ASCE MAGAZINE ISSN: 3073–1178

Revista ASCE Magazine, Periodicidad: Trimestral, Volumen: 5, Número: 4, Año: 2026 pág. 796-827

Doi: https://doi.org/10.70577/asce.v5i4.1158

Recibido: 2026-10-25

Aceptado: 2026-11-11

Publicado: 2026-11-25

Impacto de la inteligencia artificial generativa como tutor personalizado en el rendimiento académico en competencias algebraicas de estudiantes de bachillerato en la Unidad Educativa San Rafael del cantón Portoviejo

Impact of generative artificial intelligence as a personalized tutor on academic performance in algebraic competencies among high school students at the San Rafael Educational Unit, Portoviejo canton

Autor(s)

Rayniero Cely Castro [1]
racelyc@ube.edu.ec
https://orcid.org/0009-0000-0296-9403
Universidad Bolivariana del Ecuador
Guayas – Ecuador

Dayron Rumbaut Rangel [2]
drumbautr@ube.edu.ec
https://orcid.org/0009-0001-9087-0979
Universidad Bolivariana del Ecuador
Guayas – Ecuador

Lorenzo Cevallos Torres [3]
ljcevallost@ube.edu.ec
https://orcid.org/0000-0002-7211-2891
Universidad Bolivariana del Ecuador
Guayas – Ecuador

Como Citar:

Cely Castro , R., Rumbaut Rangel , D., & Cevallos Torres , L. (2026). Impacto de la inteligencia artificial generativa como tutor personalizado en el rendimiento académico en competencias algebraicas de estudiantes de bachillerato en la Unidad Educativa San Rafael del cantón Portoviejo. ASCE MAGAZINE, 5(4), 796-827. https://doi.org/10.70577/asce.v5i4.1158

https://creativecommons.org/licenses/by-nc-sa/4.0/

Resumen

El propósito de este trabajo fue valorar qué efecto tiene la inteligencia artificial generativa, empleada como tutor personalizado, sobre el rendimiento académico en competencias algebraicas de un grupo de jóvenes que cursaban el segundo año de bachillerato en la Unidad Educativa San Rafael, de Portoviejo, durante el período lectivo 2025–2026. Se trabajó con un enfoque cuantitativo y un diseño cuasiexperimental con medición inicial y final (*pretest* y *postest*) y un grupo control no equivalente, con 60 participantes divididos en dos grupos de 30. El rendimiento se entendió aquí como la competencia que el propio estudiante percibe. Se empleó un cuestionario Likert de 12 ítems agrupados en cuatro bloques, cuya validez de contenido avalaron seis expertos (3,87/4,00; CVR = 1,00). Como su consistencia interna resultó baja (α entre −,32 y ,13 según grupo y momento), los datos se examinaron por bloque y con carácter exploratorio, sin recurrir a la puntuación total. Luego de cuatro semanas de trabajo con ChatGPT (GPT-5) como tutor, la media del bloque de rendimiento algebraico percibido subió en el grupo experimental de M = 2,59 (DE = 0,73) a M = 4,28 (DE = 0,61), mientras que en el control fue de M = 2,38 (DE = 0,77) a M = 2,99 (DE = 0,67). En la medición final, la distancia entre grupos llegó a 1,29 puntos (IC 95% [0,96; 1,62]; $t(58) = 7{,}81$; $p < ,001$; d = 2,02) y persistió tras controlar la medición inicial (ANCOVA). Los tres bloques restantes mostraron diferencias de 1,18 a 1,36 puntos. El grupo que usó la herramienta la calificó de forma positiva (Bloque C, medición final: M = 4,17). Todo indica que la IA generativa puede apoyar la enseñanza y elevar la competencia algebraica percibida de quienes tienen acceso a internet, si bien el diseño, el tamaño de la muestra y la corta duración impiden generalizar.

Palabras clave: inteligencia artificial generativa; tutoría personalizada; competencias algebraicas; rendimiento académico; aprendizaje adaptativo; diseño cuasiexperimental.

https://creativecommons.org/licenses/by-nc-sa/4.0/

Abstract

This research aims to evaluate the impact of generative artificial intelligence as a personalized tutor on academic performance in algebraic competencies among second-year high school students at the San Rafael Educational Unit in Portoviejo, during the 2025–2026 school year. A quantitative approach was applied using a quasi-experimental pretest-posttest design with a non-equivalent control group; 60 students participated (30 per group). In this study, performance is operationalized as self-perception. The 12-item Likert self-perception instrument, organized in four blocks, was validated through expert judgment (six experts; 3.87/4.00; CVR = 1.00); its internal consistency was low (α between −.32 and .13 by group and time point), so the analysis was exploratory and by block, not with the total score. After four weeks of intervention with ChatGPT (GPT-5) as personalized tutor, in the self-perceived algebraic performance block the experimental group rose from M = 2.59 (SD = 0.73) to M = 4.28 (SD = 0.61) and the control group from M = 2.38 (SD = 0.77) to M = 2.99 (SD = 0.67); the posttest difference was 1.29 points (95% CI [0.96, 1.62]; $t(58) = 7.81$; $p < .001$; $d = 2.02$) and remained after adjusting for the pretest (ANCOVA). In the other three blocks, differences ranged from 1.18 to 1.36 points. The experimental group rated the tool favorably (Block C, posttest: M = 4.17). The results suggest that generative AI can be a useful pedagogical support for improving self-perceived algebraic competencies among students with internet access, although the design, sample size and duration limit generalization.

Keywords: generative artificial intelligence; personalized tutoring; algebraic competencies; academic performance; adaptive learning; quasi-experimental design.

https://creativecommons.org/licenses/by-nc-sa/4.0/

Introducción

Desde hace aproximadamente diez años, la inteligencia artificial generativa (IA generativa) se ha incorporado de manera progresiva a las prácticas de enseñanza. El álgebra resulta especialmente apta para un trabajo personalizado: al ser una materia lógica y ordenada por pasos, los errores conceptuales de los alumnos tienden a repetirse, pueden reconocerse con facilidad y se corrigen mejor con una respuesta inmediata. Tanto la UNESCO (2025) como la Organización de las Naciones Unidas (2024) señalan que estas herramientas podrían favorecer la inclusión y la calidad educativa en contextos vulnerables, y acercar la tutoría individual a quienes enfrentan obstáculos económicos o geográficos. Con ese marco, el presente estudio analiza qué efecto tiene la IA generativa, usada como tutor personalizado, en el rendimiento algebraico de quienes cursaban el segundo año de bachillerato en la Unidad Educativa San Rafael (Portoviejo) en el año lectivo 2025–2026.

Los datos de PISA-D 2018, examinados por Tene-Tenempaguay et al. (2025), ubicaron a los estudiantes ecuatorianos de 15 años en un nivel apenas básico de competencia matemática. A ese rezago se añade otro problema: no todos los colegios disponen de recursos tecnológicos que refuercen la enseñanza, y donde existen no siempre se aprovechan. En su trabajo con alumnos de bachillerato, Mendoza-Zambrano et al. (2023) describieron obstáculos derivados de la brecha digital y del escaso acceso a tecnologías educativas, y concluyeron que conviene ofrecer opciones tecnológicas que ayuden a aprender.

Portoviejo no es una excepción. Vélez-Basurto y Vélez-Loor (2024) documentaron que, en el bachillerato de esta zona, los alumnos con problemas de aprendizaje en Matemáticas no reciben estrategias ajustadas a sus necesidades, y que esa carencia se refleja en sus calificaciones. Las consecuencias llegan hasta la universidad: al evaluar a 4.306 aspirantes admitidos en la Universidad Técnica de Manabí, Montes Rodríguez et al. (2022) detectaron vacíos en conocimientos matemáticos elementales. De ahí la necesidad de reforzar el acompañamiento académico y de ensayar formas de aprendizaje personalizado que fortalezcan las competencias algebraicas.

La dificultad de fondo es práctica: con más de 35 alumnos por aula, el profesor no alcanza a corregir a cada uno en el momento. Una herramienta de IA generativa puede cubrir parte de ese vacío, porque explica de acuerdo con el nivel de quien pregunta, propone ejemplos a la medida y señala

https://creativecommons.org/licenses/by-nc-sa/4.0/

los errores al instante. En un entorno digital, esto se parece a contar con un tutor particular, pero sin depender de horarios ni de la presencia física del docente (Mujica-Sequera, 2024; Serrano y Moreno-García, 2024).

A esto se suma que las condiciones de los colegios del país no son iguales: la conexión a internet y el equipamiento varían mucho de una institución a otra. Entre los estudiantes de bachillerato siguen existiendo diferencias de acceso a la red (Mendoza-Zambrano et al., 2017), y en una muestra de egresados de ese nivel apenas uno de cada cinco, el 20%, contaba con señal de internet dentro del aula (Bravo Guerrero y Quezada Matute, 2021). Por esta razón, algunos autores ven en la IA generativa un recurso a favor de la equidad: quien tiene conexión recibe explicaciones adaptadas a su caso sin necesitar que alguien en su hogar le ayude con las tareas (Holmes et al., 2019).

Desde el punto de vista teórico, este uso de la IA generativa encuentra respaldo en el conectivismo y en la noción de Zona de Desarrollo Próximo de Vygotsky. De acuerdo con ese planteamiento, se aprende mejor cuando la ayuda que se recibe corresponde al nivel actual de comprensión y empuja al estudiante hacia el paso conceptual inmediato (Ayuso-del Puerto y Gutiérrez-Esteban, 2022; Hwang y Tu, 2021). Los tutores inteligentes basados en IA generativa operan de forma parecida: estiman lo que el alumno ya domina, localizan sus fallos concretos y le ofrecen explicaciones que actúan como un apoyo graduado y personal (Panqueban y Huincahue, 2024).

Los modelos de lenguaje entienden y producen lenguaje natural, de modo que la conversación con ellos resulta más próxima que la que permiten los programas educativos de otras épocas. Gracias a ello, la IA puede adaptar el aprendizaje a cada persona y responder de inmediato (Loayza Solórzano y Moya Martínez, 2024). Mientras una plataforma fija solo ofrece contenidos cerrados, la IA generativa atiende preguntas abiertas y presenta un mismo concepto desde distintos ángulos. Investigaciones recientes sobre tutores inteligentes (Villegas-Ch et al., 2025) muestran que estos sistemas pueden plantear ejercicios de dificultad creciente y comentar cada respuesta en particular, aun cuando el alumno se exprese de forma imprecisa. Ese intercambio que se ajusta al estudiante marca la diferencia con la tutoría automática de generaciones previas y explica por qué vale la pena someterlo a prueba experimental en la escuela.

