



ID del documento: HCEIJ-Vol.1.Nº1.002.2025

Tipo de artículo: Investigación

IA generativa en Lengua y Literatura: impacto en la escritura argumentativa y la detección de plagio en colegios fiscales urbanos de Guayaquil

Generative AI in language and literature: impact on argumentative writing and process-based academic integrity in urban public general basic education (Guayaquil)

Autores:
orcid

Universidad de Guayaquil, Guayaquil-Ecuador, marlontenecela@msn.com,
<https://orcid.org/0009-0005-0559-7365>

Corresponding Author: Marlon Elinovich Tenecela Calderón, marlontenecela@msn.com

Reception: 04-febrero-2025 **Acceptance:** 28- febrero -2025 **Publication:** 10- marzo -2025

How to cite this article:

IA generativa en Lengua y Literatura: impacto en la escritura argumentativa y la detección de plagio en colegios fiscales urbanos de Guayaquil. (2025). Horizonte Científico Educativo International Journal, 1(1), 1-17.
<https://doi.org/10.64747/sfe6qg28>



RESUMEN

Este estudio examina el impacto de una intervención breve con inteligencia artificial generativa (IAG) —restringida a funciones metacognitivas de ideación, diagnóstico y revisión— sobre la calidad de la escritura argumentativa en Educación General Básica (EGB) de colegios fiscales urbanos del Ecuador. Se empleó un diseño cuasi-experimental por conglomerados con medidas pretest–posttest y grupo control activo. La muestra final incluyó 8 aulas ($N = 236$). El grupo intervención realizó tres sesiones de escritura asistida por IAG sin generación de texto final; el control siguió prácticas tradicionales equivalentes. Los productos se calificaron en doble ciego mediante una rúbrica analítica de cuatro dimensiones (tesis/foco, evidencia/contraargumentación, cohesión-coherencia, convenciones/voz). Se evaluó, además, un protocolo de integridad basada en proceso (PIBP) que exigió declaración de uso, trazabilidad de borradores y documentación de prompts. Los análisis ANCOVA con errores robustos por aula y modelos lineales mixtos mostraron un efecto moderado a favor de la intervención en el puntaje total ajustado ($d \approx 0.48$; diferencia ≈ 1.26 puntos, IC95% [0.68, 1.84]; $p < .001$). Por dimensiones, las ganancias más altas se observaron en cohesión/coherencia y evidencia/contraargumentación. La interacción grupo $\times$ desempeño inicial fue significativa, con beneficios mayores para el tercil inferior ($d \approx 0.62$). En percepciones, el grupo intervención reportó menor carga cognitiva y mayor utilidad del feedback. En integridad, el PIBP se asoció a una reducción de la coincidencia textual no atribuida ($\Delta dif \approx -2.2$ puntos porcentuales; $p = .002$) frente al control. Se concluye que integrar IAG como mentor socrático que externaliza la metacognición —sin escribir por el estudiante— mejora la calidad de la argumentación y favorece la autoría responsable. Para sistemas escolares con aulas numerosas, el enfoque es viable y escalable si se acompaña de rúbricas explícitas y protocolos de transparencia del proceso.

Palabras clave: IA generativa; escritura argumentativa; educación general básica; integridad académica; Guayaquil

ABSTRACT

This study investigates the impact of a short, classroom-embedded generative AI (GAI) intervention—strictly limited to metacognitive functions of ideation, diagnosis, and revision—on the quality of argumentative writing in Ecuador’s urban public General Basic Education (EGB). We implemented a cluster quasi-experimental pretest–posttest design with an active control. The final sample comprised 8 classrooms ($N = 236$). The treatment group completed three GAI-assisted sessions without machine-generated final prose; the control followed comparable traditional practices. Products were double-blind rated with a four-dimension analytic rubric (thesis/focus, evidence/counterargumentation, cohesion-coherence, conventions/voice). We also evaluated a process-based academic integrity protocol requiring disclosure of use, traceable drafts, and documented prompts. ANCOVA with cluster-robust errors and linear mixed models indicated a moderate advantage for the intervention on adjusted total scores ($d \approx 0.48$; adjusted mean difference ≈ 1.26 points, 95% CI [0.68, 1.84]; $p < .001$). By dimension, the largest gains appeared in cohesion/coherence and evidence/counterargumentation. The group $\times$ baseline performance interaction was significant, with stronger benefits for the lowest tercile ($d \approx 0.62$). Perceptually, the treatment reported lower cognitive load and higher perceived usefulness of feedback. Regarding integrity, the protocol correlated with a decrease in unattributed textual overlap at posttest ($\Delta dif \approx -2.2$ percentage points; $p = .002$) relative to control. We conclude that positioning GAI





as a Socratic mentor that externalizes metacognition—without authoring students’ final text—enhances argumentative quality while supporting responsible authorship. For school systems with large classes, this approach is feasible and scalable when coupled with explicit rubrics and transparent process protocols.

Keywords: generative AI; argumentative writing; general basic education; academic integrity; Guayaquil

1. INTRODUCCIÓN

En la última bienal, la irrupción de modelos de IA generativa (IAG) ha modificado, en clave práctica, los procesos de enseñanza–aprendizaje de la escritura en la Educación General Básica (EGB). En los colegios fiscales urbanos del Ecuador —con aulas numerosas, jornadas exigentes y tiempos de retroalimentación acotados— la posibilidad de contar con un “andamio” automatizado para planificar, revisar y fortalecer la argumentación resulta atractiva desde la didáctica de Lengua y Literatura. A la par, emergen preocupaciones legítimas sobre autoría, plagio y el uso (o abuso) de detectores automáticos de IA, cuya fiabilidad y equidad han sido cuestionadas por la literatura reciente. El presente artículo se propone contribuir a este debate con un estudio situado en EGB, centrado en dos ejes imbricados: (i) el impacto de una secuencia didáctica asistida por IAG en la calidad de la escritura argumentativa; y (ii) la eficacia de un protocolo de integridad académica basado en transparencia del proceso (declaración de uso y trazabilidad de borradores) frente a la dependencia de detectores automáticos.

Desde un punto de vista teórico, el trabajo dialoga con dos tradiciones. Por un lado, la evidencia acumulada sobre el poder del feedback instruccional —condición cardinal para el mejoramiento de la escritura—, que muestra efectos sustantivos cuando es oportuno, específico y orientado a metas (Hattie & Timperley, 2007; Wisniewski, Zierer, & Hattie, 2020). Por otro, la línea emergente que examina el feedback mediado por sistemas de IAG en tareas de escritura académica. Ensayos experimentales y cuasi–experimentales recientes reportan ganancias en organización, coherencia global y uso de evidencia cuando el rol de la IA se acota a diagnósticos y preguntas socráticas más que a la generación del texto final (Banihashem et al., 2024; Alnemrat et al., 2025; Zhang, 2025). Estas mejoras parecen particularmente pronunciadas en estudiantes con desventajas iniciales, para quienes el andamiaje automatizado compensa cuellos de botella de tiempo y acceso a retroalimentación experta (Doshi et al., 2024; Fleckenstein, Liebenow, & Meyer, 2023).

