

Efecto del uso pedagógico de ChatGPT en la adquisición de vocabulario en estudiantes de 10.º de básica: un estudio mixto con medidas repetidas.

Effect of the Pedagogical Use of ChatGPT on Vocabulary Acquisition in 10th-Grade Basic Education Students: A Mixed Study with Repeated Measures.

Coraima Lizbeth Armijos-Guamán¹, Isabel Petita Guamán-Villa², Ruth Elizabeth Peñafiel-Jurado³ y Carlos Iván Andrade-Bayona⁴

¹ Institución Fiscal Libertador Bolívar, provincia del Azuay.

² Institución Fiscal Sevilla de Oro, provincia del Azuay.

³ Institución Fiscal Miguel Moreno Ordoñez, provincia del Azuay.

⁴ Asesoría en Investigación Académica (AiA), Cuenca-Ecuador.

Correspondencia: carlosivanandradebayona@gmail.com

Recepción: 19 de mayo de 2026 – **Aceptación:** 9 de agosto de 2026 – **Publicación:** 15 de agosto de 2026.

RESUMEN

Se evaluó el efecto del uso pedagógico de ChatGPT en la adquisición de vocabulario receptivo en estudiantes de décimo año de Educación General Básica de una institución rural ecuatoriana. Se aplicó un diseño cuasi experimental de grupos no equivalentes, con pretest y posttest, en una muestra de 15 estudiantes: siete en el grupo experimental y ocho en el grupo de control. El grupo experimental trabajó durante una semana con actividades guiadas mediante ChatGPT, orientadas al reconocimiento de significados, uso contextualizado de palabras, sinónimos, antónimos y producción breve; el grupo de control continuó con enseñanza habitual. Los resultados mostraron incrementos leves en ambos grupos. El grupo experimental pasó de una media de 21,00 a 21,43 puntos, mientras que el grupo de control pasó de 20,88 a 22,00 puntos. Las pruebas de Wilcoxon y U de Mann-Whitney no evidenciaron diferencias

estadísticamente significativas ni dentro de los grupos ni entre ellos. Se concluye que, en esta aplicación breve y contextualizada, ChatGPT no produjo una mejora diferencial significativa frente a la enseñanza habitual, aunque tampoco se observaron retrocesos en el grupo experimental.

Palabras clave: inteligencia artificial; ChatGPT; vocabulario receptivo; educación básica; enseñanza de lenguas.

ABSTRACT

This study evaluated the effect of the pedagogical use of ChatGPT on receptive vocabulary acquisition among tenth-grade students in a rural Ecuadorian school. A quasi-experimental design with non-equivalent groups, pretest and posttest was applied to a sample of 15 students: seven in the experimental group and eight in the control group. The experimental group worked for one week with guided activities supported by ChatGPT, focused on meaning recognition, contextualized word use, synonyms, antonyms, and short written production; the control group continued with regular instruction. The results showed slight increases in both groups. The experimental group moved from a mean score of 21.00 to 21.43 points, while the control group increased from 20.88 to 22.00 points. Wilcoxon and Mann-Whitney U tests did not show statistically significant differences either within groups or between them. It is concluded that, in this brief and contextualized application, ChatGPT did not produce a significant differential improvement compared with regular instruction, although no decline was observed in the experimental group.

Keywords: artificial intelligence; ChatGPT; receptive vocabulary; basic education; language teaching.

INTRODUCCIÓN

La adquisición de vocabulario es condición decisiva para la comprensión lectora, la producción escrita y la participación académica en las distintas áreas del currículo. En décimo año de Educación General Básica (EGB), esta dimensión tiene principal importancia porque el estudiante además de reconocer palabras frecuentes, debe inferir significados por contexto, distinguir matices

semánticos y emplear el léxico con precisión en producciones breves.

En instituciones rurales del Ecuador, este desafío se intensifica por desigualdades asociadas al acceso a conectividad, dispositivos, recursos didácticos y experiencias sostenidas de alfabetización digital. Por ello, cualquier intervención apoyada en inteligencia artificial debe analizarse como una mediación pedagógica situada. Alvarado et al. (2025) advierten que la educación digital rural en Ecuador continúa atravesada por brechas de infraestructura y formación, mientras que Morocho et al. (2023) subrayan que la integración de IA en el sistema educativo ecuatoriano exige condiciones institucionales, técnicas y docentes que todavía son desiguales.

El vocabulario receptivo puede entenderse como la capacidad de reconocer y comprender palabras en diversos contextos comunicativos. Esta competencia implica relacionar términos con campos semánticos, interpretar usos contextualizados y seleccionar significados adecuados según la situación. Luo et al. (2021) muestran que la amplitud y la profundidad del vocabulario inciden de manera diferenciada en la comprensión, dependiendo del tipo de tarea empleada para medirla.

