



Asesorías y Tutorías para la Investigación Científica en la Educación Puig-Salabarría S.C.
José María Pino Suárez 400-2 esq a Lerdo de Tejada, Toluca, Estado de México. 7223898473
 RFC: ATI120618V12

Revista Dilemas Contemporáneos: Educación, Política y Valores.

<http://www.dilemascontemporaneoseducacionpoliticayvalores.com/>

Año: XIV Número: 1 Artículo no.:32 Período: 1 de septiembre al 31 de diciembre del 2026

TÍTULO: Efecto de una estrategia con software libre e inteligencia artificial sobre el rendimiento en álgebra universitaria: un estudio cuasiexperimental.

AUTORES:

1. Dr. Aníbal Zaldívar Colado.
2. Lic. Hannia Zaldívar Nava.
3. Dr. Jorge Lizárraga Reyes.
4. Dra. Gloria María Peña García.

RESUMEN: Se evaluó el efecto de una estrategia didáctica fundamentada en software libre (GeoGebra y Octave) e IA (ChatGPT y Wolfram Alpha) sobre el rendimiento académico en álgebra de estudiantes universitarios. Se planteó que dicha estrategia produciría un rendimiento significativamente superior al de la enseñanza tradicional. Se aplicó un diseño cuasiexperimental con grupos no equivalentes, modalidad pretest-posttest, durante 16 semanas, con 105 estudiantes (experimental $n=55$; control $n=50$). El primero obtuvo una ganancia de 27.61 puntos frente a 14.66 del segundo, con diferencia significativa ($t(103)=6.66$, $p<0.001$; $d=1.30$) y ANCOVA confirmatorio (η^2 parcial=0.303). La integración planificada de recursos digitales abiertos y asistentes generativos mejoró sustancialmente el aprendizaje del álgebra universitaria.

PALABRAS CLAVES: rendimiento académico, álgebra, software libre, inteligencia artificial, educación superior.

TITLE: Effect of an open-source software and artificial intelligence strategy on university algebra performance: A quasi-experimental study.

AUTHORS:

1. PhD. Aníbal Zaldívar Colado.
2. Bach. Hannia Zaldívar Nava.
3. PhD. Jorge Lizárraga Reyes.
4. PhD. Gloria María Peña García.

ABSTRACT: The effect of a teaching strategy based on open-source software (GeoGebra and Octave) and AI tools (ChatGPT and Wolfram Alpha) on university students' academic performance in algebra was evaluated. It was hypothesized that this strategy would produce significantly higher performance than traditional instruction. A quasi-experimental design with non-equivalent groups and a pretest-posttest format was implemented over 16 weeks with 105 students (experimental group: $n=55$; control group: $n=50$). The experimental group achieved a gain of 27.61 points, compared with 14.66 points in the control group, with a significant difference ($t(103)=6.66$, $p<0.001$; $d=1.30$) and confirmatory ANCOVA results (partial $\eta^2=0.303$). The planned integration of open digital resources and generative assistants substantially improved learning in university-level algebra.

KEY WORDS: academic performance, algebra, open-source software, artificial intelligence, higher education.

INTRODUCCIÓN.

La enseñanza del álgebra en las carreras de ingeniería y ciencias computacionales en los primeros semestres compone un reto constante para los establecimientos de educación superior. Los estudios empíricos han mostrado evidencias de altos índices de reprobación, baja retención y problemas cognitivos en los contenidos de aprendizaje como lo son: sistemas numéricos, inducción matemática, fracciones algebraicas, operaciones con polinomios, y resolución de sistemas de ecuaciones (Hillmayr *et al.*, 2020). La incorporación vinculada de herramientas tecnológicas de acceso abierto con recursos de inteligencia

artificial generativa (IAG), se ha colocado como una vía académica con alto poder para transformar la dinámica dentro del aula universitaria.

La oportunidad de investigar su impacto en contextos mexicanos reconoce a la necesidad de instaurar evidencia local sobre las posibles habilidades didácticas que optimicen los aprendizajes matemáticos, especialmente en programas educativos vinculados con las tecnologías de la información, donde el dominio algebraicamente es indefectible para unidades de aprendizaje posteriores. Una facultad de sistemas computacionales de una universidad pública en Mazatlán constituye una atmósfera propicia para inquirir este fenómeno.

Bray y Tangney (2017) expusieron a través de una revisión sistemática, que las tecnologías digitales en educación matemática causan progresos uniformes cuando se integran a estrategias con fundamento teórico. Siregar (2025) especificó las potencialidades de GeoGebra para enlazar representaciones algebraicas, geométricas y numéricas, resultado demostrado por Drijvers (2015).

Kasneci *et al.* (2023) compararon oportunidades y riesgos de ChatGPT en contextos educativos, y Cooper (2023) consiguió evidencia real acerca de la utilidad de la IA conversacional en la enseñanza de las ciencias. Los estudios que adoptan simultáneamente *software* libre e IA en la enseñanza universitaria del álgebra continúan siendo insuficientes, especialmente en el contexto mexicano.

DESARROLLO.

Pregunta de investigación.

Se plantea como pregunta de investigación: ¿En qué medida la integración didáctica de *software* libre (GeoGebra y Octave) y herramientas de inteligencia artificial (ChatGPT y Wolfram Alpha) mejora el rendimiento académico en álgebra de estudiantes del segundo semestre de las Licenciaturas en Informática e Ingeniería en Sistemas de Información, en comparación con la enseñanza tradicional?