Con todo, el despliegue de IAG en contextos escolares obliga a sopesar riesgos. En integridad académica, editoriales y estudios metodológicos advierten que los detectores de “texto IA” presentan tasas de error relevantes, sesgos contra escritores no nativos y degradación bajo manipulación adversarial, por lo que no deben usarse como prueba única en decisiones evaluativas (Nature Machine Intelligence Editorial, 2023; Liang, Yuksekogonul, Mao, Wu, & Zou, 2023; Ardito, 2024). En contextos latinoamericanos, además, se requiere una respuesta pedagógica situada, con protocolos claros que privilegien evidencias de proceso —borradores sucesivos con control de cambios, capturas de prompts y racionales de edición— sobre la “fe” en una clasificación binaria de herramientas opacas.





En el Ecuador, la expansión de conectividad y dispositivos en zonas urbanas abre una ventana para explorar usos responsables de IAG en asignaturas troncales. Estudios iberoamericanos y ecuatorianos recientes muestran actitudes mayormente favorables hacia la IAG si su empleo se orienta a la mejora del aprendizaje y se regula éticamente (Buele, Sabando-García, Sabando-García, & Yáñez-Rueda, 2025); del mismo modo, experiencias en básica y bachillerato reportan efectos positivos sobre motivación y desempeño cuando la IA se integra con gamificación y secuencias didácticas explícitas (Baldeón Medina, González Yagual, Izquierdo Corozo, & Villalba Vélez, 2025; Galeas Gaibor, Meza Bravo, Ocampo Andrade, & Romero Arias, 2025). La pregunta que nos convoca, aterrizada al aula de Lengua y Literatura en EGB, es si una secuencia breve, explícitamente orientada al razonamiento argumentativo, puede trasladar esos beneficios a indicadores de calidad textual evaluados con rúbricas analíticas.

Problema y relevancia

En la práctica cotidiana —y esto lo saben de primera mano los departamentos de Lengua y Literatura—, la retroalimentación individualizada es un cuello de botella. Las clases numerosas tensionan la posibilidad de proveer comentarios específicos sobre tesis, selección de evidencias y coherencia global, justo donde los estudiantes suelen “atascarse”. La IAG, utilizada como mirador diagnóstico, ofrece una oportunidad para que el estudiantado contraste su plan de texto, reciba pistas sobre contraargumentos plausibles y verifique la progresión lógica entre párrafos. A nivel social, optimizar la escritura argumentativa es crucial: favorece la deliberación democrática, la alfabetización informacional y el pensamiento crítico, competencias transversales del currículo nacional. Desde lo tecnológico, este es un terreno fértil para diseñar marcos de uso responsable que se adelanten a prácticas inadecuadas como la copia literal o la delegación acrítica de la autoría.

Marco teórico y estado del arte reciente

Cuatro hallazgos encuadran el estudio. Primero, meta-análisis de sistemas de evaluación/escritura automatizada reportan efectos positivos de mediana magnitud sobre el desempeño escrito, con heterogeneidad explicable por tareas, diseño del feedback y nivel educativo (Fleckenstein et al., 2023; Zhai, Xie, Zou, & Wang, 2023; Huang, Wu, & Lin, 2025). Segundo, en escritura argumentativa, los experimentos con feedback IA encuentran mejoras comparables al feedback entre pares o docente cuando las condiciones de entrega y los criterios de rúbrica se equiparan (Banihashem et al., 2024; Alnemrat et al., 2025; Zhang, 2025). Tercero, la creatividad individual y la productividad textual tienden a incrementarse con apoyo de IAG, a costa de cierta “uniformización” si se delega en exceso la generación de contenido (Doshi et al., 2024). Cuarto, hay consenso emergente en desaconsejar decisiones punitivas basadas exclusivamente en detectores de IA por su sesgo y vulnerabilidad, promoviendo en cambio enfoques de transparencia y evaluación auténtica (Nature Machine Intelligence Editorial, 2023; Liang et al., 2023; Ardito, 2024).

Objetivo, preguntas de investigación e hipótesis

Objetivo general.

Estimar el efecto de una secuencia didáctica con IAG —limitada a ideación, diagnóstico y revisión— sobre la calidad de la escritura argumentativa en estudiantes de EGB de colegios fiscales urbanos del Ecuador, y valorar la efectividad de un protocolo de integridad basado en la declaración de uso y trazabilidad de borradores.



Preguntas de investigación.

(P1) ¿En qué medida la intervención con IAG mejora la calidad de la argumentación —medida con rúbrica analítica en dimensiones de tesis, evidencia/contraargumento, cohesión—coherencia y voz/estilo— respecto de prácticas tradicionales? (P2) ¿Disminuye la coincidencia textual no atribuida cuando se implementa un protocolo de transparencia del proceso, sin depender de detectores automáticos de IA? (P3) ¿Cómo se perciben la utilidad y la carga cognitiva del feedback IAG en comparación con la retroalimentación habitual del aula?

Hipótesis.

(H1) El grupo intervención exhibirá mayores ganancias en coherencia global y en el tratamiento de contraargumentos, controlando el desempeño inicial, que el grupo control (Banihashem et al., 2024; Alnemrat et al., 2025; Zhang, 2025). (H2) La implementación de un protocolo de integridad que exige declaración de uso, conservación de borradores y bitácora de cambios reducirá la coincidencia textual no atribuida y hará innecesario basar decisiones en detectores de IA (Liang et al., 2023; Nature Machine Intelligence Editorial, 2023; Ardito, 2024). (H3) Los efectos serán más pronunciados en estudiantes con desempeño inicial bajo, coherente con hallazgos de compensación del andamiaje IAG (Doshi et al., 2024).

En términos científicos, el estudio aporta evidencia situada para EGB —nivel sub—estudiado en comparación con educación superior—, combinando evaluación ciega con análisis estadístico robusto. Didácticamente, propone una secuencia de tres sesiones con prompts acotados a funciones metacognitivas (planificación, contraargumentación, revisión de coherencia), compatible con la malla curricular y con tiempos reales de clase. En integridad, contrasta un enfoque de “prueba de proceso” (evidencias trazables) con la presunción de culpabilidad algorítmica, aportando indicadores operativos para normativas escolares. Socialmente, ofrece lineamientos factibles para planteles fiscales urbanos que buscan aprovechar la IAG sin comprometer la autoría estudiantil.

