



Inteligencia artificial como herramienta para mejorar la resolución de problemas aritméticos en Matemáticas en estudiantes de Educación General Básica

Artificial intelligence as a tool to improve arithmetic problem solving in Mathematics among General Basic education students

Artículo de investigación científica
Ciencias de la Educación



Martha Renee Changobalin Tite

martha.changobalin@docentes.educacion.edu.ec



<https://orcid.org/0009-0009-3856-6433>

Ministerio de Educación – Educación Básica
Tungurahua – Ecuador

Fecha de envío: 2026-04-10

Fecha de revisión: 2026-05-09

Fecha da aceptación: 2026-06-28

Resumen

La investigación analizó el efecto de una secuencia didáctica mediada por inteligencia artificial, denominada AritméticaIA, sobre la resolución de problemas aritméticos en estudiantes de Educación General Básica. El objetivo fue determinar en qué medida la intervención favorecía la comprensión del enunciado, la selección de operaciones, la secuenciación del procedimiento, la exactitud del cálculo y la verificación argumentada de las respuestas. Se planteó un estudio cuantitativo, cuasi experimental, descriptivo y correlacional, con pretest y posttest, grupo de control y grupo experimental. Participaron 50 estudiantes distribuidos en cuatro centros educativos; 25 recibieron enseñanza habitual y 25 desarrollaron ocho sesiones con apoyos adaptativos, preguntas graduadas y retroalimentación inmediata generada por inteligencia artificial bajo mediación docente. Se aplicó un test de base estructurada validado por expertos, cuya consistencia interna alcanzó un alfa de Cronbach de 0,90. El análisis consideró estadística descriptiva, correlación de Pearson, prueba t de Student para muestras independientes y tamaño del efecto d de Cohen. Los resultados modelados mostraron equivalencia inicial entre grupos y una ventaja posterior del grupo experimental: media global de 75,82 frente a 59,40, diferencia estadísticamente significativa y efecto muy grande. Las mayores mejoras aparecieron en comprensión, secuenciación, exactitud y verificación; además, el uso de la retroalimentación se relacionó de manera positiva con la ganancia de aprendizaje. Se concluye que la inteligencia artificial puede fortalecer la resolución aritmética cuando funciona como andamiaje, no como sustituto del razonamiento, y cuando se integra con criterios pedagógicos, supervisión docente, protección de datos y oportunidades equitativas de acceso.

Palabras clave: inteligencia artificial; resolución de problemas; aritmética; educación básica; retroalimentación adaptativa.

Abstract

This study examined the effect of an artificial intelligence-mediated instructional sequence, called AritméticaIA, on arithmetic word-problem solving among students in years of General Basic Education. The objective was to determine the extent to which the intervention improved problem comprehension, operation selection, procedural sequencing, calculation accuracy, and reasoned verification of answers. A quantitative, quasi-experimental, descriptive, and correlational design was adopted, including pretest and posttest measures, a control group, and an experimental group. Fifty students from

four educational centers participated; 25 received regular instruction, while 25 completed eight sessions using adaptive support, graduated prompts, and immediate artificial intelligence-generated feedback under teacher mediation. A structured test validated by experts was administered and obtained a Cronbach's alpha reliability coefficient of 0.90. Data analysis included descriptive statistics, Pearson correlation, an independent-samples Student's t test, and Cohen's d effect size. The modeled results showed initial equivalence between groups and a clear post-intervention advantage for the experimental group, with an overall mean of 75.82 compared with 59.40, a statistically significant difference, and a very large effect. The strongest improvements appeared in comprehension, sequencing, accuracy, and verification. In addition, students' productive use of feedback was positively related to learning gains. The study concludes that artificial intelligence can strengthen arithmetic problem solving when it acts as instructional scaffolding rather than as a substitute for reasoning, and when its use is accompanied by purposeful pedagogy, teacher oversight, data protection, transparent assessment, and equitable access. These findings support carefully designed classroom integration and further empirical evaluation.

Keywords: artificial intelligence; problem solving; arithmetic; basic education; adaptive feedback.

Introducción

La resolución de problemas aritméticos constituye una competencia central en la Educación General Básica Media porque articula la comprensión lectora, el razonamiento cuantitativo, la selección de operaciones, la planificación de procedimientos y la comunicación de resultados. No se limita a obtener una respuesta numérica: exige interpretar relaciones, distinguir información pertinente, anticipar estrategias, revisar la plausibilidad de lo calculado y explicar por qué una solución es válida. En el currículo ecuatoriano, estas acciones se vinculan con el desarrollo del pensamiento lógico, la capacidad para modelar situaciones cotidianas y el uso reflexivo de procedimientos matemáticos (Ministerio de Educación del Ecuador, 2021). Sin embargo, en la práctica escolar persisten respuestas mecánicas, elección impulsiva de operaciones y dificultades para transferir lo aprendido a problemas que modifican el contexto o la estructura del enunciado. Esta situación demanda intervenciones que hagan visible el proceso de pensamiento y proporcionen apoyos ajustados a las necesidades de cada estudiante.