Objetivo.

Como objetivo se determina: Evaluar el efecto de una estrategia didáctica basada en *software* libre (GeoGebra y Octave) e inteligencia artificial (ChatGPT y Wolfram Alpha) sobre el rendimiento académico en álgebra de estudiantes de segundo semestre de Facultad de Informática Mazatlán de la Universidad Autónoma de Sinaloa, mediante un diseño cuasiexperimental durante el ciclo escolar de enero a mayo de 2026.

Hipótesis (H₁).

Se establece como hipótesis (H₁): La aplicación de una estrategia didáctica fundamentada en *software* libre e inteligencia artificial produce un rendimiento académico en álgebra significativamente superior en estudiantes de segundo semestre, en comparación con quienes reciben enseñanza bajo un enfoque tradicional.

Estructura del documento.

El artículo se organiza en seis apartados: la introducción, el estado del arte, la metodología cuasiexperimental, la muestra y los instrumentos, los resultados cuantitativos, los hallazgos, las conclusiones, limitaciones y líneas futuras, y finalmente, se incluyen las referencias conforme a las normas APA, séptima edición.

Estado del arte.

El aprendizaje del álgebra sigue siendo uno de los ejercicios de mayor compromiso y exigencia de la educación universitaria, en virtud de la representación conceptual y la transición desde el razonamiento aritmético hacia métodos formales de generalidad.

En las carreras vinculadas con las ciencias computacionales, el dominio algebraico suele ser determinante para materias posteriores que exigen manejo simbólico, modelado y resolución de sistemas de ecuaciones.

La investigación reciente ha mostrado que el *software* dinámico de código abierto reduce las demandas

de esquematización y los inconvenientes procedimentales, beneficiando la perspicacia conceptual cuando los estudiantes interactúan con representaciones algebraicas, geométricas y numéricas de un mismo objeto (Egara & Mosimege, 2024). Esta evidencia redirige el problema: el inconveniente del álgebra no se imputa solamente al estudiante, sino que se liga con la mediación instruccional y con los recursos disponibles para representar y explorar los conceptos.

Dentro de los recursos de código abierto, GeoGebra reúne la mayor base empírica reciente. En un diseño cuasiexperimental con grupos no equivalentes, Egara y Mosimege (2024) reportaron diferencias estadísticamente significativas en la comprensión conceptual a favor del grupo que aprendió con apoyo de GeoGebra frente a la enseñanza convencional, atribuyendo el resultado a la reducción de la carga abstracta y a la conexión entre registros.

GNU Octave, también de licencia libre, ofrece un entorno de cómputo numérico y matricial con sintaxis compatible con MATLAB y capacidades adecuadas para operar con matrices y resolver sistemas de ecuaciones; las revisiones recientes acerca de tecnología digital en educación matemática (Deng *et al.*, 2024; Pepin *et al.*, 2025) han incluido pocos estudios con entornos de cómputo numérico de código abierto, lo cual configura un punto de interés metodológico para investigaciones empíricas que integren ambos tipos de herramientas dentro de una misma estrategia didáctica.

En cuanto a la IAG, la síntesis cuantitativa disponible respalda un efecto positivo sobre el aprendizaje. Deng *et al.* (2024) examinaron 69 estudios experimentales publicados entre los años 2022 y 2024, y concluyeron que las intervenciones con ChatGPT, predominantes en el nivel universitario, mejoran el rendimiento académico, los estados afectivo-motivacionales y las disposiciones de pensamiento de orden superior, además de reducir el esfuerzo mental, aunque sin un efecto significativo sobre la autoeficacia. Los propios autores advirtieron limitaciones de diseño que condicionan las inferencias causales. En educación superior de ingeniería, Noster *et al.* (2024) evaluaron varias herramientas generativas, incluida Wolfram, frente a tareas reales de evaluación en diez asignaturas, y observaron que su desempeño varía

según el tipo de tarea y que su integración exige criterios de integridad académica y orientación docente. Los dos trabajos concordaron en que el resultado de estas herramientas estriba del diseño instruccional y no de su mera disponibilidad.

En el terreno concretamente matemático, los hallazgos fueron más matizados. Pepin *et al.* (2025) esquematizaron la literatura sobre ChatGPT en educación matemática y resaltaron su utilidad para una retroalimentación pronta, los esclarecimientos personalizados, y el aprendizaje autorregulado, al tiempo en que mostraron equivocaciones esporádicas, riesgo de dependencia, y la necesidad de supervisión humana.

Korkmaz Guler *et al.* (2024) evaluaron el desempeño de ChatGPT en un examen nacional de matemáticas y encontraron resultados heterogéneos: la versión GPT-4 superó a GPT-3.5 en comprensión y detalle, pero ambas cometieron errores que aconsejan mediación pedagógica. En un contexto latinoamericano cercano al de este estudio, Rigaud Téllez *et al.* (2024) propusieron, en una universidad mexicana, un marco de generación de *prompts* para mejorar la retroalimentación que docentes y ChatGPT ofrecen ante problemas verbales de álgebra lineal, evidenciando posibilidades concretas y también desafíos en problemas de mayor complejidad.