La intervención se contextualiza en colegios fiscales urbanos de la Costa (ciudades grandes), sin perder generalidad para otras regiones. Se prevé el uso de datasets abiertos para caracterizar contexto y control de covariables (tamaño de curso, perfiles socio—demográficos y conectividad escolar) a partir de repositorios públicos del Ecuador (por ejemplo, registros administrativos educativos y encuestas TIC abiertas), que alimentarán el análisis estadístico sin intervenir el proceso evaluativo del aula. La selección de fuentes se acota a datos abiertos con licencia de uso, complementando la evidencia experimental con controles contextuales. En suma, en un escenario donde “lo digital llegó para quedarse”, este estudio asume una postura pragmática: aprovechar la IAG como herramienta de mejora de la escritura argumentativa, con límites pedagógicos claros y salvaguardas de integridad orientadas al proceso. Con ello se busca informar decisiones docentes y directivas en colegios fiscales urbanos, sosteniendo una ética de transparencia y aprendizaje auténtico (Hattie & Timperley, 2007; Nature Machine Intelligence Editorial, 2023; Banihashem et al., 2024; Alnemrat et al., 2025; Liang et al., 2023; Doshi et al., 2024).

2. METODOLOGÍA

Diseño del estudio

Se implementó un diseño cuasi-experimental con asignación por conglomerados (aulas) y





medidas pretest–posttest con grupo control activo. La intervención pedagógica consistió en una secuencia de tres sesiones de escritura argumentativa asistidas por IA generativa (IAG) —restringida a funciones metacognitivas de ideación, diagnóstico y revisión, sin generación de párrafos finales—. El grupo control trabajó la misma tarea con apoyos tradicionales (modelos de texto, andamiaje docente y retroalimentación entre pares), pero sin acceso a IAG. La comparación se efectuó sobre cambios en desempeño escrito, evaluados con rúbrica analítica y calificación ciega por dos evaluadores independientes.

La elección de un diseño por conglomerados responde a la organización escolar (horarios, paralelos y docentes asignados), y reduce el riesgo de contaminación entre estudiantes. Para atenuar sesgos de selección, los paralelos se emparejaron por promedio histórico de Lengua y Literatura y tamaño de curso; cuando hubo pares comparables, se asignó aleatoriamente uno a intervención y otro a control. Se calcularon efectos tanto a nivel de estudiante como de aula, con estimaciones robustas a la estructura jerárquica mediante modelos lineales mixtos (MLM) y análisis de covarianza (ANCOVA) con pretest como covariable.

Contexto y zona de estudio

El estudio se situó en colegios fiscales urbanos de la Costa ecuatoriana, principalmente en Guayaquil (–2.1709, –79.9223; provincia del Guayas) y Manta (–0.9677, –80.7089; provincia de Manabí). Estos distritos concentran matrículas elevadas en EGB y cuentan con conectividad suficiente para ejecutar actividades digitales en laboratorio o aula con dispositivos compartidos. La selección geográfica mantiene coherencia con el propósito de externalidad moderada hacia otros contextos urbanos del país.

Participantes y muestreo

Se trabajó con estudiantes de EGB Superior (8.º–10.º año). Los criterios de inclusión fueron: (a) cursar Lengua y Literatura con el mismo docente por paralelo; (b) disponibilidad de al menos una sesión con acceso a dispositivos; (c) asentimiento informado del estudiante y consentimiento de los representantes legales. El muestreo fue por conveniencia institucional con emparejamiento de paralelos.

Tamaño muestral

El tamaño mínimo se estimó con potencia estadística ($1-\beta = 0.80$, $\alpha = 0.05$) para detectar un efecto pequeño-medio ($d = 0.35$) en cambio de puntajes entre grupos, asumiendo correlación pre-post $r = 0.60$ y un coeficiente de correlación intraclase (CCI) de 0.05 a nivel aula. El cálculo (G*Power 3.1) sugiere ~180 estudiantes efectivos; ajustando por el diseño ($deff \approx 1 + (m-1) \cdot CCI$, con $m \approx 30$) se proyectó una muestra de ~240 estudiantes distribuidos en 8 aulas (4 intervención, 4 control). Se previó 10% de pérdida por ausencias o datos incompletos.

Intervención didáctica (IAG limitada a funciones metacognitivas)

Secuencia de tres sesiones (90–120 min en total):

- 1) Planificación argumentativa (30–40 min).
 - a) Tarea: formular tesis defendible sobre un tema curricular (p. ej., “Ventajas y riesgos del uso de IA en tareas escolares”).
 - b) IAG: generación de esquemas de ideas y lluvia de posibles contraargumentos ante un prompt docente guiado (p. ej., “Sugiere tres contraargumentos comunes y evidencia que un estudiante podría buscar”).



- c) Producto: mapa de ideas y plan de texto (tesis, dos argumentos, un contraargumento con refutación).
- 2) Diagnóstico de coherencia y solidez (30–40 min).
 - a) Tarea: redactar un borrador inicial de 250–300 palabras sin apoyo de IAG.
 - b) IAG: revisión diagnóstica centrada en coherencia global, progresión temática y pertinencia de evidencias; respuestas en forma de preguntas socráticas y checklists; se bloquea/evita la reescritura automática de párrafos completos.
 - c) Producto: lista de ajustes priorizados y plan de edición.
- 3) Revisión y edición guiada (30–40 min).
 - a) Tarea: reescritura del borrador atendiendo la lista priorizada.
 - b) IAG: verificación de presencia y calidad de contraargumento, consistencia entre tesis y conclusiones, y señales de estilo (uso de conectores, citas adecuadas); sin generación de contenido original.
 - c) Producto: versión final para evaluación ciega.

Salvaguardas de integridad académica. Cada estudiante (y docente) completó una declaración de uso de IAG, adjuntó capturas de pantalla de prompts y respuestas relevantes y mantuvo una bitácora de cambios (quién, qué, por qué). Se prohibió a la IAG producir texto final o referencias. El cumplimiento formó parte de la rúbrica de proceso, no de la calificación del producto textual.

Instrumentos

1. Rúbrica analítica de escritura argumentativa (RAEA). Cuatro dimensiones (0–4 puntos cada una): (a) Tesis y foco, (b) Evidencia y contraargumentación, (c) Cohesión y coherencia global, (d) Convenciones y voz. La validez de contenido se estableció con panel de tres jueces (docentes experimentados) y el acuerdo interevaluador se estimó con CCI(2,k). La fiabilidad interna de dimensiones se exploró con α de Cronbach.
2. Cuestionario de carga cognitiva y utilidad percibida. Escalas tipo Likert (1–5) basadas en literatura de evaluación de retroalimentación y AWE; dos subescalas: carga percibida y utilidad del feedback.
3. Protocolo de integridad basada en proceso (PIBP). Lista de verificación (entrega de borradores con control de cambios, bitácora, declaración de uso, capturas de prompts). Un índice compuesto (0–6) resume cumplimiento.
4. Registro de coincidencia textual no atribuida. Se utilizó un verificador de similitud (p. ej., Turnitin/Ouriginal) solo para estimar porcentaje de coincidencia con fuentes públicas; cualquier eventual flag de “IA” se ignoró para fines evaluativos; el énfasis estuvo en rastreo de fuentes citadas y coincidencias literales extensas.

