

Retroalimentación formativa con inteligencia artificial generativa y su efecto en el razonamiento matemático universitario

Formative feedback with generative artificial intelligence and its effect on university mathematical reasoning

Juan Jacinto Macias Hinojoza ^{1*}, Carlos Manuel Núñez Michuy ², Angel Patricio Verdezoto-Michuy ³ y Adriana Raquel Velasco Borja ⁴

¹ Universidad Técnica Estatal de Quevedo, Ecuador; <https://orcid.org/0009-0005-2964-2466>

² Universidad Técnica Estatal de Quevedo, Ecuador; <https://orcid.org/0000-0003-2298-7697> ; cnunezm@uteq.edu.ec

³ Universidad Casa Grande, Ecuador; <https://orcid.org/0009-0002-3237-6144> ; angel.patricio.verdezoto@casagrande.edu.ec

⁴ Unidad Educativa Quintiliano Sánchez Rendón, Ecuador; <https://orcid.org/0009-0002-5574-7868>

Cita: Macias Hinojoza, J. J., Núñez Michuy, C. M., Verdezoto-Michuy, A. P., & Velasco Borja, A. R. (2026). Retroalimentación formativa con inteligencia artificial generativa y su efecto en el razonamiento matemático universitario. *Horizon Nexus Journal*, 4(3), 263-274. <https://doi.org/10.70881/hnj/v4/n3/158>

* Correspondencia: jmaciash6@uteq.edu.ec

 <https://doi.org/10.70881/hnj/v4/n3/158>

Recibido: 08/07/2026
Revisado: 12/08/2026
Aceptado: 17/08/2026
Publicado: 01/09/2026



Copyright: © 2026 por los autores. Este artículo es un artículo de acceso abierto distribuido bajo los términos y condiciones de la Licencia Creative Commons, Atribución-NoComercial 4.0 Internacional. (CC BY-NC).

(<https://creativecommons.org/licenses/by-nc/4.0/>)

Resumen: El presente estudio evalúa el impacto de la retroalimentación formativa automatizada mediante inteligencia artificial generativa sobre el pensamiento matemático avanzado y la ansiedad hacia las matemáticas en la educación superior. Mediante un diseño cuantitativo cuasi-experimental con mediciones pretest y posttest, se analizó una muestra de 120 estudiantes de segundo año de ingeniería divididos equitativamente en un grupo experimental ($n = 60$), que recibió andamiaje socrático adaptativo basado en el método de Polya y la Escalera de Wilson administrado por un asistente configurado con GPT-4o, y un grupo de control ($n = 60$), expuesto a retroalimentación docente analógica convencional. Los resultados del post-test revelaron diferencias altamente significativas a favor del grupo experimental ($M = 72.96$; $DE = 10.41$) frente al de control ($M = 63.30$; $DE = 10.14$; $t = 5.147$; $p < 0.001$), demostrando un tamaño del efecto de gran magnitud ($d = 0.940$). Asimismo, se observó una disminución significativa en la ansiedad matemática y un incremento de la autoeficacia, aunque persisten reticencias debidas a desajustes cualitativos del software. Se concluye que el andamiaje algorítmico guiado optimiza la carga cognitiva y promueve el aprendizaje independiente, siempre que se restrinja la entrega directa de respuestas completas para evitar la dependencia de los estudiantes.

Palabras clave: retroalimentación formativa; inteligencia artificial generativa; razonamiento matemático; educación superior; resolución de problemas.

Abstract: This study evaluates the impact of automated formative feedback using generative artificial intelligence on advanced mathematical thinking and mathematics anxiety in higher education. Using a quantitative quasi-experimental design with pretest and posttest measurements, a sample of 120 second-year engineering students was analyzed, divided equally into an experimental group ($n = 60$), which received adaptive socratic scaffolding based on Polya's method and Wilson's Feedback Ladder managed by a custom GPT-4o assistant, and a control group ($n = 60$), exposed to conventional analog teacher feedback. The post-test results revealed highly significant differences favoring the experimental group ($M = 72.96$; $SD = 10.41$) compared to the control group ($M = 63.30$; $SD = 10.14$; $t = 5.147$; $p < 0.001$), demonstrating a large effect size ($d = 0.940$). Likewise, a significant decrease in mathematics anxiety and an increase in self-efficacy were observed, although discrepancies remain due to qualitative system errors. It is concluded that guided algorithmic scaffolding optimizes cognitive load and promotes independent learning, provided that the direct delivery of complete answers is restricted to prevent student cognitive dependence.

