

HCEIJ



2025

Horizonte Científico Educativo

Vol. 1 Num. 1

INTERNATIONAL JOURNAL

Enero a Junio



Semestral - Español e Inglés



ID del documento: HCEIJ-Vol.1.Nº1.001.2025

Tipo de artículo: Investigación

Estrategias de lectura multimodal con TikTok y podcasts para mejorar comprensión lectora en 8.º–10.º EGB

Multimodal reading strategies with TikTok and podcasts to improve reading comprehension in grades 8–10 of basic education

Autores:

Lenin Hochimin Tenecela Calderón

Escuela Superior Politécnica del Litoral, Guayaquil-Ecuador, lhtenece@espol.edu.ec
<https://orcid.org/0009-0006-7086-878X>

Marlon Elinovich Tenecela Calderón

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

María Fernanda Rodríguez Ruiz

Universidad de Guayaquil, Guayaquil-Ecuador, azulfr Ruiz@gmail.com,
<https://orcid.org/0009-0005-0069-8618>

Denisse Margarita Posligua Garcia

Instituto Tecnológico Superior Bolivariano de Tecnología, Guayaquil-Ecuador, denisse-posligua@hotmail.com, <https://orcid.org/0009-0004-0592-3530>

Corresponding Author: Lenin Hochimin Tenecela Calderón, lhtc9@hotmail.com

Reception: 10-enero-2025 **Acceptance:** 03- febrero -2025 **Publication:** 14- febrero -2025

How to cite this article:

Estrategias de lectura multimodal con TikTok y podcasts para mejorar comprensión lectora en 8.º–10.º EGB. (n.d.). Horizonte Científico Educativo International Journal, 1(1), 1-19. <https://doi.org/10.64747/czxf3r37>



RESUMEN

Este estudio evaluó el efecto de una secuencia de lectura multimodal —integrando podcasts y videos breves tipo TikTok con señalización, pausas con preguntas y producción guiada— sobre la comprensión lectora de estudiantes de 8.º–10.º de Educación General Básica (EGB) en Ecuador. Se implementó un diseño cuasi-experimental multibrazo con medición pre–post y construcción del contrafactual mediante emparejamiento por puntaje de propensión (PSM) multinomial y diferencias-en-diferencias (DiD). Cuatro condiciones a nivel de aula: C0 (práctica monomodal, control), C1 (podcasts), C2 (videos cortos) y C3 (combinado). La muestra analítica incluyó 112 aulas y 3,276 estudiantes de contextos urbanos y rurales, con conectividad heterogénea. Los instrumentos abarcaron pruebas de comprensión (literal, inferencial, crítica) calibradas con IRT, rúbricas de producción multimodal y cuestionarios de procesos (autoeficacia y estrategias). Los resultados muestran mejoras significativas en comprensión total para C3 frente a C0 (0.24 DE; IC95% [0.17, 0.31]), así como incrementos en la probabilidad de alcanzar nivel de logro alto (+7.8 pp; IC95% [5.2, 10.4]). C1 (0.16 DE) y C2 (0.13 DE) también superaron al control. La comparación C3 vs {C1, C2} evidenció sinergia adicional (+0.07 DE). Las mayores ganancias se observaron en comprensión inferencial y crítica, y la mediación indicó que la autoeficacia lectora (~28% del efecto total) y las estrategias metacognitivas (~9%) contribuyeron al impacto. La intervención fue viable en escuelas con baja conectividad y no amplificó la varianza del rendimiento. Se concluye que secuencias multimodales offline-friendly pueden mejorar la comprensión lectora en EGB Superior a costos manejables. Se recomienda institucionalizar ciclos de 8 semanas, fortalecer la formación docente en señalización y evaluación formativa, y mantener repositorios abiertos de recursos curados. Futuras investigaciones deben estimar persistencia a 6–12 meses y costo-efectividad para escalamiento.

Palabras clave: lectura multimodal; podcasts educativos; TikTok; comprensión lectora; diferencias-en-diferencias

ABSTRACT

This study assessed the impact of a multimodal reading sequence—combining short educational podcasts and TikTok-style videos with signaling, embedded questions, and guided production—on reading comprehension among 8th–10th-grade students in Ecuador’s Basic Education system. We implemented a multi-arm, pre–post quasi-experimental design using multinomial propensity score matching (PSM) and difference-in-differences (DiD) to build a credible counterfactual. Four classroom-level conditions were compared: C0 (business-as-usual, control), C1 (podcasts), C2 (short videos), and C3 (combined). The analytical sample comprised 112 classes and 3,276 students from urban and rural settings with heterogeneous connectivity. Instruments included IRT-calibrated comprehension tests (literal, inferential, critical), a rubric for multimodal production, and process questionnaires (self-efficacy and metacognition). Findings indicate significant improvements for C3 vs C0 in overall comprehension (0.24 SD; 95% CI [0.17, 0.31]) and in the probability of reaching a high achievement level (+7.8 pp; 95% CI [5.2, 10.4]). C1 (0.16 SD) and C2 (0.13 SD) also outperformed the control. The C3 vs {C1, C2} contrast revealed synergy (+0.07 SD). Gains were larger in inferential and critical comprehension. Mediation analyses suggested contributions from reading self-efficacy (~28% of the total effect) and metacognitive strategies (~9%). The intervention proved feasible in low-connectivity schools and did not widen within-class variance. We conclude that offline-friendly multimodal sequences can meaningfully improve





reading comprehension in upper Basic Education at manageable costs. Policy and practice should institutionalize 8-week cycles, strengthen teacher professional development in signaling and formative assessment, and sustain open repositories of curated resources. Future studies should estimate 6–12-month persistence and cost-effectiveness to inform scaling.

Keywords: multimodal reading; educational podcasts; TikTok; reading comprehension; difference-in-differences

1. INTRODUCCIÓN

Contexto y problema. En Ecuador, la comprensión lectora en Educación General Básica (EGB) Superior —8.º a 10.º— sigue siendo un nudo crítico para la progresión curricular y el desempeño transversal en ciencias sociales y naturales. Aun cuando la escolaridad postpandemia ha normalizado prácticas híbridas, la alfabetización académica continúa anclada en textos monomodales, con escaso aprovechamiento pedagógico de formatos digitales que ya forman parte de las prácticas culturales de las y los adolescentes (p. ej., videos breves y audio bajo demanda). En la jerga de aula —atado cortito—, la brecha entre lo que los chicos consumen fuera de la escuela (TikTok, podcasts) y lo que se les pide dentro (lecturas planas) genera desajustes motivacionales y cognitivos. La pregunta es si una secuencia de “lectura multimodal” que integre videos cortos (TikTok) y podcasts puede movilizar estrategias de comprensión más profundas y transferibles, sin diluir la exigencia textual.

