.

En resumen, el crecimiento de publicaciones científicas ha reforzado la necesidad de RS y MA como herramientas clave para avanzar en el conocimiento. En el ámbito educativo, asegurar la calidad de estos estudios es esencial. Este trabajo evalúa la adherencia a la declaración PRISMA de los estudios de MA sobre intervenciones educativas orientadas a mejorar el rendimiento académico entre 2009 y 2022. Los resultados sugieren que la adherencia a PRISMA sigue siendo baja, y que aquellos artículos que se adhieren a sus directrices se asocian a un menor RoB. Para fortalecer la investigación en educación es imprescindible fomentar la realización de meta-revisiones que permitan analizar las limitaciones de la investigación actual e identificar y promover estrategias que aumenten la adherencia real a las mejores recomendaciones disponibles, como la PRISMA 2020, para mejorar la transparencia y calidad de las publicaciones.

Financiación

Este estudio se llevó a cabo en el marco del proyecto de investigación “Transformación del aprendizaje en contextos híbridos para la inclusión educativa y laboral de colectivos especialmente vulnerables, con énfasis en UFM” (TYNDALL/UFM), referencia: PID2020-119194RB-I00, financiado por MCIN/AEI/10.13039/501100011033.

Declaración de intereses

Los/as autores/as declaran no tener conflicto de intereses.

Declaración de disponibilidad de los datos

El protocolo fue registrado previamente en OSF (Open Science Framework: <https://doi.org/10.17605/OSF.IO/PSKN6>). Todos los datos utilizados se han incluido en el artículo.

Contribuciones de autoría

MSM y FNM diseñaron el estudio. MSM y MGS participaron en el proceso de selección y extrajeron los datos de forma independiente. EOM y la

FNM participaron como terceros evaluadores cuando fue necesario. MSM y FNM realizaron el análisis de datos, la interpretación de los resultados y redactaron el manuscrito. Todos los autores lo revisaron críticamente y aprobaron la versión final para su publicación.

Referencias bibliográficas

- Ahn, S., Ames, A. J., & Myers, N. D. (2012). A Review of Meta-Analyses in Education: Methodological Strengths and Weaknesses. *Review of Educational Research*, 82(4), 436-476. <https://doi.org/10.3102/0034654312458162>
- American Psychological Association Publications and Communications Board Working Group on Journal Article Reporting Standards. (2008). Reporting standards for research in psychology: Why do we need them? What might they be? *Am. Psychol.*, 63, 839-851. <https://doi.org/10.1037/0003-066X.63.9.839>
- Anguera, M. T. (2023). Revisitando las revisiones sistemáticas desde la perspectiva metodológica. *RELIEVE-Revista Electrónica de Investigación y Evaluación Educativa*, 29(1), art. M4. <https://doi.org/10.30827/relieve.v29i1.27758>
- Appelbaum, M., Cooper, H., Kline, R. B., Mayo-Wilson, E., Nezu, A. M., & Rao, S. M. (2018). Journal article reporting standards for quantitative research in psychology: The APA Publications and Communications Board task force report. *American Psychologist*, 73(1), 3-25. <https://doi.org/10.1037/amp0000191>
- Blanco-Blanco, Á. (2018). Estado de las prácticas científicas e investigación educativa. Posibles retos para la próxima década. *Revista de Educación*, 381, 207-230. <https://doi.org/10.4438/1988-592X-RE-2017-381-386>
- Cook, D. A., Beckman, T. J., & Bordage, G. (2007). Quality of reporting of experimental studies in medical education: A systematic review. *Medical Education*, 41(8), 737-745. <https://doi.org/10.1111/j.1365-2923.2007.02777.x>
- Cook, D. A., Levinson, A. J., & Garside, S. (2011). Method and reporting quality in health professions education research: A systematic review. *Medical Education*, 45(3), 227-238. <https://doi.org/10.1111/j.1365-2923.2010.03890.x>
- Cullis, P. S., Gudlaugsdottir, K., & Andrews, J. (2017). A systematic review of the quality of conduct and reporting of systematic reviews and me-

- ta-analyses in paediatric surgery. *PloS One*, 12(4), e0175213. <https://doi.org/10.1371/journal.pone.0175213>
- Eser, M. T., & Yurtçu, M. (2020). A Review of Meta-Analysis Articles in Educational Sciences Conducted Between 2010 and 2019 in Turkey. *International Journal of Progressive Education*, 16(6), 15-32. <https://doi.org/10.29329/ijpe.2020.280.2>
- Glass, G. V. (1976). Primary, Secondary, and Meta-Analysis of Research. *Educational Researcher*, 5(10), 3-8. <https://doi.org/10.2307/1174772>
- Hederich, C., Martínez, J., & Rincón, L. (2014). Hacia una educación basada en la evidencia. *Revista Colombiana de Educación*, 66, 19-54. <https://doi.org/10.17227/01203916.66rce1954>
- Hohn, R. E., Slaney, K. L., & Tafreshi, D. (2019). Primary Study Quality in Psychological Meta-Analyses: An Empirical Assessment of Recent Practice. *Frontiers in Psychology*, 9, 2667. <https://doi.org/10.3389/fpsyg.2018.02667>
- Holmqvist, M., & Lantz Ekström, M. (2024). A systematic review of research on educational superintendents. *Cogent Education*, 11(1), 2307142. <https://doi.org/10.1080/2331186X.2024.2307142>
- Hong, Q. N., Gonzalez-Reyes, A., & Pluye, P. (2018). Improving the usefulness of a tool for appraising the quality of qualitative, quantitative and mixed methods studies, the Mixed Methods Appraisal Tool (MMAT). *Journal of Evaluation in Clinical Practice*, 24(3), 459-467. <https://doi.org/10.1111/jep.12884>
- Howley, L., Szauter, K., Perkowski, L., Clifton, M., McNaughton, N., & Association of Standardized Patient Educators (ASPE). (2008). Quality of standardised patient research reports in the medical education literature: Review and recommendations. *Medical Education*, 42(4), 350-358. <https://doi.org/10.1111/j.1365-2923.2007.02999.x>
- Innocenti, T., Feller, D., Giagio, S., Salvioli, S., Minnucci, S., Brindisino, F., Cosentino, C., Piano, L., Chiarotto, A., & Ostelo, R. (2022). Adherence to the PRISMA statement and its association with risk of bias in systematic reviews published in rehabilitation journals: A meta-research study. *Brazilian Journal of Physical Therapy*, 26(5), 100450. <https://doi.org/10.1016/j.bjpt.2022.100450>
- Ioannidis, J. P. A. (2018). Meta-research: Why research on research matters. *PLOS Biology*, 16(3), e2005468. <https://doi.org/10.1371/journal.pbio.2005468>
- Ioannidis, J. P. A., Fanelli, D., Dunne, D. D., & Goodman, S. N. (2015). Meta-research: Evaluation and Improvement of Research Methods and Practices. *PLoS Biology*, 13(10), e1002264. <https://doi.org/10.1371/journal.pbio.1002264>

- Javidan, A., Alaichi, J., Nassar, Y., Li, A., Balta, K. Y., & Naji, F. (2023). Completeness of reporting in systematic reviews and meta-analyses in vascular surgery. *Journal of Vascular Surgery*, 78(6), 1550-1558.e2. <https://doi.org/10.1016/j.jvs.2023.04.009>
- Kepes, S., McDaniel, M. A., Brannick, M. T., & Banks, G. C. (2013). Meta-analytic Reviews in the Organizational Sciences: Two Meta-analytic Schools on the Way to MARS (the Meta-analytic Reporting Standards). *Journal of Business and Psychology*, 28(2), 123-143.
- Liu, H., Zhou, X., Yu, G., & Sun, X. (2019). The effects of the PRISMA statement to improve the conduct and reporting of systematic reviews and meta-analyses of nursing interventions for patients with heart failure. *International Journal of Nursing Practice*, 25(3), e12729. <https://doi.org/10.1111/ijn.12729>
- Liu, X., Kinzler, M., Yuan, J., He, G., & Zhang, L. (2017). Low Reporting Quality of the Meta-Analyses in Diagnostic Pathology. *Archives of Pathology & Laboratory Medicine*, 141(3), 423-430. <https://doi.org/10.5858/arpa.2016-0144-OA>
- Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *BMJ*, 339, b2535. <https://doi.org/10.1136/bmj.b2535>
- Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., Shekelle, P., & Stewart, L. A. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Systematic Reviews*, 4(1). <https://doi.org/10.1186/2046-4053-4-1>
- Moreno, E. M., Montilla-Arechabala, C., & Maldonado, M. A. (2022). Effectiveness and characteristics of programs for developing oral competencies at university: A systematic review. *Cogent Education*, 9(1), 2149224. <https://doi.org/10.1080/2331186X.2022.2149224>
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., ... Moher, D. (2021a). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372(71). <https://doi.org/10.1136/bmj.n71>
- Page, M. J., & Moher, D. (2017). Evaluations of the uptake and impact of the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) Statement and extensions: A scoping review. *Systematic Reviews*, 6(1), 263, 1-14. <https://doi.org/10.1186/s13643-017-0663-8>

- Page, M. J., Moher, D., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., ... McKenzie, J. E. (2021b). PRISMA 2020 explanation and elaboration: Updated guidance and exemplars for reporting systematic reviews. *BMJ*, *372*(160). <https://doi.org/10.1136/bmj.n160>
- Pandis, N., & Fedorowicz, Z. (2011). The international EQUATOR network: Enhancing the quality and transparency of health care research. *Journal of Applied Oral Science: Revista FOB*, *19*(5). <https://doi.org/10.1590/s1678-77572011000500001>
- Panic, N., Leoncini, E., de Belvis, G., Ricciardi, W., & Boccia, S. (2013). Evaluation of the endorsement of the preferred reporting items for systematic reviews and meta-analysis (PRISMA) statement on the quality of published systematic review and meta-analyses. *PloS One*, *8*(12), e83138. <https://doi.org/10.1371/journal.pone.0083138>
- Rubio-Aparicio, M., Sánchez-Meca, J., Marín-Martínez, F., & López-López, J. A. (2018). Guidelines for Reporting Systematic Reviews and Meta-analyses. *Anales de Psicología*, *34*(2), 412. <https://doi.org/10.6018/analesps.34.2.320131>
- Ruiz-Corbella, M., Galán, A., & Diestro, A. (2014). Las revistas científicas de Educación en España: Evolución y perspectivas de futuro. *RELIEVE*, *20*(2). <https://doi.org/10.7203/relieve.20.2.4361>
- Sánchez-Martín, M., Gutiérrez-Sánchez, M., Olmedo-Moreno, E. M. & Navarro-Mateu, F. (2025). Calidad del diagrama de flujo de los meta-análisis en Educación: una revisión sistemática [Quality of flowcharts in meta-analyses in education: A systematic review]. *Espiral Cuadernos del Profesorado*, *18*(37), 89-110. <https://doi.org/10.25115/ecp.v18i37.10159>
- Sánchez-Martín, M., Gutiérrez-Sánchez, M., Olmedo-Moreno, E. M., & Navarro-Mateu, F. (2024a). A systematic review to evaluate the risk of bias of meta-analyses reporting experimental educational interventions focused on academic performance. *Cogent Education*, *11*(1), 2385785. <https://doi.org/10.1080/2331186X.2024.2385785>
- Sánchez-Martín, M., Navarro-Mateu, F., & Sánchez-Meca, J. (2022). Las revisiones sistemáticas y la educación basada en evidencias [Systematic reviews and evidence-based education]. *Espiral. Cuadernos del Profesorado*, *15*(30), 108-120. <https://doi.org/10.25115/ecp.v15i30.7860>
- Autor... (2022)
- Sánchez-Martín, M., Olmedo Moreno, E. M., Gutiérrez-Sánchez, M., & Navarro-Mateu, F. (2024b). EQUATOR-Network: una hoja de ruta para

- mejorar la calidad y transparencia de la investigación [EQUATOR-Network: a roadmap to improve the quality and transparency of research reporting]. *Espiral. Cuadernos del Profesorado*, 17(35). <https://doi.org/10.25115/ecp.v17i35.9529>
- Sánchez-Martín, M., Ponce, A., Plana, M., Ibáñez-López, F., & Navarro-Mateu, F. (2023). There are many lists of research, but... not all serve the same [Existen muchas listas en investigación, pero ... no todas sirven para lo mismo]. *Espiral. Cuadernos del Profesorado.*, 16(33), 81-91. <https://doi.org/10.25115/ecp.v16i33.9334>
- Sánchez-Meca, J., Boruch, R. F., & Petrosino, A. (2002). La Colaboración Campbell y la Práctica Basada en la Evidencia. *Papeles del Psicólogo*, 83, 44-48.
- Sánchez-Meca, J., Marín-Martínez, F., López-López, J. A., Núñez-Núñez, R. M., Rubio-Aparicio, M., López-García, J. J., López-Pina, J. A., Blázquez-Rincón, D. M., López-Ibáñez, C., & López-Nicolás, R. (2021). Improving the reporting quality of reliability generalization meta-analyses: The REGEMA checklist. *Research Synthesis Methods*, 12(4), 516-536. <https://doi.org/10.1002/jrsm.1487>
- Schalken, N., & Rietbergen, C. (2017). The Reporting Quality of Systematic Reviews and Meta-Analyses in Industrial and Organizational Psychology: A Systematic Review. *Frontiers in Psychology*, 8, 1395. <https://doi.org/10.3389/fpsyg.2017.01395>
- Sharma, S., & Oremus, M. (2018). PRISMA and AMSTAR show systematic reviews on health literacy and cancer screening are of good quality. *Journal of Clinical Epidemiology*, 99, 123-131. <https://doi.org/10.1016/j.jclinepi.2018.03.012>
- Shea, B. J., Reeves, B. C., Wells, G., Thuku, M., Hamel, C., Moran, J., Moher, D., Tugwell, P., Welch, V., Kristjansson, E., & Henry, D. A. (2017). AMSTAR 2: A critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *BMJ (Clinical Research Ed.)*, 358, j4008. <https://doi.org/10.1136/bmj.j4008>
- Stang, A. (2010). Critical evaluation of the Newcastle-Ottawa scale for the assessment of the quality of nonrandomized studies in meta-analyses. *European Journal of Epidemiology*, 25(9), 603-605. <https://doi.org/10.1007/s10654-010-9491-z>
- Struthers, C., Harwood, J., de Beyer, J. A., Dhiman, P., Logullo, P., & Schlüssel, M. (2021). GoodReports: Developing a website to help health researchers find and use reporting guidelines. *BMC Medical Research Methodology*, 21(1), 217. <https://doi.org/10.1186/s12874-021-01402-x>

- Sun, X., Wang, D., Wang, M., Li, H., & Liu, B. (2021). The reporting and methodological quality of systematic reviews and meta-analyses of nursing interventions for chronic obstructive pulmonary disease. A systematic review. *Nursing Open*, 8(3), 1489-1500. <https://doi.org/10.1002/nop2.767>
- Sun, X., Zhou, X., Zhang, Y., & Liu, H. (2019). Reporting and Methodological Quality of Systematic Reviews and Meta-Analyses of Nursing Interventions in Patients With Alzheimer’s Disease: General Implications of the Findings. *Journal of Nursing Scholarship: An Official Publication of Sigma Theta Tau International Honor Society of Nursing*, 51(3), 308-316. <https://doi.org/10.1111/jnu.12462>
- Thornton, A., & Lee, P. (2000). Publication bias in meta-analysis: Its causes and consequences. *Journal of Clinical Epidemiology*, 53(2), 207-216. [https://doi.org/10.1016/s0895-4356\(99\)00161-4](https://doi.org/10.1016/s0895-4356(99)00161-4)
- Urrútiá, G., & Bonfill, X. (2010). Declaración PRISMA: una propuesta para mejorar la publicación de revisiones sistemáticas y metaanálisis. *Med Clin (Barc)*, 135(11), 507-511.
- Wasiak, J., Tyack, Z., Ware, R., Goodwin, N., & Faggion, C. M. (2017). Poor methodological quality and reporting standards of systematic reviews in burn care management. *International Wound Journal*, 14(5), 754-763. <https://doi.org/10.1111/iwj.12692>
- Yang, Q., Xian, H., Cheng, X., Wu, X., Meng, J., Chen, W., & Zeng, Z. (2024). Methodological and reporting quality assessment of systematic reviews and meta-analyses in the association between sleep duration and hypertension. *Systematic Reviews*, 13(1), 211. <https://doi.org/10.1186/s13643-024-02622-0>

Anexos

Tabla suplementaria I. Descripción general de los estudios incluidos

Total	Muestra general (n, %)
	69 (100.0)
La revista recomienda el uso de la declaración PRISMA	
No	36 (52.2)
Si	33 (47.8)
Los autores afirman informar según declaración PRISMA	
No	45 (65.2)
Si	24 (34.8)

Total	Muestra general (n, %)
	69 (100.0)
N° recomendaciones PRISMA (Media, SD)	19.7 (4.4)
RoB *	
Alto	48 (69.6)
Bajo	21 (30.4)
Año de publicación (Media, SD)	2017.5 (2.9)
2009-2016	23 (32.9)
2017-2022	47 (67.1))
Estrategia de búsqueda con restricciones geográficas	
Regionales o nacionales	56 (80.0)
Internacionales	14 (20.0)
Países	
EEUU	18 (26.1)
Canadá	1 (1.4)
UK	6 (8.7)
Países bajos	6 (8.7)
Portugal	1 (1.4)
Noruega	1 (1.4)
Finlandia	1 (1.4)
España	3 (4.3)
Australia	1 (1.4)
Turquía	12 (17.4)
Irán	2 (2.9)
China	11 (15.9)
Taiwan	2 (2.9)
India	3 (4.3)
Filipinas	1 (1.4)
Continentes	
Australia & Asia	32 (46.4)
Europa	18 (26.1)
Norteamérica	19 (27.5)
Área de educación	
Educación	44 (62.3)
Psicología	8 (11.6)

Total	Muestra general (n, %)
	69 (100.0)
Salud	18 (26.1)
Tipo de intervención	
Estrategias metodológicas	30 (43.5)
Estrategias organizativas	17 (24.6)
Actividades físicas y otras	10 (14.5)
Combinación	12 (17.4)
Población de origen	
Primaria, secundaria y diversidad funcional	10 (14.5)
Educación superior	15 (21.7)
Todos los niveles incluidos	19 (27.5)
Combinación de niveles	24 (34.8)

Fuente: Elaboración propia

Tabla suplementaria II. Cumplimiento de las recomendaciones de la declaración PRISMA

Primer autor (año de publicación)	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	Nº recomendaciones cumplimentadas
Aydin (2021)	S	N	S	S	N	S	S	S	S	N	S	N	N	N	S	N	N	N	N	S	S	Y	N	S	S	S	N	14
Akar (2020)	S	N	S	S	N	S	S	S	N	N	S	S	S	N	N	N	N	N	N	S	S	N	S	S	S	S	N	14
Aktamis (2016)	S	N	S	S	N	S	N	S	S	N	S	S	S	S	S	S	N	N	N	N	S	S	S	S	S	S	N	16
Alegre-Ansuategui (2018)	S	N	S	S	N	S	S	S	S	N	S	N	S	S	N	S	N	S	N	S	S	N	S	S	N	N	N	16
Alvarez-Bueno (2017)	S	N	S	S	S	S	S	S	S	Y	S	S	S	S	S	S	S	S	S	S	S	S	S	S	N	N	S	25
Aspiranti (2018)	S	N	S	S	N	S	S	S	S	N	S	N	N	N	N	N	S	S	N	N	N	S	S	S	S	S	N	14
Ayaz (2015)	S	N	S	S	N	S	S	S	S	S	S	S	S	N	N	N	S	N	N	S	S	S	S	S	S	S	N	17
Bai (2020)	S	N	S	S	N	S	S	S	S	N	S	S	S	S	S	N	S	S	S	S	S	S	S	S	S	S	N	22
Balakrishnan (2021)	S	N	S	S	N	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	25
Bas (2016)	S	N	S	S	N	S	S	S	S	S	S	N	S	S	N	N	N	S	N	S	S	S	S	S	S	S	N	17
Bedard (2019)	S	N	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	25
Brierly (2021)	S	N	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	N	S	S	S	S	S	S	S	S	S	25

Primer autor (año de publicación)	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	Nº recomendaciones cumplimentadas
Capar (2015)	S	N	S	S	N	S	S	S	S	N	S	S	S	N	N	N	S	N	N	N	S	N	S	S	S	S	N	10
Cartiff (2021)	S	N	S	S	S	S	S	S	S	S	S	N	S	S	S	S	S	S	N	S	S	S	S	S	S	S	N	23
Celio (2011)	S	N	S	S	N	S	S	S	N	S	S	S	S	N	N	N	N	N	N	N	S	N	S	S	S	S	N	13
Cen (2021)	S	N	N	S	N	S	S	S	S	S	S	S	S	S	N	S	S	S	S	S	S	S	S	S	S	S	S	22
Chandran (2022)	S	N	S	S	S	S	S	S	S	S	S	S	S	N	S	S	S	S	S	S	S	S	S	S	S	S	S	25
Chauhan (2016)	S	N	S	S	N	S	N	S	S	S	S	N	S	S	N	N	S	S	N	S	S	N	S	S	S	S	N	17
Chen (2019)	S	N	S	S	N	S	S	S	S	S	S	N	S	S	S	S	N	N	N	N	S	N	S	S	S	S	N	15
Chen (2018)	S	N	S	S	N	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	25
Cheng (2018)	S	N	S	S	N	S	S	S	S	S	S	N	S	N	S	S	S	N	N	S	S	S	N	S	S	S	S	20
Cheung (2013)	S	N	S	S	N	S	S	N	S	S	S	N	S	S	N	N	N	S	N	N	S	S	S	S	S	S	N	17
Coosta (2021)	S	N	S	S	N	S	S	S	N	S	S	N	S	N	N	N	S	N	N	S	S	N	S	S	S	S	S	16
Dagva (2015)	S	N	S	S	N	S	S	S	S	S	S	N	S	N	N	N	N	N	N	S	S	S	N	S	S	S	N	15
De Boer (2014)	S	N	S	S	N	S	S	S	S	S	S	N	S	N	N	Y	N	S	N	N	S	S	S	S	S	S	N	18
De Greeff (2017)	S	N	S	S	N	S	S	S	N	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	24

Primer autor (año de publicación)	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	Nº recomendaciones cumplimentadas
Dietrichson (2017)	S	N	S	S	S	S	S	S	S	S	S	S	S	S	N	S	S	S	S	S	S	S	S	S	S	S	N	23
Donker (2013)	S	N	S	S	N	S	S	S	S	S	S	N	S	N	N	S	N	S	N	S	S	S	S	S	S	S	N	19
Double (2019)	S	N	S	S	N	S	S	S	S	S	S	N	S	S	N	S	S	S	N	S	S	S	S	S	S	S	N	21
Ergen (2017)	S	N	S	S	N	S	S	S	S	S	S	N	S	S	S	S	S	N	N	S	S	S	S	S	S	S	N	21
Faramarzi (2015)	S	N	S	S	N	S	N	N	S	S	N	N	S	N	N	N	N	S	N	N	S	N	N	S	S	S	S	12
García-Hermoso (2021)	S	N	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	26
Hew (2018)	S	S	S	S	N	S	S	S	S	S	S	N	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	25
Hu (2020)	S	N	S	S	N	S	S	S	S	S	S	N	S	S	S	S	S	S	N	S	S	S	S	S	S	S	S	23
Hu (2018)	S	S	S	S	N	S	S	S	S	S	S	S	S	S	S	S	S	S	N	S	S	S	S	S	S	S	S	26
Jacobse (2011)	S	N	S	S	N	S	S	S	S	S	S	N	S	S	N	S	N	N	N	S	S	N	S	S	S	S	N	18
Kaçar (2021)	S	N	S	S	N	S	N	N	S	S	S	N	S	S	S	N	S	N	N	S	S	S	S	S	N	S	S	18
Kalalán (2017)	S	N	S	S	N	S	S	S	S	S	S	N	S	S	S	S	N	N	N	S	S	N	S	S	S	S	S	19
Karagöli (2019)	S	N	S	S	N	S	N	N	S	S	S	N	S	N	N	N	S	N	N	S	S	S	N	S	S	S	N	14
Karich (2014)	S	N	S	S	N	S	S	S	S	S	S	S	S	N	S	N	N	N	N	S	S	S	N	S	S	S	N	21

Primer autor (año de publicación)	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	Nº recomendaciones cumplimentadas
Kim (2021)	S	N	S	S	N	S	S	S	S	S	S	N	S	S	S	S	S	S	N	S	S	S	S	S	S	S	N	21
King-Sears (2021)	S	N	S	S	N	S	S	S	S	S	S	S	N	N	N	N	S	S	S	S	S	S	S	S	S	S	S	22
Korpershoek (2016)	S	N	S	S	N	S	S	S	S	S	S	N	S	N	N	S	N	N	N	N	N	S	S	S	S	S	S	17
Lag (2019)	S	N	S	S	N	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	N	23
Leung (2015)	N	N	S	S	N	S	S	S	S	S	S	S	S	S	S	S	S	N	N	S	S	N	S	S	S	S	N	20
Linden (2018)	S	N	S	S	N	S	S	S	S	S	S	S	S	S	N	N	S	S	S	S	S	S	N	S	S	S	S	22
May (2021)	S	N	S	S	S	S	S	S	S	S	S	S	N	S	S	S	S	S	S	S	S	S	S	S	S	S	S	25
Moore (2018)	S	N	S	S	S	S	S	S	S	S	S	S	N	S	S	S	S	S	S	S	S	S	S	S	S	S	S	25
Norris (2019)	S	N	S	S	S	S	S	S	S	S	S	S	N	S	S	S	S	N	N	S	S	S	S	S	S	S	S	23
Oh-Young (2015)	S	N	S	S	N	S	S	S	S	S	S	S	S	S	N	N	N	S	S	S	S	S	N	S	S	S	N	20
Orhan (2019)	S	N	S	S	N	S	S	S	S	S	S	N	S	S	S	N	N	N	N	S	S	S	S	S	S	S	N	18
Petersen-Brown (2019)	S	N	S	S	N	S	S	S	S	S	S	S	S	S	S	S	S	N	N	S	S	S	S	S	S	S	S	24
Phelps (2019)	S	N	S	S	N	S	S	N	S	S	S	S	S	S	N	S	N	N	N	S	S	S	S	S	S	N	S	17
Robbins (2009)	S	N	S	S	N	S	S	S	S	S	S	N	S	S	S	N	N	N	N	N	S	S	S	S	S	S	N	18

Primer autor (año de publicación)	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	Nº recomendaciones cumplimentadas
Saw (2021)	S	N	S	S	N	S	S	S	S	N	S	N	S	S	S	S	S	S	N	S	S	S	S	S	S	S	S	22
Sayyah (2017)	S	N	S	S	N	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	N	S	S	24
Semerci (2015)	S	N	S	S	N	S	S	S	S	S	S	N	S	N	N	N	N	N	N	N	N	N	N	N	N	N	N	12
Sneck (2019)	S	S	S	S	N	S	S	S	S	N	S	S	S	S	N	S	S	S	S	S	S	S	S	S	S	S	S	23
Steele (2016)	S	N	S	S	N	S	S	S	S	S	S	S	S	S	S	N	S	S	S	S	S	S	N	S	S	S	S	22
Sugano (2020)	S	N	S	S	N	S	N	N	S	S	N	N	N	N	N	N	N	N	N	N	N	S	N	S	N	S	N	9
Sung (2016)	S	N	S	S	N	S	S	N	S	N	S	N	S	N	S	N	N	N	N	N	N	S	S	N	N	N	S	13
Tan (2017)	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	N	S	S	S	S	S	S	S	S	S	26
Tokac (2019)	S	N	S	S	N	S	S	S	S	N	S	N	S	S	S	S	S	S	S	S	S	S	S	S	N	S	S	21
Warren (2012)	S	N	S	S	N	S	S	S	S	N	S	N	S	S	N	S	S	N	N	N	N	N	S	S	S	S	N	15
Watson (2017)	S	N	S	S	S	S	S	S	S	S	S	S	S	N	S	N	S	S	S	S	S	S	S	S	S	S	S	24
Wilson (2019)	S	N	S	S	S	S	S	S	S	S	S	N	S	S	S	S	S	S	S	S	S	S	S	S	S	S	N	24
Sang (2020)	S	N	S	S	S	S	S	S	S	S	S	N	S	N	N	N	N	N	N	S	S	S	S	S	N	S	S	20
Yoiro (2012)	S	N	S	S	N	S	N	N	S	S	S	N	S	S	N	S	S	S	S	S	S	S	S	S	S	S	N	17

Primer autor (año de publicación)	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	Nº recomendaciones cumplimentadas
Zheng (2016)	S	N	S	S	N	S	S	S	S	N	S	N	S	S	S	N	N	N	S	S	S	N	S	S	N	S	N	17
S: Sí; N: No																												

Recomendaciones PRISMA: TÍTULO: 1. Título; RESUMEN: 2. Resumen estructurado; 3. Justificación; 4. Objetivos; MÉTODO: 5. Protocolo y registro; 6. Criterios de elegibilidad; 7. Fuentes de información; 8. Búsqueda; 9. Selección de los estudios; 10. Proceso de extracción de los datos; 11. Lista de datos; 12. RoB en los estudios individuales; 13. Medidas de resumen; 14. Síntesis de resultados; 15. RoB entre los estudios; 16. Análisis adicionales; RESULTADOS: 17. Selección de estudios; 18. Características de los estudios; 19. RoB en los estudios; 20. Resultados de los estudios individuales; 21. Síntesis de los resultados; 22. RoB entre los estudios; 23. Análisis adicionales; DISCUSIÓN: 24. Resumen de la evidencia; 25. Limitaciones; 26. Conclusiones; FINANCIACIÓN: 27. Financiación.

Fuente: Elaboración propia.

Eficacia de los Programas Psico-socio-educativos aplicados en Justicia Juvenil en España: Revisión Sistemática

Efficacy of Psycho-socio-educational Programs applied in Juvenile Justice in Spain: A Systematic Review

<https://doi.org/10.4438/1988-592X-RE-2025-412-737>

Mery Estefanía Buestán-Játiva

<https://orcid.org/0000-0002-9023-1966>

Universidad de Almería

Elena Ortega-Campos

<https://orcid.org/0000-0003-1994-7109>

Universidad de Almería

Leticia de la Fuente-Sánchez

<https://orcid.org/0000-0002-8320-3078>

Universidad de Almería

Juan García-García

<https://orcid.org/0000-0002-0123-8497>

Universidad de Almería

Resumen

El estudio sobre la intervención en la conducta antisocial penada en la adolescencia adquiere relevancia para la implantación de medidas socioeducativas dentro del sistema de justicia juvenil. Sin embargo, es escasa la información disponible sobre los factores relacionados con la eficacia de las diferentes prácticas y programas de intervención en justicia juvenil. El objetivo de esta revisión sistemática es conocer la eficacia de los programas psico-socio-educativos aplicados a menores en conflicto con la ley en España. La búsqueda de estudios se ha realizado en bases de datos electrónicas y en revistas y recursos online especializados.