Keywords: formative feedback; generative artificial intelligence; mathematical reasoning; higher education; problem solving.

1. Introducción

La incorporación de la inteligencia artificial generativa en la educación superior ha modificado las dinámicas de acceso al conocimiento y ha abierto nuevas posibilidades para la retroalimentación, la resolución de problemas y el acompañamiento del aprendizaje. En el ámbito de las matemáticas, la evidencia reciente muestra que estos sistemas pueden apoyar la comprensión y el razonamiento cuando se integran dentro de actividades pedagógicas estructuradas y con objetivos explícitos (Leiva-Guerrero et al., 2026; Liu et al., 2025; Rivas-Díaz et al., 2024). Sin embargo, el beneficio no depende únicamente de disponer de un modelo generativo, sino de la forma en que se organiza la interacción, se verifican sus respuestas y se preserva la participación intelectual del estudiante (Rigaud Téllez et al., 2024; Urhan et al., 2026). En consecuencia, el problema central consiste en determinar bajo qué condiciones la retroalimentación automatizada contribuye al razonamiento matemático sin sustituir el esfuerzo cognitivo necesario para aprender.

En el plano sociotécnico, la literatura identifica diferencias en la manera en que estudiantes y docentes perciben y utilizan la inteligencia artificial generativa. Hossein-Mohand et al. (2026) señalan que la aceptación estudiantil se relaciona con la utilidad percibida, la confianza y la experiencia de uso, mientras que Hennessey et al. (2026) destacan que la retroalimentación generada por inteligencia artificial se valora mejor cuando mantiene coherencia pedagógica y no desplaza el juicio docente. Estas evidencias sugieren que la adopción educativa no depende solamente del acceso a la herramienta, sino también de su fiabilidad, de la transparencia de sus resultados y de su articulación con la mediación humana.

En la enseñanza universitaria de las matemáticas, uno de los desafíos consiste en avanzar desde la ejecución mecánica de procedimientos hacia formas de razonamiento

que exigen argumentación, modelación y justificación conceptual. Investigaciones recientes han señalado que la inteligencia artificial puede apoyar la resolución de problemas y la comprensión cuando el estudiante mantiene un papel activo y verifica las respuestas generadas (Borja et al., 2025; Rivas-Díaz et al., 2024; Urhan et al., 2026). En cambio, la aceptación acrítica de soluciones automáticas puede favorecer errores conceptuales y dependencia del sistema, especialmente cuando el estudiante no dispone de conocimientos suficientes para evaluar la calidad matemática de la respuesta (Urhan et al., 2026).

Para reducir estos riesgos, la retroalimentación formativa debe orientar el proceso de aprendizaje y no limitarse a comunicar una respuesta final. Black y Wiliam (2018) sostienen que la retroalimentación resulta más útil cuando proporciona información que permite reconocer la distancia entre el desempeño actual y la meta de aprendizaje. En entornos con inteligencia artificial generativa, esta función puede operacionalizarse mediante preguntas, pistas graduadas y solicitudes de justificación. Rigaud Téllez et al. (2024) y Leiva-Guerrero et al. (2026) muestran que la retroalimentación automatizada puede ser pedagógicamente pertinente cuando se estructura con criterios explícitos y se somete a revisión. En este estudio se integraron las fases de resolución de Polya con la Escalera de Retroalimentación de Wilson para configurar un asistente socrático orientado a clarificar el problema, valorar el procedimiento, identificar dificultades y sugerir mejoras sin revelar directamente la solución.

El propósito principal de esta investigación consiste en evaluar la efectividad de un sistema de retroalimentación formativa automatizada, diseñado bajo las directrices del método de Polya y la Escalera de Wilson mediante inteligencia artificial generativa, sobre el razonamiento matemático de estudiantes universitarios de ingeniería, analizando en paralelo su impacto sobre la ansiedad matemática y la autoeficacia percibida. Se plantea como hipótesis de trabajo que los estudiantes expuestos a este andamiaje adaptativo estructurado demuestran un incremento estadísticamente significativo en su capacidad de razonamiento matemático independiente y una reducción más profunda de su nivel de ansiedad matemática, en comparación con aquellos alumnos que reciben retroalimentación tradicional manuscrita y colectiva.