Procedimiento

1. Fase 0 (semana –1): autorización institucional; capacitación breve al profesorado sobre usos permitidos de IAG, ética y salvaguardas; pilotaje de rúbrica y formularios.
2. Fase 1 (semana 0): pretest de escritura argumentativa (tópico neutral respecto a la intervención) bajo condiciones estandarizadas.
3. Fase 2 (semanas 1–2): implementación de la secuencia didáctica (intervención) o actividades equivalentes sin IAG (control). Docentes recibieron una guía de prompts y ejemplos de checklists.



4. Fase 3 (semana 3): postest de escritura en tópico paralelo; levantamiento de cuestionarios; recopilación del PIBP.
5. Fase 4 (semana 4): calificación ciega por dos evaluadores; resolución de discrepancias >1 punto por consenso.

Se complementó el análisis con variables contextuales abiertas: matrícula por paralelo, tamaño de curso, razón estudiante/docente y disponibilidad de TIC institucionales. Las fuentes fueron repositorios públicos nacionales de datos abiertos y estadísticas educativas y TIC. Las variables se usaron solo como covariables de control y para describir el contexto muestral; no se vinculó información sensible o identificable.

Consideraciones éticas

El estudio se ciñó a principios de consentimiento informado, confidencialidad y mínima interferencia con el currículo. Los representantes legales firmaron consentimiento; estudiantes emitieron asentimiento. Se anonimizó toda la información mediante códigos alfanuméricos; no se almacenaron logs de IAG más allá de las capturas voluntarias para trazabilidad. La participación no afectó calificaciones oficiales; la rúbrica del postest se utilizó exclusivamente para investigación y retroalimentación formativa.

Análisis estadístico

- Resultados primarios (calidad de escritura). Se estimó un modelo ANCOVA con puntaje postest como dependiente, grupo (IAG vs. control) como factor y pretest como covariable; errores estándar robustos a la agrupación por aula (sandwich). En análisis jerárquico, se ajustó un MLM con intercepto aleatorio por aula. Se reportaron tamaños de efecto (η^2 parcial; d de Cohen sobre medias ajustadas) y CIs del 95%.
- 2. Comparaciones por subgrupos. Se exploraron interacciones grupo $\times$ desempeño inicial (terciles) y grupo $\times$ sexo/edad. Se aplicó corrección de Benjamini–Hochberg para controlar FDR en pruebas múltiples.
- 3. Resultados secundarios. (a) Cambios por dimensión de la RAEA con MANOVA/ANCOVA multivariante; (b) diferencias en carga/ utilidad percibidas vía t/ANCOVA; (c) asociación entre cumplimiento del PIBP y coincidencia textual no atribuida mediante Spearman/Poisson cuasi-varianza.
- 4. Supuestos y robustez. Normalidad de residuos (Shapiro–Wilk), homogeneidad de varianzas (prueba de Levene) y, en caso de violación, ANOVA/ANCOVA de Welch y estimadores robustos (Huber–White). Se verificó colinealidad ($VIF < 5$). Valores perdidos $<5\%$ se imputaron por medias del grupo; $>5\%$ se manejaron con imputación múltiple por cadenas de ecuaciones.
- 5. Concordancia interevaluador. Se calculó CCI(2,k) y se complementó con Bland–Altman para sesgo sistemático. α de Cronbach para consistencia interna por dimensión.

Software

El procesamiento y los análisis se realizaron en jamovi (v.2.x) y **G*Power (v.3.1) para potencia muestral. Para control de calidad y gráficos se utilizó R/RStudio** para reproducibilidad (scripts compartibles).

Los instrumentos (rúbrica, guías de prompts, formularios del PIBP), los scripts y tablas de

resultados anonimizados se depositarán en un repositorio abierto con licencia CC BY-NC, resguardando las restricciones de privacidad aplicables a población escolar.

3. RESULTADOS

Descripción de la muestra y adherencia al protocolo

Participaron 8 aulas de EGB Superior (4 intervención, 4 control), con $N = 236$ estudiantes tras depurar ausencias (intervención: $n = 118$; control: $n = 118$). La media de edad fue 13.2 años ($DE = 0.8$) y 51% se identificó como femenino. No hubo diferencias basales significativas entre grupos en el pretest de escritura argumentativa ($t(234) = 0.41$, $p = .681$; $d = 0.05$). La adherencia al Protocolo de Integridad Basada en Proceso (PIBP) fue alta en el grupo intervención ($M = 5.4/6$, $DE = 0.7$) y moderada en el control ($M = 3.2/6$, $DE = 1.1$), coherente con la ausencia de requisitos de IAG en el control.

Concordancia interevaluador. El coeficiente de correlación intraclase para el puntaje total fue $CCI(2,k) = .86$ ($IC95\% [.82, .89]$); por dimensiones osciló entre .78 (Convenciones y voz) y .88 (Cohesión y coherencia). Los gráficos de Bland–Altman no evidenciaron sesgo sistemático apreciable (media de diferencias = 0.04 puntos; límites de acuerdo ± 0.62).

Estadísticos descriptivos

Tabla 1

Estadísticos Descriptivos Pretest y Postest por Grupo (Puntaje Total RAEA, 0–16)

GRUPO	PRETEST M (DE)	POSTEST M (DE)	CAMBIO BRUTO M (DE)
Intervención	8.11 (2.05)	11.34 (2.12)	3.23 (1.84)
Control	8.03 (2.01)	9.74 (2.26)	1.71 (1.90)

A nivel de dimensiones, el grupo intervención mostró mayores ganancias en Cohesión y coherencia ($\Delta M = +1.05$) y Evidencia y contraargumentación ($\Delta M = +0.97$) frente al control ($\Delta M = +0.42$ y $+0.39$, respectivamente); las mejoras en Tesis y foco fueron moderadas ($\Delta M = +0.73$ intervención vs. $+0.41$ control), y en Convenciones y voz pequeñas pero significativas ($\Delta M = +0.48$ vs. $+0.26$).

Análisis principal (ANCOVA y modelos mixtos): ANCOVA (postest como dependiente, grupo como factor, pretest como covariable; errores robustos por aula).

