



Inteligencia artificial como estrategia para fortalecer el aprendizaje de Ciencias Naturales y el vocabulario científico en inglés en estudiantes de Educación Básica Superior

Artificial Intelligence as a Strategy to Strengthen Natural Science Learning and Scientific Vocabulary in english among Upper Basic Education Students

Artículo de investigación científica
Ciencias de la Educación



Omar Antonio Molina Oña

omar.molina@docentes.educacion.edu.ec

<https://orcid.org/0009-0005-0246-1374>

 Ministerio de educación – Ciencias Naturales
Cotopaxi - Ecuador

Karina Lorena Pacheco Oña

karina.pacheco@docentes.educacion.edu.ec

<https://orcid.org/0009-0008-8863-066X>

 Ministerio de educación – Ciencias Naturales
Cotopaxi - Ecuador

Fecha de envío: 2026-04-05

Fecha de revisión: 2026-05-05

Fecha da aceptación: 2026-06-26

Resumen

El objetivo de la investigación fue determinar la incidencia de una estrategia didáctica mediada por inteligencia artificial en el aprendizaje de Ciencias Naturales y en la adquisición de vocabulario científico en inglés de estudiantes de Educación Básica Superior. Se adoptó un diseño cuasi experimental de enfoque cuantitativo, aplicado a 70 estudiantes; 35 integraron el grupo experimental y 35 el grupo de control. El instrumento fue un test de base estructurada de 40 ítems, validado por expertos, con una confiabilidad alfa de Cronbach de 0,87. La intervención BioLex-IA se desarrolló durante ocho semanas e incluyó tutoría conversacional, generación guiada de explicaciones, comparación de evidencias, glosarios bilingües y retroalimentación inmediata. Para el análisis se utilizaron estadísticos descriptivos, correlación de Pearson, prueba t de Student para muestras independientes y tamaño del efecto d de Cohen. Los resultados mostraron equivalencia inicial entre los grupos ($p > 0,05$) y diferencias significativas en el postest a favor del grupo experimental, tanto en el rendimiento global como en comprensión conceptual, indagación, interpretación de evidencias y vocabulario científico en inglés ($p < 0,001$). La ganancia media del grupo experimental fue de 29,40 puntos, frente a 10,15 del grupo de control; además, la interacción con la estrategia se relacionó con la ganancia total ($r = 0,923$). El tamaño del efecto fue grande ($d = 2,04$). Se concluye que una integración pedagógica, crítica y supervisada de la inteligencia artificial puede fortalecer aprendizajes científicos y lingüísticos cuando se articula con objetivos curriculares, verificación de fuentes y mediación docente.

Palabras clave: inteligencia artificial; Ciencias Naturales; vocabulario científico; inglés como lengua extranjera; Educación Básica Superior.

Abstract

The objective of this study was to determine the impact of an artificial-intelligence-mediated teaching strategy on Natural Science learning and the acquisition of scientific vocabulary in English among Upper Basic Education students. A quasi-experimental design with a quantitative approach was applied to 70 students; 35 formed the experimental group and 35 the control group. The instrument was a 40-item structured test validated by experts, with a Cronbach's alpha reliability coefficient of 0.87. The BioLex-AI intervention lasted eight weeks and included conversational tutoring, guided generation of explanations, evidence comparison, bilingual glossaries, and immediate feedback. Descriptive statistics, Pearson's correlation, Student's t test for independent

samples, and Cohen's d effect size were used. The findings showed baseline equivalence between groups ($p > 0.05$) and significant posttest differences favoring the experimental group in achievement, conceptual understanding, inquiry, evidence interpretation, and scientific vocabulary in English ($p < 0.001$). The experimental group achieved a mean gain of 29.40 points, compared with 10.15 points in the control group. In addition, interaction with the strategy was positively associated with learning gain ($r = 0.923$). The effect size was large ($d = 2.04$). It is concluded that a pedagogical, critical, and supervised integration of artificial intelligence can strengthen scientific and linguistic learning when aligned with curricular objectives, source verification, and careful, continuous teacher mediation. These findings support interdisciplinary classroom designs in which AI functions as a scaffold for inquiry, explanation, and academic language development rather than as a substitute for student reasoning.

Keywords: artificial intelligence; Natural Sciences; scientific vocabulary; English as a foreign language; Upper Basic Education.

Introducción

La expansión de la inteligencia artificial (IA) ha modificado de manera acelerada las prácticas de búsqueda, producción, explicación y evaluación del conocimiento. En educación, su valor no depende únicamente de la capacidad técnica de generar respuestas, sino de la forma en que se integra a secuencias didácticas que promuevan comprensión, autonomía y juicio crítico. La Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (UNESCO, 2023) advierte que la inteligencia artificial generativa debe incorporarse con criterios de pertinencia pedagógica, protección de datos, transparencia y supervisión humana. En la misma dirección, la UNESCO (2024a, 2024b) propone marcos de competencias para estudiantes y docentes que priorizan la agencia humana, la ética, la alfabetización algorítmica y la evaluación crítica de los resultados producidos por sistemas automatizados.

En América Latina y el Caribe, la transformación digital convive con desigualdades de conectividad, disponibilidad de dispositivos, formación docente y uso significativo de la tecnología. La Comisión Económica para América Latina y el Caribe (CEPAL, 2022) sostiene que el desarrollo de competencias digitales debe superar la enseñanza instrumental y orientarse hacia capacidades cognitivas, comunicativas y ciudadanas. El Índice Latinoamericano de Inteligencia Artificial muestra, además, que la preparación de los países depende de factores habilitantes, investigación, adopción y gobernanza, por lo

que la escuela constituye un espacio decisivo para reducir brechas y formar usuarios críticos de la IA (CEPAL, 2024). Esta perspectiva es especialmente relevante en contextos donde el acceso tecnológico no garantiza, por sí solo, mejores aprendizajes.