En este contexto, ChatGPT aparece como una herramienta con potencial para generar ejemplos, reformulaciones, ejercicios graduados, retroalimentación inmediata y situaciones de práctica léxica adaptadas al nivel del estudiante. No obstante, su valor educativo reside en el diseño pedagógico que organiza su uso. Kohnke et al. (2023) sostienen que ChatGPT puede apoyar la enseñanza de lenguas, aunque requiere orientación docente, criterios éticos y competencias digitales. De modo similar, Zhang y Tur (2024) señalan que en educación K-12 puede favorecer la personalización y la producción de materiales, pero su integración debe atender la calidad de las respuestas, la integridad académica y el rol activo del profesorado. En términos más amplios, Chounta et al. (2021) y Kim y Kim (2022) coinciden en que la IA puede apoyar la planificación y la enseñanza, siempre que no desplace el juicio profesional del docente.

La literatura reciente muestra, sin embargo, un campo todavía en consolidación. Li et al. (2024), en una revisión sistemática sobre ChatGPT y educación lingüística, identifican un crecimiento

acelerado de estudios empíricos, aunque con predominio de investigaciones en educación superior y menor presencia de trabajos escolares, longitudinales o situados en contextos rurales. En el ámbito del aprendizaje lexical, Mugableh (2024) encontró efectos positivos de ejercicios generados con ChatGPT en estudiantes EFL, mientras que Abdelhalim y Alsehibany (2025) reportaron mejoras en aprendizaje y retención de vocabulario. En Ecuador, Durazno et al. (2025) documentaron avances en adquisición de vocabulario en estudiantes que utilizaron ChatGPT frente a enseñanza tradicional, además de percepciones favorables sobre la experiencia. Estos hallazgos sugieren posibilidades relevantes, pero todavía requieren mayor validación en escenarios escolares concretos.

La incorporación de IA en educación secundaria también demanda cautela. Pratama et al. (2023) y Sumakul et al. (2022) destacan beneficios relacionados con la personalización y el acceso a recursos, pero Akgun y Greenhow (2022), McGrath et al. (2023) y Nazaretsky et al. (2022) advierten riesgos vinculados con privacidad, sesgos, dependencia tecnológica y confianza acrítica en sistemas automatizados. En el contexto ecuatoriano, Márquez et al. (2025) encontraron que los docentes valoran la IA por su utilidad para planificar, motivar y adaptar materiales, aunque también identifican riesgos como debilitamiento del pensamiento crítico, erosión de la autonomía estudiantil y falta de condiciones institucionales para una integración equitativa. Esta lectura coincide con Peñafiel et al. (2024), quienes plantean que el éxito de la IA educativa depende menos de su sofisticación técnica que de su apropiación ética, contextual y pedagógicamente orientada.

Estas advertencias son especialmente importantes cuando se trabaja con adolescentes. Darvishi et al. (2024) muestran que la asistencia de IA puede afectar la agencia estudiantil si los alumnos se limitan a seguir sugerencias sin construir estrategias propias. Zhai et al. (2024) asocian la sobredependencia de sistemas conversacionales con menor retención, menor resolución autónoma de problemas y debilitamiento de habilidades críticas. Walter (2024) insiste, además, en que la alfabetización en IA debe incluir pensamiento crítico y capacidad para evaluar las respuestas generadas. En consecuencia, el uso pedagógico de ChatGPT para

vocabulario debe entenderse como práctica guiada, en la que el docente selecciona palabras, regula la dificultad, promueve ejemplos contextualizados y verifica que el estudiante comprenda y use el léxico con pertinencia. García (2024) confirma esta necesidad al mostrar que ChatGPT puede generar materiales útiles para el trabajo lexical, aunque sus respuestas requieren revisión y ajuste didáctico.

A partir de la problemática descrita, el presente artículo tiene como objetivo evaluar el efecto del uso pedagógico de ChatGPT en la adquisición de vocabulario receptivo en estudiantes de décimo año de Educación General Básica de una institución rural ecuatoriana. Para ello, se considera una aplicación inicial y una aplicación posterior de un instrumento de vocabulario, entre las cuales se implementó una estrategia pedagógica con ChatGPT en el grupo experimental, mientras el grupo de control continuó con enseñanza habitual. El estudio busca determinar si esta mediación produce cambios diferenciados en el desempeño léxico y, a la vez, comprender las percepciones estudiantiles sobre una experiencia de aprendizaje en la que la IA funciona como apoyo, pero no como reemplazo de la mediación docente.