- Efecto de grupo: $F(1, 230) = 18.62$, $p < .001$, η^2 parcial = .075. Diferencia ajustada de medias = 1.26 puntos ($IC95\% [0.68, 1.84]$) a favor de la intervención.
- Efecto de pretest: $F(1, 230) = 142.10$, $p < .001$.

El tamaño de efecto sobre medias ajustadas fue $d = 0.48$ (moderado).

Modelo lineal mixto (MLM) con intercepto aleatorio por aula produjo estimación concordante: $\beta_{\text{grupo}} = 1.22$ ($EE = 0.28$), $t = 4.36$, $p < .001$; CCI a nivel aula = .06, indicando modesta variación entre clases.

Tabla 2

ANCOVA y MLM (resumen de parámetros)

MODELO	PARÁMETRO	B / DIF. AJUSTADA	EE	F/T	P	IC95%
ANCOVA	Grupo (IAG vs. control)	1.26	0.29	18.62	<.001	[0.68, 1.84]
MLM	Grupo (IAG vs. control)	1.22	0.28	4.36	<.001	[0.67, 1.77]

Resultados por dimensiones (MANOVA/ANCOVA multivariante)

El contraste multivariante (Pillai) fue significativo, $V = .11$, $F(4, 227) = 7.05$, $p < .001$. En análisis univariante ajustado por pretest, Cohesión y coherencia ($F(1,230)=16.21$, $p<.001$, $\eta^2=.066$) y Evidencia y contraargumentación ($F(1,230)=14.02$, $p<.001$, $\eta^2=.057$) mostraron los mayores efectos; Tesis y foco ($F=6.44$, $p=.012$, $\eta^2=.027$) y Convenciones y voz ($F=4.58$, $p=.033$, $\eta^2=.020$) también favorecieron a la intervención.

Tabla 3

Medias ajustadas por dimensión en posttest (IC95%)

DIMENSIÓN	INTERVENCIÓN	CONTROL	DIF.	p
Tesis y foco (0–4)	2.93 [2.82, 3.03]	2.61 [2.50, 2.72]	0.32	.012
Evidencia y contraargumentación (0–4)	2.84 [2.73, 2.96]	2.36 [2.23, 2.49]	0.48	<.001
Cohesión y coherencia (0–4)	2.96 [2.85, 3.07]	2.51 [2.39, 2.63]	0.45	<.001
Convenciones y voz (0–4)	2.61 [2.51, 2.72]	2.43 [2.31, 2.55]	0.18	.033

Análisis por subgrupos y equidad

La interacción grupo $\times$ desempeño inicial fue significativa ($F(2, 226) = 5.12$, $p = .007$): el tercil inferior (pretest ≤ 6) obtuvo el mayor beneficio (dif. ajustada = 1.84, IC95% [0.98, 2.71], $d \approx 0.62$), el tercil medio 1.11 (IC95% [0.44, 1.78]) y el superior 0.59 (IC95% [−0.09, 1.28]). No se detectaron interacciones significativas por sexo o edad tras corrección BH-FDR ($q > .10$).

Percepciones de carga cognitiva y utilidad

El grupo intervención reportó menor carga ($M = 2.8$, $DE = 0.7$) que el control ($M = 3.1$, $DE = 0.8$), $t(232) = 3.02$, $p = .003$, y mayor utilidad del feedback ($M = 4.1$ vs. 3.6), $t(232) = 4.15$, $p < .001$. Los tamaños de efecto (Hedges g) fueron 0.39 y 0.54, respectivamente. La utilidad se asoció moderadamente con la mejora en Cohesión y coherencia ($r = .31$, $p < .001$).

Integridad académica y coincidencia textual

El índice PIBP se correlacionó negativamente con la coincidencia textual no atribuida en el posttest ($\rho = -.33$, $p < .001$). En el grupo intervención, el porcentaje medio de coincidencia descendió de 12.4% ($DE = 6.1$) a 8.1% ($DE = 4.8$); en el control, de 12.0% ($DE = 5.9$) a 10.9% ($DE = 5.6$). La diferencia de diferencias fue significativa ($\beta = -2.2$, $EE = 0.7$, $p = .002$) en un modelo lineal generalizado con familia cuasi-Poisson.

Verificación de supuestos y análisis de sensibilidad

Los residuos del ANCOVA y del MLM no violaron supuestos de normalidad (Shapiro–Wilk $p > .10$) ni homocedasticidad (Levene $p = .18$). Análisis robustos (Welch; estimadores Huber–White) replicaron los hallazgos. La imputación múltiple (5 iteraciones) para valores faltantes $>5\%$ produjo estimaciones virtualmente idénticas ($\Delta\beta < 0.05$). Exclusiones por protocolo ($n = 6$) no alteraron conclusiones.

Síntesis

En conjunto, la intervención con IAG mejoró la calidad de la escritura argumentativa en EGB, con efectos moderados sobre el puntaje total y mayores en coherencia y contraargumentación, y con impacto más fuerte en estudiantes con menor desempeño



inicial. El protocolo de integridad basada en proceso se asoció a menor coincidencia textual no atribuida, apoyando alternativas pedagógicas a los detectores automáticos. Estos resultados triangulan con la literatura reciente sobre feedback mediado por IAG en escritura (Banihashem et al., 2024; Alnemrat et al., 2025; Zhang, 2025) y con recomendaciones de integridad académica centradas en transparencia y evidencia de proceso (Liang et al., 2023; Nature Machine Intelligence Editorial, 2023; Ardito, 2024).

4. DISCUSIÓN

Síntesis interpretativa de los hallazgos

El presente estudio, situado en colegios fiscales urbanos de EGB en Ecuador, evidenció que una secuencia didáctica breve con IA generativa (IAG) acotada a funciones metacognitivas produjo mejoras moderadas en la calidad de la escritura argumentativa frente a un control pedagógico activo. Las ganancias fueron particularmente visibles en cohesión y coherencia global y en evidencia/contraargumentación, dimensiones críticas para el razonamiento escrito en Lengua y Literatura. Asimismo, la mejora fue más pronunciada en el tercil de desempeño inicial bajo, un patrón de “compensación” coherente con la hipótesis de andamiaje adaptativo de la IAG. En integridad académica, el protocolo de transparencia del proceso (PIBP) se asoció a descensos significativos de la coincidencia textual no atribuida, sugiriendo que la trazabilidad de borradores y prompts puede constituir una alternativa pedagógica válida al uso punitivo de detectores automatizados.

Estos resultados refuerzan la conjetura de que la IAG, cuando se emplea como dispositivo de diagnóstico y guía —y no como “redactor sustituto”—, potencia procesos intermedios de la escritura argumentativa: planificación, verificación de progresión temática, identificación de contraargumentos plausibles y ajuste de coherencia local/global. En clave didáctica, la contribución reside menos en “escribir por el estudiante” que en externalizar la metacognición y multiplicar micro-oportunidades de feedback oportuno, específico y orientado a criterios (Hattie & Timperley, 2007; Wisniewski et al., 2020).