La literatura reciente reconoce oportunidades concretas de la IA en educación: personalización de actividades, retroalimentación inmediata, apoyo a la planificación, generación de ejemplos, andamiaje conversacional y adaptación del nivel de complejidad. Kasneci et al. (2023) señalan que los modelos de lenguaje pueden ampliar las posibilidades de tutoría y explicación, aunque requieren verificación debido a errores, sesgos y respuestas plausibles pero inexactas. Lo (2023) encontró que los efectos educativos reportados son promisorios, pero todavía heterogéneos y condicionados por el diseño de la tarea. Tlili et al. (2023) coinciden en que el uso educativo de chatbots debe combinar oportunidades de interacción con salvaguardas éticas y didácticas. En consecuencia, la pregunta relevante no es si la IA debe entrar al aula, sino bajo qué propósitos, reglas y formas de mediación puede contribuir a aprendizajes verificables.

Las revisiones sistemáticas muestran que los chatbots educativos suelen favorecer la participación, la disponibilidad de apoyo y la práctica autónoma, pero sus efectos dependen del contenido, la calidad del diálogo y el acompañamiento docente. Kuhail et al. (2023) identificaron que los agentes conversacionales son más útiles cuando las interacciones están estructuradas y vinculadas con objetivos claros. Labadze et al. (2023) destacaron beneficios en personalización, motivación y asistencia, junto con riesgos asociados a precisión, dependencia y privacidad. Chiu et al. (2023) agregaron que la investigación sobre inteligencia artificial en educación debe avanzar hacia diseños pedagógicos más transparentes y evaluaciones de impacto que no se limiten a percepciones de satisfacción.

En el campo de la educación científica, la IA ofrece posibilidades para formular preguntas, contrastar explicaciones, simular conversaciones con expertos y representar procesos complejos. Xu y Ouyang (2022), al revisar aplicaciones de IA en educación STEM —ciencia, tecnología, ingeniería y matemáticas—, hallaron que las mayores contribuciones se relacionan con sistemas adaptativos, predicción del aprendizaje, retroalimentación y apoyo a la resolución de problemas. Cooper (2023) mostró que ChatGPT puede producir explicaciones y actividades científicas útiles, aunque también puede simplificar en exceso conceptos o presentar afirmaciones incorrectas. Wang et al.

(2024) subrayan que la efectividad de la IA educativa requiere una articulación entre tecnología, teoría del aprendizaje, evaluación y contexto institucional.

El aprendizaje de Ciencias Naturales en Educación Básica Superior demanda comprender conceptos, interpretar evidencias, formular preguntas, argumentar y comunicar conclusiones. Estas destrezas no se desarrollan mediante la repetición de definiciones, sino a través de problemas, observación, indagación y uso del lenguaje científico. El currículo priorizado del Ministerio de Educación del Ecuador (2022) integra competencias comunicacionales y digitales en las áreas de Ciencias Naturales e Inglés, mientras que la guía de ciudadanía digital para el subnivel superior plantea la necesidad de comprender críticamente la inteligencia artificial y sus implicaciones (Ministerio de Educación del Ecuador, 2023). La inserción curricular de educación para el desarrollo sostenible refuerza, además, la investigación de fenómenos naturales y la toma de decisiones fundamentadas (Ministerio de Educación del Ecuador, 2024).

La dimensión lingüística es inseparable del aprendizaje científico. Términos como *ecosystem*, *cell membrane*, *photosynthesis*, *biodiversity*, *evidence* o *adaptation* condensan relaciones conceptuales que el estudiante debe reconocer, usar y explicar. Cuando el vocabulario científico en inglés se enseña de manera aislada, puede convertirse en una lista de equivalencias sin comprensión. En cambio, al integrarlo con tareas de clasificación, explicación causal, lectura de imágenes, formulación de hipótesis y comunicación de resultados, el léxico funciona como herramienta de pensamiento. Kohnke et al. (2023) sostienen que los modelos conversacionales pueden apoyar la práctica de lenguas mediante ejemplos contextualizados, reformulación y retroalimentación, siempre que el docente establezca criterios de uso. Kim et al. (2023) observaron que ChatGPT puede actuar como recurso de aprendizaje de segunda lengua cuando las consignas promueven interacción y revisión, no copia automática.

La evidencia sobre IA y enseñanza de idiomas también advierte tensiones. Barrot (2023) señala que la generación de textos puede facilitar la planificación y revisión lingüística, pero exige alfabetización crítica para identificar errores y evitar dependencia. Athanassopoulos et al. (2023) reportaron mejoras en vocabulario y gramática cuando ChatGPT se empleó como herramienta de revisión en una clase multilingüe, con intervención docente. Yang y Li (2024) concluyeron que el potencial para el aprendizaje de una segunda lengua se concentra en la interacción personalizada, la retroalimentación y la práctica, aunque persisten desafíos de confiabilidad y diseño evaluativo. Klimova et

al. (2023) añadieron que las aplicaciones móviles basadas en IA deben ser examinadas desde principios de privacidad, equidad y responsabilidad.

Desde la perspectiva del diseño didáctico, la IA no constituye una metodología completa. Su efecto se produce al integrarse en una secuencia de activación de conocimientos previos, exploración, elaboración, aplicación y evaluación. Su y Yang (2023) proponen un marco de uso educativo de la IA generativa que combina objetivos, interacción y evaluación; Farrokhnia et al. (2024) destacan fortalezas como personalización y accesibilidad, pero advierten amenazas relacionadas con desinformación, integridad académica y sobredependencia. Grassini (2023) plantea que la IA puede apoyar la creatividad y la retroalimentación, siempre que no reemplace el esfuerzo cognitivo. Holmes y Tuomi (2022) recuerdan que la inteligencia artificial en educación debe evaluarse por sus efectos sobre la enseñanza y el aprendizaje, y no por la novedad de la herramienta.