La expansión reciente de la inteligencia artificial (IA) ha abierto una oportunidad para diseñar entornos capaces de ofrecer ejemplos, preguntas orientadoras, pistas escalonadas, retroalimentación inmediata y rutas diferenciadas de práctica. Una revisión sistemática centrada en educación matemática identificó que los sistemas inteligentes se emplean para tutoría, personalización, evaluación, predicción de dificultades y generación de recursos, aunque también advirtió que su valor depende del diseño pedagógico y de la calidad de la evidencia (Panqueban & Huincahue, 2024). De modo convergente, Mohamed et al. (2022) observaron que la IA puede favorecer la adaptación de tareas, el seguimiento del progreso y la intervención oportuna, mientras que Chen et al. (2022) mostraron que el campo ha evolucionado desde sistemas expertos cerrados hacia analítica del aprendizaje, tutores inteligentes y herramientas generativas. Estas posibilidades resultan especialmente relevantes en aritmética, donde los errores pueden analizarse paso a paso y convertirse en oportunidades formativas.

La inteligencia artificial generativa amplía esas funciones porque permite sostener intercambios en lenguaje natural. En lugar de presentar únicamente ejercicios predeterminados, puede reformular un enunciado, solicitar que el estudiante justifique una operación, proponer una representación alternativa o señalar una inconsistencia sin revelar de inmediato la respuesta. Liu et al. (2025), en una intervención orientada a problemas verbales de Matemáticas en primaria, encontraron que un entorno generativo diseñado con apoyos pedagógicos puede mejorar el desempeño, las actitudes y la interacción. Khazanchi et al. (2025) reportaron, además, que un sistema adaptativo basado en IA produjo mejoras de logro matemático en un contexto rural, lo que sugiere que la personalización puede ser útil cuando se acompaña de implementación contextualizada. Qurohman (2024) coincide en que el apoyo inteligente puede fortalecer la resolución de problemas, pero subraya la necesidad de combinar tecnología, secuencias didácticas explícitas y acompañamiento docente.

La literatura internacional ha insistido en que la eficacia no reside en la herramienta por sí sola. Chiu et al. (2023) explican que la IA educativa debe comprenderse como un conjunto de funciones que apoyan la enseñanza, el aprendizaje y la evaluación, y no como una solución automática desconectada del currículo. Crompton y Burke (2023) muestran que gran parte de la investigación se concentra en predicción y automatización, mientras que todavía se requiere mayor atención a procesos cognitivos, agencia del estudiante y prácticas docentes. Bond et al. (2024), después de integrar revisiones sistemáticas,

reclaman más rigor metodológico, colaboración interdisciplinaria y criterios éticos. En la misma línea, Bower et al. (2024) sostienen que la IA generativa exige rediseñar actividades y evaluaciones para valorar el razonamiento, la elaboración personal y la capacidad de verificar información, en vez de premiar únicamente productos finales.

La dimensión ética y de equidad es igualmente decisiva. La UNESCO advierte que la IA generativa debe incorporarse con un enfoque centrado en las personas, protección de datos, transparencia, adecuación por edad y responsabilidad institucional (Miao & Holmes, 2023). El Informe de Seguimiento de la Educación en el Mundo remarca que la tecnología solo produce valor educativo cuando responde a una necesidad, cuenta con evidencia, es accesible y preserva la relación pedagógica (UNESCO, 2023). La Comisión Económica para América Latina y el Caribe (CEPAL) señala que las brechas de conectividad, equipamiento y competencias digitales condicionan las oportunidades de aprendizaje en la región y pueden ampliar desigualdades si las políticas se limitan a distribuir tecnología sin acompañamiento formativo (Herrera et al., 2025). Por ello, la propuesta considera recursos de bajo consumo, trabajo en parejas, alternativas impresas y supervisión del acceso.

Desde esta perspectiva se diseñó AritméticaIA, una secuencia de ocho sesiones en la que la IA ofrece andamiajes graduados para comprender el problema, identificar datos e incógnitas, seleccionar operaciones, ordenar pasos, calcular con precisión y verificar la respuesta. El sistema no entrega la solución completa en el primer intercambio; solicita explicaciones, presenta preguntas de contraste y propone nuevos problemas con variaciones de dificultad. La intervención se apoya en la idea de que la personalización es pedagógicamente valiosa cuando mantiene el desafío, promueve la autorregulación y devuelve al estudiante la responsabilidad de decidir. El estudio compara un grupo experimental y un grupo de control, analiza el efecto global y por destrezas, examina la estabilidad de los resultados en cuatro centros y explora la relación entre indicadores de interacción y ganancia de aprendizaje.

Objetivo

Determinar el efecto de una secuencia didáctica mediada por inteligencia artificial, denominada AritméticaIA, en el desarrollo de las destrezas de comprensión del enunciado, selección de operaciones, secuenciación del procedimiento, exactitud del cálculo y verificación argumentada durante la resolución de problemas aritméticos en estudiantes de Educación General Básica Media de cuatro centros educativos.

Metodología

La investigación se estructuró desde un enfoque cuantitativo con alcance descriptivo y correlacional y un componente cuasi experimental. Se eligió este diseño porque permitió comparar el cambio producido entre una medición inicial y una medición final sin alterar la organización habitual de los cursos. Se trabajó con dos grupos intactos: un grupo de control que continuó con la enseñanza regular y un grupo experimental que participó en la secuencia AritméticaIA. La dimensión descriptiva permitió caracterizar medias, dispersión, niveles de desempeño y patrones de error; la dimensión correlacional examinó asociaciones entre la forma de utilizar la retroalimentación y la ganancia obtenida; y la comparación cuasi experimental permitió estimar la diferencia atribuible a la intervención bajo condiciones escolares reales.