Lo expuesto evidencia un patrón estable, y a la vez, una omisión. Existe evidencia que respalda por separado el efecto del *software* libre y el de la IAG sobre el aprendizaje matemático, con coincidencia en que su eficacia depende del diseño didáctico que las integra; sin embargo, los estudios que combinan de manera deliberada ambos tipos de recursos para la enseñanza universitaria del álgebra son aún limitados, y la mayoría se concentra en la educación secundaria, en disciplinas distintas del álgebra, o en herramientas aisladas.

La escasez de diseños cuasiexperimentales que miden el rendimiento académico en álgebra al integrar *software* libre e IA en programas de informática e ingeniería en sistemas, especialmente en contextos universitarios de Latinoamérica, precisa el vacío que sustenta este trabajo. Tomar en cuenta ese vacío

justifica la pregunta de investigación y la hipótesis de que una estrategia didáctica fundamentada en GeoGebra, Octave, ChatGPT y Wolfram Alpha origina un rendimiento significativamente superior al de una enseñanza tradicional, lo que orienta el objetivo de evaluar dicho efecto mediante un diseño cuasiexperimental en el ciclo escolar de enero a mayo de 2026.

Metodología.

Enfoque y diseño.

El estudio adoptó un enfoque cuantitativo, alcance explicativo y diseño cuasiexperimental con grupos no equivalentes, modalidad pretest-posttest. La asignación de los participantes a las condiciones experimental y de control no se realizó de manera aleatoria; se respetó la estructura administrativa de los grupos académicos, acorde a los lineamientos descritos por Shadish *et al.* (2002) para investigaciones educativas en contextos reales. El diseño se estructuró bajo el esquema O1 X O2 para el grupo experimental y O3 O4 para el grupo control, donde O representa las mediciones y X la intervención didáctica con *software* libre e IA.

Participantes.

La muestra fue de 105 estudiantes de segundo semestre inscritos en la asignatura de álgebra durante el ciclo escolar de enero a mayo del año 2026, pertenecientes a la Licenciatura de Informática y la de Ingeniería en Sistemas de Información de la Facultad de Informática Mazatlán, Universidad Autónoma de Sinaloa. El Grupo 1 ($n=55$; 20 mujeres y 35 hombres) funcionó como grupo experimental; el Grupo 2 ($n=50$; 23 mujeres y 27 hombres) operó como grupo control. La edad fluctuó entre 18 y 22 años ($M=19.3$; $DE=1.2$). El muestreo fue no probabilístico por conveniencia.

Variables.

La variable independiente fue la estrategia didáctica compuesta por dos niveles: la enseñanza con *software* libre (GeoGebra y Octave) y la IA (ChatGPT y Wolfram Alpha) para el grupo experimental, y enseñanza

tradicional (exposición magistral, ejercicios en el pizarrón y en libro de texto) para el grupo control. La variable dependiente fue el rendimiento académico en álgebra, operacionalizada como puntaje de prueba objetiva usando una escala de 0 a 100 puntos. Como variables de control fueron sexo, edad, promedio académico del semestre anterior y la experiencia con *software* matemático.

Instrumentos.

Se emplearon dos instrumentos. La Prueba de Rendimiento en Álgebra (PRA) fue establecida *ad hoc* con base en los contenidos curriculares formales: fracciones algebraicas, sistemas numéricos, inducción matemática, operaciones con polinomios, y sistemas de ecuaciones, compuesta por 25 reactivos distribuidos de forma proporcional entre los cinco bloques.

La validez de contenido se determinó mediante juicio de cinco expertos en matemática educativa, y el índice de validez de contenido de Lawshe modificado por Tristán-López (2008) fue de 0.87. La confiabilidad por consistencia interna fue de $\alpha=0.84$ en una aplicación piloto con 30 estudiantes de particularidades similares. El segundo instrumento fue un cuestionario sociodemográfico con nueve reactivos relativo a sexo, edad, carrera, promedio académico previo, y experiencia tecnológica matemática.

Procedimiento.

La intervención fue desarrollada durante 16 semanas, en dos sesiones semanales de 90 minutos (48 horas por grupo). Durante la primera semana se usó el pretest y el posttest hasta la semana 16. En el grupo experimental, GeoGebra fue utilizado para visualizar funciones y representación de sistemas de ecuaciones; Octave para operaciones con matrices y resolución computacional; el ChatGPT como asistente conversacional para esclarecimientos guiados y retroalimentación, y Wolfram Alpha para comprobación simbólica y exploración paso a paso.

La secuencia se distribuyó bajo los principios del aprendizaje activo. En el grupo control, la enseñanza se desarrolló con un enfoque tradicional sin recursos tecnológicos digitales. Los dos grupos fueron

atendidos por el mismo profesor, con la finalidad de eliminar la variabilidad atribuible a diferencias de formación, estilos de enseñanza o experiencia entre profesores; dicho profesor tiene el grado académico de doctorado y experiencia de más de veinte años en la materia. Esta decisión hizo posible controlar por diseño el efecto del docente como variable de confusión; de tal forma, que la única diferencia sistemática entre condiciones fue la mediación tecnológica. En ambos grupos se mantuvieron homogéneos, tanto los contenidos, la carga horaria, los tiempos, como los criterios de evaluación.

Consideraciones éticas.

El estudio se llevó a cabo conforme a la Declaración de Helsinki (Asociación Médica Mundial, 2024) y al reglamento de investigación de la Universidad Autónoma de Sinaloa. Se obtuvo aprobación del comité académico de la Facultad de interés. Los estudiantes firmaron autorización, y se garantizó la confidencialidad y anonimato de los datos recabados. La disposición de participar no afectó ninguna de sus calificaciones académicas.