2. Materiales y Métodos

2.1. Diseño de investigación

Se empleó un enfoque cuantitativo con diseño cuasi-experimental de grupos no equivalentes y mediciones pretest y posttest. La intervención se desarrolló durante un semestre académico regular de 16 semanas en una universidad pública española. El diseño permitió comparar el cambio del razonamiento matemático entre un grupo experimental y un grupo de control, así como examinar variaciones en variables actitudinales asociadas al uso de inteligencia artificial generativa.

2.2. Participantes, muestra y criterios de elegibilidad

La población estuvo conformada por estudiantes matriculados en las asignaturas obligatorias de Álgebra Lineal y Cálculo de segundo año de la Facultad de Ingeniería. Mediante un muestreo no probabilístico por conveniencia, condicionado por la distribución institucional previa de los cursos presenciales, se conformó una muestra final de 120 participantes. La muestra se dividió en dos cohortes independientes: Grupo

Experimental ($n = 60$), que utilizó la plataforma inteligente de retroalimentación, y Grupo de Control ($n = 60$), que desarrolló las actividades bajo un esquema de retroalimentación docente tradicional.

Los criterios de inclusión fueron: estar matriculado regularmente en la asignatura por primera vez, disponer de un computador portátil o tableta con conexión estable a internet y registrar una asistencia presencial superior al 85 % de las sesiones teóricas. Como criterios de exclusión se consideraron la convalidación previa de créditos de matemáticas superiores y el abandono voluntario de la plataforma antes de completar el primer tercio del periodo lectivo.

2.3. Intervención experimental y condición de control

El Grupo Experimental utilizó la aplicación interactiva MathFormativeAI como andamiaje cognitivo durante las tareas semanales de resolución de problemas de cálculo y álgebra lineal fuera del aula. El motor de la plataforma empleó la API del modelo GPT-4o, configurada con una temperatura de 0.1 para reducir la variabilidad de las respuestas y mantener la retroalimentación dentro de las rúbricas definidas por el equipo docente. El diseño de la interacción se apoyó en principios de retroalimentación formativa y resolución guiada descritos en investigaciones previas sobre inteligencia artificial aplicada al aprendizaje universitario (Leiva-Guerrero et al., 2026; Rigaud Téllez et al., 2024).

El prompt de sistema fue sometido a validación de contenido mediante dos rondas Delphi con un panel interdisciplinario de ocho expertos en didáctica de la matemática y tecnología educativa. Se utilizó una escala ordinal de 0 a 3 y el coeficiente W de Kendall para examinar el nivel de acuerdo entre evaluadores y refinar las instrucciones. La lógica de retroalimentación se organizó en cuatro acciones: aclarar, valorar, expresar inquietudes y sugerir mejoras, articuladas con las fases de resolución de Polya. El prompt prohibía revelar el resultado final de los problemas y priorizaba preguntas, pistas y solicitudes de justificación para sostener una interacción dialógica con el estudiante.

El Grupo de Control resolvió el mismo conjunto de problemas semanales de álgebra y cálculo. La retroalimentación fue proporcionada por el docente mediante correcciones escritas sobre las entregas manuscritas, con un desfase promedio de siete días, y se complementó con discusiones en la pizarra y actividades de corrección entre compañeros. De esta forma, ambos grupos trabajaron contenidos equivalentes, diferenciándose principalmente en el tipo y la inmediatez de la retroalimentación recibida.

2.4. Instrumentos de recolección de datos

Para evaluar el efecto de la intervención se aplicaron dos instrumentos en las fases de pretest y posttest: una Prueba de Razonamiento Matemático y la Escala de Actitudes hacia la Inteligencia Artificial en el Aprendizaje de las Matemáticas (IAMAT).

La Prueba de Razonamiento Matemático estuvo compuesta por diez problemas de alta demanda cognitiva orientados a evaluar modelación conceptual, análisis de demostraciones formales y deducción lógica en geometría analítica, cálculo multivariable y álgebra lineal. Las respuestas abiertas se analizaron mediante una adaptación del esquema argumentativo de Toulmin, considerando datos, aserción,

garantía y respaldo, procedimiento que ha sido utilizado para examinar el razonamiento matemático en estudios recientes (Urhan et al., 2026). La calificación se realizó en una escala analítica de 0 a 100 puntos y las respuestas fueron revisadas de forma ciega por dos evaluadores externos mediante rúbricas estandarizadas.