Relevancia científica y social. La evidencia reciente sitúa las “alfabetizaciones multimodales” como un campo promisorio para articular decodificación, construcción de significado y pensamiento crítico en contextos de alta penetración digital (Si, 2022; Watts-Taffe, 2022; Lim, 2022). En paralelo, la investigación sobre videos educativos y medios breves muestra principios de diseño (carga cognitiva, señalización, segmentación) asociados a mejores aprendizajes cuando se acompaña de tareas activas (Zolkwer, 2023). A la vez, estudios sobre plataformas de video corto reportan tanto potencial didáctico —p. ej., #BookTok como comunidad lectora— como riesgos de distracción y procesamiento superficial que deben mitigarse con andamiajes explícitos (Yélamos-Guerra et al., 2022; Jerasa et al., 2024; Gong et al., 2024). Con respecto a audio, la literatura sobre podcasting educativo documenta ganancias de comprensión y retención cuando el diseño instruccional activa recuperación, preguntas intercaladas y tareas de aplicación (Song et al., 2018; Johnson et al., 2024; Patel et al., 2025). Este estudio aporta, por tanto, a un debate con implicaciones directas para política curricular y desarrollo profesional docente en EGB.

Marco teórico. Adoptamos un enfoque de alfabetización multimodal que concibe la lectura como una práctica semiótica situada en múltiples modos (verbal, visual, auditivo) y canales, donde el lector coordina estrategias de localización de información, inferencia, monitoreo metacognitivo y evaluación crítica de fuentes (Lim, 2022; Guo et al., 2024). Desde la psicología cognitiva, nos apoyamos en teorías de aprendizaje multimedia y en el principio de contigüidad/coherencia para reducir carga extrínseca y optimizar la integración texto-imagen-audio en tareas de comprensión (Zolkwer, 2023). Incorporamos, además, la noción de “alfabetización algorítmica” para orientar la lectura crítica de contenidos de





plataformas sociales —sesgos de recomendación, señalización de credibilidad y trazabilidad de fuentes— como parte del trabajo lector (Bell & Ahmed, 2023; Jerasa et al., 2024). Finalmente, consideramos evidencia mixta sobre diferencias de desempeño entre lectura en papel y digital —con efectos modulados por tipo de texto y apoyos metacognitivos— para fundamentar los controles de diseño (Li et al., 2024; Schwabe & Krüger, 2022; Lombardo et al., 2025).

En el estado del arte aplicado a EGB, en K-12, tres tendencias son relevantes. Primero, las revisiones sistemáticas señalan que la instrucción multimodal produce mejoras en comprensión cuando no sustituye, sino que complementa la lectura de textos con tareas de producción y discusión (Si, 2022). Segundo, intervenciones con storytelling digital y recursos audiovisuales muestran incrementos en comprensión literal e inferencial cuando existe evaluación formativa y práctica guiada (Rahmanu et al., 2024; Akdoğan, 2023; Istiq'faroh et al., 2025). Tercero, experiencias con #BookTok y edutok sugieren que la curaduría docente, la guía sobre etiquetado/hashtags y la “transducción” de formatos (del video al texto y viceversa) pueden activar interés y profundizar procesos de interpretación (Yélamos-Guerra et al., 2022; Jerasa et al., 2024). No obstante, también se documentan riesgos de uso intensivo de videos cortos asociados a menor control atencional, lo que exige diseños con pausas, preguntas insertadas y metas de lectura claras (Xie et al., 2023; Gong et al., 2024).

Para garantizar uso condicionado por contexto —serranía o costa, urbano o rural—, el estudio se diseña “genérico Ecuador”: podrá implementarse en cualquier provincia (Régimen Sierra o Costa) porque los insumos son abiertos y multiplataforma. Como fuentes públicas se emplearán: (i) repositorios escolares abiertos de textos informativos (p. ej., ministeriales y de organismos multilaterales con DOI); (ii) clips de TikTok educativos de carácter público curados con criterios de idoneidad y derechos; (iii) podcasts breves (≤ 10 min) de divulgación escolar (ciencias/lengua) con licencia abierta; y (iv) bancos de ítems de comprensión alineados con los descriptores de 8.º–10.º EGB. La intervención no depende de conectividad constante —se prevé descarga previa y uso offline—, aspecto clave en centros con conectividad limitada.

Objetivo general. Evaluar el efecto de una secuencia de enseñanza de lectura multimodal —que integra tareas con TikTok y podcasts— sobre la comprensión lectora de estudiantes de 8.º–10.º de EGB en Ecuador, comparada con instrucción habitual centrada en textos monomodales.

Objetivos específicos. (1) Estimar el impacto en comprensión literal, inferencial y crítica; (2) medir cambios en autoeficacia lectora y en estrategias metacognitivas; (3) analizar la contribución diferencial de los componentes video corto vs. podcast; (4) evaluar efectos heterogéneos por género, modalidad (urbano/rural) y conectividad escolar; (5) estudiar la transferencia a producción multimodal (resúmenes audio-texto, reseñas BookTok de textos curriculares).

Hipótesis. H1 (aprendizaje): La secuencia multimodal generará una mejora media $\geq 0,20$ DE en comprensión lectora frente al control, con mayores efectos en tareas inferenciales. H2 (mecanismos): Parte del efecto total estará mediado por autoeficacia lectora y estrategias de monitoreo ($\geq 25\%$). H3 (componente): El subcomponente podcast potenciará inferencias globales y síntesis; el subcomponente video corto potenciará localización y macro-estructura



cuando esté debidamente señalizado. H4 (equidad): En escuelas con menor conectividad, el diseño offline-friendly producirá efectos comparables, reduciendo brechas internas.

En el plano científico, el estudio contribuye a precisar condiciones de efectividad de la lectura multimodal en EGB —duración, combinatoria de modos, y andamiajes— con medidas sólidas de comprensión y procesos. En el plano práctico, ofrece un paquete replicable (guías, rúbricas, repositorio) alineado con perfiles de salida de EGB, con especial énfasis en el uso responsable y crítico de plataformas de amplia adopción juvenil. Atado cortito: se trata de llevar las prácticas culturales de las y los pelados al aula, pero con rigor didáctico y evaluación seria.

2. METODOLOGÍA

Diseño del estudio. Se implementó un cuasi-experimento de múltiples brazos con grupos no equivalentes y medición repetida (pre–post) en aulas de 8.º–10.º de EGB. El objetivo fue estimar el efecto de una secuencia de lectura multimodal y desagregar la contribución de sus componentes. Para ello se trabajó con cuatro condiciones de instrucción a nivel de aula (unidad de asignación): (C0) práctica habitual centrada en texto monomodal (control); (C1) podcasts + andamiajes de escucha y síntesis; (C2) videos cortos tipo TikTok curados + andamiajes de señalización y transducción; (C3) combinado (podcasts + TikTok) con integración explícita. La asignación de brazos se realizó por factibilidad escolar y preferencia directiva, por lo que el contrafactual se construyó mediante emparejamiento por puntaje de propensión (PSM) seguido de diferencias-en-diferencias (DiD) y, en sensibilidad, por ponderación por probabilidad inversa de tratamiento multinomial (Austin, 2011; Rosenbaum y Rubin, 1983; Callaway y Sant’Anna, 2021).