Análisis estadístico.

El procesamiento se realizó con IBM SPSS Statistics, versión 27 (IBM Corp, 2020). Se utilizaron estadísticos descriptivos, prueba de Kolmogorov-Smirnov con corrección de Lilliefors para normalidad, y prueba de Levene para homogeneidad de varianzas. Comprobados los supuestos paramétricos, se aplicó prueba t de Student para muestras independientes y relacionadas, y análisis de covarianza (ANCOVA) con el pretest como covariable. Se computarizó el tamaño del efecto con la d de Cohen y el η^2 parcial (Daniel, 2016). El nivel de significancia se instituyó en $p < 0.05$.

Resultados.

Caracterización y verificación de supuestos.

La muestra final conservó la integridad inicial de 105 estudiantes, sin pérdidas muestrales y la tabla 1 muestra la distribución sociodemográfica.

Las pruebas de chi cuadrado sobre las variables categóricas no revelaron diferencias significativas entre grupos (χ^2 sexo=1.04, $p=0.308$; χ^2 carrera=0.62, $p=0.431$). La prueba de Kolmogorov-Smirnov con corrección de Lilliefors demostró distribuciones compatibles con la normalidad (todos los $p>0.05$). La prueba de Levene ratificó la homogeneidad de varianzas en pretest ($F=0.06$, $p=0.811$) y postest ($F=0.28$, $p=0.598$), respaldando la aplicación de pruebas paramétricas.

Tabla 1. Características sociodemográficas de los participantes. Fuente propia.

Variable	Grupo experimental ($n=55$)	Grupo control ($n=50$)	Total ($n=105$)
Sexo: mujeres	20 (36.4%)	23 (46.0%)	43 (41.0%)
Sexo: hombres	35 (63.6%)	27 (54.0%)	62 (59.0%)
Edad (DE)	19.4 (1.1%)	19.2 (1.3%)	19.3 (1.2%)
Lic. en Informática	29 (52.7%)	24 (48.0%)	53 (50.5%)
Ing. en Sistemas	26 (47.3%)	26 (52.0%)	52 (49.5%)

Fuente propia.

La prueba t de Student para muestras independientes sobre los puntajes del pretest mostró equivalencia inicial estadística entre los grupos, $t(103)=-0.10$, $p=0.924$. El análisis intragrupal con prueba t para muestras relacionadas reveló incrementos significativos en ambos grupos. La ganancia experimental fue de 27.61 puntos, $t(54)=14.88$, $p<0.001$; la ganancia control fue de 14.66 puntos, $t(49)=6.78$, $p<0.001$. La magnitud de la mejora resultó notoriamente superior en el grupo expuesto a la estrategia didáctica con *software* libre e IA (Tabla 2).

Tabla 2. Estadísticos descriptivos y comparaciones intragrupales pretest-postest.

Grupo	Momento	M	DE	Ganancia	t	gl	p
Experimental ($n=55$)	Pretest	51.24	9.54	-	-	-	-
	Postest	78.85	9.23	27.61	14.88	54	< 0.001
Control ($n=50$)	Pretest	51.42	9.52	-	-	-	-
	Postest	66.08	10.43	14.66	6.78	49	< 0.001

Comparación intergrupos y ANCOVA. Fuente propia.

La prueba t para muestras independientes sobre el posttest reveló una diferencia altamente significativa a favor del grupo experimental, $t(103)=6.66$, $p<0.001$, con diferencia de medias de 12.77 puntos.

El tamaño del efecto d de Cohen fue 1.30, considerado grande (Cohen, 1988). El ANCOVA, utilizando el pretest como covariable, evidenció un efecto significativo del grupo sobre el posttest, $F(1, 102)=44.44$, $p<0.001$, con η^2 parcial=0.303 (Tabla 3). Las medias ajustadas fueron 78.84 (experimental) y 66.09 (control) con diferencia ajustada de 12.75 puntos.

Tabla 3. Resultados del análisis de covarianza con pretest como covariable. Fuente propia.

Fuente de variación	SC	gl	CM	F	p	η^2 parcial
Pretest (covariable)	17.94	1	17.94	0.19	0.668	0.002
Grupo	4,226.81	1	4 226.81	44.44	< 0.001	0.303
Error	9,700.45	102	95.10	-	-	-
Total corregido	13,945.20	104	-	-	-	-

Fuente propia.

Las comparaciones intergrupales por bloque temático evidenciaron diferencias estadísticamente significativas a favor del grupo experimental en los cinco bloques, con tamaños del efecto entre 0.79 y 1.17 (Tabla 4).

Tabla 4. Comparación intergrupar por bloque temático en el posttest. Fuente propia.

Bloque temático	Exp. M (DE)	Ctrl. M (DE)	t (103)	p	d de Cohen
Sistemas numéricos	16.74 (2.29)	14.53 (2.14)	5.11	<0.001	1.00
Inducción matemática	14.88 (1.93)	11.92 (3.06)	5.97	<0.001	1.17
Fracciones algebraicas	15.48 (2.52)	12.61 (2.80)	5.54	<0.001	1.08
Operaciones con polinomios	15.83 (2.23)	13.97 (2.37)	4.13	<0.001	0.81
Sistemas de ecuaciones	14.35 (3.05)	12.07 (2.68)	4.06	<0.001	0.79

Fuente propia.

