Revista Multidisciplinaria Perspectivas Investigativas
Multidisciplinary Journal Investigative Perspectives
Vol. 6(4), 35-43, 2026
https://doi.org/10.62574/rmpi.v6i4.526

Estrategia IOBAS con chatbots de inteligencia artificial y aprendizaje de las matemáticas en básica secundaria

IOBAS strategy with artificial intelligence chatbots and mathematics learning in lower secondary education

Johann Verney Méndez-Gamba
johann-v.mendez-g@up.ac.pa
Universidad de Panamá, Panamá, Provincia de Panamá, Panamá
https://orcid.org/0009-0008-6332-4850

RESUMEN

El estudio analiza la incidencia de chatbots de inteligencia artificial integrados a la estrategia didáctica IOBAS en el aprendizaje de las matemáticas de estudiantes de básica secundaria de una institución técnica de Samacá, Colombia. Se empleó un diseño mixto secuencial explicativo, con un componente cuasiexperimental pretest-postest (30 estudiantes en el grupo experimental y 34 en el control), una rúbrica de resolución de problemas, una encuesta de percepción, observaciones de aula y artefactos de aprendizaje. La ganancia global fue de 8,60 puntos en el grupo experimental y de 5,46 en el control (g = 1,21), y la diferencia se mantuvo tras ajustar por el pretest (η^2p = ,251), con mayor consistencia en las dimensiones conceptual e interpretativa que en la procedimental. Se concluye que la condición integrada se asoció con un mejor desempeño, sin que el efecto pueda atribuirse a los chatbots de forma aislada.

Descriptores: inteligencia artificial; matemáticas; métodos de enseñanza. (Fuente: Tesauro UNESCO).

ABSTRACT

The study analyzes the incidence of artificial intelligence chatbots integrated into the IOBAS teaching strategy on mathematics learning among lower secondary students of a technical school in Samacá, Colombia. A sequential explanatory mixed design was used, with a quasi-experimental pretest-posttest component (30 students in the experimental group and 34 in the control group), a problem-solving rubric, a perception survey, classroom observations and learning artifacts. The overall gain was 8.60 points in the experimental group and 5.46 in the control group (g = 1.21), and the difference remained after adjusting for the pretest (η^2p = .251), with greater consistency in the conceptual and interpretative dimensions than in the procedural one. It is concluded that the integrated condition was associated with better performance, although the effect cannot be attributed to the chatbots in isolation.

Descriptors: artificial intelligence; mathematics; teaching methods. (Source: UNESCO Thesaurus).

Recibido: 05/10/2026. Revisado: 12/10/2026. Aprobado: 18/10/2026. Publicado: 05/11/2026.
Sección artículos de investigación

INTRODUCCIÓN

El aprendizaje de las matemáticas en la educación básica secundaria depende en buena medida de la capacidad de los estudiantes para resolver problemas, esto es, para comprender una situación, reconocer la información pertinente, elegir una estrategia y justificar el resultado. En este orden de ideas, Benítez-Pérez, et al., (2023) vinculan la comprensión del enunciado con los procesos inductivos y deductivos que intervienen en la resolución, mientras que Pérez-González, et al., (2021) proponen una estrategia didáctica orientada a dirigir el proceso de enseñanza y aprendizaje de la resolución de problemas matemáticos. Desde esta perspectiva, las dificultades no se limitan al cálculo, sino que se extienden a las fases iniciales de lectura, organización y planificación, lo que justifica el interés por estrategias que hagan visible cada momento del proceso.

Una de ellas es IOBAS, propuesta de origen peruano para la resolución de problemas, que estructura el trabajo en cinco momentos: identificación del problema, organización de saberes, búsqueda y elección de estrategias, aplicación, y síntesis o socialización de los resultados. Su fundamento en el aprendizaje basado en problemas la hace compatible con el uso de herramientas que ofrezcan apoyo conversacional, de manera que los chatbots de inteligencia artificial pueden actuar como andamiaje en cada fase. Con relación al andamiaje, van de Pol, et al., (2019) analizaron cómo los estudiantes aprovechan el apoyo docente en el trabajo en pequeños grupos, y Sweller, et al., (2019) sostienen que la arquitectura cognitiva humana obliga a estructurar la información que se presenta a quienes aprenden un contenido nuevo.