Los criterios de inclusión utilizados para la selección de los trabajos han sido: estudios en los que se implante un programa de intervención psico-socio-educativo en el sistema de justicia juvenil español; la muestra debía estar formada por menores que estuvieran cumpliendo una medida socioeducativa bajo la Ley Orgánica 5/2000 durante la realización del programa psico-socio-educativo; y se debía evaluar la eficacia del programa y presentar datos cuantitativos sobre la eficacia del mismo. En este trabajo se incluyen 18 estudios independientes con una muestra total de 6.753 jóvenes, que aportan resultados de áreas como satisfacción, disminución de conductas agresivas, mayor asunción de responsabilidad o mejoras en el ámbito familiar, entre otros aspectos. Del total de estudios, once presentan datos cuantitativos que indican una reducción general de la reiteración de la conducta delictiva, entre un 3,4% y un 28,8%. Esto subraya la importancia de evaluar los programas psico-socio-educativos dirigidos a menores en conflicto con la ley y de recabar información sobre los indicadores de eficacia para la elaboración de programas adecuados en el ámbito de justicia juvenil.

Palabras clave: Eficacia, Jóvenes Infractores, Justicia Juvenil, Programas Psico-socio-educativos, Reiteración Delictiva.

Abstract

The study of intervention in juvenile antisocial behavior acquires relevance for the implementation of socio-educational measures within the juvenile justice system. However, there is little information available on the factors related to the effectiveness of different practices and intervention programs in juvenile justice. The aim of this systematic review is to determine the effectiveness of psycho-socio-educational programs applied to juveniles in conflict with law in Spain. The search for studies was carried out in electronic databases and specialized journals and online resources. The inclusion criteria used for the selection of papers in this systematic review were: studies in which a psycho-socio-educational intervention program was implemented in the Spanish juvenile justice system; the sample had to be made up of juveniles who were completing a psycho-socio-educational measure under Organic Law 5/2000 during the psycho-socio-educational program; and the effectiveness of the program had to be evaluated and quantitative data on its effectiveness had to be presented. This work includes a total of 18 independent studies with a total sample of 6753 juveniles, which provide results in areas such as satisfaction, reduction of aggressive behavior, greater assumption of responsibility or improvements in the family environment, among other aspects. Of the total number of studies, eleven studies present quantitative data indicating a overall reduction in the repetition of criminal behavior, among 3.4% and 28.8%. This underlines the importance of evaluating psycho-socio-educational programs aimed at juveniles in conflict with the law and of gathering information on effectiveness indicators for the development of appropriate programs in the juvenile justice field.

Keywords: Effectiveness, Juvenile Justice, Psycho-socio-educational Programs, Repeated Offenses, Young Offenders.

Introducción

La conducta antisocial penada comprende el conjunto de acciones que transgreden las normas sociales y derechos, constituyendo un quebrantamiento de la ley (De la Peña, 2005; García et al., 2010; Rutter et al., 2000; Villafuerte-Díaz et al., 2022). El estudio de la conducta antisocial penada adolescente es relevante para la planificación e implementación de políticas públicas relativas a la intervención y prevención de la violencia, siendo la perspectiva psicoeducativa la más efectiva (Badasa et al., 2019; García et al., 2012; Organización Mundial de la Salud, 2020; Souverein et al., 2019).

En España, durante el año 2024, fueron condenados por sentencia firme un total de 13.491 menores con edades comprendidas entre 14 y 17 años, lo que supuso un aumento del 3,6% respecto al año anterior. El 79,7% de los menores condenados eran varones. Los delitos de mayor incidencia fueron: lesiones (31,6%), robos (15,8%) y amenazas (9,8%) (Instituto Nacional de Estadística, 2024). La Ley Orgánica 5/2000 regula la responsabilidad penal de los menores, entre 14 y 17 años, que han cometido hechos tipificados como delitos en el Código Penal. La mencionada ley subraya que los menores que han cometido conductas antisociales penadas deben recibir una educación integral y tienen la obligación de participar en actividades formativas, educativas y laborales. Las medidas sancionadoras-educativas que se adoptan están orientadas hacia la efectiva reinserción social del menor desde una óptica psicoeducativa (Barroso-Hurtado y Serrano, 2019; García et al., 2012).

El problema de la violencia en menores no es un fenómeno aislado, por el contrario, está vinculado a otros comportamientos problemáticos y factores de riesgo psicosociales (Moffitt, 2018; Young et al., 2017). Algunos de los factores de riesgo relacionados con la conducta antisocial penada en menores incluyen situaciones sociales desfavorecidas, fracaso escolar, absentismo, iguales disociales, consumo de sustancias, familias disfuncionales o conductas antisociales previas (Aazami et al., 2023; Anjaswarni et al., 2019; Jolliffe et al., 2017; Lahey et al., 2003; Mampaso et al., 2014; Salazar et al., 2011; Van der Put et al., 2011). Los modelos utilizados para la explicación y abordaje del comportamiento antisocial penado adolescente se basan en la identificación de los factores de riesgo y protección que presenta cada menor en conflicto con la ley (Andrews & Bonta, 2010; Ward & Brown, 2004).

El modelo *Risk-Need-Responsivity* (Andrews & Bonta, 2010) hace hincapié en la evaluación del riesgo y las necesidades que presenta el menor con el objetivo de planificar una intervención adaptada a sus necesidades

específicas. Complementariamente, el *Good Lives Model* (Ward & Brown, 2004) subraya la necesidad de desarrollar las capacidades y fortalezas de los menores que realizan conductas antisociales penadas. La tendencia actual está enfocada en intervenciones psicoeducativas dirigidas a trabajar con las necesidades y fortalezas que presenta cada menor, identificadas en las evaluaciones previas a las intervenciones llevadas a cabo en justicia juvenil (Calvo, 2007; Lipsey, 2014; Puig y Capdevila, 2016).

En este contexto, el movimiento *What Works?* surge en el sistema de justicia para adultos en respuesta al informe de Martinson (1974), dónde se presenta una revisión de 231 programas de rehabilitación en prisiones, concluyendo que “nada funciona” en la reducción de la reincidencia. La investigación sobre *What works?* empezó a popularizarse en el sistema de justicia juvenil en la década de 1980 y 1990 (Krisberg, 2005), planteando el análisis de las intervenciones realizadas en justicia de adultos con el objetivo de evaluar la eficacia de los programas aplicados e identificar los elementos o pilares fundamentales para que la intervención sea exitosa (Lab & Whitehead, 1988). Este pensamiento se ha trasladado al ámbito de justicia juvenil con la premisa de evaluar y determinar la eficacia de las intervenciones realizadas, así como, su contribución a la reinserción psicoeducativa del menor (Dowden & Andrews, 1999; Evans-Chase & Zhou, 2014; Pappas & Dent, 2023; Zahn et al., 2009). Este movimiento surge como respuesta a la necesidad de fundamentar las decisiones en el sistema de justicia juvenil en datos y resultados empíricos (Krisberg, 2005; Latessa, 2004). El objetivo del *What Works* es identificar y realizar intervenciones, programas y políticas basadas en la evidencia científica que han demostrado ser efectivas en la reducción del comportamiento antisocial penado (Krisberg, 2005).

El desarrollo del *What Works* está ligado al incremento de la investigación evaluativa y a la búsqueda de respuestas sobre la eficacia y eficiencia de las prácticas y políticas en el sistema de justicia (Elliott et al., 2020; Latessa, 2004; Smith, 2005). Este modelo ha revelado la importancia de la identificación de los factores de riesgo y protección asociados con la reiteración de la conducta antisocial penada, fundamentales para la planificación y creación de programas específicos para menores en conflicto con la ley (Garrido et al., 2006a; Lipsey, 2009; Lipsey et al., 2007; Ortega et al., 2014; Pappas & Dent, 2023; Redondo et al., 2011).

Siguiendo las premisas del modelo *What Works?* se han realizado diversos estudios, entre los que destacan el meta-análisis de Lipsey (2009) dónde se analizó la eficacia de 548 programas aplicados a menores infractores. Los resultados mostraron diferencias en la reiteración de la conducta antisocial penada según el tipo de intervención reci-

bida, siendo menor la reiteración en los programas de orientación y asesoramiento (coeficiente Phi $\phi=0,066$), programas multidisciplinares ($\phi=0,062$), programas de desarrollo de habilidades prosociales y gestión del comportamiento ($\phi=0,060$) y programas de justicia restaurativa ($\phi=0,050$). Posteriormente, Pappas & Dent (2023) evaluaron el efecto de los programas psico-socio-educativos aplicados en justicia juvenil, encontrando una reducción estadísticamente significativa de la reiteración de la conducta antisocial penada ($\phi=0,09$) para aquellos menores que participaron en un programa frente a aquellos menores que no realizaron ningún programa psicoeducativo.

Redondo y Sánchez-Meca (2003) llevaron a cabo un meta-análisis que incluyó 17 programas europeos aplicados a un total de 2775 menores infractores, obtuvieron un tamaño del efecto global de $r=0,18$, lo que supuso una reducción de dieciocho puntos de la reiteración delictiva en los grupos de intervención. Los resultados más relevantes se obtuvieron en programas educativos ($r=0,42$), programas de derivación bajo el sistema judicial ($r=0,29$), programas en comunidades terapéuticas ($r=0,14$) y programas cognitivo-conductuales ($r=0,14$). Un par de años más tarde, Garrido et al. (2006a) analizaron la efectividad de 30 programas aplicados a menores en conflicto con la ley, obteniendo un tamaño de efecto global $d=0,14$, lo que supuso una reducción de siete puntos en la conducta antisocial penada. Las intervenciones de tipo cognitivo-conductual ($d=0,22$) y cognitivo ($d=0,12$) demostraron una mayor eficacia.

Tradicionalmente la intervención se ha centrado en los factores de riesgo que presentan los menores en conflicto con la ley, sin embargo, la investigación actual enfatiza la necesidad de intervenciones basadas en las fortalezas que presenta cada menor, con el objetivo de lograr su reinserción social (Day & Howells, 2002; Looman & Abracen, 2013; Menon & Cheung, 2018; Zelaya, 2020). La inclusión de una perspectiva psicoeducativa, crucial en situaciones de riesgo y exclusión social, adquiere mayor relevancia en el contexto de justicia juvenil, orientando la reeducación de estos menores según sus circunstancias personales, familiares y sociales (Caride, 2005; Dionne y Altamirano, 2012).

Uno de los factores tradicionalmente estudiados como indicador de la eficacia de los programas aplicados en menores en conflicto con la ley ha sido la reiteración de la conducta delictiva (García et al., 2010, 2012; Ortega et al., 2014). En este sentido, la evaluación del riesgo de reiteración delictiva conforma un eje fundamental durante y después del proceso de intervención (Carbonell et al., 2016; Redondo y Martínez, 2013; Schwalbe, 2007). Se han desarrollado instrumentos específicos para medir el riesgo de reiteración de la conducta antisocial penada en

menores, destacando entre los más utilizados el *Youth Level of Service Case Management Inventory* (Hoge y Andrews, 2006), traducido al castellano como *Inventario de Gestión e Intervención con jóvenes (IGI-J)* (Garrido et al., 2006b) y el *Structured Assessment of Violence Risk in Youth (SAVRY)* (Borum et al., 2003), adaptado al castellano por Vallés y Hilterman (2006).

A pesar del desarrollo de programas psico-socio-educativos en el sistema de justicia juvenil español, el conocimiento sobre la eficacia de estos programas es limitado, lo que hace necesaria la investigación sobre qué indicadores pueden ser tomados como referencia de la efectividad de los programas aplicados, con el objetivo de proporcionar información sobre los factores clave para el éxito de las intervenciones (Barroso-Hurtado y Serrano, 2019; Redondo et al., 2011, 2012a). La falta de estudios sobre la eficacia de los programas psicoeducativos en menores cobra especial relevancia de cara a la planificación de intervenciones para prevenir y/o reducir la conducta antisocial penada en menores en conflicto con la ley.

Por ello, el objetivo de este trabajo es realizar una revisión sistemática sobre la eficacia de los programas psico-socio-educativos aplicados a menores en conflicto con la ley en España, prestando especial atención a la evaluación realizada y a los indicadores de eficacia, como la reiteración delictiva.

Método

Para la elaboración de este estudio de revisión sistemática se han seguido las recomendaciones de la declaración PRISMA (Page et al., 2021).

Estrategia de búsqueda

Las fuentes de información usadas en la búsqueda de estudios han sido: a) bases de datos electrónicas (*Web of Science*, *Scopus*, *Science Direct*, *Redalyc*, *PubMed*, *Psycinfo* y *Dialnet*); b) revistas especializadas (*Bienestar y Protección Infantil*, *Boletín Criminológico*, *Revista Española de Investigación Criminológica* y *Anuario de Psicología Jurídica*); c) recursos online especializados (*Observatorio de la Infancia y Adolescencia*, *Observatorio de la Delincuencia*, *Observatorio de Justicia Penal Juvenil* e *International Juvenile Justice Observatory*); d) informes de los organismos o entidades competentes en justicia juvenil en las comunidades autónomas (*Agencia para la Reeducción y Reinserción del Menor Infractor*; *Centro de Estudios Jurídicos y Formación Especializada*; *Fundación Diagrama*, *GINSO* y *ADIS-Meridianos*); y e) revisión de las referencias de los estudios incluidos en este trabajo, así como de otros trabajos de revi-

siones sistemáticas y meta-análisis realizados. La búsqueda de estudios se ha realizado entre los meses de octubre a diciembre de 2023.

Palabras clave

Para llevar a cabo la búsqueda de estudios se emplearon descriptores seleccionados específicamente para esta revisión sistemática. Las palabras clave utilizadas fueron: Jóvenes Infractores, Eficacia, Evaluación, Intervención, Justicia Juvenil y Programas. En inglés se utilizaron las *keywords*: *Young offenders, Effectiveness, Evaluation, Intervention, Juvenile Justice* y *Programs*. Los operadores booleanos usados fueron AND y OR.

Criterios de elegibilidad

Los criterios de inclusión utilizados para la selección de los trabajos en esta revisión sistemática fueron los siguientes: (1) Estudios en los que se lleve a cabo un programa de intervención psico-socio-educativo en el sistema de justicia juvenil español; (2) Estudios en los que la muestra sean menores en conflicto con la ley que estuvieran cumpliendo una medida socioeducativa bajo la Ley Orgánica 5/2000 durante la realización del programa psico-socio-educativo; (3) Los estudios deben evaluar la eficacia del programa; y (4) Deben ofrecer resultados cuantitativos acerca de la eficacia del programa.

Complementariamente, los criterios de exclusión de esta revisión sistemática han sido: (1) Estudios en los que no se ofreciera información sobre la eficacia del programa psicoeducativo llevado a cabo; (2) Estudios en los que los menores estuvieran cumpliendo medida bajo legislaciones anteriores a la Ley Orgánica 5/2000; y (3) Estudios con muestras mixtas, es decir, aquellos en los que la muestra estuviera compuesta por adolescentes con y sin medidas judiciales, así como también, si estuviera compuesta por adolescentes y adultos.

Codificación de estudios

Las variables extraídas de los estudios seleccionados fueron: autores y año de publicación, tipo de publicación (informe, artículo, tesis doctoral, libro o manual), comunidad autónoma en la que se realiza el estudio, tipo de estudio (experimental, cuasi-experimental, prospectivo, retrospectivo), tamaño muestral, edad, sexo (% hombres), nacionalidad (% españoles), régimen de la medida educativa impuesta (abierto, cerrado o mixto), medida judicial impuesta, tipología del delito, tipo de programa de intervención, instrumentos de evaluación, antecedentes de conducta antisocial penada (%) y reiteración de la conducta antisocial penada (%).

Procedimiento de búsqueda

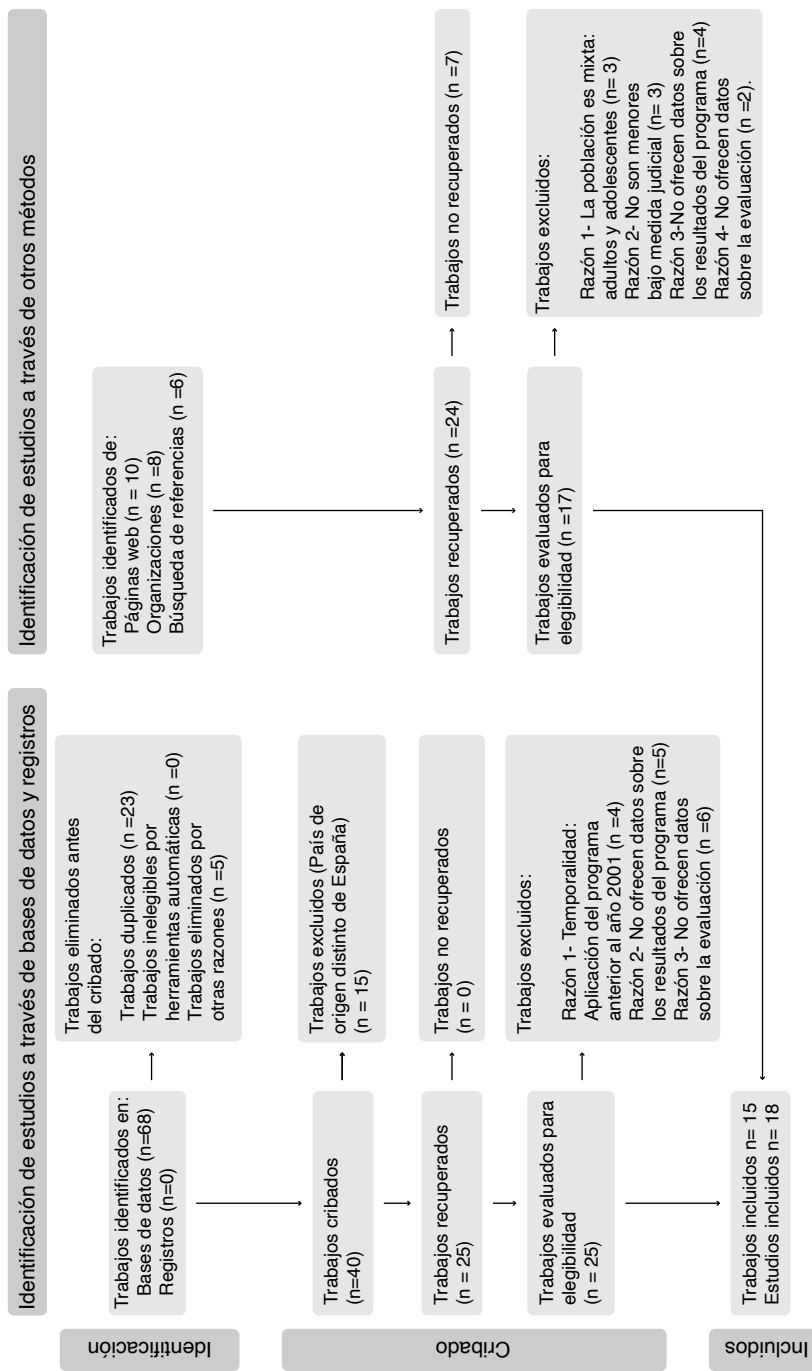
La búsqueda realizada permitió extraer un total de 15 trabajos con 18 estudios independientes sobre programas aplicados a menores bajo medida sancionadora-educativa en justicia juvenil. La búsqueda se inició en las bases de datos electrónicas, obteniendo un total de 68 resultados. Después de eliminar duplicados y aquellos trabajos que no eran estudios empíricos se identificaron 40 trabajos para cribar. De estos, se examinaron 25 trabajos, excluyendo 15 por diversos motivos: aplicación del programa bajo una ley anterior a la ley 5/2000 ($n=4$), no ofrecían resultados sobre los resultados del programa ($n=5$) o no ofrecían datos sobre la evaluación del programa ($n=6$). Consiguiendo diez trabajos independientes.

Por otro lado, se identificaron 24 trabajos a través de consultas a revistas y organismos autonómicos especializados como el Centro de Estudios Jurídicos y Formación Especializada o la Agencia para la Reeducción y Reinserción del Menor Infractor y la revisión de las referencias de los estudios incluidos en este trabajo. Se evaluaron un total de 17 trabajos para decidir sobre su elegibilidad, se excluyeron 12 por los siguientes motivos: uso de muestra mixta (adolescentes y adultos) ($n=3$), los menores no estaban bajo medida judicial ($n=3$), no ofrecían datos sobre los resultados del programa ($n=4$), o no ofrecían datos sobre la evaluación del programa ($n=2$). Se obtuvieron cinco trabajos independientes. El proceso de búsqueda de estudios se indica mediante un diagrama de flujo (Figura 1).

Análisis de Datos

Se ha realizado un análisis descriptivo (cualitativo y cuantitativo) de los resultados más relevantes de las variables incluidas en la revisión sistemática. En la parte cuantitativa, para cada estudio o muestra independiente se ha estimado un tamaño del efecto como medida de la eficacia de los programas psico-socio-educativos aplicados. Se presenta un gráfico forest-plot que incluye los tamaños del efecto estimados para cada muestra independiente y un tamaño del efecto promedio, con el objetivo de realizar un análisis exploratorio de la eficacia de los programas de intervención en justicia juvenil. Se han estimado los tamaños del efecto tomando como referencia las proporciones de la reiteración de la conducta antisocial penada, utilizando el modelo de efectos fijos, que es el más adecuado para el tamaño muestral utilizado y la aproximación meta-analítica descriptiva al fenómeno objeto de estudio. Para el análisis de datos se ha utilizado el software Jamovi 2.5 (The Jamovi Project, 2024).

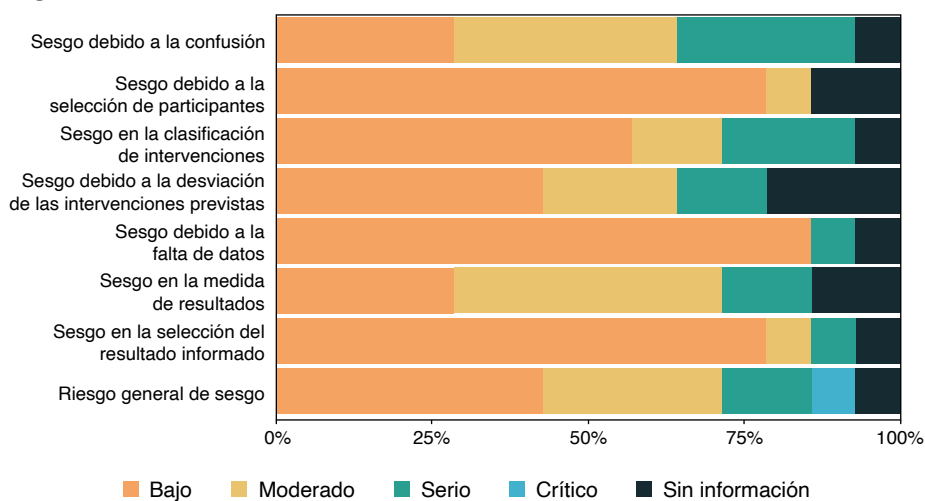
Figura 1. Diagrama de flujo



Evaluación del riesgo de sesgo

El riesgo de sesgo en revisiones sistemáticas se refiere a la posibilidad de errores sistemáticos en los estudios primarios que puedan dar lugar a resultados y conclusiones inexactos o erróneos (Drucker et al., 2016). En este trabajo la evaluación del riesgo de sesgo se realizó a través de la herramienta *ROBINS-I tool (Risk of Bias in Non-randomized Studies-of Interventions)* (Sterne et al., 2016), que se ocupa de evaluar el riesgo de sesgo en los resultados de estudios no aleatorizados de intervenciones que comparan efectos sobre dos o más intervenciones. Se eligió esta herramienta con el fin de tener un criterio del rigor metodológico de los estudios incluidos en esta revisión sistemática. Los resultados obtenidos muestran un riesgo de sesgo entre moderado y bajo (figura II). Se ha utilizado el programa *Robvis* para la creación del gráfico de riesgo de sesgo (McGuinness & Higgins, 2021).

Figura II. Gráfico resumen



Nota. Gráfico de barras de la distribución del riesgo de sesgo en cada área estudiada.

Resultados

El proceso de búsqueda ha permitido identificar 15 trabajos que dan lugar a un total de 18 estudios independientes sobre programas socio-educativos aplicados en el sistema de justicia juvenil español con un tamaño muestral de 6753 adolescentes. Las principales características de

cada estudio, así como los resultados obtenidos se reflejan en las tablas I y II, así como en la figura III.

Características globales de los estudios

En primer lugar, en la tabla I, se presentan de forma global e integrada, las características de los 18 estudios incluidos en la revisión sistemática, respecto a las siguientes variables: tipo, año de publicación, lugar de realización, tamaño muestral, tipo de programa implantado, sexo (% chicos), nacionalidad (% españoles), medida impuesta, existencia de antecedentes del menor y medida de reiteración delictiva.

Respecto al tipo de estudio, nueve de los dieciocho estudios están publicados como informes (Alba et al., 2007; Barnes et al., 2018; Blanch y Mancho, 2023; Camps y Cano, 2006a,b; Capdevila et al., 2017a,b,c; Junta de Extremadura, 2021), cuatro son tesis doctorales (Fernández, 2016; Luna et al., 2015; Pintado, 2012; Sánchez, 2008), cuatro son artículos de investigación (Burcet i Solé et al., 2019; Ocáriz, 2013; Palanques-Alegre et al., 2022; Redondo et al., 2012b) y uno es un manual de tratamiento (González et al., 2013). Los estudios se publicaron entre 2006 y 2023, mayoritariamente entre 2011 y 2020; en concreto, el 28% de los estudios se publicaron entre 2011-2015 y el 33% entre 2016-2020. Cataluña es la comunidad autónoma con más estudios realizados, el 55%, seguida de Madrid y Valencia con un 11% cada una de ellas, mientras que en el resto de CCAA se ha realizado un único estudio. Respecto al tamaño muestral de los estudios, este oscila entre 9 y 2022 menores (Capdevila et al., 2017; Pintado, 2012), presentando la mayoría de los estudios muestras pequeñas. En concreto, el 44% de los estudios tiene menos de 50 participantes, el 22% presenta entre 51-100 participantes y entre 101-1000 y >1000 un 17% de estudios cada uno.

En relación con las características de los adolescentes, el sexo predominante fue el masculino (81,23%), el 28% de los estudios tiene entre un 71-80% de chicos, el 22% tiene entre 81-90% y el 17% tiene más del 91% de chicos. En relación con la nacionalidad de los menores en el conjunto de estudios, el 73% eran españoles; en concreto, el 28% de los estudios tenían menos del 75% de españoles, el 22% tenía entre 76-90% y el 11% de los estudios tiene más del 91%. La media de edad de los menores fue de 16,8 años.

En cuanto a la medida sancionadora-educativa aplicada a los menores, en el 33% de los estudios se aplicó el programa de intervención psicoeducativo estando los menores en medio cerrado o internamiento (Camps y Cano, 2006b; Fernández, 2016; González et al., 2013; Junta de Extremadura, 2021; Palanques-Alegre et al., 2022; Pintado, 2012), en el 61% de los estudios los menores estaban en medio abierto o sin privación de

libertad (Alba et al., 2007; Blanch y Mancho, 2023; Burcet i Solé et al., 2019; Camps y Cano, 2006a; Capdevila et al., 2017a,b,c; Luna et al., 2015; Ocáriz, 2013; Redondo et al., 2012b; Sánchez, 2008) y en un estudio se incluían menores de ambos sistemas (Barnes et al., 2018).

Según el ámbito de intervención, el 28% de los estudios implantan y evalúan programas de Mediación y Reparación (MRM) (Blanch y Mancho, 2023; Capdevila et al., 2017a,b,c; Ocáriz, 2013), el 22% llevan a cabo programas específicos de Violencia Filio-parental y ámbito familiar (Barnes et al., 2018; Burcet i Solé et al., 2019; González et al., 2013; Sánchez, 2008), otro 22% implantan programas de Competencias sociales y pensamiento prosocial (Alba et al., 2007; Luna et al., 2015; Pintado, 2012; Redondo et al., 2012b), el 17% programas de Autocontrol e impulsividad (Camps y Cano, 2006a,b; Palanques-Alegre et al., 2022) y el 11% programas de Consumo de sustancias y adicciones (Fernández, 2016; Junta de Extremadura, 2021).

Atendiendo a las variables criminológicas, el 61% de los estudios indican los antecedentes delictivos de los menores en conflicto con la ley (Blanch y Mancho, 2023; Camps y Cano, 2006a,b; Capdevila et al., 2017a,b,c; Fernández, 2016; González et al., 2013; Ocáriz, 2013; Palanques-Alegre et al., 2022; Redondo et al., 2012b). Igualmente, el 61% de los estudios indican si los menores reiteran en la conducta antisocial penada (Alba et al., 2007; Barnes et al., 2018; Blanch y Mancho, 2023; Burcet i Solé et al., 2019; Camps y Cano, 2006a; Capdevila et al., 2017a,b,c; Fernández, 2016; González et al., 2013; Ocáriz, 2013).

Por otra parte, en relación con el diseño del estudio, sólo dos de los dieciocho estudios utilizaron un diseño cuasi-experimental con grupos de control (Palanques-Alegre et al., 2022; Redondo et al., 2012b), empleando el resto diseños retrospectivos. Respecto a los instrumentos de evaluación utilizados han sido diversos, desde instrumentos específicos de valoración del riesgo de reiteración de la conducta antisocial penada como el IGI-J (Garrido et al., 2006b) o el SAVRY (Vallés y Hilterman, 2006); junto con instrumentos que evalúan otros aspectos relacionados con determinados aspectos tratados en los programas como la Escala de Habilidades Sociales de Gismero (2000), el Índice de Severidad de la Adicción en Adolescentes (T-ASI) (Kaminer et al., 1991), el Historial Criminológico y Social: Versión Juvenil (HCS-J) (Graña y González, 2008), o generales como el MACI (Millon, 1993) o el Inventario de Inteligencia Emocional, versión para jóvenes (EQi-YV) (Baron & Parker, 2000).

Tabla I. Características de los estudios incluidos en la revisión sistemática

Variable	N(%)	Variable	N(%)
Tipo de publicación		Año publicación	
Informe	9(50)	2006-2010	4(22)
Artículo	4(22)	2011-2015	5(28)
Tesis doctoral	4(22)	2016-2020	6(33)
Libro o manual	1(6)	2021-2023	3(17)
Sexo (% chicos)		Nacionalidad (% españoles)	
<70%	4(22)	<60%	2(11)
71-80%	5(28)	61 - 75%	3(17)
81-90%	4(22)	76 – 90%	4(22)
>91%	3(17)	>91%	2(11)
No indicado	2(11)	No indicado	7(39)
Lugar de Realización		Tipo de programa	
Cataluña	10(55)	Mediación y Reparación	5(28)
Madrid	2(11)	Violencia Filio-parental y en el ámbito familiar	4(22)
Valencia	2(11)	Competencias sociales y pensamiento prosocial	4(22)
País Vasco	1(6)	Autocontrol e impulsividad	3(17)
Extremadura	1(6)	Consumo de sustancias y adicciones	2(11)
Asturias	1(6)	Tamaño muestral	
Varias CCAA	1(6)	<50	8(44)
Medida impuesta		51 – 100	4(22)
Abierto	11(61)	101 – 1000	3(17)
Cerrado	6(33)	<1001	3(17)
Mixto	1(6)		
Informan Antecedentes		Informan Reiteración	
SI	11(61)	SI	11(61)
NO	7(39)	NO	7(39)

Nota: Los datos incluidos en la tabla responden a los 18 estudios incluidos en la revisión sistemática.

