

**LATAM Revista Latinoamericana de Ciencias Sociales y
Humanidades, Asunción, Paraguay**

ISSN en línea: 2789-3855, 2026

Inteligencia artificial generativa y aprendizaje matemático: revisión sistemática de variables mediadoras cognitivas, metacognitivas y afectivas

Generative artificial intelligence and mathematics learning: systematic review of cognitive, metacognitive, and affective mediating variables

Melvin Ramírez Bogantes

meramirez@itcr.ac.cr

<https://orcid.org/0000-0001-5516-0085>

Tecnológico de Costa Rica

Heredia – Costa Rica

DOI: <https://doi.org/10.56712/latam.v7i2.5747>

Artículo recibido: 18 de diciembre de 2025.

Aceptado para publicación: 24 de abril de 2026.

Conflictos de Interés: Ninguno que declarar.

**Redilat**
Red de Investigadores
Latinoamericanos

**LATAM**
Revista Latinoamericana de
Ciencias Sociales y Humanidades

VOLUMEN VII

DOI: <https://doi.org/10.56712/latam.v7i2.5747>

Inteligencia artificial generativa y aprendizaje matemático: revisión sistemática de variables mediadoras cognitivas, metacognitivas y afectivas

Generative artificial intelligence and mathematics learning: systematic
review of cognitive, metacognitive, and affective mediating variables

Melvin Ramírez Bogantes

meramirez@itcr.ac.cr

<https://orcid.org/0000-0001-5516-0085>

Tecnológico de Costa Rica

Heredia – Costa Rica

Artículo recibido: 18 de diciembre de 2025. Aceptado para publicación: 24 de abril de 2026.
Conflictos de Interés: Ninguno que declarar.

Resumen

La integración de la inteligencia artificial generativa en la educación matemática ha producido resultados científicos heterogéneos desde el año 2022. Se realizó una revisión sistemática con protocolo PRISMA 2020 que sintetizó 165 artículos publicados entre 2020 y primer trimestre 2026, identificados en siete bases de datos: ERIC, Dialnet, Redalyc, Scopus, Web of Science, IEEE Xplore y SciELO. El análisis se estructuró en cuatro preguntas de síntesis: efectos cognitivos, metacognitivos, afectivos y moderadores del impacto. Los resultados revelan un efecto positivo moderado sobre el rendimiento cuando el diseño pedagógico incorpora salvaguardas ($g = 0.534$; Liu et al., 2026), pero un efecto negativo significativo cuando el acceso es irrestricto (-17% ; Bastani et al., 2025). Se identificó la paradoja de la autorregulación: los estudiantes más vulnerables son quienes más necesitan y más riesgo tienen con la herramienta sin soporte metacognitivo. Los efectos afectivos muestran reducción consistente de la ansiedad matemática en tres contextos culturales distintos, aunque la motivación global no alcanza significación estadística. Se propone el modelo de impacto mediado: inteligencia artificial generativa \rightarrow [diseño pedagógico \times autorregulación del aprendizaje \times alfabetización en inteligencia artificial] \rightarrow aprendizaje matemático, con nueve moderadores identificados.


Palabras clave: inteligencia artificial generativa, educación matemática, revisión sistemática, autorregulación del aprendizaje, ansiedad matemática

Abstract

The integration of generative artificial intelligence in mathematics education has produced heterogeneous scientific results since 2022. A PRISMA 2020 systematic review synthesized 165 articles published between 2020 and the first quarter of 2026, identified in seven databases: ERIC, Dialnet, Redalyc, Scopus, Web of Science, IEEE Xplore, and SciELO. Analysis was organized around four synthesis questions: cognitive, metacognitive, and affective effects, and moderators of impact. Results reveal a moderate positive effect on performance when pedagogical design incorporates guardrails ($g = 0.534$; Liu et al., 2026), but a significant negative effect when access is unrestricted (-17% ; Bastani et al., 2025). The self-regulation paradox is identified: the most vulnerable students need the tool most and are most at risk without metacognitive support. Affective effects show consistent mathematics anxiety reduction across three cultural contexts, though global motivation

effects are not statistically significant. A mediated impact model is proposed: generative AI → [pedagogical design × self-regulated learning × AI literacy] → mathematics learning, with nine identified moderators.

Keywords: generative artificial intelligence, mathematics education, systematic review, self-regulated learning, mathematics anxiety

Todo el contenido de LATAM Revista Latinoamericana de Ciencias Sociales y Humanidades, publicado en este sitio está disponibles bajo Licencia Creative Commons. 

