



Diseño personalizado de problemas matemáticos para fortalecer el pensamiento lógico en Educación General Básica mediante modelos inteligentes.

Personalized design of mathematical problems to strengthen logical thinking in General Basic Education through intelligent models.

**Artículo de investigación científica
Ciencias de la Educación**



Ximena Maribel Machado Vaca ^I

<https://orcid.org/0009-0004-7277-3851>

ximena.machado@docentes.educacion.edu.ec

Ministerio de Educación – Educación Básica
Tungurahua - Ecuador



César Augusto Enriquez Álvarez ^{II}

cesar.enriqueza@docentes.educacion.edu.ec

<https://orcid.org/0009-0008-6170-6767>

Ministerio de Educación – Física
Cotopaxi - Ecuador



Ana Victoria Bastidas Freire ^{III}

ana.bastidasf@docentes.educacion.edu.ec

<https://orcid.org/0009-0000-5710-4183>

Ministerio de Educación – Química
Cotopaxi - Ecuador



Alba Guadalupe Guanotoa Cuyo ^{IV}

alba.guanotoa@educacion.gob.ec

<https://orcid.org/0009-0009-1327-1807>

Ministerio de Educación – Educación Básica
Cotopaxi - Ecuador



Evelyn Johana Quilumba Álvarez ^V

evelyn.quilumba@docentes.educacion.edu.ec

<https://orcid.org/0009-0009-8740-3948>

Ministerio de Educación – Lenguaje y Literatura
Cotopaxi - Ecuador

Fecha de envío: 2026-06-17

Fecha de revisión: 2026-07-20

Fecha da aceptación: 2026-08-21

Resumen

El estudio examinó el potencial de modelados inteligentes como estrategia didáctica para personalizar problemas matemáticos y fortalecer el pensamiento lógico en estudiantes de Educación General Básica (EGB). El objetivo fue determinar en qué medida una secuencia de aprendizaje mediada por modelos inteligentes, con supervisión docente y tareas ajustadas al nivel de desempeño, favorece el reconocimiento de patrones, la inferencia lógica, la representación de problemas, la selección de estrategias y la justificación de procedimientos. Se planteó un diseño cuasi experimental, cuantitativo y de alcance descriptivo-correlacional, con 40 estudiantes distribuidos en un grupo control y un grupo experimental, procedentes de cuatro centros educativos. La intervención comprendió doce sesiones de cuarenta y cinco minutos durante seis semanas. Se utilizó un test de base estructurada de 30 puntos, validado por juicio de expertos y con una confiabilidad ilustrativa de alfa de Cronbach de 0,90. Para el contraste se consideraron la prueba t de Student para muestras independientes, la correlación de Pearson y el tamaño del efecto d de Cohen. Como escenario estadístico simulado para este borrador, el grupo experimental alcanzó una media postest de 23,2 puntos frente a 17,4 del control; la diferencia resultó significativa, $t(38)=6,01$, $p<0,001$, con un efecto grande, $d=1,90$. La personalización se asoció positivamente con la ganancia de aprendizaje ($r=0,72$). Se concluye que la IAG puede aportar valor cuando no sustituye el razonamiento del estudiante, sino que diversifica situaciones, gradúa ayudas y promueve verificación crítica. Los resultados numéricos deben reemplazarse por los datos reales antes de cualquier envío editorial.

Palabras clave: inteligencia artificial generativa; personalización del aprendizaje; pensamiento lógico; resolución de problemas matemáticos; Educación General Básica.

Abstract

This study examined the potential of generative artificial intelligence (GenAI) as a teaching strategy to personalize mathematical problems and strengthen logical thinking among students in General Basic Education. The objective was to determine the extent to which a GenAI-mediated learning sequence, implemented under teacher supervision and with tasks adjusted to students' performance, supports pattern recognition, logical inference, problem representation, strategy selection, and justification of procedures. A quantitative quasi-experimental, descriptive-correlational design involved 40 students in control and experimental groups across four centers, with twelve 45-minute sessions over

six weeks. A 30-point structured test was considered, supported by expert content review and an illustrative Cronbach's alpha reliability coefficient of 0.90. Independent-samples Student's t test, Pearson's correlation, and Cohen's d were used as the main analytical procedures. In the simulated statistical scenario prepared for this draft, the experimental group achieved a posttest mean of 23.2 points compared with 17.4 in the control group; the difference was significant, $t(38)=6.01$, $p<.001$, with a large effect, $d=1.90$. Personalization was positively associated with learning gain ($r=.72$). The findings suggest that GenAI may provide educational value when it does not replace students' reasoning but instead diversifies problem contexts, provides graduated support, and encourages critical verification of solutions. The numerical results are illustrative and must be replaced with actual study data before any editorial submission.

Keywords: generative artificial intelligence; personalized learning; logical thinking; mathematical problem solving; General Basic Education.

Introducción

La expansión de la inteligencia artificial generativa (IAG) ha modificado la conversación educativa porque, a diferencia de herramientas digitales centradas en recuperar información o ejecutar rutinas predefinidas, los modelos generativos pueden producir explicaciones, ejemplos, preguntas, contraejemplos y variaciones de una misma tarea en tiempo real. En educación matemática, esta capacidad abre una posibilidad especialmente relevante: transformar un problema estático en una familia de problemas graduados según el nivel, los errores y el contexto del estudiante. Jauhiainen y Garagorry Guerra (2024) describen esta personalización dinámica como una de las aplicaciones pedagógicas con mayor potencial, mientras que Kasneci et al. (2023) advierten que el valor educativo de los grandes modelos de lenguaje depende de un diseño que preserve la actividad cognitiva del aprendiz y limite la delegación acrítica. La discusión, por tanto, no debería reducirse a decidir si la IAG se usa o no, sino a establecer bajo qué condiciones didácticas su intervención fortalece procesos de pensamiento que siguen perteneciendo al estudiante. La enseñanza de la matemática exige más que obtener una respuesta correcta. Resolver problemas implica reconocer regularidades, seleccionar información pertinente, construir representaciones, anticipar relaciones, justificar decisiones y revisar la coherencia del resultado. En ese sentido, el pensamiento lógico funciona como una arquitectura de relaciones que permite pasar de datos aislados a argumentos. Wardat et al. (2023) mostraron que ChatGPT puede generar explicaciones y apoyar la enseñanza de