Eficacia de los programas

La eficacia de los programas ha sido informada de diferentes formas en los estudios revisados. En la tabla II se presentan estos datos segregados para cada estudio. Así, de los once estudios que informan de datos

cuantitativos sobre la reiteración de la conducta antisocial penada de los menores como indicador de eficacia; este dato oscila entre 3,40% y 28,80% de reiteración en el grupo que recibía el programa (Alba et al., 2007; Barnes et al., 2018; Blanch y Mancho, 2023; Burcet i Solé et al., 2019; Camps y Cano, 2006a; Capdevila et al., 2017a,b,c; Fernández, 2016; González et al., 2013; Ocáriz, 2013). Se encontró una disminución del nivel de riesgo del 37,79%, siendo la media pretest 23,10 y la media postest 14,37 (Palanques-Alegre et al., 2022). En el estudio de Barnes et al. (2018) se informa de un 19,5% de reiteración de la conducta antisocial penada, de los cuales, un 7,3% correspondió a situaciones de maltrato familiar.

Por otra parte, cuatro de los estudios han tomado la asistencia al programa como medida de eficacia, presentando datos de asistencia que oscilan entre el 77% y el 90,61%, en ninguno de los estudios se trató de asistencia obligatoria (Alba et al., 2007; Barnes et al., 2018; Burcet i Solé et al., 2019; Camps y Cano, 2006), mientras que otros autores informan sobre el nivel de satisfacción con el programa, como Camps y Cano (2006a,b) que obtuvieron una puntuación de 2.6 y 2.8, en una escala de 1 a 3, o los estudios de Luna et al. (2015) y la Junta de Extremadura (2021), que informan que el 95% de los participantes se encuentran satisfechos con el programa realizado.

En los programas específicos para menores con problemas de violencia filio-parental, el trabajo de Sánchez (2008) muestra que, tras la intervención, el 53% de los menores asume parte de la responsabilidad de lo sucedido, el 34,8% trata a sus padres con respeto, mientras que el 25,8% ha vuelto a insultar esporádicamente, el 18,2% insulto habitual, el 13,6% agresión leve, mientras que el 3% ha roto algo en casa o ha amenazado con arma y el 4,5% ha cometido una agresión grave. El estudio de Barnes et al. (2018) muestra que mejora la relación con los padres en un 52,4% de los casos, mientras que la autopercepción de cambio en los padres llegó a ser del 66,70%.

En los programas de intervención para mejorar el autocontrol e impulsividad de los menores, el 70.5% reflejó un incremento en la formación sobre el control de la conducta violenta (Camps y Cano, 2006a,b). En el estudio de Palanques-Alegre et al. (2022) se redujo la propensión a la impulsividad en un 23,81%, siendo la puntuación media pretest de 80,02 y postest de 60,96, aumentó la puntuación en resolución de conflictos, pasando de 52,18 a 60,89, y de la misma manera ocurrió con el afrontamiento positivo, pasando de 51,75% a 55,53%.

En las intervenciones para mejorar el pensamiento prosocial y las habilidades sociales, Redondo et al. (2012b) observaron un efecto sig-

nificativo en habilidades sociales, agresividad y autoestima. Alba et al. (2007) registraron menor agresividad en el 40% de los casos y el 60% de los menores consiguió verbalizar sus sentimientos. En Pintado (2012) se indica que el 55,5% de los menores que finalizaron el programa reflejaron cambios positivos en el ámbito personal.

En los programas psicoeducativos para el tratamiento de adicciones, el programa PIMICA (Junta de Extremadura, 2021) registró una evolución favorable de las altas. En Fernández (2016) se indicó que en el grupo cuasi-experimental el tamaño del efecto encontrado fue alto ($\eta p^2 = 0,55$) y fue menor la reiteración de la conducta antisocial penada (28,80%) con respecto al grupo que no participó en el programa (37,50%). El IGI-J reflejó una disminución del 10% en el Riesgo Total.

Respecto a la eficacia de los programas de Mediación y Reparación, los resultados reflejan que el 82,6% de las mediaciones finalizadas en el año 2012 fueron positivas, reincidiendo un 5,9%, frente a la tasa del 18% de los que finalizaron negativamente las mediaciones (Ocáriz, 2013). En Capdevila et al. (2017a,b,c) se informa de la reiteración de la conducta antisocial penada relativa a los años 2005, 2008 y 2010, tomando valores del 14,40%, 26,10% y 27,50% respectivamente. Por último, Blanch y Mancho (2023) reportan una reiteración de la conducta antisocial penada del 26,70%.

Tabla II. Características de los estudios y reiteración delictiva

Autores y año	Participantes	Medida Judicial	Tipo de delito	Programa	Evaluación	Antecedentes	Reiteración
Camps y Cano (2006a)	N = 19 Edad = 18,6 Nacionalidad= 63,2	Tareas socioeducativas y libertad vigilada	Lesiones 15,79%; agresión sexual 47,37%; homicidio 26,32%; violencia intrafamiliar 5,26%	Control de la conducta violenta	Formulario impresiones clínicas T-ASI, IGI-J, HCS-J.	68,4%	10%
Camps y Cano (2006b)	N = 10 Edad = 19 Nacionalidad= 80	Internamiento	Lesiones 80%	Autocontrol y Asertividad	Formulario impresiones clínicas T-ASI, IGI-J, HCS-J.	90%	
Alba et al. (2007)	N = 10 (5 GC y 5 GT) Edad = 16,7 Sexo = 100	Tareas socioeducativas y libertad vigilada		PPS-VCJ	Cuestionarios indicadores externos y valoración, IGI-J.		20% GT y 80% GC
Sánchez (2008)	N = 85 (19 GC y 66 GT) Edad = 15,5 Sexo = 67,1	Tareas socioeducativas	Amenaza o insulto 100%; romper cosas 83,5%; amenaza con arma 20%; agresión leve 75,3%; Agresión grave 24,7%	Intervención familiar	Entrevista, cuestionario dinámico, test de inteligencia, cuestionario de personalidad.		
Pintado (2012)	N = 9 Edad = 14-20 Sexo = 88,8	Internamiento		PPS-VCJ	Cuestionario estilos de aprendizaje, cuestionario valoración final, cuestionario indicadores externos, IGI-J.		
Redondo et al. (2012b)	N = 28 (11 GC y 17 GT) Edad = 17,7 GT y 17,4 GC Sexo = 63,6 GT y 70,6 GC	Tareas socio-educativas y libertad condicional	Delito violento GT 76,5%; Delito violento GC 33,3%	PPS	Entrevista, escala de autoestima Rosenberg, IRI, EHS, ATIPV, AGO, BIS-10.	41,2% (GT) 85,7% (GC)	

Autores y año	Participantes	Medida Judicial	Tipo de delito	Programa	Evaluación	Antecedentes	Reiteración
González et al. (2013)	N = 263 Edad = 17 Sexo = 65 Nacionalidad = 82	Internamiento		Programa de tratamiento Educativo y Terapéutico por maltrato familiar ascendente	Fichas evaluación del progreso.	56%	3,40%
Ocaíz (2013)	N = 408 Edad = 17 Sexo = 77,5		Robo 9,55%; amenazas 10,78%; hurto 17,15%; daños 22%; lesiones 28,2%; delitos a materiales 4,4%; delitos a personas 7,8%	MRM	Informes	8,3%	24,20%
Luna et al. (2015)	N = 33 Edad = 16-18 Sexo = 79 Nacionalidad= 57	Libertad vigilada y tareas socio-educativas		Programa de Competencias sociales	Informes		
Fernández (2016)	N = 92 (40 GC y 52 GT) Edad = 17 Sexo = 100 Nacionalidad= 39,1		Delito violento con robo 32,7%; contra la libertad sexual 17,44%; delito violento sin robo 5,45%; salud pública 1,1%; violencia familiar 1,1%	Programa de Tratamiento Educativo y Terapéutico para el Consumo de Drogas	T-ASI, IGI-J	33,7%	28,8% GT 38% GC
Capdevila et al. (2017a)	N = 934 Edad = 15,7 Sexo = 87,2 Nacionalidad= 91,9		Contra las personas 20,1%; contra la propiedad violento 7,8%; contra la propiedad no violento 40,5%	MRM	SAVRY	22,8%	14,40%

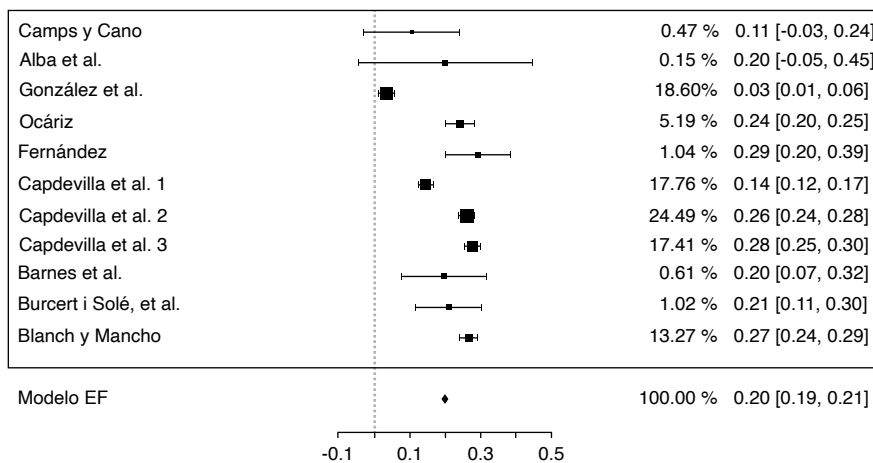
Autores y año	Participantes	Medida Judicial	Tipo de delito	Programa	Evaluación	Antecedentes	Reiteración
Capdevila et al. (2017b)	N = 2022 Edad = 16 Sexo = 80,5 Nacionalidad = 76,4		Contra las personas 36,7%; contra la propiedad violento 8,3%; contra la propiedad no violento 26,9%	MRM	SAVRY	26,7%	26,10%
Capdevila et al. (2017c)	N = 1486 Edad = 15,2 Sexo = 79 Nacionalidad = 71,1		Contra las personas 38,2%; contra la propiedad violento 8,1%; contra la propiedad no violento 27,2%	MRM	SAVRY	27,5%	27,50%
Barnes et al. (2018)	N = 41 Edad = 16 Sexo = 82,9 Nacionalidad = 85,4	Libertad vigilada, tratamiento terapéutico e internamiento	Maltrato familiar 31,7%; maltrato de pareja 2,4%	Programa Miralla Violencia filio-parental	Diario de las sesiones, registros, cuestionario de satisfacción, SAVRY, URICA.		19,50%
Burceti Solé et al. (2019)	N = 72 Edad = 15-18 Sexo = 76			Programa Miralla Violencia filio-parental	Ficha de observación		21,42%
Junta de Extremadura (2021)	N = 36 Edad = 16,5 Sexo = 91,67		Robo 36,11%; quebrantamiento 25%; filio-parental 13,89%; lesiones 13,89%; agresión sexual 5,56%; salud pública 2,78%; violencia de género 2,78%	PIMICA	Registros de actividades, entrevistas, altas.		

Autores y año	Participantes	Medida Judicial	Tipo de delito	Programa	Evaluación	Antecedentes	Reiteración
Palanques-Alegre et al. (2022)	<p>N = 93 (48 GC y 45 GT)</p> <p>Edad = 17,1 GT y 16,8 GC</p> <p>Sexo = 81,73 GT y 85,4 GC</p> <p>Nacionalidad= 91,1 (GT) y 87,9 (GC)</p>	Internamiento terapéutico	<p>GT: 40% maltrato familiar, 17,8% contra la vida, 13,3% robo con violencia, 11,1% libertad sexual, 6,7% robo con fuerza, 6,7% quebrantamiento y 4,4% seguridad colectiva.</p> <p>GC: 25,5% maltrato familiar, 23,2% contra la intimidad, 12,8% libertad sexual, 12,8% robo con fuerza, 12,8% seguridad colectiva, 4,3% robo con violencia, 4,3% contra la libertad y 4,3% contra la vida.</p>	<p>TCC</p> <p>apoyada en RV para el control de impulsos</p>	YLS/CMI, MACI, EQi: YV, ACS.	53,3% GT 68,7% GC	
Blanch y Mancho (2023)	<p>N = 1112</p> <p>Edad = 14-17</p> <p>Sexo = 79,8%</p> <p>Nacionalidad = 65,4%</p>		<p>Delitos contra las personas 40,1%; delitos contra la propiedad violento 14,9%; delitos contra la propiedad no-violento 27,6%; Otros 17,4%</p>	MRM	SAVRY	47%	26,70%

Nota. N: Tamaño muestral; GT: Grupo Tratamiento; GC: Grupo Comparación; Sexo (%): Porcentaje de hombres; Nacionalidad (%): Porcentaje de españoles; ACS: Escala de Afrontamiento para Adolescentes; AGQ: *Aggression Questionnaire*; ATIPV: *Attitude Toward Interpersonal Peer Violence Scale*; BIS-10: Escala de impulsividad de Barrat; IRI: Índice de Reactividad Interpersonal; EHS: Escala de Habilidades Sociales; URICA: *University Rhode Island Change Assessment Scale*; PPS_VCI: Programa de Pensamiento Prosocial Versión Corta para Jóvenes; PIMICA: Programa de Intervención con Menores Infractores con Conductas Adictivas; TCC: Terapia Cognitivo-Conductual; RV: Realidad Virtual; MRM: Programa de Mediación y Reparación de Menores

Para una mejor visualización y comprensión de la eficacia encontrada en los programas psicoeducativos aplicados en el sistema de justicia juvenil español, se presenta, de manera descriptiva, un gráfico *forest plot* con la representación de los diferentes tamaños del efecto y su efecto promedio, estimados a través de un modelo de efectos fijos. Para la realización de este gráfico se han tomado como tamaños del efecto las proporciones de reiteración de la conducta antisocial penada en el grupo que ha participado en los programas psicoeducativos. Se ha encontrado un tamaño del efecto global de 0,20 ($p < 0.001$) cuyo intervalo de confianza al 95% toma valores entre 0,19 y 0,21 (figura III).

Figura III: Forest Plot



Nota. Modelo EF: Modelo de efectos fijos

Discusión

El propósito de este trabajo ha sido analizar la información referente a la eficacia de los programas psicoeducativos implementados en el ámbito de justicia juvenil en España. Se han incluido un total de 15 trabajos con 18 estudios independientes que aportan datos, tanto cualitativos como cuantitativos, sobre su eficacia. Las tesis doctorales (Alba et al., 2007; Fernández, 2016; Pintado, 2012; Sánchez, 2008) y los estudios publicado por el CEJFE destacan por su mayor elaboración y respaldo empírico, la mitad de los programas han sido llevados a cabo en centros de menores en Cataluña, en gran parte debido a la existencia del CEJFE, agencia dedicada a la investigación e intervención en Justicia (Capdevila et al., 2017).

El elemento central de esta revisión es la evaluación de la eficacia de los programas aplicados a menores en conflicto con la ley, cuestión estrechamente relacionada con el movimiento *What Works?* (Pappas & Dent, 2023). A este respecto, la reiteración de la conducta delictiva es uno de los indicadores más utilizados en la evaluación de la eficacia de los programas de intervención con menores infractores. La no entrada del menor en el sistema de justicia juvenil se considera un indicador de éxito del proceso de reinserción del menor (Carbonell et al., 2016; Horcajo-Gil et al., 2019; Ortega et al., 2014; Redondo et al., 2011). Sin embargo, no todos los estudios revisados informan de este dato; sólo el 61%, reflejando una reducción global respecto al grupo de comparación o medida pretest moderada, aunque relevante. Los estudios que no presentan información sobre la reiteración de la conducta delictiva presentan otros indicadores de resultado como porcentaje de participantes que finalizaron el programa, asistencia, y satisfacción e interés en el mismo. Aunque estas medidas no son, desde nuestro punto de vista, el indicador de eficacia más directo, representan factores relacionados con la eficacia del programa y, por tanto, son aspectos para tener en cuenta en su evaluación (Redondo et al., 2011).

En el estudio de la eficacia de los programas psicoeducativos realizados en justicia juvenil, es esencial considerar la metodología de los estudios y el uso de instrumentos de evaluación (Pappas & Dent, 2023; Redondo et al., 2011). Como se ha visto, algunos de los instrumentos más frecuentes a lo largo de los diferentes programas fueron el IGI-J y el SAVRY, ampliamente utilizados en la evaluación de factores de riesgo y protección en justicia juvenil (Redondo et al., 2011). La evaluación minuciosa de las necesidades de los menores en conflicto con la ley proporciona información relevante en el marco de los factores de riesgo y protección del menor, lo cual respalda el desarrollo de intervenciones adaptadas a sus necesidades específicas (Andrews & Bonta, 2010; Ward & Brown, 2004).

La investigación centrada exclusivamente en los factores de protección todavía está en desarrollo (Lipsey, 2014; Pyle et al., 2019). De todos los estudios incluidos, solo han tenido en cuenta los factores protectores durante la intervención Barnes et al. (2018) y Burcet i Solé et al. (2019), con la aplicación paralela de un programa a los padres de los menores participantes, y la Junta de Extremadura (2021) con la inclusión de actividades de ocio saludables. Los estudios realizados por Alba et al. (2007), Pintado (2012) y Redondo et al. (2012b) se centraron en trabajar y fomentar las habilidades sociales, claves en la reinserción social del menor (Lipsey, 2009; Redondo et al., 2011). A este respecto, se destacan

los programas Mirall, PPS-VCJ y MRM, ya que han sido evaluados en múltiples ediciones y sobre los cuales se continúa trabajando para su mejora (Alba et al., 2007; Barnes et al., 2018; Blanch y Mancho, 2023; Burcet i Solé et al., 2019; Capdevila et al., 2017a,b,c; Ocáriz, 2013; Pintado, 2012).

Conclusiones

En relación con las limitaciones presentadas en este estudio, es necesario mencionar las dificultades inherentes a las revisiones, como el sesgo de publicación, que ha impedido poder obtener datos suficientes para su cálculo; la heterogeneidad de documentos encontrados, puesto que se han incluido tesis doctorales, informes, manuales y artículos de investigación. El sesgo de selección es otra consideración importante, ya que los trabajos se obtuvieron de diversas fuentes, como bases de datos electrónicas, entidades u organismos especializados y referencias de trabajos previos. Otra limitación ha sido la dificultad para encontrar artículos de investigación sobre la evaluación de eficacia de programas en España, lo que nos indica que existe poca cultura de evaluación de eficacia de las intervenciones en los ámbitos aplicados de las políticas públicas. A pesar de las limitaciones, este trabajo de revisión sistemática proporciona conclusiones relevantes sobre la eficacia de los programas de intervención psicoeducativos aplicados con menores en conflicto con la ley, con implicaciones prácticas útiles en la elaboración de futuros programas de intervención en justicia juvenil.

Los resultados obtenidos en los diferentes programas respaldan la idea de que la intervención basada en la reinserción social del menor y con un enfoque psicoeducativo funciona, es decir, reduce la reincidencia del comportamiento antisocial penado, como se ha evidenciado en estudios previos (Barroso-Hurtado y Serrano, 2019; Redondo et al., 2012a). Se resalta la evaluación detallada del menor con el fin de identificar factores de riesgo y protección, lo que permite abordar sus necesidades específicas mediante la adaptación de programas (Lipsey, 2014; Puig y Capdevila, 2016; Wilson y Lipsey, 2024). En esta línea, resulta imprescindible incorporar metodología y evaluación del programa, indicando las fases método e instrumentos empleados (Redondo et al., 2011). En concreto, se destaca el uso de medidas pre-post, la comparación de grupos y evaluaciones de seguimiento post-intervención para una valoración completa de la eficacia del programa. Sin embargo, la eficacia de un programa no depende solo de la metodología, sino también de la implementación por lo que es importante la capacitación del personal y supervisión continuada. Por otro lado, la inclusión de factores protecto-

res surge como un pilar fundamental en la intervención, destacando la inclusión de la familia, la escuela y la comunidad, ya que intervienen de forma integral con el menor y no solo con su comportamiento (Aazami et al., 2023; Pappas & Dent, 2023).

Estas conclusiones tienen implicaciones significativas tanto para la investigación como para la práctica en justicia juvenil. Se sugiere explorar nuevas líneas de investigación, como el estudio de indicadores de eficacia en la implementación de programas en justicia juvenil, la evaluación e intervención en factores de riesgo y protección en el comportamiento antisocial y el análisis de perfiles de menores en conflicto con la ley. La difusión de los resultados obtenidos después de la aplicación de los programas es crucial para conocer los factores de éxito y las limitaciones en la intervención con menores en conflicto con la ley (Barroso-Hurtado y Serrano, 2019). Los investigadores deben estar dispuestos a realizar acciones dirigidas a modificar, mejorar y desarrollar programas adaptados a las necesidades cambiantes y específicas de la población de menores en el circuito de justicia juvenil. Asimismo, es esencial profundizar en el estudio de indicadores de eficacia, como la reiteración de la conducta antisocial penada, u otros, y seguir trabajando en la creación de programas psicoeducativos con protocolos de evaluación de la eficacia, con el objetivo de proporcionar herramientas y recursos necesarios para los profesionales que trabajan en justicia juvenil.

Referencias bibliográficas

- Aazami, A., Valek, R., Ponce, A.N., & Zare, H. (2023). Risk and Protective Factors and Interventions for Reducing Juvenile Delinquency: A Systematic Review. *Social Sciences*, 12(9), 474. <https://doi.org/10.3390/socsci12090474>
- Andrews, D.A., & Bonta, J. (2010). *The psychology of criminal conduct*. Routledge.
- Anjaswarni, T., Nursalam, N., Widati, S., & Yusuf, A. (2019). Analysis of the Risk Factors Related to the Occurrence of Juvenile Delinquency Behavior. *Jurnal Ners*, 14(2), 129-136. <http://doi.org/10.20473/jn.v14i2.12465>
- Badasa, G., Gameda, A., Gaduda, B. & Wondimu, B. (2019) Juvenile Delinquency: A Need to Multiple Explanations and Interventions. *Open Access Library Journal*, 6, 1-10. <http://dx.doi.org/10.4236/oalib.1105904>.
- Baron, R., & Parker, J. (2000). *EQI: YV Baron Emotional Quotient Inventory: Youth version. Technical manual*. Multi-Health Systems Inc.
- Barroso-Hurtado, D., y Serrano, J.B. (2019). Revisión de los factores de éxito en la promoción de comportamientos prosociales como estrate-

- gia preventiva en la justicia juvenil en España. *Revista Complutense de Educación*, 30(1), 75-89. <http://dx.doi.org/10.5209/RCED.55509>
- Borum, R., Bartel, P., & Forth, A. (2003). *Structured Assessment of Violence Risk in Youth. Professional Manual*. Pearson.
- Calvo, S. G. (2007). Evolución de la intervención con menores infractores. *En la calle: revista sobre situaciones de riesgo social*, 8, 4-6.
- Caride, A. (2005). Educación social y valores cívicos. *Revista de ciencias de la educación*, 201, 25-43.
- Carbonell, A., Gil-Salmerón, A., y Margaix, E. (2016). Evaluación del riesgo de reincidencia en menores infractores: Herramientas para la mejora de estrategias reeducativas en España. *Revista Internacional de Trabajo Social y Bienestar*, 5, 79-88.
- Day, A., & Howells, K. (2002). Psychological treatments for rehabilitating offenders: Evidence based practice comes of age, *Australian Psychologist*, 37, 39-47.
- De la Peña, M.E. (2005). *Conducta antisocial en adolescentes: Factores de riesgo y protección*. [Tesis, Universidad Complutense de Madrid].
- Dowden, C., & Andrews, D. A. (1999). What works for female offenders: a meta-analytic review. *Crime and Delinquency*, 45, 438-452.
- Dionne, J. y Altamirano, C. (2012). Los desafíos de un verdadero sistema de justicia juvenil: una visión psicoeducativa. *Universitas Psychologica*, 11(4), 1055-1064.
- Drucker, A. M., Fleming, P., & Chan, A.-W. (2016). Research Techniques Made Simple: Assessing Risk of Bias in Systematic Reviews. *Journal of Investigative Dermatology*, 136(11), e109-e114. <https://doi.org/10.1016/j.jid.2016>.
- Elliott, D. S., Buckley, P. R., Gottfredson, D. C., Hawkins, J. D., & Tolan, P. H. (2020). Evidence-based juvenile justice programs and practices: A critical review. *Criminology & Public Policy*, 19(4), 1305-1328. <https://doi.org/10.1111/1745-9133.12520>
- Evans-Chase, M., & Zhou, H. (2014). A Systematic Review of the Juvenile Justice Intervention Literature: What It Can (and Cannot) Tell Us About What Works with Delinquent Youth. *Crime & Delinquency*, 60(3), 451-470. <https://doi.org/10.1177/0011128712466931>
- García, J., Ortega, E., & de la Fuente, L. (2010). Juvenile offenders' recidivism in Spain: A quantitative revision. In M. Frías-Armenta & V. Corral-Verdugo (Eds.), *Bio-psycho-social perspectives on interpersonal violence* (pp. 333-353). Nova Science Publishers.
- García, J., Zaldívar, F., De la Fuente, L., Ortega, E., y Sáinz-Cantero, B. (2012). El Sistema de Justicia Juvenil de Andalucía: descripción

- y presentación de resultados a través de la investigación empírica. *Edupsykhé*, 11(2), 287-316.
- Garrido, V., Anyela, L., & Sánchez-Meca, J. (2006a). What works for serious juvenile offenders? A systematic review. *Psicothema*, 18(3), 611-619.
- Garrido, V., López, E., Silva, T., López, M., y Molina, P. (2006b). *El Modelo de la competencia social de la ley de menores: Cómo predecir y evaluar para la intervención educativa*. Tirant lo Blanch.
- Gismero, E. (2000). *Manual de la Escala de Habilidades Sociales*. TEA.
- Graña, J.L., y González, L. (2008). *Historial Criminológico y Social-Verisión Juvenil: HCS-J*. Agencia de la Comunidad de Madrid para la Reeducación y Reinserción del Menor Infractor.
- Hoge, R., & Andrews, D. (2006). *Youth Level of Service/Case Management Inventory (YLS/CMI). User's manual*. Multi-Health Systems
- Horcajo-Gil, P.J., Dujo-López, V., Andreu-Rodríguez, J.M., y Marín-Rullán, M. (2019). Valoración y gestión del riesgo de reincidencia delictiva en menores infractores: una revisión de instrumentos. *Anuario de Psicología Jurídica*, 29, 41-53. <https://doi.org/10.5093/apj2018a15>
- Instituto Nacional de Estadística (2024). *Estadísticas de Menores Condenados*. Instituto Nacional de Estadística.
- Jolliffe, D., Farrington, D.P., Pirquero, A.R., Loeber, F., & Hill, K.G. (2017). Systematic review of early risk factors for life-course-persistent, adolescence-limited, and late-onset offenders in prospective longitudinal studies. *Aggression and Violent Behavior*, 33, 15-23. <http://dx.doi.org/10.1016/j.avb.2017.01.009>
- Kaminer, Y., Bukstein, O., & Tarte, R.E. (1991). The Teen-Addiction Severity Index: rationale and reliability. *International Journal on Addiction*, 26(2), 219-26. <https://doi.org/10.3109/10826089109053184>
- Krisberg, B. (2005). Lo que funciona en la justicia juvenil. En *Lo que funciona en la justicia juvenil* (pp. 145-162). Publicaciones SAGE, Inc., <https://doi.org/10.4135/9781483328836.n8>
- Lab, S.P., & Whitehead, J.T. (1988). An Analysis of Juvenile Correctional Treatment. *Crime & Delinquency*, 34(1), 60-83. <https://doi.org/10.1177/0011128788034001004>
- Lahey, B., Moffitt, T.E. & Caspi, A. (2003). *The causes of conduct disorder and serious juvenile delinquency*. Guilford.
- Latessa, E.J. (2004). From theory to practice, what works in reducing recidivism? In *State of Crime and Justice in Ohio*. Ohio Office of Criminal Justice Services. https://www.uc.edu/content/dam/uc/ccjr/docs/articles/Theory_Practice.pdf

- Ley Orgánica 5/2000 de 12 de enero, reguladora de la responsabilidad penal de los menores. *Boletín Oficial del Estado*, 11, de 13 de enero del 2000.
- Lipsey, M.W. (2009). The Primary Factors that Characterize Effective Interventions with Juvenile Offenders: A Meta-Analytic Overview. *Victims & Offenders*, 4(2), 124–147. <https://doi.org/10.1080/15564880802612573>
- Lipsey, M.W. (2014). Interventions for Juvenile Offenders: A Serendipitous Journey. *Criminology Public Policy*, 13(1), 1-14.
- Lipsey, M.W., Landenberger, N.A., & Wilson, S.J. (2007). Effects of cognitive-behavioral programs for criminal offenders. *Campbell Systematic Reviews*, 3(1), 1-27.
- Looman, J., & Abracen, J. (2013). The risk need responsivity model of offender rehabilitation: Is there really a need for a paradigm shift? *International Journal of Behavioral Consultation and Therapy*, 8(3-4), 30-36. <http://dx.doi.org/10.1037/h0100980>
- Mampaso, J., Pérez-Fernández, F., Corbí, B., González, M., y Bernabé, B. (2014). Factores de riesgo y de protección en menores infractores. *Análisis y prospectiva. Psychologia Latina*, 5(1), 11-20.
- Martinson, R. (1974). ¿Qué funciona? Preguntas y respuestas sobre la reforma penitenciaria. *The Public Interest*, 10, 22-54.
- Menon, S.E., & Cheung, M. (2018). Desistance-Focused Treatment and Asset-Based Programming for Juvenile Offender Reintegration: A Review of Research Evidence. *Child and Adolescent Social Work Journal*, 35(5), 459-476. <https://doi.org/10.1007/s10560-018-0542-8>
- Millon, T. (1993). *Manual of Millon Adolescent Clinical Inventory*. NCS.
- Moffitt, T.E. (2018) Male antisocial behaviour in adolescence and beyond. *Nature Human Behaviour*, 2, 177–186. <https://pubmed.ncbi.nlm.nih.gov/30271880/>
- McGuinness, L.A., & Higgins, J.P.T. (2021) Risk-of-bias VISualization (robvis): An R package and Shiny web app for visualizing risk-of-bias assessments. *Research Synthesis Methods*, 12(1), 55-61. <https://doi.org/10.1002/jrsm.1411>
- Organización Mundial de la Salud (2020). *Informe sobre la situación mundial de la prevención de la violencia contra los niños*. Organización Mundial de la Salud.
- Ortega, E., García, J., y Frías, M. (2014). Meta-análisis de la reincidencia criminal en menores: estudio de la investigación española. *Revista Mexicana de Psicología*, 31(2), 111-123.
- Page, M.J., McKenzie, J.E., Bossuyt, P.M., Boutron, I., Hoffmann, T.C., Mulrow, C.D., et al. (2021). The PRISMA 2020 statement: an updated