Llevar estas herramientas a la educación pública también plantea exigencias éticas y de privacidad. Corresponde a las autoridades educativas locales asegurar que la información de los menores esté

https://creativecommons.org/licenses/by-nc-sa/4.0/

protegida (Vilchez Ruiz, 2024). Es preciso, asimismo, vigilar que los algoritmos no repitan prejuicios de género o de nivel socioeconómico que perjudiquen a ciertos alumnos. Cuanto más claro sea el funcionamiento de las herramientas, mayor será la confianza de la comunidad de la Unidad Educativa San Rafael y menor el riesgo de que los estudiantes caigan en la pereza intelectual o pierdan autonomía para decidir por sí mismos (Ahmad et al., 2023; Yudrowsky, 2023).

El estudio buscó responder la siguiente pregunta: ¿cómo influye el empleo de la inteligencia artificial generativa como tutor personalizado en el rendimiento académico en competencias algebraicas de quienes cursaron el segundo año de bachillerato en la Unidad Educativa San Rafael, cantón Portoviejo, en el período lectivo 2025–2026?

Se planteó que, luego de cuatro semanas de tutoría con IA generativa, los estudiantes del grupo experimental puntuarían en la medición final por encima de los del grupo control en cada uno de los bloques del cuestionario de competencia percibida. Expresado en términos estadísticos, H_0: en la medición final, las medias de cada bloque no diferirán entre los grupos; H_1: en la medición final, el grupo experimental tendrá medias más altas en cada bloque.

Como objetivo general se propuso determinar el efecto que tiene la inteligencia artificial generativa, empleada como tutor personalizado, en el rendimiento académico en competencias algebraicas de quienes cursaron el segundo año de bachillerato en la Unidad Educativa San Rafael, del cantón Portoviejo, en el período lectivo 2025–2026. De él se derivan cuatro objetivos específicos:

a) Establecer, antes de introducir la IA generativa como tutor personalizado, cómo perciben los alumnos de segundo de bachillerato su rendimiento en álgebra, su aprendizaje autónomo y sus competencias digitales, sin entrar a valorar la equidad en el acceso a internet o a dispositivos.

b) Poner en práctica, durante cuatro semanas, la tutoría personalizada con inteligencia artificial generativa en el grupo experimental y contrastar sus resultados con los del grupo control.

c) Examinar qué opinan los alumnos, una vez terminada la intervención, sobre la facilidad de uso, la accesibilidad y la contribución pedagógica de la inteligencia artificial generativa como tutor personalizado.

https://creativecommons.org/licenses/by-nc-sa/4.0/

d) Objetivo metodológico: comprobar la validez de contenido y la fiabilidad del cuestionario mediante el criterio de expertos y el Alfa de Cronbach. Al tratarse de un objetivo de método, no se juzga con los mismos criterios que los demás; el procedimiento seguido se explica en Material y métodos.

Material y métodos

Material

Para el estudio se dispuso de tres recursos. El primero fue ChatGPT (modelo GPT-5 de OpenAI, versión de agosto de 2025), la herramienta de IA generativa que sirvió de tutor al grupo experimental. El segundo, el cuestionario que se detalla en el apartado Instrumento de recolección de datos. El tercero, los programas para el análisis estadístico: Python 3.12.3 junto con las librerías libres NumPy 2.5.4, SciPy 1.18.1, pandas 3.0.6 y statsmodels 0.15.0, que hacen posible repetir el cálculo de los descriptivos, los contrastes de hipótesis, los tamaños del efecto y el ANCOVA.

Instrumento de recolección de datos

Se elaboró un cuestionario de 12 ítems con respuestas en escala Likert de cinco opciones, repartidos en cuatro bloques de tres ítems cada uno: A, rendimiento académico percibido en competencias algebraicas (ítems 1 a 3); B, aprendizaje autónomo (ítems 4 a 6); C, uso y percepción de la IA generativa como tutor personalizado (ítems 7 a 9), y D, competencias digitales e inclusión educativa (ítems 10 a 12). Lo que se recoge es la opinión del alumno sobre su propia competencia en álgebra, no su desempeño observado; por ello, cualquier ganancia indica que cambió la forma en que el estudiante se ve a sí mismo y no prueba que haya mejorado en exámenes de álgebra. Seis especialistas en estadística educativa, didáctica de la matemática, tecnología educativa e inclusión revisaron el contenido y le asignaron un promedio global de 3,87 sobre 4,00, con una razón de validez de contenido (CVR) de 1,00 en cada ítem. Los ítems 7 a 9 se aplicaron igualmente en la medición inicial, cuando los alumnos aún no habían usado la herramienta, con el fin de conocer sus expectativas; únicamente la medición final se toma como juicio sobre la experiencia, y los valores iniciales de esos ítems no se usan para medir cambio. El Anexo A presenta la estructura del cuestionario (12 ítems, escala Likert de 5 puntos) y el Anexo B, los resultados globales de la revisión de expertos. La redacción completa de los ítems y la matriz de calificación de cada experto pueden pedirse al autor de correspondencia mediante solicitud razonable. Para estimar la fiabilidad

https://creativecommons.org/licenses/by-nc-sa/4.0/

se calculó el Alfa de Cronbach de los 12 ítems y de cada bloque, por grupo y por momento; los valores obtenidos figuran en Resultados.

Métodos

Tipo y diseño de investigación

En esta investigación, el rendimiento se midió a través de la percepción que cada estudiante tiene de su competencia algebraica. Se optó por un enfoque cuantitativo, orientado a cuantificar cuánto cambiaba ese rendimiento percibido tras trabajar con la IA generativa como tutor personalizado y a recoger, con los ítems 7 a 9 del Bloque C, la opinión de los alumnos sobre la herramienta, que se describe sin más pretensión. Como esas opiniones solo se resumen de forma descriptiva y no constituyen un componente cualitativo propiamente dicho, el estudio se mantiene en el terreno cuantitativo. Al no aplicarse una prueba de desempeño en álgebra sino un cuestionario Likert, los resultados se refieren al rendimiento percibido. El diseño escogido fue cuasiexperimental, con medición inicial y final (*pretest* y *postest*) y grupo control no equivalente, un esquema común en el ámbito escolar cuando los cursos ya están conformados y no es posible repartir a los alumnos al azar (Hernández-Sampieri y Mendoza, 2023, p. 178).

Población y muestra

Se tomó como población de referencia al conjunto de alumnos inscritos en el bachillerato general unificado de la Unidad Educativa San Rafael, cantón Portoviejo, en el año lectivo 2025–2026. A partir de ella se eligió, de manera intencional y no probabilística, a 60 estudiantes de segundo de bachillerato, que se distribuyeron en dos grupos de 30. Al grupo experimental se le asignó la tutoría personalizada con IA generativa orientada a reforzar el álgebra; el grupo control continuó con la enseñanza habitual. Para participar se exigió estar matriculado de forma regular en segundo de bachillerato, tener a su disposición un equipo con acceso a la red, haber entregado firmado el formulario de consentimiento y no presentar ninguna condición que dificultara la participación activa en la intervención (Otzen y Manterola, 2017, p. 230).

Escenario o contexto de aplicación

El trabajo de campo tuvo lugar en la Unidad Educativa San Rafael, colegio de nivel medio situado en la parroquia Colón del cantón Portoviejo, capital de Manabí, en la costa central ecuatoriana. Lo

https://creativecommons.org/licenses/by-nc-sa/4.0/

que ocurre allí se repite en buena parte de los planteles públicos de la provincia: cursos de más de 35 alumnos, una conexión a internet elemental y estudiantes que necesitarían ayuda individual en álgebra fuera de clase, pero que no la tienen. Estas características convierten al colegio en un escenario realista para comprobar si la IA generativa, gracias al aprendizaje adaptativo, puede reemplazar o reforzar la atención personal que el profesor no consigue dar a grupos tan grandes y, de ese modo, elevar el rendimiento en situaciones reales de aula (Villota Sarmiento et al., 2025, p. 605).

Procedimiento de intervención

La intervención se organizó en tres fases secuenciales (diagnóstico, intervención y medición final) y se aplicó de forma diferenciada a los dos grupos.

Antes de comenzar, los 60 participantes respondieron el cuestionario inicial durante la jornada escolar, bajo la vigilancia del docente del colegio. Con esta medición se fijó el punto de partida del rendimiento algebraico percibido y se pudo comparar el nivel con el que llegaba cada grupo.

Durante cuatro semanas seguidas, el grupo experimental (n = 30) utilizó ChatGPT (modelo GPT-5 de OpenAI, al que se accedió por la web y por la aplicación móvil con una cuenta educativa gratuita) como tutor personalizado. El trabajo fue presencial y se llevó a cabo en el laboratorio de computación del colegio, en tres sesiones semanales de una hora (12 sesiones y 12 horas en total) dentro del horario pactado con las autoridades. Cada alumno dispuso de una computadora propia conectada a internet. Las sesiones estuvieron a cargo de un profesor de matemática del plantel, formado de antemano en el uso didáctico de la herramienta, que acompañó la interacción, atendió los problemas técnicos y se aseguró de que los diálogos no se apartaran del álgebra. Se controló la asistencia en las 12 sesiones y, mediante una lista de cotejo, el mismo profesor comprobó que se respetara el protocolo.

Las sesiones se distribuyeron en cuatro unidades tomadas del currículo de segundo de bachillerato: expresiones algebraicas y polinomios (sesiones 1 a 3); ecuaciones lineales y sistemas de ecuaciones (sesiones 4 a 6); factorización y productos notables (sesiones 7 a 9), y funciones cuadráticas y resolución de problemas (sesiones 10 a 12). Todas seguían la misma secuencia: una breve explicación del concepto, ejercitación guiada con ChatGPT y, al final, ejercicios para afianzar lo aprendido. En la Tabla 1 se resume la planificación por unidad.

https://creativecommons.org/licenses/by-nc-sa/4.0/

Tabla 1

Plan de sesiones de la intervención con IA generativa (grupo experimental)

Unidad	Sesiones	Objetivo	Prompt tipo	Actividad	Producto esperado
1. Expresiones algebraicas y polinomios	1 a 3	Operar con expresiones algebraicas y polinomios	Diagnóstico, andamiaje, retroalimentación y transferencia, aplicados de forma progresiva	Explicación conceptual breve, práctica guiada con ChatGPT y ejercicios de consolidación	Bitácora de la sesión: concepto trabajado, error identificado y explicación recibida
2. Ecuaciones lineales y sistemas de ecuaciones	4 a 6	Resolver ecuaciones lineales y sistemas de ecuaciones	Diagnóstico, andamiaje, retroalimentación y transferencia, aplicados de forma progresiva	Explicación conceptual breve, práctica guiada con ChatGPT y ejercicios de consolidación	Bitácora de la sesión: concepto trabajado, error identificado y explicación recibida
3. Factorización y productos notables	7 a 9	Factorizar expresiones y aplicar productos notables	Diagnóstico, andamiaje, retroalimentación y transferencia, aplicados de forma progresiva	Explicación conceptual breve, práctica guiada con ChatGPT y ejercicios de consolidación	Bitácora de la sesión: concepto trabajado, error identificado y explicación recibida
4. Funciones cuadráticas y resolución de problemas	10 a 12	Analizar funciones cuadráticas y resolver problemas algebraicos	Diagnóstico, andamiaje, retroalimentación y transferencia, aplicados de forma progresiva	Explicación conceptual breve, práctica guiada con ChatGPT y ejercicios de consolidación	Bitácora de la sesión: concepto trabajado, error identificado y explicación recibida

Nota. Cada unidad comprendió tres sesiones presenciales de una hora (12 sesiones en total). Los cuatro tipos de instrucción se describen en el texto.