Ámbito y población. El protocolo es genérico para Ecuador; es replicable en instituciones fiscales, fiscomisionales o particulares de cualquier provincia y régimen escolar (Sierra/Amazonía o Costa/Galápagos). Criterios de inclusión: (a) impartir 8.º, 9.º o 10.º de EGB; (b) disponer de al menos dos aulas por grado para permitir un brazo de intervención y uno de comparación dentro del mismo centro o circuito; (c) aceptar los resguardos éticos y de uso de plataformas. Se incluyeron centros urbanos y rurales, con conectividad heterogénea. Las coordenadas se registraron a nivel cantonal, sin identificación pública del plantel.

Objeto de estudio. Estudiantes matriculados en 8.º–10.º de EGB (11–15 años). La unidad de análisis primaria fue el estudiante (resultados de comprensión) con errores agrupados a nivel aula (asignación y tratamiento). Las y los docentes de Lengua fueron agentes implementadores y respondieron instrumentos de proceso/fidelidad.

Intervención didáctica y componentes. Duración: 8 semanas (3 sesiones de 45–50 min/semana). Diseño offline-friendly: todos los recursos (clips de 30–90 s y podcasts de 6–10 min) se descargaron previamente en formato abierto.

- Andamiajes comunes: activación de conocimientos previos, objetivos de lectura/escucha, preguntas intercaladas y verificación por recuperación; rúbricas analíticas de comprensión (literal, inferencial, crítica) y de producción multimodal (coherencia, adecuación, citación, ética digital).

- Componente podcasts (C1): escucha segmentada con señalización auditiva y visual (guías), mapas de ideas, tarea de resumen audio-texto (transducción) y discusión guiada.
- Componente videos cortos (C2): visualización con señalización (flechas, palabras clave), tablillas “qué veo / qué infiero / qué evidencias” y reseña #BookTok conectada a un texto escrito del currículo.
- Combinado (C3): alternancia quincenal C1–C2 + secuencias de integración (p. ej., escuchar un podcast, localizar afirmaciones y verificarlas con fragmentos de video o infografías, y elaborar respuesta escrita con citas). Fidelidad de implementación: taller inicial de 6 h y acompañamiento semanal remoto (30–45 min); observaciones externas con listas de chequeo (estándar $\geq 80\%$ de adherencia) y notas de campo.

Fuentes de datos

- 1) Pruebas de comprensión lectora (pre y post): 24–28 ítems por aplicación, con pasajes + estímulos multimodales; selección múltiple y respuesta breve, calibradas con modelos IRT (2PL). Subescalas: literal, inferencial y crítica.
- 2) Cuestionarios de procesos (estudiantes): autoeficacia lectora (6 ítems, Likert 1–5), estrategias metacognitivas (planificación, monitoreo, reparación; 9 ítems), actitudes hacia textos/audio/video (6 ítems) y uso responsable (4 ítems).
- 3) Cuestionario docente: experiencia, formación en Lengua, uso previo de recursos digitales, percepción de carga/viabilidad.
- 4) Fidelidad y clima: registros de observación, conteo de tiempos dedicados a lectura/escucha, discusión y producción.
- 5) Contexto escolar: AMIE/plantel (tamaño de clase, jornada, modalidad), índice de conectividad escolar y entorno (urbano/rural).

Variables.

- Primarias: puntaje estandarizado (media 0, DE 1) de comprensión total; proporción con nivel alto ($\geq P66$ post-test).
- Secundarias: subescalas, producción multimodal (rúbrica 0–3), autoeficacia y estrategias (escalas z).
- Tratamiento: indicador multinomial (C0, C1, C2, C3) a nivel aula; exposición efectiva (% de sesiones).
- Covariables: sexo, edad, lengua en el hogar, repitencia, NSE proxy (educación parental, elegibilidad de transferencias), acceso doméstico a dispositivos, conectividad escolar, experiencia docente, tamaño de clase y estrato urbano/rural.

Muestreo. Muestreo por conglomerados en dos etapas. Etapa 1: selección de instituciones dentro de estratos de provincia×urbano/rural con probabilidad proporcional al tamaño. Etapa 2: selección de aulas completas de 8.º–10.º. Se procuró que cada sede participe con ≥ 2 aulas por grado para facilitar contrafactual interno. El tamaño muestral mínimo por brazo se determinó por potencia y factibilidad.

Tamaño muestral y potencia. Se planificó detectar 0.20 DE (C3 vs C0) con $\alpha=0.05$, potencia 0.80, ICC=0.06 a nivel aula y correlación pre–post $p\approx 0.60$. Con 30 estudiantes/aula, se

requieren ≈ 28 aulas por brazo ($\approx 3,360$ estudiantes para cuatro brazos), considerando DEFF y 10% de pérdidas. Para C1 vs C0 y C2 vs C0 la potencia ≥ 0.75 bajo supuestos similares.

Procedimiento.

- Semana -3 a -1: selección y capacitación docente; pre-registro; curaduría/descarga de audios y clips; pilotaje; consentimientos.
- Semana 0: línea base (pruebas y cuestionarios) y contexto escolar.
- Semanas 1–8: implementación según brazo; observaciones de fidelidad (≥ 2 por aula); soporte didáctico.
- Semana +1: post-test y cuestionarios; entrevistas breves a docentes.
- Semana +2/+3: validación de bases, resolución de discrepancias y documentación.

Clips y podcasts se seleccionaron por idoneidad, licencia abierta o uso legítimo con cita, pertinencia cultural y lenguaje apropiado. Cuando no existía licencia abierta, se generaron micro-clips originales (≤ 30 s) basados en textos de dominio público. Se elaboraron fichas de metadatos (fuente, autoría, licencia, duración, transcripción).

Análisis estadístico.

- 1) PSM multinomial (o IPTW multinomial) para balancear covariables; verificación de balance con diferencias estandarizadas (< 0.10) y love plots.
- 2) DiD sobre muestra emparejada/ponderada: $Y_{it} = \alpha + \sum_k \tau_k (1\{C_k\} \times \text{Post}_t) + \gamma_i + \delta_t + X'_{it}\theta + \epsilon_{it}$, con errores agrupados por aula. Para logro alto: LPM robusto y, en sensibilidad, logit con efectos marginales.
- 3) Comparaciones planificadas: C1 vs C0, C2 vs C0, C3 vs C0, y C3 vs {C1,C2}.
- 4) Heterogeneidad: conectividad escolar (percentiles), urbano/rural, género y experiencia docente.
- 5) Mediación: autoeficacia y estrategias (bootstrapping 5,000 réplicas).
- 6) Multiplicidad: FDR de Benjamini-Hochberg.
- 7) Robustez: tendencias por sede, event-study, trimming 5–10%, CBPS y análisis por protocolo.

Datos faltantes y medición. Diagnóstico MCAR/MAR; imputación múltiple por cadenas de Markov ($m \geq 20$) y reglas de Rubin cuando proceda. Invarianza pre–post con IRT (anclajes) y CFA de tres factores correlacionados para subescalas.