METODOLOGÍA

El estudio se desarrolló con un enfoque cuantitativo, con diseño cuasi experimental de grupos no equivalentes, pretest-postest y medidas repetidas, el cual resulta pertinente cuando no es posible asignar aleatoriamente a los participantes, pero sí comparar el cambio observado entre un grupo que recibe una intervención y otro que continúa con la enseñanza habitual (Shadish et al., 2002).

La medición antes y después permitió estimar la variación intraindividual, mientras que la comparación entre grupos permitió aproximarse al efecto diferencial de la estrategia pedagógica. Dado que se trabajó con una muestra pequeña y natural de aula, el estudio se asumió como una investigación aplicada de alcance explicativo-preliminar, orientada a valorar indicios de efecto más que a establecer generalizaciones poblacionales amplias.

La población estuvo conformada por estudiantes de décimo año de EGB de una institución educativa rural ecuatoriana. Se trabajó con la totalidad de estudiantes disponibles, por lo

que la muestra fue censal, no probabilística e intencional. Participaron 15 estudiantes, distribuidos en dos grupos previamente constituidos: grupo de control, con 8 estudiantes, y grupo experimental, con 7 estudiantes. La ausencia de asignación aleatoria responde a las condiciones reales de organización escolar; por ello, el análisis consideró la comparación de los puntajes iniciales para identificar posibles diferencias de base entre los grupos.

La variable independiente fue la estrategia pedagógica con ChatGPT, aplicada únicamente al grupo experimental durante una semana, en cinco horas de clase. Esta intervención se diseñó como una mediación docente directamente vinculada con las habilidades evaluadas en el instrumento: reconocimiento de significados, identificación de sinónimos y antónimos, inferencia contextual y producción breve de vocabulario.

El grupo experimental trabajó con actividades guiadas en las que ChatGPT fue utilizado para generar definiciones sencillas, ejemplos contextualizados, ejercicios de opción múltiple, oraciones incompletas y retroalimentación sobre producciones breves. El docente mantuvo el control pedagógico de la actividad, revisó las respuestas generadas por la IA y promovió que los estudiantes justificaran sus elecciones léxicas. El grupo de control, en cambio, continuó con la enseñanza habitual de vocabulario, basada en explicación docente, ejercicios escritos, revisión colectiva y trabajo con significados y oraciones, sin uso de ChatGPT.

La variable dependiente fue el nivel de vocabulario receptivo, operacionalizado mediante el puntaje obtenido en la Prueba de Vocabulario Receptivo para décimo año de EGB. El instrumento evalúa reconocimiento y comprensión de palabras mediante tres secciones: opción múltiple, completar oraciones y redacción breve. La primera sección incluye 10 ítems de selección múltiple; la segunda, 10 ítems de completar con palabras como *innovador*, *hostil*, *desolación*, *arduo*, *íntegro*, *fugaz*, *cordial*, *perseverar*, *trivial* y *eufórico*; y la tercera solicita producciones breves con palabras como *melancólico*, *emprender*, *trivial*, *cordial*, *hostil* y *crucial*. El puntaje total es de 35 puntos: 10 en opción múltiple, 10 en completar y 15 en redacción breve.

El procedimiento se organizó en tres fases: 1) aplicación del pretest a ambos grupos, bajo condiciones similares de tiempo, instrucciones y

materiales; 2) implementación de la estrategia pedagógica con ChatGPT solo en el grupo experimental, mientras el grupo de control continuó con enseñanza habitual; 3) nueva aplicación del mismo instrumento como posttest a ambos grupos. Para reducir sesgos de aplicación, se mantuvieron las mismas condiciones de administración: tiempo estimado de 45 a 60 minutos, uso de material impreso, prohibición de diccionarios y dispositivos electrónicos durante la prueba, y corrección con la misma escala.

El análisis de datos se realizó en dos niveles. Primero, se calcularon estadísticos descriptivos: media, mediana, desviación estándar, rango, puntaje mínimo y máximo, tanto para el puntaje total como para cada sección del instrumento. También se estimó la ganancia individual mediante la diferencia entre posttest y pretest. Segundo, debido al tamaño reducido de la muestra, se priorizaron pruebas no paramétricas: Wilcoxon para comparar pretest y posttest dentro de cada grupo, y U de Mann-Whitney para comparar las ganancias entre grupo experimental y grupo de control. Además, se reportaron tamaños del efecto, ya que estos permiten interpretar la magnitud práctica de los cambios más allá del valor de significación estadística (Fritz et al., 2012; Lakens, 2013).

En cuanto a los criterios éticos, se consideró la participación voluntaria, el consentimiento informado de los representantes legales, la confidencialidad de los datos y el uso exclusivo de la información con fines académicos. Los resultados se presentan de manera agregada, sin identificación nominal de estudiantes ni comparación individual pública.