- guideline for reporting systematic reviews. *BMJ*, 372, n71. <http://dx.doi.org/10.1136/bmj.n71>
- Pappas, L.N., & Dent, A.L. (2023). The 40-year debate: a meta-review on what works for juvenile offenders. *Journal of Experimental Criminology*, 19, 1-30. <https://doi.org/10.1007/s11292-021-09472-z>
- Puig, M.F., y Capdevila, M.C. (2016). Intervenciones con jóvenes infractores en el marco de una medida judicial. *Infancia, juventud y ley: Revista de divulgación científica del trabajo con menores*, 7, 61-69.
- Pyle, N., Flower, A., Williams, J., & Fall, A.M. (2019). Social Risk Factors of institutionalized Juvenile Offenders: A Systematic Review. *Adolescente Research Review*, 5, 173-186. <https://doi.org/10.1007/s40894-019-00120-2>.
- Redondo, S., y Martínez, A. (2013). Evaluación criminológica de la justicia juvenil en España. *Cuadernos de Política Criminal*, 2(110), 189-220.
- Redondo, S., Martínez, A., y Andrés, A. (2011). *Factores de éxito asociados a los programas de intervención con menores infractores*. Ministerio de Sanidad, Política Social e Igualdad.
- Redondo, S., Martínez, A., y Andrés-Pueyo, A. (2012a). Intervenciones con delincuentes juveniles en el marco de justicia: investigación y aplicaciones. *EduPsykbé*, 11(2), 143-169.
- Redondo, S., y Sánchez-Meca, J. (2003). Guía de tratamientos psicológicos eficaces para la delincuencia juvenil. En Pérez, M., Fernández Hermida, J.R., Fernández Rodríguez, C., y Amigo Vázquez, I. *Guía de tratamientos psicológicos eficaces III. Infancia y adolescencia* (págs. 183-214). Pirámide.
- Rutter, M., Giller, H., & Hagell, A. (2000). *La conducta antisocial de los jóvenes*. Cambridge University Press.
- Salazar J.G., Torres-López, T.M., Reynaldos, C., Figueroa, N.S., y Araiza, A. (2011). Factores asociados a la delincuencia en adolescentes de Guadalajara, Jalisco. *Papeles de población*, 17(68), 103-126.
- Souverein, F., Dekkers, T., Bulanovaitė, E., Doreleijers, T., Hales, H., Kalliala-Heino, R., & Van Der Pol, T. (2019). Overview of European forensic youth care: Towards an integrative mission for prevention and intervention strategies for juvenile offenders. *Child and adolescent psychiatry and mental health*, 13, 1-6.
- Schwalbe, C.S. (2007). Risk assessment for juvenile justice: A meta-analysis. *Law and Human Behavior*, 31, 449-462.
- Smith, D. J. (2005). The effectiveness of the juvenile justice system. *Criminal Justice*, 5(2), 181-195. <https://doi.org/10.1177/1466802505053497>
- Sterne, J.A., Hernán, M.A., Reeves, B.C., Savović, J., Berkman, N.D., Viswanathan, M., et al., (2016). ROBINS-I: A tool for assessing risk of bias in

- non-randomised studies of interventions. *BMJ*, 355, i4919. <https://doi.org/10.1136/bmj.i4919>
- The Jamovi Project (2024). Jamovi (Version 2.5) [Computer Software]. Retrieved from <https://www.jamovi.org>
- Vallés, M.D., y Hilterman, E.L.B. (2006). SAVRY: *Manual per a la valoració estructurada de risc de violència en joves*. Centro de Estudios Jurídicos y Formación Especializado.
- Van Der Put, C.E., Deković, M., Stams, G.J.J., Van Der Laan, P.H., Hoeve, M., & Van Amelsfort, L. (2011). Changes in risk factors during adolescence: Implications for risk assessment. *Criminal Justice and Behavior*, 38(3), 248-262.
- Villafuerte-Díaz, A., Ramos, P., Rivera, F., y Moreno, C. (2022). Conducta antisocial en adolescentes españoles prevalencia y relación con su salud global percibida. *Psicología Conductual*, 30(2), 641-661. <https://doi.org/10.51668/bp.8322303s>
- Ward, T., & Brown, M. (2004). The good lives model and conceptual issues in offender rehabilitation. *Psychology, Crime & Law*, 10(3), 243-257, <http://dx.doi.org/10.1080/10683160410001662744>
- Wilson, D.B., & Lipsey, M.W. (2024). Scaling up effective juvenile delinquency programs by focusing on change levers: Evidence from a large meta-analysis. *Criminology & Public Policy*, 23(2), 261-286. <https://doi.org/10.1111/1745-9133.12663>
- Young, S., Greer, B. & Church, R. (2017). Juvenile delinquency, welfare, justice and therapeutic interventions: a global perspective. *National Library of Medicine*, 41 (1), 21-29. <https://doi.org/10.1192%2Fpb.bp.115.052274>
- Zahn, M.A., Day, J.C., Mihalic, S.F., & Tichavsky, L. (2009). Determining What Works for Girls in the Juvenile Justice System: A Summary of Evaluation Evidence. *Crime & Delinquency*, 55(2), 266–293. <https://doi.org/10.1177/0011128708330649>
- Zelaya, A.M. (2020). *A Review of Common Treatments and Their Effectiveness to Reduce Juvenile Recidivism*. [Tesis, The Chicago School of Professional Psychology].

Sembrando el proyecto vital desde la infancia: Aprendizaje Sociocultural en *Eskola Txikiak*¹

Sowing the Seeds of a Lifelong Journey that Begins in Childhood: Sociocultural Learning in *Eskola Txikiak*

<https://doi.org/10.4438/1988-592X-RE-2025-412-738>

Garazi Ormazabal-Arizkorreta

<https://orcid.org/0000-0002-2554-9484>

Universidad del País Vasco

Andoni Arguiñano Madrazo

<https://orcid.org/0000-0002-9585-8303>

Universidad del País Vasco

Inaki Karrera Xuarros

<https://orcid.org/0000-0003-4448-7749>

Universidad del País Vasco

Resumen

Dentro del mundo plural, cambiante y complejo en el que nos encontramos, la escuela rural, desarrollando claves pedagógicas de referencia, se ha convertido en un contexto de gran interés por su capacidad de innovación y mejora. En esta línea, el presente estudio nace con la intención de contribuir al conocimiento de buenas prácticas en un contexto particular como es el de la educación rural. Situándonos en cuatro escuelas de la red *Eskola Txikiak* (concepto parecido a las escuelas rurales) del País Vasco, la presente investigación pretende analizar la actividad pedagógica principal de las *Eskola Txikiak*, destacando los beneficios que ésta aporta tanto al alumnado como al profesorado. La recogida de información se ha completado mediante, 420 horas de sesiones de observación,

¹ **DETALLES DE LA FINANCIACIÓN.** Este trabajo ha sido financiado por la [UPV/EHU] con la subvención [código: PIF 21/217]; el [Gobierno Vasco] con la subvención [código: PIBA_2023_1_0025]; y el [Gobierno Vasco, Departamento de Educación] con la subvención [código: hezkuntza_23/12].

8 Focus Groups y el análisis de la documentación interna. Después de categorizar y codificar la información, los resultados obtenidos muestran varias de las claves que subyacen en los procesos de aprendizaje sociocultural que surgen en los talleres pedagógicos. En concreto, los resultados obtenidos en este estudio indican la participación activa en el proceso de aprendizaje, así como la distribución creativa del espacio escolar, la experimentación y la disponibilidad de una amplia gama de recursos materiales y personales con los que los niños y niñas pueden interactuar, promueven un aprendizaje centrado en el estudiante, creativo, personalizado y de alta calidad.

Palabras clave: alumnado, empoderamiento, bienestar, escuelas rurales, talleres.

Abstract

In the pluralistic, ever-changing, and complex world we inhabit, rural schools, through the development of foundational pedagogical approaches, have become a setting of great interest due to their capacity for innovation and improvement. In this context, the present study aims to contribute to the understanding of good practices within a particular context—rural education. Focusing on four schools within the *Eskola Txikiak* network (a concept similar to rural schools) in the Basque Country, this research seeks to analyse the core pedagogical activity of *Eskola Txikiak*, highlighting the benefits it brings to both students and teachers. Data collection was completed through 420 hours of observation sessions, 8 focus groups, and an analysis of internal documentation. After categorizing and coding the information, the results obtained reveal several of the key factors underlying sociocultural learning processes that arise in educational workshops. Specifically, the results obtained in this study indicate that active participation in the learning process, as well as the creative distribution of school space, experimentation, and the availability of a wide range of material and personal resources with which children can interact, promote child-centered, creative, personalized, and high-quality learning.

Keywords: students, empowerment, wellbeing, rural schools, workshops.

Introducción

Como señalaban Bauman (2003) y Díaz (2012), estamos inmersos en un mundo plural, cambiante y complejo. Hace algunas décadas, se utilizaba el término VICA para describir la sociedad del momento, que aludía a lo Volátil, Incierto, Complejo y Ambiguo (Urbano-Carazo, 2022). No obstante, actualmente, este concepto no refleja adecuadamente el nuevo paradigma. Por esta razón, para describir tanto la sociedad actual como los acontecimientos futuros, Cascio (2023) propuso el término BANI, haciendo alusión a entornos que son Frágiles, Ansiosos, No lineales e

Incomprensibles. Siendo este nuestro contexto global, cualquier propuesta innovadora en el ámbito educativo debe tener en cuenta estas características para evitar quedarse al margen de la vida cotidiana de los ciudadanos (Martínez y Rogero, 2021). Tal y como señala, Vergara (2016, p.6) “necesitamos ciudadanos capaces de entender la complejidad de situaciones y el incremento exponencial de la información, así como de adaptarse creativamente a la velocidad del cambio y a la incertidumbre que lo acompaña”. Así pues, ante este nuevo marco, es necesario desarrollar un pensamiento creativo que ayude en la toma de decisiones y aborde el planteamiento y solución de problemas (Pérez, 2019). Para que, de tal manera, seamos capaces de reinventarnos, actualizarnos y readaptarnos a las actuales y futuras características y/o necesidades (Mena, 2018).

Respecto al pensamiento creativo, es deber de la escuela fomentar la curiosidad investigadora desde la infancia (Comisión Europea, 2016; Freinet, 1979; Tonucci, 2012), puesto que “la curiosidad es una forma de motivación intrínseca clave para fomentar el aprendizaje activo y la exploración espontánea” (Oudeyer et al., 2016, p.1). La curiosidad es el inicio del conocimiento ya que nos impulsa a observar, examinar, preguntar, descubrir, probar e interactuar con el mundo (Ortiz y Cervantes, 2015). Por este motivo, resulta necesario reconocer y respetar la disposición innata del niño y la niña hacia la investigación, incorporando procesos de aprendizaje científico en el entorno escolar (Tonucci, 2001; 2012). El desarrollo de la competencia investigadora permite a los niños y las niñas potenciar aptitudes como la creatividad, la indagación, la innovación y la capacidad de crear nuevos conocimientos (Oquendo, 2019). Factores indispensables para responder de manera eficaz a los retos de la era actual (Unesco, 2015).

Pero ¿De qué manera podríamos promover esta disposición científica? Los niños y las niñas desde la infancia van construyendo teorías sobre lo que ocurre a su alrededor de un modo parecido al que lo hacen los científicos; a través de la experimentación y del ensayo y error (Tonucci, 1995). En este sentido, es necesario transmitirles la confianza de modo que tomen conciencia de sus conocimientos actuales y del potencial que poseen para adquirir nuevos saberes (Tonucci, 1995; 2001). Esto implica, reconocer a los niños y niñas como seres con derechos y con un gran potencial de desarrollo (Decroly, 2006; Freinet, 1979; Hall et al., 2014; Malaguzzi, 1996; Montessori, 1982; Tonucci, 1995, 2001, 2012, 2017).

En la misma línea, si pretendemos mantener viva esa actitud investigadora, cobra gran importancia la opinión del niño y de la niña; ya que como se señala en la Convención sobre los Derechos del Niño, los niños y las niñas tienen derecho a opinar, expresar su punto de vista, tomar

decisiones y a ser escuchados en todo lo que les concierne (Unicef, 2015). Por consiguiente, es vital escuchar al alumnado, dar lugar a su voz y conocer lo que les interesa (Basasoro, 2021; Decroly, 2006; Freire, 1997; Rabadan, 2021). En tal sentido, tal y como señalaban diferentes referentes del movimiento de la Escuela Nueva, es imprescindible que los niños y niñas se conviertan en participantes activos de su proceso de aprendizaje (Decroly, 2006; Freinet, 1979; Hall et al., 2014; Hoyuelos, 2021; Malaguzzi, 1996; Montessori, 1937, 1982; Tonucci, 1995, 2001, 2012, 2017). Dado que, si se presentan diversas situaciones donde deban poner en práctica distintas estrategias y tomar sus propias decisiones, desarrollarán las habilidades necesarias para enfrentarse a cualquier problema. Es decir, autorregularse e informarse para poder valorar y determinar qué estrategias necesitarán poner en marcha para su resolución (Guamán y Espinoza, 2022). Así pues, la escuela ha de convertirse en un entorno rico donde el niño y la niña tenga la oportunidad de elegir entre diferentes opciones (Freinet, 1979; Hall et al., 2014; Malaguzzi, 1996; Montessori, 1982; Tonucci, 1995, 2001, 2012, 2017) y dar comienzo a diversos procesos de aprendizaje.

De acuerdo con la teoría sociocultural o el aprendizaje basado en la interacción de Vygotsky (1978), el aprendizaje y el pensamiento se desarrollan mediante la interacción social. Es decir, el aprendizaje no es un proceso individual, sino que está influido por un contexto social: la cultura y las relaciones que establecemos con nuestro entorno. Desde esta perspectiva, la interacción social y participación en actividades de la vida cotidiana resultan ser unas de las experiencias culturales más poderosas; pues mediante ellas accedemos a la información, a la cultura y a la imaginación, entre otros (Gómez et al., 2020; Lave & Packer, 2001). Sin embargo, hemos de revisar las condiciones para fomentarlos (Gómez et al., 2020), debido a que tal y como señalaba Freire (1970) “no hay ignorantes absolutos ni sabios absolutos hay hombres que en comunicación buscan saber más” (p. 87). Así pues, el dialogo igualitario es necesario, ya que en ese principio de horizontalidad, los niños y niñas, docentes y agentes educativos construyen conjuntamente el proceso de enseñanza-aprendizaje.

La escuela rural, vista como un contexto de innovación y mejora, aprovecha aquellas potencialidades que históricamente se han considerado como deficiencias, desarrollando claves pedagógicas de referencia (Santos, 2011). Muestra de ello son los diversos estudios nacionales como internacionales destacan que las escuelas rurales promueven claves pedagógicas ejemplares, entre ellas entre ellas: otro tipo de agrupaciones de niños y niñas que promueven la diversidad de edades (Åberg-Bengtsson, 2009; Boix, 2011; Little, 2001; Santos 2011); un aprendizaje autónomo,

contextualizado (Abós, 2015; Boix, 2011) e interdisciplinar (Boix, 2011); un aprendizaje basado en la colaboración (Abós, 2015; Boix, 2011; Santos, 2011); una diversa dinámica relacional entre niños y niñas (Matías y Vigo, 2020) y relaciones horizontales entre niños y niñas y el profesorado (Santos, 2011); un nuevo rol docente (Vigo y Soriano, 2015); una nueva forma de relacionarse y usar el entorno como recurso pedagógico (Abós, 2019; Boix, 2011); mayor participación familiar (Vigo y Soriano, 2015); un curriculum emergente (Arguiñano, 2019; Karrera y Arguiñano, 2018).

No obstante, varios estudios muestran que las escuelas rurales no solo pertenecen a una realidad poco conocida, sino que también generan escaso interés tanto en la administración educativa como en el ámbito académico (Little, 2001). Si nos situamos en el contexto de este estudio, se puede concluir que los estudios realizados hasta el momento son indicadores de lo mencionado. Esto se refleja en los estudios realizados en este contexto específico, los cuales se han centrado tanto en la evolución histórica y legislativa (Aguirregoitia-Güenechea, 2024) como en las claves inclusivas de dichos centros (Arguiñano, 2019; Karrera et al., 2020). Por lo tanto, siguiendo las palabras de Little (2001), este estudio pretende ‘hacer visible lo invisible’, ya que, como han señalado diversas investigaciones, es fundamental profundizar en el conocimiento de las experiencias educativas de diversas escuelas rurales (Little, 2001; Santos-Gelvasio, 2022). Es más, dado que se trata de entornos con características únicas, los procesos pedagógicos que se desarrollan en estos contextos aún requieren de una comprensión más profunda (Abós, 2020; De la Vega, 2020).

Por ello, con el objetivo de contribuir al conocimiento de buenas prácticas en un contexto particular, este estudio se centra en diversas escuelas de la red educativa *Eskola Txikiak* en el País Vasco. Así pues, teniendo en cuenta los rasgos pedagógicos de estas escuelas, esta investigación pretende:

- Analizar las prácticas pedagógicas principales de las *Eskola Txikiak*.
- Destacar los beneficios que éstas aportan tanto a los niños y niñas como al profesorado.

Método

Nos encontramos ante un estudio etnográfico de carácter crítico (Karrera, 2008). La etnografía es un método apropiado para desarrollar estudios sobre cualquier fenómeno relacionado con la organización escolar, la vida del aula o las relaciones entre los centros escolares y los entor-

nos socioculturales (Maturana y Garzón, 2015). En este sentido, considerando que el objetivo de este estudio es visibilizar experiencias educativas innovadoras e inspiradoras de las *Eskola Txikiak*, se ha optado por el método mencionado. Así, además de otorgar valor y relevancia a las emociones, pensamientos y vivencias de los niños, las niñas y el profesorado, mediante esta investigación se han generado espacios de reflexión para tomar conciencia sobre el valor intrínseco de estos participantes. Puesto que tal y como señalan Beach y Vigo-Arazola (2021), toda persona, tiene conocimiento, así como la capacidad de utilizarlo. Del mismo modo, aludiendo a la naturaleza crítica de esta investigación, la etnografía crítica va más allá de la descripción y la comprensión: está comprometida con el cambio y la transformación social (McLaren, 2005). Así pues, como indica Rodríguez (2022, p. 13) “la etnografía crítica debe ser aprovechada por nosotros los educadores para convertir nuestros recintos y nuestro quehacer docente en espacios de reflexión acerca de las necesidades de las y los estudiantes”. El carácter crítico de esta investigación ha facilitado una colaboración dialógica con la comunidad educativa (Denzin, 2018), contribuyendo a establecer metas compartidas que mejoran el aprendizaje tanto para la comunidad educativa e investigadora.

Muestra

Contexto y participantes

Las *Eskola Txikiak* es un concepto semejante a las escuelas rurales, son escuelas de la red pública situadas en un entorno rural y que, en base a su realidad y/o contexto, ofrecen diversas propuestas educativas. Las escuelas participantes se encuentran en zonas rurales de municipios con menos de 2,000 habitantes, en el territorio histórico de Guipúzcoa, País Vasco. En relación con las características socioeconómicas y culturales de estos municipios, los datos del Instituto Vasco de Estadística (Eustat, 2021) indican que se trata de localidades con un alto índice socioeconómico y una tasa de inmigración cercana al 10%. Ese mismo año, la tasa de inmigración de la CAPV era ligeramente más baja que la de los pueblos objeto de estudio siendo esta de un 8,5% (Eustat, 2022).

En cuanto a la estructura económica, en la mayoría de los municipios predomina el sector industrial, desempeñando un papel clave en el desarrollo económico local (Eustat, 2024). No obstante, es importante destacar que un estudio realizado para caracterizar estos y otros municipios rurales de la Comunidad Autónoma del País Vasco clasifica estas áreas como zonas de especial atención por sus escasos servicios en términos de calidad de vida y bienestar (Gobierno Vasco, 2022).

En cuanto a las escuelas participantes, se organizan en base a talleres integrales (Trueba, 1989), donde los conocimientos curriculares y los aprendizajes convergen de manera globalizada e interdisciplinar. Las aulas se convierten en lugares de utilización común, dedicadas a un área específica, entre otras, las áreas de arte, ciencia e idiomas. De este modo, tanto el espacio como el material se reorganiza de una forma determinada. Durante la jornada escolar, el alumnado va rotando por los diversos talleres y cada docente se sitúa en uno. Otro dato a destacar es que, en dichos espacios, el alumnado de la escuela –de 3 a 12 años–, convive e interactúa conjuntamente (*Eskola Txikiak*, 2024). Así pues, en estas escuelas los niños y las niñas tienen opción, en todo momento, a elegir en qué taller actuar y qué labor o proyecto realizar en dicho espacio.

Tras realizar dos encuentros con las coordinadoras y el excoordinador de la red educativa, se seleccionaron cuatro escuelas para realizar este estudio. En esta selección se consideraron los siguientes aspectos: la trayectoria educativa, las características y el enfoque pedagógico de las diversas escuelas pertenecientes a la red *Eskola Txikiak*. En la Tabla 1 se muestran los centros elegidos y algunas características relevantes de dichas escuelas.

Tabla 1. Centros educativos participantes y características relevantes para la selección

Centros Educativos	Características Relevantes	Talleres integrales de cada centro
CEIP Larraul (Larraul, Guipúzcoa)	<ul style="list-style-type: none"> – Niños y niñas de 3 a 12 años – Nº de niños y niñas: 19 	<ul style="list-style-type: none"> 28. Biblioteca 29. Ciencia 30. Arte 31. Carpintería 32. Expresión corporal y musical 33. Espacio del juego simbólico
CEIP Zerain (Zerain, Guipúzcoa)	<ul style="list-style-type: none"> – Niños y niñas de 3 a 7 años – Nº de niños y niñas: 17 	<ul style="list-style-type: none"> 1. Biblioteca 2. Ciencia 3. Arte 4. Expresión corporal 5. Carpintería 6. Cocina
CEIP Olaberria (Olaberria, Guipúzcoa)	<ul style="list-style-type: none"> – Niños y niñas de 3 a 12 años – Nº de niños y niñas: 33 	<ul style="list-style-type: none"> 1. Biblioteca 2. Ciencia 3. Arte 4. Carpintería 5. Expresión corporal 6. Movimiento

Centros Educativos	Características Relevantes	Talleres integrales de cada centro
CEIP Ezkio-Itsaso (Ezkio, Guipúzcoa)	<ul style="list-style-type: none">– Niños y niñas de 3 a 11 años– Nº de niños y niñas: 51	<ol style="list-style-type: none">1. Lengua2. Ciencia3. Creatividad (aula dirigida al arte)4. El rincón de los artistas (aula dirigida a la expresión corporal y musical)5. Carpintería6. Taller de costura7. Aula de las historias (aula dirigida al juego simbólico)

Instrumentos

Técnicas de recogida de información

Estas fueron las técnicas seleccionadas para la recogida de información:

- Observaciones participantes. Se llevaron a cabo un total de 420 horas de observación participante (Hammersley, 2017). Para ello, el equipo investigador visitó cada escuela durante tres semanas, realizando una jornada completa junto al alumnado y profesorado. Durante estas sesiones, se observó el centro educativo en su conjunto, a los participantes y las diversas prácticas pedagógicas que se desarrollan en estos centros. El equipo investigador participó activamente en las diversas actividades de los talleres para adquirir un conocimiento exhaustivo tanto de las tareas de los alumnos como del contexto que los rodeaba. La información de las sesiones de observación se registró mediante notas de campo y representaciones gráficas, organizadas en un cuaderno de observación con un sistema categorial. Asimismo, se incluyó una sección separada para la información emergente.
- Documentación: se analizaron los documentos internos de cada centro y de la red educativa: los proyectos educativos de cada centro y los documentos que recogen los principios pedagógicos de la red educativa.
- *Focus Groups*. Siguiendo las ideas de Hammersley (1999), este estudio acoge la experiencia, las perspectivas y las acciones de los niños y niñas y del profesorado de dichos centros. Para ello, se realizaron dos *Focus Groups* en cada centro, uno con alumnado y otro con profesorado. Así pues, se realizaron un total de 8 *Focus Groups*. En el estudio participaron 28 niños y niñas y 11 docentes en los *Focus Groups*. La selección del alumnado se realizó bajo la orientación de los docentes de cada centro, dada la necesidad de una participación activa. Se consideraron los siguientes factores para la selección: paridad de género y diversidad de edades. En

los *Focus Groups* con el alumnado, participaron entre seis y ocho niños y niñas. En cuanto al profesorado, se priorizó la participación del profesorado con mayor conocimiento y experiencia en el proyecto educativo del centro. Dado que las escuelas cuentan con plantillas reducidas, en cada *Focus Group* participaron un máximo de tres docentes.

Procedimiento

Análisis de la información

El análisis de contenido se basó en las fases propuestas por Karrera (2008). En la primera fase, se determinó cuál era la información de interés recabada mediante las sesiones de observación, documentación y *Focus Groups*, para consecuentemente proceder a una categorización de carácter inductivo y, con la ayuda del programa *Nvivo v.12*, a la posterior codificación de la información. Tras completar este paso, se inició el análisis de contenido y al proceso de triangulación del mismo. A continuación, se señalan las temáticas emergentes obtenidas mediante la recogida de la información:

- Los aprendizajes y proyectos que emergen en los talleres integrales.
- Bases y aportaciones pedagógicas del proceso de aprendizaje sociocultural.

Se completó la codificación creando y uniendo los diferentes códigos a las temáticas, fuentes de información y participantes correspondientes:

Tabla II. Códigos referentes a las temáticas, fuentes de información y participantes

LAS TEMÁTICAS Y SUS CÓDIGOS	
Los aprendizajes y proyectos que emergen en los talleres integrales	A
Bases y aportaciones pedagógicas del proceso de aprendizaje sociocultural	IIP
LAS FUENTES DE INFORMACIÓN Y SUS CÓDIGOS	
Notas de campo de las sesiones de observación participante	LO
Focus Group niños/niñas	IKTE
Focus Group profesorado	IRTE
Documentos	DOC
LOS Y LAS PARTICIPANTES Y SUS CÓDIGOS	
Docente	IR
Niños/niñas y edad	IK. Edad correspondiente entre "comillas" y en cursiva

Igualmente, se completó la devolución del documento a las personas participantes para asegurar la veracidad de los resultados obtenidos. Tal y como señala Karrera (2008), para realizar un análisis significativo es necesario ir más allá de los citados procedimientos y realizar diferentes triangulaciones. Así garantizando el valor, la fiabilidad y la aplicabilidad de la investigación (Vallejos y Finol de Franco, 2009). Con todo ello se asegura la veracidad de los resultados obtenidos y los posibles sesgos de interpretación (Karrera, 2008). En este sentido, es imprescindible señalar que esta investigación cumple con los requisitos éticos establecidos por el Comité de Ética para la Investigación con Seres Humanos (CEI-SH-UPV/EHU), ya que el mismo comité aprobó su aprobación (código: M10/2022/339). En congruencia con ello, se aseguró la aprobación y el consentimiento informado por parte de todas las personas participantes, así como de sus tutores legales, en el marco de este estudio.

Resultados

En este apartado se presentan los resultados obtenidos empleando las temáticas definidas previamente.

Los aprendizajes y proyectos que emergen en los talleres integrales

Ante la variedad de talleres integrales presentados en la Tabla 1, cada escuela brinda a los niños y a las niñas la oportunidad de elegir el espacio al que desean asistir, así como la tarea o proyecto que desean realizar. Esto implica que los y las infantes también deciden cómo abordarlo, eligiendo los recursos materiales y personales que utilizarán para completar su trabajo. Así es como dan comienzo a un día cualquiera en uno de estos centros escolares: *“A la mañana el alumnado y profesorado del centro se ha reunido en la entrada de la escuela para realizar el recibimiento y dar comienzo al día. Una docente ha dado los buenos días a los niños y niñas y ha preguntado si alguien tenía algo que compartir o que decir al grupo. Los niños y niñas han dicho que no, y la docente les ha explicado cuál era el plan para esa mañana: ha explicado cuáles van a ser las propuestas de lectura de ese día y qué talleres se iban a abrir”* (A_OLO_1).

Son diversas las razones por las que eligen ir a un taller o a otro. Por un lado, la mayoría de los niños y niñas comentan que escogen ciertos espacios porque dentro de ellos pueden completar actividades que les gustan. Asimismo, una niña añade que escoge asistir a ciertos talleres porque, además de encontrar en ellos varios elementos que le agradan,

se siente más inspirada y, como resultado, se percibe más hábil en las tareas que realiza: *“Mis preferidos son el taller de creatividad, ciencia y movimiento. Creatividad porque me gusta un montón y se me da bien y además me encantan las cosas que hay ahí. Voy mucho y me vienen muchas ideas para hacerlas ahí”* (A_OIKTE_ IK8. “11” _10). Por lo contrario, el alumnado también expresa que a veces acude a estos talleres cuando considera que tiene algún aspecto que mejorar: *“Hombre a mí no es que me guste lo que más el taller de la biblioteca, pero suelo ir bastante porque tengo que aprender”* (A_LIKTE_ IK16. “12” _6-7). Señalando que mediante la constancia llegan a disfrutar y aprender en este proceso: *“Voy tantas veces al taller de ciencia y lengua que ya les he cogido el gusto y eso...”* (A_EIKTE_ IK40. “12” _1).

Los niños y niñas emplean diversas fuentes de inspiración para dar comienzo a un nuevo proyecto. En tal sentido, es posible que un niño o niña traiga una idea desde casa: *“Pues igual veo que en casa necesito un perchero y hago un perchero y si no, mira qué bonito, pues hago eso”* (IIP_EIKTE_ IK39. “11” _2-3). En otros casos, los mismos compañeros y compañeras se convierten en fuente de inspiración: *“Igual veo que está haciendo algo y me gusta la idea y entonces la hago”* (IIP_LIKTE_ IK22. “9” _27-28). En el caso de que no saber qué es lo que quieren realizar, los niños y niñas cuentan que suelen necesitar un tiempo para pensar en ello, siendo conscientes de que este paso también tiene un límite: *“Hombre pensando... muchas veces sí, a ver, no todo el día, pero...”* (IIP_LIKTE_ IK22. “9” _32). Del mismo modo, es muy común que los niños y niñas busquen inspiración en los libros, materiales, etc. del taller pedagógico: *“Pues entras al taller de arte y no sabes qué hacer, pues coges unos libros, coges ideas y sacas de ahí...”* (IIP_LIKTE_ IK15. “10” _32). Asimismo, es posible que, en vez de dar comienzo a un nuevo proyecto o reto, asuman el rol de ayudar a sus compañeros o compañeras: *“Dos alumnos estaban realizando un circuito (con piezas de madera) que se les ha caído y se les ha roto. Una compañera observaba todo el tiempo a su alrededor. El profesor le ha propuesto ayudar y la niña ha empezado a ayudar a los dos compañeros”* (IIP_ELO_48).

Una vez que los niños y niñas deciden qué es lo que quieren realizar, son varias las técnicas que cada centro emplea para que los infantes realicen una planificación precisa del proyecto y/o trabajo escogido: *“En el taller de ciencia para realizar un proyecto primero, todos los alumnos tienen que rellenar una ficha que se encuentra en el mismo taller. (...). En la ficha tienen que responder las siguientes cuestiones: ¿Qué quiero saber?, ¿Qué sé?, ¿Qué necesito?, las actividades realizadas, la producción final, ¿Qué he aprendido?”* (IIP_ELO_16). En estos casos,

una docente comenta que en los primeros proyectos que realiza el alumnado es imprescindible la ayuda del docente: *“empiezan, pero sin planificarlo o... ¿no? Abí hay que ayudarles en los pasos que tienen que dar”* (IIP_OIRTE_IR2_6).