contenidos matemáticos, aunque su uso requiere verificación; Getenet (2024) observó que la interacción con el sistema puede ampliar el repertorio de estrategias de resolución en la formación docente; y Korkmaz Guler et al. (2024) documentaron un desempeño desigual de la herramienta en tareas matemáticas, recordando que la fluidez lingüística de una respuesta no garantiza validez matemática. Esta tensión es didácticamente fértil: la IAG puede convertirse tanto en andamiaje como en objeto de crítica, dependiendo del tipo de tarea que organice el docente.

La literatura reciente también sugiere que el aporte de la IAG no se limita a explicar procedimientos. Wahba et al. (2024) reportaron mejoras en razonamiento estadístico al integrar aprendizaje basado en ChatGPT; Jia et al. (2024) compararon orientaciones para la resolución de problemas producidas por IAG y por docentes, mostrando que la calidad de las ayudas debe juzgarse por su pertinencia pedagógica; y Lee y Yeo (2022) desarrollaron un chatbot para practicar enseñanza responsiva en matemáticas, una línea que anticipa el uso de sistemas conversacionales para adaptar preguntas y retroalimentación. Estos trabajos coinciden en un punto: la tecnología resulta más valiosa cuando las respuestas generadas se insertan dentro de una secuencia de decisión, argumentación y revisión, y no cuando funcionan como un solucionario automático.

No obstante, la personalización no debe confundirse con una adaptación automática basada únicamente en dificultad. Un problema verdaderamente personalizado puede variar el contexto semántico, el grado de abstracción, el tipo de representación, la cantidad de información irrelevante, el nivel de apoyo y el modo en que se solicita justificar la respuesta. En esta línea, la revisión de Pepin et al. (2025) identifica un crecimiento rápido de investigaciones sobre ChatGPT en educación matemática, pero también señala la necesidad de comprender con mayor precisión cómo se configura su uso en tareas, prácticas y roles docentes. Almarashdi et al. (2024), mediante una revisión sistemática, destacan posibilidades para explicar conceptos, crear recursos y apoyar el aprendizaje, al tiempo que subrayan riesgos asociados con exactitud, dependencia y diseño de evaluación. El problema científico no es, entonces, si un modelo puede producir un ejercicio matemático, sino si una secuencia de ejercicios generados bajo criterios didácticos produce cambios observables en destrezas lógicas.

También es necesario considerar la alfabetización en inteligencia artificial. Walter (2024) sostiene que la alfabetización en IA, la ingeniería de instrucciones y el pensamiento crítico se vuelven componentes centrales de una educación contemporánea; Chan y Hu

(2023) muestran que los estudiantes reconocen beneficios de eficiencia y apoyo, pero también expresan preocupaciones sobre precisión y uso ético. Chiu (2024) analiza cómo la IAG está reconfigurando prácticas, políticas y agendas de investigación, mientras que Javaid et al. (2023) destacan oportunidades para enriquecer la enseñanza cuando el uso se articula con objetivos formativos. De este modo, pedir al estudiante que verifique una respuesta de la IAG no es una actividad periférica: puede convertirse en una tarea de argumentación matemática y, simultáneamente, en una práctica de alfabetización digital crítica.

La misma herramienta que puede apoyar el razonamiento también puede debilitarlo si el estudiante externaliza procesos que debería ejecutar. Zhai et al. (2024) revisaron efectos de la sobredependencia de sistemas de diálogo y encontraron preocupaciones sobre capacidades cognitivas cuando la tecnología reemplaza actividades intelectuales; Abbas et al. (2024) evidencian que el uso de IAG puede relacionarse con consecuencias no deseadas si se convierte en dependencia; Eke (2023) y Jarrah et al. (2023) sitúan la integridad académica como un problema que exige rediseñar las tareas y transparentar el uso. Por ello, la intervención propuesta en este artículo restringe la entrega directa de respuestas finales y privilegia pistas graduadas, preguntas de autoexplicación, comparación de estrategias y verificación de errores.

La Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (UNESCO) plantea una visión centrada en las personas para la IAG en educación, con énfasis en políticas, capacidades humanas, protección de datos y uso pedagógico responsable (Miao & Holmes, 2023). Esta orientación es coherente con la necesidad de conservar la agencia docente y el juicio profesional. En el contexto latinoamericano, la Comisión Económica para América Latina y el Caribe (CEPAL) ha señalado que las oportunidades de la transformación digital conviven con brechas de acceso, uso y competencias, por lo que la incorporación de tecnologías avanzadas no puede analizarse sin considerar equidad y formación digital (CEPAL, 2024). En Ecuador, la Guía metodológica de competencias matemáticas del Ministerio de Educación del Ecuador (2022) orienta el trabajo docente hacia el desarrollo de competencias y estrategias que movilizan razonamiento, resolución y aplicación en situaciones significativas. Estas referencias institucionales convergen en una idea: la innovación tecnológica solo cobra sentido educativo cuando fortalece capacidades y no amplía desigualdades. A partir de este marco, persiste la necesidad de evaluar en EGB una estrategia concreta que conecte

personalización, destrezas lógicas y medidas de efecto, en lugar de limitarse a describir percepciones sobre la herramienta.

Objetivo

Determinar el efecto de una estrategia didáctica mediada modelos inteligentes para la personalización de problemas matemáticos sobre el fortalecimiento del pensamiento lógico en estudiantes de Educación General Básica, considerando el desempeño en reconocimiento de patrones, inferencia lógica, representación de problemas, selección de estrategias y justificación-verificación, así como la relación entre los indicadores de personalización y la ganancia de aprendizaje.