RESULTADOS

A continuación se exponen los resultados obtenidos a partir de la comparación entre el pretest y el posttest aplicados a los grupos experimental y de control. El instrumento se asumió con un puntaje total de 0 a 35 puntos, de acuerdo con la estructura de la *Prueba de Vocabulario Receptivo* utilizada en la medición inicial. La presentación se organiza en función de los datos individuales, los estadísticos descriptivos, las ganancias pretest-posttest, las comparaciones intragrupalas, las comparaciones entre grupos y los tamaños del efecto.

En el grupo experimental (tabla 1), los puntajes del pretest oscilaron entre 18 y 24 puntos, y los del posttest se mantuvieron también dentro del mismo rango. A nivel individual, tres participantes del grupo experimental aumentaron un punto entre el pretest y el posttest, mientras que cuatro participantes no registraron variación. No se observó ningún caso de disminución en este grupo.

Tabla 1

Datos del pretest y posttest

Código	Grupo	Pretest	Posttest	Ganancia
E01	Exp	18	18	0
E02	Exp	21	22	1
E03	Exp	24	24	0
E04	Exp	19	20	1
E05	Exp	22	22	0
E06	Exp	20	21	1
E07	Exp	23	23	0
C01	Cont	18	19	1
C02	Cont	20	22	2
C03	Cont	23	23	0
C04	Cont	19	21	2
C05	Cont	21	21	0
C06	Cont	22	23	1
C07	Cont	24	24	0
C08	Cont	20	23	3

En el grupo de control (tabla 1), los puntajes del pretest también se ubicaron entre 18 y 24 puntos. En el posttest, el rango fue de 19 a 24 puntos. En este grupo, cinco participantes presentaron algún incremento entre ambas mediciones, con ganancias que oscilaron entre uno y tres puntos; tres participantes no registraron cambio. Al igual que en el grupo experimental, no se observó disminución de puntajes en ningún caso.

Los estadísticos descriptivos permiten observar que ambos grupos partieron de valores iniciales muy próximos (tabla 2). En el pretest, el grupo experimental obtuvo una media de 21,00 puntos, con una desviación estándar de 2,16, mientras que el grupo de control presentó una media de 20,88 puntos, con una desviación estándar de 2,03. Las medianas también fueron cercanas: 21,00 en el grupo experimental y 20,50 en el grupo de control. Estos datos indican que, antes de la intervención, los puntajes promedio y la dispersión de ambos grupos eran similares.

Tabla 2

Estadísticos descriptivos por grupo

Grupo	Momento	n	Media	DE	Mediana	RIQ	Mín.-Máx.
Exp	Pretest	7	21,00	2,16	21,00	3,00	18-24
Exp	Postest	7	21,43	1,99	22,00	2,00	18-24
Cont	Pretest	8	20,88	2,03	20,50	2,50	18-24
Cont	Postest	8	22,00	1,60	22,50	2,00	19-24

En el postest, el grupo experimental alcanzó una media de 21,43 puntos, con una desviación estándar de 1,99 y una mediana de 22,00 (tabla 2). Por su parte, el grupo de control obtuvo una media de 22,00 puntos, con una desviación estándar de 1,60 y una mediana de 22,50. En ambos grupos se registró un aumento de la media respecto del pretest; sin embargo, el incremento observado en el grupo de control fue mayor que el registrado en el grupo experimental.

Tabla 3

Ganancia pretest-postest por grupo

Grupo	n	Media de ganancia	DE	Mediana	RIQ	Mín.-Máx.	Cambio porcentual aproximado
Exp	7	0,43	0,53	0,00	1,00	0-1	2,04 %
Cont	8	1,13	1,13	1,00	2,00	0-3	5,39 %

El grupo experimental presentó una ganancia media de 0,43 puntos, con una desviación estándar de 0,53, una mediana de 0,00 y un rango de 0 a 1 punto (tabla 3). El cambio porcentual aproximado fue de 2,04 %. En el grupo de control, la ganancia media fue de 1,13 puntos, con una desviación estándar de 1,13, una mediana de 1,00 y un rango de 0 a 3 puntos. El cambio porcentual aproximado fue de 5,39 %. Por tanto, los datos muestran que ambos grupos tuvieron incrementos leves, aunque la ganancia promedio fue superior en el grupo de control.

Para la comparación intragrupal entre pretest y postest se utilizó la prueba de Wilcoxon. En el grupo experimental, el estadístico W fue 0,00, con una significación exacta de $p = .250$. Este resultado indica que la diferencia entre el pretest y el postest no alcanzó significación estadística. En el grupo de control, el estadístico W también fue 0,00, con una significación exacta de $p = .063$. Aunque el valor

de p fue menor que en el grupo experimental, tampoco alcanzó el criterio convencional de significación estadística de $p < .05$ (tabla 4).

