



UNIVERSIDAD ESTATAL DE MILAGRO

REPÚBLICA DEL ECUADOR

**UNIVERSIDAD ESTATAL DE MILAGRO
FACULTAD DE POSGRADO**

VICERRECTORADO DE INVESTIGACIÓN Y POSGRADO

ARTÍCULOS PROFESIONALES DE ALTO NIVEL

PREVIO A LA OBTENCIÓN DEL TÍTULO DE:

**MAGÍSTER EN EDUCACIÓN CON MENCIÓN EN DOCENCIA E
INVESTIGACIÓN EN EDUCACIÓN SUPERIOR**

TEMA:

**ANÁLISIS DE LOS EFECTOS DE CHATGPT EN EL APRENDIZAJE DE LOS
ESTUDIANTES: UNA REVISIÓN SISTEMÁTICA**

Autor:

PAULO EMANUEL VELEZ LEON

Director:

MSC. JOHANNA ALEXANDRA BONILLA GUACHAMÍN

Milagro, 2026

RESUMEN

Antecedentes: La incorporación de ChatGPT a la enseñanza ha producido resultados divergentes sobre rendimiento académico, percepción del aprendizaje y pensamiento de orden superior. **Objetivo:** Estimar sus efectos y examinar condiciones pedagógicas moderadoras. **Elegibilidad:** Se incluyeron estudios experimentales o cuasiexperimentales en inglés, publicados entre noviembre de 2022 y febrero de 2025, con grupo experimental que utilizó ChatGPT, grupo control sin ChatGPT y datos suficientes para calcular tamaños de efecto. **Fuentes:** Web of Science, ScienceDirect, Taylor & Francis, Wiley Online Library y Scopus. **Métodos:** La selección siguió PRISMA 2020; se realizó el cribado y se calcularon diferencias de medias estandarizadas mediante g de Hedges bajo modelos de efectos aleatorios. **Resultados:** Se sintetizaron 51 estudios y 72 estimaciones declaradas. ChatGPT mostró un efecto significativo sobre el rendimiento académico ($g = 0.867$) y efectos moderados sobre la percepción del aprendizaje ($g = 0.456$) y el pensamiento de orden superior ($g = 0.457$). El tipo de curso, el modelo de aprendizaje, la duración y el rol pedagógico explicaron parte de la heterogeneidad. **Conclusiones:** ChatGPT puede favorecer el aprendizaje cuando se integra con diseño didáctico, andamiaje, evaluación formativa y supervisión docente; la evidencia no respalda una adopción indiscriminada.

PALABRAS CLAVE: ChatGPT, Inteligencia Artificial Generativa, Rendimiento Académico, Percepción del Aprendizaje, Pensamiento de Orden Superior

ABSTRACT

Background: The educational integration of ChatGPT has yielded divergent findings regarding academic achievement, learning perception, and higher-order thinking. **Objective:** To estimate its effects and examine pedagogical moderators. **Eligibility:** English-language experimental or quasi-experimental studies published between November 2022 and February 2025 were included when the experimental group used ChatGPT, the control group did not, and sufficient data were available to calculate effect sizes. **Sources:** Web of Science, ScienceDirect, Taylor & Francis, Wiley Online Library, and Scopus. **Methods:** Study selection followed PRISMA 2020; two reviewers screened records, and standardized mean differences were calculated as Hedges's g using random-effects models. **Results:** The review declared 51 studies and 72 effect estimates. ChatGPT produced a large effect on academic achievement ($g = 0.867$) and moderate effects on learning perception ($g = 0.456$) and higher-order thinking ($g = 0.457$). Course type, learning model, intervention duration, and pedagogical role explained part of the heterogeneity. **Conclusions:** ChatGPT can support learning when embedded in purposeful instructional design, scaffolding, formative assessment, and teacher oversight; the evidence does not justify indiscriminate adoption.

KEYWORDS: ChatGPT, Generative Artificial Intelligence, Academic Achievement, Learning Perception, Higher-Order Thinking.

1. INTRODUCCIÓN

1.1. Fundamento educativo y problema de investigación

El aprendizaje no consiste en la recepción pasiva de información, sino en la construcción progresiva de significados mediante la interacción entre conocimientos previos, tareas, lenguaje, retroalimentación y contexto. Desde una perspectiva constructivista, el estudiante reorganiza sus esquemas cognitivos al enfrentarse con problemas, explicaciones y formas de mediación que le permiten atribuir sentido a la experiencia. En esa lógica, una tecnología educativa adquiere valor cuando favorece la actividad intelectual del aprendiz, amplía sus posibilidades de participación y facilita procesos de regulación, no cuando sustituye el esfuerzo cognitivo que sostiene la comprensión (Piaget, 1970; Vygotskij & Hanfmann, 1966).

ChatGPT constituye una modalidad de inteligencia artificial generativa capaz de producir texto, responder preguntas, reformular contenidos, ofrecer ejemplos y mantener intercambios conversacionales. En educación, estas funciones han sido asociadas con acceso inmediato a explicaciones, retroalimentación personalizada, apoyo a la escritura, práctica lingüística, resolución de problemas y generación de recursos. Al mismo tiempo, la literatura advierte riesgos relacionados con respuestas inexactas, sesgos, dependencia tecnológica, reducción de la interacción humana, plagio y debilitamiento del juicio crítico cuando el sistema se utiliza sin criterios pedagógicos ni verificación de las respuestas (Ali et al., 2024; Kasneci et al., 2023; Yan et al., 2024).

La evidencia empírica disponible no es uniforme. Algunos experimentos informan mejoras en desempeño, autorregulación, motivación, escritura y resolución creativa de problemas cuando ChatGPT se integra a actividades estructuradas; otros no encuentran diferencias significativas frente a la enseñanza convencional o registran efectos adversos sobre autoeficacia, flujo, interés o creatividad. Estas divergencias pueden depender menos de la mera presencia de la herramienta que del diseño de la intervención, el tipo de tarea, la duración, el nivel educativo, el papel atribuido a ChatGPT y la calidad del acompañamiento docente (Donald et al., 2024; Lu et al., 2024; Niloy et al., 2023; Urban et al., 2024; Yang et al., 2025).

Las revisiones previas han descrito oportunidades y riesgos, pero varias se apoyan principalmente en estudios teóricos, percepciones o experiencias sin grupo de comparación. Otras se concentran en un solo resultado, como compromiso estudiantil, aprendizaje informal de inglés o educación escolar, lo cual limita la posibilidad de estimar una respuesta integrada para distintos contextos educativos. Aunque un metaanálisis reciente examinó los efectos de ChatGPT sobre el aprendizaje, la selección de estudios incluyó intervenciones cuya

comparabilidad pedagógica fue cuestionada, porque en algunos casos el grupo experimental y el grupo control no desarrollaron el mismo método de aprendizaje (Deng et al., 2025; Guan et al., 2024; Heung & Chiu, 2025; Zhang & Tur, 2024).

La presente síntesis se diferencia de esos antecedentes al restringir la muestra a estudios experimentales o cuasiexperimentales en los que el grupo experimental utilizó ChatGPT, el grupo control no lo utilizó y ambos grupos siguieron un método de aprendizaje comparable. Además, integra tres dominios complementarios: rendimiento académico, percepción del aprendizaje y pensamiento de orden superior. El rendimiento expresa resultados verificables mediante pruebas o productos; la percepción comprende actitudes, motivación, autoeficacia, interés o valoración de la experiencia; y el pensamiento de orden superior abarca procesos como análisis, evaluación, creatividad, metacognición y resolución de problemas (Lu et al., 2024; Urban et al., 2024; Zheng et al., 2021).

1.2. Marco conceptual

El valor pedagógico de ChatGPT puede interpretarse a partir del constructivismo y de los enfoques de aprendizaje activo. Como mediador, el sistema puede ofrecer explicaciones graduadas, ejemplos alternativos y preguntas que ayuden a conectar conocimientos previos con nuevos contenidos. Sin embargo, el beneficio depende de que la interacción conduzca a elaborar, contrastar y aplicar información. Cuando el estudiante se limita a copiar una respuesta, la herramienta reduce la actividad cognitiva; cuando debe justificar, corregir, comparar fuentes o revisar estrategias, puede convertirse en un andamiaje que amplía la zona de desarrollo y la reflexión sobre el propio aprendizaje (Piaget, 1970; Vygotskij & Hanfmann, 1966).