Metodología

La investigación se estructuró como un estudio cuantitativo con diseño cuasi experimental y alcance descriptivo-correlacional. El diseño permitió comparar el cambio entre un grupo experimental, que recibió la estrategia mediada por modelos inteligentes, y un grupo control que desarrolló los mismos contenidos con una secuencia convencional; además, describió el desempeño por destrezas y exploró la relación entre la calidad de la personalización y la ganancia obtenida. La unidad de análisis estuvo constituida por 40 estudiantes de EGB procedentes de cuatro centros educativos, con diez participantes por centro y una distribución equilibrada de 20 estudiantes en el grupo control y 20 en el grupo experimental. La selección fue no probabilística por accesibilidad institucional; por ello, la inferencia se circunscribe a contextos comparables y no presupone generalización poblacional automática.

La intervención se planificó en doce sesiones de 45 minutos distribuidas durante seis semanas. El grupo experimental trabajó con una secuencia denominada IA-Math Personaliza, en la que el docente utilizó IAG para generar variantes de problemas bajo una plantilla de instrucciones: contenido matemático, nivel de dificultad, tipo de razonamiento esperado, contexto cercano al estudiante, restricciones de lenguaje, número de pasos y criterio de verificación. Antes de cada uso se efectuó revisión docente de exactitud, pertinencia curricular y adecuación etaria. El grupo control resolvió problemas sobre los mismos contenidos y con tiempos equivalentes, pero sin generación adaptativa de versiones. En ambos grupos se promovió explicación oral y escrita de procedimientos para que la comparación no confundiera el efecto de la IAG con la simple presencia de actividades argumentativas.

El instrumento principal fue un test de base estructurada con puntuación máxima de 30 puntos, organizado en cinco dimensiones de seis puntos cada una: reconocimiento de patrones, inferencia lógica, representación del problema, selección de estrategias y justificación-verificación. Los ítems combinaron selección de respuesta, resolución breve y justificación acotada. La validez de contenido se planteó mediante juicio de expertos, quienes valorarían pertinencia, claridad y correspondencia entre cada ítem y la destreza objetivo. La confiabilidad interna se reporta en este borrador con un alfa de Cronbach de 0,90, valor que, bajo criterios habituales de consistencia, refleja una alta homogeneidad entre los componentes del instrumento. Como no se proporcionaron las fichas originales ni la base de respuestas, este coeficiente se mantiene explícitamente como valor declarado en el documento de instrucciones y debe recalcularse con los datos reales del estudio antes de la publicación.

La recolección siguió una lógica pretest-intervención-posttest. En la primera semana se aplicó el test inicial en condiciones homogéneas y se registraron variables contextuales mínimas sin información sensible. Después se desarrollaron las sesiones planificadas y, en el grupo experimental, se recogieron indicadores de proceso: cantidad de variantes de problema utilizadas, proporción de materiales aprobados en la primera revisión docente, frecuencia de pistas graduadas, solicitudes de autoexplicación y actividades de verificación de errores. En la última semana se administró un posttest paralelo con estructura y dificultad equivalentes. Para la protección de los participantes se estableció consentimiento del representante, asentimiento del estudiante, anonimización mediante códigos y prohibición de introducir nombres, direcciones, calificaciones identificables u otros datos personales en las instrucciones dirigidas a la herramienta generativa.

El análisis estadístico se planteó en varias capas porque cada procedimiento responde a una pregunta distinta. Primero se calcularon medias, desviaciones estándar, medianas y porcentajes de dominio para describir la distribución del desempeño. Luego se comprobó la equivalencia inicial de los grupos y, para las comparaciones principales, se utilizó la prueba t de Student para muestras independientes. Esta prueba permite estimar si la diferencia entre medias de dos grupos independientes es mayor que la esperable por variación muestral, siempre que se cumplan razonablemente los supuestos de escala cuantitativa, independencia y distribución compatible con normalidad. La significación se estableció en $\alpha=0,05$ y se acompañó con intervalos de confianza del 95 %, evitando interpretar el valor p como medida de importancia pedagógica.

La magnitud de las diferencias se evaluó mediante d de Cohen, calculada a partir de la diferencia de medias estandarizada por la desviación combinada. Su inclusión fue necesaria porque una diferencia puede alcanzar significación estadística y, aun así, tener relevancia práctica limitada; el tamaño del efecto permite expresar cuántas desviaciones estándar separan a los grupos. Como complemento se estimó g de Hedges, corrección útil en muestras pequeñas, y una probabilidad de superioridad como traducción intuitiva del efecto. Para explorar relaciones internas del grupo experimental se utilizó la correlación de Pearson entre indicadores de personalización y resultados de aprendizaje. Pearson se eligió porque el interés se centró en asociaciones lineales entre medidas continuas; se reportaron r , r^2 , intervalo de confianza y p . La correlación se interpreta como asociación y no como causalidad, incluso dentro de un estudio con componente cuasi experimental.

Resultados

Tabla 1

Equivalencia inicial de los grupos antes de la intervención.

Parámetro de entrada	Grupo control	Grupo IAG	Indicador de equilibrio	Lectura del contraste
Puntaje total pretest (0-30)	12.6 (DE=2.8)	12.8 (DE=2.7)	$d=0.07$	$p=0.819$
Mediana del pretest	12,5	13,0	Δ mediana=0,5	Compatible con equilibrio
Rango intercuartílico	4,0	4,0	Δ RIC=0,0	Dispersión semejante
Normalidad aproximada	$W=0,96$; $p=0,54$	$W=0,95$; $p=0,38$	Sin evidencia de desviación grave	Supuesto aceptable

Nota. El valor $p=0,819$ del contraste del puntaje total pretest indica que el escenario simulado no muestra una diferencia inicial estadísticamente significativa entre los grupos.