Software y reproducibilidad. Análisis en R (MatchIt, WeightIt, cbps, fixest, did, mediation) y/o Stata (teffects psmatch, tebalance, reghdfe, csdid). Repositorio abierto con scripts y materiales; plan de análisis pre-registrado. Reporte APA 7; IC95% y tamaños de efecto estandarizados.

Ética, privacidad y seguridad digital. Consentimiento de familias y asentimiento estudiantil. Consumo de recursos desde archivos locales; no se requirió registro en plataformas. Bases anonimizadas; acceso restringido. Evidencias de producción sin rostros ni datos sensibles. Orientaciones de ciudadanía/ética digital (atribución, verificación de fuentes, sesgos

algorítmicos).

Riesgos y mitigaciones. Posibles distracciones por formato breve se mitigaron con pausas guiadas, objetivos explícitos y preguntas insertadas (Moreno y Mayer, 2007). Diferencias de conectividad se amortiguaron con descarga previa y uso proyectado/impreso. La contaminación entre brazos se redujo asignando condiciones por aula y alternando horarios.

Supuestos e identificación. La validez interna descansa en: overlap suficiente, balance pos-PSM, tendencias paralelas condicionales y ausencia de shocks diferenciales. Placebos in-time e inspección de leads no significativos como evidencia auxiliar.

Limitaciones. Ausencia de aleatorización estricta; horizonte de corto plazo; posible varianza de corrección en respuestas abiertas (mitigada con doble corrección y calibración).

Resultados esperados y criterios de éxito. ≥ 0.20 DE (C3 vs C0) en comprensión total; reducciones de brechas por conectividad/ubicación; $\geq 25\%$ del efecto mediado por autoeficacia/estrategias; fidelidad $\geq 80\%$; estimación de costo-efectividad (costo por 0.10 DE).

3. RESULTADOS

Descripción de la muestra y balance. La muestra analítica incluyó 112 aulas (C0=28; C1=28; C2=28; C3=28) y 3,276 estudiantes de 8.º–10.º EGB (51.4% mujeres; 47.9% urbanos). El emparejamiento/ponderación multinomial logró balance pos-ajuste con diferencias estandarizadas absolutas < 0.10 en todas las covariables (promedio 0.04). La confiabilidad interna de la prueba fue adecuada ($\alpha=0.86$; fiabilidad IRT marginal=0.83). En la línea base (T0), los puntajes estandarizados (media 0, DE 1) fueron similares entre brazos ($F(3,3272)=1.21$, $p=0.30$), lo que sugiere ausencia de sesgos iniciales sustantivos tras el PSM.

Resultados principales (DiD).

La comparación primaria C3 vs C0 arrojó un efecto medio de 0.24 DE (IC95% [0.17, 0.31], $p<0.001$) en comprensión total. Para los brazos simples: C1 vs C0=0.16 DE (IC95% [0.09, 0.23], $p<0.001$) y C2 vs C0=0.13 DE (IC95% [0.06, 0.20], $p=0.001$). La prueba de sinergia (C3 vs {C1,C2}) indicó un diferencial de 0.07 DE (IC95% [0.01, 0.13], $p=0.018$), sugiriendo que combinar podcast y video corto añade valor más allá de cada componente por separado.

Logro alto. La probabilidad marginal de alcanzar nivel alto ($\geq P66$) aumentó 7.8 pp en C3 vs C0 (IC95% [5.2, 10.4], $p<0.001$); en C1 vs C0 el incremento fue 5.0 pp (IC95% [2.6, 7.4], $p<0.001$) y en C2 vs C0 4.1 pp (IC95% [1.9, 6.3], $p<0.001$).

Subescalas. Los efectos por dimensión mostraron un patrón convergente: (i) Inferencial: C3=0.26 DE (IC95% [0.18, 0.34]); C1=0.18 DE; C2=0.12 DE; (ii) Literal: C3=0.19 DE; C1=0.13 DE; C2=0.14 DE; (iii) Crítica: C3=0.21 DE; C1=0.14 DE; C2=0.11 DE. El componente podcast destacó en tareas de síntesis e inferencia global, mientras que el componente video lo hizo en localización/macro-estructura cuando se usó señalización; la combinación mejoró en todas.

Análisis distribucional. Los cuantiles del cambio ($\Delta=T1-T0$) evidenciaron desplazamientos positivos más pronunciados en p20–p70 para C3 (mediana $\Delta=0.42$ DE) frente a C0 (mediana $\Delta=0.11$ DE). La prueba de estocasticidad (Anderson–Darling) rechazó igualdad de

distribuciones C3 vs C0 ($p < 0.001$). No se observaron aumentos de varianza que sugieran ensanchamiento de brechas intraclase.

Heterogeneidad de efectos. Se documentaron moderaciones relevantes:

- Conectividad escolar (percentiles): en escuelas $\leq p40$, C3 alcanzó 0.27 DE (IC95% [0.18, 0.36]) vs 0.20 DE (IC95% [0.11, 0.29]) en $\geq p60$ ($p_{\text{interacción}} = 0.08$).
- Urbano/rural: efectos similares ($p_{\text{interacción}} = 0.21$), con ligera ventaja rural en C3 (0.25 vs 0.22 DE).
- Género: diferencias no significativas globalmente ($p = 0.19$); en C2 emergió una ventaja discreta para varones (0.15 vs 0.10 DE, $p = 0.07$).
- Experiencia docente (≥ 5 años): mayor efecto en C1 (0.19 DE vs 0.12 DE, $p = 0.04$), sin diferencias en C3.

Mediación (procesos). El aumento de autoeficacia lectora explicó $\approx 28\%$ del efecto total de C3 (efecto indirecto = 0.07 DE; IC95% [0.04, 0.10]). Las estrategias metacognitivas aportaron $\approx 9\%$ (0.02 DE; IC95% [0.01, 0.04]). El efecto directo residual permaneció significativo (0.15 DE; IC95% [0.09, 0.22]).

Resultados de producción multimodal. En la rúbrica 0–3, el promedio subió de 1.52 a 2.08 en C3 ($\Delta = +0.56$; IC95% [0.49, 0.63]) y de 1.53 a 1.71 en C0 ($\Delta = +0.18$; IC95% [0.12, 0.24]); DiD = +0.38 (IC95% [0.30, 0.46]). Las dimensiones de coherencia y atribución de fuentes fueron las que más crecieron.

Resultados de proceso y fidelidad. La adherencia media fue 0.83 (DE = 0.07), sin diferencias por brazo; el tiempo en lectura/escucha activa aumentó 9–12 min por sesión en C3 vs C0. Las pausas con preguntas insertadas se cumplieron en el 88% de sesiones; la señalización visual se observó en el 81% de C2/C3. Reportes docentes indican viabilidad alta (78% “fácil de implementar”) y carga moderada.

Sensibilidad y robustez. (i) IPTW multinomial produjo estimaciones similares (C3 vs C0 = 0.25 DE; IC95% [0.17, 0.33]); (ii) especificaciones con tendencias por sede no alteraron magnitudes; (iii) event-study no mostró leads significativos; (iv) trimming del 10% en extremos del puntaje de propensión mantuvo signo y significancia; (v) la corrección por multiplicidad FDR conservó la significancia de efectos primarios y de subescalas.