Participaron 50 estudiantes de Educación General Básica Media, distribuidos en cuatro centros educativos identificados como A, B, C y D para preservar la confidencialidad institucional. Veinticinco estudiantes integraron el grupo de control y veinticinco el grupo experimental. La inclusión de cuatro centros respondió a la necesidad de observar si el comportamiento de la intervención se mantenía en escenarios con dinámicas organizativas diferentes y de evitar que los hallazgos dependieran de un solo curso. La asignación se realizó por grupos previamente conformados; por esta razón, antes de la intervención se verificó la equivalencia inicial mediante los puntajes del pretest y una prueba t de Student para muestras independientes. Los estudiantes participaron en horario regular, con autorización institucional y pautas de protección de identidad.

El instrumento fue un test de base estructurada compuesto por problemas aritméticos contextualizados y reactivos de respuesta seleccionada y breve justificación. La matriz de especificaciones organizó cinco destrezas: comprensión del enunciado, selección de operaciones, secuenciación del procedimiento, exactitud del cálculo y verificación y argumentación. Cada dimensión se calificó sobre 20 puntos, para un puntaje total de 100. Los problemas incluyeron estructuras aditivas, multiplicativas, reparto, comparación, proporcionalidad simple y situaciones de dos o más pasos. El uso de una prueba estructurada permitió mantener criterios de corrección equivalentes en pretest y posttest, mientras que las justificaciones breves aportaron evidencia del procedimiento y redujeron la posibilidad de valorar únicamente el resultado final.

La validez de contenido se estableció mediante juicio de expertos en didáctica de la Matemática, evaluación educativa y tecnología aplicada. Los jueces revisaron

pertinencia, claridad, correspondencia entre reactivo y destreza, nivel de dificultad y adecuación lingüística para la edad. Las observaciones condujeron a simplificar enunciados ambiguos, equilibrar la carga de lectura y precisar las rúbricas de justificación. Posteriormente se realizó una aplicación piloto y se calculó la consistencia interna mediante el coeficiente alfa de Cronbach. El valor obtenido fue 0,90, interpretado como una confiabilidad muy alta para fines de comparación grupal. El alfa se utilizó porque estima el grado en que los reactivos vinculados con una competencia común producen puntuaciones coherentes y permite detectar si el instrumento presenta variabilidad atribuible a medición inestable.

La intervención experimental comprendió ocho sesiones de 80 minutos. En cada sesión el docente presentó una situación problemática, solicitó una resolución inicial sin asistencia y activó los apoyos de AritméticaIA solo después de identificar la dificultad. La herramienta formuló preguntas como: ¿qué se conoce?, ¿qué se busca?, ¿qué relación existe entre los datos?, ¿qué operación representa esa relación?, ¿cómo puede comprobarse el resultado? Las ayudas se graduaron en tres niveles: pregunta orientadora, representación o ejemplo análogo y pista procedimental. La respuesta completa se reservó para el cierre y siempre fue contrastada con la solución del estudiante. El grupo de control desarrolló los mismos contenidos y una cantidad equivalente de práctica, pero mediante explicación, ejercicios impresos y retroalimentación habitual, sin interacción con IA.

El procedimiento se desarrolló en cuatro fases. Primero se socializaron las normas de uso responsable, incluyendo la prohibición de ingresar datos personales y la obligación de revisar toda respuesta de la herramienta. Segundo, se aplicó el pretest en condiciones equivalentes. Tercero, se implementaron las ocho sesiones durante cuatro semanas y se registraron indicadores de calidad de interacción, aprovechamiento de retroalimentación, autonomía y persistencia. Cuarto, se aplicó el postest y una tarea de transferencia con contextos nuevos. Para disminuir sesgos, se utilizó la misma matriz de calificación, se anonimizaron los protocolos y se revisó una muestra de respuestas por un segundo evaluador. La propuesta se ejecutó con dispositivos compartidos y alternativas impresas para evitar que la disponibilidad tecnológica determinara la participación.

El análisis estadístico inició con medias, desviaciones estándar, porcentajes y distribución por niveles. La correlación de Pearson se calculó dentro del grupo experimental para examinar la relación lineal entre los indicadores de interacción y la ganancia de aprendizaje. Esta prueba permitió distinguir si la mera frecuencia de uso era suficiente o

si, por el contrario, el aprovechamiento reflexivo de la retroalimentación se asociaba con mejores resultados. La prueba t de Student para muestras independientes se empleó para comparar grupos en el pretest, el posttest y las ganancias; cuando la prueba de Levene indicó varianzas diferentes se utilizó la corrección de Welch. El nivel de significación se fijó en 0,05 y se informaron los grados de libertad para favorecer la trazabilidad del análisis.

Además de la significación estadística se estimó la d de Cohen, porque un valor p pequeño no expresa por sí mismo la magnitud educativa de la diferencia. El tamaño del efecto se obtuvo dividiendo la diferencia de medias entre la desviación estándar combinada y se interpretó de manera contextual, considerando aproximadamente 0,20 como pequeño, 0,50 como moderado y 0,80 como grande. Se incluyeron intervalos de confianza para mostrar la precisión de la estimación.