El mayor efecto correspondió a inducción matemática ($d=1.17$), seguido de fracciones algebraicas ($d=1.08$) y sistemas numéricos ($d=1.00$). La distribución de los efectos propone un impacto coherente de la estrategia didáctica sobre los contenidos curriculares evaluados con mayor intensidad en temas que demandan reflexión simbólico formal.

Discusión.

El propósito del estudio consistió en evaluar el efecto de una estrategia didáctica fundada en *software* libre (GeoGebra y Octave) e IA (ChatGPT y Wolfram Alpha) con el rendimiento académico en álgebra de alumnos de segundo semestre de la Facultad de Informática Mazatlán de la Universidad Autónoma de Sinaloa. Los resultados responden de manera afirmativa a la pregunta de investigación: el grupo expuesto a la intervención logró una ganancia pretest-postest de 27.61 puntos frente a 14.66 del grupo bajo enseñanza usual o tradicional con diferencia intergrupar estadísticamente significativa en el postest ($t(103)=6.66, p<0.001$) y tamaño del efecto grande ($d=1.30$) acorde con los criterios de Cohen (1988).

El ANCOVA con el pretest como covariable ratificó un efecto del grupo sustancial ($F(1, 102)=44.44, p<0.001; \eta^2 \text{ parcial}=0.303$), situación que respaldó la atribución causal de la mejora a la condición experimental y no a diferencias iniciales entre los grupos.

La equivalencia inicial comprobada por la prueba t sobre el pretest ($t(103)=-0.10, p=0.924$) y la ausencia de diferencias sociodemográficas significativas entre grupos consintieron condiciones adecuadas para interpretar la magnitud del efecto observado. La ganancia experimental casi duplicó la del grupo control, indicando que la mediación con recursos digitales operó como un factor diferencial dentro de un contexto donde los dos grupos recibieron carga horaria, contenidos y criterios de evaluación similares. La covariable pretest no resultó estadísticamente significativa en el modelo ($F=0.19, p=0.668$), lo que sugiere que el desempeño previo no determinó el resultado final y fortalece la idea de que la variación entre grupos respondió a la intervención y no a un sesgo de partida.

El análisis por bloques temáticos aportó datos adicionales. Las mayores magnitudes del efecto se concentraron en inducción matemática ($d=1.17$), fracciones algebraicas ($d=1.08$) y sistemas numéricos ($d=1.00$); es decir, en los contenidos que demanda conceptualización, generalización y tratamiento sintáctico.

Los bloques de operaciones con polinomios ($d=0.81$) y sistemas de ecuaciones ($d=0.79$) tuvieron registros de efectos algo menores, aunque también amplios. Este patrón resulta coherente con la idea de que las herramientas dinámicas y los asistentes generativos contribuyen con mayor fuerza en aquellos contenidos cuya dificultad pedagógica se asocia a la esquematización formal; una interpretación congruente con lo planteado por Egara y Mosimege (2024) al atribuir la mejora a la reducción de la carga abstracta y a la conexión entre registros algebraicos, geométricos y numéricos.

La convergencia con la literatura previa es clara. La magnitud del efecto encontrada ($d=1.30$) superó el tamaño reportado por Hillmayr *et al.* (2020) en su metaanálisis sobre tecnologías digitales en matemáticas y ciencias en secundaria ($g=0.65$), lo cual sugiere que la combinación de *software* libre y asistentes basados en IA podría ofrecer un retorno didáctico mayor que el documentado para herramientas digitales utilizadas de manera aislada.

La línea trazada por Bray y Tangney (2017) acerca de la dependencia del éxito tecnológico respecto a su marco teórico-pedagógico se reflejó también en estos resultados: la intervención no se sustentó en la mera incorporación de cuatro recursos, sino en una secuencia didáctica con principios de aprendizaje activo y mediación docente sostenida durante 16 semanas.

Los resultados coinciden con la evidencia reciente sobre GeoGebra; Egara y Mosimege (2024) reportaron, mediante un diseño cuasiexperimental con grupos no equivalentes, mejoras significativas en la comprensión conceptual atribuibles al trabajo con representaciones múltiples; el presente estudio extendió ese hallazgo al terreno del álgebra universitaria y a una muestra mexicana con perfil tecnológico. La descripción inicial de GeoGebra como entorno para enlazar representaciones algebraicas, geométricas y

numéricas (Siregar, 2025) y la caracterización general de las condiciones de éxito de las tecnologías digitales en matemáticas formulada por Drijvers (2015) brindan un marco interpretativo coherente con las ganancias observadas en bloques con alta demanda visual y simbólica.

En lo concerniente a la IAG, los resultados se alinean con la síntesis cuantitativa de Deng *et al.* (2024), quienes documentaron mejoras en el rendimiento académico, los estados afectivo-motivacionales, y el pensamiento de orden superior en intervenciones con ChatGPT, particularmente en el nivel universitario. La revisión panorámica de Pepin *et al.* (2025) sobre ChatGPT en educación matemática señaló su utilidad para retroalimentación inmediata, explicaciones personalizadas, y aprendizaje autorregulado, pero advirtió sobre inexactitudes y dependencia.

En el estudio aquí presentado, la mediación docente, la verificación cruzada con Wolfram Alpha, y la integración con GeoGebra y Octave parecen haber atenuado esos riesgos, lo cual respalda la lectura de que el diseño instruccional, más que la herramienta en sí, condiciona el efecto. La perspectiva de Kasneci *et al.* (2023) acerca de las oportunidades y los riesgos de los modelos de lenguaje, y la evidencia de Cooper (2023) en la enseñanza de las ciencias, se ven matizadas por estos hallazgos al mostrar que la convergencia de herramientas dispares puede producir un efecto agregado mayor que el documentado para cada recurso por separado.