Cómo citar: Ramírez Bogantes, M. (2026). Inteligencia artificial generativa y aprendizaje matemático: revisión sistemática de variables mediadoras cognitivas, metacognitivas y afectivas. *LATAM Revista Latinoamericana de Ciencias Sociales y Humanidades* 7 (2), 1647 – 1661.
<https://doi.org/10.56712/latam.v7i2.5747>

INTRODUCCIÓN

Contexto y justificación

El lanzamiento de ChatGPT en noviembre de 2022 inauguró un período de transformación acelerada en la educación matemática. Por primera vez, un sistema de inteligencia artificial generativa basado en un modelo de lenguaje de gran escala (LLM) estuvo disponible de forma masiva, gratuita e inmediata para estudiantes y docentes de todos los niveles. Desde entonces, la producción científica sobre el tema ha crecido de forma acelerada: la búsqueda sistemática realizada para este estudio identificó 535 registros en siete bases de datos internacionales, reflejando la heterogeneidad metodológica y la frecuente contradicción de resultados que caracterizan un campo en formación

La educación matemática latinoamericana enfrenta este desafío tecnológico en un contexto de brechas estructurales que amplifican tanto las oportunidades como los riesgos. La posibilidad de acceder a retroalimentación personalizada e inmediata podría democratizar el apoyo pedagógico; sin embargo, si la integración no va acompañada de diseño instruccional adecuado y formación docente sólida, existe el riesgo de ampliar las desigualdades educativas existentes.

METODOLOGÍA

Enfoque y diseño

Esta revisión sistemática sigue el estándar PRISMA 2020 (Page et al., 2021). El diseño es de tipo síntesis mixta: cuantitativa para los meta-análisis disponibles en el corpus y narrativa estructurada para los estudios primarios. El protocolo fue definido antes de ejecutar las búsquedas, garantizando la independencia entre el diseño metodológico y los resultados hallados.

Estrategia de búsqueda y fuentes

Las búsquedas sistemáticas se ejecutaron en siete bases de datos internacionales: ERIC (acceso abierto), Dialnet (acceso abierto), Redalyc (acceso abierto), Scopus (acceso institucional TEC), Web of Science (acceso institucional TEC), IEEE Xplore (acceso institucional TEC) y SciELO (acceso abierto). La búsqueda cubrió el período enero de 2020 a marzo de 2026, ejecutada en cuatro rondas entre enero y marzo de 2026.

Las ecuaciones de búsqueda combinaron los términos: ("generative artificial intelligence" OR "ChatGPT" OR "large language model" OR "LLM" OR "inteligencia artificial generativa") AND ("mathematics education" OR "educación matemática" OR "mathematics learning" OR "aprendizaje matemático") AND ("cognitive" OR "metacognitive" OR "affective" OR "anxiety" OR "self-regulated learning" OR "motivation"). Las ecuaciones se adaptaron a la sintaxis específica de cada base. La búsqueda identificó 535 registros brutos; tras eliminar 10 duplicados entre bases, quedaron 525 registros únicos para cribado.

Criterios de elegibilidad

La Tabla 1 presenta los criterios de inclusión y exclusión aplicados. Su definición siguió las recomendaciones PICOS (Población, Intervención, Comparación, Resultado, Diseño) adaptadas a revisiones en educación tecnológica.

Tabla 1

Criterios de inclusión y exclusión – protocolo PRISMA 2020

Tipo	Código	Criterio
INCLUSIÓN	CI1	Período 2020-2026
	CI2	Participantes: estudiantes de secundaria, universitaria o docentes en formación en contexto de educación matemática
	CI3	Intervención: sistema de inteligencia artificial generativa (LLM: ChatGPT, GPT-4, Gemini, Claude u otro modelo de lenguaje grande) usado en actividades matemáticas
	CI4	Resultado: al menos una variable cognitiva, metacognitiva o afectiva relacionada con matemáticas, O revisión sistemática/meta-análisis sobre el tema
	CI5	Documento revisado por pares en revista indexada
	CI6	Idioma: español o inglés
EXCLUSIÓN	CE1	Solo técnico: benchmarks sin estudiantes reales
	CE2	Sin foco matemático específico: IA generativa en otras disciplinas sin componente matemático
	CE3	Sin datos empíricos: editorial u opinión sin síntesis sistemática
	CE4	Revisión no sistemática: sin protocolo PRISMA o equivalente
	CE5	IA no generativa: sistemas expertos o tutores inteligentes pre-LLM
	CE6	Preprint sin revisión por pares certificada
	CE7	Duplicado entre bases de datos

Nota: CI = criterio de inclusión; CE = criterio de exclusión. Los criterios se aplicaron en orden secuencial, deteniéndose en el primer CE aplicable para cada registro.

Fuente: elaboración propia.

Flujo de selección

La Tabla 2 presenta el flujo completo de selección de estudios. El principal motivo de exclusión fue la ausencia de foco matemático específico (CE2, n = 223), seguido de la identificación de sistemas de inteligencia artificial no generativos, anteriores a los modelos de lenguaje grande (CE5, n = 89).

Tabla 2

Flujo de selección de estudios – PRISMA 2020

Fase	N	Descripción
Identificación total	535	Registros identificados en siete bases de datos (incluyendo duplicados entre bases)
Tras deduplicación	525	Registros únicos sometidos a cribado
Excluidos en cribado	360	CE2 (sin foco matemático, n=223); CE5 (no IA-G, n=89); CE1/CE3/CE4/CE6/CE7 (n=48)
ELEGIBLES	165	Artículos que cumplen todos los criterios CI1-CI6
Para síntesis narrativa	127	Estudios primarios cualitativos, mixtos y encuestas
Para síntesis cuantitativa	38	Meta-análisis y revisiones con datos de efecto

Nota: k = número de artículos; EM = educación matemática; HOT = pensamiento de orden superior. La categoría 'Síntesis cuantitativa' incluye únicamente artículos con datos de tamaño de efecto o estimaciones estadísticas suficientes para síntesis numérica.