Paralelamente, la investigación sobre inteligencia artificial generativa en educación matemática crece con rapidez. Debets, et al., (2025) revisaron sistemáticamente los objetivos, la base tecnológica y teórica, los criterios de evaluación y los impactos de los chatbots en educación, Pepin, et al., (2025) mapearon los estudios sobre ChatGPT en la enseñanza de las matemáticas, y Rutherford, et al., (2025) exploraron el uso docente de esta herramienta para diseñar lecciones con apoyo motivacional. Por su parte, el metaanálisis de Wang & Fan (2025) reporta efectos favorables de ChatGPT sobre el rendimiento, la percepción y el pensamiento de orden superior, con diferencias según las características de la implementación. No obstante, la retroalimentación automatizada no garantiza por sí misma un mejor aprendizaje: Kluger & DeNisi (1996) mostraron que sus efectos son variables y que en una parte de los casos disminuyen el desempeño, y Wisniewski, et al., (2020) encontraron que resulta más eficaz cuando aporta información sobre la tarea y el proceso, en línea con la revisión de Shute (2008) sobre la retroalimentación formativa.

En este contexto, la evidencia que combine una estrategia estructurada de resolución de problemas con chatbots, que contraste grupos y que a la vez examine los procesos de aula, resulta escasa en las fuentes revisadas, y los estudios tienden a describir potencialidades antes que a comparar condiciones. Con base en lo expuesto, el presente artículo, derivado de una tesis doctoral concluida, se propone analizar la incidencia de los chatbots de inteligencia artificial, integrados a la estrategia didáctica IOBAS, en el aprendizaje de las matemáticas de estudiantes de básica secundaria de la Institución Educativa Técnica Nacionalizada de Samacá, Colombia.

MÉTODO

La investigación adoptó el paradigma pragmático y un diseño mixto secuencial explicativo (CUAN → cual), en el que la fase cuantitativa precedió y orientó a la cualitativa (Creswell & Plano Clark, 2018). El componente cuantitativo fue cuasiexperimental, con medición pretest y postest en un grupo experimental, que trabajó con IOBAS mediada por chatbots, y un grupo control, que recibió enseñanza tradicional equivalente en tiempo y objetivos. Los grupos correspondieron a cursos intactos y no se asignaron de manera aleatoria. El componente cualitativo examinó los procesos de aula mediante observaciones y artefactos de aprendizaje, y la integración se realizó con una matriz de despliegue conjunto y meta-inferencias.

El estudio se desarrolló en la Institución Educativa Técnica Nacionalizada de Samacá, Colombia, con estudiantes de básica secundaria. La prueba de desempeño contó con 30 estudiantes del grupo experimental con registros completos en pretest y postest, y con 34 del grupo control en el postest, de los cuales 33 dispusieron también de pretest. Un registro del control sin

puntuaciones en el pretest se excluyó únicamente de los análisis que requerían la medición basal. No se imputaron datos ausentes. En cuanto a los aspectos éticos, se contó con el consentimiento de los acudientes y el asentimiento de los estudiantes, se garantizó la participación voluntaria y se seudonimizaron los datos y las trazas de interacción con los chatbots.

En el grupo experimental, los estudiantes resolvieron situaciones problema siguiendo las cinco fases de IOBAS y consultaron chatbots como apoyo, con normas de uso que priorizaban la verificación docente de las respuestas y la retirada gradual de las ayudas. El grupo control trabajó los mismos contenidos con la metodología habitual. Las actividades se centraron principalmente en la función lineal y la recta en el plano cartesiano, con atención a la pendiente, el punto de corte y la representación gráfica y algebraica.

Los instrumentos cuantitativos fueron una prueba de desempeño aplicada antes y después de la intervención, cuyo puntaje global resulta de sumar las dimensiones conceptual, procedimental e interpretativa; una rúbrica analítica de resolución de problemas IOBAS, con cinco criterios (identificación del problema, organización de saberes, procedimiento y aplicación de estrategias, verificación y justificación) puntuados de 1 a 4, de modo que el puntaje global varía entre 5 y 20 y se clasifica en niveles muy bajo (5 a 8), bajo (9 a 12), medio (13 a 16) y alto (17 a 20); y una encuesta de percepción de 11 ítems en escala de 1 a 5, aplicada al grupo experimental. Las fuentes cualitativas fueron una guía de observación del discurso y la práctica en aula, que consideró el marco ICAP, las normas sociomatemáticas, el andamiaje y la retroalimentación, los errores y la interacción entre estudiante, chatbot y docente, y el análisis de los artefactos de aprendizaje con sus trazas de interacción.