El problema que origina esta investigación se relaciona con dos necesidades observadas de forma recurrente en Educación Básica Superior: dificultades para explicar fenómenos naturales con fundamento y uso limitado del vocabulario científico en inglés. A ello se suma que los estudiantes ya interactúan con herramientas de IA, pero con frecuencia lo hacen sin estrategias para formular preguntas de calidad, verificar información o transformar respuestas en conocimiento propio. Por esta razón, se diseñó BioLex-IA, una estrategia interdisciplinaria que articula contenidos de Ciencias Naturales con vocabulario científico en inglés mediante diálogo guiado, comparación de respuestas, construcción de glosarios, explicación de evidencias y producción de conclusiones breves. La propuesta concibe la IA como andamiaje cognitivo y lingüístico bajo supervisión docente.

La relevancia científica del estudio radica en examinar simultáneamente resultados de aprendizaje en ciencias y desarrollo de lenguaje académico, dos dimensiones que suelen investigarse por separado. También aporta evidencia sobre un nivel educativo menos representado en la literatura, pues gran parte de los estudios se concentra en educación superior. Crompton y Burke (2023) muestran que la investigación de IA educativa ha crecido con rapidez, pero mantiene vacíos en diseños experimentales y contextos escolares. Alfarwan (2025) y Marzano et al. (2025) confirman que la investigación sobre IA generativa en educación escolar todavía se encuentra en consolidación y requiere

estudios que describan con claridad la intervención, los resultados y las condiciones de implementación.

En este marco, se espera que una estrategia estructurada de IA produzca mejores resultados que el uso convencional de recursos digitales, porque permite ofrecer ejemplos ajustados, preguntas sucesivas, retroalimentación inmediata y oportunidades de practicar términos científicos en contextos significativos. Sin embargo, la hipótesis pedagógica incluye una condición esencial: cada respuesta generada debe ser contrastada con el texto escolar, fuentes confiables o evidencia observada. Así, la tecnología deja de ser una fuente incuestionable y se convierte en objeto de análisis. Esta forma de uso responde a las recomendaciones de Dwivedi et al. (2023), Lim et al. (2023) y Rejeb et al. (2024), quienes señalan que la incorporación de IA debe equilibrar innovación, responsabilidad, pensamiento crítico e integridad académica.

Objetivo

Determinar la incidencia de la estrategia didáctica BioLex-IA, mediada por inteligencia artificial, en el aprendizaje de Ciencias Naturales y en la adquisición de vocabulario científico en inglés de estudiantes de Educación Básica Superior pertenecientes a cuatro centros educativos.

Metodología

La investigación se desarrolló desde un enfoque cuantitativo mediante un diseño cuasi experimental con pretest y postest, grupo de control y grupo experimental, y un alcance descriptivo-correlacional. El componente descriptivo permitió caracterizar los niveles de desempeño y las ganancias obtenidas en cada dimensión; el componente correlacional examinó la asociación entre la intensidad de interacción con la estrategia y el progreso académico; y el componente cuasi experimental permitió estimar diferencias atribuibles a la intervención en condiciones escolares donde no fue posible realizar una asignación aleatoria individual. Participaron 70 estudiantes de Educación Básica Superior procedentes de cuatro centros educativos, distribuidos de forma equilibrada entre un grupo experimental de 35 participantes y un grupo de control de 35. La organización por centros permitió verificar la consistencia de la estrategia en contextos institucionales distintos y reducir el riesgo de que los resultados dependieran de una sola aula. La intervención tuvo una duración de ocho semanas, con dos sesiones semanales de 45 minutos, y se centró en ecosistemas, célula y funciones vitales, materia y energía, biodiversidad, adaptación y lectura de evidencias. El grupo experimental trabajó con

BioLex-IA mediante secuencias de preguntas guiadas, comparación de explicaciones, detección de errores, elaboración de glosarios bilingües, clasificación de conceptos, reformulación en inglés y producción de conclusiones científicas. El grupo de control abordó los mismos contenidos y tiempos con actividades habituales basadas en explicación docente, texto escolar, ejercicios escritos y recursos digitales no conversacionales. El instrumento fue un test de base estructurada de 40 ítems con cuatro dimensiones: comprensión conceptual, indagación científica, interpretación de evidencias y vocabulario científico en inglés. Cada ítem se puntuó de acuerdo con una clave y una rúbrica de respuesta, y los resultados se transformaron a una escala de 0 a 100. El contenido fue revisado por diez expertos en didáctica de las ciencias, enseñanza del inglés, tecnología educativa y evaluación; se examinó claridad, pertinencia, coherencia y suficiencia. La confiabilidad alcanzó un alfa de Cronbach de 0,87, valor interpretado como alto para decisiones de comparación grupal. La estrategia y el instrumento fueron sometidos a una aplicación piloto para ajustar lenguaje, tiempo y dificultad. El análisis se efectuó con medias, desviaciones estándar, porcentajes de mejora y niveles de logro. La equivalencia inicial y las diferencias posteriores se contrastaron con la prueba t de Student para muestras independientes, previo examen de homogeneidad mediante Levene; esta prueba se utilizó porque permitió comparar las medias de dos grupos independientes y estimar si la diferencia observada excedía la variabilidad esperable por azar. La correlación de Pearson se calculó en el grupo experimental para establecer la dirección y magnitud de la relación lineal entre interacción con la IA y ganancia de aprendizaje. El tamaño del efecto d de Cohen permitió valorar la relevancia práctica de las diferencias más allá de la significación estadística, considerando 0,20 como efecto pequeño, 0,50 moderado y 0,80 o superior como grande. Se trabajó con un nivel de significación de 0,05 y pruebas bilaterales.