La escala IAMAT se adaptó a partir del instrumento descrito por Hossein-Mohand et al. (2026). El cuestionario estuvo conformado por once reactivos valorados en una escala Likert de cinco puntos y organizados en dos dimensiones: utilidad y confianza en el andamiaje algorítmico, e incertidumbre ante posibles errores o imprecisiones del sistema. En la presente muestra, la consistencia interna se evaluó mediante el coeficiente alfa de Cronbach y la adecuación muestral mediante el índice Kaiser-Meyer-Olkin antes de interpretar las puntuaciones.

2.5. Procedimiento de análisis de datos

Los datos se procesaron mediante estadística descriptiva e inferencial. Para comparar el rendimiento entre grupos se aplicó la prueba t de Student para muestras independientes en el pretest y posttest, mientras que el cambio dentro de cada grupo se examinó con la prueba t para muestras emparejadas. El tamaño del efecto de la diferencia posttest se estimó mediante la d de Cohen. Los registros de interacción con MathFormativeAI se codificaron de forma manual para identificar patrones de error y discrepancias en la retroalimentación. Los análisis cuantitativos se realizaron con el paquete estadístico SciPy.

2.6. Consideraciones éticas

Todos los participantes otorgaron consentimiento informado por escrito antes de su inclusión. Se garantizó la confidencialidad, el anonimato y el uso académico de la información. El tratamiento de los datos se efectuó de acuerdo con la aprobación del Comité de Ética de la institución y sin incorporar datos personales identificables en las transcripciones utilizadas para el análisis.

3. Resultados

Los análisis inferenciales y descriptivos revelaron un impacto estadísticamente significativo del sistema de retroalimentación interactivo sobre las variables bajo estudio al término del semestre académico.

3.1. Análisis cuantitativo y rendimiento académico

La aplicación preliminar de la Prueba de Razonamiento Matemático mostró que ambos grupos iniciaron el estudio con niveles comparables. La media del Grupo Experimental fue de 53.89 (DE = 7.72), mientras que la del Grupo de Control fue de 54.77 (DE = 7.64). La prueba t de Student para muestras independientes confirmó que la diferencia inicial no fue estadísticamente significativa ($t = -0.631$; $p = 0.5294$).

Al concluir la intervención se observó un progreso en ambos grupos, aunque la ganancia fue mayor en el Grupo Experimental. Los estudiantes que utilizaron MathFormativeAI alcanzaron una media posttest de 72.96 (DE = 10.41), frente a 63.30 (DE = 10.14) en el Grupo de Control. La comparación intergrupala mostró una diferencia estadísticamente significativa a favor del grupo experimental ($t = 5.147$; $p < 0.001$). El tamaño del efecto fue $d = 0.940$, correspondiente a una diferencia de gran magnitud.

La comparación intragrupal entre pretest y posttest también mostró avances estadísticamente significativos en el Grupo Experimental ($t = -23.918$; $p < 0.001$) y en el Grupo de Control ($t = -10.992$; $p < 0.001$). No obstante, el incremento observado en el grupo experimental fue mayor, lo que respalda la utilidad del andamiaje basado en inteligencia artificial generativa para acompañar procesos de razonamiento matemático complejo.

La Tabla 1 resume los estadísticos descriptivos e inferenciales utilizados para comparar el rendimiento en razonamiento matemático antes y después de la intervención.

Tabla 1. Comparación del rendimiento en razonamiento matemático entre el grupo experimental y el grupo de control

Variable y Momento de Medición	Grupo Experimental (n = 60) Media (DE)	Grupo de Control (n = 60) Media (DE)	Estadístico de Contraste	Valor de Significación (p)	Tamaño del Efecto (d de Cohen)
Razonamiento Matemático Pre-test	53.89 (7.72)	54.77 (7.64)	$t(118) = -0.631$	0.5294	0.114 (Mínimo)
Razonamiento Matemático Post-test	72.96 (10.41)	63.30 (10.14)	$t(118) = 5.147$	< 0.0001	0.940 (Grande)
Prueba t Emparejada (Pre vs Post)	$t(59) = -23.918$	$t(59) = -10.992$	—	< 0.0001	—

Como se observa en la Tabla 1, los grupos presentaron niveles comparables en el pretest, mientras que en el posttest el grupo experimental obtuvo una media superior. El tamaño del efecto posttest ($d = 0.940$) indica una diferencia de gran magnitud a favor de la intervención.