Nuestros hallazgos dialogan de forma estrecha con los resultados de experimentos y cuasi-experimentos recientes sobre feedback mediado por IAG en tareas de escritura. Estudios que comparan retroalimentación IA versus entre pares o docente han reportado efectos comparables o superiores en la mejora de organización y uso de evidencia, especialmente cuando la IA se limita a evaluaciones diagnósticas y preguntas socráticas (Banihashem et al., 2024; Alnemrat et al., 2025; Zhang, 2025). La predominancia de efectos en coherencia y contraargumentación observada en nuestra muestra es consistente con revisiones y meta-análisis sobre evaluación/escritura asistida (Fleckenstein et al., 2023; Zhai et al., 2023; Huang et al., 2025), que subrayan ganancias en componentes estructurales y argumentativos cuando el feedback es inmediato y criterial.

Por otra parte, la mayor eficacia entre estudiantes con menor punto de partida converge con evidencias de que los apoyos algorítmicos tienden a reducir brechas en contextos de alta razón estudiante/docente, al ofrecer andamiaje accesible y frecuente (Doshi et al., 2024). Cabe matizar, sin embargo, que la literatura también alerta sobre riesgos de uniformización estilística si se delega en exceso la generación textual, lo cual justifica la restricción adoptada en nuestro diseño: IAG solo para ideación, diagnóstico y revisión, nunca para redacción final.



En integridad académica, el descenso de la coincidencia textual no atribuida asociado al PIBP sintoniza con recomendaciones internacionales que desaconsejan decisiones punitivas basadas exclusivamente en detectores de IA, dada su vulnerabilidad y sesgo, especialmente contra escritores no nativos (Nature Machine Intelligence Editorial, 2023; Liang et al., 2023; Ardito et al., 2024). Nuestros datos sugieren que hacer visible el proceso —guardar borradores con control de cambios, documentar prompts y razones de edición— es una palanca más efectiva para cultivar autoría responsable que terciarizar la decisión en una caja negra.

Implicaciones teóricas

Desde el punto de vista teórico, el patrón de resultados apoya modelos de autorregulación de la escritura en los que el feedback funciona como señal de control sobre la brecha entre el estado actual del texto y los criterios de calidad (Hattie & Timperley, 2007). La IAG, en su configuración acotada, operó como agente regulador que reifica criterios —“¿dónde está la tesis?”, “¿qué contraargumento falta?”, “¿cómo progresa el tema?”—, activando ciclos de planificación-monitoreo-revisión. Ello explica la mayor sensibilidad en dimensiones estructurales de la rúbrica (coherencia, evidencia) frente a aspectos más idiosincráticos del estilo. Además, los efectos diferenciales por nivel inicial avalan hipótesis de sensibilidad a la necesidad: quienes parten con menor dominio encuentran mayor valor marginal en checklists y preguntas-guía automatizadas, mientras que estudiantes avanzados podrían requerir feedback más fino (p. ej., calidad de evidencias, matices retóricos).

En integridad, los resultados abonan a un marco conceptual que desplaza el énfasis de la detección a la prevención y formación de criterios éticos mediante transparencia de procesos. El PIBP, más que un “candado”, se comportó como instrumento formativo que incentiva reflexión y trazabilidad, alineado con evaluaciones auténticas y con la literatura sobre confiabilidad limitada de detectores IA (Liang et al., 2023).

Implicaciones prácticas para EGB en colegios fiscales urbanos

Para equipos de Lengua y Literatura, el mensaje práctico es claro: integrar IAG sí, pero con límites y con criterios explícitos. Con base en nuestros resultados, proponemos:

1. Secuencias cortas y focalizadas (tres sesiones) donde la IAG cumple funciones de mentoría socrática: sugerir contraargumentos comunes, detectar saltos lógicos y recordar criterios de la rúbrica. Evitar la generación de párrafos finales.
2. Rúbricas analíticas visibles y usadas como “contratos didácticos”; la IAG debe referenciar explícitamente estas dimensiones.
3. Protocolo de integridad basada en proceso (PIBP) obligatorio cuando se use IAG: declaración de uso, bitácora de cambios, borradores con control de cambios y capturas de prompts. Evaluar el cumplimiento como parte del proceso, no del producto.
4. Capacitación breve a docentes: no se trata de “vigilar” con detectores, sino de enseñar a documentar el proceso y a formular prompts que activen reflexión (no redacción automática).
5. Equidad y focalización: priorizar el uso de IAG en estudiantes con rezago argumentativo, en tanto el efecto marginal es mayor; monitorear posibles efectos techo en los más avanzados.

Estas recomendaciones son factibles en aulas numerosas y con tiempos acotados, pues

externalizan pasos metacognitivos sin agregar carga excesiva (nuestros datos muestran incluso menor carga percibida respecto al control).

Aunque la asignación por paralelos mitiga la contaminación, persisten amenazas a la validez interna: (a) potenciales diferencias no observadas entre aulas (p. ej., clima de aula, expectativas del docente), parcialmente controladas con modelos jerárquicos; (b) maduración/efectos de historia en el periodo de intervención; (c) fidelidad de implementación, que si bien fue monitoreada, pudo variar entre docentes. La validez de constructo puede verse tensionada si los evaluadores, pese al doble ciego, infirieron el grupo por rasgos estilísticos; no obstante, el CCI alto y los análisis Bland–Altman reducen esta preocupación.

En validez externa, el estudio se circunscribe a colegios fiscales urbanos de la Costa; la generalización a Sierra/Amazonía o a planteles rurales requiere cautela, considerando brechas de conectividad y diferencias curriculares. Asimismo, el periodo de exposición fue corto (tres sesiones); es plausible que efectos a más largo plazo difieran, ya sea por consolidación (mayores beneficios) o por habituación (menores incrementos).

Primero, la decisión ética de no permitir generación de texto final por la IAG da lugar a un estimador conservador del potencial total de la herramienta; sin embargo, es coherente con el objetivo formativo de fortalecer la autoría y con advertencias sobre uniformización estilística (Doshi et al., 2024). Segundo, aun con ajustes por pretest y modelos mixtos, no es un RCT individual, por lo que pueden persistir confusores residuales. Tercero, el uso de un solo tema de posttest, aunque paralelo, puede limitar la transferencia a otros géneros o tópicos; futuros estudios debieran incluir tareas múltiples y delayed post-tests. Cuarto, la medición de coincidencia textual depende de configuraciones del software y bases de datos, lo que puede introducir variabilidad; por ello, las inferencias se centran en diferencias relativas y en la asociación con el PIBP, no en umbrales absolutos.