A cada alumno se le entregó una guía impresa con instrucciones escritas para la IA (*prompts*) ya estructuradas, con el fin de encauzar el diálogo y prevenir un uso superficial o una dependencia de la herramienta. Las instrucciones eran de cuatro clases: de diagnóstico ("Explícame paso a paso cómo resolver [ecuación/polinomio] y dime en qué paso suelo equivocarme"), de apoyo graduado ("No entiendo este paso; dame otro ejemplo más sencillo y luego uno más difícil sobre el mismo concepto"), de corrección ("Revisa mi solución y señala el error conceptual, no solo el resultado") y de transferencia ("Propón un problema de la vida real que se resuelva con [concepto algebraico] y guíame sin darme la respuesta"). El profesor a cargo vigiló que las instrucciones se usaran en orden progresivo y que nadie copiara respuestas sin comprender el procedimiento. Al cerrar cada sesión, los alumnos anotaban en una bitácora corta qué concepto habían trabajado, qué error habían detectado y qué explicación habían recibido. Esas bitácoras se revisaron con una matriz de cinco criterios: tiempo real dedicado, cumplimiento de las actividades previstas, continuidad de la práctica, pruebas de que se resolvieron ejercicios de álgebra y coherencia de los registros. Gracias a esa revisión se constató que los dos grupos completaron las 12 horas de práctica fijadas y se identificaron las posibles incoherencias en los registros de cada alumno.

https://creativecommons.org/licenses/by-nc-sa/4.0/

En el grupo control (n = 30) se mantuvieron los métodos y materiales de siempre: clases presenciales con el mismo profesor de matemática que acompañó al grupo experimental, el libro Álgebra de Baldor (2008), que reúne 6.523 ejercicios y problemas resueltos, tareas en el cuaderno y corrección colectiva en clase. Este grupo no usó ChatGPT ni recibió tutoría personalizada con IA mientras duró la intervención. Con el fin de controlar el efecto de la novedad, trabajó las mismas cuatro unidades con idéntica carga de práctica en álgebra (12 horas en total), pero sin IA generativa. Se llevó la cuenta del tiempo que cada alumno dedicaba a las actividades de aprendizaje, y también aquí se usó una bitácora de práctica que se revisó con la misma matriz.

Concluidas las cuatro semanas, ambos grupos volvieron a responder el mismo cuestionario de 12 ítems, en condiciones de aplicación similares a las de la primera vez, con el objeto de medir el efecto diferencial de la intervención (Creswell y Creswell, 2023, p. 214).

Técnicas de análisis de datos

El tratamiento de los datos fue cuantitativo y partió de las puntuaciones de cada uno de los 60 alumnos. Para los 12 indicadores y para los cuatro bloques (promedio de sus tres ítems) se obtuvieron medias, desviaciones estándar y frecuencias de logro, por grupo y en ambas mediciones. Dado que la puntuación total mostró poca consistencia interna, la suma de los 12 ítems no se utilizó para contrastar hipótesis. La comparación entre el grupo experimental, que contó con tutoría de IA generativa, y el grupo control se hizo por bloque, con la prueba t de Student para muestras independientes sobre la medición final y la corrección de Holm para los cuatro contrastes, y por ítem, con la U de Mann-Whitney y la misma corrección. H_0 y H_1 se evalúan con los contrastes por bloque. Se revisó la normalidad con la prueba de Shapiro-Wilk en cada grupo y la igualdad de varianzas con la de Levene; si alguno de estos supuestos fallaba, el resultado se comprobó con la prueba de Welch y con la U de Mann-Whitney. Para examinar si los grupos partían del mismo nivel se aplicó la t de Student por bloque a la medición inicial. Puesto que ambas mediciones no son independientes, se ajustó también un análisis de covarianza (ANCOVA) por bloque, que tomó la medición final como respuesta, el grupo como factor y la medición inicial del bloque como variable de control; se comprobó que las pendientes fueran homogéneas (interacción grupo × medición inicial), que los residuos siguieran una distribución normal (Shapiro-Wilk) y que su varianza fuera constante (Breusch-Pagan). Con la ganancia normalizada de Hake (1998), g, se expresó la mejora percibida como proporción de la mejora máxima posible; al calcularse sobre

https://creativecommons.org/licenses/by-nc-sa/4.0/

percepciones, no se lee como aprendizaje alcanzado. El tamaño del efecto de cada bloque se obtuvo con la *d* de Cohen en la medición final, esto es, la diferencia de medias entre los grupos experimental y control dividida por la desviación estándar conjunta ($d = (M_{exp} - M_{ctrl}) / DE_{agrupada}$), acompañada de un intervalo de confianza del 95% calculado de forma aproximada a partir de su error estándar, y se calificó según los criterios de Cohen (1988). Este índice no se aplicó a las ganancias. La consistencia interna se examinó con el Alfa de Cronbach, el omega de McDonald (modelo de un factor por ejes principales), las correlaciones ítem-total corregidas, el alfa resultante al suprimir cada ítem y la correlación media entre ítems. Para explorar la estructura se emplearon el índice KMO, la prueba de Bartlett, el análisis paralelo y un análisis factorial exploratorio de un factor por ejes principales en cada medición, sin juntar ambas mediciones, ya que se trata de medidas repetidas.

Debido a la baja consistencia interna del cuestionario (α entre −,32 y ,13 según grupo y momento), los contrastes de hipótesis (t de Student, ANCOVA y tamaños del efecto) se realizaron por bloque y deben tomarse como exploratorios. La puntuación total de los 12 ítems solo aparece como promedio descriptivo y no sustenta ninguna conclusión estadística. En consecuencia, lo que aquí se reporta es una primera aproximación al efecto y no una confirmación de la hipótesis.

Consideraciones éticas y protección de datos

Como casi todos los participantes eran menores de edad, el estudio se guio por los principios de autonomía, beneficencia, no maleficencia y justicia. Antes de iniciar, los representantes legales de los 60 alumnos firmaron el consentimiento informado y los propios estudiantes dieron su asentimiento. Solo firmaron quienes figuraban como representantes en los registros de la Unidad Educativa San Rafael, lo cual se comprobó con la nómina de matrícula, donde consta esa representación. El asentimiento de cada alumno quedó registrado y se cotejó con la misma lista. Los estudiantes que no quisieron participar o que no tenían autorización continuaron con sus clases y no se tomaron en cuenta ni en la recolección ni en el análisis de datos. En ambos documentos se informaba el propósito del estudio, el carácter voluntario de la participación, la posibilidad de retirarse cuando se deseara sin perjuicio académico, la duración de cuatro semanas y el uso exclusivamente académico de la información.

https://creativecommons.org/licenses/by-nc-sa/4.0/

La Unidad Educativa San Rafael aprobó el protocolo con el código SR-2025-001, tras revisar que se ajustaba a la normativa nacional e internacional para estudios con menores, y permitió aplicar los cuestionarios y usar ChatGPT en el horario convenido con sus directivos. El permiso se tramitó por medio del Distrito 13D01 Portoviejo y del propio colegio. El estudio no pasó por un comité de ética institucional, de modo que la aprobación de la Unidad Educativa San Rafael fue la única autorización y cumplió ese papel. Esto se consideró adecuado porque el riesgo era mínimo: se trató de actividades de álgebra realizadas en el colegio y en horas de clase, con cuestionarios de percepción cuyos datos se codificaron, sin procedimientos invasivos ni recolección de información sensible, y con participación libre. Dado ese riesgo mínimo, el aval del Distrito 13D01 Portoviejo y de la Unidad Educativa San Rafael se estimó suficiente.

La información de los menores se resguardó con cuatro medidas. En primer lugar, nombres, apellidos, números de cédula y cualquier otro dato que permitiera identificar a los alumnos se sustituyeron por códigos alfanuméricos en todas las etapas, de la recolección al análisis. En segundo lugar, los datos sirvieron únicamente para esta investigación: no se entregaron a terceros ni se utilizaron en calificaciones oficiales. En tercer lugar, la información se almacenó en un equipo protegido con contraseña y de acceso limitado, y se borrará de forma segura una vez publicados los resultados o, a más tardar, a los dos años. Finalmente, se indicó a los alumnos que no escribieran en ChatGPT información personal, familiar o delicada; el diálogo se limitó al álgebra, y el equipo investigador no recopiló ni guardó en servidores propios las conversaciones con la IA. El manejo de los datos personales siguió la Ley Orgánica de Protección de Datos Personales del Ecuador, y las medidas descritas se corresponden con sus principios de finalidad, minimización y confidencialidad.

Resultados

Descripción de la muestra

Los resultados se exponen en seis apartados, previstos desde el diseño: (1) descriptivos de la medición inicial, (2) porcentaje de logro en ese diagnóstico, (3) comparación de los descriptivos según grupo y momento, (4) ganancia percibida del grupo experimental, (5) porcentaje de logro por indicador después de la tutoría con IA generativa y (6) opinión de los alumnos sobre ChatGPT como tutor personalizado. Participaron 60 jóvenes del segundo año de bachillerato del colegio,

https://creativecommons.org/licenses/by-nc-sa/4.0/

repartidos en un grupo experimental (n = 30) y un grupo control (n = 30). Todos contestaron los 12 ítems en las dos mediciones, por lo que no hubo deserciones ni datos faltantes.