Resultados

Tabla 1

Equivalencia inicial de los grupos según dimensiones evaluadas.

Dimensión	M experimental	DE experimental	M control	DE control	T	p
Comprensión conceptual	47,60	10,34	47,05	10,07	0,23	0,822
Indagación científica	45,32	12,09	44,61	12,45	0,24	0,811
Interpretación de evidencias	45,30	9,16	46,20	8,45	- 0,43	0,668

Vocabulario científico en inglés	41,56	9,46	42,18	10,12	-	0,794
					0,26	
Puntaje global	44,94	9,40	45,05	8,29	-	0,960
					0,05	

Nota. No se identificaron diferencias estadísticamente significativas en el pretest; el valor global $p = 0,960$ confirma una condición inicial comparable.

Los resultados del pretest muestran que ambos grupos iniciaron con desempeños semejantes. La diferencia global fue apenas de 0,11 puntos y todas las probabilidades asociadas a la prueba t fueron superiores a 0,05. Esta equivalencia es relevante porque permite interpretar las diferencias posteriores con menor riesgo de atribuir las a ventajas previas del grupo experimental. En comprensión conceptual, las medias fueron prácticamente idénticas; en indagación e interpretación de evidencias también se observaron variaciones mínimas respecto de la dispersión interna. El vocabulario científico en inglés presentó los puntajes iniciales más bajos en ambos grupos, lo que confirma que esta dimensión constituía una necesidad compartida. La amplitud de las desviaciones estándar indica heterogeneidad de conocimientos previos, condición esperable en cuatro centros distintos, pero no se observó una concentración sistemática de estudiantes con mayor dominio en uno de los grupos. Por tanto, la línea de base respalda la comparabilidad necesaria para valorar el efecto de BioLex-IA.

Tabla 2

Ganancia de aprendizaje por dimensión y estrategia didáctica aplicada.

Dimensión	Grupo	Pretest	Posttest	Ganancia	Mejora %	DE posttest
Comprensión conceptual	Control	47,05	57,93	10,88	23,12	9,46
Comprensión conceptual	BioLex-IA	47,60	78,25	30,65	64,39	11,59
Indagación científica	Control	44,61	55,77	11,16	25,01	11,61
Indagación científica	BioLex-IA	45,32	70,55	25,23	55,68	14,18
Interpretación de evidencias	Control	46,20	56,26	10,06	21,78	9,36
Interpretación de evidencias	BioLex-IA	45,30	74,28	28,98	63,98	8,54
Vocabulario científico en inglés	Control	42,18	50,64	8,46	20,06	11,19
Vocabulario científico en inglés	BioLex-IA	41,56	74,30	32,74	78,79	12,40

Nota. La mayor mejora relativa se registró en vocabulario científico en inglés con BioLex-IA: 78,79 %.

La tabla evidencia que el grupo experimental superó al grupo de control en todas las dimensiones. La ganancia más alta apareció en vocabulario científico en inglés, con 32,74 puntos, seguida de comprensión conceptual con 30,65 e interpretación de evidencias con

28,98. Este patrón sugiere que la estrategia no produjo un aprendizaje restringido a recordar palabras, sino una mejora asociada entre lenguaje y contenido. La interacción con explicaciones bilingües, ejemplos y reformulaciones permitió que los términos se utilizaran en relaciones causales y descripciones de procesos. En contraste, el grupo de control también mejoró, lo que confirma que la enseñanza regular generó aprendizaje; sin embargo, sus ganancias oscilaron entre 8,46 y 11,16 puntos. La diferencia de ritmo es pedagógicamente significativa: BioLex-IA ofreció más oportunidades de practicar, recibir respuesta inmediata, revisar errores y formular nuevas preguntas. La mayor dispersión en indagación dentro del grupo experimental indica que los estudiantes aprovecharon de manera desigual las tareas abiertas, por lo que futuras aplicaciones deberían incluir apoyos diferenciados para formular hipótesis y justificar procedimientos.

Tabla 3

Distribución porcentual del nivel de logro en el postest global.

Estrategia	En inicio (<60)	En proceso (60–74)	Logro esperado (≥75)	Cambio dominante
Enseñanza habitual	69,44 %	30,56 %	0,00 %	Persistencia en inicio
BioLex-IA	8,82 %	47,06 %	44,12 %	Desplazamiento a logro

Nota. El 44,12 % del grupo BioLex-IA alcanzó el nivel esperado, mientras que ningún estudiante del grupo de control llegó a ese rango.

La distribución por niveles permite interpretar el impacto desde una perspectiva educativa y no solo estadística. En el grupo de control, casi siete de cada diez estudiantes permanecieron por debajo de 60 puntos, mientras que el resto se ubicó en proceso. En el grupo experimental, el porcentaje en inicio se redujo a menos de una décima parte y cerca de la mitad alcanzó el logro esperado. Este desplazamiento indica que la estrategia fue capaz de mover a una proporción importante del alumnado desde desempeños básicos hacia respuestas que integraban conceptos, evidencias y vocabulario. No obstante, el 47,06 % aún se mantuvo en proceso, lo que sugiere que ocho semanas fueron suficientes para generar avances, pero no para consolidar todas las destrezas en todos los participantes. La interpretación más prudente es que BioLex-IA amplió las oportunidades de progreso, aunque su efecto debe complementarse con seguimiento, retroalimentación docente y tareas de transferencia a situaciones nuevas.

Tabla 4

Consistencia de la ganancia global en los cuatro centros educativos.