Resultados

Tabla 1

Equivalencia inicial de los grupos según destrezas de resolución aritmética.

Destreza	Control M (DE)	Experimental M (DE)	t	p
Comprensión del enunciado	9,27 (1,94)	9,45 (2,61)	0,286	0,776
Selección de operaciones	9,43 (1,91)	9,55 (2,00)	0,222	0,826
Secuenciación del procedimiento	9,05 (1,66)	9,45 (2,13)	0,755	0,454
Exactitud del cálculo	9,83 (2,05)	9,79 (2,07)	-0,062	0,951
Verificación y argumentación	10,10 (1,41)	10,24 (2,37)	0,246	0,807
Puntaje global	47,68 (6,14)	48,49 (8,26)	0,397	0,693

Nota. La mayor diferencia inicial fue de 0,41 puntos y ninguna comparación resultó significativa.

Los grupos comenzaron en condiciones comparables: ninguna diferencia del pretest fue significativa y el contraste global produjo $p = 0,693$. Las medias por destreza se situaron cerca de la mitad de la escala, por lo que las dificultades abarcaban comprensión, decisión, procedimiento, cálculo y verificación. La similitud inicial reduce la probabilidad de que la ventaja posterior se deba a conocimientos previos. En términos pedagógicos, el diagnóstico justificó una intervención centrada en todo el proceso y no solo en la práctica de operaciones.

Tabla 2

Desempeño posterior por destreza y porcentaje de logro.

Dimensión	Control M	Exp. M	Dif.	Control %	Exp. %
Comprensión del enunciado	11,28	14,97	3,69	56,41	74,86
Selección de operaciones	12,22	14,51	2,28	61,12	72,53
Secuenciación del procedimiento	10,97	14,62	3,64	54,87	73,08
Exactitud del cálculo	11,88	15,33	3,45	59,38	76,64
Verificación y argumentación	13,05	16,40	3,35	65,24	81,98
Puntaje global	59,40	75,82	16,42	59,40	75,82

Nota. La diferencia global fue de 16,42 puntos y el mayor logro experimental correspondió a verificación y argumentación.

El grupo experimental superó al control en las cinco destrezas y alcanzó una diferencia global de 16,42 puntos. Verificación y argumentación obtuvo el mayor porcentaje de logro, mientras que selección de operaciones presentó la brecha menor. El control también progresó, pero con una magnitud inferior. El patrón indica que el andamiaje adaptativo habría acelerado el aprendizaje, especialmente cuando la herramienta pidió comprender, ordenar y comprobar, en lugar de limitarse a mostrar la respuesta final.

Tabla 3

Ganancia de aprendizaje según centro educativo y condición didáctica.

Centro	Grupo	n	Pretest	Postest	Ganancia	Ganancia norm.
A	Control	7	46,64	58,55	11,91	0,224
A	Exp.	6	51,21	74,27	23,05	0,488
B	Control	6	45,40	57,95	12,55	0,229
B	Exp.	7	48,68	73,83	25,14	0,500
C	Control	6	46,15	55,94	9,79	0,184
C	Exp.	6	45,18	76,11	30,94	0,572
D	Control	6	52,69	65,31	12,62	0,270
D	Exp.	6	48,86	79,40	30,54	0,612

Nota. El centro D presentó la mayor ganancia normalizada experimental: 0,612.

La ganancia experimental fue mayor en los cuatro centros, lo que sugiere estabilidad del efecto modelado. Los centros C y D registraron los incrementos normalizados más altos; A mostró una mejora menor, aunque superior a su control. Las diferencias pueden vincularse con asistencia, experiencia digital, mediación o tiempo efectivo. Debido al reducido número de casos por centro, estas cifras no permiten atribuir efectos institucionales, pero justifican futuras muestras amplias y análisis multinivel.

Tabla 4

Cambios en indicadores del proceso de resolución.

Grupo	Errores pre	Errores post	Red. %	Multi. %	Verif. %	Transf. %
Control	52,32	40,60	22,41	8,00	32,00	8,00
Exp.	51,51	24,18	53,05	64,00	76,00	72,00

Nota. El grupo experimental redujo 53,05 % de los errores y alcanzó 76 % de verificación espontánea.

Los indicadores de proceso explican la ventaja final. El grupo experimental redujo 53,05 % de los errores, frente a 22,41 % en el control, y obtuvo mayores porcentajes de resolución multietapa, verificación espontánea y transferencia. El 76 % comprobó sus respuestas sin una orden adicional y el 72 % resolvió contextos nuevos. Estos datos sugieren una modificación de hábitos estratégicos, no solo un aumento de respuestas correctas en ejercicios familiares.

Tabla 5

Distribución del desempeño posterior por niveles de dominio.

Nivel	Control f	Control %	Exp. f	Exp. %	Cambio
Inicial (< 50)	1	4,00	0	0,00	-1
Básico (50-69)	22	88,00	7	28,00	-15
Satisfactorio (70-84)	2	8,00	14	56,00	12
Avanzado (≥ 85)	0	0,00	4	16,00	4

Nota. El 72 % del grupo experimental se ubicó en niveles satisfactorio o avanzado.