Fiabilidad del instrumento: coeficiente alfa de Cronbach

El cuestionario de 12 ítems mostró poca consistencia interna. En la muestra completa (n = 60), el Alfa de Cronbach fue α = −,06 en la medición inicial y α = ,78 en la final; dentro de cada grupo y momento varió entre α = −,32 y α = ,13. Prácticamente todos los valores están por debajo del mínimo aceptable, α = ,70 (Maldonado y Santoyo, 2024, p. 8). La única excepción, la medición final de la muestra completa, no prueba que la escala sea fiable, pues ese valor se debe sobre todo a la distancia entre los grupos y no a que los ítems varíen juntos dentro de cada uno. Por consiguiente, la puntuación total debe tratarse con prudencia, como un promedio de percepciones y no como una escala unidimensional confiable. El omega de McDonald, calculado con un modelo de un factor, dio ω = ,44 en la medición inicial y ω = ,78 en la final para la muestra completa, y entre ,46 y ,59 dentro de cada grupo y momento. Los datos tampoco respaldan una estructura factorial. En la medición inicial, el KMO fue de ,47 (inaceptable), la prueba de esfericidad de Bartlett no resultó significativa ($\chi^2(66) = 58{,}2$; p = ,743) y el análisis paralelo no conservó ningún factor. En la final, el KMO subió a ,71, la esfericidad sí fue significativa ($\chi^2(66) = 152{,}3$; $p < ,001$) y se conservó un factor, explicable en parte por la diferencia entre grupos. Con solo 60 alumnos no cabe un análisis factorial confirmatorio, y el exploratorio no permite sacar conclusiones. Por todo ello, los contrastes que se presentan enseguida tienen carácter exploratorio.

Diagnóstico adicional del instrumento. En la medición inicial de la muestra completa, 7 de los 12 ítems tuvieron correlaciones ítem-total corregidas negativas (de −,18 a ,16); en la final no hubo ninguna (de ,27 a ,56). Suprimir un ítem cualquiera casi no alteró el alfa (máximo ,048 en la medición inicial y ,775 en la final; Tabla S2). La correlación media entre ítems de la muestra completa fue de −,006 al inicio y de ,223 al final; dentro de cada grupo se movió entre −,020 y ,014, y por bloque, entre −,08 y ,01 al inicio y entre ,19 y ,26 al final. El patrón de correlaciones tampoco se mantuvo estable (correlación entre las matrices inicial y final, ,03; entre grupos, −,13 y ,05). Con un factor, el análisis factorial exploratorio dio cuenta del 7,8% de la varianza en la medición inicial (ocho ítems con carga inferior a ,30) y del 23,3% en la final (cargas de ,30 a ,63), con las reservas ya mencionadas sobre la diferencia entre grupos y el tamaño muestral. Se observó efecto techo en la medición final del grupo experimental (51,7% de respuestas con valor 5) y efecto

https://creativecommons.org/licenses/by-nc-sa/4.0/

suelo en la inicial (28,3% y 31,4% con valor 1). Así pues, antes de emplearlo de nuevo, el cuestionario debe someterse a una revisión psicométrica que incluya reescribir ítems y volver a validarlo.

Tabla 2

Correlaciones ítem-total corregidas y alfa si se elimina el ítem (muestra total, n = 60)

N°	Variable / Indicador	*r* ítem-total pretest	*α* si se elimina (pretest)	*r* ítem-total postest	*α* si se elimina (postest)
1	Comprensión procedimental algebraica	,04	−,103	,30	,774
2	Generalización simbólica	,06	−,110	,39	,765
3	Autorregulación algebraica	−,11	,006	,46	,756
4	Gestión autónoma del tiempo	−,02	−,059	,44	,759
5	Iniciativa de aprendizaje autónomo	,04	−,100	,27	,775
6	Autoeficacia digital en álgebra	,16	−,207	,56	,746
7	Utilidad pedagógica de la IA generativa	−,18	,048	,45	,759
8	Efectividad retroalimentación tutor IA	−,08	−,016	,35	,769
9	Accesibilidad del tutor IA	,05	−,108	,42	,761
10	Desarrollo de competencias digitales	−,12	,018	,52	,750
11	Percepción de equidad e inclusión	−,03	−,048	,45	,758
12	Motivación con uso responsable de la IA	−,01	−,064	,35	,769
	Alfa de Cronbach (12 ítems)		**−,064**		**,777**

Nota. *r* ítem-total = correlación del ítem con la suma de los otros 11. El *α* del postest de la muestra total no indica fiabilidad de la escala: refleja en parte la diferencia entre los grupos. Dentro de cada grupo, *α* osciló entre −,32 y ,13.

https://creativecommons.org/licenses/by-nc-sa/4.0/

Análisis de los resultados

Estadísticos descriptivos del pretest según la escala de estimación

Tabla 3

Estadísticos descriptivos del pretest por indicador (n = 60)

N°	Variable / Indicador	Bloque	*M*	*DE*	Nivel
1	Comprensión procedimental algebraica	A	2,50	1,32	Moderado
2	Generalización simbólica	A	2,23	1,23	Bajo
3	Autorregulación algebraica	A	2,72	1,32	Moderado
4	Gestión autónoma del tiempo	B	2,53	1,27	Moderado
5	Iniciativa de aprendizaje autónomo	B	2,45	1,33	Bajo
6	Autoeficacia digital en álgebra	B	2,73	1,44	Moderado
7	Utilidad pedagógica de la IA generativa	C	2,57	1,21	Moderado
8	Efectividad retroalimentación tutor IA	C	2,40	1,32	Bajo
9	Accesibilidad del tutor IA	C	2,48	1,24	Bajo
10	Desarrollo de competencias digitales	D	2,65	1,44	Moderado
11	Percepción de equidad e inclusión	D	2,75	1,30	Moderado
12	Motivación con uso responsable de la IA	D	2,52	1,24	Moderado
	Promedio general		**2,54**	**0,37**	**Moderado**

Nota. *M* = media; *DE* = desviación estándar. En la fila Promedio general, M y DE corresponden a la puntuación total individual (media de los 12 ítems). Escala de estimación: 1,00–1,49 Muy bajo; 1,50–2,49 Bajo; 2,50–3,49 Moderado; 3,50–4,49 Alto; 4,50–5,00 Muy alto. El Bloque A (ítems 1 a 3) mide el rendimiento académico que percibe el propio estudiante, no el desempeño observado, y se evalúa con una escala Likert.

Al comienzo, las medias de los 12 indicadores se situaron entre 2,23 y 2,75 (Tabla 2), con 8 indicadores en el nivel Moderado y 4 en el Bajo de la escala de estimación. El promedio general fue *M* = 2,54 (*DE* = 0,37) y ningún indicador alcanzó el umbral de logro de *M* = 3,50. Es decir, antes de la intervención los alumnos tenían un nivel insuficiente en competencias algebraicas, aprendizaje autónomo y competencias digitales, lo cual justifica haber puesto en marcha la propuesta. Más adelante se compara el punto de partida de cada grupo (Tablas 3 y 4).

https://creativecommons.org/licenses/by-nc-sa/4.0/

Porcentaje de estudiantes que alcanzan cada indicador en el pretest

Tabla 4

Porcentaje de logro por indicador en el pretest (n = 60)

N°	Variable / Indicador	Frec. logro	% Total	% Ctrl	% Exp
1	Comprensión procedimental algebraica	19	31,7%	30,0%	33,3%
2	Generalización simbólica	11	18,3%	20,0%	16,7%
3	Autorregulación algebraica	21	35,0%	36,7%	33,3%
4	Gestión autónoma del tiempo	14	23,3%	16,7%	30,0%
5	Iniciativa de aprendizaje autónomo	16	26,7%	26,7%	26,7%
6	Autoeficacia digital en álgebra	22	36,7%	30,0%	43,3%
7	Utilidad pedagógica de la IA generativa	14	23,3%	23,3%	23,3%
8	Efectividad retroalimentación tutor IA	14	23,3%	26,7%	20,0%
9	Accesibilidad del tutor IA	14	23,3%	20,0%	26,7%
10	Desarrollo de competencias digitales	18	30,0%	23,3%	36,7%
11	Percepción de equidad e inclusión	20	33,3%	40,0%	26,7%
12	Motivación con uso responsable de la IA	14	23,3%	26,7%	20,0%
	Promedio general	—	**27,4%**	**26,7%**	**28,1%**

Nota. Criterio de logro: puntaje ≥ 4 en escala Likert (De acuerdo / Totalmente de acuerdo). Frec. = frecuencia de estudiantes con puntaje ≥ 4.

Los porcentajes de logro de la medición inicial confirman que los dos grupos empezaban desde un nivel bajo (Tabla 3). Considerando a los 60 alumnos (n = 60), la proporción con puntaje ≥ 4 en cada indicador osciló entre 18,3% (ítem 2) y 36,7% (ítem 6), con una media de 27,4%. El grupo experimental arrancó con un logro medio de 28,1% y el control con 26,7%, una brecha reducida que apunta a un nivel inicial parecido; la equivalencia estadística se examina más adelante (Tabla 4). Antes de introducir la IA generativa como tutor, en promedio el 72,6% de los alumnos no llegaba al puntaje esperado en cada indicador.

https://creativecommons.org/licenses/by-nc-sa/4.0/

Estadísticos descriptivos por grupo y momento de medición

Tabla 5

Comparación de medias por grupo y momento de medición (pretest y postest)

Bloque / Dimensión	Ctrl Pretest *M* (*DE*)	Ctrl Postest *M* (*DE*)	Δ Ctrl	Exp. Pretest *M* (*DE*)	Exp. Postest *M* (*DE*)	Δ Exp.
A. Rendimiento algebraico	2,38 (0,77)	2,99 (0,67)	+0,61	2,59 (0,73)	4,28 (0,61)	+1,69
B. Aprendizaje autónomo	2,43 (0,71)	2,89 (0,70)	+0,46	2,71 (0,71)	4,11 (0,57)	+1,40
C. Uso y percepción IA	2,44 (0,69)	2,99 (0,86)	+0,54	2,52 (0,65)	4,17 (0,55)	+1,64
D. Competencias digitales	2,70 (0,72)	2,81 (0,89)	+0,11	2,58 (0,74)	4,17 (0,59)	+1,59
Promedio de los 12 ítems (solo descriptivo)	**2,49 (0,39)**	**2,92 (0,37)**	**+0,43**	**2,60 (0,34)**	**4,18 (0,30)**	**+1,58**

Nota. *M* = media (desviación estándar entre paréntesis); Δ = diferencia postest–pretest. Ctrl = grupo control; Exp. = grupo experimental. El Bloque A mide el rendimiento que percibe el propio estudiante (escala Likert), no el desempeño observado en una prueba.