Los datos se describieron mediante media, desviación estándar y mediana. Las comparaciones entre grupos se realizaron con la prueba t de Welch y, de manera complementaria, con la prueba U de Mann-Whitney, previa verificación de la normalidad con Shapiro-Wilk. Los cambios dentro de cada grupo se contrastaron con la prueba t para muestras relacionadas y con Wilcoxon, y se estimaron el d de Cohen para medidas relacionadas (dz) y la g de Hedges para la diferencia de ganancias. Para controlar el nivel inicial se ajustó un análisis de covarianza (ANCOVA) con el postest como variable dependiente, el grupo como factor y el pretest como covariable, tras comprobar la homogeneidad de las pendientes, y se informó el eta cuadrado parcial (η^2p). La consistencia interna de la encuesta se estimó con el alfa de Cronbach y su relación con las ganancias individuales con el coeficiente ρ de Spearman. El nivel de significación fue $\alpha = ,05$. El análisis cualitativo comparó las evidencias de las observaciones y los artefactos a partir de las categorías de los instrumentos, sin introducir categorías ajenas a los registros.

RESULTADOS

Los resultados se presentan en el orden de la secuencia analítica: el desempeño inicial, el desempeño final y las ganancias, el análisis ajustado por el pretest, la rúbrica IOBAS, las percepciones estudiantiles y los hallazgos cualitativos.

Desempeño inicial y final

En el pretest, ambos grupos partieron de un nivel bajo. El grupo experimental obtuvo una media global de 1,07 (DE = 0,89) y el control de 1,62 (DE = 1,50). La diferencia no alcanzó significación mediante la prueba t de Welch, $t(52,90) = -1,81$; $p = ,075$, ni mediante Mann-Whitney, $U = 393,0$; $p = ,162$. En las dimensiones conceptual (0,84 frente a 0,90) y procedimental (0,10 frente a 0,39) tampoco se encontraron diferencias significativas. En la dimensión interpretativa, en cambio, la prueba de Welch mostró una diferencia entre las medias (0,13 frente a 0,33), $t(54,48) = -2,13$; $p = ,038$, que Mann-Whitney no confirmó, $U = 376,0$; $p = ,072$. Descriptivamente, el grupo control presentó medias superiores en el puntaje global y en las tres dimensiones.

En el postest, el grupo experimental alcanzó una media global de 9,67 (DE = 2,49) y el control de 6,94 (DE = 2,94). La diferencia fue estadísticamente significativa, $t(61,90) = 4,02$; $p < ,001$, y se confirmó con Mann-Whitney, $U = 782,0$; $p < ,001$. Por dimensiones, la conceptual fue de 3,36 (DE = 0,87) frente a 2,01 (DE = 0,92), $t(61,70) = 6,07$; $p < ,001$, y la interpretativa de 2,88 (DE = 1,14) frente a 2,05 (DE = 1,29), $t(62,00) = 2,71$; $p = ,009$. En la dimensión procedimental, las medias fueron 3,44 (DE = 1,26) y 2,89 (DE = 1,33), diferencia que no alcanzó significación, $t(61,67) = 1,70$; $p = ,095$; $U = 651,0$; $p = ,059$.

Cambios pretest-postest y ganancias

Ambos grupos mejoraron entre las dos mediciones. En el grupo experimental, el puntaje global aumentó 8,60 puntos (DE = 2,36; IC 95 % [7,72; 9,49]), t(29) = 19,93; p < ,001; dz = 3,64. En el control, el aumento fue de 5,46 puntos (DE = 2,75; IC 95 % [4,48; 6,43]), t(32) = 11,41; p < ,001; dz = 1,99; en este grupo se registraron algunos cambios individuales negativos, con un mínimo de −0,88. Las pruebas de Wilcoxon coincidieron con estas decisiones. Los valores se resumen en la Tabla 1.

Tabla 1. *Puntaje global de la prueba de desempeño en el pretest, el postest y la ganancia por grupo*.