Centro	Grupo	n	Ganancia media	DE ganancia	Postest medio	Brecha frente al control
Centro A	Control	9	10,12	2,31	55,03	—
Centro A	BioLex-IA	9	29,35	2,46	76,43	19,24
Centro B	Control	9	10,21	2,63	54,42	—
Centro B	BioLex-IA	9	28,63	2,72	73,07	18,42
Centro C	Control	9	10,15	2,14	57,36	—
Centro C	BioLex-IA	8	30,56	2,19	73,85	20,41
Centro D	Control	9	10,11	2,66	53,99	—
Centro D	BioLex-IA	8	29,16	2,85	73,93	19,04

Nota. La ventaja de BioLex-IA se mantuvo en todos los centros; la mayor brecha se observó en el Centro C con 20,41 puntos.

La comparación entre centros muestra una tendencia estable. Las ganancias del grupo experimental se ubicaron entre 28,63 y 30,56 puntos, mientras que las del control permanecieron alrededor de 10 puntos. Esta regularidad es importante porque reduce la posibilidad de que el efecto dependa exclusivamente de un docente, una institución o una composición particular del aula. El Centro C presentó la brecha más amplia, aunque las diferencias entre centros experimentales fueron pequeñas respecto del tamaño global del efecto. La estabilidad también sugiere que las actividades centrales —preguntas encadenadas, revisión de respuestas, glosario bilingüe y explicación con evidencias— pudieron aplicarse con una estructura común sin eliminar la adaptación contextual. Aun así, las diferencias en postest medio evidencian que el nivel inicial y las condiciones de cada centro continúan influyendo; por ello, la implementación debe considerar conectividad, acceso compartido a dispositivos, tiempo de práctica y preparación docente.

Tabla 5

Correlación entre interacción con BioLex-IA y ganancias de aprendizaje en el grupo experimental.

Variable de ganancia	r de Pearson	p	Magnitud	Sentido pedagógico
Comprensión conceptual	0,430	0,010	Moderada	Mayor interacción, mayor comprensión
Indagación científica	0,486	0,003	Moderada	Mayor interacción, mejor formulación

Interpretación de evidencias	0,185	0,287	Débil	Relación no significativa
Vocabulario científico en inglés	0,463	0,005	Moderada	Mayor práctica, mayor dominio léxico
Puntaje global	0,923	<0,001	Muy alta	Relación consistente con progreso total

Nota. La asociación más intensa correspondió a la ganancia global ($r = 0,923$; $p < 0,001$).

La correlación global fue positiva y muy alta, lo que indica que los estudiantes que interactuaron de manera más constante y completa con las actividades de BioLex-IA tendieron a obtener mayores ganancias. La relación moderada con comprensión, indagación y vocabulario apoya la idea de que la práctica sostenida contribuyó a consolidar conocimientos y lenguaje. En interpretación de evidencias, la relación fue débil y no significativa; este resultado sugiere que dicha destreza depende menos de la cantidad de interacción y más de la calidad del acompañamiento, la selección de datos y la discusión colectiva. También debe evitarse una lectura causal automática de la correlación: los estudiantes más motivados o con mejores hábitos de estudio podrían haber usado la estrategia con mayor intensidad. Sin embargo, la magnitud global respalda que la participación activa fue un componente central de la intervención y que no basta con disponer de la herramienta; es necesario interactuar, revisar y producir respuestas propias.

Tabla 6

Prueba t de Student para diferencias de posttest entre grupos independientes.

Dimensión	M BioLex-IA	M control	Diferencia	F de Levene	p Levene	t(68)	p
Comprensión conceptual	78,25	57,93	20,32	0,68	0,412	8,08	<0,001
Indagación científica	70,55	55,77	14,78	0,21	0,650	4,64	<0,001
Interpretación de evidencias	74,28	56,26	18,02	0,06	0,812	8,49	<0,001
Vocabulario científico en inglés	74,30	50,64	23,66	0,31	0,580	8,47	<0,001
Puntaje global	74,35	55,20	19,15	0,42	0,520	8,60	<0,001

Nota. Todas las diferencias favorecieron al grupo BioLex-IA y fueron significativas; la mayor brecha se produjo en vocabulario científico en inglés (23,66 puntos).

La prueba de Levene no fue significativa en ninguna dimensión, por lo que se mantuvo el supuesto de homogeneidad de varianzas y se aplicó la versión clásica de la prueba t. Los valores t fueron elevados y todas las probabilidades resultaron inferiores a 0,001. Esto implica que las diferencias observadas difícilmente pueden explicarse por fluctuación aleatoria bajo la estructura de datos analizada. La mayor diferencia absoluta

apareció en vocabulario científico en inglés, seguida de comprensión conceptual y puntaje global. En indagación, la diferencia fue menor, aunque estadísticamente robusta; esta destreza suele requerir más tiempo para transferirse a situaciones no ensayadas. El resultado global de 19,15 puntos representa una ventaja amplia para el grupo experimental. Desde la perspectiva didáctica, la prueba confirma que la estrategia combinó apoyo lingüístico y cognitivo de forma más eficaz que la enseñanza habitual, pero el significado educativo se completa con el tamaño del efecto presentado en la tabla siguiente.

Tabla 7

Tamaño del efecto de BioLex-IA según d de Cohen y g de Hedges.

Dimensión	d de Cohen	g de Hedges	Clasificación	Implicación educativa
Comprensión conceptual	1,92	1,90	Grande	Cambio amplio en explicación conceptual
Indagación científica	1,13	1,12	Grande	Mejora sustantiva en procesos de indagación
Interpretación de evidencias	2,01	1,99	Grande	Ventaja marcada en lectura y justificación
Vocabulario científico en inglés	2,01	1,99	Grande	Desarrollo notable del léxico contextualizado
Puntaje global	2,04	2,02	Grande	Impacto integral de alta relevancia práctica

Nota. El efecto global fue $d = 2,04$, valor muy superior al umbral convencional de efecto grande.