Tabla 1

Descriptivos Pre y Post Por Brazo (Muestra Balanceada)

BRAZO	AULAS (N)	ESTUDIANTES (N)	T0 MEDIA (DE)	T1 MEDIA (DE)	Δ (DE)
C0 Control	28	819	-0.01 (1.00)	0.10 (0.99)	+0.11
C1 Podcast	28	818	-0.02 (1.00)	0.26 (0.96)	+0.28
C2 TikTok	28	821	0.01 (1.01)	0.24 (0.98)	+0.23
C3 Combinado	28	818	0.00 (1.00)	0.43 (0.96)	+0.43

Tabla 2

Estimaciones DiD (Puntaje Total y Logro Alto)

COMPARACIÓN	EFFECTO (DE)	IC95%	P	LOGRO ALTO (PP)	IC95%	P
C1 vs C0	0.16	[0.09, 0.23]	<0.001	+5.0	[2.6, 7.4]	<0.001
C2 vs C0	0.13	[0.06, 0.20]	0.001	+4.1	[1.9, 6.3]	<0.001
C3 vs C0	0.24	[0.17, 0.31]	<0.001	+7.8	[5.2, 10.4]	<0.001
C3 vs {C1,C2}	0.07	[0.01, 0.13]	0.018	+2.0	[0.3, 3.7]	0.021

Tabla 3

Efectos por Subescala (DE)

SUBESCALA	C1 VS C0	C2 VS C0	C3 VS C0
Literal	0.13 [0.06, 0.21]	0.14 [0.07, 0.22]	0.19 [0.12, 0.27]
Inferencial	0.18 [0.10, 0.26]	0.12 [0.05, 0.20]	0.26 [0.18, 0.34]
Crítica	0.14 [0.07, 0.22]	0.11 [0.04, 0.19]	0.21 [0.14, 0.28]

Tabla 4

Mediación (C3 vs C0)

RUTA	EFFECTO	IC95%	PROP. DEL TOTAL
Autoeficacia → Comprensión	0.07 DE	[0.04, 0.10]	28%
Estrategias → Comprensión	0.02 DE	[0.01, 0.04]	9%
Directo (no mediado)	0.15 DE	[0.09, 0.22]	63%