La percepción del aprendizaje puede explicarse también mediante el modelo de aceptación tecnológica. La utilidad percibida y la facilidad de uso influyen en la disposición a adoptar una herramienta, pero una valoración favorable no equivale necesariamente a un aprendizaje profundo. ChatGPT puede aumentar el interés por su rapidez y accesibilidad, aunque su falta de sensibilidad emocional y la posible mecanización de la experiencia limitan su capacidad para sustituir la relación pedagógica. La aceptación debe interpretarse junto con indicadores de desempeño y procesos cognitivos, no como evidencia aislada de eficacia educativa (Birenbaum, 2023; Venkatesh & Bala, 2008).

El pensamiento de orden superior exige algo más que obtener respuestas correctas. Supone formular problemas, evaluar argumentos, detectar supuestos, transferir conocimientos, producir alternativas y supervisar estrategias. ChatGPT puede presentar perspectivas diversas

y ofrecer retroalimentación inmediata, pero sus respuestas proceden de patrones aprendidos y pueden contener errores o razonamientos plausibles sin fundamento. Por ello, su contribución depende de actividades que obliguen a verificar, argumentar, contrastar y reconstruir las respuestas generadas, especialmente en cursos que demandan resolución de problemas o diseño de productos complejos (Li et al., 2024; Lu et al., 2024; Niloy et al., 2023).

1.3. Objetivo y preguntas de investigación

El objetivo fue estimar el efecto de ChatGPT sobre el rendimiento académico, la percepción del aprendizaje y el pensamiento de orden superior, e identificar las condiciones educativas que moderan dichos efectos. En términos de un PICO adaptado, la población estuvo formada por estudiantes de primaria, secundaria y educación superior; la intervención consistió en el uso de ChatGPT dentro de actividades de aprendizaje; la comparación correspondió a grupos sin ChatGPT y con método pedagógico equivalente; y los resultados fueron desempeño, percepción y pensamiento de orden superior.

La revisión respondió dos preguntas: ¿qué magnitud tiene el efecto de ChatGPT sobre el rendimiento académico, la percepción del aprendizaje y el pensamiento de orden superior?; y ¿en qué medida el nivel educativo, el tipo de curso, el modelo de aprendizaje, la duración, el rol de ChatGPT y el área de aplicación modifican esos efectos? Estas preguntas permiten distinguir entre eficacia promedio y eficacia condicionada por decisiones de diseño pedagógico.

§2. MÉTODOS

2.1. Diseño, protocolo y registro

La investigación adoptó un diseño de revisión sistemática con metaanálisis de estudios experimentales y cuasiexperimentales. La identificación, el cribado, la evaluación de elegibilidad y la inclusión se organizaron conforme a PRISMA 2020, con el propósito de hacer explícitas las decisiones de búsqueda y selección. El metaanálisis integró resultados cuantitativos comparables y examinó heterogeneidad, sesgo de publicación, sensibilidad y variables moderadoras (Moher et al., 2009; Page et al., 2021; Page & Moher, 2017).

La estrategia de búsqueda disponible y las variables de codificación se detallan a continuación. Los parámetros de búsqueda se aplicaron en Web of Science Core Collection, ScienceDirect, Taylor & Francis, Wiley Online Library y Scopus para el periodo comprendido entre noviembre de 2022 y febrero de 2025: ("ChatGPT" OR "chat generative pre-trained

transformer" OR "GPT-3.5" OR "GPT-4" OR "GPT-4o" OR "GenAI" OR "generative AI" OR "generative artificial intelligence" OR "Artificial Intelligence Generated Content" OR "AIGC") AND ("educat*" OR "learn*" OR "teach*") AND ("experiment*" OR "randomised controlled trial*" OR "randomized controlled trial*" OR "RCT*" OR "quasi-experiment*" OR "intervention").

Las variables de codificación se detallan y analizan en la siguiente tabla:

Tabla 1
Características de los estudios incluidos

<u>Autor y año</u>	<u>EG</u>	<u>CG</u>	<u>Nivel</u>	<u>Tipo de curso</u>	<u>Duración</u>	<u>Intervención codificada</u>	<u>Resultado principal</u>
Emran et al. (2024)	34	34	Educación superior	Aprendizaje de lenguas y escritura académica	> 8 semanas	Aprendizaje basado en problemas; Herramienta inteligente aprendizaje; Tutoría	Rendimiento g = 2.730
Almohesh (2024)	75	75	Primaria	Desarrollo de habilidades y competencias	≤ 1 semana	Aprendizaje personalizado; Tutor inteligente; Mixto	Rendimiento g = 1.117
Avello-Martínez et al. (2023)	20	21	Educación superior	Desarrollo de habilidades y competencias	≤ 1 semana	Aprendizaje personalizado; Herramienta inteligente aprendizaje; Recomendación personalizada	Rendimiento g = 0.062
Boudouaia et al. (2024)	37	39	Educación superior	Aprendizaje de lenguas y escritura académica	> 8 semanas	Aprendizaje reflexivo; Tutor inteligente; Evaluación y retroalimentación	Rendimiento g = 0.797
Bašić et al. (2023a)	12	12	Educación superior	STEM y cursos afines	≤ 1 semana	Aprendizaje basado en problemas; Herramienta inteligente aprendizaje; Tutoría	Rendimiento g = 0.993
Chen and Chang (2024)	31	30	Educación superior	STEM y cursos afines	1–4 semanas	Mixto; Herramienta inteligente aprendizaje; Mixto	Rendimiento g = 1.235; Percepción g = 1.290
Chen and Chang (2024)	63	70	Secundaria	STEM y cursos afines	1–4 semanas	Mixto; Herramienta inteligente aprendizaje; Tutoría	Rendimiento g = 0.169
Song and Song (2023)	25	25	Educación superior	Aprendizaje de lenguas y escritura académica	> 8 semanas	Mixto; Herramienta inteligente aprendizaje; Mixto	Rendimiento g = 0.790
Ng et al. (2024)	37	37	Secundaria	STEM y cursos afines	1–4 semanas	Mixto; Tutor inteligente; Tutoría	Percepción g = 0.864; Pensamiento superior g = 0.943
Sun et al. (2024)	43	39	Educación superior	STEM y cursos afines	4–8 semanas	Mixto; Herramienta inteligente aprendizaje; Tutoría	Rendimiento g = 0.280

<u>Autor y año</u>	<u>EG</u>	<u>CG</u>	<u>Nivel</u>	<u>Tipo de curso</u>	<u>Duración</u>	<u>Intervención codificada</u>		<u>Resultado principal</u>
Darmawansah et al. (2024)	34	33	Educación superior	Desarrollo de habilidades y competencias	4–8 semanas	Mixto; Mixto	Mixto; Mixto	Rendimiento $g = 0.697$
Donald et al. (2024)	16	17	Educación superior	STEM y cursos afines	1–4 semanas	Aprendizaje personalizado; Mixto; Mixto		Rendimiento $g = -0.334$; Percepción $g = -0.457$
Li (2023)	42	39	Educación superior	STEM y cursos afines	4–8 semanas	Mixto; inteligente; Mixto	Tutor Mixto	Rendimiento $g = 0.493$; Percepción $g = 0.476$
Ba et al. (2024)	39	38	Educación superior	Desarrollo de habilidades y competencias	1–4 semanas	Aprendizaje contextual; Tutor inteligente; Tutoría		Rendimiento $g = 0.067$
Hsu (2024)	20	20	Educación superior	Desarrollo de habilidades y competencias	> 8 semanas	Aprendizaje personalizado; Mixto; Mixto		Rendimiento $g = 2.420$
Hu (2024)	67	68	Educación superior	Desarrollo de habilidades y competencias	4–8 semanas	Aprendizaje basado en problemas; Compañero inteligente; Mixto		Rendimiento $g = 1.8985$
Huang et al. (2024)	60	59	Educación superior	STEM y cursos afines	≤ 1 semana	Mixto; Tutor inteligente; Recomendación personalizada		Rendimiento $g = 0.162$; Percepción $g = 0.180$
Huesca et al. (2024)	214	128	Educación superior	STEM y cursos afines	> 8 semanas	Aprendizaje personalizado; Herramienta inteligente de aprendizaje; Tutoría		Rendimiento $g = 0.453$
Roganović (2024)	16	44	Educación superior	STEM y cursos afines	≤ 1 semana	Aprendizaje basado en problemas; Herramienta inteligente de aprendizaje; Tutoría		Rendimiento $g = 0.037$
Lu et al. (2024)	107	108	Educación superior	Desarrollo de habilidades y competencias	1–4 semanas	Aprendizaje contextual; Mixto; Mixto		Rendimiento $g = 0.683$; Percepción $g = 0.403$; Pensamiento superior $g = 0.039$
Escalante et al. (2023)	23	25	Educación superior	Aprendizaje de lenguas y escritura académica	4–8 semanas	Aprendizaje reflexivo; Tutor inteligente; Evaluación y retroalimentación		Rendimiento $g = 0.500$
Mahapatra (2024)	35	37	Educación superior	Aprendizaje de lenguas y escritura académica	4–8 semanas	Aprendizaje personalizado; Mixto; Mixto		Rendimiento $g = 1.175$
Urban et al. (2024)	77	68	Educación superior	Desarrollo de habilidades y competencias	≤ 1 semana	Aprendizaje basado en problemas; Herramienta inteligente de aprendizaje; Mixto		Rendimiento $g = 0.612$; Percepción $g = 0.337$; Pensamiento superior $g = 0.366$
	20	20	Educación superior	STEM y cursos afines	> 8 semanas	Aprendizaje basado en problemas; Herramienta inteligente de aprendizaje; Tutoría		Rendimiento $g = 2.420$