En la Tabla 4 figuran, por bloque, las medias y desviaciones estándar de cada grupo en las dos mediciones. El grupo control avanzó de forma moderada, con ganancias que fueron de Δ = +0,11 (Bloque D; IC del 95% [−0,38; 0,60]) a Δ = +0,61 (Bloque A; IC del 95% [0,28; 0,94]); en el Bloque D el intervalo contiene el cero. En el grupo experimental, las ganancias se ubicaron entre Δ = +1,40 (Bloque B; IC del 95% [1,08; 1,72]) y Δ = +1,69 (Bloque A; IC del 95% [1,30; 2,08]), siendo el Bloque A, de rendimiento algebraico percibido, el que más creció. Solo a título descriptivo, el promedio de los 12 ítems subió de 2,60 a 4,18 en el grupo experimental y de 2,49 a 2,92 en el control; este promedio no intervino en los contrastes.

Equivalencia inicial. En la primera medición, el grupo experimental tuvo valores algo mayores que el control en los Bloques A (+0,21), B (+0,28) y C (+0,08), y algo menores en el D (−0,12). Ninguna de estas diferencias alcanzó significación estadística (*p* entre ,134 y ,656; *d* de −0,17 a 0,39). Ahora bien, la falta de significación no basta para afirmar que los grupos sean equivalentes, y el experimental llevaba una pequeña ventaja en tres de los cuatro bloques; por ese motivo, la medición final se ajustó por la inicial en cada bloque. Al no ser grupos equivalentes ni formados al azar, y al repetirse el mismo cuestionario en los dos momentos, la diferencia observada al final es coherente con un efecto de la tutoría con IA generativa, aunque no lo prueba.

Comparación en la medición final. Al comparar las medias con la prueba *t* para grupos independientes, se observó que el grupo experimental obtuvo puntuaciones más altas que el control en todos los bloques (Tabla 5), con diferencias que iban de 1,18 puntos en el Bloque C a 1,36

https://creativecommons.org/licenses/by-nc-sa/4.0/

puntos en el Bloque D, todas con $p < ,001$ incluso después de aplicar la corrección de Holm. En el Bloque A, que recoge el rendimiento algebraico percibido, el grupo experimental promedió M = 4,28 (DE = 0,61) frente a M = 2,99 (DE = 0,67) del control, una distancia de 1,29 puntos (IC del 95% [0,96; 1,62]; $t(58) = 7,81$; $p < ,001$). Con estos contrastes se pone a prueba la hipótesis en cada bloque. Valores de $p < ,001$ hacen difícil sostener H_0 (ausencia de diferencias entre los grupos en las medias del bloque en la medición final), y el sentido de las diferencias, favorable al grupo experimental, coincide con lo que anticipa H_1. Aun así, al tratarse de un análisis exploratorio, no se interpreta como prueba concluyente de la hipótesis. La igualdad de varianzas se cumplió en los Bloques A, B y D (Levene, p entre ,09 y 1,00), no así en el C (p = ,04); además, la normalidad del grupo experimental fue cuestionable en los Bloques A, C y D (Shapiro-Wilk, $p < ,05$). Por esa razón, los resultados se verificaron con la prueba de Welch (Bloque C: t = 6,34; $p < ,001$) y con la U de Mann-Whitney (todas $p < ,001$).

Tabla 6

Comparación entre grupos en el postest, tamaños del efecto y ANCOVA por bloque

Bloque / Dimensión	Dif. de medias postest [IC 95%]	*t(58)*	*p*	*d* [IC 95%]	Dif. ajustada (ANCOVA) [IC 95%]
A. Rendimiento algebraico	1,29 [0,96; 1,62]	7,81	< ,001	2,02 [1,40; 2,64]	1,28 [0,95; 1,62]
B. Aprendizaje autónomo	1,22 [0,89; 1,55]	7,40	< ,001	1,91 [1,30; 2,52]	1,26 [0,92; 1,59]
C. Uso y percepción IA	1,18 [0,81; 1,55]	6,34	< ,001	1,64 [1,05; 2,22]	1,16 [0,79; 1,54]
D. Competencias digitales	**1,36 [0,97; 1,75]**	**6,96**	**< ,001**	**1,80 [1,20; 2,40]**	**1,31 [0,94; 1,68]**

Nota. Dif. = media experimental − media control en el postest; IC = intervalo de confianza del 95%; d = d de Cohen con desviación estándar agrupada; Dif. ajustada = diferencia entre grupos ajustada por el pretest del bloque (ANCOVA). Prueba t de Student para muestras independientes (n = 30 por grupo). La puntuación total de los 12 ítems no se usó para la inferencia.

Tamaño del efecto y ajuste por la medición inicial. Con los criterios de Cohen (1988), el efecto resultó muy grande en los cuatro bloques: d = 2,02 (IC del 95% [1,40; 2,64]) en el A, 1,91 en el B, 1,64 en el C y 1,80 en el D (Tabla 5). Magnitudes de este orden piden una lectura prudente. Al controlar la medición inicial de cada bloque, el ANCOVA conservó las diferencias, que quedaron entre 1,16 y 1,31 puntos (η^2 parcial de ,41 a ,51; todas $p < ,001$); en el Bloque A, la diferencia ajustada fue de 1,28 puntos (IC del 95% [0,95; 1,62]; $F(1, 57) = 58,2$). En ningún bloque fue significativa la interacción entre grupo y medición inicial (p entre ,06 y ,90). Los residuos se

https://creativecommons.org/licenses/by-nc-sa/4.0/

distribuyeron de forma normal en tres bloques (Shapiro-Wilk, $p > ,05$; en el Bloque D, $p = ,048$) y mostraron varianza constante en tres (Breusch-Pagan, $p > ,05$; en el Bloque C, $p = ,014$).

Análisis por ítem. Puesto que la puntuación total no constituye una escala fiable, se repitió la comparación con cada ítem por separado. En la medición final, el grupo experimental puntuó por encima del control en los 12 ítems (diferencias entre 0,67 y 1,60 puntos; *d* de 0,65 a 1,34; U de Mann-Whitney con corrección de Holm, *p* ajustado ≤ ,018; Tabla S1), mientras que en la inicial ningún ítem mostró diferencias entre grupos. Se decidió no suprimir ítems para respetar el cuestionario tal como fue aplicado y validado por los expertos.

Tabla S1

Comparación entre grupos por ítem en el postest (U de Mann-Whitney con corrección de Holm)

N°	Variable / Indicador	Exp. *M* (*DE*)	Ctrl *M* (*DE*)	Dif.	*d*	*p* (*U*)	*p* Holm	*p* pretest
1	Comprensión procedimental algebraica	4,13 (1,01)	2,80 (1,35)	1,33	1,12	< ,001	,001	,383
2	Generalización simbólica	4,47 (0,86)	3,40 (1,43)	1,07	0,90	,001	,005	,775
3	Autorregulación algebraica	4,23 (1,01)	2,77 (1,28)	1,47	1,28	< ,001	< ,001	,257
4	Gestión autónoma del tiempo	4,13 (0,97)	2,73 (1,23)	1,40	1,26	< ,001	< ,001	,331
5	Iniciativa de aprendizaje autónomo	3,97 (0,96)	3,30 (1,09)	0,67	0,65	,018	,018	,909
6	Autoeficacia digital en álgebra	4,23 (0,97)	2,63 (1,38)	1,60	1,34	< ,001	< ,001	,262
7	Utilidad pedagógica de la IA generativa	4,30 (0,84)	2,97 (1,33)	1,33	1,20	< ,001	< ,001	,957
8	Efectividad retroalimentación tutor IA	4,17 (1,12)	3,13 (1,36)	1,03	0,83	,003	,008	,982
9	Accesibilidad del tutor IA	4,03 (0,96)	2,87 (1,36)	1,17	0,99	< ,001	,005	,790
10	Desarrollo de competencias digitales	4,07 (0,98)	2,63 (1,40)	1,43	1,19	< ,001	< ,001	,355
11	Percepción de equidad e inclusión	4,20 (1,03)	2,67 (1,47)	1,53	1,21	< ,001	< ,001	,195
12	Motivación con uso responsable de la IA	4,23 (1,01)	3,13 (1,48)	1,10	0,87	,003	,008	,215

Nota. Exp. = grupo experimental; Ctrl = grupo control; *M* = media; *DE* = desviación estándar; Dif. = diferencia de medias; *d* = *d* de Cohen con desviación estándar agrupada; *p* (*U*) = valor *p* de la prueba *U* de Mann-Whitney; *p* Holm = valor *p* ajustado por Holm para los 12 contrastes; *p* pretest = valor *p* de la misma comparación en el pretest.

Ganancia percibida entre la medición inicial y la final en el grupo experimental

Tanto la ganancia absoluta (Δ) como la ganancia normalizada de Hake (g) de cada indicador se obtuvieron a partir de las medias grupales, según lo previsto en el método. Reflejan, por tanto, el cambio del grupo en su conjunto y no el progreso individual de cada alumno. El ANCOVA, que ajusta por la línea base, se presentó en el apartado previo.

https://creativecommons.org/licenses/by-nc-sa/4.0/

Tabla 7

Ganancia absoluta y normalizada (Hake) por indicador. Grupo experimental (n = 30)

N°	Variable / Indicador	*M* Pretest *(DE)*	*M* Postest *(DE)*	Δ Abs.	*g* Hake	Nivel
1	Comprensión procedimental algebraica	2,63 (1,30)	4,13 (1,01)	+1,50	0,63	Moderado
2	Generalización simbólica	2,20 (1,27)	4,47 (0,86)	+2,27	0,81	Alto
3	Autorregulación algebraica	2,93 (1,26)	4,23 (1,01)	+1,30	0,63	Moderado
4	Gestión autónoma del tiempo	2,73 (1,41)	4,13 (0,97)	+1,40	0,62	Moderado
5	Iniciativa de aprendizaje autónomo	2,47 (1,33)	3,97 (0,96)	+1,50	0,59	Moderado
6	Autoeficacia digital en álgebra	2,93 (1,34)	4,23 (0,97)	+1,30	0,63	Moderado
7	Utilidad pedagógica de la IA generativa	2,57 (1,19)	4,30 (0,84)	+1,73	0,71	Alto
8	Efectividad retroalimentación tutor IA	2,43 (1,41)	4,17 (1,12)	+1,73	0,68	Moderado
9	Accesibilidad del tutor IA	2,57 (1,38)	4,03 (0,96)	+1,47	0,60	Moderado
10	Desarrollo de competencias digitales	2,83 (1,51)	4,07 (0,98)	+1,23	0,57	Moderado
11	Percepción de equidad e inclusión	2,53 (1,31)	4,20 (1,03)	+1,67	0,68	Moderado
12	Motivación con uso responsable de la IA	2,37 (1,38)	4,23 (1,01)	+1,87	0,71	Alto
	Promedio general	**2,60 (0,34)**	**4,18 (0,30)**	**+1,58**	**0,66**	**Moderado**

Nota. g = ganancia normalizada de Hake (1998): $g = (M_{post}–M_{pre})/(5–M_{pre})$. M = media (desviación estándar entre paréntesis). Interpretación: $g < 0{,}30$ Bajo; $0{,}30 \leq g < 0{,}70$ Moderado; $g \geq 0{,}70$ Alto. En los ítems 7 a 9 (Bloque C), el pretest se aplicó antes del uso de la herramienta y representa expectativa previa; solo el postest se interpreta como valoración, por lo que la ganancia de esos ítems no se interpreta como cambio en la percepción.