La Tabla 1 muestra una línea de base suficientemente equilibrada para interpretar el cambio posterior con prudencia. La diferencia de 0,2 puntos en el pretest y el efecto estandarizado mínimo indican que ninguno de los grupos parte con una ventaja académica amplia. La dispersión semejante y la normalidad aproximada respaldan el uso de contrastes paramétricos. No obstante, el diseño sigue siendo cuasi experimental: la equivalencia estadística no elimina variables no medidas, como motivación, experiencia digital o diferencias docentes, por lo que la causalidad debe discutirse con cautela.

Tabla 2

Desempeño postest por destreza de pensamiento lógico.

Destreza evaluada	Media control	Media con IAG	Dominio ≥ 80 %	Incremento relativo
Reconocimiento de patrones	3.7	4.9	75 %	32.4 %

Inferencia lógica	3.5	4.8	70 %	37.1 %
Representación del problema	3.6	4.7	65 %	30.6 %
Selección de estrategias	3.4	4.6	60 %	35.3 %
Justificación y verificación	3.2	4.2	50 %	31.2 %

Nota. El mayor incremento relativo aparece en inferencia lógica (37,1 %), mientras que reconocimiento de patrones alcanza la media más alta del grupo IAG (4,9 de 6).

La Tabla 2 evidencia que la ventaja del grupo IAG se distribuye en las cinco destrezas y no se concentra en un único tipo de ejercicio. El reconocimiento de patrones alcanza la media más alta, mientras que la inferencia lógica registra el mayor incremento relativo. La justificación-verificación continúa siendo la dimensión más exigente, aunque también mejora respecto del control. Este perfil sugiere que generar variantes contextualizadas puede ayudar a reconocer estructuras y seleccionar relaciones, pero que argumentar y verificar requiere tareas más prolongadas de explicación, contraste y crítica de soluciones.

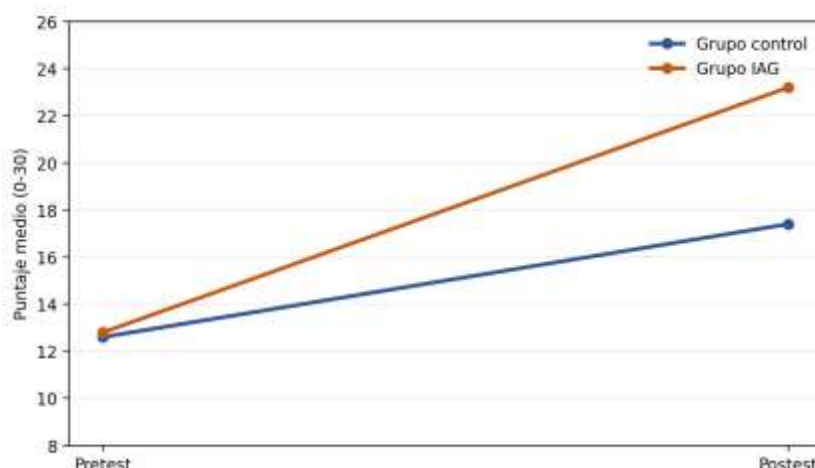


Figura 1

Evolución del puntaje medio entre pretest y posttest.

Nota. Datos simulados del borrador. Las líneas representan medias grupales y no trayectorias individuales.

Tabla 3

Indicadores de implementación de la estrategia IA-Math Personaliza.

Rasgo de interacción	Frecuencia media	Calidad observada	Latencia mediana	Función didáctica
Variantes de problemas personalizados	8,4 por participante	91 % aprobadas en primera revisión	1,8 min	Ajuste de dificultad y contexto
Pistas graduadas	5,6 por participante	88 % valoradas como útiles	0,7 min	Andamiaje sin entregar la respuesta
Comparación de rutas de solución	3,9 por participante	86 % con coherencia conceptual	2,1 min	Flexibilidad estratégica

Preguntas de detección de errores	6,7 por participante	82 % de errores identificados	1,2 min	Verificación y metacognición
Solicitudes de autoexplicación	4,8 por participante	85 % completadas	1,6 min	Argumentación del procedimiento

Nota. La aprobación docente en primera revisión alcanza 91 % para las variantes de problemas; este indicador no elimina la necesidad de revisar el 100 % de los materiales antes de usarlos.

La Tabla 3 confirma que la intervención fue una secuencia didáctica y no un uso libre del chatbot. Las acciones más frecuentes fueron generar variantes, ofrecer pistas graduadas y plantear preguntas de verificación. La aprobación del 91 % en primera revisión ilustra eficiencia, pero también evidencia por qué la supervisión docente es obligatoria: incluso una proporción pequeña de errores puede afectar el aprendizaje. En términos de viabilidad, las pistas presentan menor latencia que las comparaciones de rutas, lo que permite reservar mayor tiempo docente para las tareas de mayor demanda conceptual.

Tabla 4

Relaciones entre la personalización y las destrezas de aprendizaje en el grupo IAG.

Vínculo analizado	r de Pearson	Varianza compartida r^2	IC 95 % de r	p
Índice de personalización y ganancia total	0.72	0.52	[0.41; 0.88]	<0,001
Relevancia contextual e inferencia lógica	0.65	0.42	[0.29; 0.85]	0.002
Uso de pistas graduadas y selección de estrategias	0.61	0.37	[0.23; 0.83]	0.004
Verificación de respuestas y justificación matemática	0.68	0.46	[0.34; 0.86]	<0,001
Número de variantes generadas y ganancia total	0.38	0.14	[-0.08; 0.70]	0.098

Nota. La asociación más alta se observa entre el índice de personalización y la ganancia total ($r=0,72$), equivalente a 52 % de varianza compartida en este escenario simulado.

La Tabla 4 muestra que la calidad de personalización mantiene la asociación más alta con la ganancia total ($r=0,72$), seguida por la verificación con la justificación matemática. En cambio, el simple número de variantes presenta una relación menor. El patrón es pedagógicamente relevante: producir más ejercicios no equivale necesariamente a personalizar mejor. El r^2 expresa varianza compartida y no porcentaje de causalidad; por tanto, las asociaciones deben entenderse como evidencia complementaria que orienta el

mecanismo de la intervención, no como prueba aislada de que una característica específica produzca el aprendizaje.