<u>Autor y año</u>	<u>EG</u>	<u>CG</u>	<u>Nivel</u>	<u>Tipo de curso</u>	<u>Duración</u>	<u>Intervención codificada</u>	<u>Resultado principal</u>
Ghafouri, et al. (2024)	24	24	Educación superior	Aprendizaje de lenguas y escritura académica	> 8 semanas	Aprendizaje basado en problemas; Herramienta inteligente de aprendizaje; Evaluación y retroalimentación	Rendimiento g = 1.746; Percepción g = 1.622
Mugableh (2024)	30	30	Educación superior	Aprendizaje de lenguas y escritura académica	> 8 semanas	Aprendizaje personalizado; Tutor inteligente; Evaluación y retroalimentación	Rendimiento g = 1.553
Karaman and Goksu (2024)	24	15	Secundaria	STEM y cursos afines	4–8 semanas	Mixto; Herramienta inteligente de aprendizaje; Tutoría	Rendimiento g = 0.070
Jafarian et al. (2025)	120	153	Educación superior	STEM y cursos afines	> 8 semanas	Aprendizaje personalizado; Herramienta inteligente de aprendizaje; Tutoría	Rendimiento g = 0.047
Pan et al. (2025)	30	31	Educación superior	Aprendizaje de lenguas y escritura académica	> 8 semanas	Aprendizaje personalizado; Herramienta inteligente de aprendizaje; Recomendación personalizada	Rendimiento g = 0.798; Pensamiento superior g = 0.706
Qi et al. (2024)	23	23	Educación superior	Aprendizaje de lenguas y escritura académica	4–8 semanas	Aprendizaje basado en problemas; Tutor inteligente; Evaluación y retroalimentación	Rendimiento g = 1.105
Alneyadi and Wardat (2023)	58	64	Secundaria	STEM y cursos afines	> 8 semanas	Aprendizaje basado en problemas; Herramienta inteligente de aprendizaje; Tutoría	Rendimiento g = 0.973
Rakap (2023)	11	11	Educación superior	Desarrollo de habilidades y competencias	≤ 1 semana	Aprendizaje reflexivo; Herramienta inteligente de aprendizaje; Tutoría	Rendimiento g = 2.857
An et al. (2025)	55	55	Educación superior	Desarrollo de habilidades y competencias	4–8 semanas	Aprendizaje basado en proyectos; Mixto; Mixto	Rendimiento g = 1.374
Küchemann et al. (2023)	13	13	Secundaria	STEM y cursos afines	≤ 1 semana	Aprendizaje basado en problemas; Herramienta inteligente de aprendizaje; Tutoría	Rendimiento g = 0.688
Chan et al. (2024)	342	576	Educación superior	Aprendizaje de lenguas y escritura académica	≤ 1 semana	Aprendizaje basado en proyectos; Tutor inteligente; Evaluación y retroalimentación	Rendimiento g = 0.195; Percepción g = 0.180
Wu et al. (2024)	11	12	Educación superior	STEM y cursos afines	4–8 semanas	Aprendizaje basado en proyectos; Compañero inteligente; Tutoría	Rendimiento g = 1.537

<u>Autor y año</u>	<u>EG</u>	<u>CG</u>	<u>Nivel</u>	<u>Tipo de curso</u>	<u>Duración</u>	<u>Intervención codificada</u>	<u>Resultado principal</u>
Kosar et al. (2024)	99	83	Educación superior	STEM y cursos afines	> 8 semanas	Mixto; Herramienta inteligente aprendizaje; Mixto	Rendimiento $g = -0.034$
Yang et al. (2025)	76	77	Secundaria	STEM y cursos afines	1-4 semanas	Aprendizaje personalizado; Mixto; Mixto	Rendimiento $g = 0.315$; Percepción $g = 0.486$
Wang et al. (2024a, 2024b)	36	38	Educación superior	Desarrollo de habilidades y competencias	> 8 semanas	Aprendizaje personalizado; Herramienta inteligente aprendizaje; Evaluación y retroalimentación	Rendimiento $g = 0.712$
Gan et al. (2024)	55	55	Educación superior	STEM y cursos afines	1-4 semanas	Aprendizaje personalizado; Herramienta inteligente aprendizaje; Tutoría	Rendimiento $g = 0.281$
Ji et al. (2023)	21	15	Educación superior	STEM y cursos afines	4-8 semanas	Mixto; Compañero inteligente; Tutoría	Rendimiento $g = 4.009$
Basic et al. (2023b)	9	9	Educación superior	Aprendizaje de lenguas y escritura académica	≤ 1 semana	Aprendizaje basado en problemas; Herramienta inteligente aprendizaje; Tutoría	Rendimiento $g = -0.516$
Hui et al. (2025)	21	21	Educación superior	Desarrollo de habilidades y competencias	1-4 semanas	Aprendizaje basado en problemas; Compañero inteligente; Mixto	Rendimiento $g = 1.146$
Ji et al. (2023)	21	15	Educación superior	STEM y cursos afines	4-8 semanas	Aprendizaje basado en proyectos; Herramienta inteligente aprendizaje; Mixto	Percepción $g = 0.342$
Lu et al. (2024)	107	108	Educación superior	Desarrollo de habilidades y competencias	1-4 semanas	Aprendizaje contextual; Mixto; Mixto	Rendimiento $g = 0.683$; Percepción $g = 0.403$; Pensamiento superior $g = 0.039$
Liao et al. (2024)	26	26	Educación superior	STEM y cursos afines	4-8 semanas	Aprendizaje basado en proyectos; Herramienta inteligente aprendizaje; Tutoría	Percepción $g = -0.007$
Shang et al. (2024)	55	54	Educación superior	STEM y cursos afines	≤ 1 semana	Aprendizaje basado en problemas; Herramienta inteligente aprendizaje; Tutoría	Percepción $g = 0.320$
Wu et al. (2024)	35	35	Secundaria	STEM y cursos afines	1-4 semanas	Aprendizaje personalizado; Herramienta inteligente aprendizaje; Tutoría	Percepción $g = 0.474$

<u>Autor y año</u>	<u>EG</u>	<u>CG</u>	<u>Nivel</u>	<u>Tipo de curso</u>	<u>Duración</u>	<u>Intervención codificada</u>	<u>Resultado principal</u>
Yilmaz et al. (2023)	21	24	Educación superior	STEM y cursos afines	4–8 semanas	Aprendizaje personalizado; Herramienta inteligente aprendizaje; Tutoría	Percepción $g = 0.296$; Pensamiento superior $g = 0.710$
Essel et al. (2024)	60	65	Educación superior	Aprendizaje de lenguas y escritura académica	> 8 semanas	Aprendizaje basado en proyectos; Herramienta inteligente aprendizaje; Tutoría	Pensamiento superior $g = 0.157$
Kareem et al. (2025)	150	150	Educación superior	Desarrollo de habilidades y competencias	1–4 semanas	Aprendizaje basado en proyectos; Herramienta inteligente aprendizaje; Tutoría	Pensamiento superior $g = 0.448$

Nota. La tabla reproduce las 51 filas de la matriz. El país no fue consignado y no se infirió. La columna “Intervención codificada” resume, sin añadir información externa, el modelo de aprendizaje, el rol de ChatGPT y el área de aplicación.