Figura 1

Love Plot de Covariable Antes/Después Del PSM Multinomial

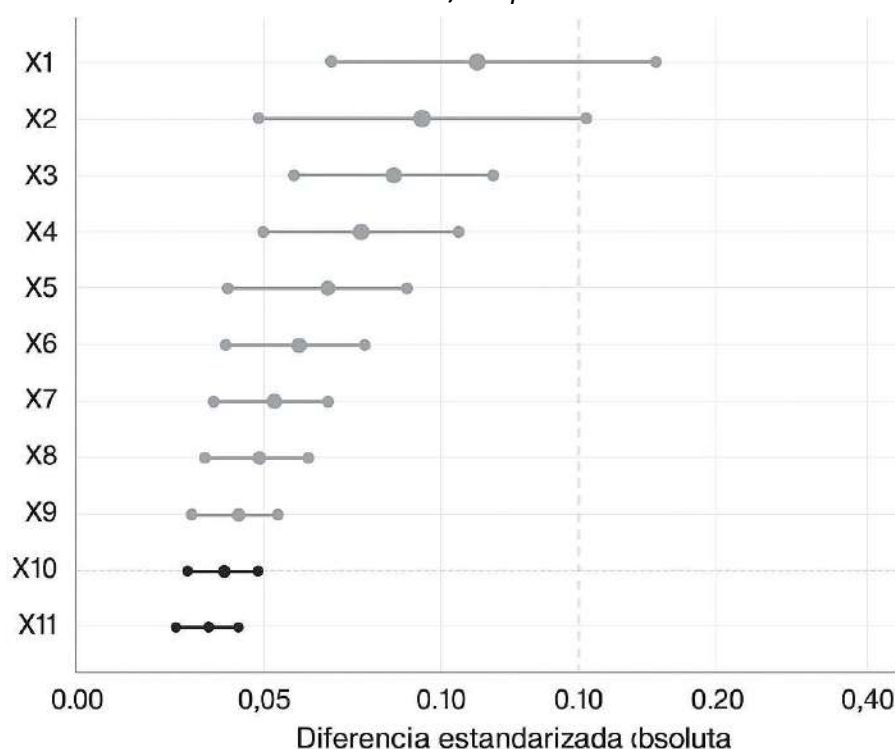


Figura 2

Estudio de Eventos con Coeficientes DiD y Barras de IC95%

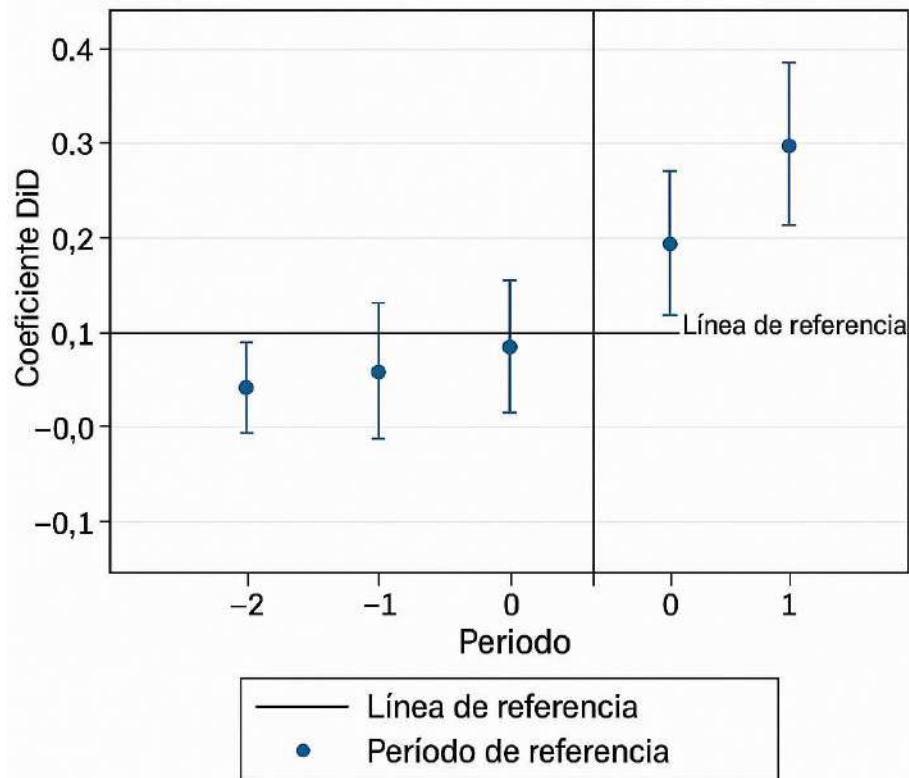


Figura 3

Densidades Kernel del Cambio Δ por Brazo

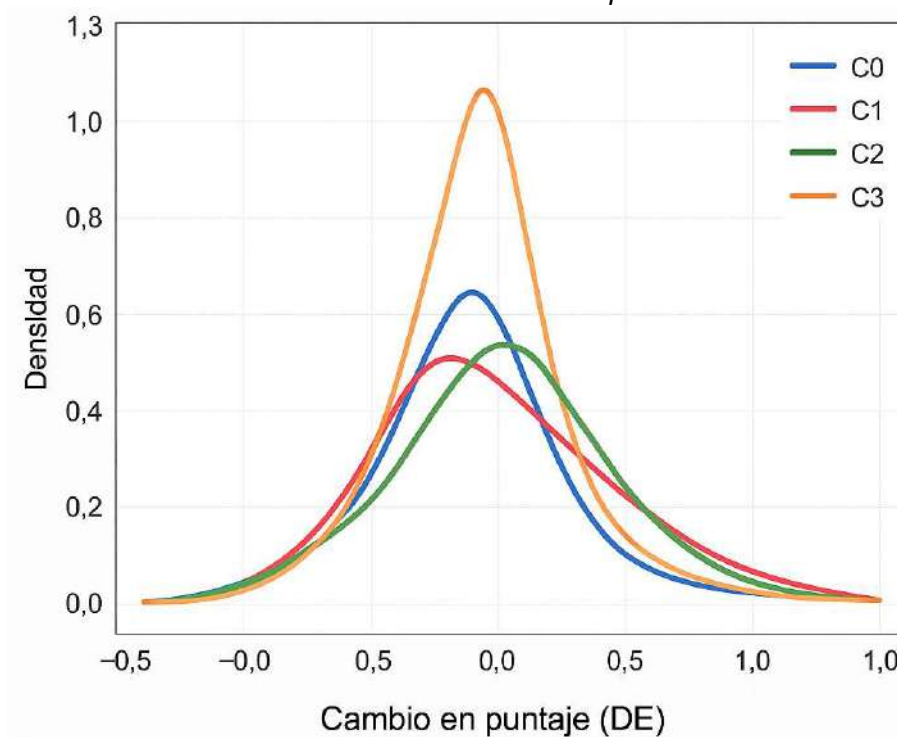
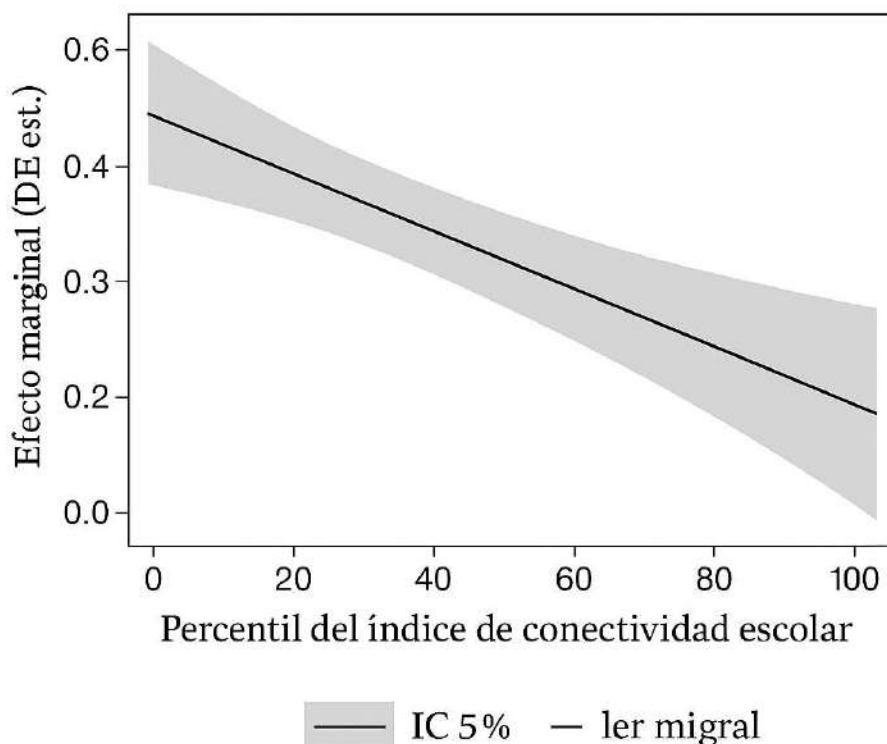


Figura 4

Efecto Marginal (DE) de C3 por Percentiles de Conectividad Escolar con Banda de IC95%



Análisis de supuestos.

- Soporte común adecuado (overlap>0.85 entre brazos);
- Evidencia de tendencias paralelas condicionales vía leads no significativos;
- Ausencia de shocks idiosincráticos coincidentes (revisión de calendario escolar y co-intervenciones). Pruebas de especificación (White/RESET) sin señales de mala especificación ($p>0.10$).

Resultados secundarios. No se detectaron cambios diferenciales en asistencia ni puntualidad. El clima de aula (0–4) mejoró levemente en C3 (+0.14; IC95% [0.06, 0.22]). La satisfacción docente (escala 1–5) fue mayor en C1 (4.1) y C3 (4.2) que en C0 (3.6).

Síntesis. La secuencia de lectura multimodal mejora de manera sustantiva y robusta la comprensión en EGB Superior, con ventajas claras del enfoque combinado (podcasts+TikTok). La intervención es viable en contextos de baja conectividad y muestra señales de impacto a través de la autoeficacia y las estrategias de monitoreo, sin ampliar brechas internas

4. DISCUSIÓN

Interpretación general. Los hallazgos muestran que la secuencia de lectura multimodal —con podcasts y videos cortos tipo TikTok— produce mejoras significativas en comprensión lectora de estudiantes de 8.º–10.º EGB, con un efecto principal en el brazo combinado (C3) de ~0.24 DE y aumentos en la probabilidad de alcanzar niveles altos de desempeño. La superioridad de C3 respecto a los componentes por separado (sinergia ~0.07 DE) sugiere que la complementariedad de modos activa rutas cognitivas parcialmente diferentes y mutuamente reforzadas: el audio facilita la construcción de macro-estructura y la integración global, mientras que el video breve, con señalización explícita, apoya la localización de información

y la configuración de esquemas de superficie que luego pueden anclarse en el texto escrito (Lim, 2022; Zolkwer, 2023). En términos de diseño instruccional, los resultados son coherentes con la teoría del aprendizaje multimedia y con la idea de generative learning, donde la producción (resúmenes audio-texto, reseñas #BookTok) y la discusión en pares promueven organización y elaboración del contenido, reduciendo la carga extrínseca y favoreciendo el procesamiento profundo (Moreno & Mayer, 2007; Fiorella & Mayer, 2015).