Fuente: elaboración propia.

Extracción de datos y calidad metodológica

La extracción se realizó mediante un instrumento estructurado de 36 campos distribuidos en seis módulos: (A) identificación bibliográfica; (B) características del estudio; (C) descripción de la intervención con inteligencia artificial generativa; (D) variables de resultado y tamaños de efecto; (E) moderadores y contexto geográfico; y (F) evaluación de calidad metodológica. La calidad se evaluó con el instrumento CASP (Critical Appraisal Skills Programme), adaptado al tipo de diseño de cada artículo, en una escala de 0 a 20 puntos. El cribado inicial fue realizado por el investigador principal; un segundo revisor verificó de forma independiente una muestra aleatoria del 20% de las decisiones de inclusión, obteniendo un acuerdo del 91.3% ($\kappa = 0.84$).

La síntesis cuantitativa reporta los tamaños de efecto originales de los meta-análisis disponibles (Hedges' g, d de Cohen, SMD) sin reestimación propia. La síntesis narrativa organiza los hallazgos según las cuatro preguntas de síntesis, destacando patrones de convergencia y condiciones moderadoras.

Consideraciones éticas

Este estudio sintetiza la investigación publicada y revisada por pares. No involucra participantes humanos directos y no requirió aprobación del comité de ética. Las fuentes citadas se atribuyen con fidelidad a sus autores originales y se respetan los criterios de propiedad intelectual en todos los casos.

DESARROLLO

Los meta-análisis disponibles sobre inteligencia artificial generativa en educación general reportan tamaños de efecto positivos de magnitud moderada (SMD = 0.45 (Han et al., 2025) y $d = 0.68$ (Ma & Zhong, 2025)), pero incluyen las matemáticas como uno de varios dominios sin especificidad en los moderadores. El meta-análisis más específico en educación matemática (Liu et al., 2026) reporta $g = 0.534$ sobre 22 estudios primarios ($n = 5.232$), con el dominio de la geometría y el modo de transformación creativa como moderadores de mayor tamaño de efecto. En sentido contrario, el ensayo controlado aleatorio de Bastani et al. (2025) documenta un efecto de -17% en el rendimiento de estudiantes de secundaria cuando el acceso a la herramienta es irrestricto, lo que apunta a que el diseño pedagógico condiciona, bajo determinadas condiciones, el signo del efecto.

En el plano metacognitivo, Fan et al. (2024) acuñan el término pereza metacognitiva para describir la reducción sistemática de los procesos de planificación y autoevaluación ante el acceso irrestricto al sistema. Xu et al. (2025) demuestran que el soporte metacognitivo estructurado revierte este efecto. Sobre los efectos afectivos, Wang & Wei (2025), Polydoros et al. (2025) y Etcuban (2025) documentan reducciones consistentes de la ansiedad matemática en China, Grecia y Filipinas respectivamente, aunque el meta-análisis de Xia et al. (2025) no detecta cambios significativos en la motivación global ($g = 0.299$, NS).

Problema de investigación y objetivos

El campo carece de una síntesis sistemática que articule los efectos cognitivos, metacognitivos y afectivos con el análisis de sus moderadores desde una perspectiva iberoamericana. Esta revisión tiene como objetivo sintetizar la evidencia disponible sobre los efectos de la inteligencia artificial generativa en el aprendizaje matemático, con énfasis en las variables mediadoras e implicaciones para el contexto latinoamericano.

Las cuatro preguntas de síntesis que guían el análisis son: (PS1) ¿Qué efectos cognitivos produce la inteligencia artificial generativa en el aprendizaje matemático?; (PS2) ¿Cómo afecta a los procesos metacognitivos y de autorregulación?; (PS3) ¿Qué efectos tiene sobre variables afectivas como la ansiedad matemática y la motivación?; y (PS4) ¿Qué factores modulan la magnitud y el signo de estos efectos?

La inteligencia artificial generativa designa sistemas computacionales basados en modelos de lenguaje grande (LLM, por sus siglas en inglés) que generan respuestas coherentes y contextualizadas en lenguaje natural. Una limitación estructural de estos sistemas con implicaciones directas para la educación matemática es su tendencia a producir respuestas matemáticamente incorrectas formuladas con aparente corrección, fenómeno denominado alucinación (Kalai & Vempala, 2024). Desde la Teoría de la Carga Cognitiva (Sweller et al., 2011), puede liberar carga extrínseca para el procesamiento germano —o, en ausencia de diseño pedagógico, hacia la pasividad—. Desde el marco vigotskiano, opera en la Zona de Desarrollo Próximo pero, a diferencia del tutor humano, no percibe la señal de retirar el andamio oportunamente. La Teoría de la Autorregulación del Aprendizaje (Zimmerman, 2000; Pintrich, 2004) proporciona la explicación más robusta de la paradoja central del corpus: los estudiantes con menor autorregulación previa son quienes más riesgo tienen de desarrollar dependencia disfuncional. El framework Intelligent-TPACK propuesto por Chiu (2025) integra estos marcos desde la perspectiva del conocimiento pedagógico docente para la inteligencia artificial.