Tabla 4

Comparación intragrupal

Grupo	Comparación	Prueba	Estadístico W	p exacta
Exp	Pretest vs. postest	Wilcoxon	0,00	.250
Cont	Pretest vs. postest	Wilcoxon	0,00	.063

Las comparaciones entre grupos se realizaron mediante la prueba U de Mann-Whitney (tabla 5). En la comparación de los puntajes del pretest, el estadístico U fue 29,00, con $p = .955$. Este resultado muestra que no existió una diferencia estadísticamente significativa entre ambos grupos antes de la intervención. En la comparación de los puntajes del postest, el estadístico U fue 23,00, con $p = .613$, lo que indica que tampoco se identificó una diferencia significativa entre los grupos al finalizar el proceso.

Tabla 5

Comparación entre grupos

Comparación	Prueba	Estadístico	p
Pretest exp. vs. pretest cont.	U de Mann-Whitney	29,00	.955
Postest exp. vs. postest cont.	U de Mann-Whitney	23,00	.613
Ganancia exp. vs. ganancia cont.	U de Mann-Whitney	18,00	.281

A su vez, la comparación de las ganancias entre el grupo experimental y el grupo de control arrojó un estadístico U de 18,00, con $p = .281$ (tabla 5). Este resultado muestra que, aunque la ganancia media del grupo de control fue descriptivamente mayor que la del grupo experimental, la diferencia entre ambas ganancias no fue estadísticamente significativa. Por tanto, no se evidencia una diferencia significativa entre los grupos en términos de cambio pretest-postest.

Finalmente, los tamaños del efecto permiten complementar la descripción de la magnitud de los cambios observados (tabla 6). El cambio pre-post en el grupo experimental presentó un valor $d_{av} = 0,21$, mientras que en el grupo de control el valor

fue $d_{av} = 0,61$. Para la diferencia de ganancias entre el grupo experimental y el grupo de control se obtuvo una d de Cohen de $-0,77$ y un Delta de Cliff de $-0,36$. Estos valores indican que la magnitud descriptiva del cambio fue menor en el grupo experimental que en el grupo de control, aunque las pruebas de significación estadística no confirmaron diferencias significativas entre ambos grupos.

Tabla 6

Tamaños del efecto descriptivos

Comparación	Tamaño del efecto	Valor
Cambio pre-post en grupo exp.	d_{av}	0,21
Cambio pre-post en grupo cont.	d_{av}	0,61
Diferencia de ganancias: exp. vs. cont.	d de Cohen	$-0,77$
Diferencia de ganancias: exp. vs. cont.	Delta de Cliff	$-0,36$

DISCUSIÓN

El uso pedagógico de ChatGPT no produjo una mejora estadísticamente significativa en la adquisición de vocabulario receptivo en los estudiantes del grupo experimental. Aunque este grupo no presentó retrocesos y registró una ganancia media positiva, el incremento fue reducido: pasó de una media de 21,00 puntos en el pretest a 21,43 en el posttest, con una ganancia media de 0,43 puntos. En contraste, el grupo de control, que continuó con la enseñanza habitual, pasó de 20,88 a 22,00 puntos, con una ganancia media de 1,13 puntos. Esta diferencia descriptiva favoreció al grupo de control, aunque la comparación de ganancias entre ambos grupos no alcanzó significación estadística. Por tanto, los datos no permiten afirmar que la estrategia con ChatGPT haya tenido un efecto diferencial positivo frente a la enseñanza habitual.

Este hallazgo debe leerse con cautela frente a una parte de la literatura reciente, que ha reportado efectos favorables del uso de ChatGPT en el aprendizaje de vocabulario. Mugableh (2024), por ejemplo, encontró que los ejercicios generados con ChatGPT podían contribuir al desarrollo del conocimiento lexical en estudiantes de inglés como lengua extranjera. De manera similar, Abdelhalim y Alsehibany (2025) reportaron mejoras en el aprendizaje y la retención de vocabulario mediante actividades apoyadas en ChatGPT. En el contexto ecuatoriano, Durazno et al. (2025) documentaron avances en adquisición de vocabulario en estudiantes que utilizaron esta herramienta frente a modelos de

enseñanza tradicional. Frente a esos antecedentes, los resultados del presente estudio no confirman una ventaja clara del grupo experimental. Más bien, sugieren que la incorporación de ChatGPT, al menos en las condiciones específicas de esta intervención breve, no fue suficiente para producir una mejora superior a la obtenida mediante la enseñanza convencional.