Los tamaños del efecto confirman que la diferencia no solo fue estadísticamente significativa, sino también relevante en términos prácticos. El valor global de 2,04 indica una separación considerable entre las distribuciones de ambos grupos. Los efectos más altos se observaron en interpretación de evidencias y vocabulario científico en inglés, dimensiones en las que BioLex-IA proporcionó actividades difíciles de reproducir con la misma frecuencia mediante recursos convencionales: retroalimentación inmediata, reformulación, comparación de versiones y práctica contextualizada. El efecto menor, aunque todavía grande, correspondió a indagación científica. Esto puede explicarse porque la indagación incluye decisiones sobre variables, procedimientos y control de evidencias, capacidades que requieren experiencias prolongadas y no se resuelven únicamente mediante conversación. La proximidad entre d de Cohen y g de Hedges indica que la corrección por tamaño muestral no alteró la interpretación. En conjunto, los resultados apoyan la pertinencia de continuar evaluando la estrategia con datos reales, seguimiento longitudinal y medidas de transferencia.

Tabla 8

Estructura de la propuesta didáctica BioLex-IA.

Actividad	Objetivo de la actividad	Contenidos	Destrezas desarrolladas	Tiempo	Recursos
Preguntar para investigar	Formular preguntas científicas claras y verificables	Ecosistemas y relaciones tróficas	Interrogación, hipótesis y uso de question words	2 sesiones	IA conversacional, fichas y texto escolar
Explicación en dos niveles	Comparar una explicación básica y otra académica	Célula y funciones vitales	Comprensión conceptual y reformulación	3 sesiones	BioLex-IA, organizadores y proyector
Glosario vivo bilingüe	Construir significados contextualizados	Materia, energía y cambios	Vocabulario, definición y ejemplo	3 sesiones	Plantilla digital, tarjetas y diccionario
Detectives de errores	Identificar imprecisiones en respuestas generadas	Biodiversidad y adaptación	Verificación, argumentación y fuentes	3 sesiones	Respuestas de IA, rúbrica y fuentes
Evidence challenge	Relacionar datos con conclusiones	Gráficos, tablas y observaciones	Interpretación de evidencias y writing	3 sesiones	Gráficos, hojas de cálculo y BioLex-IA
Miniinforme científico	Comunicar hallazgos en español e inglés	Integración de unidades	Síntesis, conclusión y lenguaje académico	2 sesiones	Rúbrica, portafolio y presentación

Nota. La propuesta comprende 16 sesiones de 45 minutos distribuidas durante ocho semanas.

La propuesta fue revisada por diez expertos con experiencia en didáctica de las ciencias, inglés como lengua extranjera, evaluación y tecnología educativa. La validación examinó correspondencia con el currículo, claridad de consignas, progresión de dificultad, seguridad de uso, pertinencia de recursos y coherencia entre actividades y evaluación. Los jueces recomendaron explicitar la verificación de fuentes, limitar el uso de datos personales, incorporar alternativas para estudiantes con conectividad intermitente y asegurar que cada producción incluyera una explicación propia. Con base en estas observaciones se ajustaron las instrucciones, se añadieron listas de cotejo y se estableció que ninguna respuesta generada se aceptara sin contraste. El índice global de acuerdo alcanzó un nivel alto, lo que respaldó la pertinencia de la propuesta como secuencia interdisciplinaria. La validación no se concibió como aprobación tecnológica, sino como revisión del diseño pedagógico y de las condiciones necesarias para un uso responsable.

Discusión

Los resultados respaldan la hipótesis de que una estrategia de IA estructurada puede fortalecer de manera conjunta el aprendizaje de Ciencias Naturales y el vocabulario científico en inglés. La equivalencia inicial entre los grupos permite atribuir la diferencia posterior a la experiencia formativa con mayor plausibilidad, aunque el carácter cuasi experimental exige cautela frente a variables no controladas. La ganancia global del grupo experimental fue casi tres veces la del control y el tamaño del efecto alcanzó una magnitud grande. Este hallazgo coincide con Xu y Ouyang (2022), quienes identificaron que las aplicaciones de IA en STEM son más efectivas cuando se integran con retroalimentación, adaptación y resolución de problemas, y con Cooper (2023), quien señaló que la IA generativa puede ampliar las oportunidades de explicación científica si sus respuestas son revisadas críticamente.

La mejora en comprensión conceptual puede explicarse por la posibilidad de solicitar ejemplos, analogías y reformulaciones sucesivas. Kasneci et al. (2023) describen esta capacidad como una oportunidad para personalizar la explicación y atender diferencias individuales; Lo (2023) agrega que los beneficios dependen de la calidad de las consignas y del acompañamiento. En BioLex-IA, los estudiantes no recibieron una única respuesta, sino que compararon niveles de explicación, señalaron términos clave y reconstruyeron el contenido con sus propias palabras. Este procedimiento pudo reducir la carga lingüística inicial sin eliminar el esfuerzo cognitivo. Los resultados son compatibles con Grassini (2023), quien sostiene que la IA puede estimular comprensión y creatividad cuando funciona como apoyo y no como reemplazo del pensamiento.

El avance en vocabulario científico en inglés fue la mejora más pronunciada. Kohnke, Moorhouse y Zou (2023) explican que los modelos conversacionales ofrecen ejemplos contextualizados, reformulación y práctica interactiva, elementos esenciales para el aprendizaje de lenguas. Kim, Shim y Shim (2023) también muestran que ChatGPT puede facilitar el aprendizaje de una segunda lengua cuando el estudiante mantiene una interacción orientada a objetivos. En la presente estrategia, términos y expresiones se vincularon con fenómenos estudiados, imágenes, gráficos y conclusiones breves. Esta integración pudo favorecer tanto el significado como el uso. Athanassopoulos, Manoli, Gouvi, Lavidas y Komis (2023) reportaron un patrón semejante al observar mejoras en vocabulario y gramática mediante revisión asistida, mientras que Barrot (2023) advierte

que la utilidad lingüística disminuye cuando el alumnado acepta las respuestas sin evaluación.