2.2. Criterios de elegibilidad

Se incluyeron investigaciones relacionadas con ChatGPT y al menos uno de los tres resultados de interés: rendimiento académico, percepción del aprendizaje o pensamiento de orden superior. Los trabajos debían estar escritos en inglés y publicados entre noviembre de 2022, fecha de lanzamiento de ChatGPT, y febrero de 2025. Se admitieron participantes de primaria, secundaria y educación superior; la evidencia disponible no incorporó educación infantil y solo incluyó un estudio de primaria.

Los diseños elegibles fueron experimentales o cuasiexperimentales con al menos un grupo experimental y un grupo control. El grupo experimental debía utilizar ChatGPT durante una experiencia de aprendizaje, mientras que el control no podía emplear ChatGPT ni tecnologías emergentes con funciones inteligentes. Ambos grupos debían seguir el mismo método pedagógico, de modo que la diferencia estimada pudiera atribuirse con mayor plausibilidad al uso de ChatGPT y no a una combinación desigual de estrategias educativas.

Se excluyeron estudios no experimentales, investigaciones sobre intención de uso, experiencias en las que ChatGPT se empleó únicamente como herramienta instrumental sin proceso de aprendizaje y comparaciones con realidad aumentada, realidad virtual, metaverso, gemelos digitales u otros sistemas inteligentes. También se excluyeron trabajos sin medias, desviaciones estándar, tamaños muestrales o estadísticos equivalentes que permitieran calcular g de Hedges. No se estableció una restricción explícita por modalidad presencial, híbrida o virtual, pero todos los estudios debían situarse en una experiencia educativa identificable.

2.3. Fuentes de información y estrategia de búsqueda

La búsqueda se realizó en Web of Science Core Collection, ScienceDirect, Taylor & Francis, Wiley Online Library y Scopus. El periodo cubrió desde noviembre de 2022 hasta febrero de 2025. No se realizaron búsquedas en ERIC, Dialnet, Redalyc ni repositorios dedicados de literatura gris. La ecuación combinó términos de la tecnología, del contexto educativo y del diseño de intervención: (ChatGPT OR chat generative pre-trained transformer OR GPT-3.5 OR GPT-4 OR GPT-4o OR GenAI OR generative AI OR generative artificial intelligence OR Artificial Intelligence Generated Content OR AIGC) AND (educat* OR learn* OR teach*) AND (experiment* OR randomised controlled trial* OR randomized controlled trial* OR RCT* OR quasi-experiment* OR intervention). La sintaxis íntegra se detalla en 2.1 para favorecer la trazabilidad (Page et al., 2021).

2.4. Selección y extracción de datos

Dos investigadores especializados en tecnología educativa desarrollaron los criterios de inclusión y exclusión y seleccionaron de manera independiente los trabajos que cumplían las condiciones de codificación. Las discrepancias se resolvieron mediante discusión y consenso. Se reportó un coeficiente alfa de Cronbach de 0.96 como indicador de acuerdo; aunque este coeficiente se utiliza habitualmente para consistencia interna y no es la medida interjueces más específica, el valor informado sugiere una elevada coincidencia entre las decisiones de los evaluadores (Page et al., 2021).

La extracción registró autoría, año, tamaño de los grupos experimental y control, nivel educativo, tipo de curso, duración, modelo de aprendizaje, rol de ChatGPT, área de aplicación y resultado. Cuando un estudio contenía varios experimentos, cada experimento se trató como una estimación independiente; cuando un mismo constructo se medía mediante varias dimensiones, se calculó un promedio para evitar multiplicar indebidamente los tamaños de efecto de una sola muestra.

2.5. Calidad metodológica y riesgo de sesgo

La revisión aplicó controles de elegibilidad relacionados con diseño, comparabilidad de grupos, equivalencia del método de aprendizaje y suficiencia de los datos. También examinó sesgo de publicación mediante gráficos de embudo y N de seguridad, y evaluó estabilidad con análisis de eliminación de un estudio (Borenstein et al., 2009; Schulz et al., 1995).

2.6. Codificación y métodos de síntesis

Las variables moderadoras se organizaron en seis grupos. El nivel educativo distinguió primaria, secundaria y universidad. El tipo de curso separó áreas STEM, aprendizaje de lenguas

y escritura académica, y desarrollo de habilidades y competencias. El modelo de aprendizaje comprendió aprendizaje personalizado, basado en problemas, basado en proyectos, contextual, reflexivo y mixto. La duración se clasificó en una semana o menos, una a cuatro semanas, cuatro a ocho semanas y más de ocho semanas.

El rol de ChatGPT se codificó como tutor inteligente, compañero inteligente, herramienta inteligente de aprendizaje o función mixta. El área de aplicación distinguió evaluación y retroalimentación, tutoría, recomendación personalizada y uso mixto. Estas categorías permiten examinar si el efecto depende de la arquitectura pedagógica de la intervención y no únicamente del acceso a la tecnología.

El tamaño de efecto fue la diferencia de medias estandarizada corregida mediante g de Hedges, calculada a partir de medias, desviaciones estándar y tamaños muestrales de los grupos o, cuando esos datos no estaban disponibles, de estadísticos χ^2 , t o F . La corrección de Hedges reduce el sesgo de la diferencia estandarizada en muestras pequeñas. La magnitud se interpretó con los puntos de referencia de 0.10 muy pequeño, 0.20 pequeño, 0.50 mediano, 0.80 grande, 1.20 muy grande y 2.00 enorme (Hedges, 1981; Sawilowsky, 2009).

La heterogeneidad se evaluó con Q de Cochran e I^2 . Se consideró heterogeneidad baja entre 0 % y 25 %, moderada entre 25 % y 75 % y alta por encima de 75 %. Debido a que los tres resultados presentaron heterogeneidad estadísticamente significativa, se aplicaron modelos de efectos aleatorios, adecuados cuando se asume que los estudios estiman efectos verdaderos distintos y que la variación procede tanto del error muestral como de diferencias reales entre contextos (Borenstein et al., 2009; Higgins et al., 2003).

3. RESULTADOS

3.1. Selección de estudios

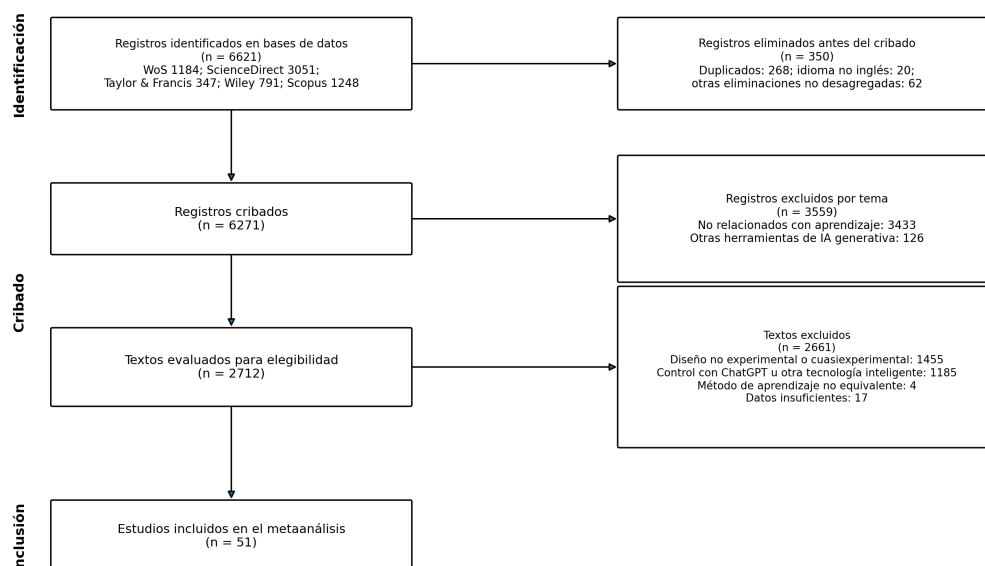
La búsqueda identificó 6621 registros: 1184 en Web of Science, 3051 en ScienceDirect, 347 en Taylor & Francis, 791 en Wiley Online Library y 1248 en Scopus. Antes del cribado se eliminaron 350 registros; 268 eran duplicados, 20 estaban en idiomas distintos del inglés y 62 no fueron desagregados. Se cribaron 6271 títulos y resúmenes, de los cuales 3559 fueron excluidos por falta de relación con el aprendizaje mediado por ChatGPT o por utilizar otras herramientas de inteligencia artificial generativa.