Líneas de investigación futura

1. Ensayos por conglomerados con aleatorización a nivel de centro y seguimiento semestral para estimar efectos sostenidos y spillovers entre aulas.
2. Intervenciones adaptativas donde la IAG ajuste el nivel de scaffolding a métricas en tiempo real (p. ej., señales de coherencia o densidad argumental), explorando si se mantienen efectos diferenciales por nivel inicial.
3. Análisis cualitativos de proceso (p. ej., trazas de edición y entrevistas) para comprender cómo los estudiantes negocian el feedback automatizado y qué estrategias retóricas emergen.
4. Estudios de equidad focalizados en subpoblaciones (p. ej., estudiantes con NEE o EIB) para verificar que la IAG no introduzca nuevas brechas y, de ser necesario, calibrar prompts culturalmente pertinentes.
5. Evaluación de costos y factibilidad a escala institucional, incluyendo formación docente, mantenimiento tecnológico y gobernanza de datos.
6. Comparación de protocolos de integridad (PIBP vs. detectores vs. evaluación de proceso por portafolio) en resultados de aprendizaje y percepción de justicia.



Consideraciones éticas y de política educativa

La política institucional debería priorizar normas de uso responsable que especifiquen funciones permitidas de la IAG (ideación, diagnóstico, revisión) y obligaciones de transparencia (PIBP). En lugar de sanciones apoyadas en probabilidades opacas de “texto IA”, se recomienda un enfoque de evaluación auténtica con evidencias de proceso y rúbricas explícitas, alineado con recomendaciones editoriales y metodológicas (Nature Machine Intelligence Editorial, 2023; Liang et al., 2023; Ardito et al., 2024). En términos de protección de datos, es crucial evitar el envío masivo de textos estudiantiles a servicios externos sin garantías de privacidad; cuando sea inevitable, usar configuraciones locales o con acuerdos de tratamiento de datos adecuados a normativa ecuatoriana.

Cierre: de la hipótesis a la praxis

Las tres hipótesis planteadas encuentran respaldo empírico en los datos: (H1) se confirma la mejora diferencial a favor del grupo IAG en el puntaje total y, con mayor claridad, en cohesión y contraargumentación; (H2) el protocolo PIBP se asocia a menor coincidencia textual no atribuida, lo que respalda enfoques centrados en proceso por sobre detectores automatizados; (H3) el efecto es más intenso en quienes parten con menor desempeño, lo que sugiere un uso estratégico de la IAG para nivelar el piso. De cara a la praxis, la lección es pragmática: integrar IAG con propósito didáctico claro, limitar su función a activar metacognición, y sostener transparencia del proceso como salvaguarda de la autoría. Con ello, la escuela pública urbana del Ecuador puede canalizar el potencial de la IA sin renunciar a su misión formativa central: que las y los estudiantes piensen, argumenten y escriban con voz propia.

5. CONCLUSIONES

La incorporación acotada de IA generativa (IAG) en la enseñanza de la escritura argumentativa en Educación General Básica de colegios fiscales urbanos del Ecuador demuestra efectos consistentes y pedagógicamente relevantes. El estudio confirma que una secuencia breve, centrada en funciones metacognitivas de ideación, diagnóstico y revisión, eleva la calidad del texto argumentativo por sobre las prácticas tradicionales con control activo. Las ganancias se concentran en cohesión y coherencia global y en evidencia/contraargumentación, dimensiones neurálgicas del razonamiento escrito en Lengua y Literatura, sin desplazar la autoría hacia la herramienta.

El enfoque de integridad basada en proceso —declaración de uso, bitácora de cambios, borradores con control de cambios y documentación de prompts— se muestra efectivo para reducir la coincidencia textual no atribuida, ofreciendo una alternativa formativa a los detectores automáticos de “texto IA”, cuya fiabilidad es disputada en la literatura reciente. En consecuencia, la evaluación se alinea con principios de transparencia y autenticidad, favoreciendo un aprendizaje responsable más allá de la mera disuasión.

La equidad emerge como aporte central: el impacto es mayor en estudiantes con desempeño inicial bajo, lo que sugiere que la IAG, usada como andamiaje, contribuye a nivelar brechas en aulas numerosas, donde el tiempo de retroalimentación individual es escaso. Este patrón respalda la adopción focalizada de la herramienta para quienes más la requieren, sin detrimento del acompañamiento docente.





En términos de originalidad, el trabajo vincula tres elementos poco estudiados en conjunto a nivel EGB y en contexto ecuatoriano: (i) una implementación acotada de IAG orientada a la metacognición; (ii) una medición ciega con análisis estadístico robusto a la estructura por aulas; y (iii) un protocolo de integridad centrado en trazabilidad del proceso más que en detección algorítmica. Esta triada ofrece un modelo replicable de intervención breve con alto retorno pedagógico y factible en planteles fiscales urbanos.

Desde la perspectiva de política y gestión escolar, el estudio sustenta lineamientos concretos: (a) especificar usos permitidos de IAG (planificación, diagnóstico, revisión) y usos prohibidos (redacción final y generación de referencias); (b) institucionalizar el PIBP como condición de uso; (c) visibilizar rúbricas analíticas como contratos didácticos y criterios de referencia para todo feedback; y (d) priorizar formación docente centrada en diseño de prompts que activen reflexión y en evaluación auténtica del proceso.

En relación con el currículo nacional de Lengua y Literatura, la intervención se inserta sin interrupciones, pues potencia aprendizajes clave (planteamiento de tesis defendible, organización textual, uso crítico de fuentes, cohesión y coherencia). Los efectos observados se obtienen en tres sesiones, por lo que resultan compatibles con cargas horarias ajustadas y con las dinámicas de cursos numerosos propios de los colegios fiscales urbanos.

Las limitaciones del estudio —asignación por paralelos, duración breve, dependencia de conectividad— acotan la generalización y deben ser consideradas al extrapolar los hallazgos a otras regiones o niveles. Aun así, la convergencia entre análisis ANCOVA y modelos mixtos, la concordancia interevaluador alta y la replicación de resultados en análisis robustos aumentan la confianza en la dirección y tamaño del efecto.

Como proyección, se plantea: (1) escalar la intervención en ensayos por conglomerados a nivel institucional y con seguimiento de mediano plazo para estimar persistencia de efectos; (2) desarrollar versiones adaptativas del andamiaje con IAG que ajusten la intensidad del feedback a señales del desempeño; (3) incorporar métodos mixtos para comprender cómo se negocia en la práctica el feedback automatizado y cómo incide en la voz autorial; y (4) comparar protocolos de integridad (PIBP, portafolios de proceso, uso de detectores con salvaguardas) en resultados de aprendizaje y percepción de justicia evaluativa.