También es necesario destacar que hay veces que los proyectos no salen tal y como se lo esperan, en estos casos, admiten que experimentando y trabajando en un mismo proyecto, llegan a mejorar y a completar los proyectos: *“E pues en lengua hice un proyecto sobre Joxe Miguel Barandiarán y bueno me salió un poco mal, pero luego bien... luego lo mejoré y me salió bien y al final los demás aprendieron y yo también aprendí”* (IIP_EIKTE_IK43. “11”_36).

Todos los centros proponen un espacio para que los niños y niñas compartan y presenten el trabajo realizado: *“Una alumna después de desarrollar un trabajo sobre algunos de los países de Europa, se ha comprometido a realizar una presentación del trabajo a toda la escuela. En la presentación le han escuchado todos los alumnos de infantil y primaria. La docente también le ha realizado varias preguntas para que pudiese compartir todo lo aprendido”* (IIP_OLO_2). Convirtiéndose en referentes y fuente de aprendizaje para sus compañeros y compañeras: *“es verdad, porque, las presentaciones también te sirven para aprender porque imagínate, yo he hecho un proyecto sobre los pájaros y una compañera no sabe algo sobre un pájaro, pues viene a mi presentación y lo aprende...”* (IIP_EIKTE_IK39. “11”_30).

Bases y aportaciones pedagógicas del proceso de aprendizaje sociocultural

El profesorado remarca la importancia de que los niños y niñas tengan la oportunidad de experimentar para que puedan llegar a comprender y aprender diferentes aspectos: *“Al final es hacer y actuar, tienen que probar, tienen que experimentar haciendo y actuando, y el hecho de actuar no significa producir, a veces es observar, a veces es una producción escrita, otras veces es la experimentación pero tienen que probar (...) porque al final, no se trata de interiorizar contenidos, si no de obtener ciertas competencias, y las competencias se obtienen haciendo, no hay más”* (IIP_LIRTE_IR12_2). Remarcando la importancia de que los niños y niñas tengan la oportunidad de tocar, sentir y vivenciar el trabajo que están realizando: *“que sea vivencial, ¿no? ¿Hay una frase, ¿no? Aprenderé todo lo que vivencie ¿no? Entonces...”* (IIP_EIRTE_IR35_3-4).

Los niños y niñas también son conscientes de esta forma de aprender, ya que remarcan el valor de la práctica: *“Porque las cosas se aprenden haciéndolas”* (IIP_ZIKTE_IK33. “8” 5-6). Es más, argumentan cómo

mediante la práctica obtienen diversos aprendizajes: *“En la cocina... cuando vamos a la cocina si no cogemos bien las medidas sale mal y entonces ahí aprendemos que hay que coger bien las medidas para que algo salga bien. También en la carpintería o en el taller de creatividad”* (IIP_OIKTE_IK8. “11”_2). Igualmente, los niños y niñas valoran el hecho de poder dedicar tiempo y esfuerzo a un mismo proyecto: *“Por ejemplo en un proyecto estas repitiendo constantemente lo mismo entonces, se te mete en la mente y entonces, lo tienes aprendido”* (IIP_EIKTE_IK39. “11”_9).

Asimismo, varios y varias docentes remarcan la importancia de ofrecer propuestas atractivas y divertidas; destacando la función que cumple el juego durante el proceso aprendizaje: *“todo eso a veces tiene relación con el juego o que el trabajo que haga el alumno sea un trabajo que lo haga con ganas ¿No?”* (IIP_LIRTE_IR12_4). Subrayando la idea de: *“Cualquier actividad que sea dinámica puede ser un juego, pero sí es posible aprender de ello, sobre todo en la infancia diría que se aprende mediante el juego, que no hay otra manera...”* (IIP_EIRTE_IR37_3-4).

Ante este planteamiento, el profesorado expone que: *“Pues para mí ellos no saben, no saben exactamente dónde está el límite ¿no? A veces lo que nosotros consideramos trabajo, puede ser juego para ellos y lo que nosotros consideramos juego para ellos trabajo”* (IIP_LIRTE_IR13_5). La gran mayoría de los niños y niñas consideran que sí es posible aprender y jugar a la vez, siendo conscientes del valor pedagógico de los atractivos y divertidos procesos de aprendizaje: *“Aprendemos jugando”* (IIP_LIKTE_IK15. “10”_1). En la misma línea, es importante señalar que el alumnado es consciente del valor interdisciplinar de cada aprendizaje: *“Porque, por ejemplo, en todas las cosas... e... en todas las cosas por ejemplo hay matemática, todo está dentro de todo, cuando haces una casa, tienes las medidas, y luego... tienes que sumar y medir bien y todo eso es matemática...”* (IIP_OIKTE_IK7. “12”_5).

Asimismo, algunos niños y niñas empiezan a identificar estos puntos de interés y a tomar decisiones desde edades muy tempranas: *“niños de 2-3 años, ahora algunos de 3 años han empezado a elegir un taller porque les gusta, la biblioteca... aun sabiendo que la docente de infantil estará en el taller de expresión ellos acuden a la biblioteca y ya ahí, con sólo 3 años hay una elección, o sea es...”* (A_OIRTE_IR3_8). En este sentido, los niños y niñas relacionan la oportunidad de aprender algo que les interesa con el esfuerzo que realizan en antes y durante proceso de aprendizaje: *“Porque si te toca algún tema que no te gusta pues igual no lo haces tan a gusto y no pones tanto esfuerzo (...) pero si te gusta...vas a hacerlo con ganas”* (A_OIKTE_IK7. “12”_4). Igualmente, varios niños y

niñas dicen que suelen organizarse con anterioridad para poder realizar todos los trabajos previamente planteados, puesto que no siempre tienen la posibilidad de ir a todos los talleres: *“Según cuando... por ejemplo el taller de los artistas está abierto los lunes y los martes y entonces pienso si ir los lunes o los martes, porque si no... no tengo opción”* (A_EIKTE_IK43_“11”_6).

Así, los niños y niñas de los cuatro centros coinciden en que les agrada tener la opción de elegir a dónde ir, qué trabajo realizar y con quién llevarlo a cabo. Consideran que esto facilita su experiencia, ya que no lo perciben como una tarea estrictamente definida o dirigida: *“Al final haces cosas que te gustan, sobre cosas que te gustan”* (A_LIKTE_IK17_“11”_10). Simultáneamente, tanto el profesorado como el alumnado señalan que perciben que los niños y niñas se sienten más tranquilos al trabajar de esta forma: *“Está trabajando en algo que le gusta y eso... es una tranquilidad, es una tranquilidad para ellos”* (A_OIRTE_IR2_3).

Por último, una docente comenta que, de esta manera, tanto el profesorado como los niños y niñas descubren cuáles son las fortalezas de cada persona: *“Sí, y además eso en el mismo taller, sí, se ve claro ¿no? Cada niño ¿no? Que cada alumno tiene su fortaleza ¿no?”* (A_OIRTE_IR4_5). Reduciendo las comparaciones: *“mi trabajo no es mejor que el tuyo, yo no soy más listo que tú, yo no soy más que tu... porque tú tienes tu proyecto y yo tengo el mío y yo cuando presento el mío se valora el mío y cuando tú presentas tu proyecto se valora el tuyo”* (A_LIRTE_IR12_6).

Discusión y conclusiones

El estudio actual, fundamentado en la participación de niños, niñas y personal docente de cuatro escuelas rurales, ha buscado contribuir al escaso conocimiento sobre diversas experiencias educativas situadas en el entorno rural educativo, con el propósito de fomentar la reflexión y promover la mejora del sistema educativo (Rodríguez, 2022). Tal y como expone Tonucci (1996, p. 90), “la escuela rural podría convertirse en un interesante laboratorio experimental de reforma educativa del que podrían surgir indicaciones extrapolables, incluso a las escuelas de las ciudades”. Así pues, es imperativo optimizar las oportunidades inherentes a la investigación etnográfica crítica, ya que, como sostiene Denzin (2017), es nuestro deber promover una transformación global que impulse una democracia plena, inclusiva y participativa.

Asimismo, es necesario destacar que las conclusiones que se desarrollarán en estas líneas han nacido de las vivencias y opiniones de niños y niñas de 6 a 12 años. Este aspecto otorga relevancia a lo expresado, dado

que son ellos y ellas quienes vivencian todos los procesos pedagógicos que ocurren en los centros educativos. Esto nos muestra que, si pretendemos proponer experiencias pedagógicas acordes a la realidad y/o necesidades de los niños y niñas, es vital escuchar y aprender de sus vivencias (Tonucci, 1995, 2001, 2012, 2017). En este sentido, las conclusiones de este estudio demuestran claramente que situar al niño y a la niña en el centro del proceso educativo tiene una influencia fundamental en todo esfuerzo pedagógico. Mostrando cómo, tal y como reflejan las propuestas pedagógicas de Decroly, Freinet, Malaguzzi, Montessori y Tonucci; tales esfuerzos deben comenzar por un acto de escucha (Decroly, 2006; Freinet, 1979; Hall et al., 2014; Hoyuelos, 2021; Malaguzzi, 1996; Montessori, 1982; Tonucci, 1995, 2001, 2012, 2017). Según Freire (2006, p. 71), “Los niños deben tener asegurado el derecho a aprender a decidir, cosa que solo se hace decidiendo”. Como se ha demostrado en este estudio, los procesos de aprendizaje van encaminados a ello, ya que cada proyecto comienza con el proceso de escucha activa. *“Naturalmente, lo importante es que tengan derecho a decidir en todo momento.”* (A_LIRTE_IR12_2).

Es evidente que no todos los niños y niñas, sienten, piensan y actúan de forma similar; pues no todos tienen las mismas necesidades e inquietudes (Decroly, 2006; Freinet, 1979; Hall et al., 2014; Hoyuelos, 2021; Malaguzzi, 1996; Montessori, 1982; Tonucci, 1995, 2001, 2012, 2017). Tal y como señala Robinson (2009), todos poseemos destrezas y capacidades que podemos desarrollar; aptitudes que están por identificar y potenciar. Este estudio evidencia que desde el ámbito escolar no solo es viable aceptar la diversidad intrínseca a la naturaleza humana, sino que, al proporcionar experiencias de aprendizaje sociocultural, cada individuo puede llegar a conocer dichas aptitudes y desarrollarse de manera global (Karrera et al., 2021; Tonucci, 2001, 2012, 2017).

Teniendo en cuenta el significado que guarda la curiosidad en el proceso de aprendizaje del niño y de la niña (Decroly, 2006; Freinet, 1979; Ortiz y Cervantes, 2016; Ouyeder et al., 2016; Tonucci, 1995), la escuela ha de fomentar diversos contextos y/o experiencias para satisfacer y preservar la curiosidad inherente de los niños y de las niñas (Tonucci, 2001, 2012). Los procesos de aprendizaje sociocultural que se presentan en esta investigación han resultado ser experiencias educativas interesantes para alcanzar dicho fin, ya que todos los proyectos parten de las inquietudes de los niños y niñas. Asimismo, basándonos en las ideas de Gómez et al. (2020), podemos concluir que a través de proyectos creativos presentados en el estudio, los infantes van interactuando con diversas personas, experiencias u objetos, y obtienen conocimientos y comprenden el mundo que les rodea. Asimismo, se amplía el espectro

de aprendizajes a los que puede acceder cada niño o niña, puesto que cualquier propuesta puede ser abordada de infinitas maneras: *“O sea que no tienes que hacer siempre lo que te diga el profesor, puedes pensar en algo y hacerlo”* (A_LIKTE_IK22. “9”_12). Lo que aporta a la atención a la diversidad, puesto que el hecho de proporcionar diferentes experiencias o recursos pedagógicos, hace el proceso de aprendizaje se adapte a las necesidades de cada niño o niña y se garantice que todos y todas lleguen a comprender y aprender en dicho proceso (Tonucci, 2001, 2012, 2017).

En la senda al bienestar, es evidente que cuántas más experiencias educativas vivan los niños y niñas, más oportunidades tendrán de encontrar las que más congenien, empiecen a experimentar y aprendan en ese camino (Freinet, 1976, 1979; Tonucci, 2001, 2012, 2017). Así, se concluye que dichas experiencias aportan al proceso de autoconocimiento del niño y de la niña, ya que, tal como señalaron diversos referentes del movimiento pedagógico de la Escuela Nueva, mediante la experimentación los niños y niñas toman conciencia de sus aptitudes y, en consecuencia, son capaces de desarrollarse de manera más plena (Al, et al., 2012; Freinet, 1976, 1979; Karrera et al., 2021; Tonucci, 2001, 2012, 2017). De forma paralela, los miembros de la comunidad educativa reconocen la potencialidad inherente de cada individuo, apreciando la singularidad de sus cualidades, y evitando realizar comparaciones en diferentes situaciones.

Por otro lado, vemos que ofrecer la oportunidad y la confianza de elegir entre escenarios creativos en los que iniciar varios proyectos favorece la motivación y el esfuerzo que ponen en cada proyecto (Freinet, 1976, 1979; Tonucci, 2001, 2012, 2017). Cuando el niño o la niña realiza una tarea vinculada a sus intereses, experimenta una sensación de inspiración y bienestar. Como consecuencia, se percibe a sí misma como una persona hábil y válida (Arguiñano, 2019), lo que le permite sentirse capaz de alcanzar sus metas y completar la tarea con mayor facilidad y de forma más productiva (Yañez, 2016).

Con respecto a las fuentes de inspiración que emplean los niños y niñas a la hora de realizar un proyecto, se concluye que, la diversidad de edades que promueven estas escuelas, ha demostrado ser un recurso pedagógico invaluable que mejora significativamente el aprendizaje (Boix, 2011; Santos 2011; Tonucci, 2012). Los niños y niñas trabajan e interactúan junto con el profesorado y con el resto infantiles diferentes edades, por lo que en varios casos en los que los propios compañeros y compañeras se convierten en una fuente de estímulo creativo. Tal y como señalan Freinet (1976, 1979) y Tonucci (Tonucci, 2001, 2012, 2017, 2022) concluimos que este aspecto también favorece el trabajo del docente, ya

que, cuando los niños y niñas se ayudan unos a otros, no es necesario que el o la docente esté constantemente atento a las necesidades de todo niño y niña.

Asimismo, al poder iniciar proyectos basados en reflexiones que surgen fuera del horario escolar, se concluye que, de este modo, el proceso de aprendizaje se integra con la vida cotidiana. Dejando de ser concebido como algo separado al que solo se puede acceder durante la jornada escolar para abordarlo con materiales didácticos específicos (Freinet, 1979). Es más, vemos como en estas experiencias se diluyen los diversos saberes de las diferentes disciplinas, favoreciendo así una visión más amplia de la realidad y fortificando la capacidad de establecer conexiones entre las distintas disciplinas (Gúzman et al., 2019). Esto se puede relacionar con la perspectiva global sobre el conocimiento y su adquisición de Decroly (2006), ya que los proyectos que surgen de la conexión entre varios talleres respetan la concepción del mundo que tiene el niño y la niña, fomentando la comprensión de que los sujetos y el conocimiento están interconectados. Es por ello que, se concluye que mediante el aprendizaje que surge en dichas experiencias el niño y la niña desarrollan su imaginación y pensamiento creativo (Zuloeta, et al., 2021). Cualidades indispensables para afrontar el mundo cada vez más incierto y globalizado en el que estamos (Pérez, 2019; Vergara, 2016).

Prosiguiendo con el valor pedagógico de los procesos de aprendizaje sociocultural que surgen en estos escenarios, la experimentación que surge en estos contextos cobra gran importancia (Lave y Packer, 2011), puesto que, mediante una experiencia activa, el niño y niña de forma autónoma llega a obtener diversas competencias, reflexiona y comprende los diferentes aprendizajes que subyacen en cada proyecto: *“Pues ella no lo sabía, pero entonces está aprendiendo algo que anteriormente no sabía hacer”* (IIP_LIKTE_IK22. “9”_2). En este contexto, las niñas y los niños desarrollan una mayor conciencia de sí mismos y un compromiso más profundo con su aprendizaje, ya que entienden que, aunque los proyectos no siempre se concreten según la idea original, la persistencia y el esfuerzo les permiten mejorar sus creaciones (Tonucci, 2001). Al mismo tiempo, aprenden de sus propios errores y encuentran nuevas soluciones: *“Se trata de mejorar. Al principio, se prueban diferentes cosas, y luego se empieza a ser más serio y a tomar mejores medidas, y a la tercera vez, ya está bien planificado”* (IIP_OIKTE_IK7 “8”_4). Igualmente, es interesante destacar el momento de la presentación y explicación del trabajo realizado, ya que se ha demostrado que en este instante no sólo aprende quien ha realizado el proyecto sino que los oyentes también aprenden de sus compañeros y compañeras. En definitiva, este tipo de

escuela fomenta la cooperación y el aprendizaje mutuo debido a la interacción que se produce entre las personas que conviven y trabajan en un mismo espacio (Abós, 2015; Boix, 2011; Santos, 2011; Trueba, 1989; Tonucci, 1995, 2001, 2012).

Considerando que las emociones influyen directamente en el aprendizaje disminuyendo o aumentando nuestra conciencia, atención y/o memoria (Immordino-Yang y Damasio, 2007), en este estudio se ha mostrado que es imprescindible crear situaciones de aprendizaje que sean agradables para los niños y niñas. Desde esta perspectiva, es crucial destacar el valor que tiene el juego en el proceso de aprendizaje (Rabadan, 2021; Trueba, 1989; Tonucci, 2012, 2017). En estos contextos, las líneas entre el aprendizaje y el juego se difuminan, lo que facilita la adquisición de conocimientos. La posibilidad de aprender a través de experiencias reales, involucrándose activamente en el proceso educativo, contribuye a que el aprendizaje sea no solo efectivo, sino también divertido, atractivo y motivador (Karrera et al., 2021). Así pues, aunque en varias ocasiones los niños y niñas perciban el aprendizaje como un juego, son conscientes del valor pedagógico que existe detrás de estas vivencias: “*Aprendemos, pero de otra manera*” (IIP_OIKTE_IK7. “12”_1).

Por lo tanto, en este contexto global en el que nos situamos, existen numerosas ventajas al situar a los niños y niñas en el centro del proceso de aprendizaje y convertirlo en el protagonista de la vida escolar (Rico-Gómez y Ponce, 2022). En este sentido, los procesos de aprendizaje sociocultural se presentan como una respuesta interesante, ya que al vivir experiencias alineadas con su personalidad y ejercitar competencias como el esfuerzo, la constancia y la organización, los niños y niñas desempeñan un papel activo fundamental en la construcción de su propio proyecto personal (Pérez y Ahedo, 2020). “Prepararse para la vida significa saber dónde y cómo queremos situarnos en ella y, sin duda, las experiencias vividas son fundamentales. Encontrar, conocer y fortalecer los propios talentos debería ser importante” (Rabadan, 2021, p. 66).

Limitaciones

Dado que este estudio se basa en pequeñas comunidades educativas, el número y la diversidad de participantes es limitado, lo que puede haber influido en los resultados obtenidos. Asimismo, respecto a la diversidad de los participantes, sería interesante incluir a las personas que acaban de llegar a estas escuelas o que no se ajustan al modelo de trabajo de estas; es decir, al profesorado que no se han adaptado a este nuevo rol docente. De esta manera se podrían conocer las inquietudes de quienes trabajan en estas escuelas o identificar aspectos que dificultan la

adaptación a estas propuestas pedagógicas. Creando así la posibilidad de proponer soluciones y estrategias para afrontar estos problemas. La investigación no plantea un modo único de actuar, sino que, en línea con Malaguzzi (1996), propone usar las ideas como inspiración para que cada escuela adapte el enfoque a su propio contexto. Esto puede resultar útil tanto para centros con características similares —incluso condicionadas por su ubicación geográfica (Hall et al., 2014)— como para aquellos con realidades diferentes (Tonucci, 1996).

Líneas futuras

Los autores Abós y Boix (2017) señalan que se pueden desarrollar planteamientos de evaluación alternativos en las experiencias educativas que aúnan la diversidad de edades y las prácticas pedagógicas activas. En este sentido, dado que estamos ante escuelas que organizan y desarrollan prácticas escolares, espacios y tiempos de forma totalmente diferente, sería interesante profundizar en cómo se realiza la evaluación y el seguimiento de los niños y niñas en estas escuelas. A pesar de que durante el proceso de investigación se ha visto que estas escuelas utilizan estrategias de evaluación innovadoras e interesantes, no se ha profundizado en ello. Por ello, en líneas futuras sería interesante estudiar cómo se desarrolla la cuestión de la evaluación en estas experiencias de aprendizaje sociocultural. Todo ello, además de dar más valor a lo mencionado en este estudio, puede suponer ofrecer modelos de utilidad para el resto de las escuelas.

Referencias bibliográficas

Referencias

- Åberg-Bengtsson, L. (2009). The smaller the better? A review of research on small rural schools in Sweden. *International Journal of Educational Research*, 48(2), 100-108. <https://doi.org/10.1016/j.ijer.2009.02.007>
- Abós, P. (2015). El Modelo de Escuela Rural ¿Es un Modelo Transferible a Otro Tipo de Escuela?. *Educação & Realidade*, 40(3), 667-684. <https://doi.org/10.1590/2175-623645781>
- Abós, P. (2020). La escuela ubicada en territorios rurales: una escuela diferente, un reto pedagógico. *Aula*, 26(29), 41-52.
- Aguirregoitia-Güenechea, M. (2024). Antecedents de les escoles rurals a Euskadi (Eskola Txikiak) des de principis del segle XX fins a l'actualitat. *Temps d'Educació*, 66, 177-195. <http://dx.doi.org/10.1344/TempsEducacio2024.66.11>

- Arguiñano, A. (2019). *Hezkuntza berrikuntza: Antzuola Herri Eskolako bilakaera berrikuntza pedagogikoaren bidean* [Tesis Doctoral, UPV/EHU]. ADDI. <https://addi.ehu.es/handle/10810/42035>
- Basasoro, M. (2021). *Ikasleen parte hartzea eskola demokratikoak erakitzeke giltzarri. Irakasleen begirada berritu batetik abiatutako Antzuola Herri Eskolako hezkuntza proiektua*. [Tesis doctoral, UPV/EHU]. Archivo digital docencia investigación.
- Bauman, Z. (2003). *Modernidad líquida*. Fondo de Cultura Económica.
- Brown, L.M. y Flaumenhaft, J. (2019). Student-empowered curricular change. *Phi Delta Kappan*, 100 (6), 13-19. <https://doi.org/10.1177/0031721719834023>
- Boix, R. (2011). ¿Qué queda de la escuela rural? Algunas reflexiones sobre la realidad pedagógica del aula multigrado. *Profesorado. Revista de Currículum y Formación de Profesorado*, 15(2), 13-23. <https://dialhnet.unirioja.es/servlet/articulo?codigo=4089448>
- Cascio, J. (27 de febrero 2023). *Facing the Age of Chaos*. Medium. <https://medium.com/@cascio/facing-the-age-of-chaos-b00687b1f51d>
- Comisión Europea (2016). *EntreComp: The Entrepreneurship Competence Framework*. <http://publications.jrc.ec.europa.eu/repository/bitstream/JRC101581/lfna27939enn.pdf>
- Decroly, O. (2006). *La función de globalización y la enseñanza y otros ensayos*. Biblioteca Nueva.
- De la Vega, L. (2020). Investigación sobre enseñanza y desarrollo profesional docente en escuelas rurales: una revisión. *REXE*, 20(43), 307-325.
- Díaz, F. (2012). *Las claves para educar en tiempos de crisis. Cómo transformar las dificultades en estímulo para una mejor educación*. Toromítico.
- Denzin, N. K. (2018). *The qualitative manifesto: A call to arms*. Routledge. <https://doi.org/10.4324/9780429449987>
- Eskola Txikiak (26 de noviembre, 2024). *Sare Izaera. Printzipio eta baloreak*. Eskola Txikiak. <https://www.eskolatxikiak.eus/>
- Eustat (2021). *Renta familiar media de la C. A. de Euskadi por ámbitos territoriales, según tipo de renta (euros)*. 2021. Recuperado de la base de datos de Eustat. https://www.eustat.eus/elementos/ele0005700/renta-familiar-media-de-la-ca-de-euskadi-por-ambitos-territoriales-segun-tipo-de-renta-euros/tbl0005791_c.html
- Eustat (2022). *Estadística municipal de habitantes*. https://www.eustat.eus/elementos/not0020444_c.pdf
- Eustat (2024). *Udalmap. Departamento de Economía, Trabajo y Empleo*. <https://www.euskadi.eus/indicadores-municipales/web01-a2ogaeko/es/>

- Fielding, M. (2004). Transformative approaches to student voice: Theoretical underpinnings, recalcitrant realities. *British Educational Research Journal*, 2, 295–311.
- Freinet, C. (1979). *Las invariantes pedagógicas*. Morata (2º ed.).
- Freire, P. (1970). *Pedagogía del Oprimido*. Siglo XXI.
- Freire, P. (1997). *Pedagogía de la autonomía. Saberes necesarios para la práctica educativa*. Siglo XXI.
- Gobierno Vasco (2022). *Caracterización de las zonas rurales de Euskadi*. Departamento de Alimentación, Desarrollo Rural, Agricultura y Pesca. https://www.euskadi.eus/contenidos/informacion/caracterizacion_zonas_rurales/es_def/adjuntos/Caracterizacion-zonas-rurales.pdf
- Gómez, L. V., Valverde, K., y Villón, A. M. (2020). El aprendizaje dialógico en los procesos de interacción social del contexto educativo. *Dilemas contemporáneos: Educación, Política y Valores*. 1, 1-13. <https://doi.org/10.46377/dilemas.v35i1.2222>
- Guamán, V. J., y Espinoza, E. E. (2022). Aprendizaje basado en problemas para el proceso de enseñanza aprendizaje. *Revista Universidad y Sociedad*, 14(2), 124-131. <http://scielo.sld.cu/pdf/rus/v14n2/2218-3620-rus-14-02-124.pdf>
- Gúzman, I., Settati, A., y Marín, R. (2019). Transdisciplinariedad y la transversalidad: una experiencia para religar la práctica educativa. *Cultura Educación Y Sociedad*, 10(2), 73–84. <https://doi.org/10.17981/cultedusoc.10.2.2019.06>
- Hall, K., Horgan, M., y Ridgway, A. (2014). *Loris Malaguzzi and the Reggio Emilia Experience*. Bloomsbury.
- Hammersley, M. (1999). "Introduction." In M. Hammersley (Ed.), *Researching School Experience: Ethnographic Studies of Teaching and Learning*, (pp. 1-14). Routledge.
- Hammersley, M. (2017). What is ethnography? Can it survive? Should it?. *Ethnography and Education*. 10.1080/17457823.2017.129845.
- Immordino-Yang, M. H., y Damasio, A. (2007), We Feel, Therefore We Learn: The Relevance of Affective and Social Neuroscience to Education. *Mind, Brain, and Education*, 1, 3-10. <https://doi.org/10.1111/j.1751-228X.2007.00004.x>
- Karrera, I. (2008). El método cualitativo como herramienta crítica válida en la resolución de conflictos escolares [Tesis doctoral, UPV/EHU]. ADDI. <https://ekoizpen-zientifikoa.ehu.es/documentos/60cd611ec63d7a5247e8d17b7?lang=eu>
- Karrera, I., y Arguiñano, A. (2018). Escuelas inclusivas en el País Vasco: organización escolar y currículum emergentes y flexibles. perspectivas

- y desafíos. En H. Arancibia, P. Castillo y J. Saldaña (eds). *Innovación Educativa: perspectivas y desafíos*, (139-174). Universidad de Valparaíso.
- Karrera, I., Arguiñano, A., Basasoro, M., Castillo, P. (2020) Innovative pedagogical experiences at the Basque Country inclusive school. *British Journal of Educational Studies*, 68, 753 - 770. <https://doi.org/10.1080/00071005.2020.171903>
- Karrera, I., Basasoro, M. & Arguiñano, A. (2021) Pedagogical keys to inclusion and democracy in a Basque Country school: gaining knowledge from the educational project at Antzuola Herri Eskola. *Educational Review*, 75(3). <https://doi.org/10.1080/00131911.2021.1923460>
- Tonucci, F. (1995). *El niño y la ciencia. Con ojos de maestro*. Troquel.
- Tonucci, F. (1996). Un modelo para el cambio. *Cuadernos de Pedagogía*, 247, 48-51.
- Tonucci, F. (2001). ¿Cómo introducir la investigación escolar?. *Revista Investigación en la Escuela*, 43, 39-50. <http://hdl.handle.net/11441/60305>
- Tonucci, F. (2012). La escuela que queremos: "Aprender con besitos". *Revista Investigación en la Escuela*, 77, 13-24. <http://hdl.handle.net/11441/60090>
- Tonucci, F. (2017). A modo de introducción...: La diversidad como valor en una escuela que cambia. *Aula Abierta*, 46(2), 9-12. <https://reunido.uniovi.es/index.php/AA/article/view/11980>
- Malaguzzi, L. (1996). *Los cien lenguajes del niño*. Reggio Children. https://www.educarjuntos.com.ar/wp-content/imagenes/los_cien_lenguajes_del_nino.pdf
- Martínez, J., y Rogero, J. (2021). El entorno y la innovación educativa. *REICE. Revista Iberoamericana sobre Calidad, Eficacia y Cambio en Educación*, 19(4), 71-81. <https://doi.org/10.15366/reice2021.19.4.004>
- Matías, E., y Vigo, B. (2020). El valor del lugar en las relaciones de inclusión y exclusión en un colegio rural agrupado. Un estudio etnográfico. *Márgenes, Revista de Educación de la Universidad de Málaga*, 1(2), 90-106. <https://doi.org/10.24310/mgnmar.v1i2.8457>
- Maturana, G. A., y Garzón, C. (2015). La etnografía en el ámbito educativo: una alternativa metodológica de investigación al servicio docente. *Revista de Educación y Desarrollo Social*, 9 (2), 192-205. <https://dialnet.unirioja.es/servlet/articulo?codigo=5386223>
- McLaren, P. (2005). *La vida en las escuelas. Una introducción a la pedagogía crítica en los fundamentos de la educación*. Siglo XXI.
- Mena, L. (2018). *El desarrollo de la autonomía en la infancia. Programa de aplicación en el aula*. [Tesis doctoral, UPV/EHU]. ADDI. https://addi.ehu.es/bitstream/handle/10810/36145/TESIS_MENA_LUCIA_INES.pdf