Pasaron a evaluación de elegibilidad 2712 textos. Se excluyeron 1455 por no presentar un diseño experimental o cuasiexperimental, 1185 porque el grupo control utilizaba ChatGPT u

otra tecnología inteligente, cuatro porque los grupos no siguieron el mismo método de aprendizaje y 17 por datos insuficientes. La muestra final declarada fue de 51 estudios publicados entre noviembre de 2022 y febrero de 2025.

Figura 1

Diagrama de flujo PRISMA 2020 de la selección de estudios



Nota. La categoría de 62 eliminaciones no desagregadas se obtiene por diferencia aritmética entre 350 registros eliminados y los 288 motivos especificados. Elaboración propia a partir de los datos de búsqueda y selección.

3.2. Características de la evidencia

La evidencia se concentró en educación superior. La matriz incluyó un estudio de primaria, varios de secundaria y una mayoría universitaria; no hubo experiencias de educación infantil. Los cursos se agruparon en STEM, lenguas y escritura académica, y formación de habilidades o competencias. Las intervenciones abarcaron desde actividades breves de una semana o menos hasta usos superiores a ocho semanas, con predominio de aprendizaje personalizado, basado en problemas, por proyectos, contextual, reflexivo o mixto.

ChatGPT desempeñó funciones de tutor, compañero, herramienta de aprendizaje o combinación de roles. Las aplicaciones incluyeron tutoría, evaluación y retroalimentación, recomendaciones personalizadas y usos mixtos. La matriz no consignó país ni una descripción narrativa homogénea de la intervención para cada fila; por rigor, esas variables no se completaron mediante inferencias geográficas. La tabla 1 reproduce las características

disponibles y formula la intervención de manera descriptiva a partir del modelo, el rol y el área codificados.

3.3. Efectos globales y heterogeneidad

El efecto global sobre el rendimiento académico fue significativo y positivo: $g = 0.867$, IC del 95 % [0.683, 1.052], $Z = 8.458$, $p < .001$. La heterogeneidad fue alta, $Q = 399.725$, $gl = 43$, $p < .001$, $I^2 = 89.243$ %. El valor Q de 399.725 corrige una mención textual inconsistente de 339.725, porque solo el primero coincide con la tabla analítica y con el I^2 informado. El resultado sugiere una mejora promedio relevante, aunque con variaciones sustanciales entre intervenciones.

La percepción del aprendizaje presentó un efecto moderado y positivo: $g = 0.456$, IC del 95 % [0.287, 0.625], $Z = 5.289$, $p < .001$. La heterogeneidad fue moderada y próxima al límite superior de esa categoría, $Q = 65.646$, $gl = 18$, $p < .001$, $I^2 = 72.580$ %. En conjunto, ChatGPT se asoció con mejoras en actitudes, interés, autoeficacia u otras valoraciones del aprendizaje, aunque la magnitud fue menor que la observada en desempeño (Venkatesh & Bala, 2008).

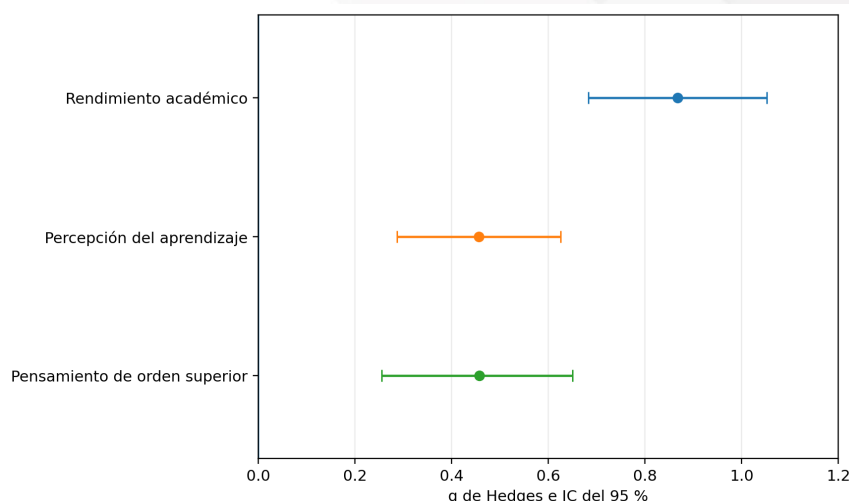
El pensamiento de orden superior alcanzó igualmente un efecto moderado: $g = 0.457$, IC del 95 % [0.255, 0.650], $Z = 4.471$, $p < .001$. La heterogeneidad fue moderada, $Q = 18.105$, $gl = 8$, $p < .05$, $I^2 = 55.813$ %. Este resultado indica que el uso de ChatGPT puede apoyar procesos de creatividad, pensamiento crítico, metacognición o resolución de problemas, pero no autoriza a suponer que la herramienta los desarrolla automáticamente en cualquier actividad (Li et al., 2024; Lu et al., 2024).

Tabla 2
Efectos globales de ChatGPT sobre los resultados educativos

Resultado	k	g	IC 95 %	Z	p	Q	I²
Rendimiento académico	44	0.867	[0.683, 1.052]	8.458	< .001	399.725	89.243 %
Percepción del aprendizaje	19	0.456	[0.287, 0.625]	5.289	< .001	65.646	72.580 %
Pensamiento de orden superior	9	0.457	[0.255, 0.650]	4.471	< .001	18.105	55.813 %

Nota. k = número de estimaciones declaradas; g = g de Hedges; IC = intervalo de confianza. Se aplicaron modelos de efectos aleatorios (Hedges, 1981; Higgins et al., 2003).

Figura 2
Tamaños de efecto globales e intervalos de confianza



3.4. Sesgo de publicación y sensibilidad

Los gráficos de embudo mostraron una distribución relativamente equilibrada, aunque algunos puntos quedaron fuera de los límites esperados. El N de seguridad fue 5490 para rendimiento, 399 para percepción y 101 para pensamiento de orden superior, cifras superiores a los umbrales $5k + 10$ de 230, 105 y 55, respectivamente. Estos resultados reducen la preocupación de que pocos estudios no publicados anulen los efectos, pero no sustituyen una evaluación formal y completa del sesgo de publicación (Borenstein et al., 2009; Schulz et al., 1995).

El análisis de eliminación de un estudio mostró que los intervalos de confianza permanecieron positivos bajo modelos de efectos fijos y aleatorios. En rendimiento, el intervalo del modelo aleatorio se mantuvo entre 0.683 y 1.052; en percepción, entre 0.287 y 0.625; y en pensamiento de orden superior, entre 0.255 y 0.650. La estabilidad indica que ningún estudio individual modificó por sí solo la dirección general de los resultados (Borenstein et al., 2009).

3.5. Moderadores del rendimiento académico

El tipo de curso moderó significativamente el rendimiento, $Q_B = 64.249$, $p < .001$. El mayor efecto se observó en formación de habilidades y competencias, $g = 0.874$, seguido por aprendizaje de lenguas y escritura académica, $g = 0.531$, y STEM, $g = 0.311$. La ventaja en cursos de competencias puede relacionarse con tareas delimitadas, procedimientos observables y oportunidades de retroalimentación inmediata, condiciones en las que ChatGPT puede proporcionar guía específica y ejemplos adaptados (Rakap, 2023; Urban et al., 2024).

El modelo de aprendizaje también fue significativo, $Q_B = 76.220$, $p < .001$. El aprendizaje basado en problemas obtuvo el efecto más alto, $g = 1.113$; le siguieron el aprendizaje reflexivo, $g = 0.866$, el personalizado, $g = 0.535$, el contextual, $g = 0.517$, el mixto, $g = 0.330$, y el basado en proyectos, $g = 0.239$. La superioridad del enfoque basado en problemas es coherente con la

capacidad de ChatGPT para ofrecer explicaciones, hipótesis y rutas de solución, mientras que los proyectos complejos requieren integración, ejecución y validación que exceden la generación conversacional de respuestas (Küchemann et al., 2023; Urban et al., 2024).