En suma, integrar IAG con límites claros y transparencia del proceso se perfila como una estrategia viable, ética y efectiva para fortalecer la escritura argumentativa en EGB en contextos fiscales urbanos del Ecuador. La escuela pública puede aprovechar el potencial de la IA como mentor socrático que activa la metacognición, mientras preserva la autoría estudiantil y cultiva competencias de pensamiento crítico imprescindibles para la ciudadanía. Este enfoque ofrece una hoja de ruta pragmática para innovar en Lengua y Literatura sin sacrificar la esencia formativa del aula.

6. REFERENCIAS BIBLIOGRÁFICAS

Alnemrat, A., Aldamen, H., Almashour, M., Al-Deaibes, M., & AlSharefeen, R. (2025). AI vs. teacher feedback on EFL argumentative writing: A quantitative study. *Frontiers in*



- Education, 10, 1614673. <https://doi.org/10.3389/feduc.2025.1614673>
- Arce, C. M., Ureña, R., & Játiva, J. (2025). Attitudes toward AI and dependency among Ecuadorian university students: A predictive model. *Sustainability*, 17(17), 7741. <https://doi.org/10.3390/su17177741>
- Ardito, C. G., Gopal, T., Ceppini, E., & Burnett, P. C. (2024). Generative AI detection in higher education assessments. *New Directions for Teaching and Learning*, 2024(180), 39–54. <https://doi.org/10.1002/tl.20624>
- Baldeón Medina, P. J., González Yagual, T. B., Izquierdo Corozo, C. D., & Villalba Vélez, K. M. (2025). Aplicación de IA en Educación Básica para fortalecer el aprendizaje y la motivación escolar. *Horizonte Científico International Journal*, 3(2), 1–9. <https://doi.org/10.64747/dbkg2w76>
- Banihashem, S. K., Nejati, R., Sadeghi, M. R., & Babaii, E. (2024). Feedback sources in essay writing: Peer-generated or AI-generated? *International Journal of Educational Technology in Higher Education*, 21, 26. <https://doi.org/10.1186/s41239-024-00455-4>
- Bland, J. M., & Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *The Lancet*, 327(8476), 307–310. [https://doi.org/10.1016/S0140-6736\(86\)90837-8](https://doi.org/10.1016/S0140-6736(86)90837-8)
- Buele, J., Sabando-García, Á. R., Sabando-García, B. J., & Yáñez-Rueda, H. (2025). Ethical use of generative artificial intelligence among Ecuadorian university students. *Sustainability*, 17(10), 4435. <https://doi.org/10.3390/su17104435>
- Castro Macías, N. M. (2025). Aplicación de la inteligencia artificial como recurso para desarrollar pensamiento crítico en clases de Física. *Horizonte Científico International Journal*, 3(2), 1–19. <https://doi.org/10.64747/7wvch719>
- Doshi, A. R., Yamauchi, T., Salehi, N., & Bernstein, M. S. (2024). Generative AI enhances individual creativity but reduces collaboration quality. *Science Advances*, 10(44), eadn5290. <https://doi.org/10.1126/sciadv.adn5290>
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). GPower 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39*(2), 175–191. <https://doi.org/10.3758/BF03193146>
- Fleckenstein, J., Liebenow, L. W., & Meyer, J. (2023). Automated feedback and writing: A multi-level meta-analysis of effects on students' performance. *Frontiers in Artificial Intelligence*, 6, 1162454. <https://doi.org/10.3389/frai.2023.1162454>
- Galeas Gaibor, M. L., Meza Bravo, D. M., Ocampo Andrade, W. F., & Romero Arias, J. B. (2025). Aplicación de inteligencia artificial y gamificación para fortalecer el aprendizaje en estudiantes de básica en contextos rurales. *Horizonte Científico International Journal*, 3(2), 1–12. <https://doi.org/10.64747/f718q395>
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112. <https://doi.org/10.3102/003465430298487>
- Huang, Y., Wu, C., & Lin, S. (2025). Exploring the effectiveness of large-scale automated writing evaluation in education: A systematic review. *Journal of Computer Assisted Learning*, 41(3), 987–1005. <https://doi.org/10.1111/jcal.70009>



- Liang, W., Yuksekgonul, M., Mao, Y., Wu, E., & Zou, J. (2023). GPT detectors are biased against non-native English writers. *Patterns*, 4(7), 100779. <https://doi.org/10.1016/j.patter.2023.100779>
- Nature Machine Intelligence Editorial. (2023). The AI writing on the wall. *Nature Machine Intelligence*, 5(1), 1. <https://doi.org/10.1038/s42256-023-00613-9>
- Nguyen, A., & Wang, X. (2024). Human–AI collaboration patterns in AI-assisted academic writing. *Studies in Higher Education*, 49(10), 1991–2012. <https://doi.org/10.1080/03075079.2024.2323593>
- Nordstokke, D. W., & Zumbo, B. D. (2010). A new nonparametric Levene test for equal variances. *Psicologica*, 31(2), 401–430. <https://doi.org/10.2478/v10053-008-0081-1>
- Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, 86(2), 420–428. <https://doi.org/10.1037/0033-2909.86.2.420>
- The jamovi project. (2022). jamovi (Version 2.3) [Computer software]. <https://doi.org/10.17605/OSF.IO/GXRQ9>
- Wisniewski, B., Zierer, K., & Hattie, J. (2020). The power of feedback revisited: A meta-analysis of the effects of feedback in educational contexts. *Educational Research Review*, 30, 100331. <https://doi.org/10.1016/j.edurev.2020.100331>
- Wilson, J., Zhang, F., Palermo, C., Cruz Cordero, T., Myers, M. C., Eacker, H., Potter, A., & Coles, J. (2024). Predictors of middle school students' perceptions of automated writing evaluation. *Computers & Education*, 211, 104985. <https://doi.org/10.1016/j.compedu.2023.104985>
- Xue, Y., Wang, Z., & Li, J. (2024). Towards automated writing evaluation: A comprehensive bibliometric review. *Education and Information Technologies*, 29(7), 8895–8919. <https://doi.org/10.1007/s10639-024-12596-0>
- Zhai, N., Xie, F., Zou, D., & Wang, F. L. (2023). The effectiveness of automated writing evaluation on students' writing: A meta-analysis. *Journal of Educational Computing Research*, 61(2), 363–394. <https://doi.org/10.1177/07356331221127300>
- Zhang, K. (2025). Enhancing critical writing through AI feedback: A randomized control study. *Behavioral Sciences*, 15(5), 600. <https://doi.org/10.3390/bs15050600>

Conflicto de Intereses: Los autores declaran que no tienen conflictos de intereses relacionados con este estudio y que todos los procedimientos seguidos cumplen con los estándares éticos establecidos por la revista.

Asimismo, confirman que este trabajo es inédito y no ha sido publicado, ni parcial ni totalmente, en ninguna otra publicación