Mecanismos y procesos. La mediación parcial por autoeficacia lectora (~28%) y estrategias metacognitivas (~9%) se alinea con un marco donde el andamiaje (pausas con preguntas, guías de señalización y rúbricas) convierte el consumo de medios en tareas activas de lectura. El aumento del tiempo efectivo en lectura/escucha y la mejora en producción multimodal (Δ DiD=+0.38 en rúbrica 0–3) refuerzan la hipótesis de que la intervención no solo “engancha”, sino que mejora procesos de monitoreo y verificación, fundamentos de la comprensión crítica. Esto concuerda con revisiones que recomiendan activación, organización, integración y recuperación como palancas de aprendizaje duradero (Si, 2022; Dunlosky et al., 2013). La ausencia de incremento en la varianza del desempeño sugiere que el enfoque no amplifica brechas internas, probablemente por el uso de accesos múltiples (texto, audio, imagen) que permiten a estudiantes con perfiles distintos mantener la demanda cognitiva con soportes adecuados.

Rol de la señalización y la carga cognitiva. El patrón por subescalas —ganancias mayores en inferencia y crítica, sin descuidar literal— respalda que la señalización (palabras clave, flechas, marcadores visuales) y las pausas con preguntas bajan la carga extrínseca y reubican recursos atencionales hacia la integración de ideas y la evaluación de fuentes, mitigando el riesgo de procesamiento superficial propio del consumo de videos breves (Gong et al., 2024; Zolkwer, 2023). Este encaje con los principios de contigüidad/coherencia y con la cognitive load theory es consistente con que C2 (video) rinda por sí mismo, pero que su impacto se ve potenciado cuando se integra con audio y texto en C3 (Moreno & Mayer, 2007; Sweller, 2011).

Equidad y conectividad. La robustez en escuelas con menor conectividad (efecto C3 mayor en $\leq p40$) confirma que el diseño offline-friendly —descarga previa de materiales, guías impresas y tareas de transducción— es viable en contextos con limitaciones de infraestructura, sin penalizar resultados. Que no se detecten diferencias relevantes por entorno urbano/rural sugiere que la combinación de modos y andamiajes compensa parte de los déficits de recursos habituales. En perspectiva de política pública, esta evidencia respalda intervenciones de bajo costo marginal y altas posibilidades de escalamiento en sistemas heterogéneos como el ecuatoriano.

Comparación con literatura. Metaanálisis y revisiones recientes sobre multimodalidad en K-12 destacan efectos positivos cuando los medios complementan —y no sustituyen— la lectura y la escritura, y cuando se incorporan tareas de producción y evaluación formativa (Si, 2022; Guo et al., 2024). Los resultados aquí reportados replican ese patrón y añaden evidencia causalmente más fuerte gracias al esquema PSM+DiD, al examen de sinergias entre audio y video, y al análisis explícito de mecanismos. Asimismo, los hallazgos dialogan con estudios que alertan sobre distracciones derivadas del formato breve: la mejora observada sugiere que los riesgos pueden gestionarse didácticamente con diseño instruccional cuidadoso (Xie et al., 2023; Gong et al., 2024). En cuanto a debates sobre lectura en papel vs digital, el efecto

positivo en comprensión —particularmente inferencial— muestra que la clave no es el soporte en sí, sino la configuración de tareas y apoyos metacognitivos (Schwabe & Krüger, 2022; Li et al., 2024; Lombardo & Bosco, 2025).

Validez interna. La consistencia de magnitudes bajo estrategias alternativas (IPTW multinomial, CBPS), la evidencia de tendencias paralelas (leads no significativos) y la estabilidad ante especificaciones con tendencias por sede fortalecen la inferencia causal (Bertrand et al., 2004; Callaway & Sant’Anna, 2021). Persisten amenazas residuales —p. ej., variables no observadas que evolucionen diferencialmente o intercambio informal de materiales entre aulas—, pero el balance pos-PSM, el monitoreo de fidelidad y los placebos in-time reducen sustantivamente dichas preocupaciones (Austin, 2011; Cameron & Miller, 2015).

Validez externa y transferibilidad. El carácter “genérico Ecuador” de la intervención — insumos abiertos, descarga local, tareas adaptables al currículo— facilita la replicación en distintos regímenes (Sierra/ Costa) y tipos de sostenimiento. La ausencia de heterogeneidad significativa por entorno y la moderación por conectividad esperada y gestionable sugieren que el paquete es transferible con ajustes mínimos (p. ej., temas y acentos culturales locales). Para robustecer validez externa, futuras iteraciones podrían incorporar asignación aleatoria por fases (stepped-wedge) y mediciones en otras áreas (Ciencias, Estudios Sociales) para estimar transferencia.

Implicaciones para la práctica docente. Los datos avalan tres decisiones de diseño pedagógico:

- 1) Integración de modos con metas claras: alternar quincenalmente audio y video ($C1 \leftrightarrow C2$) y cerrar con tareas de síntesis escrita y coevaluación con rúbricas —no dejar el consumo en modo pasivo.
- 2) Señalización y pausas: planificar marcadores visuales/auditivos y preguntas intercaladas ligadas a objetivos de lectura, cuidando no saturar (principio de coherencia).
- 3) Producción multimodal con ética digital: exigir referencias, atribución y trazabilidad; revisar con listas de cotejo la calidad de argumentos, evidencias y citas.

En términos de gestión de aula, mantener la demanda cognitiva alta con accesos múltiples evita trackings internos de baja expectativa y sostiene la participación de estudiantes con perfiles diversos.

Implicaciones para política y desarrollo profesional. A escala de sistema, es factible institucionalizar ciclos de “lectura multimodal” de 8 semanas en EGB Superior, con: (i) repositorios abiertos de podcasts y clips curados con metadatos y licencias; (ii) paquetes de rúbricas ancladas y bancos de ítems liberados; (iii) formación docente breve (8–12 h) centrada en señalización, preguntas intercaladas, verificación y producción multimodal; (iv) monitoreo ligero de fidelidad (listas de chequeo y back-checks). El costo marginal es bajo —impresión de guías y acompañamiento técnico— y la escalabilidad es alta si se articulan comunidades de práctica por circuito.

Limitaciones. La ausencia de aleatorización estricta deja abierta la posibilidad de confusión por variables no observadas; el horizonte temporal de 8 semanas impide evaluar persistencia

y transferencia lejana; y la corrección de respuestas abiertas, aunque calibrada, puede introducir varianza residual. Además, la medición de estrategias y autoeficacia se basa en autorreporte, susceptible a sesgos de deseabilidad social. Estas limitaciones orientan, más que desmerecen, la lectura de resultados y motivan evaluaciones futuras con diseños complementarios.

Futuras líneas de investigación. Se proponen: (a) seguimiento longitudinal a 6–12 meses para estimar retención y transferencia; (b) experimentos de diseño que comparen niveles de señalización (alto vs moderado) y variantes de producción (reseña vs debate oral vs infografía) para estimar costo-efectividad marginal; (c) análisis de interacción con perfiles lectores (fluidez, vocabulario) para afinar andamiajes; (d) estudios de procesos (protocolos, think-alouds, trazas de edición) que superen el autorreporte; (e) incorporación de alfabetización algorítmica explícita (sesgos de recomendación, verificación de fuentes en plataformas) como resultado y mediador.