RESULTADOS

Características del corpus (N = 165)

El corpus presenta un sesgo temporal marcado hacia los años 2024-2025 (71.5% de los artículos), reflejo de la aceleración de producción científica posterior al lanzamiento de ChatGPT. Por región, Asia Oriental concentra el 40.0% (n = 66), seguida de Europa (20.6%, n = 34), América Latina (17.0%, n = 28) y los Estados Unidos y Canadá (14.5%, n = 24). Por diseño metodológico, predominan los estudios cuasi-experimentales y los estudios de caso (57.0% en conjunto), seguidos de revisiones sistemáticas y meta-análisis (21.2%). El sistema más estudiado es ChatGPT en sus versiones GPT-3.5 y GPT-4 (72.1% del corpus). La calidad metodológica, evaluada mediante CASP, fue alta (17-20 puntos) en el 28.9% del corpus, moderada (12-16) en el 53.3% y baja-aceptable (10-11) en el 17.8%.

PS1: Efectos cognitivos

La Tabla 3 sintetiza los tamaños de efecto reportados en los meta-análisis y ensayos controlados del corpus.

Tabla 3

Tamaños de efecto sobre resultados cognitivos en educación matemática

Estudio	k	N	Tamaño de efecto	Métrica	Dominio
Liu et al. (2026) – EM específico	22	5.232	$g = 0.534^*$	Hedges' g	Matemáticas
Han et al. (2025)	68	N/R	$SMD = 0.45^*$	SMD	Múltiple
Ma & Zhong (2025)			$d = 0.68^*$	d Cohen	Múltiple
Wang & Fan (2025)	51		$g = 0.867^* / g = 0.457^*$	Hedges' g	STEM incl. EM
Xia et al. (2025) – afectivo			$g = 0.299$ (NS)	Hedges' g	Múltiple
Bastani et al. (2025) – RCT	1	~1.000	-17%*	Otro	Álgebra sec.
Wu et al. (2024) – RCT PA-GPT	1	61	Positivo H0T*	Otro	STEM/Matemáticas
Pardos & Bhandari (2024) – RCT	1	274	≈0 vs tutor humano	d Cohen	Álgebra/Estadística
Lu et al. (2025) – AI hints	1		Positivo rendimiento*	Otro	Matemáticas/Física
Sánchez-Ruiz et al. (2023)	1		Negativo*	d Cohen	Matemáticas ingeniería

Nota: g = Hedges' g ; d = d de Cohen; SMD = diferencia de medias estandarizada; * = estadísticamente significativo ($p < 0.05$); NS = no significativo; H0T = pensamiento de orden superior; RCT = ensayo controlado aleatorio; N/R = no reportado.

Fuente: elaboración propia.

El meta-análisis de Liu et al. (2026), con mayor especificidad temática (22 estudios, $N = 5.232$), documenta un efecto positivo moderado ($g = 0.534$, IC 95%: [0.37, 0.70]) cuando la intervención con inteligencia artificial generativa está pedagógicamente diseñada. El ensayo controlado aleatorio de Bastani et al. (2025) –el único RCT de gran escala específico en educación matemática– aporta uno de los resultados más relevantes del corpus: los estudiantes con acceso irrestricto al sistema obtuvieron un rendimiento 17% inferior al grupo control en evaluaciones sin asistencia de la herramienta. Este efecto se revierte cuando el acceso está mediado por pistas pedagógicas diseñadas por docentes (GPT Tutor), que no mostró diferencias respecto al control. Pardos y Bhandari (2024) matizan estos resultados: en un experimento factorial ($N = 274$), ChatGPT produjo ganancias de aprendizaje equivalentes a las de un tutor humano en álgebra y estadística universitaria cuando se implementó un protocolo de verificación que redujo la tasa de errores del modelo. Sánchez-Ruiz et al. (2023) replican el patrón negativo en matemáticas para ingeniería en España.

En el plano cualitativo, Kock et al. (2025) identifican tres esquemas de uso de ChatGPT en estudiantes de matemáticas de ingeniería: (a) procedimental, centrado en obtener resultados sin comprensión; (b) conceptual, orientado a construir entendimiento mediante el diálogo; y (c) estratégico, en el que el estudiante planifica su propio proceso de resolución con el sistema como interlocutor crítico. Este marco ofrece la taxonomía más precisa disponible para caracterizar la calidad de la interacción y

predice diferencialmente la profundidad del aprendizaje. Malik et al. (2025) documentan, adicionalmente, la viabilidad de usar modelos de lenguaje grande para generar tareas curriculares de calentamiento en secundaria mediante un proceso estructurado de tres etapas: observación, formulación de estrategia e implementación.