Grupo	n pre	Pretest M (DE)	n post	Postest M (DE)	n Δ	Ganancia Δ M (DE)
Experimental	30	1,07 (0,89)	30	9,67 (2,49)	30	8,60 (2,36)
Control	33	1,62 (1,50)	34	6,94 (2,94)	33	5,46 (2,75)

Nota: elaboración propia. M = media; DE = desviación estándar; Δ = postest − pretest. La ganancia del control se calculó con los 33 casos con ambas mediciones, cuya media en el postest fue de 7,08 (DE = 2,87).

La comparación directa de las ganancias mostró una diferencia de 3,15 puntos a favor del grupo experimental (IC 95 % [1,86; 4,44]), t(60,83) = 4,88; p < ,001; g = 1,21, resultado respaldado por Mann-Whitney, U = 801,5; p < ,001. Como se observa en la Tabla 2, la diferencia fue significativa en las tres dimensiones, con la mayor magnitud en la conceptual y la menor en la procedimental.

Tabla 2. *Comparación de las ganancias pretest-postest entre grupos, por dimensión*.

Variable	Experimental Δ M (DE)	Control Δ M (DE)	Diferencia	IC 95 %	t	p	g
Puntaje global	8,60 (2,36)	5,46 (2,75)	3,15	[1,86; 4,44]	4,88	< ,001	1,21
Conceptual	2,52 (0,97)	1,14 (0,99)	1,38	[0,88; 1,87]	5,58	< ,001	1,39
Procedimental	3,34 (1,19)	2,54 (1,44)	0,81	[0,14; 1,47]	2,42	,019	0,60
Interpretativa	2,75 (1,13)	1,78 (1,22)	0,96	[0,37; 1,56]	3,24	,002	0,81

Nota: elaboración propia. Δ = postest − pretest; M = media; DE = desviación estándar; IC 95 % = intervalo de confianza de la diferencia; t = t de Welch (gl entre 60,5 y 61,0); g = g de Hedges. n = 30 en el grupo experimental y 33 en el control.

Análisis ajustado por el desempeño basal

Las pendientes de regresión resultaron homogéneas en el puntaje global, F(1, 59) = 0,13; p = ,715, y en las tres dimensiones, por lo que se estimaron los modelos sin el término de interacción. En el puntaje global, la diferencia persistió tras controlar el pretest, F(1, 60) = 20,14; p < ,001; η^2p = ,251, con medias ajustadas de 9,88 para el grupo experimental y de 6,90 para el control (diferencia de 2,98; IC 95 % [1,65; 4,31]). El modelo explicó el 27,8 % de la variabilidad del postest. La Tabla 3 presenta los resultados por dimensión.

Tabla 3. *Resultados del ANCOVA sobre el postest, con el pretest como covariable*.

Variable	M ajustada experimental	M ajustada control	Diferencia	IC 95 %	F(1, 60)	p	η^2p
Puntaje global	9,88	6,90	2,98	[1,65; 4,31]	20,14	< ,001	,251
Conceptual	3,37	2,03	1,33	[0,89; 1,77]	36,42	< ,001	,378
Procedimental	3,50	2,87	0,63	[−0,03; 1,29]	3,65	,061	,057
Interpretativa	2,95	2,04	0,91	[0,29; 1,53]	8,61	,005	,125

Nota: elaboración propia. M = media; IC 95 % = intervalo de confianza de la diferencia ajustada; η^2p = eta cuadrado parcial. n = 30 en el grupo experimental y 33 en el control.

En la dimensión procedimental, el efecto del grupo no alcanzó significación convencional (p = ,061), el intervalo de confianza incluyó el cero y los residuos se apartaron de la normalidad (W = ,950; p = ,012), de modo que este resultado difiere del obtenido al comparar las ganancias (p = ,019). En el modelo global y en los de las dimensiones conceptual e interpretativa, los residuos no mostraron desviaciones de la normalidad ni heterogeneidad de varianzas.

Desempeño en la rúbrica IOBAS

La rúbrica reunió 26 registros de actividades de aula, que no equivalen a 26 estudiantes independientes, pues algunos corresponden a estudiantes individuales y otros a parejas o grupos. Su puntaje global medio fue de 12,96 (DE = 3,97; mediana = 12,50), con valores entre 6

y 20. Cuatro registros (15,4 %) se ubicaron en el nivel muy bajo, nueve (34,6 %) en el bajo, seis (23,1 %) en el medio y siete (26,9 %) en el alto. Los criterios con menores medias fueron la identificación del problema y la organización de saberes, mientras que la verificación y la justificación presentaron las mayores (Tabla 4).