La similitud con la literatura es también parcial en algunos puntos; por ejemplo, Korkmaz Guler *et al.* (2024) describieron un desempeño heterogéneo de ChatGPT en un examen nacional de matemáticas, con errores frecuentes incluso en GPT-4, lo cual podría insinuar un efecto educativo modesto. La diferencia con los hallazgos de este estudio se explica plausiblemente porque allí el *chatbots* operó como solucionador de problemas, mientras que aquí actuó como asistente mediado por el docente para explicaciones guiadas y retroalimentación, función en la que la literatura ha reportado los mayores beneficios.

En la misma dirección, Rigaud Téllez *et al.* (2024), en una universidad mexicana, mostraron que un marco estructurado de generación de *prompts* mejora la calidad de la retroalimentación en álgebra lineal, lo que respalda la decisión metodológica de integrar ChatGPT bajo orientación docente y no como herramienta de uso libre. Noster *et al.* (2024) advirtieron, en evaluaciones de ingeniería con varias herramientas generativas y Wolfram, que el desempeño varía según el tipo de tarea y que la incorporación requiere criterios de integridad académica; los resultados de este trabajo sugieren que esta implementación es viable cuando se realiza con planificación didáctica explícita.

La hipótesis de investigación fue confirmada plenamente, debido a que el rendimiento académico en álgebra fue significativamente superior en el grupo experimental con efectos grandes en la comparación directa, en el modelo ajustado por pretest, y en cada uno de los cinco bloques temáticos. La uniformidad de la dirección y la magnitud del efecto a través de distintas pruebas estadísticas, incluida la verificación de supuestos paramétricos (Kolmogorov-Smirnov con corrección de Lilliefors y prueba de Levene), refuerza la credibilidad de la inferencia. La hipótesis nula de igualdad entre condiciones fue rechazada con un margen amplio, lo cual sustenta la conclusión sobre el efecto positivo de la estrategia.

En cuanto a las implicaciones del estudio, en el plano teórico, los resultados contribuyen con evidencia empírica a favor de una visión integrada de la mediación tecnológica en la educación matemática, en la que el *software* de visualización, el entorno de cómputo numérico, el asistente conversacional, y el motor de conocimiento computacional ayudan aportando funciones distintas pero articuladas para obtener un objetivo de aprendizaje común bajo una misma guía.

En el plano metodológico, el trabajo muestra la viabilidad de emplear diseños cuasiexperimentales con ANCOVA y cálculo de tamaño del efecto en contextos universitarios reales, acorde a los lineamientos de Shadish *et al.* (2002) para investigaciones educativas con grupos intactos. En el plano práctico, los resultados amparan políticas institucionales encaminadas a incorporar herramientas de licencia abierta y

asistentes generativos con mediación docente en los primeros semestres de las carreras de informática y de sistemas, donde el dominio algebraico condiciona el avance curricular.

Se reconocen las restricciones inherentes a la investigación, y en lo concerniente al muestreo, fue no probabilístico por conveniencia y los grupos se respetaron en su conformación administrativa; la muestra de 105 estudiantes de una sola institución del noroeste de México limita la generalización de los resultados a contextos de mayor diversidad cultural, lingüística e institucional. El diseño cuasiexperimental, frente a un experimento puro, deja abierta la posibilidad de variables no controladas, pese a la equivalencia inicial verificada y al uso del ANCOVA. Si bien la asignación de un mismo profesor a ambos grupos controló por diseño el efecto del instructor, esta misma condición no excluye un posible sesgo de expectativa del docente hacia la condición experimental, ni el efecto de novedad asociado a la introducción de herramientas tecnológicas no habituales para los alumnos. Ambos fenómenos podrían haber contribuido a la magnitud del efecto observado y no fueron medidos de manera independiente, por lo que conviene considerarlos en futuras réplicas mediante protocolos de enmascaramiento parcial o intervenciones de mayor duración que atenúen la novedad.

La duración de la intervención (16 semanas) cubre un ciclo escolar completo (un semestre), pero no es posible estimar la persistencia del aprendizaje en el mediano plazo. La medición del rendimiento se centró en una prueba objetiva validada ($IVC=0.87$; $\alpha=0.84$) según el procedimiento de Tristán-López (2008), pero no incorporó indicadores cualitativos de procesos cognitivos, afectivos, o de uso de las herramientas. El estudio en todo momento se ajustó a la Declaración de Helsinki y al reglamento institucional, aunque sin componente cualitativo que documente la experiencia subjetiva de los colaboradores.

Las líneas futuras se desprenden de las restricciones señaladas; por lo que resulta adecuado replicar el diseño con muestras de mayor tamaño y origen multiinstitucional, incluir un postest diferido para evaluar retención, e incorporar variables como autoeficacia matemática, motivación y carga cognitiva. La inclusión de métodos mixtos pudiese mejorar la perspicacia de cómo los estudiantes interactúan con cada

herramienta, qué tipos de *prompts* generan mejor retroalimentación y cómo el profesor regulariza el uso del asistente generativo. Una línea adicional reside en aislar el efecto de cada herramienta a través de diseños factoriales, dado que esta investigación evaluó el efecto agregado de la integración. La extensión del estudio a otras unidades de aprendizaje con alta carga simbólica, como cálculo diferencial y álgebra lineal, podría ratificar si la magnitud del efecto se conserva en contenidos distintos.