Esta discrepancia no necesariamente invalida el potencial pedagógico de ChatGPT, pero sí obliga a relativizarlo. Kohnke et al. (2023) señalan que ChatGPT puede apoyar la enseñanza de lenguas, siempre que su uso esté mediado por orientación docente, criterios éticos y competencias digitales. Zhang y Tur (2024), en la misma línea, advierten que su empleo en educación K-12 puede favorecer la personalización y la generación de materiales, pero requiere cuidar la calidad de las respuestas, la integridad académica y el rol activo del profesorado. Desde esta perspectiva, los resultados obtenidos son coherentes con una lectura no tecnocéntrica, que afirmaría que la herramienta, por sí sola, no garantiza un efecto pedagógico superior. Su eficacia dependería de la duración de la intervención, la pertinencia de las actividades, el grado de apropiación docente, la familiaridad de los estudiantes con la herramienta y la articulación entre práctica, retroalimentación y evaluación.

También es importante considerar que la adquisición de vocabulario receptivo no se reduce al reconocimiento inmediato de significados. Luo et al. (2021) sostienen que la amplitud y la profundidad del vocabulario inciden de manera diferenciada en la comprensión, dependiendo del tipo de tarea utilizada para medirla. El instrumento aplicado en este estudio combinó reconocimiento, completación de oraciones y producción breve; por tanto, evaluó tanto identificación lexical básica, como uso contextualizado. En tal razón, es posible que una intervención de una semana y cinco horas de clase fuera insuficiente para producir cambios consistentes en dimensiones más profundas del vocabulario. Los resultados muestran incrementos leves, pero no una transformación clara del desempeño lexical. Esta evidencia es compatible con la idea de que el aprendizaje de vocabulario, especialmente cuando implica inferencia contextual y empleo pertinente de palabras, requiere práctica sostenida y seguimiento pedagógico continuo.

Los hallazgos también dialogan con la revisión de Li et al. (2024), quienes advierten que la investigación sobre ChatGPT y educación lingüística

se encuentra todavía en consolidación, con predominio de estudios en educación superior y menor presencia de investigaciones escolares, longitudinales o situadas en contextos rurales. El presente estudio se ubica justo en un escenario donde la evidencia aún es limitada: estudiantes de décimo año de EGB en una institución rural ecuatoriana. Por ello, sus resultados aportan con la señal empírica de que, en contextos escolares concretos, los efectos de ChatGPT pueden ser más modestos que los observados en investigaciones realizadas con otros perfiles de estudiantes, otros niveles educativos o intervenciones de mayor duración.

Esta lectura se refuerza al considerar las condiciones estructurales de la educación rural ecuatoriana. Alvarado et al. (2025) advierten que la educación digital rural en Ecuador continúa atravesada por desigualdades de conectividad, infraestructura, dispositivos y formación. Morocho et al. (2023) sostienen, además, que la integración de la inteligencia artificial en el sistema educativo ecuatoriano exige condiciones institucionales, técnicas y docentes que todavía son desiguales. El resultado no significativo del grupo experimental puede entenderse como la advertencia de que no basta con introducir una herramienta de IA en una secuencia didáctica para obtener mejoras verificables. La mediación tecnológica necesita condiciones de implementación suficientemente estables, familiaridad operativa, continuidad y un diseño instruccional capaz de convertir la interacción con la IA en aprendizaje observable.

Los hallazgos también se relacionan con los trabajos que llaman a una integración prudente de la IA en educación secundaria. Pratama et al. (2023) y Sumakul et al. (2022) destacan beneficios potenciales asociados con la personalización y el acceso a recursos; sin embargo, Akgun y Greenhow (2022), McGrath et al. (2023) y Nazaretsky et al. (2022) advierten riesgos vinculados con privacidad, sesgos, dependencia tecnológica y confianza acrítica en sistemas automatizados. En el presente estudio no se evaluaron directamente estos riesgos, por lo que no corresponde afirmar que hayan influido en los resultados. No obstante, la ausencia de un efecto significativo sí respalda la conclusión, aunque todavía, de que la IA educativa debe no ser asumida de antemano como una innovación necesariamente superior.

Asimismo, los resultados son consistentes con lo señalado por Márquez et al. (2025), quienes encontraron que los docentes valoran la IA por su

utilidad para planificar, motivar y adaptar materiales, pero también reconocen riesgos como debilitamiento del pensamiento crítico, erosión de la autonomía estudiantil y falta de condiciones institucionales para una integración equitativa. Peñafiel et al. (2024) coinciden en que el éxito de la IA educativa depende menos de su sofisticación técnica que de su apropiación ética, contextual y pedagógica. En el presente estudio, aunque la intervención fue diseñada como mediación docente guiada, los resultados no permiten sostener que esa mediación haya producido una mejora diferencial. Esto sugiere que la planificación pedagógica con IA debe fortalecerse mediante intervenciones más prolongadas, tareas más progresivas y mecanismos de seguimiento más finos sobre el proceso lexical.