Nota. Las puntuaciones se calcularon sobre un máximo de 100 puntos. DE = desviación estándar; d = d de Cohen.

3.2. Clasificación de discrepancias del asistente de inteligencia artificial

A pesar del efecto positivo sobre el rendimiento, el monitoreo de las interacciones reveló desajustes y limitaciones conceptuales en algunas respuestas del asistente. Este tipo de inconsistencias ha sido señalado como un aspecto que debe controlarse cuando los modelos generativos se utilizan para producir retroalimentación educativa (Leiva-Guerrero et al., 2026; Urhan et al., 2026). En el presente estudio, las incidencias afectaron principalmente la coherencia entre el comentario cualitativo y la valoración asignada, así como la precisión de ciertas orientaciones.

Las transcripciones fueron codificadas manualmente y los desajustes se agruparon en cuatro categorías. La Tabla 2 presenta su frecuencia absoluta, el porcentaje de ocurrencia y una descripción de su posible impacto pedagógico.

Tabla 2. Frecuencia y características de los desajustes detectados en la retroalimentación generada por inteligencia artificial

Tipología de desajuste	Frecuencia	Ocurrencia (%)	Descripción e impacto pedagógico
Incongruencia entre el feedback cualitativo y la nota	127	15.0	El modelo podía validar verbalmente un procedimiento incorrecto o asignar una puntuación que no coincidía con las observaciones formuladas.
Vaguedad conceptual y ambigüedad en la guía	47	5.6	Se generaron comentarios generales poco vinculados con los pasos específicos del procedimiento del estudiante, reduciendo su utilidad orientativa.
Exigencia de elementos ya incluidos	31	3.7	El asistente señalaba como ausentes demostraciones o justificaciones que sí estaban presentes en el desarrollo del estudiante.
Omisión de fallos procedimentales	4	0.7	El modelo pasó por alto errores menores en pasos algebraicos intermedios, con riesgo de generar una falsa sensación de corrección.

Como se observa en la Tabla 2, la incongruencia entre el comentario cualitativo y la puntuación fue el desajuste más frecuente (15.0 %), seguida de la vaguedad conceptual (5.6 %). En conjunto, estas incidencias muestran que la salida del modelo requería verificación antes de utilizarse como criterio de evaluación o como orientación definitiva para el estudiante.

3.3. Actitudes y perfiles de confianza tecnológica

El análisis de la escala IAMAT mostró una valoración favorable de la utilidad del asistente para recibir orientación inmediata, acompañada de incertidumbre cuando la plataforma emitía retroalimentación ambigua o incongruente. Esta combinación es coherente con estudios sobre actitudes hacia la inteligencia artificial generativa en matemáticas, en los que la utilidad percibida puede coexistir con reservas respecto a la confiabilidad y a la automatización de la evaluación (Hosseini-Mohand et al., 2026). Además, la reducción observada de la ansiedad matemática en el grupo experimental coincide con evidencia que relaciona el uso guiado de ChatGPT con menores niveles de ansiedad y una mejor percepción de la capacidad para resolver problemas (Wahbeh et al., 2026).

La Tabla 3 sintetiza las principales tendencias actitudinales observadas tras la intervención y su interpretación pedagógica.

Tabla 3. Síntesis de las actitudes estudiantiles frente al uso de MathFormativeAI

Dimensión	Tendencia observada	Interpretación pedagógica
Utilidad y confianza	Valoración favorable del acceso inmediato a pistas y orientaciones durante la resolución.	La disponibilidad del asistente favorece la continuidad del trabajo autónomo cuando la respuesta es clara y pertinente.
Ansiedad matemática	Reducción observada del bloqueo y de la tensión durante la resolución de problemas.	El acompañamiento paso a paso puede disminuir la fricción asociada a pedir ayuda y fortalecer la percepción de autoeficacia.
Incertidumbre ante errores	La confianza disminuyó cuando aparecieron respuestas ambiguas o inconsistentes.	La retroalimentación generada por IA requiere mecanismos de verificación y criterios de calidad explícitos.
Mediación docente	Se mantuvo la necesidad de validar procedimientos y resultados de mayor complejidad.	La herramienta funciona mejor como apoyo formativo que como evaluador autónomo del razonamiento matemático.