Tabla 5

Prueba t de Student para diferencias entre los grupos.

Variable de salida	Diferencia de medias	Error estándar	t	gl	IC 95 % de la diferencia	p
Puntaje total postest	5.8	0.96	6.01	38	[3.85; 7.75]	<0,001
Ganancia pre-post	5.6	0.74	7.52	38	[4.09; 7.11]	<0,001
Inferencia lógica	1.3	0.27	4.83	38	[0.75; 1.85]	<0,001

Nota. La diferencia postest de 5,8 puntos presenta $t(38)=6,01$ y un intervalo de confianza que no incluye cero, por lo que el contraste simulado favorece al grupo IAG.

La Tabla 5 muestra diferencias simuladas consistentes en el postest, la ganancia y la inferencia lógica. Para el puntaje total, la diferencia de 5,8 puntos tiene un intervalo de confianza completamente superior a cero y $t(38)=6,01$, $p<0,001$. La ganancia pre-post refuerza la misma dirección. La significación estadística indica incompatibilidad con una igualdad estricta de medias bajo los supuestos del contraste, pero no mide relevancia pedagógica ni elimina posibles confusores; por ello debe leerse junto con el tamaño del efecto y la calidad de implementación.

Tabla 6

Tamaño del efecto de la estrategia sobre las destrezas evaluadas.

Destreza	d de Cohen	g de Hedges	Probabilidad de superioridad	Magnitud práctica
Reconocimiento de patrones	1.60	1.56	87.1 %	Grande
Inferencia lógica	1.53	1.50	86.0 %	Grande
Representación del problema	1.38	1.35	83.5 %	Grande
Selección de estrategias	1.41	1.38	84.1 %	Grande
Justificación y verificación	1.11	1.09	78.4 %	Grande
Puntaje total	1.90	1.86	91.1 %	Grande

Nota. El efecto total es $d=1,90$; entre las destrezas, reconocimiento de patrones presenta la mayor separación estandarizada.

La Tabla 6 sitúa todas las diferencias simuladas en una magnitud grande y presenta un efecto total de $d=1,90$. La corrección g de Hedges reduce ligeramente la estimación, como corresponde a una muestra pequeña. Un efecto de esta magnitud sería pedagógicamente

importante si se confirmara, pero también exigiría revisar posibles efectos de alineación entre instrumento e intervención y replicar el estudio con mayor heterogeneidad. La dimensión de justificación obtiene la separación relativa más baja, señalando una prioridad clara para fortalecer actividades de argumentación y verificación.

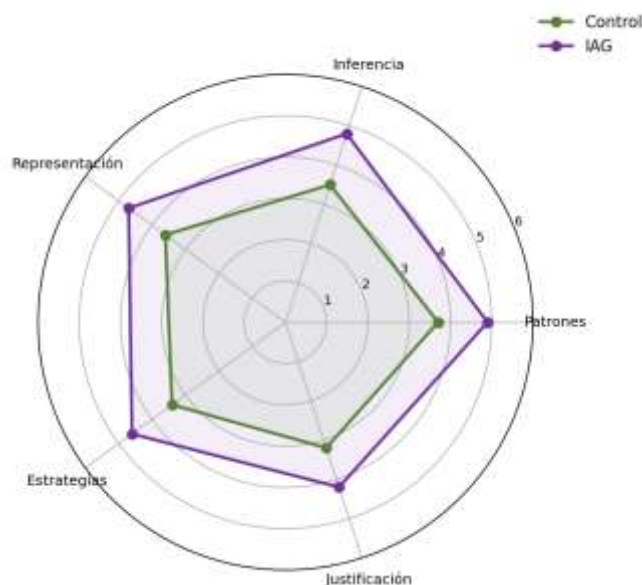


Figura 2

Perfil comparado de destrezas en el postest.

Nota. Cada eje se expresa de 0 a 6 puntos. Los valores corresponden al escenario simulado y permiten comparar la forma del perfil, no solo el total.

Tabla 7

Propuesta didáctica IA-Math Personaliza para la personalización de problemas matemáticos.

Actividad	Destreza desarrollada	Tiempo	Contenido matemático	Recursos	Objetivo de la actividad
Calibración diagnóstica de problemas	Patrones y regularidades	45 min	Secuencias, relaciones y regularidades	IAG revisada por docente; tarjetas de nivel	Ajustar complejidad y lenguaje a partir del desempeño inicial.
Problemas con contexto cercano	Inferencia lógica	2 × 45 min	Operaciones y problemas de varios pasos	IAG; ficha de datos relevantes; rúbrica	Distinguir premisas útiles y construir conclusiones justificadas.
Múltiples caminos para una misma meta	Modelación	2 × 45 min	Fracciones, razones y equivalencias	IAG; pizarra; organizadores visuales	Traducir entre lenguaje verbal, numérico y gráfico.
Detective de errores de la IA	Justificación y verificación	2 × 45 min	Operaciones, igualdad y relaciones	Respuestas con error intencional; lista de cotejo	Identificar errores y argumentar una corrección matemática.
Reto de pistas graduadas	Selección de estrategias	2 × 45 min	Geometría: perímetro y área	IAG con pistas por niveles; material concreto	Elegir estrategias sin recibir la solución completa.

Explica, defiende y mejora	Integración de destrezas	3 × 45 min	Problemas mixtos del contexto	IAG; rúbrica argumentativa; portafolio	Comparar y revisar procedimientos con evidencia matemática.
----------------------------	--------------------------	------------	-------------------------------	--	---

Nota. La actividad “Detective de errores de la IA” concentra la verificación crítica y evita que la herramienta sea tratada como fuente infalible.