Los registros procedieron de dos fechas de aplicación, 15 del 20 de mayo de 2026 y 11 del 21 de mayo de 2026. El puntaje global medio fue de 10,20 (DE = 2,21) en la primera y de 16,73 (DE = 2,37) en la segunda, y la diferencia se presentó en los cinco criterios. Dado que las unidades observadas no fueron equivalentes entre fechas, esta comparación es solamente descriptiva y no constituye una medición pareada.

Tabla 4. *Puntuaciones de la rúbrica IOBAS por criterio y fecha de aplicación*.

Criterio	Total M (DE)	Mediana	20 de mayo M	21 de mayo M
Identificación del problema	2,12 (1,11)	2,00	1,40	3,09
Organización de saberes	2,31 (1,12)	2,00	1,60	3,27
Procedimiento y estrategias	2,62 (0,75)	3,00	2,20	3,18
Verificación	2,92 (0,80)	3,00	2,47	3,55
Justificación	3,00 (1,10)	3,00	2,53	3,64
Puntaje global (5 a 20)	12,96 (3,97)	12,50	10,20	16,73

Nota: elaboración propia. M = media; DE = desviación estándar. Cada criterio se puntuó de 1 a 4. El total corresponde a 26 registros, 15 del 20 de mayo y 11 del 21 de mayo de 2026, no equivalentes entre sí.

Percepciones estudiantiles

La encuesta reunió 25 formularios completos, sin valores perdidos, con una consistencia interna de α = ,819. Los códigos 16 y 33 aparecieron dos veces y uno de los códigos no tenía correspondencia en la base del grupo experimental, por lo que los formularios no se asumieron como 25 estudiantes distintos. La media conjunta de las 275 respuestas fue de 4,06; el 25,1 % correspondió a la categoría 5, el 57,5 % a la 4, el 16,0 % a la 3 y el 1,5 % a la 2, sin respuestas en la categoría 1, de modo que el 82,5 % se concentró en las categorías 4 y 5. Las medias por ítem oscilaron entre 3,68 y 4,32 y la mediana fue 4 en todos (Tabla 5).

Tabla 5. *Resultados descriptivos de la encuesta de percepción sobre IOBAS y el uso de chatbots*.

Ítem	Contenido abreviado	M	DE
1	Comprensión de los temas matemáticos mediante chatbots	4,16	0,75
2	Organización de conocimientos previos mediante IOBAS	3,88	0,78
3	Elección de estrategias mediante chatbots	4,12	0,60
4	Dinamismo del aprendizaje con chatbots	4,12	0,67
5	Seguridad para aplicar estrategias de resolución	3,96	0,68
6	Aprendizaje mediante la socialización de resultados	3,68	0,63
7	Utilidad y personalización de la retroalimentación del chatbot	4,04	0,61
8	Actitud positiva hacia el aprendizaje de las matemáticas	4,00	0,65
9	Confianza en el uso de herramientas tecnológicas	4,32	0,69
10	Percepción de mejora del aprendizaje mediante IOBAS	4,16	0,75
11	Percepción de mejora del aprendizaje mediante IA y chatbots	4,24	0,60

Nota: elaboración propia. M = media; DE = desviación estándar. Escala de 1 a 5; 25 formularios completos.

La confianza en el uso de herramientas tecnológicas (ítem 9) obtuvo la mayor media y la socialización de los resultados (ítem 6) la menor. Al vincular las respuestas con los resultados individuales se logró una correspondencia inequívoca en 22 participantes; en los dos códigos duplicados se promediaron las respuestas por ítem, y el uso alternativo de la primera o de la segunda respuesta no modificó la significación. La puntuación global de percepción tuvo una media de 4,08 (DE = 0,41). Su correlación con la ganancia total fue baja y no significativa (ρ = ,169; p = ,451), igual que con las ganancias conceptual (ρ = −,219; p = ,327), procedimental (ρ = ,271; p = ,222) e interpretativa (ρ = ,328; p = ,137).