Todos los indicadores mejoraron en el grupo experimental (Tabla 6). En promedio, la ganancia absoluta fue de Δ = +1,58 puntos; la menor correspondió al ítem 10, Desarrollo de competencias digitales (Δ = +1,23), y la mayor al ítem 2, Generalización simbólica (Δ = +2,27). La ganancia normalizada global de Hake fue g = 0,66, que corresponde a la categoría Moderado y queda próxima al límite de ganancia alta ($g \geq 0{,}70$). Los ítems 2, 7 y 12 superaron ese límite, y el resto se situó entre g = 0,57 y g = 0,68.

https://creativecommons.org/licenses/by-nc-sa/4.0/

Porcentaje de logro por indicador en el grupo experimental

Tabla 8

Porcentaje de logro pretest vs. postest por indicador. Grupo experimental (n = 30)

N°	Variable / Indicador	% Logro Pretest	% Logro Postest	Incremento (pp)	Impacto
1	Comprensión procedimental algebraica	33,3%	70,0%	+36,7%	Alto
2	Generalización simbólica	16,7%	83,3%	+66,7%	Muy alto
3	Autorregulación algebraica	33,3%	73,3%	+40,0%	Muy alto
4	Gestión autónoma del tiempo	30,0%	73,3%	+43,3%	Muy alto
5	Iniciativa de aprendizaje autónomo	26,7%	66,7%	+40,0%	Alto
6	Autoeficacia digital en álgebra	43,3%	70,0%	+26,7%	Alto
7	Utilidad pedagógica de la IA generativa	23,3%	76,7%	+53,3%	Muy alto
8	Efectividad retroalimentación tutor IA	20,0%	66,7%	+46,7%	Muy alto
9	Accesibilidad del tutor IA	26,7%	63,3%	+36,7%	Alto
10	Desarrollo de competencias digitales	36,7%	70,0%	+33,3%	Alto
11	Percepción de equidad e inclusión	26,7%	76,7%	+50,0%	Muy alto
12	Motivación con uso responsable de la IA	20,0%	80,0%	+60,0%	Muy alto
	Promedio general	**28,1%**	**72,5%**	**+44,4%**	**Muy alto**

Nota. Criterio de logro: puntaje ≥ 4. Incremento = % Logro Postest – % Logro Pretest, en puntos porcentuales (pp). Nivel de impacto: Leve Δ < 10%; Moderado Δ ≥ 10%; Alto Δ ≥ 25%; Muy alto Δ ≥ 40%. En los ítems 7 a 9 (Bloque C), el pretest se aplicó antes del uso de la herramienta y representa expectativa previa; solo el postest se interpreta como valoración, por lo que la ganancia de esos ítems no se interpreta como cambio en la percepción.

La evolución del porcentaje de logro se muestra en la Tabla 7. En el grupo experimental, el logro medio subió de 28,1% al inicio a 72,5% al final, es decir, +44,4 puntos porcentuales, lo que en la escala adoptada equivale a un impacto Muy alto; en el grupo control, el aumento fue de 26,7% a 35,8%. El ítem 2 registró el mayor avance (+66,7 puntos porcentuales) y el ítem 6, el menor (+26,7). Al final, ningún indicador quedó por debajo del 63,3% de logro.

Percepción de los estudiantes sobre la IA generativa como tutor personalizado

Para conocer la opinión del grupo experimental sobre ChatGPT se usaron los ítems 7, 8 y 9 del Bloque C, referidos a la utilidad pedagógica, la calidad de la retroalimentación y la accesibilidad del tutor (Tablas 4, 6 y 7). En la medición final, el Bloque C obtuvo M = 4,17 (DE = 0,55), valor que corresponde al nivel Alto de la escala de estimación. Por ítem, la utilidad pedagógica alcanzó M = 4,30, la retroalimentación M = 4,17 y la accesibilidad M = 4,03. Respecto al logro, dieron un puntaje ≥ 4 el 76,7% de los alumnos en utilidad, el 66,7% en retroalimentación y el 63,3% en accesibilidad, que fue el aspecto menos apreciado del bloque. Estos mismos ítems se aplicaron en la medición inicial, cuando los alumnos todavía no conocían la herramienta (Bloque C del grupo

https://creativecommons.org/licenses/by-nc-sa/4.0/

experimental: M = 2,52). Aquellos valores expresan expectativas y no una experiencia de uso, de modo que solo la medición final sirve como juicio sobre el tutor. Por la misma razón, la diferencia entre ambos momentos no se interpreta como un cambio de opinión.

Vistos en conjunto, estos datos muestran que, al cabo de 12 sesiones, los alumnos consideraron la herramienta útil para aprender álgebra y acertada en sus correcciones, y que su utilidad pedagógica fue el aspecto más apreciado. La opinión positiva alcanzó también la equidad y el uso responsable: el ítem 11 (Percepción de equidad e inclusión) subió de M = 2,53 a M = 4,20 (Δ = +1,67) y el ítem 12 (Motivación con uso responsable de la IA), de M = 2,37 a M = 4,23 (Δ = +1,87), con un logro final de 76,7% y 80,0%, respectivamente. Tomando como referencia la medición final, que es la única que expresa una valoración, se da por cumplido el tercer objetivo específico: los alumnos juzgaron de manera favorable la facilidad de uso, la accesibilidad y la contribución pedagógica de la IA generativa.

Discusión

Según estos resultados, haber usado la IA generativa como tutor personalizado durante cuatro semanas, dentro de un diseño cuasiexperimental, se acompañó de mejoras en los 12 indicadores evaluados. Fuentes Cabrera et al. (2025) ya habían sugerido la evaluación formativa con apoyo de IA como vía para ofrecer retroalimentación oportuna. En el presente caso, el Bloque A (rendimiento algebraico percibido) aumentó Δ = +1,69 puntos en el grupo experimental y Δ = +0,61 en el control; en la medición final la brecha fue de 1,29 puntos (IC del 95% [0,96; 1,62]; $t(58)$ = 7,81; p < ,001) y se sostuvo al ajustar por la medición inicial (diferencia ajustada = 1,28; IC del 95% [0,95; 1,62]). En los demás bloques, las diferencias se movieron entre 1,18 y 1,36 puntos. Con la cautela de un análisis exploratorio, estos valores apuntan en el sentido de H_1 (el grupo experimental obtendrá medias superiores en la medición final), sin que ello equivalga a confirmarla. No se hallaron trabajos con igual población, instrumento y diseño que permitieran contrastar directamente la magnitud del efecto.

En los metaanálisis disponibles sobre IA y aprendizaje de las matemáticas, los efectos van de pequeños a moderados; Hwang (2022), por ejemplo, encontró un efecto pequeño (g = 0,351) en alumnos de primaria. En el nivel secundario, Xing et al. (2025), mediante un estudio de diseño en tres fases que terminó en un cuasiexperimento, observaron que los alumnos que interactuaron con

https://creativecommons.org/licenses/by-nc-sa/4.0/

un agente de IA generativa ganaron significativamente más conocimiento que el grupo control. Por su parte, Trocado et al. (2026), en una experiencia cualitativa en la que 26 estudiantes de décimo grado usaron ChatGPT como tutor de funciones cuadráticas, destacaron su potencial para el razonamiento matemático. Las magnitudes halladas aquí superan con mucho el efecto informado por Hwang (2022), pero provienen de percepciones y, por tanto, no pueden equipararse con efectos medidos en pruebas de rendimiento. Esta prudencia concuerda con una revisión sistemática reciente, que encontró que los estudios con IA generativa del tipo de ChatGPT informan sobre todo de opiniones favorables y de mayor implicación, pero aportan escasa evidencia de mejoras directas en el logro (Molemane et al., 2026). En el grupo experimental, el porcentaje de logro pasó de 28,1% a 72,5% entre ambas mediciones (+44,4 puntos porcentuales), y el tamaño del efecto fue muy grande en los cuatro bloques (*d* entre 1,64 y 2,02) conforme a los criterios de Cohen (1988). Conviene recordar que el colegio tiene cursos de más de 35 alumnos, en los que resulta complicado atender a cada estudiante de forma individual.

Dado que la consistencia interna del cuestionario fue baja (α entre −,32 y ,13 dentro de cada grupo y momento), estos efectos solo pueden leerse como exploratorios. Los 12 ítems no varían de forma conjunta lo suficiente como para constituir una escala unidimensional, y el puntaje total agrupa percepciones de naturaleza distinta. Por otra parte, resulta poco creíble que una intervención de cuatro semanas produzca un efecto de esa magnitud; es posible que influyan la tendencia a dar respuestas socialmente aceptables, las expectativas o la familiaridad con la herramienta, más que una mejora real de la competencia. Con todo, el mismo patrón aparece en los bloques, en cada ítem y después de ajustar por la medición inicial. Vilchez Ruiz (2024) recomienda establecer normas, formar éticamente a docentes y alumnos y promover la autonomía para evitar la dependencia de la tecnología. Esa recomendación no puede valorarse con este estudio, que no midió la supervisión del profesor ni el uso ético de la herramienta. El Bloque A (rendimiento algebraico percibido) presentó la mayor ganancia (Δ = +1,69) y el Bloque D (competencias digitales e inclusión educativa) registró Δ = +1,59, pero estas cifras solo describen el cambio en cada bloque y nada dicen sobre el efecto de la supervisión. La buena calificación que recibió ChatGPT al final (Bloque C: *M* = 4,17; utilidad pedagógica con 76,7% de logro) muestra que la herramienta fue bien aceptada y pudo usarse sin dificultad por alumnos con internet y una conexión elemental. De ello no se desprende que la IA generativa reduzca la brecha digital: disponer de un dispositivo conectado era

https://creativecommons.org/licenses/by-nc-sa/4.0/

requisito para participar, por lo que la herramienta se probó solo con quienes ya tenían acceso, y no se estudió qué sucede con quienes carecen de conexión. De igual forma, el aumento en la percepción de equidad e inclusión (ítem 11) expresa la opinión de estos alumnos y no constituye una medida de la equidad en el acceso. Para saber si estos efectos se repiten en otros contextos, los próximos estudios tendrían que abarcar más colegios de Manabí y del resto del Ecuador, aplicar una prueba de álgebra y registrar la supervisión del profesor.