Por último, los hallazgos deben interpretarse en relación con las advertencias de Darvishi et al. (2024), Zhai et al. (2024) y Walter (2024) sobre agencia estudiantil, sobredependencia y alfabetización crítica en IA. Aunque este estudio no midió dependencia tecnológica ni pensamiento crítico, sus resultados refuerzan la necesidad de no reemplazar la construcción activa del vocabulario por la simple exposición a respuestas generadas por IA. García (2024) señala que ChatGPT puede generar materiales útiles para el trabajo lexical, pero estos requieren revisión y ajuste didáctico. En suma, se mostró que, en esta aplicación específica, el efecto del ChatGPT fue limitado y no superior al de la enseñanza habitual.

Limitaciones del estudio

La principal limitación fue el tamaño muestral reducido, con 15 estudiantes distribuidos en dos grupos naturales, lo que restringe la potencia estadística y la posibilidad de generalizar los resultados. Además, la intervención tuvo una duración breve, de una semana y cinco horas de clase, lo que pudo limitar la aparición de cambios más estables en vocabulario receptivo. También debe considerarse que no se incorporaron mediciones de seguimiento para valorar retención. Por ello, los hallazgos deben entenderse como evidencia preliminar, situada y exploratoria, no como una conclusión definitiva sobre la eficacia general de ChatGPT en la adquisición de vocabulario.

CONCLUSIÓN

El uso pedagógico de ChatGPT, aplicado durante una intervención breve y guiada, no generó una mejora

estadísticamente significativa en la adquisición de vocabulario receptivo frente a la enseñanza habitual. Aunque el grupo experimental presentó una ganancia positiva, esta fue menor que la observada en el grupo de control y no alcanzó significación estadística. Estos hallazgos sugieren que la incorporación de inteligencia artificial en el aula no garantiza, por sí misma, mejores resultados de aprendizaje. Su efectividad depende de la duración de la intervención, la mediación docente, la calidad de las actividades y las condiciones institucionales de implementación. Por tanto, ChatGPT puede considerarse un recurso pedagógico potencialmente útil, pero requiere integración planificada, seguimiento sostenido y evaluación empírica rigurosa antes de atribuirle efectos educativos concluyentes.