- Montessori, M. (1982). *El niño: el secreto de la infancia*. Diana.
- Lave, J., y Packer, M. (2011). Hacia una ontología social del aprendizaje. *Revista de Estudios Sociales*, 40, 12–22. <https://doi.org/10.7440/res40.2011.02>
- Little, A. W. (2001). Multigrade teaching: towards an international research and policy agenda. *International Journal of Educational Development*, 21(6), 481–497. [http://dx.doi.org/10.1016/S0738-0593\(01\)00011-6](http://dx.doi.org/10.1016/S0738-0593(01)00011-6)
- Oquendo, S. (2019). Estrategia para el desarrollo de la competencia investigativa en estudiantes de básica primaria. *Encuentros*, 17(2), 95–107. <https://www.redalyc.org/journal/4766/476661510009/html/>
- Ortiz, G., y Cervantes, M. L. (2016). La formación científica en los primeros años de escolaridad. *Panorama*, 9(17), 10–23. <https://doi.org/10.15765/pnrm.v9i17.788>
- Oudeyer, P.-Y., Gottlieb, J., y Lopes, M. (2016). Intrinsic motivation, curiosity, and learning: Theory and applications in educational technologies. *Prog Brain Res*, 229, 257–284. <https://pubmed.ncbi.nlm.nih.gov/27926442/>
- Pérez, Á. I. (2019). Ser docente en tiempos de incertidumbre y perplejidad. *Márgenes, Revista de Educación de la Universidad de Málaga*, 3–17, <https://revistas.uma.es/index.php/mgn/article/view/6497>
- Pérez, J., y Ahedo, J. (2020). La educación personalizada según García Hoz. *Revista complutense de educación*, 31(2), 153–161.
- Rabadan, K. (2021). *Bidean*. Txalaparta.
- Rico-Gómez, M. L., y Ponce, A. I. (2022). El docente del siglo XXI: perspectivas según el rol formativo y profesional. *Revista mexicana de investigación educativa*, 27(92), 77–101. <https://www.scielo.org.mx/pdf/rmie/v27n92/1405-6666-rmie-27-92-77.pdf>
- Robinson, K. (2009). *El elemento. Descubrir tu pasión lo cambia todo*. Grijalbo.
- Rodríguez, M. E. (2022). La etnografía crítica: un transmétodo rizomático transcomplejo en la transmodernidad. *Perspectivas Metodológicas*, 22, 21–21. <https://doi.org/10.18294/pm.2022.3992>
- Santos, L. (2011). Aulas multigrado y circulación de los saberes: especificidades didácticas de la escuela rural. *Revista de currículum y formación de profesorado*, 15(2), 71–91).
- Santos-Gelvasio, A. (2022). Inclusión y atención a la diversidad en el aula rural multigrado: un estudio de caso. *Revista Caribeña de Investigación Educativa*, 6(2), 15–34. <https://doi.org/10.32541/recie.2022.v6i2.pp15-34>
- Urbano-Carazo, M. I. (2022). Factores Determinantes de la Gestión de Conocimiento y su incidencia en el Desempeño Organizacional en

- entornos BANI: Una revisión teórica. *Revista Científica Anfibios*, 5(1), 11-25. <https://doi.org/10.37979/afb.2022v5n1.99>
- Unesco (2015). *Educación 2030. Declaración de Incheón y marco de acción para la realización del Objetivo de Desarrollo Sostenible n° 4*. https://unesdoc.unesco.org/ark:/48223/pf0000245656_spa
- Unicef (2015). *Convención sobre los derechos del niño*. <https://www.unicef.es/publicacion/convencion-sobre-los-derechos-del-nino>
- Vallejos, R., y Finol de Franco, M. (2009). La triangulación como procedimiento de análisis para investigaciones educativas. *REDHECS: Revista electrónica de Humanidades, Educación y Comunicación Social*, 4(7), 117-133. <https://dialnet.unirioja.es/servlet/articulo?codigo=3063110>
- Vergara, J. J. (con Pérez Gómez, A. I.). (2016). *Aprendo porque quiero*. Ediciones SM.
- Vigo-Arrazola, B., y Beach, D. (2022). Las escuelas rurales ante las políticas de mercado: Un análisis meta-etnográfico. *Archivos Analíticos de Políticas Educativas*, 30(173). <https://doi.org/10.14507/epaa.30.6962>
- Vigo, B., y Soriano, J. (2015). Family involvement in creative teaching practices for all in small rural schools. *Ethnography and Education*, 10(3), 325-339. <https://doi.org/10.1080/17457823.2015.1050044>
- Vygotsky, L. (1978). *El desarrollo de los procesos psicológicos superiores*. Austral
- Yañez, P. (2016). El proceso de aprendizaje: fases y elementos fundamentales. *Revista San Gregorio*, 11, 70-81. <https://dialnet.unirioja.es/servlet/articulo?codigo=5585727>
- Zuloeta, E. J., Rojas, N. R., y Caramutti, V. (2021). La creatividad en estudiantes educación inicial: una revisión bibliográfica. *Revista Conrado*, 17(82), 260-267. <https://conrado.ucf.edu.cu/index.php/conrado/article/view/1957>

Comparación entre el pensamiento visual y el aula invertida en la enseñanza de las relaciones internacionales

Comparing Visual Thinking and Flipped Classroom in International Relations Teaching

<https://doi.org/10.4438/1988-592X-RE-2025-412-739>

Andrea Betti

<https://orcid.org/0000-0002-4629-4572>

Universidad Pontificia Comillas

Pablo Biderbost

<https://orcid.org/0000-0002-4086-3658>

Universidad de Salamanca

Esther Vaquero

<https://orcid.org/0000-0001-5196-3462>

Universidad Pontificia Comillas

Resumen

Los avances en la innovación docente han experimentado recientemente un notable impulso hacia formatos capaces de incrementar y mejorar el aprendizaje activo del estudiantado. Una amplia variedad de formatos de enseñanza basados en este concepto se ha implementado en diversas disciplinas universitarias, desde las ciencias naturales hasta las ciencias sociales. La mayoría de los análisis empíricos se han centrado hasta ahora en comparar los efectos de los formatos de enseñanza tradicionales y activos sobre el rendimiento académico del alumnado. Esto ha permitido ampliar el conocimiento científico sobre los beneficios y limitaciones de los distintos formatos de enseñanza.

No obstante, contrastar los formatos tradicionales de clase magistral con enfoques basados en actividades orientadas al aprendizaje activo conlleva el riesgo de generar dicotomías simplistas y poco útiles entre la instrucción «tradicional» y la «activa». En este artículo seguimos una estrategia distinta. Mediante el uso de

pruebas no paramétricas, comparamos dos formatos de enseñanza basados en dos estrategias diferentes de aprendizaje activo: el aula invertida (*flipped classroom*) y el pensamiento visual (*visual thinking*).

Se comparan los resultados de los dos grupos, integrados por estudiantes de grado matriculados en un doble programa en Relaciones Internacionales y Comunicación Global, tanto en términos de rendimiento académico —es decir, sus calificaciones— como en un conjunto de habilidades blandas, tales como la percepción del aprendizaje, la autoeficacia y el trabajo en equipo. El estudio detecta un efecto relevante en las calificaciones, ya que el estudiantado que siguió el formato de aula invertida obtiene resultados superiores a quienes fueron enseñados mediante el formato de pensamiento visual. Por otro lado, el estudio identifica únicamente un efecto significativo en la habilidad blanda de percepción del aprendizaje, con puntuaciones más altas entre el estudiantado del formato de aula invertida en comparación con el del formato de pensamiento visual.

Si bien esto no indica necesariamente que los formatos de aula invertida sean más eficaces que los de pensamiento visual, la diferencia observada en las calificaciones y en la percepción del aprendizaje puede señalar que distintos formatos de enseñanza pueden resultar, de manera diversa pero igualmente útil, adecuados según el objetivo o la necesidad docente.

Palabras clave: Aprendizaje activo, Relaciones Internacionales, Pensamiento Visual, Películas, Documentales, Clase Invertida, Educación superior.

Abstract

Developments in teaching innovation have recently seen a great push towards formats that can increase and improve students' active learning. A wide array of teaching formats based on this concept has been implemented across various university disciplines, from natural to social sciences. Most empirical analyses have so far focused on comparing the effects of traditional and active teaching formats on students' academic performance. This has improved scientific knowledge about the benefits and drawbacks of several teaching formats. Nevertheless, contrasting traditional lecturing formats with activity-based approaches to active learning risks yielding simplistic and unhelpful dichotomies between "traditional" and "active" instruction. In this article, we follow a different strategy. By using non-parametric tests, we compare two different teaching formats based on two different active learning strategies, such as flipped classroom and visual thinking. Students' performances in the two groups, composed of undergraduate students enrolled in a dual degree in International Relations and Global Communication, are compared in terms of academic results, that is their grades, and a set of soft skills, such as learning perception, self-efficacy, and teamwork. The study detects a relevant effect in terms of grades, with students through the flipped format outperforming students taught through the visual thinking format. On the other hand, the study detects only one relevant effect in the soft skill of learning perception, with students in the flipped format scoring higher than students in the visual thinking format. While this does not necessarily indicate

that FC formats are more effective than VT formats, this difference in students' grades and learning perception can signal that different teaching formats can be differently but equally useful depending on the teaching goal or need.

Keywords: Active learning, International Relations, Visual Learning, Films, Documentaries, Flipped Classroom, Higher Education.

Introducción

En las últimas décadas, la educación universitaria ha experimentado un importante impulso hacia formatos de enseñanza que pueden mejorar el «aprendizaje activo» de los estudiantes. En lugar de los formatos tradicionales, en los que el profesor imparte clases magistrales y los estudiantes toman apuntes que utilizarán después de clase para intentar memorizar los contenidos (Bligh, 1998), cada vez se exige más que la enseñanza y el aprendizaje sean colaborativos, «basados en problemas», ricos en recursos digitales y capaces de fomentar el pensamiento crítico y el trabajo en equipo (Collins y Halverson, 2010). Se supone que esto mejora no solo las habilidades de orden inferior de los estudiantes, como la «transferencia» y la «memorización» de «conocimientos básicos» (Omelicheva y Avdeyeva, 2008), sino también sus «habilidades de orden superior», como el análisis, la síntesis y la evaluación (Anderson y Krathwohl, 2001).

Muchos profesores han respondido a esta llamada implementando estrategias de aprendizaje activo en sus clases, tanto en ciencias sociales como en ciencias naturales (Strelan et al., 2020; Galindo-Domínguez, 2021). Como resultado, se ha vuelto más común diseñar estudios destinados a comprender y evaluar empíricamente la eficacia de tales estrategias en el aprendizaje de los estudiantes (Hussain et al., 2023; Naing et al., 2023). Los primeros ejemplos de estos estudios tendían a comparar los formatos de enseñanza tradicionales y activos (Cheng et al., 2018; Talbert y Bergmann, 2017; Bergmann y Sams, 2012;). Estas comparaciones han aumentado significativamente el conocimiento colectivo sobre sus posibles ventajas e inconvenientes. No obstante, la mera oposición entre los formatos de enseñanza tradicionales y activos conlleva el riesgo de simplificar en exceso un debate importante, exagerar las ventajas de las estrategias de aprendizaje activo y llegar a la conclusión insuficiente de que el aprendizaje activo es mejor que el tradicional, a menudo descartado como aprendizaje pasivo.

Por este motivo, siguiendo una tendencia al alza (Betti et al., 2022; Jensen et al., 2015; Lai y Hwang, 2016), este artículo investiga la eficacia del aprendizaje activo en el ámbito de la enseñanza de las relaciones

internacionales (RI), comparando dos formatos de enseñanza basados en estrategias de aprendizaje activo. Se impartió la misma clase de RI a dos grupos de estudiantes. A uno se le enseñó mediante un formato de clase invertida (FC en adelante por su denominación en inglés, *Flipped Classroom*), mientras que al otro se le enseñó mediante un formato de pensamiento visual (VT en adelante por su denominación en inglés, *Visual Thinking*) basado en películas y documentales. Se compararon sus resultados en términos de calificaciones y un conjunto de habilidades blandas, como el trabajo en equipo, la autoeficacia y la percepción del aprendizaje.

Por un lado, nuestros resultados identificaron un efecto relevante en términos de calificaciones, ya que los estudiantes a los que se impartió la clase mediante el formato FC obtuvieron mejores resultados que los estudiantes a los que se impartió mediante el formato VT. Por otro lado, la única habilidad blanda en la que el estudio detectó un efecto relevante fue la percepción del aprendizaje, en la que los estudiantes de FC obtuvieron puntuaciones más altas que los de VT.

Clase invertida y pensamiento visual

El FC se basa en la idea de que «lo que tradicionalmente se hace en clase ahora se hace en casa, y lo que tradicionalmente se hace como tarea ahora se completa en clase» (Bergmann y Sams, 2012, p. 13). Las clases tradicionales en el aula, basadas en la enseñanza frontal de los contenidos por parte del profesor, se sustituyen por materiales, normalmente videoclases pregrabadas, que los estudiantes deben estudiar antes de la clase. Así, el tiempo de clase se dedica íntegramente a resolver las dudas de los alumnos sobre los contenidos y a realizar ejercicios prácticos, como debates basados en problemas, presentaciones o trabajos en grupo, diseñados para aplicar lo que los alumnos han aprendido a través de las videoconferencias. «Al trasladar la impartición de contenidos fuera del aula», el tiempo de clase puede utilizarse de forma más eficiente para realizar actividades que prometen fomentar las habilidades de resolución de problemas y trabajo en equipo de los alumnos (Jenkins, 2015, p. 607). El objetivo general es convertir el aula «en un entorno de aprendizaje dinámico e interactivo», en el que los profesores ayuden a los alumnos en su propio proceso de aprendizaje, fomentando una participación más creativa con la materia (Berge y Nederveld, 2015, p. 163).

El VT se definió originalmente como «un método basado en la investigación», destinado a desarrollar «las habilidades de pensamiento de los estudiantes mediante la observación y el debate del arte» (Tisham et al., 1999, p. 1). La idea principal es que, a través de la observación de imágenes, el aprendizaje puede ser más interactivo y entretenido, lo que

fomenta la capacidad de los estudiantes para comprender conceptos abstractos y difíciles a través de una experiencia visual. Lejos de limitarse a imágenes artísticas, los formatos de enseñanza de VT pueden basarse en una amplia gama de materiales visuales, como películas, documentales (Van Munster y Silvest, 2015; Heck, 2017), cómics (Schmid, 2020) o videojuegos (Valeriano y Habel, 2016), integrando instrumentos de enseñanza tradicionales, como lecturas y clases magistrales.

Estudios previos

Los estudios que midieron los efectos de las técnicas de aprendizaje activo en el aprendizaje de los estudiantes suelen operativizar el rendimiento académico como las calificaciones obtenidas en las tareas y los exámenes. Sin embargo, varios estudios también tuvieron como objetivo evaluar los efectos del aprendizaje activo en las habilidades blandas, entendidas como «cualidades interpersonales... y atributos personales que uno posee» (Robles, 2012, p. 453). El trabajo en equipo es una de las habilidades blandas más estudiadas. Se suele referir a la capacidad de «dirigir y coordinar las actividades de otros miembros del equipo», «anticipar las necesidades de otros miembros del equipo» y «aplicar» y «ajustar estrategias» (Salas et al., 2005, pp. 558-559). La percepción del aprendizaje se define como el «esfuerzo cognitivo necesario durante el aprendizaje» (Deslauriers et al., 2019: 19251), que suele traducirse en la satisfacción o insatisfacción de los estudiantes con los formatos de enseñanza. Por último, la autoeficacia se define a menudo como «las creencias de los estudiantes sobre su capacidad para mostrar un determinado comportamiento de aprendizaje» (Baars y Wijnia, 2018, p. 127) basadas en sus convicciones sobre su propia capacidad para «alcanzar los niveles de rendimiento designados» (Bandura, 1994, p. 2).

A diferencia de otras disciplinas, especialmente las ciencias naturales (Naing et al., 2023), ha habido relativamente menos interés por las técnicas de aprendizaje activo en Ciencias Políticas (CP) y Relaciones Internacionales (RI), especialmente en lo que se refiere a los análisis empíricos sobre la eficacia de dichas técnicas en el aprendizaje de los estudiantes. En cuanto a FC, hay algunos estudios de CP y RI que compararon los formatos de enseñanza tradicionales, centrados en el profesor y la exposición de contenidos durante el tiempo de clase, con formatos invertidos, basados en la sustitución de la enseñanza tradicional en el aula por videoconferencias pregrabadas que se ven antes de la clase, de modo que el tiempo de clase se puede dedicar a ejercicios y aplicaciones. Algunos detectaron una mejora en las calificaciones de los estudiantes (Touchton, 2015), mientras que otros no encontraron ninguna diferencia

significativa entre el formato tradicional y el invertido en términos de rendimiento académico de los estudiantes (Lambach et al., 2017). Se pueden encontrar resultados divergentes similares en los estudios que compararon los formatos tradicionales y invertidos en términos de las habilidades blandas de los estudiantes. Mientras que algunos identificaron una mejora en las habilidades de trabajo en equipo de los estudiantes a los que se enseñaba mediante FC (Cit., p. 563), otros detectaron un empeoramiento de las habilidades de trabajo en equipo de los estudiantes en FC (Jenkins, 2015, p. 610). En cuanto a la percepción del aprendizaje, los estudios disponibles vuelven a divergir. Mientras que algunos observaron que los estudiantes se dividían fundamentalmente entre los que preferían un formato invertido y los que preferían uno tradicional (Lambach et al., 2017), otros detectaron una preferencia de los estudiantes por las «sesiones de clase mixtas», en lugar de «tener todas las sesiones de clase invertidas» (Jenkins, 2015, p. 610). Por último, según nuestro conocimiento, no hay estudios de CP o RI disponibles sobre la relación entre el FC y la autoeficacia de los estudiantes.

No obstante, siguiendo la estela procedente de otras disciplinas, varios estudios de CP y RI consideraron más útil comprobar la eficacia del formato FC frente a «un modelo de control que utiliza el aprendizaje activo» (Jensen et al., 2015, p. 11). Esto reduce el riesgo de examinar formatos de enseñanza demasiado diferentes para poder compararlos y aumenta las oportunidades de «analizar los efectos y determinar un factor causal específico» (Cit., p. 2). Algunos identificaron una mejora en las calificaciones de los estudiantes que recibieron enseñanza a través de un FC, en comparación con los estudiantes que recibieron clases sincrónicas en línea (Whitman Cobb, 2016). De manera similar, algunos descubrieron que las calificaciones de los estudiantes pueden mejorar con FC, si se diseña en combinación con otras formas de enseñanza a distancia, tanto *offline* como *online* (Van der Zwan y Afonso, 2019). En cuanto a las habilidades blandas, uno de los pocos estudios que pudimos encontrar que comparaba el formato invertido con otras estrategias de aprendizaje activo detectó una mejora simultánea en el rendimiento académico y la percepción del aprendizaje, y los estudiantes expresaron su satisfacción con FC (Whitman Cobb, 2016).

A diferencia de otras disciplinas (Yen et al., 2018; Lin et al., 2019), en las que existe un consenso más claro a favor de los beneficios del formato de enseñanza invertida tanto en el rendimiento académico de los estudiantes como en sus habilidades blandas, en los campos de la educación física y las relaciones internacionales se observa menos acuerdo: diferentes estudios han llegado a conclusiones dispares y parcialmente contradictorias. Esto se suma a la impresión de que se trata de una dis-

ciplina en la que el interés y la curiosidad por la innovación docente siguen estando empañados por la prudencia y el escepticismo. Por lo tanto, el primer objetivo de este estudio es investigar si un formato de enseñanza basado en el FC puede mejorar el rendimiento de los estudiantes en términos de calificaciones y habilidades blandas.

Estas tendencias son aún más visibles en el estudio de la eficacia de los formatos de enseñanza VT en las asignaturas de ciencias sociales. En disciplinas como la historia, el uso de materiales visuales, incluidas las películas, ha sido habitual (Rose, 2016). Aunque la mayoría de los estudios disponibles han sido en gran medida anecdóticos y no se han basado en análisis empíricos sistemáticos, algunos historiadores han descubierto que el uso de películas en las clases de historia tiene un impacto positivo en la comprensión de situaciones complejas por parte de los estudiantes (Wagner, 2018). Sin embargo, esos mismos historiadores también informaron de la dificultad de utilizar materiales visuales, como películas, en una clase de Historia, ya sea por la dificultad de los alumnos para discernir los hechos de las interpretaciones visuales y artísticas (Stoddard, 2012) o por la falta de habilidades analíticas de los medios visuales por parte de los profesores (Wagner, 2018).

Se puede observar aún más «reluctancia» (Dean, 2019, p. 257) en el uso de formatos VT en CP y RI. Para algunos, el escepticismo sobre «las dinámicas afectivas y emocionales que sustentan las formas cotidianas de participación y compromiso políticos» fomentaría una «reluctancia» de estas disciplinas a utilizar materiales visuales (Cit.). Posiblemente por ello, no son muchos los profesores de CP y RI que se han mostrado dispuestos a utilizar materiales visuales en sus clases, aunque esta tendencia parece estar cambiando. Estudios recientes, por ejemplo, detectaron la preferencia de los estudiantes por las clases impartidas mediante diapositivas ricas en imágenes en lugar de diapositivas con mucho texto, ya que con el primer formato se sentirían más comprometidos con los contenidos de la clase. Además, asociarían una mayor sobrecarga cognitiva con las diapositivas con mucho texto, ya que generarían un dilema constante entre leer las diapositivas y escuchar al profesor. Las imágenes mejorarían, por tanto, su satisfacción y percepción del aprendizaje en la clase. Sin embargo, el mismo estudio también advierte sobre el abuso de imágenes durante la explicación de los contenidos, no solo porque su contenido puede ser a veces controvertido, sino también porque no siempre son suficientes para comunicar conceptos o datos complejos. El riesgo de simplificación excesiva podría reducir la capacidad de los estudiantes para leer y analizar críticamente la información presentada en libros y artículos especializados (Roberts, 2017).

Otros estudios han comparado los formatos tradicionales y VT basados en el uso de películas. La mayoría de ellos han identificado efectos positivos en las habilidades blandas de los estudiantes, como el pensamiento creativo (Valeriano, 2013; Kuzma y Haney, 2001; Weber, 2001) y las habilidades analíticas (Lieberfeld, 2007). Las películas pueden aumentar la capacidad de análisis de los estudiantes y ayudarles a conectar y contextualizar hechos y personajes históricos (Gokcek y Howard, 2013). Por otro lado, otros han advertido sobre las posibles desventajas del uso de películas con fines didácticos, por ejemplo, cuando simplifican en exceso la realidad mediante representaciones inexactas de los hechos (Kuzma y Haney, 2002, p. 93) o cuando contienen representaciones estereotipadas y sesgadas de culturas específicas (Giglio, 2002). Aunque la mayoría de estos estudios no se basan en análisis empíricos sistemáticos, algunos también han tratado de observar los posibles efectos de las películas en el rendimiento académico y la percepción del aprendizaje de los estudiantes. Algunos identificaron beneficios en el uso de películas, como una mejora en la capacidad de los estudiantes para comprender y retener conceptos (Sunderland et al., 2009), gracias al hecho de que las películas ayudarían a «hacer más comprensibles las teorías y conceptos abstractos» (Gokcek y Howard, 2013, p. 441; Kiasatpour, 1999), entretenidas (Kuzma y Haney, 2001, p. 47) o atractivas (Gokcek y Howard, 2013; Iretzberger, 2021).

Sin embargo, otros estudios señalaron que la percepción del aprendizaje de los estudiantes no necesariamente se beneficia del uso de películas. En ocasiones, las películas pueden aumentar la sensación de una mayor carga de trabajo (Kiasatpour, 1999, p. 85) o incluso de confusión, especialmente porque a los estudiantes les puede resultar difícil distinguir entre la realidad y la ficción (Lieberfeld, 2007, p. 573). Aunque en ocasiones las películas pueden mejorar la comprensión de los estudiantes de procesos complejos, como los relacionados con la toma de decisiones, es posible que no mejore su comprensión de «los conceptos teóricos y cómo se pueden aplicar» (Inoue y Krain, 2014, p. 20). Uno de los pocos análisis empíricos que uso métodos cuantitativos identificó resultados contradictorios. Las películas pueden aumentar la comprensión, el compromiso y las calificaciones de los estudiantes, pero también pueden tratar los temas de forma simplificada en aras de la representación artística, lo que reduce la capacidad de los estudiantes para ver una relación clara entre el contenido académico y las representaciones ficticias (Swimelar, 2013, p. 24). Todos estos análisis se basaron en comparaciones entre los formatos de enseñanza tradicionales y los de VT. La única excepción que pudimos encontrar fue un estudio que comparaba dos formatos de enseñanza diferentes, ambos basados en estrategias de aprendizaje activo,

uno basado en películas y otro en juegos, y que detectó que este último tenía un efecto más beneficioso en las calificaciones y la percepción del aprendizaje de los estudiantes que el primero (Brandle, 2020).

En conclusión, a diferencia de otras materias en las que se puede encontrar un consenso más claro sobre los beneficios del uso de películas para el rendimiento y la satisfacción de los estudiantes, los estudios de CP y RI ofrecen una gama más amplia de resultados. Estas divergencias podrían estar motivadas en parte por la relativa escasez de estudios. Pocos de ellos se basan en análisis sistemáticos y empíricos de los efectos del uso de películas en las calificaciones y las habilidades blandas de los estudiantes. Además, la mayoría de ellos se basan en comparaciones básicas entre las clases magistrales tradicionales y el uso de películas. Esto lleva a la necesidad de comprobar de forma más sistemática los posibles beneficios e inconvenientes del uso de formatos VT frente a otros formatos de aprendizaje activo. CP y RI han sido menos propensos a utilizar formatos VT en el aula, lo que hace que el estudio de su posible potencial para fines didácticos siga estando relativamente poco explorado. El segundo objetivo de este estudio es, por tanto, investigar si un formato de enseñanza basado en VT puede mejorar el rendimiento de los estudiantes en términos de calificaciones y habilidades blandas.

Por esta razón, en este estudio proponemos analizar dos formatos de enseñanza diferentes, ambos basados en estrategias de aprendizaje activo, uno diseñado de forma invertida y el otro en torno al uso de materiales visuales, como películas y documentales. El objetivo es añadir datos empíricos sobre la eficacia de dos importantes estrategias de aprendizaje activo para el aprendizaje de los estudiantes, contribuyendo a un debate más amplio, aún parcialmente subdesarrollado en CP e RI, sobre qué formatos de enseñanza pueden implementarse para mejorar el rendimiento y la experiencia de los estudiantes.

Diseño y métodos de la investigación

Nuestro objetivo es descubrir si las estrategias docentes basadas en la FC y la VT pueden conducir a una mejora en el rendimiento académico de los estudiantes, tanto en lo que respecta a las habilidades técnicas, es decir, sus calificaciones, como a las habilidades blandas, tales como el trabajo en equipo, la autoeficacia y la percepción del aprendizaje.

Este estudio cuasiexperimental se llevó a cabo en dos secciones diferentes de una asignatura troncal obligatoria de segundo curso de grado, denominada «Sistemas políticos comparados», como parte de una doble titulación en Relaciones Internacionales y Comunicación Global. A un grupo se le impartió la mitad de la clase (7 de las 14 semanas de clase) mediante una

metodología FC, mientras que al otro grupo se le impartió la misma mitad de la clase mediante una metodología VT. Ambos formatos de enseñanza se basaron en los mismos contenidos, fueron impartidos por el mismo profesor y se llevaron a cabo durante las siete últimas semanas de la clase.

Para el formato FC, el profesor grabó primero una serie de videoconferencias, con el apoyo de la Unidad de Innovación Docente de la Universidad donde se llevó a cabo el estudio. Los vídeos incluían explicaciones del contenido de la clase y secciones específicas en las que se destacaban los conceptos clave. Los estudiantes debían verlos atentamente antes de asistir a clase. A partir de esas videoconferencias, los estudiantes debían realizar diversas actividades en el aula, bajo la supervisión del profesor. Por ejemplo, el profesor hacía preguntas para estimular la reflexión individual o en grupo sobre cómo definir y relacionar conceptos y categorías políticas específicas. En otros casos, los estudiantes debían leer y luego comentar lecturas específicas seleccionadas por el profesor y relacionadas con las videoconferencias. Por último, se pidió a los alumnos que trabajaran individualmente o en grupo para presentar al profesor y al resto de sus compañeros los resultados de pequeñas tareas de investigación que debían realizar previamente en casa. El objetivo de estas actividades era promover la reflexión individual y colectiva sobre cuestiones complejas, fomentar el aprendizaje cooperativo y aumentar la posibilidad de que el profesor proporcionara comentarios instantáneos y públicos.

Para el formato VT, los estudiantes tenían que ver una variedad de contenidos visuales antes de la clase, como documentales y películas, seleccionados por el profesor y relacionados con algunos de los conceptos clave de la clase. Los contenidos visuales se asignaban semanalmente. Después de cada tarea, los estudiantes debían asistir a clase y participar en actividades diseñadas por el profesor y destinadas a reforzar el conocimiento y la comprensión de los conceptos principales de la clase. Estas actividades incluían debates en clase y breves presentaciones basadas en pequeñas tareas de investigación sobre el contenido de los materiales visuales asignados. Además, trabajando en grupo fuera del aula, se pidió a los alumnos que produjeran breves contenidos visuales, como vídeos grabados o selecciones de fotografías, a través de los cuales profundizaran en conceptos y fenómenos específicos relacionados con el contenido de la clase. De esta manera, los alumnos no actuaban simplemente como consumidores pasivos de contenido visual. Las películas y los documentales se convirtieron en instrumentos para analizar, evaluar y utilizar como puntos de partida para una reflexión más profunda basada en pequeñas tareas de investigación. Esto les permitió experimentar con habilidades como el análisis, la comparación, el contraste y la síntesis.

Además, a través de las tareas que implicaban la producción propia de contenidos visuales, los estudiantes pudieron entrenarse para comunicar visualmente los resultados y hallazgos de su proceso de aprendizaje. Así, los contenidos visuales se utilizaron tanto como recursos didácticos como contribuciones propias de los estudiantes.

Las primeras siete semanas de clase se impartieron mediante un formato semi-tradicional basado en una combinación de clases magistrales y técnicas de aprendizaje activo, como presentaciones, actividades en equipo y debates. Se dedicaron dos horas semanales a las clases magistrales, mientras que las otras dos se dedicaron a que los alumnos realizaran diversas actividades en clase basadas en su participación activa, como presentaciones de pequeñas tareas de investigación, debates o proyectos en equipo.

Como muestra la tabla 1, no pudimos recopilar los datos de todos los alumnos matriculados en los dos grupos, ya que algunos no respondieron a ninguna de las encuestas.

Tabla 1. Descripción de la muestra

	Grupo 1: Metodología FC	Grupo 2: Metodología VT
Estudiantes matriculados	45	62
Muestra	31 (69 %)	42 (68 %)

Se analizaron ambos grupos para averiguar si existían diferencias significativas en cuanto a su rendimiento académico antes del inicio de la clase que se utilizó para el estudio. Tras observar que las series son normales, la prueba t de muestras independientes muestra que no hubo diferencias significativas ni en las calificaciones medias ni en la varianza entre los dos grupos (véase la tabla 2). Esto permitió detectar que no había diferencias significativas entre los dos grupos en cuanto al rendimiento académico previo. Tampoco se encontraron diferencias significativas en cuanto al nivel socioeconómico de los estudiantes. Por último, todos los estudiantes son de nacionalidad española y han sido formados en el sistema educativo tradicional español, basado en la enseñanza frontal tradicional. De este modo, se pudo establecer que los dos grupos eran comparables.

Para responder a las dos preguntas de investigación siguientes, este estudio emplea una metodología basada en la comparación de los resultados medios de cada grupo de estudiantes, tanto en habilidades duras como blandas. El objetivo es descubrir si el formato de enseñanza influye en los resultados. Como se explica a continuación, el tipo de prueba utilizada depende de si las series de datos son normales o no.