Como se resume en la Tabla 3, la aceptación del asistente no fue uniforme: la utilidad y la disponibilidad inmediata favorecieron su uso, mientras que los errores cualitativos redujeron la confianza y reforzaron la necesidad de validación docente.

En conjunto, los resultados de las Tablas 1, 2 y 3 muestran que la retroalimentación adaptativa estructurada con inteligencia artificial generativa se asoció con un mejor rendimiento en razonamiento matemático y con una percepción favorable de apoyo al aprendizaje. Al mismo tiempo, las inconsistencias detectadas en el feedback y la necesidad de mediación humana evidencian que la herramienta debe utilizarse como apoyo formativo y no como sustituto del juicio docente.

4. Discusión

4.1. Efecto sobre el razonamiento matemático

El principal hallazgo fue la diferencia del postest a favor del Grupo Experimental, con un tamaño del efecto de gran magnitud ($d = 0.940$). Este resultado sugiere que un sistema de inteligencia artificial generativa puede favorecer el razonamiento matemático cuando la interacción obliga al estudiante a explicar, planificar y revisar su procedimiento. La dirección del efecto es coherente con metaanálisis recientes que reportan mejoras en los resultados de aprendizaje cuando la inteligencia artificial generativa se integra con objetivos pedagógicos definidos y actividades de participación activa (Liu et al., 2025; Ma & Zhong, 2025). No obstante, la magnitud observada debe interpretarse dentro del diseño cuasi-experimental y del contexto específico de la intervención, por lo que no se asume como un efecto universal para todas las asignaturas, poblaciones o plataformas.

4.2. Andamiaje formativo y preservación del esfuerzo cognitivo

La intervención fue diseñada para impedir que el asistente entregara soluciones completas y para mantener al estudiante dentro de una secuencia de resolución guiada. Este principio coincide con investigaciones que señalan que los modelos generativos pueden apoyar el razonamiento matemático, pero no sustituyen la necesidad de

comprobar pasos, justificar decisiones y evaluar la coherencia de la respuesta (Urhan et al., 2026). Asimismo, la generación automatizada de retroalimentación en álgebra lineal muestra mayor valor pedagógico cuando se organiza mediante criterios claros y se utiliza como recurso formativo, no como respuesta final (Rigaud Téllez et al., 2024). En este sentido, la combinación de preguntas socráticas, pistas graduadas y restricciones para revelar resultados parece haber contribuido a mantener el esfuerzo cognitivo durante la resolución.

4.3. Ansiedad matemática y autoeficacia

Un segundo hallazgo relevante fue la mejora del componente socioafectivo. La disponibilidad de un canal de consulta permanente permitió que los estudiantes solicitaran orientación sin esperar a la siguiente sesión presencial y sin la presión social asociada a exponer públicamente una dificultad. Este resultado es consistente con estudios que relacionan el uso de asistentes conversacionales con una reducción de la ansiedad matemática y una mayor percepción de capacidad para resolver problemas cuando la herramienta se emplea como apoyo complementario (Sosas et al., 2026; Wahbeh et al., 2026). Sin embargo, la disminución de la ansiedad no debe interpretarse como evidencia suficiente de aprendizaje: el acompañamiento debe seguir exigiendo explicación, verificación y toma de decisiones por parte del estudiante.

4.4. Fiabilidad de la retroalimentación y supervisión humana

Los beneficios observados coexistieron con inconsistencias en la retroalimentación generada. La Tabla 2 mostró que la incongruencia entre comentarios y puntuaciones constituyó el desajuste más frecuente, seguida de orientaciones vagas y de la identificación incorrecta de elementos supuestamente ausentes. Estos hallazgos coinciden con la literatura que advierte que la calidad de la retroalimentación de inteligencia artificial depende de la claridad de las instrucciones, del tipo de tarea y de la capacidad del usuario para verificar la respuesta (Hennessey et al., 2026; Leiva-Guerrero et al., 2026). Por ello, el sistema no debe operar como un evaluador autónomo en tareas de razonamiento complejo; la supervisión docente continúa siendo necesaria para validar criterios, resolver contradicciones y corregir errores conceptuales.