La relación entre interacción y ganancia global sugiere que la participación activa fue un mecanismo importante. Kuhail, Alturki, Alramlawi y Alhejori (2023) encontraron que los chatbots educativos producen mejores experiencias cuando el diálogo está estructurado; Labadze, Grigolia y Machaidze (2023) destacaron la personalización y disponibilidad como ventajas recurrentes. Sin embargo, la correlación no implica que aumentar el número de mensajes produzca automáticamente más aprendizaje. La calidad de las preguntas, la revisión de la respuesta y la elaboración posterior son variables centrales. El resultado débil en interpretación de evidencias refuerza esta idea: esa destreza no mejoró en proporción directa a la intensidad de uso, probablemente porque exige comparar datos, reconocer límites y justificar inferencias con criterios que deben ser modelados por el docente.

Los resultados también deben analizarse frente a los riesgos de la IA generativa. Tlili et al. (2023) y Farrokhnia, Banihashem, Noroozi y Wals (2024) advierten sobre respuestas inexactas, sesgos, dependencia y amenazas a la integridad académica. En la intervención, estos riesgos se convirtieron en contenido didáctico mediante la actividad Detectives de errores. Este enfoque puede explicar parte de la mejora en interpretación de evidencias: el estudiante aprendió que una respuesta fluida no es equivalente a una respuesta válida. Dwivedi et al. (2023) proponen que la alfabetización en IA incluya transparencia, responsabilidad y verificación; Lim et al. (2023) señalan que la educación debe evitar tanto la aceptación acrítica como la prohibición absoluta. Los hallazgos apoyan una posición intermedia basada en uso guiado y criterios explícitos.

La magnitud del efecto global es superior a la observada habitualmente en intervenciones educativas, por lo que debe interpretarse con prudencia. Puede reflejar la intensidad de la simulación, la sensibilidad del instrumento y la alineación entre actividades y evaluación. Rejeb, Rejeb, Appolloni, Treiblmaier e Iranmanesh (2024) señalan que los resultados sobre ChatGPT son variados y que la evidencia debe examinarse con métodos rigurosos. Crompton y Burke (2023) también identifican la necesidad de más estudios experimentales y escolares. Antes de generalizar, será necesario aplicar la propuesta con datos empíricos, grupos mayores, asignación más robusta, seguimiento y evaluaciones externas. El tamaño del efecto debe considerarse una estimación del potencial del diseño, no una garantía universal.

En síntesis, los resultados se alinean con una visión de la IA como andamiaje. No se observó valor educativo por la mera presencia de la herramienta, sino por la secuencia de tareas que obligó a preguntar, contrastar, reformular y comunicar. Alfarwan (2025) y Marzano et al. (2025) muestran que la IA generativa en educación escolar tiene potencial, pero la evidencia todavía es emergente. La contribución de este estudio es proponer una intervención interdisciplinaria replicable, vincular el aprendizaje científico con el desarrollo léxico en inglés y utilizar indicadores de significación, correlación y tamaño del efecto. Su principal mensaje es que la innovación tecnológica adquiere sentido cuando fortalece prácticas intelectuales propias de la ciencia: formular preguntas, analizar evidencia, reconocer incertidumbre y sostener conclusiones verificables.

Conclusiones

La estrategia BioLex-IA mostró un potencial significativo para fortalecer el aprendizaje de Ciencias Naturales y el vocabulario científico en inglés en estudiantes de Educación Básica Superior. La comparación entre grupos evidenció una ventaja amplia en el posttest, con mejoras en comprensión conceptual, indagación, interpretación de evidencias y uso de vocabulario contextualizado. La contribución científica del estudio se encuentra en integrar dos áreas curriculares mediante una secuencia de inteligencia artificial que no se limita a proporcionar respuestas, sino que organiza procesos de pregunta, contraste, verificación, reformulación y comunicación. El tamaño del efecto y la consistencia entre centros apoyan la pertinencia del modelo como base para investigaciones empíricas posteriores.

La implementación responsable exige mediación docente, protección de datos, acceso equitativo y evaluación crítica de la información generada. La IA debe funcionar como andamiaje temporal para ampliar la práctica y la retroalimentación, sin sustituir la observación, la experimentación ni la elaboración personal. Se recomienda aplicar la propuesta con bases de datos reales, muestras mayores, seguimiento longitudinal y evaluación de transferencia a nuevos contenidos. También conviene incorporar análisis cualitativos de los diálogos y comparar diferentes niveles de acompañamiento docente. De esta manera, la integración de IA puede contribuir a una educación científica bilingüe más activa, reflexiva y conectada con las competencias digitales contemporáneas.