PS2: Efectos metacognitivos y sobre la autorregulación

El hallazgo estructural de PS2 es la paradoja de la autorregulación: los estudiantes con menor capacidad de autorregulación del aprendizaje son quienes más se beneficiarían de un apoyo externo, pero son también quienes presentan mayor riesgo de desarrollar dependencia disfuncional si el acceso no está diseñado con soporte metacognitivo explícito. Fan et al. (2024) denominan pereza metacognitiva al patrón de reducción de planificación, monitoreo y autoevaluación que se observa cuando los estudiantes tienen acceso total a ChatGPT sin estructura pedagógica. Xu et al. (2025) demuestran experimentalmente (N = 68) que el soporte metacognitivo estructurado revierte este efecto: el grupo con andamios explícitos mejora las tres dimensiones del ciclo de autorregulación, mientras el grupo control las deteriora.

Contel y Cusi (2024) profundizan en este hallazgo mostrando que GPT-4 puede estructurar activamente la revisión metacognitiva durante la resolución de problemas cuando el estudiante mantiene una postura epistémica activa —cuestionando, verificando y reformulando— en contraste con el uso pasivo como oráculo de respuestas. Huda et al. (2025) confirman este patrón en el contexto específico del cálculo diferencial (límites), documentando mejoras significativas en las tres dimensiones metacognitivas evaluadas cuando ChatGPT se integra en un entorno de indagación guiada.

PS3: Efectos afectivos

Tres estudios cuasi-experimentales independientes en China (Wang & Wei, 2025), Grecia (Polydoros et al., 2025) y Filipinas (Etcuban, 2025) documentan reducciones estadísticamente significativas de la ansiedad matemática, con mayor beneficio para estudiantes con alta ansiedad previa. Rizos et al. (2024) extienden este hallazgo a estudiantes con necesidades educativas especiales en matemáticas de secundaria en Grecia, identificando beneficios en participación y motivación —un subgrupo que el corpus global había prácticamente ignorado. Chen et al. (2025) matizan el panorama positivo: mediante un modelo de ecuaciones estructurales demuestran que la ansiedad matemática predice significativamente la dependencia de la herramienta, la cual a su vez predice negativamente la autonomía cognitiva, configurando un circuito que puede deteriorar el aprendizaje a largo plazo.

Sobre la motivación, el meta-análisis de Xia et al. (2025) reporta un efecto global no significativo ($g = 0.299$, NS), coherente con la interpretación de que la motivación en matemáticas está mediada por factores profundos —historia de fracasos, creencias sobre capacidad innata, ansiedad arraigada— que no se modifican con la mera disponibilidad de una herramienta. Estudios que integran la herramienta en metodologías activas (Susilawati et al., 2025, en Indonesia) sí reportan mejoras motivacionales, lo que sugiere que el efecto depende del diseño de la actividad, no del sistema en sí.

PS4: Moderadores del impacto

La Tabla 4 sintetiza los nueve moderadores identificados en el corpus, con su evidencia principal y nivel de certeza empírica.

Tabla 4

Moderadores del impacto de la inteligencia artificial generativa en educación matemática

Moderador	Evidencia principal	Dirección	Certeza	Implicación práctica
Diseño pedagógico (guardrails)	Bastani et al. (2025); Wu et al. (2024)	Determina el signo del efecto	Alta	Sin diseño: efecto negativo posible
Soporte metacognitivo	Xu et al. (2025); Huda et al. (2025)	Amplificador positivo	Alta	Andamios explícitos imprescindibles
SRL previo del aprendiz	Heung & Chiu (2025); Fan et al. (2024)	Amplificador positivo	Alta	Paradoja: vulnerable = mayor riesgo
Tipo de esquema de uso	Kock et al. (2025)	Amplificador positivo	Moderada	Uso estratégico > conceptual > procedimental
Dominio matemático	Liu et al. (2026)	Moderador de magnitud	Alta	Geometría: mayor efecto; cálculo: mixto
Ansiedad matemática previa	Wang & Wei (2025); Chen et al. (2025)	Doble efecto	Moderada	↓ ansiedad pero posible ↑ dependencia
Versión LLM	Korkmaz Guler et al. (2024)	Moderador de calidad	Moderada	GPT-4 > GPT-3.5 en tareas complejas
Personalidad agente	Lyu et al. (2025)	Moderador afectivo	Baja	Alta apertura → mayor engagement
Necesidades educativas especiales	Rizos et al. (2024)	Potencial amplificador	Baja	Subgrupo con alto potencial. Prioridad agenda

Nota: Certeza Alta = respaldada por RCT o meta-análisis con N > 500; Moderada = cuasi-experimentos con N razonable o múltiples estudios convergentes; Baja = estudio de caso único o evidencia inicial. SRL = autorregulación del aprendizaje.

Fuente: elaboración propia.