4.5. Diferencias individuales en la apropiación de la inteligencia artificial

La utilidad de la herramienta también depende de la capacidad del estudiante para interpretar críticamente la retroalimentación. Urhan et al. (2026) muestran que el uso de ChatGPT durante tareas de razonamiento matemático exige evaluar la coherencia de los argumentos y no aceptar automáticamente las respuestas generadas. En el presente estudio, las inconsistencias detectadas sugieren que los estudiantes con menor dominio conceptual pueden requerir apoyos adicionales para identificar errores y contrastar procedimientos. Este aspecto se relaciona con las dimensiones de utilidad, confianza e incertidumbre descritas por Hossein-Mohand et al. (2026): una herramienta puede ser percibida como útil y, al mismo tiempo, generar reservas cuando su funcionamiento no es suficientemente predecible. En consecuencia, la alfabetización en inteligencia artificial debe incluir competencias de verificación, formulación de preguntas y evaluación crítica de resultados.

4.6. Implicaciones pedagógicas y limitaciones del estudio

Los resultados respaldan un modelo híbrido en el que la inteligencia artificial automatiza parte de la retroalimentación inmediata, mientras el docente conserva la responsabilidad sobre la orientación conceptual, la evaluación y la toma de decisiones pedagógicas. Esta integración requiere formación para diseñar instrucciones, verificar resultados y seleccionar tareas en las que el uso de inteligencia artificial aporte valor real al aprendizaje (Celik, 2022; Hennessey et al., 2026). También deben considerarse las limitaciones del estudio: la muestra fue obtenida por conveniencia, los grupos no se asignaron aleatoriamente, la intervención se desarrolló en una sola institución durante 16 semanas y los resultados corresponden a una configuración específica de MathFormativeAI. Además, no se realizó una medición diferida que permitiera determinar la permanencia del aprendizaje. Futuros estudios deberían incorporar muestras multicéntricas, asignación aleatoria cuando sea viable, seguimiento longitudinal y comparación entre diferentes configuraciones de modelos generativos.

5. Conclusiones

La retroalimentación formativa adaptativa estructurada mediante inteligencia artificial generativa ejerce un impacto altamente positivo y estadísticamente significativo sobre el rendimiento en razonamiento matemático de estudiantes universitarios de ingeniería, en comparación con los métodos tradicionales de corrección docente manuscrita e individualizada. Sin embargo, este beneficio cognitivo no es un resultado intrínseco de la capacidad de procesamiento lógico de las herramientas, sino que depende de manera fundamental de la rigurosidad pedagógica de su diseño heurístico interactivo, requiriendo sistemas de prompts optimizados que bloqueen de forma estricta la entrega inmediata de resultados y guíen socráticamente el pensamiento reflexivo del estudiante.

En el plano socio-afectivo, la incorporación de un asistente formativo permanente reduce sustancialmente el sentimiento de ansiedad hacia las matemáticas y optimiza la autoeficacia y confianza de los alumnos. Al constituirse en un canal de consulta privado de bajo impacto de fricción social, la inteligencia artificial generativa neutraliza el temor al juicio de los pares académicos, permitiendo al estudiante ensayar y refinar de forma continua sus hipótesis analíticas y desarrollos procedimentales.

No obstante, las recurrentes inconsistencias, errores cualitativos y alucinaciones lógicas detectadas en la plataforma automatizada, sumadas a la asimetría en la capacidad crítica de los alumnos con menor rendimiento para detectar dichos fallos conceptuales, impiden asumir a la inteligencia artificial como un evaluador autónomo y confiable del razonamiento abstracto. Se concluye que el diseño de entornos de aprendizaje de matemáticas de nivel superior debe avanzar hacia modelos educativos de carácter híbrido, en los que el andamiaje algorítmico automatizado se integre con la validación epistemológica experta del docente humano, asegurando que la tecnología opere como un mediador crítico y no como un atajo cognitivo para el estudiantado.

Contribución de los autores: Conceptualización: J.M.H.; Metodología: J.M.H.; C.N.M.; A.V.M.; A.V.B.; Análisis formal J.M.H.; C.N.M.; A.V.M.; A.V.B.; Investigación: J.M.H.; C.N.M.; A.V.M.; A.V.B.; Curación de datos: J.M.H.; C.N.M.; A.V.M.; A.V.B.; Visualización: J.M.H.; C.N.M.; A.V.M.; A.V.B.; Redacción borrador original: A.V.B.; Redacción, revisión y edición: J.M.H.; C.N.M.; A.V.M.; A.V.B.; Supervisión: J.M.H.; C.N.M.; A.V.M.; Los autores han leído y aprobado la versión publicada del manuscrito.