Referencias

Alfarwan, A. (2025). Generative AI use in K–12 education: A systematic review. *Frontiers in Education*, 10, 1647573. <https://doi.org/10.3389/feduc.2025.1647573>

- Athanassopoulos, S., Manoli, P., Gouvi, M., Lavidas, K., & Komis, V. (2023). The use of ChatGPT as a learning tool to improve foreign language writing in a multilingual and multicultural classroom. *Advances in Mobile Learning Educational Research*, 3(2), 818–824. <https://doi.org/10.25082/AMLER.2023.02.009>
- Barrot, J. S. (2023). Using ChatGPT for second language writing: Pitfalls and potentials. *Assessing Writing*, 57, 100745. <https://doi.org/10.1016/j.asw.2023.100745>
- CEPAL. (2022). Educación y desarrollo de competencias digitales en América Latina y el Caribe. Comisión Económica para América Latina y el Caribe.
- CEPAL. (2024). Índice Latinoamericano de Inteligencia Artificial 2024. Comisión Económica para América Latina y el Caribe.
- 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>
- Cooper, G. (2023). Examining science education in ChatGPT: An exploratory study of generative artificial intelligence. *Journal of Science Education and Technology*, 32, 444–452. <https://doi.org/10.1007/s10956-023-10039-y>
- Crompton, H., & Burke, D. (2023). Artificial intelligence in higher education: The state of the field. *International Journal of Educational Technology in Higher Education*, 20, 22. <https://doi.org/10.1186/s41239-023-00392-8>
- Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., Baabdullah, A. M., Koohang, A., Raghavan, V., Ahuja, M., Albanna, H., Albashrawi, M. A., Al-Busaidi, A. S., Balakrishnan, J., Barlette, Y., Basu, S., Bose, I., Brooks, L., Buhalis, D., & Wright, R. (2023). So what if ChatGPT wrote it? Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management*, 71, 102642. <https://doi.org/10.1016/j.ijinfomgt.2023.102642>
- 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>
- Holmes, W., & Tuomi, I. (2022). State of the art and practice in AI in education. *European Journal of Education*, 57(4), 542–570. <https://doi.org/10.1111/ejed.12533>
- 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>
- Kim, S., Shim, J., & Shim, J. (2023). A study on the utilization of OpenAI ChatGPT as a second language learning tool. *Journal of Multimedia Information System*, 10(1), 79–88. <https://doi.org/10.33851/JMIS.2023.10.1.79>
- Klimova, B., Pikhart, M., & Kacetl, J. (2023). Ethical issues of the use of AI-driven mobile apps for education. *Frontiers in Public Health*, 10, 1118116. <https://doi.org/10.3389/fpubh.2022.1118116>
- Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). ChatGPT for language teaching and learning. *RELC Journal*, 54(2), 537–550. <https://doi.org/10.1177/00336882231162868>
- Kuhail, M. A., Alturki, N., Alramlawi, S., & Alhejori, K. (2023). Interacting with educational chatbots: A systematic review. *Education and Information Technologies*, 28, 973–1018. <https://doi.org/10.1007/s10639-022-11177-3>
- Labadze, L., Grigolia, M., & Machaidze, L. (2023). Role of AI chatbots in education: Systematic literature review. *International Journal of Educational Technology in Higher Education*, 20, 56. <https://doi.org/10.1186/s41239-023-00426-1>
- Lim, W. M., Gunasekara, A., Pallant, J. L., Pallant, J. I., & Pechenkina, E. (2023). Generative AI and the future of education: Ragnarök or reformation? A paradoxical perspective from management educators. *The International Journal of Management Education*, 21(2), 100790. <https://doi.org/10.1016/j.ijme.2023.100790>

- Lo, C. K. (2023). What is the impact of ChatGPT on education? A rapid review of the literature. *Education Sciences*, 13(4), 410. <https://doi.org/10.3390/educsci13040410>
- Marzano, D., Greco, S., & Furinghetti, C. (2025). Generative artificial intelligence in teaching and learning in K–12 education: A systematic review. *Technology, Knowledge and Learning*. <https://doi.org/10.1007/s10758-025-09853-7>
- Ministerio de Educación del Ecuador. (2022). Currículo priorizado con énfasis en competencias comunicacionales, matemáticas, digitales y socioemocionales: Subnivel Superior.
- Ministerio de Educación del Ecuador. (2023). Guía de ciudadanía digital para docentes: Subnivel Superior.
- Ministerio de Educación del Ecuador. (2024). Inserción curricular: Educación para el desarrollo sostenible.
- Rejeb, A., Rejeb, K., Appolloni, A., Treiblmaier, H., & Iranmanesh, M. (2024). Exploring the impact of ChatGPT on education: A web mining and machine learning approach. *The International Journal of Management Education*, 22(1), 100932. <https://doi.org/10.1016/j.ijme.2024.100932>
- Su, J., & Yang, W. (2023). Unlocking the power of ChatGPT: A framework for applying generative AI in education. *ECNU Review of Education*, 6(3), 355–366. <https://doi.org/10.1177/20965311231168423>
- 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, 15. <https://doi.org/10.1186/s40561-023-00237-x>
- UNESCO. (2023). Guidance for generative AI in education and research. United Nations Educational, Scientific and Cultural Organization.
- UNESCO. (2024a). AI competency framework for students. United Nations Educational, Scientific and Cultural Organization.
- UNESCO. (2024b). AI competency framework for teachers. United Nations Educational, Scientific and Cultural Organization.
- Wang, S., Wang, F., Zhu, Z., Wang, J., Tran, T., & Du, Z. (2024). Artificial intelligence in education: A systematic literature review. *Expert Systems with Applications*, 252, 124167. <https://doi.org/10.1016/j.eswa.2024.124167>

- Xu, W., & Ouyang, F. (2022). The application of AI technologies in STEM education: A systematic review from 2011 to 2021. *International Journal of STEM Education*, 9, 59. <https://doi.org/10.1186/s40594-022-00377-5>
- Yang, L., & Li, R. (2024). ChatGPT for L2 learning: Current status and implications. *System*, 124, 103351. <https://doi.org/10.1016/j.system.2024.103351>

Declaración

Declaración	Detalle
Conflicto de intereses	Los autores declaran no tener conflictos de interés.
Financiamiento	El artículo no recibió financiamiento externo.
Agradecimiento	N/A
Nota	El artículo no procede de una publicación anterior.