La distribución por niveles muestra un desplazamiento educativo relevante: 72 % del grupo experimental alcanzó los niveles satisfactorio o avanzado, frente a 8 % del control. Quince estudiantes salieron del nivel básico y cuatro llegaron al avanzado. Aun así, siete participantes permanecieron en el nivel básico, por lo que la media favorable no elimina la necesidad de apoyo diferenciado. Los resultados deben utilizarse para planificar la siguiente etapa de enseñanza y no para retirar ayudas de forma general.

Tabla 6

Relación entre indicadores de interacción y ganancia en el grupo experimental.

Indicador	r de Pearson	p	Magnitud	Decisión
Calidad de interacción	0,316	0,124	Positiva baja	No significativa
Uso del feedback	0,455	0,022	Positiva moderada	Significativa
Autonomía	0,127	0,545	Positiva muy baja	No significativa
Persistencia	0,141	0,502	Positiva muy baja	No significativa

Nota. El uso de la retroalimentación mostró la única asociación significativa, $r = 0,455$.

El uso de la retroalimentación se relacionó de manera moderada con la ganancia, $r = 0,455$, $p = 0,022$. En cambio, frecuencia de interacción, autonomía y persistencia no mostraron asociaciones significativas. El hallazgo indica que aprender depende menos de consultar muchas veces que de convertir una orientación en una decisión matemática.

Aunque la correlación no prueba causalidad, respalda actividades en las que el estudiante explica qué cambió y demuestra la corrección del nuevo procedimiento.

Tabla 7

Prueba t de Student para diferencias entre grupos.

Variable	t	gl	p	Levene p	Criterio
Global postest	7,635	48	< 0,001	0,078	Igual
Ganancia global	13,197	37,779	< 0,001	0,015	Welch
Comprensión	5,357	48	< 0,001	0,111	Igual
Selección	3,123	48	0,003	0,207	Igual
Secuenciación	4,761	48	< 0,001	0,095	Igual
Exactitud	4,675	48	< 0,001	0,181	Igual
Verificación	4,215	48	< 0,001	0,072	Igual

Nota. La ganancia global presentó el contraste más alto; se aplicó Welch por desigualdad de varianzas.

La prueba t confirmó diferencias significativas en el puntaje global, la ganancia y cada destreza. En la ganancia se aplicó la corrección de Welch porque Levene indicó varianzas distintas; esta decisión protege la validez del contraste. Selección de operaciones presentó el menor valor t, coherente con su diferencia descriptiva, pero mantuvo $p = 0,003$. La convergencia de los contrastes apoya la hipótesis de un mejor desempeño experimental, sin confundir significación estadística con relevancia educativa.

Tabla 8

Tamaño del efecto de la intervención mediante d de Cohen.

Resultado	d	IC 95 % inferior	IC 95 % superior	Interpretación
Global postest	2,159	1,457	2,862	Muy grande
Ganancia	3,733	2,803	4,663	Muy grande
Comprensión	1,515	0,883	2,147	Muy grande
Selección	0,883	0,301	1,465	Grande
Secuenciación	1,347	0,730	1,963	Muy grande
Exactitud	1,322	0,708	1,937	Muy grande
Verificación	1,192	0,589	1,796	Grande

Nota. La d global fue 2,159 y la d de la ganancia alcanzó 3,733.

Los tamaños del efecto fueron grandes o muy grandes. La d global alcanzó 2,159 y la de la ganancia 3,733; selección de operaciones presentó la magnitud menor, aunque todavía grande. Los intervalos permanecieron por encima de cero, pero fueron amplios por el tamaño muestral. Valores tan elevados requieren replicación, especialmente porque proceden de datos modelados. En una investigación real deben informarse pérdidas, fidelidad, equivalencia docente y análisis de sensibilidad antes de generalizar el efecto.

Propuesta didáctica Aritmética IA

A partir de las dificultades diagnosticadas se organizó la propuesta AritméticaIA como una secuencia gradual y replicable. Cada actividad conserva una meta matemática

explícita, un momento de resolución autónoma, una interacción guiada con la herramienta y un cierre de explicación o comprobación. La IA se utiliza para formular preguntas y adaptar la dificultad, mientras que el docente decide cuándo activar las ayudas, revisa la calidad de las respuestas y promueve el diálogo entre estudiantes.

Tabla 9

Secuencia de actividades, destrezas, contenidos, recursos y objetivos.

Actividad	Destreza	Contenido	Mi n	Recursos	Objetivo
Diagnóstico	Comprensión	Datos e incógnita	80	Fichas e IA	Identificar datos.
Operaciones	Selección	Estructuras aritméticas	80	Tarjetas	Justificar operación.
Ruta	Secuenciación	Problemas de 2–3 pasos	80	Organizador	Ordenar acciones.
Laboratorio	Exactitud	Cálculo y estimación	80	Banco de errores	Corregir errores.
Revisión	Verificación	Inversa y estimación	80	Rúbrica	Comprobar solución.
Contextos	Transferencia	Dinero, tiempo y medida	80	Casos locales	Aplicar lo aprendido.
Desafío	Argumentación	Estrategias	80	Roles y registro	Defender solución.
Cierre	Autocontrol	Autoevaluación	80	Portafolio	Reconocer avances.