Hallazgos cualitativos

La guía de observación contiene los registros de las sesiones 01, 02, 03 y 05; no se identificó el correspondiente a la sesión 04. Además, el docente y el observador figuran con el mismo nombre, aunque el instrumento preveía una observación no participante. En la primera sesión, realizada sin chatbot, predominó una participación pasiva, centrada en escuchar y tomar

apuntes. En la segunda se registraron intercambio de ideas sobre estrategias, uso de los chatbots como apoyo y un nivel predominante constructivo, con justificaciones todavía parciales y propuestas correctas junto a otras incorrectas. La sesión 03 mantuvo el nivel constructivo con mayor discusión entre pares y, en la sesión 05, el nivel predominante fue interactivo, con justificación de respuestas, valoración de estrategias y validación argumentada de resultados.

La apropiación de las fases de IOBAS fue gradual. En la primera sesión las cinco fases se valoraron con nivel 1, dado su carácter introductorio. En la sesión 02, identificación y búsqueda obtuvieron nivel 1 y organización y síntesis nivel 2, con incertidumbre sobre lo que debía hacerse en cada fase. En la sesión 03 las cinco fases alcanzaron nivel 2 y en la sesión 05 se registraron niveles 3 en identificación, búsqueda, aplicación y síntesis, y 4 en organización. Los artefactos mostraron variabilidad: algunos productos del 20 de mayo presentaron niveles bajos de identificación y organización y procedimientos sin una secuencia documentada, mientras que otros del 21 de mayo alcanzaron niveles 3 y 4, con organización completa de los datos, reconocimiento de la pendiente y del punto de corte, y argumentación consistente de las soluciones.

El uso de los chatbots fue heterogéneo. En un primer acercamiento, varios estudiantes introdujeron los enunciados para obtener la solución y, en un artefacto, esa consulta se clasificó con pertinencia baja, uso parcial de la respuesta, dependencia media y reelaboración baja. Otros grupos consultaron si el procedimiento para obtener la ecuación de una recta era correcto, preguntaron por el significado de la pendiente, por las rectas verticales, horizontales, paralelas y perpendiculares, o por la corrección de sus tablas de valores, con pertinencia alta, dependencia baja y reelaboración alta en este último caso. En la sesión 05 las preguntas fueron más puntuales y explicativas y disminuyeron las consultas para obtener soluciones directas, aunque algunos estudiantes continuaron recurriendo a la herramienta sin desarrollar antes el proceso.

Persistieron dificultades específicas. Se observaron problemas para asociar la pendiente con la tasa de cambio, distinguir la condición inicial y relacionarla con el punto de corte, manejar signos y unidades, contextualizar los resultados y traducir entre la representación gráfica y la ecuación. En la sesión 05, los errores conceptuales, procedimentales e interpretativos se clasificaron con frecuencia baja, mientras que los representacionales permanecieron en frecuencia media, principalmente por las escalas de los ejes. En cuanto a la mediación, las explicaciones directas del docente fueron frecuentes en la primera sesión y de frecuencia baja en la quinta, cuando adquirieron mayor presencia las preguntas orientadoras y la retroalimentación formativa.

DISCUSIÓN

Los resultados indican que la condición que integró IOBAS y chatbots se asoció con un desempeño matemático superior al de la enseñanza tradicional. Ambos grupos mejoraron, pero el experimental lo hizo en 3,15 puntos más ($g = 1,21$), y la diferencia se mantuvo tras controlar el pretest ($\eta^2p = ,251$), de modo que no depende únicamente de la ventaja o desventaja inicial. Esta dirección coincide con la del metaanálisis de Wang & Fan (2025), que reporta efectos favorables de ChatGPT sobre el rendimiento, y con la revisión de Debets, et al., (2025), que sitúa a los chatbots como un campo con impactos documentados. Sin embargo, la condición experimental combinó la estrategia IOBAS, el trabajo entre pares, la mediación docente y los chatbots, y el diseño cuasiexperimental no permite separar la contribución de cada componente, razón por la cual la diferencia debe leerse como una asociación con la configuración pedagógica completa y no como un efecto atribuible a la inteligencia artificial. A ello se suma que los grupos no eran necesariamente equivalentes al inicio, pues el control mostró medias superiores y una discrepancia entre pruebas en la dimensión interpretativa, y que la ausencia de significación en el pretest no equivale a una prueba de equivalencia.