La duración produjo diferencias significativas, $Q_B = 55.998$, $p < .001$. Las intervenciones de cuatro a ocho semanas alcanzaron $g = 0.999$; las superiores a ocho semanas, $g = 0.531$; las de una a cuatro semanas, $g = 0.428$; y las de una semana o menos, $g = 0.332$. El patrón sugiere que se requiere tiempo para aprender a formular instrucciones, evaluar respuestas e incorporar la herramienta a rutinas de estudio, pero una exposición prolongada sin renovación pedagógica puede favorecer dependencia o uso superficial (Sung et al., 2017; Tian & Zheng, 2023; White et al., 2023).

No se observaron diferencias significativas por nivel educativo, $Q_B = 0.627$, $p = .428$; rol de ChatGPT, $Q_B = 6.261$, $p = .100$; ni área de aplicación, $Q_B = 6.210$, $p = .102$. El estudio de primaria fue excluido del análisis de nivel por constituir una sola observación. La falta de significación no demuestra equivalencia absoluta, sino que los datos disponibles fueron insuficientes para detectar diferencias consistentes en esas categorías.

3.6. Moderadores de la percepción y el pensamiento de orden superior

En percepción del aprendizaje, la duración fue el único moderador significativo, $Q_B = 19.839$, $p < .001$. Los efectos aumentaron desde $g = 0.203$ en una semana o menos hasta $g = 0.404$ entre una y cuatro semanas, $g = 0.493$ entre cuatro y ocho semanas y $g = 1.054$ por encima de ocho semanas. El uso sostenido puede mejorar familiaridad, utilidad percibida y confianza, siempre que la experiencia continúe proporcionando retroalimentación relevante y no se reduzca a una rutina mecánica (Guan et al., 2024; Venkatesh & Bala, 2008).

No fueron significativos para percepción el nivel educativo, el tipo de curso, el modelo de aprendizaje, el rol ni el área de aplicación. En pensamiento de orden superior, todos los estudios procedieron de educación superior, por lo que no se analizó el nivel. El tipo de curso sí produjo diferencias, $Q_B = 7.811$, $p < .05$: STEM alcanzó $g = 0.737$, lenguas y escritura $g = 0.332$, y habilidades y competencias $g = 0.295$.

El rol de ChatGPT moderó el pensamiento de orden superior, $Q_B = 4.872$, $p < .05$. Como tutor inteligente alcanzó $g = 0.945$, mientras que como herramienta inteligente de aprendizaje obtuvo $g = 0.428$. Los roles mixtos y de compañero fueron excluidos por contar con una sola estimación. El modelo de aprendizaje, la duración y el área de aplicación no alcanzaron

significación estadística, aunque el reducido número de estimaciones limita la potencia de estas comparaciones (Borenstein et al., 2009; Escalante et al., 2023).

Tabla 3

Síntesis de moderadores estadísticamente significativos

<u>Resultado</u>	<u>Moderador</u>	<u>Q B</u>	<u>Categoría con mayor efecto</u>
Rendimiento	Tipo de curso	64.249***	Habilidades y competencias, $g = 0.874$
Rendimiento	Modelo de aprendizaje	76.220***	Basado en problemas, $g = 1.113$
Rendimiento	Duración	55.998***	4–8 semanas, $g = 0.999$
Percepción	Duración	19.839***	> 8 semanas, $g = 1.054$
Pensamiento de orden superior	Tipo de curso	7.811*	STEM, $g = 0.737$
Pensamiento de orden superior	Rol de ChatGPT	4.872*	Tutor inteligente, $g = 0.945$

Nota. * $p < .05$. *** $p < .001$. Los restantes moderadores no alcanzaron significación estadística (Borenstein et al., 2009).

4. DISCUSIÓN

4.1. Interpretación de los hallazgos

El hallazgo principal es que ChatGPT se asocia con una mejora grande del rendimiento académico y con mejoras moderadas de la percepción del aprendizaje y el pensamiento de orden superior. La diferencia entre resultados sugiere que la herramienta facilita con mayor rapidez tareas observables —comprender conceptos, corregir textos, acceder a ejemplos o resolver problemas— que procesos afectivos o cognitivos complejos, los cuales requieren tiempo, interacción humana, práctica deliberada y criterios de evaluación (Deng et al., 2025; Li et al., 2024; Lu et al., 2024).

Desde el constructivismo, el efecto positivo no debe atribuirse a la generación automática de contenido, sino a la interacción que se establece alrededor de ese contenido. ChatGPT puede actuar como andamiaje cuando formula preguntas, ofrece retroalimentación, presenta alternativas y permite revisar explicaciones; pierde valor cuando reemplaza la elaboración del estudiante. La magnitud superior en aprendizaje basado en problemas respalda la idea de que el sistema resulta más útil cuando se integra a una tarea que exige interpretar, decidir y aplicar conocimientos (Piaget, 1970; Urban et al., 2024; Vygotskij & Hanfmann, 1966).

La duración óptima de cuatro a ocho semanas para rendimiento puede interpretarse como un periodo de apropiación pedagógica. Al inicio, los estudiantes todavía desarrollan habilidades para formular preguntas y evaluar respuestas; posteriormente, la práctica permite integrar la herramienta a estrategias de estudio. Cuando la intervención supera ocho semanas, el descenso

relativo del efecto puede reflejar habituación, dependencia o falta de renovación de las actividades. Este patrón no aconseja limitar mecánicamente el uso, sino variar tareas, incorporar comprobaciones y revisar periódicamente la autonomía del estudiante (Sung et al., 2017; White et al., 2023).

4.2. Percepción del aprendizaje

El efecto moderado sobre percepción indica que la rapidez y la personalización pueden mejorar la experiencia, pero no sustituyen la dimensión relacional de la enseñanza. ChatGPT puede aumentar utilidad percibida, control y confianza cuando responde de forma inmediata; también puede generar frustración cuando ofrece información errónea, respuestas genéricas o retroalimentación sin sensibilidad contextual. La mejora progresiva con la duración sugiere que la aceptación se construye mediante experiencias consistentes de utilidad y no solo por novedad tecnológica (Birenbaum, 2023; Venkatesh & Bala, 2008).

4.3. Pensamiento de orden superior

El efecto moderado sobre pensamiento de orden superior confirma un potencial condicionado. En STEM, ChatGPT puede apoyar la comparación de procedimientos, la generación de hipótesis y la exploración de soluciones, lo que explica el efecto mayor. Sin embargo, la herramienta no posee una garantía interna de verdad ni de rigor argumentativo; por ello, las actividades deben exigir evidencias, contraste con fuentes, detección de errores y justificación de decisiones (Kasneci et al., 2023; Li et al., 2024).

El efecto especialmente alto cuando ChatGPT actúa como tutor sugiere que la estructura de la interacción importa. Un tutor no se limita a entregar una respuesta: orienta, pregunta, retroalimenta y ayuda a regular el proceso. Para aproximarse a esa función, las consignas deben pedir al sistema que explique pasos, proponga contraejemplos, cuestione supuestos y ofrezca pistas graduales, mientras el docente verifica la pertinencia de la retroalimentación y evita delegar la evaluación final (Escalante et al., 2023; Lu et al., 2024).

4.4. Implicaciones para la práctica y la política educativa

En el aula, ChatGPT debería incorporarse como recurso subordinado a objetivos de aprendizaje explícitos. Antes de usarlo, el docente necesita definir qué proceso cognitivo se pretende promover, qué evidencias demostrarán aprendizaje y qué parte de la tarea debe realizarse sin asistencia. En actividades de resolución de problemas, la herramienta puede generar casos, ofrecer pistas o comparar estrategias; en escritura, puede apoyar revisión y retroalimentación,

siempre que el estudiante conserve la responsabilidad autoral y justifique los cambios (Escalante et al., 2023; Song & Song, 2023; Urban et al., 2024).

El andamiaje debe incluir alfabetización en instrucciones, verificación de información, citación, integridad académica y reconocimiento de incertidumbre. Los estudiantes necesitan aprender que una respuesta fluida no equivale a una respuesta correcta. Una secuencia adecuada puede pedir primero una solución individual, luego una consulta a ChatGPT, después una comparación crítica y, finalmente, una versión revisada con explicación de las decisiones. Esta estructura transforma la herramienta en objeto de análisis y reduce el riesgo de sustitución cognitiva (Gallent-Torres et al., 2023; Kasneci et al., 2023; White et al., 2023).