Nota. Las ocho actividades totalizan 640 minutos y articulan intento autónomo, apoyo gradual y cierre metacognitivo.

La propuesta se sometió a una validación de contenido modelada con diez expertos en didáctica de la Matemática, tecnología educativa, evaluación y educación básica. Se valoraron pertinencia, coherencia, claridad, viabilidad, accesibilidad, protección de datos y adecuación a la edad. Las observaciones llevaron a limitar pistas consecutivas, incluir una alternativa impresa, añadir explicación oral y exigir verificación al cierre. La valoración fue de alta pertinencia; no obstante, antes de una publicación esta validación debe ejecutarse y documentarse con expertos reales.

Discusión

Los resultados modelados respaldan la idea de que la IA puede contribuir a la resolución de problemas aritméticos cuando se integra en una arquitectura pedagógica explícita. La ventaja del grupo experimental no se restringió al puntaje total, sino que apareció en comprensión, selección, secuenciación, cálculo y verificación. Este patrón coincide con Liu et al. (2025), quienes mostraron que un entorno generativo diseñado para problemas verbales puede mejorar desempeño e interacción, y con Khazanchi et al. (2025), que identificaron avances matemáticos mediante un sistema adaptativo en un contexto rural. Asimismo, Qurohman (2024) relacionó el apoyo inteligente con el fortalecimiento de la

resolución, mientras que Panqueban y Huincahue (2024) encontraron que la personalización, la tutoría y la retroalimentación son funciones recurrentes en la literatura. La coincidencia entre fuentes sugiere que el beneficio surge cuando la tecnología interviene sobre operaciones cognitivas concretas y no cuando se usa como repositorio de respuestas.

La selección de operaciones presentó el efecto más bajo, aunque todavía grande. Este resultado es razonable porque reconocer la estructura de un problema requiere construir relaciones semánticas y no únicamente responder a una pista superficial. Un estudiante puede comprender qué ocurre y aun dudar entre sumar, multiplicar o combinar operaciones. Por ello, el apoyo de IA debe incluir contrastes entre problemas parecidos con estructuras diferentes y solicitar justificaciones. Lo (2023) advierte que las herramientas generativas pueden ofrecer explicaciones útiles, pero también respuestas plausibles que ocultan errores; Kasneci et al. (2023) señalan el mismo riesgo y recomiendan desarrollar verificación crítica. Grassini (2023) sostiene que la educación debe orientar el uso de IA hacia pensamiento y creatividad, no hacia la delegación mecánica. La menor magnitud en esta destreza indica un límite pedagógico: ninguna pista sustituye la construcción progresiva de conceptos aritméticos.

El avance en secuenciación y exactitud puede explicarse por la externalización del procedimiento. La herramienta pedía anticipar cada paso, comparar el resultado parcial con una estimación y corregir el punto exacto donde aparecía la inconsistencia. Esta dinámica se aproxima a la retroalimentación formativa descrita por Bower et al. (2024), quienes proponen rediseñar tareas para promover razonamiento verificable en la era generativa. Chan y Hu (2023) reportan que los estudiantes perciben valor en la disponibilidad de explicaciones personalizadas, mientras que Chiu et al. (2023) destacan el potencial de la IA para apoyar enseñanza, aprendizaje y evaluación. No obstante, el efecto depende de que la ayuda sea contingente. Una explicación extensa ofrecida antes del intento inicial puede reducir la carga productiva; una pregunta breve después de un error puede orientar sin apropiarse de la solución.

La verificación y argumentación alcanzó el porcentaje experimental más alto y mostró un efecto muy grande. Este hallazgo sugiere que la IA puede ser particularmente útil para activar una fase que suele omitirse cuando el estudiante considera terminado el ejercicio al obtener un número. El sistema solicitó comprobar por operación inversa, estimar el orden de magnitud y explicar si la respuesta tenía sentido en el contexto. Tlili et al. (2023)

sostienen que las herramientas conversacionales obligan a replantear evaluación y responsabilidad, y Farrokhnia et al. (2024) señalan que sus fortalezas se aprovechan mejor cuando se contrarrestan riesgos de inexactitud y dependencia. Rahman y Watanobe (2023) también subrayan que la evaluación en escenarios con IA debe valorar proceso y autenticidad. Así, la posibilidad de error del propio sistema se convirtió en un recurso didáctico: verificar dejó de ser un trámite y pasó a ser una condición para confiar en la respuesta.

La correlación entre uso de la retroalimentación y ganancia de aprendizaje aporta una explicación adicional. No fue la interacción en abstracto la que se relacionó significativamente con el progreso, sino la capacidad para incorporar una orientación al siguiente paso. Esta distinción coincide con la perspectiva de Chiu (2024), para quien la adopción educativa depende de expectativas, apoyo y diseño de la experiencia. Crompton et al. (2024) resaltan que el profesorado necesita preparación para transformar la información generada por IA en decisiones pedagógicas. Korkmaz Güler et al. (2024) muestran que las actitudes docentes y las condiciones de implementación influyen en el uso. En consecuencia, aumentar el número de consultas no constituye un indicador suficiente de calidad; es preferible registrar si el estudiante revisa, explica y transfiere lo recibido.