Los resultados sostienen que la implementación planificada de *software* libre e IAG, con mediación docente y diseño didáctico explícito, origina mejoras académicas importantes en el aprendizaje del álgebra en el nivel superior. La evidencia conseguida en la Facultad de Informática Mazatlán favorece a un campo en el que la generalidad de las publicaciones recientes documenta el efecto de cada herramienta por separado o en niveles distintos del universitario. La coherencia entre la magnitud del efecto observado, la dirección de los resultados en los cinco bloques, y la afinidad con la literatura internacional amparan la conveniencia de seguir explorando este modelo de intervención en contextos universitarios de Latinoamérica.

CONCLUSIONES.

El estudio evaluó el efecto de una estrategia didáctica fundamentada en *software* libre (GeoGebra y Octave) e IA (ChatGPT y Wolfram Alpha) con el rendimiento académico en álgebra de los alumnos de segundo semestre de Licenciatura en Informática y de Ingeniería en Sistemas de Información de la Facultad de Informática Mazatlán, durante el periodo escolar de enero a mayo del 2026.

El objetivo general se consumó mediante un diseño cuasiexperimental con grupos no equivalentes, pretest-posttest con muestra de 105 estudiantes y comprobación de los supuestos paramétricos intimados. La pregunta de investigación obtuvo una respuesta afirmativa y cuantificada: la integración didáctica de las cuatro herramientas mejoró el rendimiento académico en álgebra de manera estadísticamente significativa frente a la enseñanza tradicional con una diferencia ajustada de 12.75 puntos en una escala de 0 a 100 y con un tamaño del efecto grande según los criterios convencionales ($d=1.30$; η^2 parcial=0.303).

La hipótesis de investigación (H_1) se comprobó completamente. El grupo expuesto a la intervención alcanzó una ganancia pretest-posttest de 27.61 puntos, casi el doble de los 14.66 puntos logrados por el grupo bajo enseñanza tradicional. La superioridad del grupo experimental se sostuvo tanto en la comparación directa ($t(103)=6.66, p<0.001$) como en el modelo ANCOVA con el pretest como covariable ($F(1, 102)=44.44, p<0.001$) y se replicó en los cinco bloques temáticos evaluados con tamaños del efecto entre 0.79 y 1.17. La uniformidad de la dirección y de la magnitud del efecto, junto con la equivalencia inicial comprobada entre grupos, aporta evidencia consistente de que las diferencias observadas se asocian con la condición experimental y ampara el rechazo de la hipótesis nula.

Los resultados por bloque temático le añaden un matiz pedagógico de interés. La intervención favoreció mejoras en inducción matemática, fracciones algebraicas, y sistemas numéricos; es decir, en los contenidos cuya dificultad se asocia con razonamiento simbólico y generalización formal.

Las mejoras en operaciones con polinomios y sistemas de ecuaciones, aunque algo menores, igualmente fueron sustanciales. Este patrón propone que la combinación entre visualización dinámica, cómputo numérico, asistencia conversacional, y verificación computacional maniobra con mayor fuerza allí donde la enseñanza tradicional encuentra sus límites más conocidos.

Es conveniente precisar el alcance de las conclusiones, debido a que la muestra fue no probabilística por conveniencia, los grupos se respetaron en su estructura administrativa, y la investigación se efectuó en solo una facultad del noroeste de México, lo cual ajusta la generalización de los resultados a contextos universitarios de perfil tecnológico similar. La duración de la intervención cubrió un ciclo escolar, lo que imposibilita estimar la persistencia del aprendizaje en el mediano plazo.

La medición del rendimiento se basó en una prueba objetiva validada, y aunque suficiente para los objetivos planteados, no incorpora indicadores cualitativos sobre procesos cognitivos o de la experiencia de uso de cada herramienta. El diseño cuasiexperimental, pese a la equivalencia inicial verificada y al

control estadístico mediante ANCOVA, deja abierta la posibilidad de variables no medidas que conviene examinar en futuras investigaciones.

La investigación contribuye con evidencia empírica de que una estrategia didáctica basada en GeoGebra, Octave, ChatGPT y Wolfram Alpha, planeada bajo principios de aprendizaje activo y sostenida por mediación docente, origina mejoras académicas amplias y coherentes en el aprendizaje del álgebra escolar, con efecto superior al de la enseñanza tradicional. El trabajo responde de modo afirmativo a la pregunta de investigación, confirma la hipótesis, y cumple el objetivo general.

Los resultados obtenidos en la Facultad de Informática Mazatlán contribuyen a un campo en el que la mayoría de la literatura reciente documenta el efecto de cada herramienta por separado o en niveles distintos del universitario, y abren una agenda investigativa pertinente para la educación matemática en universidades latinoamericanas.

REFERENCIAS BIBLIOGRÁFICAS.