La formación docente debe superar la capacitación meramente operativa. Se requiere diseño de actividades, selección de roles, evaluación de calidad, gestión de privacidad y análisis de sesgos. El hallazgo favorable para intervenciones de cuatro a ocho semanas permite planificar unidades didácticas en las que el uso se introduzca gradualmente, se acompañe con evaluación formativa y se revise con evidencias de progreso. En periodos más largos conviene variar la función de ChatGPT y alternar tareas asistidas y no asistidas (Lu et al., 2024; Sung et al., 2017).

Las políticas públicas deberían promover investigación longitudinal, recursos abiertos de alfabetización en inteligencia artificial y marcos de evaluación que combinen aprendizaje, bienestar, equidad e integridad. La evidencia disponible se concentra en educación superior y en intervenciones breves, por lo que no permite generalizar con seguridad a educación infantil, primaria o contextos con menor infraestructura. Cualquier expansión debe acompañarse de evaluación local y de mecanismos para retirar o rediseñar usos que no produzcan beneficios verificables (Hays et al., 2024; Zhang & Tur, 2024).

4.5. Limitaciones

La revisión presenta limitaciones de cobertura y calidad. Solo se incluyeron estudios en inglés y no se incluyó literatura gris. La concentración en educación superior, la presencia de un único estudio de primaria y la ausencia de educación infantil restringen la generalización. Además, percepción y pensamiento de orden superior se apoyaron en 19 y 9 estimaciones, respectivamente, muestras limitadas para análisis extensos de moderación (Page et al., 2021; Zhang & Tur, 2024).

La heterogeneidad fue alta en rendimiento, lo que indica que el promedio sintetiza intervenciones diferentes. Tampoco se examinaron variables como país, cultura, experiencia

previa, política institucional o características del docente, porque no estaban disponibles de forma consistente (Borenstein et al., 2009; Higgins et al., 2003).

5. CONCLUSIONES

ChatGPT mostró un efecto significativo sobre el rendimiento académico y efectos moderados sobre la percepción del aprendizaje y el pensamiento de orden superior. La eficacia no fue uniforme: el rendimiento varió según tipo de curso, modelo y duración; la percepción dependió de la duración; y el pensamiento de orden superior cambió según tipo de curso y rol de la herramienta. Los mejores resultados se asociaron con aprendizaje basado en problemas, intervenciones de cuatro a ocho semanas y uso de ChatGPT como tutor en tareas cognitivamente exigentes.

La conclusión principal es que ChatGPT no constituye una intervención autosuficiente. Su contribución depende de objetivos claros, tareas auténticas, andamiaje, verificación de respuestas, evaluación formativa y supervisión docente. La herramienta puede ampliar oportunidades de práctica y retroalimentación, pero también facilitar dependencia, respuestas acríticas o pérdida de autoría cuando se emplea sin regulación pedagógica.

La aplicación responsable exige combinar actividades asistidas y no asistidas, evaluar transferencia y autonomía, y adaptar las políticas al contexto. La evidencia respalda una integración prudente y deliberada, no una adopción generalizada por novedad. Futuras revisiones deberían registrar protocolos, ampliar idiomas y niveles educativos, aplicar herramientas formales de riesgo de sesgo, depurar la matriz de extracción y publicar datos y códigos para que los resultados puedan ser verificados y actualizados.

Acknowledgments

Research ethics: Not applicable. **Informed consent:** Not applicable. **Research funding:** None declared. **Note on the nature and source of this paper:** The present paper constitutes an academic translation and synthesis of the article by Wang and Fan (2025). Accordingly, this work is not presented as original or previously unpublished research, nor are the results, data, analyses, or conclusions of the source article claimed as the author's own. It has been prepared exclusively for academic, educational, and illustrative purposes, as part of the requirements for the High-Level Professional Articles submitted prior to the award of the Master's Degree in Education, with a specialisation in Higher Education Teaching and Research. The original

source is expressly acknowledged and constitutes the substantive basis of the material presented herein. **Author contributions:** PV is the sole author of the present translation and academic synthesis and was exclusively responsible for its preparation, translation, synthesis, and final revision. PV has read and approved the final manuscript and assumes responsibility for the accuracy and integrity of the translation and synthesis. JW and WF are the authors of the original source article and are not authors of the present work. **Competing interests:** The author declares no competing interests. **Use of large language models, AI and machine learning tools:** Generative artificial intelligence tools were used for literature searches, preliminary organization, linguistic correction, translation, and auxiliary checks. Web of Science Research Assistant, Scopus AI, and ScienceDirect AI in LeapSpace supported literature discovery; OpenAI's ChatGPT and Google's NotebookLM assisted with outlining, preliminary syntheses, and consistency checks; Paperpal and QuillBot were used for stylistic review; and DeepL Translator was used for subsequently verified translations. **Data, code, and materials availability:** Not applicable. **Supplementary materials:** Not applicable.

References

- Ali, D., Fatemi, Y., Boskabadi, E., Nikfar, M., Ugwuoke, J., & Ali, H. (2024). ChatGPT in teaching and learning: A systematic review. *Education Sciences*, 14(6), 643. <https://doi.org/10.3390/educsci14060643>
- Alneyadi, S., & Wardat, Y. (2023). ChatGPT: Revolutionizing student achievement in the electronic magnetism unit for eleventh-grade students in Emirates schools. *Contemporary Educational Technology*, 15(4). <https://doi.org/10.30935/cedtech/13417>
- Birenbaum, M. (2023). The chatbots' challenge to education: Disruption or destruction? *Education Sciences*, 13(7), 711. <https://doi.org/10.3390/educsci13070711>
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Wiley.
- Deng, R., Jiang, M., Yu, X., Lu, Y., & Liu, S. (2025). Does ChatGPT enhance student learning? A systematic review and meta-analysis of experimental studies. *Computers & Education*, 227, 105224. <https://doi.org/10.1016/j.compedu.2024.105224>
- Donald, M. J., Will, D., & Christopher, M. E. (2024). Using ChatGPT with novice Arduino programmers: Effects on performance, interest, self-efficacy, and programming ability.

- Journal of Research in Technical Careers, 8(1), 1–17. <https://doi.org/10.9741/2578-2118.1152>
- Escalante, J., Pack, A., & Barrett, A. (2023). AI-generated feedback on writing: Insights into efficacy and ENL student preference. *International Journal of Educational Technology in Higher Education*, 20(1). <https://doi.org/10.1186/s41239-023-00425-2>
- Gallent-Torres, C., Zapata-González, A., & Ortego-Hernando, J. L. (2023). The impact of generative artificial intelligence in higher education: A focus on ethics and academic integrity. *RELIEVE*, 29(2). <https://doi.org/10.30827/relieve.v29i2.29134>
- Guan, L., Li, S., & Gu, M. M. (2024). AI in informal digital English learning: A meta-analysis of its effectiveness on proficiency, motivation, and self-regulation. *Computers & Education: Artificial Intelligence*, 7, 100323. <https://doi.org/10.1016/j.caeai.2024.100323>
- Hays, L., Jurkowski, O., & Sims, S. K. (2024). ChatGPT in K–12 education. *TechTrends*, 68(2), 281–294. <https://doi.org/10.1007/s11528-023-00924-z>
- Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics*, 6(2), 107–128. <https://doi.org/10.2307/1164588>
- Heung, Y. M. E., & Chiu, T. K. F. (2025). How ChatGPT impacts student engagement from a systematic review and meta-analysis study. *Computers & Education: Artificial Intelligence*, 8, 100361. <https://doi.org/10.1016/j.caeai.2025.100361>
- Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *BMJ*, 327(7414), 557–560. <https://doi.org/10.1136/bmj.327.7414.557>
- Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, 102274. <https://doi.org/10.1016/j.lindif.2023.102274>
- Küchemann, S., Steinert, S., Revenga, N., Schweinberger, M., Dinc, Y., Avila, K. E., & Kuhn, J. (2023). Can ChatGPT support prospective teachers in physics task development? *Physical Review Physics Education Research*, 19(2), 020128. <https://doi.org/10.1103/PhysRevPhysEducRes.19.020128>