La equidad constituye una condición, no un complemento. La CEPAL señala que las brechas de conectividad y competencias digitales en América Latina y el Caribe afectan la posibilidad de beneficiarse de tecnologías educativas (Herrera et al., 2025). La UNESCO (2023) insiste en que la tecnología debe ser pertinente, accesible y respaldada por evidencia, y Miao y Holmes (2023) proponen protección de datos, adecuación por edad y supervisión humana. Estas orientaciones son decisivas para Educación General Básica Media. Una intervención puede producir una media favorable y, al mismo tiempo, excluir a estudiantes con menor acceso, dificultades lectoras o necesidades específicas. Por eso, las sesiones deben prever modalidades de baja conectividad, lenguaje claro, apoyos visuales y evaluación de efectos distributivos, no solo promedios.

Los tamaños de efecto fueron muy grandes, circunstancia que debe interpretarse con prudencia. Las intervenciones intensivas, las muestras pequeñas y la estrecha alineación entre enseñanza e instrumento pueden elevar la estimación. Bond et al. (2024) alertan sobre limitaciones de rigor en parte de la literatura, y Zhai et al. (2024) observan que los estudios de IA generativa todavía incluyen numerosos diseños exploratorios. Nguyen y

Truong (2025) solicitan mayor evidencia empírica en educación matemática y análisis de habilidades específicas. En un estudio definitivo sería necesario registrar pérdidas, controlar el efecto docente, verificar supuestos, usar medidas estandarizadas adicionales y, cuando la muestra lo permita, modelar la estructura jerárquica de estudiantes dentro de centros. La *d* de Cohen debe leerse junto con intervalos, medidas descriptivas y relevancia curricular.

En síntesis, los hallazgos modelados apoyan una integración complementaria. La IA fue útil como andamiaje que ofrece disponibilidad, variación y retroalimentación; el docente mantuvo la responsabilidad de seleccionar contenido, interpretar errores y cuidar el clima de aprendizaje; y el estudiante conservó la obligación de resolver y verificar. Zhao et al. (2025) relacionan el uso bien diseñado de IA con oportunidades para el pensamiento de orden superior, mientras que Tuomi (2024) recuerda que sus efectos exceden la herramienta y alcanzan el currículo y las instituciones. Las futuras investigaciones deberían comparar tipos de ayuda, analizar retención a mediano plazo, estudiar estudiantes con perfiles diversos y documentar costos y condiciones de sostenibilidad. También deben incorporar bases reales, protocolos preregistrados y materiales abiertos que permitan replicar los resultados.

Conclusiones

La secuencia AritméticaIA mostró, en el modelo analítico desarrollado, una ventaja consistente frente a la enseñanza habitual en la resolución de problemas aritméticos. La equivalencia del pretest y la convergencia entre puntajes posteriores, ganancias, niveles de desempeño, indicadores de proceso, prueba *t* y tamaño del efecto sugieren que el andamiaje adaptativo puede fortalecer componentes que suelen enseñarse de manera fragmentada. Las mejoras más visibles se concentraron en comprensión, secuenciación, exactitud y verificación, mientras que la selección de operaciones continuó siendo la destreza que demanda mayor profundización conceptual. La contribución principal consiste en proponer una forma de integración en la que la IA no entrega respuestas como producto terminado, sino que organiza preguntas graduadas, promueve el análisis del error y exige una comprobación independiente.

Desde una perspectiva científica y educativa, el estudio subraya que el efecto de la IA depende de la calidad de la secuencia, la mediación docente, el uso productivo de la retroalimentación y las condiciones de equidad. La tecnología no reemplaza la enseñanza de conceptos ni la interacción humana; amplía las oportunidades de práctica y hace

posible una ayuda más inmediata cuando se utiliza con criterios claros. La propuesta ofrece una base para investigaciones posteriores con datos empíricos, muestras mayores, seguimiento longitudinal y control de la fidelidad de implementación. Antes de una publicación, las estimaciones modeladas deben sustituirse por resultados obtenidos en campo y someterse a verificación independiente. Con esa condición, AritméticaIA puede aportar un marco replicable para estudiar cómo la inteligencia artificial contribuye a una Matemática escolar más reflexiva, argumentada y accesible.

Referencias

- Bond, M., Khosravi, H., De Laat, M., Bergdahl, N., Negrea, V., Oxley, E., Pham, P., Chong, S. W., & Siemens, G. (2024). A meta systematic review of artificial intelligence in higher education: A call for increased ethics, collaboration, and rigour. *International Journal of Educational Technology in Higher Education*, 21, Article 4. <https://doi.org/10.1186/s41239-023-00436-z>
- Bower, M., Torrington, J., Lai, J. W. M., Petocz, P., & Alfano, M. (2024). How should we change teaching and assessment in response to increasingly powerful generative artificial intelligence? Outcomes of the ChatGPT teacher survey. *Education and Information Technologies*, 29, 15403-15439. <https://doi.org/10.1007/s10639-023-12405-0>
- 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, Article 43. <https://doi.org/10.1186/s41239-023-00411-8>
- Chen, X., Zou, D., Xie, H., Cheng, G., & Liu, C. (2022). Two decades of artificial intelligence in education: Contributors, collaborations, research topics, challenges, and future directions. *Educational Technology & Society*, 25(1), 28-47. <https://doi.org/10.2307/48647028>
- Chiu, T. K. F. (2024). The impact of generative AI 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>
- Chiu, T. K. F., Xia, Q., Zhou, X., Chai, C. S., & Cheng, M. (2023). Systematic literature review on opportunities, challenges, and future research recommendations of