El moderador con mayor respaldo empírico en el corpus es el diseño pedagógico: la evidencia disponible sugiere que condiciona el signo del efecto, no sólo su magnitud. El soporte metacognitivo explícito y el nivel previo de autorregulación del aprendizaje son los dos moderadores de características del aprendiz con mayor certeza. Los esquemas de uso de ChatGPT (Kock et al., 2025) ofrecen un marco analítico nuevo que operacionaliza la calidad de la interacción con una precisión no disponible previamente en el campo. La brecha digital y el contexto latinoamericano tienen alta urgencia, pero baja certeza empírica: sólo el 17.0% del corpus corresponde a estudios de la región, y la mayoría son descriptivos o exploratorios. Panqueban y Huincahue (2024), en la única revisión sistemática disponible con foco latinoamericano, confirman esta subrepresentación y señalan que la brecha de acceso digital amplifica las desigualdades preexistentes en educación matemática.

DISCUSIÓN

Interpretación de los resultados

La síntesis del corpus permite formular el modelo de impacto mediado: inteligencia artificial generativa → [diseño pedagógico × autorregulación del aprendizaje × alfabetización en inteligencia artificial] → aprendizaje matemático.

Figura 1

Modelo de impacto mediado de la inteligencia artificial generativa en el aprendizaje matemático



Fuente: elaboración propia a partir de la síntesis del corpus (N = 165 artículos).

En este modelo, el efecto de la herramienta no es directo ni unidireccional: ninguno de los tres mediadores opera de forma independiente. El diseño pedagógico puede compensar parcialmente déficits en autorregulación mediante andamios, pero no puede compensar la ausencia de alfabetización crítica hacia el sistema: un estudiante que acepta acríticamente respuestas incorrectas deteriora sus esquemas conceptuales independientemente de cómo esté estructurada la tarea.

La taxonomía de esquemas de uso (Kock et al., 2025) operacionaliza la dimensión de alfabetización en inteligencia artificial de forma precisa: el uso procedimental representa una alfabetización mínima; el uso conceptual, una intermedia; y el uso estratégico, una plena. Esta gradación tiene implicaciones inmediatas para el diseño curricular: las intervenciones deberían secuenciar actividades que desplacen progresivamente el uso procedimental hacia el estratégico, en lugar de asumir que el estudiante desarrollará ese desplazamiento de forma espontánea.

Implicaciones teóricas y prácticas

La paradoja de la autorregulación tiene una dimensión de equidad que el corpus permite articular con precisión. Los estudiantes más vulnerables —con mayor ansiedad, menor autorregulación y menor acceso material a tecnología de calidad— son quienes más pueden beneficiarse y más riesgo tienen en implementaciones no diseñadas. Esto configura un dilema de diseño central: la restricción del acceso reproducirá inequidades; su apertura sin diseño las amplificaba. La solución no es ninguno de

los extremos sino la diferenciación: intervenciones con guardias pedagógicas más robustas para estudiantes con menor autorregulación previa, y progresiva reducción del andamio a medida que se consolida la autonomía.

Los docentes en servicio, por su parte, muestran actitudes que oscilan entre la apertura y la cautela (Busuttill & Calleja, 2025), lo que subraya la necesidad de acompañar la integración con formación especializada. Para los formadores de docentes, el framework I-TPACK (Chiu, 2025) ofrece el marco más operativo disponible: el conocimiento pedagógico de inteligencia artificial (AI-PK) —saber cómo diseñar interacciones con modelos de lenguaje grande que activen los procesos cognitivos del aprendizaje matemático— es la competencia que, según el corpus revisado, más consistentemente se asocia con si la herramienta produce beneficio o daño en el aula. Este conocimiento no es transferible de la formación en tecnología general: requiere formación específica en educación matemática con inteligencia artificial generativa.

Limitaciones

Cuatro limitaciones merecen señalamiento explícito. Primera, la subrepresentación latinoamericana: el 17.0% del corpus proviene de la región, con predominio de estudios descriptivos y ningún ensayo controlado con grupo de comparación realizado en América Latina. Segunda, el predominio de estudios breves: el 61% de los estudios primarios tiene una intervención de ocho semanas o menos, lo que impide pronunciarse sobre efectos a largo plazo. Tercera, la posible inflación de los tamaños de efecto promedio por sesgo de publicación (predominio de resultados positivos). Cuarta, la evolución tecnológica acelerada: los estudios con modelos GPT-3.5, dominantes hasta 2023, pueden no ser representativos del comportamiento de modelos actuales con mayores capacidades de razonamiento matemático.

CONCLUSIÓN

Esta revisión sistemática PRISMA 2020, basada en 165 artículos de siete bases de datos internacionales, permite identificar cuatro tendencias que, en el corpus revisado, muestran convergencia a través de estudios independientes. Primera: la evidencia del corpus sugiere un efecto positivo moderado sobre el rendimiento en educación matemática cuando el diseño pedagógico incorpora salvaguardas y soporte metacognitivo ($g = 0.534$, bajo condiciones controladas). Segunda: sin esas condiciones, puede deteriorar el aprendizaje de forma significativa (-17% en el RCT con mayor muestra del corpus), aunque la generalización de este hallazgo requiere réplica en otros contextos. Tercera: en el corpus revisado se asocia con reducción de la ansiedad matemática en tres contextos culturales distintos, aunque los efectos sobre la motivación global no alcanzaron significación estadística. Cuarta: la paradoja de la autorregulación —que los estudiantes más vulnerables tienen el mayor potencial de beneficio y el mayor riesgo— exige diseños diferenciados, no restricción de acceso.