- Li, H.-F. (2023). Effects of a ChatGPT-based flipped learning guiding approach on learners' courseware project performances and perceptions. *Australasian Journal of Educational Technology*, 39(5), 40–58. <https://doi.org/10.14742/ajet.8923>
- Li, T., Ji, Y., & Zhan, Z. (2024). Expert or machine? Comparing the effect of pairing student teacher with in-service teacher and ChatGPT on their critical thinking, learning performance, and cognitive load in an integrated-STEM course. *Asia Pacific Journal of Education*, 44(1), 45–60. <https://doi.org/10.1080/02188791.2024.2305163>
- Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Sage.
- Lu, J., Zheng, R., Gong, Z., & Xu, H. (2024). Supporting teachers' professional development with generative AI: The effects on higher-order thinking and self-efficacy. *IEEE Transactions on Learning Technologies*, 17, 1279–1289. <https://doi.org/10.1109/TLT.2024.3369690>
- Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Annals of Internal Medicine*, 151(4), 264–269. <https://doi.org/10.7326/0003-4819-151-4-200908180-00135>
- Niloy, A. C., Akter, S., Sultana, N., Sultana, J., & Rahman, S. I. U. (2023). Is ChatGPT a menace for creative writing ability? An experiment. *Journal of Computer Assisted Learning*, 40(2), 919–930. <https://doi.org/10.1111/jcal.12929>
- OpenAI. (2022). *Introducing ChatGPT*.
- Page, M. J., & Moher, D. (2017). Evaluations of the uptake and impact of the PRISMA statement and extensions: A scoping review. *Systematic Reviews*, 6(1), 263. <https://doi.org/10.1186/s13643-017-0663-8>
- 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. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372, n71. <https://doi.org/10.1136/bmj.n71>
- Piaget, J. (1970). *Structuralism*. Basic Books.
- Rakap, S. (2023). Chatting with GPT: Enhancing individualized education program goal development for novice special education teachers. *Journal of Special Education Technology*, 39(3), 339–348. <https://doi.org/10.1177/01626434231211295>

- Sawilowsky, S. S. (2009). New effect size rules of thumb. *Journal of Modern Applied Statistical Methods*, 8(2), 597–599. <https://doi.org/10.22237/jmasm/1257035100>
- Schulz, K. F., Chalmers, I., Hayes, R. J., & Altman, D. G. (1995). Empirical evidence of bias: Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA*, 273(5), 408–412. <https://doi.org/10.1001/jama.273.5.408>
- Song, C., & Song, Y. (2023). Enhancing academic writing skills and motivation: Assessing the efficacy of ChatGPT in AI-assisted language learning for EFL students. *Frontiers in Psychology*, 14, 1260843. <https://doi.org/10.3389/fpsyg.2023.1260843>
- Sung, Y.-T., Yang, J.-M., & Lee, H.-Y. (2017). The effects of mobile-computer-supported collaborative learning: Meta-analysis and critical synthesis. *Review of Educational Research*, 87(4), 768–805. <https://doi.org/10.3102/0034654317704307>
- Tian, Q., & Zheng, X. (2023). Effectiveness of online collaborative problem-solving method on students' learning performance: A meta-analysis. *Journal of Computer Assisted Learning*, 40(1), 326–341. <https://doi.org/10.1111/jcal.12884>
- Urban, M., Dechterenko, F., Lukavský, J., Hrabalová, V., Svacha, F., Brom, C., & Urban, K. (2024). ChatGPT improves creative problem-solving performance in university students: An experimental study. *Computers & Education*, 215, 105031. <https://doi.org/10.1016/j.compedu.2024.105031>
- Venkatesh, V., & Bala, H. (2008). Technology acceptance model 3 and a research agenda on interventions. *Decision Sciences*, 39(2), 273–315. <https://doi.org/10.1111/j.1540-5915.2008.00192.x>
- Vygotskij, L. S., & Hanfmann, E. (1966). *Thought and language*. MIT Press.
- Wang, J. and Fan, W. (2025), The effect of ChatGPT on students' academic performance, learning perception, and higher-order thinking: insights from a meta-analysis, published in *Humanities and Social Sciences Communications*, 12, 621. <https://doi.org/10.1057/s41599-025-04787-y>.
- White, J., Fu, Q., Hays, S., Sandborn, M., Olea, C., Gilbert, H., Elnashar, A., Spencer-Smith, J., & Schmidt, D. C. (2023). A prompt pattern catalog to enhance prompt engineering with ChatGPT. In *Proceedings of the 30th Conference on Pattern Languages of Programs*. The Hillside Group.
- Yan, L., Sha, L., Zhao, L., Li, Y., Martinez-Maldonado, R., Chen, G., Li, X., Jin, Y., & Gašević, D. (2024). Practical and ethical challenges of large language models in education: A

systematic scoping review. *British Journal of Educational Technology*, 55(1), 90–112.
<https://doi.org/10.1111/bjet.13370>

Yang, T.-C., Hsu, Y.-C., & Wu, J.-Y. (2025). The effectiveness of ChatGPT in assisting high school students in programming learning: Evidence from a quasi-experimental research. *Interactive Learning Environments*. <https://doi.org/10.1080/10494820.2025.2450659>

Zhang, P., & Tur, G. (2024). A systematic review of ChatGPT use in K–12 education. *European Journal of Education*, 59(2). <https://doi.org/10.1111/ejed.12599>

Zheng, L., Niu, J., Zhong, L., & Gyasi, J. F. (2021). The effectiveness of artificial intelligence on learning achievement and learning perception: A meta-analysis. *Interactive Learning Environments*, 31(9), 5650–5664. <https://doi.org/10.1080/10494820.2021.2015693>.



INSTITUTO STVDIA HVMANITATIS
Pasaje de Salamanca 5,1º,5
28702 San Sebastián de los Reyes
Madrid, España
studiahumanitatis.eu

May 15, 2026

Paulo Vélez León, MA.
Universidad Estatal de Milagro – UNEMI
Milagro, Ecuador. —

Ref.: Decision on submission to Disputatio (Spain)

We are delighted to let you know that the manuscript entitled "**Análisis de los efectos de ChatGPT en el aprendizaje de los estudiantes: una revisión sistemática / Analysis of the effects of ChatGPT on student's learning: a systematic review**" by Paulo Vélez León, on which you were a Contributing Author, has been **accepted** for publication in *Disputatio* (Spain, ISSN: 2254-0601). This decision has been supported by a thorough and favourable review by our expert referees.

As the corresponding author of an accepted manuscript, your next steps will be to complete an Open Access Licence to publish on behalf of all authors, confirm your institutional affiliation, and arrange payment of your article processing charge (APC).

Once we've prepared your manuscript for publication, you will receive a proof. For the main text, only errors that have been introduced during the production process, or those that directly compromise the scientific integrity of the paper, may be corrected. As the corresponding author, you are responsible for the accuracy of all content, including spelling of names and current affiliations.

To ensure prompt publication, your proofs should be returned within two working days. The final version of your manuscript and the completed editorial sheet must be submitted to our official email (disputatio@studiahumanitatis.eu). Publication is typically within one to two weeks after the proofing process and the publishing agreements have been completed. Please note that your article cannot proceed to publication if the publishing agreements and proofing process have not been completed.

Within few days, you will receive an email with payment options and instructions from Studia Humanitatis e-commerce partner, Ediciones Studia-Humanitatis S.A.S. You will be able to choose either to pay by credit card or bank transfer. If you are not making the payment yourself, you may forward the email to the person or organisation that will be paying on your behalf. Your article will not be processed by production until you have paid the article processing charge.

Our journal is covered by the following indexing services: Emerging Sources Citation Index - ESCI (JCI, Q3/WoS), Scopus (Q3, Elsevier), Qualis (A1, Capes), Philosopher's Index with Full Text (EBSCO), MLA International Bibliography (EBSCO), ProQuest Central, Philosophy Collection, Philosophy Research Collection (Clarivate Analytics - Proquest), PhilPapers, Linguistic Bibliography (Brill), ERIH+ Plus, among others.

We appreciate your trust in Disputatio and look forward to future collaboration.

Warm regards, and we await your response with our best wishes.

Dr. Alba Moreira Pinargote
STUDIA HUMANITATIS

UNEMI

UNIVERSIDAD ESTATAL DE MILAGRO

¡Evolución académica!

@UNEMIEcuador