El avance no fue homogéneo entre las dimensiones. Las evidencias fueron más consistentes en la conceptual ($\eta^2p = ,378$) y la interpretativa ($\eta^2p = ,125$), mientras que la procedimental dependió del análisis: la ganancia fue mayor en el grupo experimental ($p = ,019$), pero el ANCOVA no alcanzó significación ($p = ,061$) y los residuos se apartaron de la normalidad. Por tanto, no cabe afirmar un efecto procedimental concluyente. Los registros cualitativos son compatibles con esta cautela, porque disminuyeron los errores operativos, pero continuaron los fallos en signos, unidades y secuenciación de pasos. Una lectura plausible, no verificada en este diseño, es que el apoyo conversacional favorece la explicación y la interpretación de relaciones más que la

ejecución autónoma de procedimientos, aspecto que Sweller, et al., (2019) permiten relacionar con la necesidad de retirar de manera gradual las ayudas a medida que aumenta la experticia.

Las diferencias entre las fases de IOBAS aportan una información que no ofrece el resultado final de la prueba. En la rúbrica, la identificación del problema (2,12) y la organización de saberes (2,31) obtuvieron las menores medias, y los artefactos mostraron estudiantes que ejecutaban operaciones apropiadas sin explicitar el objetivo ni la organización de los datos. Este patrón concuerda con la importancia que Benítez-Pérez, et al., (2023) otorgan a la comprensión del enunciado y con la propuesta de Pérez-González, et al., (2021) de dirigir la resolución como un proceso, de modo que un resultado correcto no equivale a un proceso explícito. La mejora entre el 20 y el 21 de mayo (de 10,20 a 16,73) es compatible con una apropiación progresiva de la estrategia, aunque las unidades observadas no fueron equivalentes y no puede interpretarse como un cambio individual.

Respecto de los chatbots, los registros muestran funciones diferentes: obtener la solución, orientar el procedimiento, aclarar conceptos y verificar un trabajo propio. Esta heterogeneidad ofrece una explicación plausible, sin demostración causal, de por qué la percepción favorable (4,06 sobre 5) no se asoció con la ganancia individual (ρ = ,169; p = ,451), puesto que dos estudiantes pueden valorar igual la herramienta y utilizarla de forma muy distinta. Lo anterior concuerda con la literatura sobre retroalimentación, según la cual sus efectos varían y a veces perjudican el desempeño (Kluger & DeNisi, 1996), y su eficacia aumenta cuando orienta sobre la tarea y el proceso (Wisniewski, et al., 2020). En la misma línea, Pepin, et al., (2025) y Rutherford, et al., (2025) muestran que el valor de ChatGPT en educación matemática depende de cómo se incorpora a la enseñanza, y van de Pol, et al., (2019) destacan el papel del apoyo docente en el trabajo grupal, coherente con el cambio observado desde explicaciones directas hacia preguntas orientadoras. Además, los datos de la encuesta describen valoraciones y no una medida objetiva de aprendizaje, por lo que el 82,5 % de respuestas en las categorías 4 y 5 no puede utilizarse como evidencia de mejora.

La transición de una participación pasiva a niveles constructivos e interactivos puede leerse desde el marco ICAP, según el cual los modos interactivo y constructivo se asocian con mejores resultados que los modos pasivo y activo (Chi & Wylie, 2014). No obstante, las observaciones describen dinámicas de aula y no puntuaciones individuales, la sesión 04 no se registró y el observador estaba vinculado con la implementación, de modo que esta correspondencia es interpretativa. Entre las limitaciones se encuentran la conformación no aleatoria de los grupos, las diferencias basales, las pérdidas y duplicados en algunas bases, la unidad de registro mixta de la rúbrica, el tamaño reducido de la muestra vinculada para las correlaciones y el hecho de tratarse de una única institución, por lo que los resultados no deben generalizarse a otros grados, contenidos o contextos sin evidencia adicional.