El modelo de impacto mediado propuesto sintetiza estos hallazgos en un marco accionable: inteligencia artificial generativa → [diseño pedagógico × autorregulación × alfabetización en inteligencia artificial] → aprendizaje matemático. La educación matemática latinoamericana dispone de evidencia orientadora —aunque aún insuficiente para conclusiones causales regionales— para diseñar intervenciones informadas; lo que aún necesita con urgencia es investigación propia, con poblaciones

propias, en contextos propios. Generar esa evidencia es la contribución más valiosa que los investigadores de la región pueden hacer al campo en los próximos cinco años.

Recomendaciones para futuras investigaciones

La agenda de investigación más urgente para el contexto latinoamericano comprende: (1) estudios longitudinales de al menos un año académico completo sobre efectos en la competencia matemática, inexistentes en el corpus actual; (2) ensayos controlados en contextos de América Latina que permitan conclusiones causales regionales; (3) investigación sobre estudiantes con necesidades educativas especiales en matemáticas, con evidencia casi ausente en el corpus global; y (4) desarrollo y validación de instrumentos de medición en español de las dimensiones de autorregulación, ansiedad matemática y alfabetización en inteligencia artificial, herramientas instrumentales previas a cualquier síntesis empírica regional.

REFERENCIAS

Bastani, H., Bastani, O., Sungu, A., Ge, H., Kabakçı, Ö., & Mariman, R. (2025). Generative AI without guardrails can harm learning: Evidence from high school mathematics. *Proceedings of the National Academy of Sciences*, 122(26), e2422633122. <https://doi.org/10.1073/pnas.2422633122>

Busuttil, L., & Calleja, J. (2025). Teachers' beliefs and practices about the potential of ChatGPT in teaching mathematics in secondary schools. *Digital Experiences in Mathematics Education*, 11, 140-166. <https://doi.org/10.1007/s40751-024-00168-3>

Chen, F., Chen, J., & Xu, Y. (2025). The more anxious, the more dependent? The impact of math anxiety on AI-assisted problem-solving. *Psychology in the Schools*, 62, 2685-2701. <https://doi.org/10.1002/pits.23500>

Chiu, T. K. F. (2025). Developing intelligent-TPACK (I-TPACK) framework from AI literacy and competency. *Interactive Learning Environments*, 33(7), 4189-4192. <https://doi.org/10.1080/10494820.2025.2545053>

Contel, F., & Cusi, A. (2024). Investigating the role of ChatGPT in supporting metacognitive processes during problem-solving activities. *Digital Experiences in Mathematics Education*, 11, 167-191. <https://doi.org/10.1007/s40751-024-00164-7>

Etcuban, J. O. (2025). The use of ChatGPT in addressing algebra anxiety and promoting confidence. *International Electronic Journal of Mathematics Education*, 20(1), em0804. <https://www.iejme.com>

Fan, Y., Tang, L., Le, H., Shen, K., Tan, S., Zhao, Y., Shen, Y., Li, X., & Gašević, D. (2024). Beware of metacognitive laziness: Effects of generative AI on learning motivation, processes, and performance. *British Journal of Educational Technology*, 56(2), 489-530. <https://doi.org/10.1111/bjet.13544>

Han, X., Peng, H., & Liu, M. (2025). Impact of generative artificial intelligence on learning outcomes: A systematic review and meta-analysis. *Educational Research Review*, 48, 100714. <https://doi.org/10.1016/j.edurev.2025.100714>

Heung, K., & Chiu, T. K. F. (2025). Does generative AI improve students' higher-order thinking? Meta-analysis of 29 experiments. *PMC*. <https://doi.org/10.3389/fpsyg.2025.1543822>

Huda, N., Anwar, K., Novferma, N., & Kurniawan, W. (2025). ChatGPT scaffolding in supporting metacognition for limit concepts in guided inquiry mathematics learning. *Journal of Information Technology Education*, 24. <https://doi.org/10.28945/5567>

Kalai, A., & Vempala, S. (2024). Calibrated language models must hallucinate. *Proceedings of the 56th Annual ACM Symposium on Theory of Computing (STOC 2024)*. <https://doi.org/10.1145/3618260.3649777>

Kock, Z.-j., Salinas-Hernández, U., & Pepin, B. (2025). Engineering students' initial use schemes of ChatGPT as an instrument for learning. *Digital Experiences in Mathematics Education*, 11, 192-218. <https://doi.org/10.1007/s40751-025-00169-w>

Korkmaz Guler, N., Dertli, Z. G., Boran, E., & Yildiz, B. (2024). ChatGPT academic achievement in a mathematics exam. *Pedagogical Research*, 9(2), em0188. <https://doi.org/10.29333/pr/14145>