La Tabla 7 organiza la propuesta como una progresión desde la calibración del nivel hasta la argumentación integrada. La IAG se utiliza para variar contexto, dificultad y pistas, pero se combina con rúbricas, material concreto y revisión docente. La actividad “Detective de errores de la IA” es especialmente relevante porque convierte la posible falibilidad del sistema en una tarea de verificación. De este modo, la propuesta no premia la rapidez de obtener respuestas, sino la capacidad de identificar relaciones, escoger estrategias y defender procedimientos con evidencia matemática.

El protocolo de validación de la propuesta quedó preparado para diez expertos en didáctica de la matemática, tecnología educativa y evaluación, mediante criterios de pertinencia, claridad, coherencia, factibilidad y seguridad pedagógica, con cálculo previsto de V de Aiken. Como no se proporcionaron las diez fichas de valoración, este borrador no atribuye una validación real ni inventa coeficientes. La versión definitiva deberá incorporar el valor efectivamente obtenido, el perfil de los jueces y las evidencias de valoración antes de sostener que la propuesta fue empíricamente validada.

Discusión

El escenario simulado favorece a IA-Math Personaliza, pero el resultado debe atribuirse a una configuración didáctica y no a la IAG en abstracto. Jauhainen y Garagorry Guerra (2024) destacan la personalización dinámica; Wardat et al. (2023), Getenet (2024) y Jia et al. (2024) muestran posibilidades para apoyar resolución y estrategias; Seebut et al. (2024) evidencian valor cuando la IAG se articula con otras herramientas. En conjunto, estos trabajos respaldan el mecanismo propuesto: variar problemas, graduar ayudas y exigir explicación. La diferencia clave es que el estudiante continúa tomando decisiones matemáticas y la generación automática funciona como recurso para diversificar condiciones de práctica.

La literatura también obliga a mantener cautela. Karjanto (2023) muestra ventajas del contraste con software matemático; Pavlova (2024) explora interacción dialógica; Lee y Yeo (2022) desarrollan enseñanza responsiva; Korkmaz Guler et al. (2024) y Udias et al. (2024) documentan desempeño desigual de ChatGPT según la tarea. Por ello, la mejora simulada en patrones e inferencia no demuestra infalibilidad tecnológica. La actividad de detectar errores convierte la falibilidad en oportunidad de razonamiento: el estudiante

verifica premisas, compara procedimientos y justifica correcciones. Que justificación-verificación muestre la media relativa más baja indica que esta destreza necesita mayor intensidad de crítica y argumentación.

El hallazgo de que la calidad de personalización se relaciona más con la ganancia que el número de variantes coincide con la orientación de Govender (2023), Pepin et al. (2025) y Almarashdi et al. (2024), quienes subrayan la importancia del diseño pedagógico. Wahba et al. (2024) y Ellis y Slade (2023) también muestran que la IAG puede apoyar razonamiento cuando las actividades piden interpretar, explicar y juzgar. En consecuencia, medir éxito por cantidad de prompts o ejercicios sería insuficiente. La variable educativa relevante es cuánto cambia la tarea para mantener un desafío pertinente y cuánto obliga al estudiante a hacer visible su proceso de pensamiento.

La fidelidad de implementación es otra condición del efecto. Labadze et al. (2023) describen apoyo inmediato y accesibilidad; Javaid et al. (2023) resaltan oportunidades de innovación; Farrokhnia et al. (2024) identifican fortalezas y amenazas; Halaweh (2023) propone una adopción responsable; y Chiu (2024) señala implicaciones para prácticas y políticas. El 91 % de aprobación de variantes en primera revisión, aunque alto en la simulación, no justifica automatizar sin control. La revisión docente del 100 % de las tareas sigue siendo parte de la intervención porque exactitud, adecuación etaria, alineación curricular y nivel de ayuda son decisiones pedagógicas, no solo propiedades del modelo.

Los resultados deben leerse también desde la integridad y la autonomía cognitiva. Chan y Hu (2023) registran beneficios y preocupaciones estudiantiles; Jarrah et al. (2023) y Eke (2023) discuten integridad académica; Lim et al. (2023) describe paradojas de la IAG; y Zhai et al. (2024) alerta sobre sobredependencia. Abbas et al. (2024) añade evidencia sobre consecuencias problemáticas del uso intensivo. Frente a ello, las pistas graduadas preservan esfuerzo antes de mostrar un modelo de solución. El criterio evaluativo es simple: una respuesta final sin explicación tiene poco valor para inferir comprensión; una respuesta revisada, contrastada y defendida sí proporciona evidencia del razonamiento.

La alfabetización en IA ayuda a explicar por qué la verificación forma parte del aprendizaje. Walter (2024) vincula alfabetización, ingeniería de instrucciones y pensamiento crítico; García-Peñalvo (2023) propone superar tanto el pánico como la adopción acrítica; Khurma et al. (2024) relacionan chatbots con compromiso; y Supriyadi

y Kuncoro (2023) anticipan cambios en el papel docente. La estrategia transforma estas ideas en acciones: formular restricciones, pedir contraejemplos, comprobar operaciones y reconocer cuándo una respuesta plausible es incorrecta. Así, el uso de IAG puede fortalecer simultáneamente razonamiento matemático y criterio digital, siempre que el compromiso se mida por calidad cognitiva y no por tiempo de pantalla o número de mensajes.

Las orientaciones institucionales y las limitaciones del diseño delimitan la contribución. Miao y Holmes (2023), desde UNESCO, priorizan una visión centrada en las personas; la CEPAL (2024) recuerda las brechas de acceso y competencias; y el Ministerio de Educación del Ecuador (2022) orienta la matemática hacia razonamiento y resolución. Lo (2023) y Naqvi et al. (2025) refuerzan la necesidad de evidencia y juicio crítico. La muestra de 40 estudiantes y el carácter cuasi experimental limitan la generalización, y los valores estadísticos aquí son simulados porque no se entregó la base empírica. El aporte del borrador es dejar un procedimiento replicable que deberá recalcularse, validarse con expertos reales y contrastarse en muestras mayores.