Resultados

Pregunta de investigación 1 (PI1): Rendimiento académico (habilidades duras)

¿Mejora una estrategia de enseñanza basada en FC o VT el rendimiento académico de los estudiantes medido en términos de calificaciones?

Este es el primer objetivo de nuestro estudio. Para evaluar el impacto de FC y VT en los resultados académicos de los estudiantes medidos en términos de calificaciones, consideramos:

- la media de las calificaciones académicas de los estudiantes de cada grupo,
- los resultados académicos obtenidos en un examen que los estudiantes realizaron al final de las primeras siete semanas, período en el que recibieron la clase a través de un formato de enseñanza semi-tradicional,
- los resultados obtenidos en un examen que se realizó al final de los dos formatos de enseñanza objeto de estudio, respectivamente, a través de FC y VT,
- y los resultados obtenidos en el examen final realizado al final de la clase. Este examen final se basó principalmente en preguntas destinadas a evaluar la capacidad de los estudiantes para retener los contenidos de la clase.¹

Debido a que algunas de estas series no eran normales², realizamos la prueba no paramétrica de Krusal-Wallis para comparar los resultados académicos obtenidos por los estudiantes de un grupo con los obtenidos por los estudiantes del otro grupo. Además, para analizar la evolución de las calificaciones de los alumnos a lo largo del estudio, realizamos una prueba de Wilcoxon. Los resultados aparecen en la tabla 2.

Tabla 2. Comparación del rendimiento académico en los formatos de enseñanza FC y VT en términos de habilidades “duras” (calificaciones)

(1)	(2)	Valor p	Tamaño del efecto
Antecedentes académicos del grupo de VT	Antecedentes académicos del grupo de FC	0,072	0,328

¹ Los datos oscilan entre 0,00, que es la calificación más baja, y 10,00, que es la calificación más alta. La calificación mínima para aprobar un examen es 5,00.

² Prueba de normalidad.

(1)	(2)	Valor p	Tamaño del efecto
Examen parcial del grupo FC antes de la metodología FC	Examen parcial del grupo VT antes de la metodología VT	0,315	0,106
Examen parcial del grupo FC	Examen parcial del grupo VT	0,000**	0,576
Examen parcial del grupo FC	Calificación final del grupo FC	0,695	0,070
Examen parcial del grupo VT	Calificación final del grupo VT	0,000**	0,784
Calificación final del grupo VT	Calificación final del grupo de FC	0,013*	0,262
Antecedentes académicos del grupo de FC	Calificación final del grupo de FC	0,005	0,505
Antecedentes académicos del grupo de VT	Calificación final del grupo VT	0,361	0,119

Como muestra la tabla 2, nuestro análisis detecta una mejora significativa en el rendimiento académico de los estudiantes, medido en términos de calificaciones, cuando reciben la modalidad FC. La mejora de los estudiantes es el resultado del formato de enseñanza invertida utilizado. Esto se demuestra por el hecho de que sus calificaciones en el examen final de la primera sección (donde los dos grupos de estudiantes recibieron la clase en el mismo formato semitradicional) son estadísticamente similares, con un tamaño del efecto pequeño. Esto significa que la mejora en sus calificaciones, al recibir la clase en el formato invertido, puede atribuirse al formato de enseñanza.

En este sentido, nuestro análisis detecta una diferencia estadísticamente significativa en las calificaciones de los exámenes que se realizaron al final de los dos formatos de enseñanza y que se administraron, respectivamente, a través de FC y VT. En este examen, se evaluó a los estudiantes sobre los mismos contenidos, con la única diferencia de que dichos contenidos se habían

Prueba de normalidad de las habilidades "duras"							
	Kolmogorov-Smirnova			Shapiro-Wilk			
	Estadístico	gl	p valor	Estadístico	gl	p valor	
Examen parcial del grupo VT antes de la metodología VT	0,160	59	0,001**	0,880	59	0,000**	
Examen parcial del grupo VT	0,117	59	0,045*	0,947	59	0,013	
Calificación final del grupo VT	0,112	59	0,064	0,960	59	0,051	
Antecedentes académicos del grupo de VT	0,096	59	0,2	0,971	59	0,166	
Examen parcial del grupo FC	0,180	31	0,012	0,932	31	0,048*	
Antecedentes académicos del grupo de FC	0,096	31	0,2	0,969	31	0,496	
Calificación final del grupo de FC	0,108	31	0,2	0,942	31	0,094	
Examen parcial del grupo FC antes de la metodología FC	0,148	31	0,082	0,956	31	0,235	

explicado a través de FC a un grupo y de VT al otro grupo. También en este caso, los estudiantes que recibieron la clase en formato invertido obtuvieron mejores calificaciones que los estudiantes que recibieron la clase en formato visual. Como se puede observar, el tamaño del efecto es relativamente grande. Esto supone que no solo hay una diferencia en las calificaciones, sino que además la diferencia es significativa.

Por último, esta mejora también se observa al analizar las calificaciones finales de la clase, que son la media de todas las calificaciones obtenidas a lo largo de todas las secciones de la clase. Sin embargo, como se puede observar, la mejora es relativamente pequeña.

A partir de estos resultados, se puede afirmar que el formato FC fue más eficaz que el formato VT para mejorar las calificaciones de los estudiantes.

Pregunta de investigación 2 (PI2): Habilidades blandas

¿Una estrategia de enseñanza basada en FC o VT mejora el rendimiento académico de los estudiantes en términos de habilidades blandas?

Este es el segundo objetivo de nuestro estudio. Para evaluar el impacto de la FC y la VT en las habilidades blandas de los estudiantes, operacionalizadas en términos de una encuesta basada en diez variables relacionadas con la autoeficacia, seis variables relacionadas con el trabajo en equipo y una variable relacionada con la percepción del aprendizaje (véase la tabla 3), realizamos dos pruebas en cada grupo, una al principio y otra al final de la clase, para averiguar si se producía algún cambio en cada una de las variables consideradas.

Tabla 3. Encuesta sobre habilidades blandas³

	Autoeficacia	Totalmente en desacuerdo	Totalmente de acuerdo
X1	Creo que este año voy a sacar muy buenas notas.		
X2	Si me esfuerzo, creo que tengo la capacidad suficiente para obtener un buen expediente académico.		
X3	Creo que soy capaz de comprender incluso los temas más difíciles de este curso.		
X4	Creo que tengo la capacidad suficiente para comprender una materia de forma rápida y adecuada.		
X5	Creo que puedo aprobar los cursos con bastante facilidad e incluso obtener buenas calificaciones.		
X6	Aunque los profesores son exigentes y estrictos, tengo mucha confianza en mi propia capacidad académica.		

³ El Instituto de Educación Científica de la universidad donde se llevó a cabo el estudio proporcionó el cuestionario. Se basó en preguntas establecidas que este instituto había utilizado anteriormente. Se realizó una prueba previa con cinco estudiantes para validar su redacción y comprensión.

Autoeficacia		Totalmente en desacuerdo	Totalmente de acuerdo
X7	Creo que estoy preparado y bien cualificado para alcanzar el éxito académico.		
X8	Cuando me piden que haga proyectos o deberes, estoy seguro de que los haré bien.		
X9	Trabajo eficazmente en cualquier equipo, sin importar quiénes sean mis compañeros.		
X10	Teniendo en cuenta la dificultad de la carrera, lo que estoy aprendiendo y mis propias capacidades, creo que me irá bien cuando termine (la carrera).		
Trabajo en equipo⁴		Percepción muy negativa	Percepción muy positiva
X11	Participación en el trabajo en equipo compartiendo información, conocimientos y experiencias.		
X12	Aceptación y cumplimiento de las normas acordadas en el grupo (plazos, partes del trabajo, formato, etc.).		
X13	Acción para afrontar los conflictos del equipo en esta materia.		
X14	Compromiso con la gestión y el funcionamiento del equipo.		
X15	Gestión eficaz de las reuniones.		
X16	Comunicación y cohesión dentro del grupo.		
Percepción del aprendizaje		Poca	Mucho
X17	Independientemente de tus resultados en los exámenes, piensa en cuánto aprenderás en esta asignatura sobre Sistemas Políticos Comparados.		

Tras observar este cambio, definido como «C» para cada habilidad social, analizamos si dicho cambio era el mismo en el caso de utilizar el FC o el VT. De este modo, comparamos el cambio observado en cada una de las habilidades blandas en el formato de enseñanza basado en el FC con los cambios observados en cada una de las habilidades blandas en el formato de enseñanza basado en el VT. Debido a que no todas las series eran normales⁵, con el fin de homogeneizar el análisis, utilizamos

⁴ Antes del estudio, los estudiantes ya habían asistido a una clase obligatoria del diploma en «Desarrollo de habilidades personales, comunicativas y profesionales», impartida por la universidad donde se llevó a cabo el estudio. Esta clase se denominaba «Trabajo en equipo». En ella se les proporcionó una percepción inicial de su capacidad para trabajar en equipo. Esta percepción inicial se midió al comienzo del estudio. Presupusimos que esta percepción era superior a cero, debido a la asistencia previa de los estudiantes a la clase de «Trabajo en equipo». Al finalizar el estudio, volvimos a medir esta percepción, considerando si había mejorado o empeorado en comparación con la inicial.

⁵ Prueba de normalidad, donde V indica que el cambio se calcula dentro del formato VT y F se calcula dentro del formato FC.

técnicas de análisis no paramétricas que son menos restrictivas, como la prueba de la mediana, la prueba U de Mann-Whitney y la prueba de Kruskal-Wallis. Los resultados aparecen en la tabla 4.

Prueba de normalidad de las habilidades blandas						
	Kolmogorov-Smirnova			Shapiro-Wilk		
	Estadístico	gl	p valor	Estadístico	gl	p valor
C1V	0,323	31	0,000**	0,799	31	0,000**
C2V	0,257	31	0,000**	0,841	31	0,000**
C3V	0,203	31	0,002**	0,922	31	0,027**
C4V	0,303	31	0,000**	0,834	31	0,000**
C5V	0,302	31	0,000**	0,852	31	0,001**
C6V	0,286	31	0,000**	0,847	31	0,000**
C7V	0,284	31	0,000**	0,847	31	0,000**
C8V	0,277	31	0,000**	0,849	31	0,000**
C9V	0,264	31	0,000**	0,857	31	0,001**
C10V	0,228	31	0,000**	0,900	31	0,007**
C11V	0,238	31	0,000**	0,852	31	0,001**
C12V	0,281	31	0,000**	0,875	31	0,002**
C13V	0,321	31	0,000**	0,829	31	0,000**
C14V	0,291	31	0,000**	0,835	31	0,000**
C15V	0,308	31	0,000**	0,824	31	0,000**
C16V	0,259	31	0,000**	0,855	31	0,001**
C17V	0,221	31	0,001**	0,879	31	0,002**
C1F	0,370	31	0,000**	0,731	31	0,000**
C2F	0,323	31	0,000**	0,827	31	0,000**
C3F	0,297	31	0,000**	0,848	31	0,000**
C4F	0,371	31	0,000**	0,718	31	0,000**
C5F	0,266	31	0,000**	0,846	31	0,000**
C6F	0,227	31	0,000**	0,870	31	0,001**
C7F	0,261	31	0,000**	0,886	31	0,003**
C8F	0,184	31	0,009**	0,932	31	0,051
C9F	0,339	31	0,000**	0,742	31	0,000**
C10F	0,292	31	0,000**	0,880	31	0,002**
C11F	0,328	31	0,000**	0,765	31	0,000**
C12F	0,394	31	0,000**	0,729	31	0,000**
C13F	0,210	31	0,001**	0,937	31	0,067
C14F	0,310	31	0,000**	0,834	31	0,000**
C15F	0,278	31	0,000**	0,836	31	0,000**
C16F	0,258	31	0,000**	0,854	31	0,001**
C17F	0,268	31	0,000**	0,859	31	0,001**

Tabla 4. Comparación de los cambios en las habilidades blandas entre los formatos de enseñanza VT y FC

Variable	Prueba de la mediana	Prueba U de Mann-Whitney	Prueba de Kruskal-Wallis
C1	0,169	0,174	0,174
C2	0,538	0,329	0,329
C3	0,057	0,073	0,073
C4	0,849	0,618	0,618
C5	0,311	0,244	0,244
C6	0,653	0,735	0,735
C7	0,834	0,476	0,476
C8	0,744	0,866	0,866
C9	0,815	0,995	0,995
C10	0,412	0,249	0,249
C11	0,255	0,347	0,347
C12	0,466	0,861	0,861
C13	0,417	0,101	0,101
C14	0,849	0,981	0,981
C15	0,972	0,428	0,428
C16	0,925	0,699	0,699
C17	<,001***	0,002***	0,002***

Como muestra la tabla 4, a diferencia de lo que ocurre con las calificaciones, nuestro análisis no detecta ninguna mejora en las habilidades blandas de los estudiantes, ni en la sección impartida a través de FC ni en la impartida a través de VT. No se observan diferencias estadísticamente significativas en ninguno de los dos formatos. La única excepción es la diferencia que se observa en términos de percepción del aprendizaje, entendida como la forma en que los estudiantes esperan que el formato de enseñanza pueda afectar a sus calificaciones. Los que recibieron la clase a través del formato invertido tienden a percibir que van a obtener mejores calificaciones, en comparación con las percepciones de los que recibieron la clase en el formato visual.

Discusión

Nuestro estudio corrobora otros anteriores que detectaron un impacto positivo de la FC tanto en las calificaciones de los estudiantes como en la percepción del aprendizaje (Whitman Cobb, 2016; Van der Zwan y

Afonso, 2019). Además, al detectar una falta de mejora similar entre los estudiantes a los que se les impartió la clase a través de VT, nuestro estudio va en una dirección similar a los que expresaron prudencia a la hora de adaptar clases completas o secciones de ellas en torno a materiales visuales (Lieberfeld, 2007, p. 573; Inoue y Krain, 2014, p. 20). El formato FC, que hace hincapié en la posibilidad de ver las videoconferencias en casa y discutir las dudas y los aspectos poco claros de la clase en el aula con el profesor y los compañeros, resultó útil para que los estudiantes consolidaran su comprensión de conceptos complejos y propusieran ideas sobre cómo aplicarlos. No se puede decir lo mismo del formato basado en VT.

Sin embargo, lejos de indicar una falta de utilidad en los formatos VT, esta diferencia en las calificaciones y la percepción del aprendizaje de los estudiantes puede indicar que los diferentes formatos de enseñanza activa pueden ser útiles de manera diferente según el objetivo o la necesidad de la enseñanza. Hay varias razones que pueden explicar por qué el formato FC resultó más eficaz que el VT para las calificaciones de los estudiantes.

En primer lugar, aunque como técnica docente el FC es sin duda más reciente (Bergmann y Sams, 2012) que las proyecciones de películas y documentales en el aula, que se utilizan desde hace mucho tiempo en diversas ciencias sociales (Rose, 2016), el primero se basa en una tradición más amplia en términos de estudio sobre su diseño, implementación y evaluación empírica. Esto ha dado lugar a una amplia bibliografía que tiene como objetivo comprobar empíricamente sus puntos fuertes y sus limitaciones en cuanto a la mejora del rendimiento de los estudiantes, tanto en lo que se refiere a las calificaciones como a las habilidades blandas. Aunque probablemente tenga una historia más larga, el uso de películas y documentales aún no ha dado lugar a la misma producción en términos de análisis empíricos de su eficacia para el aprendizaje de los estudiantes.

En segundo lugar, el FC solo puede implementarse mediante una reorganización integral del tiempo y el espacio de clase. Esto se debe a sus características, basadas en eliminar las clases magistrales presenciales y sustituirlas por videoconferencias que se ven fuera de clase y ejercicios y aplicaciones que se realizan en clase bajo la supervisión del profesor (Jenkins 2015, 607; Berge y Nederveld 2015, 163). Esto implica preparar contenidos que deben ponerse a disposición de los alumnos antes de la clase y diseñar ejercicios y actividades para la clase basados en los contenidos que se supone que los alumnos han aprendido con las videoconferencias. De esta manera, el tiempo de clase puede utilizarse de forma valiosa para resolver las dudas de los alumnos y ayudarles a dominar las aplicaciones de lo que han aprendido. En este sentido, el FC puede utili-

zar o integrar una variedad de técnicas de enseñanza, como presentaciones, juegos, simulaciones y materiales visuales. Esto tiende a convertir al FC en una metodología de enseñanza integral, y no solo en un conjunto de materiales que se pueden utilizar en el aula. Por otra parte, aunque las películas y los documentales se utilizan desde hace mucho tiempo en las aulas, aún no se reconocen como un formato de enseñanza independiente y autosuficiente. Probablemente deberían considerarse materiales de referencia para complementar y apoyar la enseñanza, no solo en su formato tradicional de clase magistral, sino también en cualquier otro formato que tenga como objetivo hacer que el aprendizaje de los alumnos sea más activo y atractivo, incluido el FC. Cuando se reorganiza toda una clase o una parte de ella en términos de visionado de películas y documentales para los alumnos, como en nuestro estudio, esto no tiene por qué mejorar necesariamente la comprensión y los conocimientos de los alumnos (Kiasatpour, 1999, p. 85; Lieberfeld, 2007, p. 573; Inoue y Krain, 2014, p. 20; Swimelar, 2013, p. 24; Brandle, 2020). Esta conclusión parece mantenerse cuando se evalúa el rendimiento de los estudiantes mediante una medida relativamente más objetiva, es decir, sus calificaciones, y otra más subjetiva, es decir, su percepción del aprendizaje. Análisis previos basados en la VT mostraron que las películas y los documentales pueden obstaculizar la percepción del aprendizaje de los estudiantes, ya que no siempre logran transmitirles la impresión de que las representaciones ficticias son útiles para dominar conceptos académicos complejos (Betti et al. 2024; Lieberfeld, 2007, p. 573; Inoue y Krain, 2014, p. 20; Swimelar, 2013, p. 24; Brandle, 2020).

Varios estudios han advertido que hay que tener cuidado al tomar al pie de la letra las percepciones de los estudiantes sobre la eficacia de los formatos de enseñanza (Deslauriers et al. 2019). A veces, a los estudiantes les puede resultar difícil evaluar qué y cuánto han aprendido de una materia. Sus percepciones podrían estar relacionadas con «una serie de factores diferentes», incluida la «novedad», que no son necesariamente relevantes en términos pedagógicos (McNally et al. 2017, p. 292). No obstante, cuando se trata del aprendizaje, las percepciones de los estudiantes sobre lo que aprenden y lo que esperan aprender son importantes. Las percepciones pueden ser engañosas, especialmente cuando no van acompañadas de alguna medida objetiva de lo que los estudiantes han aprendido efectivamente. Sin embargo, esto no es motivo suficiente para no incluirlas en un análisis empírico sobre el impacto del aprendizaje activo en su rendimiento. Si los estudiantes son los principales actores del proceso de aprendizaje, sus percepciones son una fuente de datos relevante para comprender qué formatos pueden ser eficaces para mejorar

el aprendizaje. Mejorar la percepción del aprendizaje puede ser, por lo tanto, muy útil en un proceso de enseñanza, ya que puede tener consecuencias positivas en la motivación y el compromiso de los estudiantes.

Con todo esto, este estudio no sugiere que el FC sea necesariamente mejor que el VT para el rendimiento académico de los estudiantes. Por ejemplo, ni el VT ni el FC mejoraron otras habilidades blandas importantes de los estudiantes, como la autoeficacia y el trabajo en equipo. Este resultado difiere de los estudios que observaron una mejora en el rendimiento de los estudiantes en este tipo de habilidades (Whitman Cobb, 2016) y coincide con otros que no encontraron ninguna diferencia significativa entre el uso de un formato de enseñanza u otro en términos de habilidades blandas de los estudiantes (Jenkins 2015, p. 610).

Estos resultados nos llevan más bien a la conclusión de que, en lugar de considerar la FC y la VT, o cualquier otro formato de enseñanza de aprendizaje activo, como mutuamente excluyentes o incompatibles, los futuros enfoques pedagógicos deberían basarse en una integración complementaria de diferentes estrategias, técnicas y materiales didácticos. En lugar de hacer que el aprendizaje de los estudiantes dependa de una única innovación docente, los profesores deben estar preparados para combinar diferentes técnicas, en función de los objetivos y las características de cada clase (Jenkins 2015, p. 610; Betti et al. 2020; Betti et al., 2022). Los diferentes formatos de enseñanza pueden ofrecer diferentes ventajas y beneficios. Esto puede fomentar procesos de aprendizaje distintos, pero igualmente significativos. Por un lado, en un formato FC, los estudiantes pueden ver el contenido de las clases antes del comienzo de las mismas. Basándose en esos contenidos, deben realizar ejercicios en clase, bajo la supervisión del profesor y en colaboración con sus compañeros. Este proceso puede dar lugar a un diálogo constructivo y mutuamente enriquecedor entre los estudiantes y entre estos y el profesor, capaz de generar un aprendizaje más profundo y duradero. Por otro lado, en un formato VT, los estudiantes deben estudiar fenómenos políticos y sociales a través de materiales visuales que deben analizar antes de la clase y a través de contenidos visuales que deben producir en casa y presentar durante la misma. Esto también puede crear las condiciones para un diálogo iterativo entre ellos y con el profesor. Ambos formatos pueden, por lo tanto, proporcionar un contexto para procesos de aprendizaje diferentes pero complementarios, ambos potencialmente capaces de producir un aprendizaje significativo.

Además, ningún formato de enseñanza activa debe pretender sustituir los elementos tradicionales de la enseñanza, como las clases magistrales de los profesores o los apuntes de los alumnos. Por un lado, hay momen-

tos en una clase en los que la complejidad de los conceptos puede requerir una clase magistral. Esto puede permitir a los profesores evaluar con mayor precisión las respuestas de los alumnos a las explicaciones de conceptos específicos. Por otro lado, cuando es necesario realizar ejercicios para aplicar conceptos o resolver problemas sociales y políticos, puede resultar más práctico pedir a los alumnos que estudien esos conceptos viendo en casa una explicación pregrabada del profesor y que luego acudan a clase con hipótesis e ideas que puedan utilizar en esos ejercicios.

Sin ignorar las dificultades para lograrlo, lo ideal sería que un aula pudiera integrar diferentes estrategias, técnicas y materiales didácticos a lo largo del mismo año o semestre (Betti et al. 2024; Betti et al. 2022). Dependiendo de la necesidad, esto debería incluir clases magistrales tradicionales, videoconferencias pregrabadas, ejercicios y actividades en clase, como debates, presentaciones y simulaciones, exámenes y tareas, y una combinación de materiales visuales y no visuales, como libros de texto, apuntes, imágenes, películas y documentales. En este sentido, creemos que no puede haber un formato «único» que satisfaga todos los contextos y necesidades docentes.

Conclusiones, limitaciones e investigaciones futuras

Si bien la clase invertida ha sido objeto de muchos estudios destinados a evaluar empíricamente sus efectos en el aprendizaje de los estudiantes, lo que hay hasta ahora sobre pensamiento visual son principalmente descripciones de su aplicación en la enseñanza. El número de estudios empíricos aplicados al uso de películas y documentales en CP o RI sigue siendo menor que el que se puede encontrar sobre la clase invertida. Se necesitan más análisis para comprender mejor los aspectos cognitivos del rendimiento de los estudiantes en los que la VT podría ser útil. El objetivo de este estudio es contribuir a esta literatura. Nuestra conclusión es que el FC debe considerarse un formato de enseñanza más completo y exhaustivo que puede ayudar a reestructurar el tiempo y el espacio de clase de manera que se maximicen los beneficios de la enseñanza y el aprendizaje. La VT a través de películas y documentales debe considerarse más bien como un recurso didáctico que puede proporcionar a los profesores una gran variedad de materiales visuales para utilizar en clase, con el fin de complementar las explicaciones del contenido o gestionar actividades evaluadas, como comentarios, debates o pequeños proyectos de investigación, que se realizarán en clase o en casa. A su vez, somos conscientes de las limitaciones de este estudio: se llevó a cabo durante un periodo de tiempo limitado, de solo siete semanas; además,

en él participaron estudiantes con características contextuales específicas, como estar matriculados en una institución privada y proceder de un entorno socioeconómico similar de clase media-alta. Somos conscientes, por tanto, de la necesidad de ampliar el estudio con muestras más grandes a lo largo de períodos de tiempo más largos y de incluir entornos socioeconómicos seleccionados de forma más aleatoria. Sin embargo, también creemos que este estudio proporciona un valioso punto de partida para evaluar comparativamente la utilidad y la eficacia de diferentes formatos de enseñanza basados en el aprendizaje activo.

Los estudios futuros deberán aportar más datos para corroborar esta conclusión. Por ejemplo, deberán comparar más exhaustivamente las diferentes estrategias de enseñanza para mejorar nuestra comprensión de los mejores aspectos y prácticas que se pueden extraer de cada una de ellas, con el objetivo de integrarlas de forma productiva en la enseñanza. Además, deberán tratar de comprender las soluciones más eficaces para evaluar el aprendizaje de los estudiantes en función del tipo de formato o técnica de enseñanza aplicada. Por último, deben seguir desarrollando estrategias de investigación para incluir de forma fiable las percepciones de los estudiantes sobre el aprendizaje en el proceso docente. Esto podría implicar el uso de entrevistas, encuestas o grupos de discusión. Esto sería útil para comprender cómo se pueden adaptar los diferentes formatos de enseñanza a las necesidades de aprendizaje de los diferentes estudiantes. En este sentido, también sería importante examinar las percepciones de los profesores para comprender cómo los diferentes estilos de enseñanza pueden adaptarse a diferentes formatos basados en el aprendizaje activo. Debido a las constantes transformaciones que caracterizan los ámbitos de la enseñanza y el aprendizaje, la experimentación con aplicaciones y evaluaciones empíricas de las clases basadas en FC, VT y otras estrategias de enseñanza potencialmente innovadoras sigue siendo la forma más eficaz de enriquecer nuestra comprensión sobre cómo mejorar el aprendizaje de los alumnos.

En conclusión, los estudios futuros no solo deberían mejorar nuestra comprensión empírica de la eficacia de los formatos de enseñanza basados en el aprendizaje activo, sino también incluir en el debate sus principales aspectos pedagógicos. Los análisis futuros deberían, por ejemplo, considerar la importancia de la inclusividad en la enseñanza. Esto implica evaluar si los formatos de enseñanza pueden garantizar la igualdad de acceso a la tecnología, reducir el impacto de las desigualdades socioeconómicas y tener en cuenta la diversidad de las necesidades de aprendizaje. Esta es la única manera de garantizar que los formatos de enseñanza puedan mejorar los perfiles de los diferentes estudiantes. Esto

también requiere que las universidades dediquen los recursos financieros y humanos adecuados para implementar de manera eficaz formatos de enseñanza innovadores, de modo que los profesores puedan adquirir las competencias necesarias para integrarlos en sus clases. En este sentido, será esencial examinar constantemente las percepciones de los profesores y los estudiantes para identificar posibles problemas y diseñar soluciones. Para alcanzar estos objetivos, las universidades deberán estimular debates abiertos e informados basados en asociaciones y redes de conocimiento formadas por profesores, educadores, estudiantes, familias e instituciones públicas y privadas. Estas redes deberán utilizar los formatos de enseñanza, los recursos multimedia y las herramientas digitales más adecuados para fomentar el aprendizaje de los estudiantes.

Materiales complementarios

Los materiales complementarios para el análisis cuantitativo, junto con la lista completa de películas y documentales utilizados para el estudio, no se incluyen por razones de espacio, pero pueden facilitarse previa solicitud.

Normas éticas

El estudio recibió la autorización del Comité de Evaluación de la Innovación Docente de la Universidad donde se llevó a cabo. El Comité concedió fondos para el proyecto «Implementación y evaluación de estrategias de pensamiento visual en la doble titulación en Relaciones Internacionales y Comunicaciones Globales», para el período académico 2021-2023. Los estudiantes aceptaron participar en el estudio rellenando y firmando un formulario. Durante el proceso, recibieron información completa sobre los objetivos de la investigación. Sus datos personales no se compartieron con ninguna persona o institución. Su participación fue totalmente voluntaria.

Referencias

- Anderson, L. W., y Krathwohl, D. (2001). *A Taxonomy for Learning, Teaching, and Assessing. A Revision of Bloom's Taxonomy of Educational Objectives*. New York: Longman.
- Baars, M., y Wijnia, L. (2018). The relation between task-specific motivational profiles of self-regulated learning skills. *Learning and Individual Differences*, (64), 125-137. <https://doi.org/10.1016/j.lindif.2018.05.007>.

- Bandura, A. (1994). Self-efficacy. In R. Vylayanur (ed.), *Encyclopaedia of Human Behaviour* (pp. 71-81). London: Academic Press.
- Berge, Z. L. y Nederveld, A. (2015). Flipped Learning in the Workplace. *Journal of Workplace Learning*, 27(2), 162-172. <https://doi.org/10.1108/JWL-06-2014-0044>.
- Bergmann, J. y Sams, A. (2012). *Flip Your Classroom: Reach Every Student in Every Class Every Day*. Eugene: ISTE.
- Betti, A., Biderbost, P. y García Domonte, A. (2020). Flipping the Classroom in Political Science: Student Achievement and Perceptions. *Revista de Ciencia Política*, 40(3), 589-615. <https://doi.org/10.4067/S0718-090X2020005000102>.
- (2022). Developing Students' "Soft Skills" through the Flipped Classroom: Evidence from an International Studies Class. *International Studies Perspectives*, 23(1), 1-24. <https://doi.org/10.1093/isp/ekab014>.
- Betti, A., Biderbost, P., Vaquero E. (2024). Is Visual Thinking Useful for Teaching International Relations? Evidence from an Implementation through Movies and Documentaries in an Undergraduate Class. In Seda Bayrakdar / P Seda Bayrakdar / Patrycja Chodnicka-Jaworska (eds.). *Current Dynamics in Business and Economics: Theoretical Approaches and Empirical Discoveries* (pp. 351-360). Berlin: Peter Lang.
- Bligh, D.A. (1998). *What's the Use of Lectures?* Bristol: Intellect.
- Brandle, S. M. (2020). Games, Movies, and Zombies: Making IR Fun for Everyone. *Journal of Political Science Education*, 16(4), 459-78. <https://doi.org/10.1080/15512169.2019.1568880>.
- Cheng, L., Ritzhaupt, A.D., y Antonenko, P. (2018). Effects of the flipped classroom instructional strategy on students' learning outcomes: a meta-analysis. *Education Technology Research Development*, <https://doi.org/10.1007/s11423-018-9633-7>.
- Collins, A., y Halverson, R. (2010). The Second Education Revolution: rethinking education in the age of technology. *Journal of Computer Assisted Learning*, 26, 18-27. <https://doi.org/10.1111/j.1365-2729.2009.00339.x>
- Dean, J. (2019). Sorted for Memes and Gifs: Visual Media and Everyday Digital Politics. *Political Studies Review*, 17(3), 255-66. <https://doi.org/10.1177/1478929918807483>.
- Deslauriers, L., Logan, S., McCarthy, K.M., Callaghan, K., y Kestin, G. (2019). "Measuring actual learning versus feeling of learning in response to being actively engaged in the classroom." *PNAS* 116(39): 19251-19257. <https://doi.org/10.1073/pnas.1821936116>.