- artificial intelligence in education. *Computers and Education: Artificial Intelligence*, 4, 100118. <https://doi.org/10.1016/j.caeai.2022.100118>
- Crompton, H., & Burke, D. (2023). Artificial intelligence in higher education: The state of the field. *International Journal of Educational Technology in Higher Education*, 20, Article 22. <https://doi.org/10.1186/s41239-023-00392-8>
- Crompton, H., Jones, M. V., & Burke, D. (2024). Affordances and challenges of artificial intelligence in K-12 education: A systematic review. *Journal of Research on Technology in Education*, 56(3), 248-268. <https://doi.org/10.1080/15391523.2022.2121344>
- 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>
- Grassini, S. (2023). Shaping the future of education: Exploring the potential and consequences of AI and ChatGPT in educational settings. *Education Sciences*, 13(7), 692. <https://doi.org/10.3390/educsci13070692>
- Herrera, D., Huepe, M., & Trucco, D. (2025). Educación y desarrollo de competencias digitales en América Latina y el Caribe. *Comisión Económica para América Latina y el Caribe*. LC/TS.2025/3.
- 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>
- Khazanchi, R., Di Mitri, D., & Drachsler, H. (2025). The effect of AI-based systems on mathematics achievement in rural context: A quantitative study. *Journal of Computer Assisted Learning*, 41(1), e13098. <https://doi.org/10.1111/jcal.13098>
- Liu, J., Sun, D., Sun, J., Wang, J., & Yu, P. L. H. (2025). Designing a generative AI enabled learning environment for mathematics word problem solving in primary schools: Learning performance, attitudes and interaction. *Computers and Education: Artificial Intelligence*, 9, 100438. <https://doi.org/10.1016/j.caeai.2025.100438>

- 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>
- Matzakos, N., Doukakis, S., & Moundridou, M. (2023). Learning mathematics with large language models. *International Journal of Emerging Technologies in Learning*, 18(20), 51-71. <https://doi.org/10.3991/ijet.v18i20.42979>
- Miao, F., & Holmes, W. (2023). Guidance for generative AI in education and research. UNESCO.
- Ministerio de Educación del Ecuador. (2021). Currículo priorizado con énfasis en competencias comunicacionales, matemáticas, digitales y socioemocionales: Educación General Básica Media. Autor.
- Mohamed, M. Z. B., Hidayat, R., Suhaizi, N. N. B., Sabri, N. B. M., Mahmud, M. K. H. B., & Baharuddin, S. N. B. (2022). Artificial intelligence in mathematics education: A systematic literature review. *International Electronic Journal of Mathematics Education*, 17(3), em0694. <https://doi.org/10.29333/iejme/12132>
- Nguyen, T. N., & Truong, H. T. (2025). Trends of artificial intelligence in mathematics education: A systematic review. *Eurasia Journal of Mathematics, Science and Technology Education*, 21(3), em2580. <https://doi.org/10.29333/ejmste/16124>
- Panqueban, D., & Huincahue, J. (2024). Artificial intelligence in mathematics education: A systematic review. *Uniciencia*, 38(1), 1-17. <https://doi.org/10.15359/ru.38-1.20>
- Qurohman, M. T. (2024). Artificial intelligence-based learning to improve mathematical problem-solving skills. *International Journal of Trends in Mathematics Education Research*, 7(4), 384-393. <https://doi.org/10.33122/ijtmer.v7i4.358>
- Rahman, M. M., & Watanobe, Y. (2023). ChatGPT for education and research: Opportunities, threats, and strategies. *Applied Sciences*, 13(9), 5783. <https://doi.org/10.3390/app13095783>
- Tlili, A., Shehata, B., Adarkwah, M. A., Bozkurt, A., Hickey, D. T., Huang, R., & Agyemang, B. (2023). What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. *Smart Learning Environments*, 10, Article 15. <https://doi.org/10.1186/s40561-023-00237-x>
- Tuomi, I. (2024). Beyond mastery: Toward a broader understanding of AI in education. *International Journal of Artificial Intelligence in Education*, 34, 20-45. <https://doi.org/10.1007/s40593-023-00343-4>

- UNESCO. (2023). Global education monitoring report 2023: Technology in education: A tool on whose terms? UNESCO. <https://doi.org/10.54676/UZQV8501>
- Zhai, X., 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, Article 28. <https://doi.org/10.1186/s40561-024-00316-7>
- Zhao, Y., Pinto Llorente, A. M., & Sánchez Gómez, M. C. (2025). The impact of generative artificial intelligence on higher-order thinking: A systematic review. *Journal of Intelligence*, 13(12), 160. <https://doi.org/10.3390/jintelligence13120160>

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	N/A
Nota	El artículo no es producto de una publicación anterior. Los análisis estadísticos incluidos corresponden a un modelo académico y requieren sustitución por datos empíricos antes de su publicación.