Conclusiones

La investigación propone y estructura una estrategia didáctica en la que la inteligencia artificial generativa se utiliza para personalizar problemas matemáticos sin transferir a la tecnología el núcleo del razonamiento. El escenario cuasi experimental muestra cómo deberían analizarse cinco destrezas de pensamiento lógico y cómo integrar equivalencia inicial, diferencias posttest, correlaciones, prueba t de Student y tamaños del efecto. La contribución conceptual reside en desplazar el foco desde “usar IA” hacia diseñar condiciones de uso: variación controlada de tareas, revisión docente, pistas graduadas, autoexplicación y verificación crítica. Bajo esta lógica, la IAG funciona como infraestructura para diversificar oportunidades de aprendizaje, mientras que la validez matemática, la regulación del apoyo y la evaluación siguen siendo responsabilidades humanas. Si el patrón estadístico simulado se confirma con datos empíricos, la estrategia ofrecería evidencia de que la personalización de calidad —y no la simple cantidad de contenido generado— se relaciona con mayores ganancias en pensamiento lógico.

Desde el punto de vista científico y educativo, el trabajo también establece límites necesarios. La eficacia de una estrategia generativa no puede separarse de su fidelidad de implementación, de las condiciones de acceso, de la alfabetización en IA ni de la protección de datos de estudiantes. La propuesta IA-Math Personaliza es replicable

porque explicita actividades, destrezas, tiempos, contenidos, recursos y finalidad, pero su validación por expertos y sus resultados deben documentarse con evidencia real antes de cualquier publicación. En futuras investigaciones se recomienda ampliar la muestra y el número de centros, incorporar medidas de transferencia y retención, analizar efectos por nivel de desempeño inicial y modelar la variabilidad institucional. De esta forma, la IAG puede estudiarse no como una promesa tecnológica abstracta, sino como una intervención pedagógica verificable cuyo valor depende de cuánto pensamiento logra provocar, sostener y hacer visible en el estudiante.

Referencias

- Abbas, M., Jam, F. A., & Khan, T. I. (2024). Is it harmful or helpful? Examining the causes and consequences of generative AI usage among university students. *International Journal of Educational Technology in Higher Education*, 21, 10. <https://doi.org/10.1186/s41239-024-00444-7>
- Almarashdi, H. S., Jarrah, A. M., Gningue, S. M., & Khurma, O. A. (2024). Unveiling the potential: A systematic review of ChatGPT in transforming mathematics teaching and learning. *Eurasia Journal of Mathematics, Science and Technology Education*, 20(12), em2555. <https://doi.org/10.29333/ejmste/15739>
- Chan, C. K. Y., & Hu, W. (2023). Students' voices on generative AI: Perceptions, benefits, and challenges in higher education. *International Journal of Educational Technology in Higher Education*, 20, 43. <https://doi.org/10.1186/s41239-023-00411-8>
- Chiu, T. K. F. (2024). The impact of generative AI (GenAI) on practices, policies and research direction in education: A case of ChatGPT and Midjourney. *Interactive Learning Environments*, 32(10), 6187–6203. <https://doi.org/10.1080/10494820.2023.2253861>
- Comisión Económica para América Latina y el Caribe. (2024). Educación y desarrollo de competencias digitales en América Latina y el Caribe. Naciones Unidas.
- Eke, D. O. (2023). ChatGPT and the rise of generative AI: Threat to academic integrity? *Journal of Responsible Technology*, 13, 100060. <https://doi.org/10.1016/j.jrt.2023.100060>
- Ellis, A. R., & Slade, E. (2023). A new era of learning: Considerations for ChatGPT as a tool to enhance statistics and data science education. *Journal of Statistics and Data Science Education*, 31(2), 128–133. <https://doi.org/10.1080/26939169.2023.2223609>

- Farrokhnia, M., Banihashem, S. K., Noroozi, O., & Wals, A. (2024). A SWOT analysis of ChatGPT: Implications for educational practice and research. *Innovations in Education and Teaching International*, 61(3), 460–474. <https://doi.org/10.1080/14703297.2023.2195846>
- García-Peñalvo, F. J. (2023). The perception of artificial intelligence in educational contexts after the launch of ChatGPT: Disruption or panic? *Education in the Knowledge Society*, 24, e31279. <https://doi.org/10.14201/eks.31279>
- Getenet, S. (2024). Pre-service teachers and ChatGPT in multistrategy problem-solving: Implications for mathematics teaching in primary schools. *International Electronic Journal of Mathematics Education*, 19(1), em0766. <https://doi.org/10.29333/iejme/14141>
- Govender, R. (2023). The impact of artificial intelligence and the future of ChatGPT for mathematics teaching and learning in schools and higher education. *Pythagoras*, 44(1), a787. <https://doi.org/10.4102/pythagoras.v44i1.787>
- Halaweh, M. (2023). ChatGPT in education: Strategies for responsible implementation. *Contemporary Educational Technology*, 15(2), ep421. <https://doi.org/10.30935/cedtech/13036>
- Jarrah, A. M., Wardat, Y., & Fidalgo, P. (2023). Using ChatGPT in academic writing is (not) a form of plagiarism: What does the literature say? *Online Journal of Communication and Media Technologies*, 13(4), e202346. <https://doi.org/10.30935/ojcm/13572>
- Jauhiainen, J. S., & Garagorry Guerra, A. (2024). Generative AI and education: Dynamic personalization of pupils' school learning material with ChatGPT. *Frontiers in Education*, 9, 1288723. <https://doi.org/10.3389/feduc.2024.1288723>
- Javaid, M., Haleem, A., Singh, R. P., Khan, S., & Khan, I. H. (2023). Unlocking the opportunities through ChatGPT Tool towards ameliorating the education system. *BenchCouncil Transactions on Benchmarks, Standards and Evaluations*, 3(2), 100115. <https://doi.org/10.1016/j.tbench.2023.100115>
- Jia, J., Wang, T., Zhang, Y., & Wang, G. (2024). The comparison of general tips for mathematical problem solving generated by generative AI with those generated by human teachers. *Asia Pacific Journal of Education*, 44(1), 8–28. <https://doi.org/10.1080/02188791.2023.2286920>