1. Asociación Médica Mundial. (2024). Declaración de Helsinki de la AMM: Principios éticos para las investigaciones médicas en seres humanos. <https://www.wma.net/es/policies-post/declaracion-de-helsinki-de-la-amm-principios-eticos-para-las-investigaciones-medicas-en-seres-humanos/>
2. Bray, A., y Tangney, B. (2017). Technology usage in mathematics education research: A systematic review of recent trends. Computers & Education, 114, 255-273. <https://doi.org/10.1016/j.compedu.2017.07.004>
3. Cohen, J. (1988). Statistical Power Analysis for the Behavioral Sciences. Second Edition. Lawrence Erlbaum Associates, Publishers. Library of Congress. Publication Data. <https://utstat.toronto.edu/~brunner/oldclass/378f16/readings/CohenPower.pdf>
4. Cohen, J. (2013). Statistical power analysis for the behavioral sciences. Routledge.

5. Cooper, G. (2023). Examining science education in ChatGPT: An exploratory study of generative artificial intelligence. *Journal of Science Education and Technology*, 32(3), 444-452.
<http://doi.org/10.1007/s10956-023-10039-Y>
6. Daniel, W. W. (2016). *Bioestadística: Base para el análisis de las ciencias de la salud*. Limusa Wiley.
7. Deng, R., Jiang, M., Yu, X., Lu, Y., y Liu, S. (2024). Does ChatGPT enhance student learning? A systematic review and meta-analysis of experimental studies. *Computers & Education*, 227, 1-35.
<https://doi.org/10.1016/j.compedu.2024.105224>
8. Drijvers, P. (2015). Digital technology in mathematics education: Why it works (or doesn't). En S. J. Cho (Ed.), *Selected regular lectures from the 12th International Congress on Mathematical Education* (pp. 135-151). Springer. https://doi.org/10.1007/978-3-319-17187-6_8
9. Egara, F. O., & Mosimege, M. (2024). Exploring the Integration of Artificial Intelligence-Based ChatGPT into Mathematics Instruction: Perceptions, Challenges, and Implications for Educators. *Education Sciences*, 14(7), 742. <https://doi.org/10.3390/educsci14070742>
10. Hillmayr, D., Ziernwald, L., Reinhold, F., Hofer, S. I., y Reiss, K. M. (2020). The potential of digital tools to enhance mathematics and science learning in secondary schools: A context-specific meta-analysis. *Computers & Education*, 153, 1-25. <https://doi.org/10.1016/j.compedu.2020.103897>
11. IBM Corp. (2020). IBM SPSS. Statistics for Windows, (Version 27.0) [Computer software].
<https://www.ibm.com/products/spss-statistics>
12. Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günemann, 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>

13. Korkmaz Guler, N., Dertli, Z. G., Boran, E., y 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>
14. Noster, N., Gerber, S., & Siller, H.-S. (2024). Pre-Service Teachers' Approaches in Solving Mathematics Tasks with ChatGPT. *Digital Experiences in Mathematics Education*, 10, 543–567. <https://doi.org/10.1007/s40751-024-00155-8>
15. Pepin, B., Buchholtz, N., y Salinas-Hernández, U. (2025). A scoping survey of ChatGPT in mathematics education. *Digital Experiences in Mathematics Education*, 11(1), 9-41. <https://doi.org/10.1007/s40751-025-00172-1>
16. Rigaud Téllez, N., Rayón Villela, P., y Blanco Bautista, R. (2024). Evaluating ChatGPT-generated linear algebra formative assessments. *International Journal of Interactive Multimedia and Artificial Intelligence*, 8(5), 75-82. <https://doi.org/10.9781/ijimai.2024.02.004>
17. Shadish, W. R., Cook, T. D., y Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin.
18. Siregar, T. (2025). Application of GeoGebra for teaching mathematics. Preprints.org. https://www.preprints.org/frontend/manuscript/b21ecc5c71e6dcb7f9dcf72ff6051d62/download_publication
19. Tristán-López, A. (2008). Modificación al modelo de Lawshe para el dictamen cuantitativo de la validez de contenido de un instrumento objetivo. *Avances en Medición*, 6(1), 37-48. https://www.humanas.unal.edu.co/lab_psicometria/application/files/9716/0463/3548/VOL_6_Articulo4_Indice_de_validez_de_contenido_37-48.pdf

DATOS DE LOS AUTORES.

1. **Aníbal Zaldívar Colado.** Doctor en Educación por la Universidad de Durango. Profesor investigador de la Facultad de Informática Mazatlán de la Universidad Autónoma de Sinaloa. México. Integrante

del SNII nivel II, Investigador honorífico del SSIT, México. Correo electrónico: azaldivar@uas.edu.mx

ORCID ID: <https://orcid.org/0000-0002-6622-6630>.

2. **Hannia Zaldívar Nava.** Licenciada en Enseñanza del Idioma Inglés. Profesora de la Universidad Autónoma de Durango, México. Correo electrónico: hanniazaldivar5@gmail.com
3. **Jorge Lizárraga Reyes.** Doctor en Educación por la Universidad del Pacífico Norte. Profesor Investigador de la Universidad Autónoma de Sinaloa, México. Correo electrónico: jorge.uas@uas.edu.mx ORCID: <https://orcid.org/0000-0001-6281-4446>.
4. **Gloria María Peña García.** Doctora en Enfermería por la Universidad Andrés Bello de Santiago de Chile. Profesora Investigadora de la Facultad de Enfermería Mazatlán de la Universidad Autónoma de Sinaloa, México. Integrante del SNII nivel II, Investigadora honorífica del SSIT, México. Correo electrónico: gpena@uas.edu.mx ORCID ID: <https://orcid.org/0000-0001-9935-608X>.

RECIBIDO: 7 de julio del 2026.

APROBADO: 29 de julio del 2026.