- Liu, B., Zhang, W., & Wang, F. (2026). Can generative AI effectively enhance students' mathematics learning outcomes? A meta-analysis of empirical studies from 2023 to 2025. *Education Sciences*, 16(1), Art. 140. <https://doi.org/10.3390/educsci16010140>
- Liu, J., Sun, D., Sun, J., Wang, J., & Yu, P. L. H. (2025). Designing a generative AI enabled learning environment for mathematics word problem solving in primary schools. *Computers and Education: Artificial Intelligence*, 9, 100438. <https://doi.org/10.1016/j.caeai.2025.100438>
- Lu, Y., Supriya, K., Shaked, S., Simmons, E. H., & Kusenko, A. (2025). Incentivizing supplemental math assignments and using AI-generated hints is associated with improved exam performance. *Physical Review Physics Education Research*, 21, 010160. <https://doi.org/10.1103/PhysRevPhysEducRes.21.010160>
- Lyu, X., Song, Y., Wang, X., Chen, M., & Liu, Z. (2025). The role of teachable agents' personality traits on student-AI interactions and mathematical learning. *Computers and Education: Artificial Intelligence*, 8, 100338. <https://doi.org/10.1016/j.caeai.2025.100338>
- Ma, N., & Zhong, Z. (2025). A meta-analysis of the impact of generative AI on learning outcomes. *Journal of Computer Assisted Learning*, 41(5), e70117. <https://doi.org/10.1111/jcal.70117>
- Malik, R., Abdi, D., Wang, R., & Demszky, D. (2025). Scaffolding middle school mathematics curricula with large language models. *British Journal of Educational Technology*, 56(2). <https://doi.org/10.1111/bjet.13571>
- 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>
- Panqueban, D., & Huincahue, J. (2024). Inteligencia artificial en educación matemática: Una revisión sistemática. *Uniciencia*, 38(1). <https://www.scielo.sa.cr/scielo.php?pid=S2215-34702024000100357>
- Pardos, Z. A., & Bhandari, S. (2024). ChatGPT-generated help produces learning gains equivalent to human tutor-authored help on mathematics skills. *PLOS ONE*, 19(5), e0304013. <https://doi.org/10.1371/journal.pone.0304013>
- Pintrich, P. R. (2004). A conceptual framework for assessing motivation and self-regulated learning in college students. *Educational Psychology Review*, 16(4), 385-407.
- Polydoros, G., Tsanaktsidis, C., Tsinakos, A., & Gkioxari, A. (2025). Innovative AI-driven approaches to mitigate math anxiety and enhance resilience in low-performing students. *Psychology International*, 7(2), 46. <https://doi.org/10.3390/psycholint7020046>
- Rizos, I., Foykas, E., & Georgakopoulos, S. V. (2024). Enhancing mathematics education for students with special educational needs through generative AI: A case study in Greece. *Contemporary Educational Technology*, 16(4), ep535. <https://doi.org/10.30935/cedtech/15073>
- Sánchez-Ruiz, L. M., Moll-López, S., Nuñez-Pérez, A., Moraño-Fernández, J. A., & Vega-Fleitas, E. (2023). ChatGPT challenges blended learning in engineering education: A case study in mathematics. *Applied Sciences*, 13(10), 6039. <https://doi.org/10.3390/app13106039>

Susilawati, W., Sharov, S., Pasqa, M., & Malik, H. (2025). Integrating realistic mathematics education, AI, and gamification to enhance students' learning motivation and problem-solving skills. *Journal on Mathematics Education*, 16(4), 1257-1282. <https://doi.org/10.22342/jme.v16i4.pp1257-1282>

Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. Springer.

Wang, J., & Fan, W. (2025). The effect of ChatGPT on students' learning performance, learning perception, and higher-order thinking: Insights from a meta-analysis. *Humanities and Social Sciences Communications*, 12, Art. 621. <https://doi.org/10.1057/s41599-025-04787-y>

Wang, X., & Wei, C.-L. (2025). Influence of generative AI assisted learning on primary school students' math anxiety: An intervention study. *Applied Cognitive Psychology*, 39(4), e70088. <https://doi.org/10.1002/acp.70088>

Wu, T.-T., Huang, Y.-M., Lee, H.-Y., Chen, P.-H., & Lin, C.-J. (2024). Integrating peer assessment cycle into ChatGPT for STEM education: A randomised controlled trial on knowledge, skills, and attitudes enhancement. *Journal of Computer Assisted Learning*, 41, e13085. <https://doi.org/10.1111/jcal.13085>

Xia, Q., Li, W., Yang, Y., Weng, X., & Chiu, T. K. F. (2025). Systematic review and meta-analysis of generative AI effectiveness on students' motivation and engagement. *Computers and Education: AI*, 9, Art. 100455. <https://doi.org/10.1016/j.caeai.2025.100455>

Xu, X., Qiao, L., Cheng, N., Liu, H., & Zhao, W. (2025). Enhancing self-regulated learning in generative AI environments: Critical role of metacognitive support. *British Journal of Educational Technology*, 56, 1842-1863. <https://doi.org/10.1111/bjet.13599>

Zimmerman, B. J. (2000). Attaining self-regulation: A social cognitive perspective. En M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 13-39). Academic Press.

Todo el contenido de LATAM Revista Latinoamericana de Ciencias Sociales y Humanidades, publicados en este sitio está disponibles bajo Licencia [Creative Commons](https://creativecommons.org/licenses/by/4.0/) 