Financiamiento: Esta investigación no recibió financiación externa.

Conflicto de intereses: Los autores declaran no tener ningún conflicto de intereses.

Declaración de disponibilidad de datos: Los datos están disponibles previa solicitud a los autores de correspondencia: jmaciash6@uteq.edu.ec

Referencias Bibliográficas

Black, P., & William, D. (2018). Classroom assessment and pedagogy. *Assessment in Education: Principles, Policy & Practice*, 25(6), 551–575.

<https://doi.org/10.1080/0969594X.2018.1441807>

Borja, N., Chasiliquin, L., Grados, M., Toro, J., & Vera, M. (2025). El uso de inteligencia artificial para fomentar la resolución de problemas en estudiantes con baja capacidad de razonamiento lógico-matemático. *Revista Veritas de Difusión Científica*, 6(1), 1649–1670. <https://doi.org/10.61616/rvdc.v6i1.476>

Celik, I. (2022). Towards systematic design of artificial intelligence education in teacher training: An AI literacy framework. *Education and Information Technologies*, 27(1), 121–144. <https://doi.org/10.1007/s10639-021-10515-3>

Hennessey, S., Mercer, N., & Warwick, P. (2026). Teacher students' experiences of AI-generated feedback in collaborative settings. *Teacher Development*, 30(1), 89–108. <https://doi.org/10.1080/10476210.2026.2655770>

Hosseini-Mohand, H., Hosseini-Mohand, H., García-Alonso, M., & Olmos-Gómez, M. C. (2026). Generative AI in university mathematics: Attitudes and academic-leisure use patterns. *Electronic Journal of e-Learning*, 24(4), 18–31. <https://doi.org/10.34190/ejel.24.4.3110>

Leiva-Guerrero, M. V., Araya Zamorano, I., Escobar Collins, R., & Silva Castro, F. (2026). Feedback on learning with generative artificial intelligence in university students [Retroalimentación de aprendizajes con inteligencia artificial generativa en estudiantes universitarios]. *RIED. Revista Iberoamericana de Educación a Distancia*, 29(1), 241–267. <https://doi.org/10.5944/ried.45547>

Liu, B., Zhang, W., & Wang, F. (2025). Can generative artificial intelligence effectively enhance students' mathematics learning outcomes? A meta-analysis of empirical studies from 2023 to 2025. *Education Sciences*, 15(6), Article 816. <https://doi.org/10.3390/educsci15060816>

Ma, Y., & Zhong, X. (2025). A meta-analysis of the impact of generative artificial intelligence on learning outcomes. *Journal of Computer Assisted Learning*, 41(5), e70117. <https://doi.org/10.1111/jcal.70117>

Rigaud Téllez, N., Rayón Villela, P., & Blanco Bautista, R. (2024). Evaluating ChatGPT-generated linear algebra formative assessments. *International Journal of Interactive Multimedia and Artificial Intelligence*, 8(5), 75–82.
<https://doi.org/10.9781/ijimai.2024.03.003>

Rivas-Díaz, J., Cevallos-Méndez, C., & Llange-Nieves, Z. (2024). Uso de modelos de inteligencia artificial en la optimización de la enseñanza de matemáticas en la educación superior. *Reincisol*, 3(6), 4334–4350.
<https://doi.org/10.56781/reincisol.v3i6.4334>

Sosas, M. R. D., Fat, A. F. G., Dangculos, J. M. M., Anoba, H. R. P., Remiscal, C. G., & Aleguen, J. M. (2026). Use of ChatGPT as a supplemental learning tool in mathematics. *Contemporary Mathematics and Science Education*, 7(1), ep26004.
<https://doi.org/10.30935/conmaths/17989>

Urhan, S., Gençaslan, O., & Dost, Ş. (2026). Is ChatGPT a rational assistant for university students during mathematical reasoning? *Science & Education*. Advance online publication. <https://doi.org/10.1007/s11191-025-00681-y>

Wahbeh, F., Ghnaim, F., & Al-orainat, L. (2026). The impact of ChatGPT usage intensity on mathematics anxiety and perceived problem-solving skills among undergraduate students. *International Journal of Engineering Pedagogy (iJEP)*, 16(2), 45–62. <https://doi.org/10.3991/ijep.v16i2.60901>