CONCLUSIÓN

La integración de chatbots de inteligencia artificial en la estrategia IOBAS, acompañada de mediación docente y trabajo entre pares, se asoció con un mayor incremento del desempeño matemático de los estudiantes de básica secundaria de Samacá frente a la enseñanza tradicional: la ganancia global fue de 8,60 puntos en el grupo experimental y de 5,46 en el control, y la diferencia persistió tras ajustar por el pretest (η^2p = ,251), con mayor consistencia en las dimensiones conceptual e interpretativa que en la procedimental. La rúbrica y los artefactos mostraron que las mayores dificultades se concentraron en la identificación del problema y la organización de saberes, y los registros cualitativos evidenciaron una participación creciente y usos de los chatbots que fueron desde la solicitud de respuestas hasta la aclaración y la verificación. La valoración estudiantil fue favorable, pero no se relacionó de forma significativa con las ganancias individuales. Estos resultados corresponden a una configuración pedagógica compuesta, cuyos componentes no pudieron separarse, y a un diseño cuasiexperimental con diferencias basales, por lo que respaldan una valoración situada de la intervención y no la superioridad general de los chatbots. El aporte del estudio radica en mostrar que su valor depende de la forma en que se integran al proceso de resolución, sometidos a la explicación, la verificación y la justificación del estudiante.

FINANCIAMIENTO

No monetario.

CONFLICTO DE INTERÉS

No existe conflicto de interés con personas o instituciones ligadas a la investigación.

AGRADECIMIENTOS

A los estudiantes y docentes de la Institución Educativa Técnica Nacionalizada de Samacá, por su participación, y al Dr. César Ignacio Báez Quintero, director de la tesis doctoral de la cual deriva este artículo.

REFERENCIAS

Benítez-Pérez, A. A., García-Rodríguez, M. L., & Flores-Cervantes, C. (2023). Acercamiento a la comprensión de textos en la resolución de problemas en matemáticas considerando los procesos inductivos y deductivos: el caso de Bobby [Approach to text comprehension in mathematical problem solving considering inductive and deductive processes: the case of Bobby]. *RIDE. Revista Iberoamericana para la Investigación y el Desarrollo Educativo, 13*(26). https://doi.org/10.23913/ride.v13i26.1423

Chi, M. T. H., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist, 49*(4), 219–243. https://doi.org/10.1080/00461520.2014.965823

Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research* (3rd ed.). SAGE Publications.

Debets, T., Banihashem, S. K., Joosten-Ten Brinke, D., Vos, T. E. J., Maillette de Buy Wenniger, G., & Camp, G. (2025). Chatbots in education: A systematic review of objectives, underlying technology and theory, evaluation criteria, and impacts. *Computers & Education, 234*, 105323. https://doi.org/10.1016/j.compedu.2025.105323

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin, 119*(2), 254–284. https://doi.org/10.1037/0033-2909.119.2.254

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

Pérez-González, A., Quero-Méndez, O. N., & Bravo-Viera, J. L. (2021). Estrategia didáctica para enseñar a dirigir el proceso de enseñanza-aprendizaje de la resolución de problemas matemáticos [Didactic strategy to teach how to direct the teaching-learning process of mathematical problem solving]. *Revista Educación, 45*(1). https://doi.org/10.15517/revedu.v45i1.42112

Rutherford, T., Rodrigues, A., Duque-Baird, S., Veng, S., Mykyta-Chomsky, R., Cao, Y., Chisholm, K., & Bergwall, E. (2025). "I just think it is the way of the future": Teachers' use of ChatGPT to develop motivationally-supportive math lessons. *Computers & Education: Artificial Intelligence, 8*(3), 100367. https://doi.org/10.1016/j.caeai.2025.100367

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research, 78*(1), 153–189. https://doi.org/10.3102/0034654307313795

Sweller, J., van Merriënboer, J. J. G., & Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. *Educational Psychology Review, 31*(2), 261–292. https://doi.org/10.1007/s10648-019-09465-5

van de Pol, J., Mercer, N., & Volman, M. (2019). Scaffolding student understanding in small-group work: Students' uptake of teacher support in subsequent small-group interaction. *Journal of the Learning Sciences, 28*(2), 206–239. https://doi.org/10.1080/10508406.2018.1522258

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

Wisniewski, B., Zierer, K., & Hattie, J. (2020). The power of feedback revisited: A meta-analysis of educational feedback research. *Frontiers in Psychology, 10*, 3087. https://doi.org/10.3389/fpsyg.2019.03087

Derechos de autor: 2026 Por los autores. Este artículo es de acceso abierto y distribuido según los términos y condiciones de la licencia Creative Commons Atribución-NoComercial-CompartirIgual 4.0 Internacional (CC BY-NC-SA 4.0)
https://creativecommons.org/licenses/by-nc-sa/4.0/