- Galindo-Domínguez, H. (2021). Flipped Classroom in the Educational System: Trend or Effective Pedagogical Model Compared to Other Methodologies? *Educational Technology y Society*, 24 (3), 44–60.
- Giglio, E. (2002). Using Film to Teach Political Concepts. *European Political Science*, 1(2), 53-8. <https://doi.org/10.1057/eps.2002.10>.
- Gokcek, G., y Howard, A. (2013). Movies to the Rescue: Keeping the Cold War Relevant for Twenty-First Century Students. *Journal of Political Science Education*, 9(4), 436-52. <https://doi.org/10.1080/15512169.2013.835561>.
- Heck, A. (2017). Analyzing Docudramas in International Relations: Narratives in the Film *A Murderous Decision*. *International Studies Perspectives*, 18(4), 365-90. <https://doi.org/10.1093/isp/ekw012>.
- Hussain, M.I., Preetha, R., Naik, M.S., Panaskar, H.C., y Das, A.D. (2023). Assessing the Effectiveness of Flipped Classroom Strategy on Student Performance. *European Chemistry Bulletin*, 12, 2883-2896. <https://doi.org/10.1186/s12909-024-06105-7>.
- Inoue, C. Y. A. y Krain, M. (2014). One World, Two Classrooms, *Thirteen Days*: Film as an Active Teaching and Learning Tool in Cross-national Perspective. *Journal of Political Science Education*, 10(4), 424-442. <https://doi.org/10.1080/15512169.2014.947422>.
- Iretzberger, M. (2021). Film and its Unused Potential for Teaching International Relations: Exploring New Possibilities in a Classroom Experiment. *Journal of Political Science Education*, 17(1), 567-82. <https://doi.org/10.1080/15512169.2020.1854774>.
- Jenkins, S. (2015). "Flipping the Introductory American Politics Class: Student Perceptions of the Flipped Classroom." *PS: Political Science and Politics*, 48(4), 607-611. <https://doi.org/10.1017/S1049096515000840>.
- Jensen, J.L., Tyler, A.K., y Godoy, P. (2015). Improvements from a Flipped Classroom May Simply be the Fruits of Active Learning. *Life Sciences Education*, 14(1), 1-12. <https://doi.org/10.1187/cbe.14-08-0129>.
- Kiasatpour, S. M. (1999). The Internet and Films: Teaching Middle East Politics Interactively. *PS: Political Science and Politics*, 32(1), 83-9. <https://doi.org/10.2307/420755>.
- Kuzma, L. M., y Haney, P. J. (2001). And...Action! Using Films to Learn about Foreign Policy. *International Studies Perspectives*, 2(1), 33-50. <https://doi.org/10.1111/1528-3577.00036>.
- (2002). Comments by Kuzma and Haney: Using Film in the Classroom, *International Studies Perspectives*, 3, 92-4. <https://doi.org/10.1111/1528-3577.00081>.
- Lai, C.L., y Hwang, G.J. (2016). A Self-Regulated Flipped Classroom Approach to Improving Students' Learning Performance in a Mathe-

- maths Course. *Computers and Education*, 100, 126-140. <https://doi.org/10.1016/j.compedu.2016.05.006>.
- Lambach, D., Karger, C., y Goerres, A. (2017). Inverting the Large Lecture Class: Active Learning in an Introductory International Relations Course. *European Political Science*, 16(4), 553-569. <https://doi.org/10.1057/s41304-016-0078-3>.
- Lieberfeld, D. (2007). Teaching about War through Film and Literature. *PS: Political Science and Politics*, 40(3), 571-4. <https://doi.org/10.1017/S1049096507070837>.
- Lin, Y-N., Hsia, L-H., Sung, M-Y., y Hwang, G-H. (2019). Effects of integrating mobile technology-assisted peer assessment into flipped learning on students' dance skills and self-efficacy. *Interactive Learning Environments*, 27(8), 995-1010. <https://doi.org/10.1080/10494820.2018.1461115>.
- McNally, B., Chipperfield, J., Dorsett, P., Del Fabbro, L., Frommolt, V., Goetz, S., Lewohl, J., Molineux, M., Pearson, A., Reddan, G., Roiko, A., y Rung, A. (2017). Flipped Classroom Experiences: Student Preferences and Flip Strategy in a Higher Education Context. *Higher Education* (73), 281-298. <https://doi.org/10.1007/s10734-016-0014-z>.
- Naing C., Whittaker, M.A., Aung, H.H., Chellapan, D.K., y Riegelman, A. (2023). The effects of flipped classrooms to improve learning outcomes in undergraduate health professional education: A systematic review. *Campbell Systematic Reviews*, <https://doi.org/10.1002/cl2.1339>.
- Omelicheva, M. Y., y Avdeyeva, O. (2008). Teaching with Lecture or Debate? Testing the Effectiveness of Traditional versus Active Learning Methods of Instruction. *PS: Political Science and Politics (July)*, 603-7. <https://doi.org/10.1017/S1049096508080815>.
- Roberts, D. (2017). The Message is in the Medium: Evaluating the Use of Visual Images to Provoke Engagement and Active Learning in Politics and International Relations Lectures. *Politics*, 38(2), 232-249. <https://doi.org/10.1177/0263395717717229>.
- Robles, M. M. (2012). Executive Perceptions of the Top 10 Soft Skills Needed in Today's Workplace. *Business Communication Quarterly*, 75(4), 453-465. <https://doi.org/10.1177/1080569912460400>.
- Rose, G. (2016). *Visual Methodologies: An Introduction to Researching with Visual Materials*, 4th Edition. London: Sage.
- Salas, E., Sims, D.E., y Burke, S.C. (2005). Is there a Big Five in Teamwork? *Small Group Research*, 36(5), 555-99. <https://doi.org/10.1177/1046496405277134>.
- Schmid, J. (2020). (Captain) America in crisis: popular digital culture and the negotiation of Americanness, *Cambridge Review of International*

- Affairs*, 33(5), 690-712. <https://doi.org/10.1080/09557571.2019.1708864>.
- Stoddard, J. D. (2012). Film as a “thoughtful” medium for teaching history. *Learning, Media, and Technology*, 37(3), 271-88. <https://doi.org/10.1080/17439884.2011.572976>.
- Strelan, P., Osborn, A., y Palmer, E. (2020). The Flipped Classroom: A Meta Analysis of effects on student performance across disciplines and education levels. *Educational Research Review* 30. <https://doi.org/10.1016/j.edurev.2020.100314>.
- Sunderland, S., Rothermel, J.C. y Lusk, A. (2009). Making Movies Active: Lessons from Simulations. *PS: Political Science and Politics*, 42(3), 543-47. <https://doi.org/10.1017/S1049096509090878>.
- Swimelar, S. (2013). Visualizing International Relations: Assessing Student Learning through Film. *International Studies Perspectives*, 14(1), 14-38. <https://doi.org/10.1111/j.1528-3585.2012.00467.x>.
- Talbert, R., y Bergmann, J. (2017). *Flipped Learning: A Guide for Higher Education Faculty*. London: Routledge.
- Tisham, S., MacGillivray, D., y Palmer, P. (1999). Investigating the Educational Impact and Potential of the Museum of Modern Art's Visual Thinking Curriculum, Final Report, Harvard Project Zero.
- Touchton, M. (2015). Flipping the Classroom and Student Performance in Advanced Statistics: Evidence from a Quasi-Experiment. *Journal of Political Science Education*, 11(1), 28-44. <https://doi.org/10.1080/15512169.2014.985105>.
- Valeriano, B. (2013). Teaching Introduction to International Politics with Film. *Journal of Political Science Education*, 9(1), 52-72. <https://doi.org/10.1080/15512169.2013.747840>.
- Valeriano, B., y Habel, P. (2016). Who are the Enemies? The Visual Framing of Enemies in Digital Games. *International Studies Review*, 18. <https://doi.org/10.1093/isr/viv007>.
- Van der Zwan, N., y Afonso, A. (2019). Activating the Research Methods Curriculum: A Blended Flipped Classroom. *PS: Political Science and Politics*, 52(4), 749-753. <https://doi.org/10.1017/S1049096519000581>.
- Van Munster, R., y Sylvest, C. (2015). *Documenting World Politics: A Critical Companion to IR and non-fiction film*. London: Routledge.
- Wagner, D-A. (2018). Teachers' Use of Film in the History Classroom: A Survey of 19 High School Teachers in Norway. *Nordidactica*, www.kau.se/nordidactica.

- Weber, C. (2001). The Highs and Lows of Teaching IR Theory: Using Popular Films for Theoretical Critique. *International Studies Perspectives*, 2(3), 281-87. <https://doi.org/10.1111/1528-3577.00058>.
- Whitman Cobb, W. N. (2016). Turning the Classroom Upside Down: Experimenting with the Flipped Classroom in American Government. *Journal of Political Science Education*, 12(1), 1-14. <https://doi.org/10.1080/15512169.2015.1063437>.
- Yen, S-C., Lo, Y., Lee, A., y Enriquez, J. (2018). Learning online, offline, and in-between: comparing student academic outcomes and course satisfaction in face-to-face, online, and blended teaching modalities. *Education and Information Technologies*, 23, 2141-2153. <https://doi.org/10.1007/s10639-018-9707-5>.

RESEÑAS

Natallia Kabiak (2025). *Aprender ruso a través del cortometraje contemporáneo*. Londres: Routledge. 134 pp. ISBN: 9781032739380

En los últimos años, la integración de medios auténticos en la pedagogía de lenguas ha transformado los enfoques del aprendizaje de lenguas extranjeras. *Learn Russian through Contemporary Short Film* (Routledge, 2025), de Natallia Kabiak, contribuye a este panorama en evolución al ofrecer un marco innovador para la enseñanza del ruso como lengua extranjera (RLE) a través del medio del cortometraje contemporáneo. El manual, diseñado para estudiantes de los niveles B1–C1 del MCER, combina rigor lingüístico, exploración cultural y alfabetización mediática dentro de una estructura coherente y adaptable.

Kabiak, profesora de Estudios Rusos en la Universidad de Melbourne, ha publicado previamente manuales que priorizan materiales culturales auténticos. En esta obra más reciente, amplía este enfoque incorporando doce cortometrajes rusos y kazajos (2014–2022) al aula. Estas películas —que van desde *Собеседование* (The Interview) hasta *Варя* (Varya)— funcionan como catalizadores pedagógicos para el desarrollo tanto de la competencia comunicativa como de la comprensión intercultural. El resultado es un manual que no solo mejora la competencia lingüística, sino que también estimula la empatía, el pensamiento crítico y la apreciación de la sociedad rusa contemporánea.

El libro está cuidadosamente organizado en doce módulos independientes, cada uno centrado en un cortometraje. Cada módulo sigue una progresión constante diseñada para andamiar el aprendizaje lingüístico:

- Las actividades previas al visionado introducen vocabulario contextual, conceptos culturales y debates predictivos.
- Las tareas de visionado implican al alumnado con estímulos audiovisuales.
- Los ejercicios posteriores al visionado desarrollan destrezas receptoras y productivas mediante preguntas de comprensión, tareas de gramática y formación de palabras, y debates estructurados.
- Las tareas de escritura y las pruebas de comprensión amplían el aprendizaje más allá del visionado.

- Las claves de respuestas permiten el estudio autónomo y la flexibilidad en la evaluación.

Por ejemplo, *The Interview* ayuda a los estudiantes a explorar el discurso profesional y los dilemas morales relacionados con la honestidad y la autopresentación, mientras que *The Breath* abre conversaciones éticas y emocionales sobre la traición y el perdón. A través de estas narrativas, Kabiak fusiona la forma lingüística, la emoción humana y la reflexión moral, convirtiendo el cortometraje en un medio para el crecimiento intelectual y comunicativo.

El diseño pedagógico de Kabiak se fundamenta en teorías de aprendizaje comunicativo, basado en tareas y sociocultural. En lugar de enseñar la gramática como reglas descontextualizadas, integra las estructuras en una comunicación significativa. Este enfoque inductivo se alinea con la Hipótesis de la Noticing de Schmidt (1990) y con el marco del aprendizaje basado en tareas de Ellis (2003), que abogan por un aprendizaje impulsado por el descubrimiento.

El uso de cortometrajes auténticos también refleja la Hipótesis del Input de Krashen (1982), al proporcionar input comprensible enriquecido por la implicación emocional. De manera más amplia, el aula de Kabiak encarna la teoría sociocultural de Vygotsky (1978), que enfatiza el aprendizaje como un proceso mediado socialmente y apoyado por la interacción y el andamiaje.

Además, el libro ilustra principios del aprendizaje multimodal (Kress y van Leeuwen, 2001): las películas combinan códigos visuales, auditivos y lingüísticos, ayudando a los estudiantes a construir significado a través de múltiples canales. De este modo, la experiencia cinematográfica no se convierte en un material suplementario, sino en un vehículo central para la adquisición lingüística y cultural.

Una fortaleza central del manual es la fusión del lenguaje con la cultura contemporánea. Cada cortometraje refleja realidades cotidianas en contextos rusos y kazajos —dificultades laborales, dinámicas familiares, conciencia medioambiental o transformación personal—. Las tareas que acompañan a estas películas fomentan el desarrollo de la competencia comunicativa intercultural (Byram, 1997), alentando a los estudiantes a interpretar, comparar y evaluar perspectivas culturales a través del lenguaje.

Los ejercicios basados en la reflexión desplazan sistemáticamente a los estudiantes desde la comprensión hacia la evaluación crítica y la respuesta personal. Al invitar a la comparación con sus propias sociedades, los estudiantes no solo practican ruso, sino que también desarrollan empatía y alfabetización cultural. En este sentido, el libro funciona tanto como un proyecto lingüístico como humanístico.

El manual de Kabiak destaca por varias razones:

- Autenticidad: el uso de textos cinematográficos sin modificar expone a los estudiantes a un habla natural, variedad dialectal y vocabulario coloquial.
- Equilibrio pedagógico: cada módulo armoniza las cuatro destrezas —escuchar, hablar, leer y escribir—.
- Adaptabilidad: el diseño modular permite su uso en diversos contextos docentes, desde programas universitarios hasta el autoaprendizaje.
- Aprendizaje autónomo: la inclusión de claves de respuestas empodera a los estudiantes más allá del aula.
- Dimensión digital: la incorporación de comentarios reales de espectadores en línea introduce al alumnado en el discurso ruso informal y digital.

Además, al enmarcar el visionado de cortometrajes como una actividad de indagación y no de entretenimiento, Kabiak reposiciona a los estudiantes como intérpretes activos del significado. Esta postura pedagógica se alinea con el movimiento hacia la alfabetización mediática crítica, en el que los estudiantes aprenden a analizar cómo el lenguaje y la imagen interactúan para transmitir ideología y emoción.

A pesar de sus numerosas fortalezas, algunos aspectos merecen consideración. Algunos estudiantes —especialmente aquellos en el nivel inferior del rango intermedio— pueden encontrar desafiante la instrucción gramatical implícita sin apoyo adicional. Notas gramaticales complementarias o un recurso digital podrían mejorar la accesibilidad.

El acceso a las películas también depende de una conectividad a internet fiable y de la disponibilidad continuada de los enlaces de streaming. Una plataforma institucional segura o un sitio complementario curado mejoraría la sostenibilidad.

Por último, la densidad cultural de las películas, rica en referencias locales y matices sociales, puede requerir la mediación del docente. No obstante, estos desafíos son inherentes al uso de materiales auténticos y subrayan la necesidad de una mediación reflexiva más que de un control prescriptivo.

Learn Russian through Contemporary Short Film presenta un argumento convincente a favor de la enseñanza basada en el cine como un espacio de educación lingüística, cultural y estética. Ejemplifica el giro interdisciplinar en la lingüística aplicada que integra estudios de medios, alfabetización digital y teoría intercultural.

El modelo de Kabiak puede inspirar a docentes de otras lenguas menos difundidas que buscan formas innovadoras de implicar al alumnado con

discursos del mundo real. El libro también ofrece un terreno fértil para futuras investigaciones sobre el procesamiento del input multimodal, la motivación del alumnado y la adquisición de lenguas basada en medios.

Más allá de su valor pedagógico inmediato, el manual invita a una reconceptualización de la educación lingüística como un acto de participación cultural: un compromiso con el arte, la empatía y la comprensión ética.

En suma, *Learn Russian through Contemporary Short Film* de Natallia Kabiak representa una contribución prospectiva y metodológicamente sólida a la pedagogía del ruso. Al fusionar la formación lingüística con la experiencia cultural auténtica, el libro revitaliza la manera en que el lenguaje, el cine y la identidad se intersectan en el aula. Su estructura modular, profundidad académica y resonancia emocional lo convierten en un recurso valioso para docentes, estudiantes e investigadores.

Para programas de Estudios Rusos, Lingüística Aplicada o Comunicación Intercultural, este manual no es simplemente una herramienta didáctica: es un modelo de cómo la educación puede humanizar, modernizar y globalizar el aprendizaje de lenguas.

Referencias

- Byram, M. (1997). Teaching and assessing intercultural communicative competence. *Multilingual Matters*.
- Ellis, R. (2003). *Task-based language learning and teaching*. Oxford University Press.
- Krashen, S. (1982). *Principles and practice in second language acquisition*. Pergamon.
- Kress, G., & van Leeuwen, T. (2001). *Multimodal discourse: The modes and media of contemporary communication*. Arnold.
- Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics*, 11(2), 129–158.
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.

Xiying Wang

Garrido Yserte, R. y Casanueva, H. (2025, Eds.). Geopolítica de la Educación Superior: La Universidad como actor global. Catarata. ISBN: 978-84-1067-419-6

El campo de la Geopolítica ha experimentado una notable transformación en los últimos años. Tradicionalmente limitada a la sociología o la política, su estudio en educación se abordaba desde dos perspectivas: la Educación Comparada (currículo e instrucción) y la Historia de la Educación. Hasta hace poco, las aproximaciones eran escasas, destacando trabajos recientes en España como los del grupo GIEEPE-URJC en Economía de la Educación. Por ello, esta visión más internacionalizada de la Geopolítica en la Educación Superior convierte al volumen en una aportación valiosa desde perspectivas educativas, sociológicas y sociales. Un vistazo a sus contenidos, en gran parte elaborados por autores de la Universidad de Alcalá, revela una variedad temática relevante para la universidad del futuro, objetivo central del libro. No se limita a describir el estado actual del debate —entre proteccionismo social y enfoque neoliberal—, sino que muestra cómo la universidad puede ser un actor global capaz de influir en los grandes cambios del siglo XXI. Aunque organizado académicamente, cada capítulo transmite la idea práctica de que las universidades deben asumir retos como transferencia, empleabilidad y complejidad en un mundo marcado por tensiones tecnológicas, ecológicas y sociales que polarizan la educación entre Occidente y Asia. El prólogo de José Vicente Sanz enfatiza este momento de cambio social y económico de la universidad con mención especial al Grupo de investigación IAES como facilitador de la investigación para la transformación del conocimiento y motor de la acción social para cambiar un patrón anquilosado y buscar nuevas dimensiones más actuales. Después, la introducción de Rubén Garrido Yserte y Héctor Casanueva mueve a reconsiderar lo que significa “universidad” y su gobernanza en un mundo lleno de incertidumbres. Creo que esta introducción ayuda bastante porque explica que la universidad no solo recibe el impacto de los cambios globales, sino que puede liderar respuestas colectivas. El libro se estructura en tres partes que abordan la relación entre educación superior y los retos globales. En la primera, Concepción Olavarrieta analiza amenazas como la crisis climática y la desigualdad, subrayando que la universidad no puede permanecer estática, mientras Héctor Casanueva presenta el “Pacto para el Futuro” de la ONU como respuesta multilateral y destaca el papel de la universidad como puente entre políticas y sociedad. La segunda parte se centra en la dimensión ética y cooperativa: María del Carmen Patricia

Morales reflexiona sobre la solidaridad intelectual y moral; Yuma Inzolia propone reimaginar la educación superior para que sea inclusiva y sostenible; Francisco Rojas Aravena y Emily Alfaro Rojas muestran cómo las universidades pueden cooperar para la paz mediante diplomacia académica; y M^a Jesús Such Devesa, M^a Jesús Salado García y Elena Mañas profundizan en el papel universitario respecto a los ODS, transformándolos en objetivos concretos. La tercera parte aborda la transformación institucional: Alfonso González Hermoso de Mendoza plantea que no basta con actualizar programas, sino que se requiere un cambio cultural; Xóchitl Arias González y Guillermina Benavides Rincón insisten en formar para la complejidad; Martina Bo aporta la perspectiva estudiantil; María Teresa Gallo Rivera y Rubén Garrido Yserte discuten la universidad con propósito y la necesidad de medir su valor social; y los capítulos de Óscar Montes Pineda, María Cristina Pineda de Carías, Yiem Ataucusi y Garrido Yserte con Natalia Usach abordan equidad digital, territorio y desarrollo regional, mostrando que la universidad debe actuar global y localmente. En conjunto, la obra no es solo un diagnóstico, sino una llamada a la acción: plantea que la universidad del futuro debe ser flexible, comprometida y consciente de su responsabilidad ética. Aunque algunos capítulos son más conceptuales, el libro mantiene coherencia y ofrece una propuesta sólida para comprender cómo la educación superior se inserta en la geopolítica y cómo puede transformarse para contribuir a un mundo más justo y sostenible.

En resumen, este es un volumen que viene a cubrir un vacío que debe ir completándose en el futuro. La geoestrategia de la educación es un aspecto llamado a tener una enorme relevancia en los próximos años y, por ello, este trabajo es un excelente punto de inicio para una investigación que deberá avanzar en el futuro ya cercano.

Jesús García Laborda,
Universidad de Alcalá

Gaukbar Niyetalina,
Turan University

Patrick Parra Pennefather (2025). *Regenerating Learning: Transforming How You Learn with Generative AI*. 565 pp. ISBN: 979-8868810602

La inteligencia artificial generativa no surgió de la nada; marca una nueva fase en la evolución de varias décadas del aprendizaje automático. Herramientas como ChatGPT están pasando de ser pilotos aislados a convertirse en el centro de la vida social y organizacional cotidiana. En todo el mundo, la integración impulsada por la IA se está convirtiendo en algo rutinario, especialmente en organizaciones bien equipadas y en entornos educativos, donde la IA generativa se está integrando sistemáticamente en los flujos de trabajo laborales, docentes y de aprendizaje. Sin embargo, este cambio también trae consigo desafíos éticos y la presión de una narrativa de “la eficiencia por encima de todo”. La IA generativa puede tanto remodelar las formas establecidas de aprender y trabajar como amplificar complejidades sociales persistentes, incluidas las parcialidades y la desinformación, lo que hace que se sienta menos como una innovación única y más como una alarma que se intensifica cada vez más. En este contexto, *Regenerating Learning: Transforming How You Learn with Generative AI* se enfrenta directamente a la rápida propagación de la IA generativa. Basándose en casos ricos y teoría educativa, rastrea cómo la interacción humana-IA está remodelando los mecanismos de aprendizaje y reexamina cuestiones clave: cómo aprendemos, qué aprendemos y qué debe actualizarse en este nuevo panorama tecnológico. Con una postura de curiosidad abierta y un discernimiento crítico necesario, el libro aclara conceptos erróneos comunes, destaca el aprendizaje iterativo a través del uso práctico y ayuda a los lectores a navegar oportunidades y riesgos con un juicio más agudo. Ofrece caminos accionables para trabajadores creativos y del conocimiento, mientras proporciona un marco más amplio para comprender la transformación del aprendizaje y la renovación de capacidades, equilibrando eficiencia con calidad y equidad, y promoviendo un ecosistema de aprendizaje más inclusivo y sostenible para el futuro.

El libro consta de veinte capítulos. El Capítulo 1 enfatiza la reescritura de la narrativa del aprendizaje con una mentalidad abierta, tratando el “aprendizaje a partir de la IA” como un punto de partida para la transformación. El Capítulo 2 aboga por reflexionar sobre y ajustar los modelos de enseñanza y aprendizaje existentes, utilizando la autoconciencia para aclarar cómo las experiencias pasadas moldean el aprendizaje de cada persona. El Capítulo 3 enmarca la IA generativa como una oportunidad

para romper con los paradigmas pedagógicos arraigados y repensar lo que significa “inteligencia”, mientras guía a los lectores a reconocer sus propios patrones de aprendizaje. El Capítulo 4 introduce el concepto de “aprendizaje interrumpido” en el lugar de trabajo. El Capítulo 5 aborda el papel del “investigador”, destacando el diseño de aprendizaje centrado en el ser humano. El Capítulo 6 se enfoca en la acción y la práctica, explorando el potencial de la IA generativa para conectar el aprendizaje de conocimiento con la aplicación en el mundo real. El Capítulo 7 resalta la necesidad de construir estándares personales para evaluar los resultados de la IA. El Capítulo 8 discute cómo adaptarse a la IA: tratar el trabajo como un proceso continuo, abrazar el fracaso, romper hábitos y aprender de los resultados generados por la IA. El Capítulo 9 se centra en el proceso de “prototipado” en la interacción humana-IA. El Capítulo 10 enfatiza el aprendizaje iterativo a través del uso de herramientas de IA. El Capítulo 11 examina cómo equilibrar eficiencia, ética y riesgo al usar IA generativa. El Capítulo 12 subraya la importancia de entender los algoritmos. El Capítulo 13 destaca el aprendizaje continuo a través de la interacción continua con los sistemas de IA. El Capítulo 14 introduce el mayor sentido de control habilitado por los chatbots personalizables. El Capítulo 15 argumenta que la alfabetización en IA es solo un punto de partida y que un aprendizaje más profundo y específico se logra mediante la práctica deliberada. El Capítulo 16 enfatiza el valor de aprender con otros. El Capítulo 17 se enfoca en el aprendizaje organizacional y la reevaluación del valor del aprendizaje a medida que las empresas integran la IA. El Capítulo 18 subraya la importancia de mantener el control narrativo y la autoría creativa en la producción de contenido mediada por IA. El Capítulo 19 analiza el techno-hype sin evidencia y las afirmaciones exageradas. El Capítulo 20 regresa a las reflexiones fundamentales sobre inteligencia, creatividad y preguntas clave relacionadas.

En general, basado en la teoría del aprendizaje y la educación, el libro ofrece un análisis matizado de cómo la IA generativa está remodelando el aprendizaje en el lugar de trabajo: cómo aprendemos, qué aprendemos, cuándo y dónde aprendemos, y por qué aprendemos. Va más allá de la tradicional “narrativa escolar”, situando el aprendizaje dentro de las realidades más amplias del trabajo y la vida cotidiana. Es importante destacar que no es simplemente una guía práctica para usar herramientas; promueve una metodología de aprendizaje más transferible. A través de la interacción exploratoria con la IA generativa, el libro invita a los lectores a reflexionar y reconfigurar sus hábitos de aprendizaje, y, mediante el “modelo del iceberg”, subraya que la competencia con las herramientas es solo la superficie visible, debajo de la cual yacen oportunidades más

profundas para los ecosistemas de aprendizaje y la transformación organizacional. A lo largo del libro, se promueve la “curiosidad con cautela”, ayudando a los lectores a separar el verdadero valor del bombo tecnológico y ofreciendo una guía oportuna para los estudiantes y trabajadores del conocimiento en todo el mundo. Al enfatizar la autorregulación y el aprendizaje iterativo mientras se enfrentan preocupaciones prácticas como el sesgo, las alucinaciones, la privacidad y la seguridad, el libro equipa a los lectores con un marco más sólido para navegar por las compensaciones entre eficiencia y riesgo.

Sin embargo, aunque el libro ofrece una descripción bastante sistemática, basada en la teoría del aprendizaje, de cómo la IA generativa está remodelando aspectos clave del aprendizaje en el lugar de trabajo, aún queda espacio para un mayor desarrollo en varios aspectos. Por un lado, el libro combina eficazmente la teoría con la práctica y se apoya en casos ricos, pero la mayoría de los ejemplos provienen de contextos con condiciones técnicas y de recursos relativamente fuertes. Como resultado, ofrece una cobertura limitada de las variaciones entre regiones, industrias y poblaciones con diferentes niveles de acceso a los recursos. Por lo tanto, trabajos futuros podrían realizar estudios comparativos en entornos más diversos y articular más claramente los caminos de integración institucional dentro de entornos educativos formales, como las escuelas. Por otro lado, aunque el libro resalta la importancia de las políticas y el apoyo organizacional, el marco de gobernanza para traducir ideas en implementación aún requiere mayor especificidad. Por ejemplo, investigaciones futuras podrían mapear cuestiones como la seguridad y protección de la privacidad, la soberanía de los datos y el cumplimiento transfronterizo en diferentes escenarios de uso según el nivel de riesgo, y desarrollar casos reutilizables y plantillas para aclarar los requisitos y hacer que sean más fáciles de ejecutar, avanzando así la discusión hacia una orientación más concreta y accionable.

Sin duda, al vincular el contexto social con la teoría educativa, el libro ofrece una descripción sistemática de cómo la IA generativa está remodelando el aprendizaje y el trabajo, los riesgos que introduce y los caminos prácticos para responder. No solo profundiza nuestra comprensión del “aprendizaje” en sí, sino que también vincula la adopción tecnológica con la transformación de los ecosistemas de aprendizaje, proporcionando un marco que está tanto teóricamente fundamentado como prácticamente útil para la investigación futura. Además, al confrontar los desequilibrios estructurales que una narrativa de “eficiencia por encima de todo” puede generar, fomenta que las organizaciones y los sistemas educativos construyan un consenso y una gobernanza más duraderos en torno a la pri-

vacidad, la seguridad y la equidad. Además, ofrece apoyo metodológico para la colaboración intercultural y el desarrollo de la competencia global, impulsando la transformación del lugar de trabajo y del aprendizaje en la era de la IA generativa y contribuyendo a un ecosistema de aprendizaje más inclusivo, bien gobernado y sostenible.

Información de financiamiento:

- Proyecto de Fondo de Investigación Científica del Departamento de Educación de la Provincia de Yunnan 2019: Investigación sobre la Ruta de Orientación de Posgrado en Disciplinas de Lenguas Extranjeras en el Este de China en el Contexto de la Construcción de los “Double First-Class” (Número de Proyecto: 2019J0217).
- Proyecto de Fondo Especial de Traducción del Proyecto Nacional de Enseñanza y Investigación de Lenguas Extranjeras de la Editorial de Enseñanza y Investigación de Lenguas Extranjeras 2019: Investigación sobre el Desarrollo y los Factores que Influyen en la Autoeficacia en los Traductores Novatos (Número de Proyecto: 2020YN0025).

**Zizhao Xiong,
Qiujun Su,
Yang Yao,**

*Facultad de Lenguas Extranjeras (Escuela Internacional),
Universidad Forestal del Suroeste, China*

Revista de Educación es una publicación científica del Ministerio de Educación y Formación Profesional español. Fundada en 1940, y manteniendo el título de Revista de Educación desde 1952, es un testigo privilegiado de la evolución de la educación en las últimas décadas, así como un reconocido medio de difusión de los avances en la investigación y la innovación en este campo, tanto desde una perspectiva nacional como internacional. La revista es editada por la Subdirección General de Atención al Ciudadano, Documentación y Publicaciones, y actualmente está adscrita al Instituto Nacional de Evaluación Educativa de la Dirección General de Evaluación y Cooperación Territorial.



GOBIERNO
DE ESPAÑA

MINISTERIO
DE EDUCACIÓN, FORMACIÓN PROFESIONAL
Y DEPORTES