BIBLIOGRAFÍA

- Abdelhalim, S. y Alsehibany, R. (2025). Integrating ChatGPT for vocabulary learning and retention: A classroom-based study of Saudi EFL learners. *Language Learning & Technology*, 29(1), 1–24. <https://doi.org/10.64152/10125/73635>
- Akgun, S. y Greenhow, C. (2022). Artificial intelligence in education: Addressing ethical challenges in K-12 settings. *AI and Ethics*, 2, 431–440. <https://doi.org/10.1007/s43681-021-00096-7>
- Alvarado, R., Alvarado, S., Govea, D. y Jaime, D. (2025). Impacto de la educación digital en las zonas rurales de Ecuador: Retos y oportunidades. *Seminars in Medical Writing and Education*, 4, 468. <https://doi.org/10.56294/mw2025468>
- Chounta, I., Bardone, E., Raudsep, A. y Pedaste, M. (2021). Exploring teachers' perceptions of artificial intelligence as a tool to support their practice in Estonian K-12 education. *International Journal of Artificial Intelligence in Education*. <https://doi.org/10.1007/s40593-021-00243-5>
- Darvishi, A., Khosravi, H., Sadiq, S., Gašević, D. y Siemens, G. (2024). Impact of AI assistance on student agency. *Computers & Education*, 210, 104967. <https://doi.org/10.1016/j.compedu.2023.104967>
- Durazno, B., Díaz, K. y Zambrano, J. (2025). Exploring the effect of ChatGPT as an educational tool to improve vocabulary acquisition in the English language. *Ciencia Latina Revista Científica Multidisciplinar*, 9(1), 4874–4886. https://doi.org/10.37811/cl_rcm.v9i1.16190
- Fritz, C., Morris, P. y Richler, J. (2012). Effect size estimates: Current use, calculations, and interpretation. *Journal of Experimental Psychology: General*, 141(1), 2–18. <https://doi.org/10.1037/a0024338>
- García, M. (2024). Inteligencia artificial y léxico: Una propuesta con ChatGPT para niveles B1 y B2 en la enseñanza del español como lengua extranjera. *RILEX. Revista sobre Investigaciones Léxicas*, 7(3), e9111. <https://doi.org/10.17561/rilex.7.3.9111>
- Kim, N. y Kim, M. (2022). Teacher's perceptions of using an artificial intelligence-based educational tool for scientific writing. *Frontiers in Education*, 7, 755914. <https://doi.org/10.3389/feduc.2022.755914>
- Kohnke, L., Moorhouse, B. y Zou, D. (2023). ChatGPT for language teaching and learning. *RELC Journal*, 54(2), 537–550. <https://doi.org/10.1177/00336882231162868>
- Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for *t*-tests and ANOVAs. *Frontiers in Psychology*, 4, 863. <https://doi.org/10.3389/fpsyg.2013.00863>
- Li, B., Lowell, V. L., Wang, C. y Li, X. (2024). A systematic review of the first year of publications on ChatGPT and language education: Examining research on ChatGPT's use in language learning and teaching. *Computers and Education: Artificial Intelligence*, 7, 100266. <https://doi.org/10.1016/j.caeai.2024.100266>
- Luo, Y., Yang, M. y Zhang, X. (2021). The effect of vocabulary depth and breadth on English listening comprehension can depend on how comprehension is measured. *Frontiers in Psychology*, 12, 657573. <https://doi.org/10.3389/fpsyg.2021.657573>
- Márquez, N., Guamán, I. y Peñafiel, R. (2025). Entre la innovación y la cautela: Voces docentes sobre la inteligencia artificial en la educación secundaria. *South American Research Journal*, 5(1), 13–25. <https://doi.org/10.5281/zenodo.15659239>
- McGrath, C., Cerratto Pargman, T., Juth, N. y Palmgren, P. (2023). University teachers' perceptions of responsibility and artificial intelligence in higher education: An experimental philosophical study. *Computers and Education: Artificial Intelligence*, 4, 100139. <https://doi.org/10.1016/j.caeai.2023.100139>
- Morocho, R., Cartuche, A., Tipan, A., Guevara, A. y Ríos, M. (2023). Integración de la inteligencia artificial en la educación. *Ciencia Latina Revista Científica Multidisciplinar*, 7(6), 2032–2053. https://doi.org/10.37811/cl_rcm.v7i6.8832
- Mugableh, A. (2024). The impact of ChatGPT on the development of vocabulary knowledge of Saudi EFL students. *Arab World English Journal, Special Issue on ChatGPT*, 265–281. <https://dx.doi.org/10.24093/awej/ChatGPT.18>
- Nazaretsky, T., Ariely, M., Cukurova, M. y Alexandron, G. (2022). Teachers' trust in AI-powered educational technology and a professional development program to improve it. *British Journal of Educational Technology*, 53(4), 914–931. <https://doi.org/10.1111/bjet.13232>
- Peñafiel, R., Márquez, N. y Guamán, I. (2024). Inteligencia artificial en la educación: Revisión sistemática de perspectivas, beneficios y desafíos en la práctica docente. *South American Research Journal*, 4(2), 5–15. <https://doi.org/10.5281/zenodo.14507789>

- Pratama, M., Sampelolo, R. y Lura, H. (2023). Revolutionizing education: Harnessing the power of artificial intelligence for personalized learning. *Klasikal: Journal of Education, Language Teaching and Science*, 5(2), 350–357. <https://doi.org/10.2656-9914>
- Shadish, W., Cook, T. y Campbell, D. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin.
- Sumakul, D., Hamied, F. y Sukyadi, D. (2022). Artificial intelligence in EFL classrooms: Friend or foe? *LEARN Journal: Language Education and Acquisition Research Network*, 15(1), 232–256. <https://so04.tci-thaijo.org/index.php/LEARN/index>
- Walter, Y. (2024). Embracing the future of artificial intelligence in the classroom: The relevance of AI literacy, prompt engineering, and critical thinking in modern education. *International Journal of Educational Technology in Higher Education*, 21, 15. <https://doi.org/10.1186/s41239-024-00448-3>
- What Works Clearinghouse. (2022). *What Works Clearinghouse procedures and standards handbook, version 5.0*. U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance. <https://ies.ed.gov/ncee/wwc/handbooks/>
- Zhai, C., Wibowo, S. y Li, L. (2024). The effects of over-reliance on AI dialogue systems on students' cognitive abilities: A systematic review. *Smart Learning Environments*, 11, 28. <https://doi.org/10.1186/s40561-024-00316-7>
- Zhang, P. y Tur, G. (2024). A systematic review of ChatGPT use in K-12 education. *European Journal of Education*, 59(2), e12599. <https://doi.org/10.1111/ejed.12599>

Declaración de uso de IA: Se usó Inteligencia Artificial exclusivamente con propósitos de edición del artículo.