Limitaciones

El cuestionario tuvo una consistencia interna baja. El Alfa de Cronbach se situó entre $\alpha = -,32$ y $\alpha = ,13$ según grupo y momento, y solo la medición final de la muestra completa rebasó el umbral nominal de ,70 ($\alpha = ,78$); ese valor, sin embargo, no revela una escala fiable, porque obedece a la diferencia entre los grupos y no a que los ítems varíen juntos dentro de cada uno. Aunque los expertos avalaron la validez de contenido (promedio 3,87/4,00; CVR = 1,00), la débil relación entre los ítems impide considerar su suma como una escala unidimensional confiable. Ni el omega de McDonald ni el análisis factorial exploratorio sostienen una estructura de un solo factor, y con 60 alumnos no es posible un análisis factorial confirmatorio. Por esa razón, los contrastes de hipótesis se hicieron por bloque y por ítem, y no con la puntuación total; tienen carácter exploratorio, y las conclusiones se limitan a una tendencia preliminar, no a un efecto confirmado.

Un diseño cuasiexperimental con cursos ya formados y sin reparto aleatorio no permite descartar amenazas a la validez interna. Los grupos no difirieron significativamente al inicio (d entre −0,17 y 0,39 por bloque, con una pequeña ventaja del experimental en tres bloques), pero esa falta de significación no demuestra equivalencia y, además, lo que se comparó fue la competencia percibida y no el dominio real del álgebra, que no se midió. La medición final se ajustó por la inicial con ANCOVA, si bien la correlación entre ambas fue débil ($r = ,09$). No todos los bloques cumplieron los supuestos de la prueba t (Levene en el Bloque C; Shapiro-Wilk en tres bloques del grupo experimental; Breusch-Pagan en el ANCOVA del Bloque C), por lo que los resultados se verificaron con las pruebas de Welch y de Mann-Whitney. El profesor, que dictó clases a un grupo y acompañó al otro, sabía a qué condición pertenecía cada uno, y no consta que los alumnos desconocieran la suya, ya que aceptaron participar en un estudio; en consecuencia, no se pueden descartar la tendencia a responder lo socialmente esperado ni el efecto Hawthorne. Tampoco es posible, con estos datos, atribuir a una causa concreta la ganancia del grupo control ($\Delta = +0,43$).

https://creativecommons.org/licenses/by-nc-sa/4.0/

Existen otras amenazas a la validez interna. Se trabajó con cursos completos y no equivalentes; un contraste no significativo no bastó para demostrar que partían del mismo nivel, y el grupo experimental empezó con una leve ventaja. Además, la novedad de la herramienta pudo motivar más a ese grupo y elevar sus respuestas; asignar al control la misma carga horaria atenúa este efecto, pero no lo suprime. Que el mismo profesor de matemática diera las clases al control y acompañara las sesiones del experimental reduce la variación debida al docente, aunque impide distinguir el efecto de la herramienta del de la presencia del profesor durante el trabajo con IA. A lo largo de cuatro semanas pudo haber cambios por maduración o por el propio avance del currículo, que el grupo control solo neutraliza en parte, y sucesos ajenos a la intervención (evaluaciones, actividades del colegio) pudieron afectar a cada grupo de manera diferente. Como se aplicó dos veces el mismo cuestionario, los resultados pueden estar influidos por la práctica, el recuerdo de las respuestas y la tendencia a contestar lo socialmente aceptable. Por último, ChatGPT (modelo GPT-5 de OpenAI, versión de agosto de 2025) recibe actualizaciones constantes, de modo que su comportamiento puede variar y nada garantiza que ofrezca las mismas respuestas en otro momento, lo cual dificulta repetir la intervención en idénticas condiciones.

La selección de la muestra fue intencional y no probabilística, y uno de los requisitos fue contar con un dispositivo con internet. Esta condición no encaja del todo con la preocupación inicial por la brecha digital y reduce la validez externa: los hallazgos no son extensibles a alumnos sin conexión ni a otros entornos socioeconómicos, y tampoco autorizan a afirmar que la IA generativa favorezca a los estudiantes cualquiera que sea su situación familiar o económica. No se registró la calidad de la conexión ni el tipo de equipo empleado. A ello se suman la breve duración (cuatro semanas), el hecho de trabajar en un solo colegio y los 30 alumnos por grupo, factores que restan potencia y estabilidad a las estimaciones.

El rendimiento se midió como competencia algebraica percibida y no como desempeño observado en una prueba de álgebra, y se usó el mismo cuestionario en ambas mediciones. Las ganancias revelan, por tanto, un cambio en la forma en que los alumnos ven su competencia, no una mejora comprobada en álgebra: sin una prueba de rendimiento no puede afirmarse que la intervención haya mejorado la ejecución real. Los ítems de opinión (7 a 9) también se respondieron al inicio, antes de conocer la herramienta, y por eso solo la medición final se considera una valoración de la

https://creativecommons.org/licenses/by-nc-sa/4.0/

experiencia. Para llegar a conclusiones confirmatorias será necesario ampliar la muestra a más instituciones, alargar la intervención, incorporar pruebas de rendimiento y depurar el cuestionario.

Conclusiones

Con la cautela que impone la baja consistencia interna del cuestionario, los datos sugieren que trabajar con la IA generativa como tutor personalizado se asoció con una mejor percepción de la competencia algebraica en el grupo experimental de la Unidad Educativa San Rafael. En el bloque de rendimiento algebraico percibido (A), el grupo experimental subió de $M = 2,59$ a $M = 4,28$ ($\Delta = +1,69$), mientras que el control pasó de $M = 2,38$ a $M = 2,99$ ($\Delta = +0,61$). Al final, la diferencia entre ambos fue de 1,29 puntos (IC del 95% [0,96; 1,62]; $t(58) = 7,81$; $p < ,001$; $d = 2,02$) y se mantuvo después de ajustar por la medición inicial; en los tres bloques restantes, las diferencias oscilaron entre 1,18 y 1,36 puntos. Se trata de una tendencia preliminar que concuerda con la hipótesis planteada, pero que no alcanza a confirmarla. Una explicación plausible es que la respuesta inmediata y ajustada de la IA generativa facilite la corrección de errores conceptuales en álgebra, aunque el diseño no permite aislar ese mecanismo (Abril Ruiz y Abril Ruiz, 2024; Serrano y Moreno-García, 2024).

En cuanto al objetivo a) (diagnóstico inicial), los 12 indicadores partieron de niveles Bajo o Moderado y ninguno llegó al umbral de $M = 3,50$ (promedio general $M = 2,54$), lo que justificó la intervención. Respecto al objetivo b), se completaron las 12 sesiones planificadas, con control de asistencia y lista de cotejo, y el logro del grupo experimental subió de 28,1% a 72,5%. En lo que toca al objetivo c), al final los alumnos calificaron favorablemente la utilidad pedagógica, la retroalimentación y la accesibilidad (medias de 4,03 a 4,30; logro de 63,3% a 76,7%), resultado que concuerda con la idea de que la tecnología puede respaldar el aprendizaje personalizado (Reyes Parra et al., 2024). Por último, en relación con el objetivo d) (metodológico), el criterio de seis expertos (3,87/4,00; CVR = 1,00) avala la validez de contenido del cuestionario, pero su consistencia interna fue baja (α entre −,32 y ,13 según grupo y momento). Su fiabilidad, por tanto, no quedó comprobada, y el instrumento necesita depurarse antes de emplearse como escala unidimensional.

Lo encontrado no constituye evidencia confirmatoria, sino una tendencia exploratoria observada en un único colegio, con 60 alumnos, cuatro semanas de trabajo y un cuestionario de percepción.

https://creativecommons.org/licenses/by-nc-sa/4.0/

Estos resultados no pueden extenderse al conjunto del bachillerato ecuatoriano ni sirven para afirmar que la IA generativa favorezca a alumnos sin conexión o de otros entornos familiares. Antes de aconsejar su adopción como estrategia de enseñanza, es necesario repetir el estudio con muestras mayores, cuestionarios depurados y pruebas de rendimiento que permitan saber si la mejora percibida se convierte en aprendizaje observable. Los trabajos recientes sobre IA en la enseñanza de la matemática destacan tanto sus posibilidades como los retos que supone incorporarla al currículo (Chérrez Ibarra y Escalona Reyes, 2024; Villota Sarmiento et al., 2025).

Referencias Bibliográficas

Abril Ruiz, R. S. y Abril Ruiz, E. A. (2024). Interpretación mediante PRISMA 2020 de la inteligencia artificial para evaluación y retroalimentación en el aula. *Mikarimin. Revista Científica Multidisciplinaria, 10*(3), 5–30. https://doi.org/10.61154/mrcm.v10i3.3643

Ahmad, S. F., Han, H., Alam, M. M., Rehmat, M. K., Irshad, M., Arraño-Muñoz, M. y Ariza-Montes, A. (2023). Impact of artificial intelligence on human loss in decision making, laziness and safety in education. *Humanities and Social Sciences Communications*, *10*, Artículo 311. https://doi.org/10.1057/s41599-023-01787-8

Ayuso-del Puerto, D. y Gutiérrez-Esteban, P. (2022). La inteligencia artificial como recurso educativo durante la formación inicial del profesorado. *RIED. Revista Iberoamericana de Educación a Distancia, 25*(2), 347–362. https://doi.org/10.5944/ried.25.2.32332

Baldor, A. (2008). *Álgebra* (2.ª ed., 1.ª reimp.). Grupo Editorial Patria.

Bravo Guerrero, F. E. y Quezada Matute, T. G. (2021). Uso de tecnologías de la información y comunicación en el bachillerato. *Revista RECUS, 6*(1), 19–27. http://revistas.utm.edu.ec/index.php/Recus

Chérrez Ibarra, R. X. y Escalona Reyes, M. (2024). La inteligencia artificial en la enseñanza de la Matemática en la carrera Administración de Empresas. *Didasc@lia: Didáctica y Educación, 15*(6, edición especial), 328–343. https://dialnet.unirioja.es/servlet/articulo?codigo=10045977

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.