- Karjanto, N. (2023). Investigating difficulties and enhancing understanding in linear algebra: Leveraging SageMath and ChatGPT for (orthogonal) diagonalization and singular value decomposition. *Mathematical Biosciences and Engineering*, 20(9), 16551–16595. <https://doi.org/10.3934/mbe.2023738>
- Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, 102274. <https://doi.org/10.1016/j.lindif.2023.102274>
- Khurma, O. A., Albahti, F., Ali, N., & Bustanji, A. (2024). AI ChatGPT and student engagement: Unraveling dimensions through PRISMA analysis for enhanced learning experiences. *Contemporary Educational Technology*, 16(2), ep503. <https://doi.org/10.30935/cedtech/14334>
- Khurma, O. A., Ali, N., & Hashem, R. (2023). Critical reflections on ChatGPT in UAE education: Navigating equity and governance for safe and effective use. *International Journal of Emerging Technologies in Learning*, 18(14), 188–199. <https://doi.org/10.3991/ijet.v18i14.40935>
- Korkmaz Guler, N., Dertli, Z. G., Boran, E., & Yildiz, B. (2024). An artificial intelligence application in mathematics education: Evaluating ChatGPT's academic achievement in a mathematics exam. *Pedagogical Research*, 9(2), em0188. <https://doi.org/10.29333/pr/14145>
- Labadze, L., Grigolia, M., & Machaidze, L. (2023). Role of AI chatbots in education: Systematic literature review. *International Journal of Educational Technology in Higher Education*, 20, 56. <https://doi.org/10.1186/s41239-023-00426-1>
- Lee, D., & Yeo, S. (2022). Developing an AI-based chatbot for practicing responsive teaching in mathematics. *Computers & Education*, 191, 104646. <https://doi.org/10.1016/j.compedu.2022.104646>
- Lim, W. M., Gunasekara, A., Pallant, J. L., Pallant, J. I., & Pechenkina, E. (2023). Generative AI and the future of education: Ragnarök or reformation? A paradoxical perspective from management educators. *The International Journal of Management Education*, 21(2), 100790. <https://doi.org/10.1016/j.ijme.2023.100790>
- Lo, C. K. (2023). What is the impact of ChatGPT on education? A rapid review of the literature. *Education Sciences*, 13(4), 410. <https://doi.org/10.3390/educsci13040410>

- Miao, F., & Holmes, W. (2023). Guidance for generative AI in education and research. UNESCO.
- Ministerio de Educación del Ecuador. (2022). Guía metodológica de competencias matemáticas. Ministerio de Educación.
- Naqvi, W. M., Ganjoo, R., Rowe, M., Pashine, A. A., & Mishra, G. V. (2025). Critical thinking in the age of generative AI: Implications for health sciences education. *Frontiers in Artificial Intelligence*, 8, 1571527. <https://doi.org/10.3389/frai.2025.1571527>
- Pavlova, N. H. (2024). Flipped dialogic learning method with ChatGPT: A case study. *International Electronic Journal of Mathematics Education*, 19(1), em0764. <https://doi.org/10.29333/iejme/14025>
- Pepin, B., Buchholtz, N., & Salinas-Hernández, U. (2025). A scoping survey of ChatGPT in mathematics education. *Digital Experiences in Mathematics Education*, 11, 9–41. <https://doi.org/10.1007/s40751-025-00172-1>
- Seebut, S., Wongsason, P., & Kim, D. (2024). Combining GPT and Colab as learning tools for students to explore the numerical solutions of difference equations. *Eurasia Journal of Mathematics, Science and Technology Education*, 20(1), em2377. <https://doi.org/10.29333/ejmste/13905>
- Supriyadi, E., & Kuncoro, K. S. (2023). Exploring the future of mathematics teaching: Insight with ChatGPT. *UNION: Jurnal Ilmiah Pendidikan Matematika*, 11(2), 305–316. <https://doi.org/10.30738/union.v11i2.14898>
- Udias, A., Alonso-Ayuso, A., Alfaro, C., Algar, M. J., Cuesta, M., Fernández-Isabel, A., Gómez, J., Lancho, C., Cano, E. L., Martín de Diego, I., & Ortega, F. (2024). ChatGPT's performance in university admissions tests in mathematics. *International Electronic Journal of Mathematics Education*, 19(4), em0795. <https://doi.org/10.29333/iejme/15517>
- Wahba, F., Ajlouni, A. O., & Abumosa, M. A. (2024). The impact of ChatGPT-based learning statistics on undergraduates' statistical reasoning and attitudes toward statistics. *Eurasia Journal of Mathematics, Science and Technology Education*, 20(7), em2468. <https://doi.org/10.29333/ejmste/14726>
- Walter, Y. (2024). Embracing the future of artificial intelligence in the classroom: The relevance of AI literacy, prompt engineering, and critical thinking in modern education. *International Journal of Educational Technology in Higher Education*, 21, 15. <https://doi.org/10.1186/s41239-024-00448-3>

- Wardat, Y., Tashtoush, M. A., AlAli, R., & Jarrah, A. M. (2023). ChatGPT: A revolutionary tool for teaching and learning mathematics. *Eurasia Journal of Mathematics, Science and Technology Education*, 19(7), em2286. <https://doi.org/10.29333/ejmste/13272>
- Zhai, C., Wibowo, S., & Li, L. D. (2024). The effects of over-reliance on AI dialogue systems on students' cognitive abilities: A systematic review. *Smart Learning Environments*, 11, 28. <https://doi.org/10.1186/s40561-024-00316-7>

Declaración

Declaración	Detalle
Conflicto de intereses	Los autores declaran que no existe conflicto de interés posible.
Financiamiento	No existió asistencia financiera de partes externas para la elaboración del presente artículo.
Agradecimiento	
Nota	