Creswell, J. W. y Creswell, J. D. (2023). *Research design: Qualitative, quantitative, and mixed methods approaches* (6th ed.). SAGE Publications. https://doi.org/10.4135/9781071948194

https://creativecommons.org/licenses/by-nc-sa/4.0/

Fuentes Cabrera, C., Tapia Zurita, O. y Tapia Herrera, D. (2025). Uso de la inteligencia artificial en la evaluación formativa y su incidencia en el desempeño académico de los estudiantes. *Revista Social Fronteriza, 5*(1), 1–17. https://doi.org/10.59814/resofro.2025.5(1)611

Hake, R. R. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics, 66*(1), 64–74. https://doi.org/10.1119/1.18809

Hernández-Sampieri, R. y Mendoza, C. P. (2023). *Metodología de la investigación: Las rutas cuantitativa, cualitativa y mixta* (2.ª ed.). McGraw-Hill. https://www.bibliotecadigitaldebogota.gov.co/resources/3650023/

Holmes, W., Bialik, M. y Fadel, C. (2019). *Artificial intelligence in education: Promises and implications for teaching and learning*. Center for Curriculum Redesign.

Hwang, G.-J. y Tu, Y.-F. (2021). Roles and research trends of artificial intelligence in mathematics education: A bibliometric mapping analysis and systematic review. *Mathematics, 9*(6), Artículo 584. https://doi.org/10.3390/math9060584

Hwang, S. (2022). Examining the effects of artificial intelligence on elementary students' mathematics achievement: A meta-analysis. *Sustainability, 14*(20), Artículo 13185. https://doi.org/10.3390/su142013185

Loayza Solórzano, M. S. y Moya Martínez, M. E. (2024). Los retos de la inteligencia artificial en el proceso de enseñanza aprendizaje. *LATAM Revista Latinoamericana de Ciencias Sociales y Humanidades*, *5*(2), 1983–1996. https://doi.org/10.56712/latam.v5i2.1989

Maldonado, N. y Santoyo, F. (2024). Validez de contenido por juicio de expertos: Integración cuantitativa y cualitativa en la construcción de instrumentos de medición. *REIRE Revista d'Innovació i Recerca en Educació, 17*(2), 1–19. https://doi.org/10.1344/reire.46238

Mendoza-Zambrano, D. M., Tirado-Morueta, R. y Marín-Gutiérrez, I. (2017). Niveles de acceso a Internet de los estudiantes del bachillerato en Ecuador. *Chasqui. Revista Latinoamericana de Comunicación*, (134). https://www.redalyc.org/articulo.oa?id=16057382023

Mendoza-Zambrano, M. G., De-la-Peña-Consuegra, G. y Linzán-Saltos, M. F. (2023). Tecnologías educativas emergentes para fortalecer el proceso de enseñanza-aprendizaje en los estudiantes de tercero Bachillerato en tiempos de pandemia. *MQRInvestigar, 7*(1), 54–73. https://doi.org/10.56048/MQR20225.7.1.2023.54-73

Molemane, N., Mosia, M. y Egara, F. O. (2026). Effectiveness of AI-based tutoring and assessment systems in mathematics education: A systematic review. *Discover Education*, *5*, Artículo 973. https://doi.org/10.1007/s44217-026-01872-5

Montes Rodríguez, C. A., Lucas Ostaiza, N. B., López Zamora, M. E. y Carrión Cano, H. A. (2022). Análisis de los conocimientos previos en matemáticas de los bachilleres admitidos a la educación superior. En *Memorias de la VI Convención Científica Internacional de la*

https://creativecommons.org/licenses/by-nc-sa/4.0/

Universidad Técnica de Manabí (p. 584). Universidad Técnica de Manabí. https://www.utm.edu.ec/ediciones/images/pdf/Libro%20Memorias%20de%20la%20VI%20Convencion%20Cientifica%20Internacional%20UTM-2022.pdf

Mujica-Sequera, R. (2024). Clasificación de las herramientas de la inteligencia artificial en la educación. *Revista Tecnológica-Educativa Docentes 2.0, 17*(1), 31–40. https://doi.org/10.37843/rted.v17i1.513

Organización de las Naciones Unidas. (2024). *Educación - desarrollo sostenible*. https://www.un.org/sustainabledevelopment/es/education/

Otzen, T. y Manterola, C. (2017). Técnicas de muestreo sobre una población a estudio. *International Journal of Morphology, 35*(1), 227–232. https://doi.org/10.4067/S0717-95022017000100037

Panqueban, D. y Huincahue, J. (2024). Inteligencia artificial en educación matemática: Una revisión sistemática. *Uniciencia, 38*(1), 1–17. https://doi.org/10.15359/ru.38-1.20

Reyes Parra, D., Rozo García, H. A. y Buitrago Espitia, J. E. (2024). Aportes de la tecnología al aprendizaje personalizado: Una revisión a la literatura. *Diá-logos, 16*(28), 9–29. https://doi.org/10.61604/dl.v16i28.352

Serrano, J. L. y Moreno-García, J. (2024). Inteligencia artificial y personalización del aprendizaje: ¿Innovación educativa o promesas recicladas? *Edutec. Revista Electrónica de Tecnología Educativa*, (89), 1–17. https://doi.org/10.21556/edutec.2024.89.3577

Tene-Tenempaguay, T., Martínez-Abad, F. y Hernández-Ramos, J. P. (2025). Factores asociados a la competencia matemática: Un estudio a partir de PISA-D 2018. *REICE. Revista Iberoamericana sobre Calidad, Eficacia y Cambio en Educación, 23*(2). https://doi.org/10.15366/reice2025.23.2.001

Trocado, A., dos Santos, J. M., Saimon, M. y Lavicza, Z. (2026). Learning quadratic functions with ChatGPT: An innovative experience in high school mathematics education. *International Journal of Education in Mathematics, Science and Technology, 14*(1), 296–314. https://eric.ed.gov/?id=EJ1494390

UNESCO. (2025). *La UNESCO destaca el papel de la inteligencia artificial en la educación en el Congreso Futuro 2025*. https://www.unesco.org/es/articles/la-unesco-destaca-el-papel-de-la-ia-en-la-educacion

Vélez-Basurto, G. Y. y Vélez-Loor, J. M. (2024). Estrategias pedagógicas para la atención a estudiantes con necesidades educativas especiales asociadas a problemas de aprendizaje en la asignatura matemática en el primer nivel de bachillerato, Portoviejo. *REICOMUNICAR, 7*(13 Ed. esp.), 11–38. https://reicomunicar.journalgestar.org/index.php/reicomunicar/article/view/259

https://creativecommons.org/licenses/by-nc-sa/4.0/

Vilchez Ruiz, M. I. (2024). Contenido educativo con inteligencia artificial: ¿Restringir o enseñar a personalizar éticamente en el ámbito educativo? *Revista Científica Ciencia y Tecnología, 24*(44). https://doi.org/10.47189/rcct.v24i44.689

Villegas-Ch, W., Buenano-Fernández, D., Maldonado Navarro, A. y Mera-Navarrete, A. (2025). Adaptive intelligent tutoring systems for STEM education: Analysis of the learning impact and effectiveness of personalized feedback. *Smart Learning Environments, 12*(1), Artículo 41. https://doi.org/10.1186/s40561-025-00389-y

Villota Sarmiento, P. O., Pérez Larco, P. R., Sigcho Segovia, M. C., Jácome Guano, L. G. y Monge Tipan, O. V. (2025). Inserciones curriculares de IA en Matemáticas: Oportunidades y desafíos en el aula a nivel de Bachillerato. *Prospherus, 2*(3), 601–619. https://doi.org/10.63535/243kan76

Xing, W., Song, Y., Li, C., Liu, Z., Zhu, W. y Oh, H. (2025). Development of a generative AI-powered teachable agent for middle school mathematics learning: A design-based research study. *British Journal of Educational Technology, 56*(5), 2043–2077. https://doi.org/10.1111/bjet.13586

Yudrowsky, E. (2023). La inteligencia artificial generativa y la evaluación: ¿Qué pasará con los exámenes? *Investigación Educación Médica, 12*(48), 1–8. https://doi.org/10.22201/fm.20075057e.2023.48.23550

Anexos

Anexo A

Estructura del cuestionario de competencia percibida

N°	Variable / Indicador	Bloque	Escala de respuesta
1	Comprensión procedimental algebraica	A. Rendimiento algebraico	Likert de 5 puntos
2	Generalización simbólica	A. Rendimiento algebraico	Likert de 5 puntos
3	Autorregulación algebraica	A. Rendimiento algebraico	Likert de 5 puntos
4	Gestión autónoma del tiempo	B. Aprendizaje autónomo	Likert de 5 puntos
5	Iniciativa de aprendizaje autónomo	B. Aprendizaje autónomo	Likert de 5 puntos
6	Autoeficacia digital en álgebra	B. Aprendizaje autónomo	Likert de 5 puntos
7	Utilidad pedagógica de la IA generativa	C. Uso y percepción IA	Likert de 5 puntos
8	Efectividad retroalimentación tutor IA	C. Uso y percepción IA	Likert de 5 puntos
9	Accesibilidad del tutor IA	C. Uso y percepción IA	Likert de 5 puntos
10	Desarrollo de competencias digitales	D. Competencias digitales	Likert de 5 puntos
11	Percepción de equidad e inclusión	D. Competencias digitales	Likert de 5 puntos
12	Motivación con uso responsable de la IA	D. Competencias digitales	Likert de 5 puntos

Nota. Todos los ítems se responden en una escala Likert de 5 puntos (1 = totalmente en desacuerdo; 5 = totalmente de acuerdo). Los enunciados completos de los 12 ítems están disponibles bajo solicitud razonable al autor de correspondencia.

https://creativecommons.org/licenses/by-nc-sa/4.0/

Anexo B

Validación por juicio de expertos: resultados globales

Indicador	Resultado
Número de expertos	6 (estadística educativa, didáctica de la matemática, tecnología educativa e inclusión)
Promedio general de la valoración	3,87 sobre 4,00
Razón de validez de contenido (CVR)	1,00 en los 12 ítems
Consistencia interna (α de Cronbach, 12 ítems)	−,32 a ,13 por grupo y momento; −,06 (pretest) y ,78 (postest) en la muestra total

Nota. CVR = razón de validez de contenido. La matriz de evaluación por ítem y por experto está disponible bajo solicitud razonable al autor de correspondencia.

Conflicto de intereses:

Los autores declaran que no existe conflicto de interés posible.

Financiamiento:

No existió asistencia financiera de partes externas al presente artículo.

Agradecimiento:

N/A

Nota:

El artículo no es producto de una publicación anterior.

https://creativecommons.org/licenses/by-nc-sa/4.0/