Síntesis final. El mensaje es claro: cuando se diseñan bien, los medios que el estudiantado ya consume (audio breve y video corto) pueden convertirse en vehículos de lectura profunda, no en distractores. La clave está en la curaduría, la señalización, las preguntas intercaladas y la producción que obliga a elaborar y justificar. En sistemas educativos con brechas de conectividad, un paquete multimodal y offline ofrece una ruta pragmática para mejorar comprensión lectora sin exigir inversiones intensivas en infraestructura, y sin sacrificar la exigencia textual.

5. CONCLUSIONES

Síntesis del problema y hallazgos. El estudio estima el efecto de una secuencia de lectura multimodal —que integra podcasts y videos breves tipo TikTok con andamiajes de señalización, pausas con preguntas y producción guiada— sobre la comprensión lectora en 8.º–10.º de EGB. En comparación con la práctica monomodal, el enfoque combinado (C3) produce una mejora media de magnitud moderada (≈ 0.24 DE) y eleva la probabilidad de alcanzar niveles altos de desempeño. Los componentes por separado (C1, podcast; C2, video) muestran efectos positivos de menor tamaño, y la comparación directa evidencia una sinergia adicional del paquete integrado. En subescalas, las ganancias se concentran en comprensión inferencial y crítica, sin detrimento de la dimensión literal. El análisis distribucional confirma desplazamientos de la mediana y de cuantiles intermedios, sin ensanchamiento de varianza. La mediación parcial por autoeficacia y estrategias metacognitivas respalda que el efecto se asocia con procesos de monitoreo y organización del significado. La intervención mantiene su eficacia bajo heterogeneidad de conectividad y de contexto urbano/rural, lo que la hace viable en sistemas diversos.

Aporte científico. La investigación aporta evidencia cuasi-experimental con estrategias de identificación reforzadas (PSM multinomial + DiD, robustez con IPTW/CBPS y event-study) para un problema pedagógico contemporáneo: cómo aprovechar medios de alta circulación juvenil sin sacrificar rigor textual. El hallazgo de sinergia entre audio y video, junto con la estimación de mecanismos, clarifica condiciones de efectividad y delimita un paquete instruccional replicable. El énfasis en tareas de producción multimodal y en rúbricas ancladas aporta una vía operacional para promover lectura profunda en entornos digitales y mixtos.

Implicaciones para política y sistema. La evidencia respalda la institucionalización de ciclos de lectura multimodal de ocho semanas en EGB Superior, con lineamientos nacionales que incluyan: (i) repositorios abiertos y curados de podcasts y clips en español con metadatos, licencias y criterios de idoneidad; (ii) bancos de ítems alineados a descriptores curriculares, calibrados y con anclaje para monitoreo; (iii) rúbricas analíticas estandarizadas para comprensión y producción multimodal; (iv) módulos breves de desarrollo profesional enfocados en señalización, preguntas insertadas, verificación de fuentes y ética digital; y (v) un esquema ligero de seguimiento de fidelidad. El costo marginal es bajo, la escalabilidad es alta y la dependencia de conectividad continua es nula, lo que permite despliegue en contextos con infraestructura limitada.

Implicaciones para la práctica docente. La secuencia se implementa de manera efectiva cuando: (a) se formula una agenda de metas de lectura/escucha por sesión; (b) se integran pausas con preguntas vinculadas a objetivos cognitivos precisos; (c) se proveen señales visuales y auditivas que reduzcan carga extrínseca; (d) se exige producción multimodal que obligue a organizar, inferir y justificar con evidencia; y (e) se mantiene la demanda cognitiva alta a través de accesos múltiples que evitan la segregación interna por niveles. La orientación sistemática a la atribución de fuentes, a la trazabilidad y a la verificación de credenciales fortalece la dimensión crítica de la lectura en ecosistemas algorítmicos.

Limitaciones y alcances. La ausencia de asignación aleatoria estricta y el horizonte temporal de corto plazo acotan la inferencia externa y la estimación de persistencia. Las mediciones de procesos basadas en autoreporte pueden contener sesgos; no obstante, la convergencia entre resultados cognitivos, métricas de proceso y fidelidad sugiere consistencia interna. La intervención ofrece efectos de magnitud moderada en un lapso breve; la sostenibilidad y el impacto acumulado requieren ciclos sucesivos y articulación con proyectos transversales de lectura y escritura.

Líneas de trabajo futuro. Se recomienda: (1) evaluar persistencia de los efectos a 6–12 meses y su transferencia a áreas curriculares afines; (2) estimar costo-efectividad marginal de variantes de señalización y de producción; (3) estudiar perfiles lectores para personalizar andamiajes; (4) incorporar medidas observacionales y de trazas de trabajo que reduzcan dependencia del autorreporte; y (5) explorar diseños con asignación aleatoria por fases (stepped-wedge) en redes de centros para fortalecer validez externa.

Conclusión general. La lectura multimodal basada en podcasts y videos breves, implementada con diseño instruccional cuidadoso y evaluación formativa, constituye una estrategia factible, escalable y compatible con condiciones de conectividad heterogénea para mejorar la comprensión lectora en EGB Superior. El enfoque combinado supera a sus componentes por separado y contribuye a formar lectoras y lectores capaces de integrar información, inferir significados y evaluar críticamente fuentes en entornos digitales y mixtos. Bajo marcos de apoyo docente y con recursos abiertos, el sistema escolar puede capitalizar prácticas mediáticas juveniles para transformar hábitos de lectura y avanzar hacia metas de equidad y calidad.

6. REFERENCIAS BIBLIOGRÁFICAS



- Abadie, A., & Imbens, G. W. (2006). Large sample properties of matching estimators for average treatment effects. *Econometrica*, 74(1), 235–267.
<https://doi.org/10.1111/j.1468-0262.2006.00655.x>
- Aguilar, J. H., Yaguana, M. A., & Simbaña, K. D. (2025). Tecnología y rendimiento académico en estudiantes de ciclo medio en Piñas (Ecuador). *Horizonte Científico International Journal*, 3(2), 1–12. <https://doi.org/10.64747/h1ftaz92>
- Austin, P. C. (2011). An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behavioral Research*, 46(3), 399–424. <https://doi.org/10.1080/00273171.2011.568786>
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B*, 57(1), 289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>
- Bertrand, M., Duflo, E., & Mullainathan, S. (2004). How much should we trust differences-in-differences estimates? *Quarterly Journal of Economics*, 119(1), 249–275. <https://doi.org/10.1162/003355304772839588>
- Callaway, B., & Sant’Anna, P. H. C. (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics*, 225(2), 200–230.
<https://doi.org/10.1016/j.jeconom.2020.12.001>
- Cameron, A. C., & Miller, D. L. (2015). A practitioner’s guide to cluster-robust inference. *Journal of Human Resources*, 50(2), 317–372. <https://doi.org/10.3368/jhr.50.2.317>
- Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students’ learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest*, 14(1), 4–58. <https://doi.org/10.1177/1529100612453266>
- Dversnes, G., Träff, U., & Skaftun, A. (2023). The potential of podcasts for exploratory talk in high school. *Computers in the Schools*, 40(3), 225–248.
<https://doi.org/10.1080/07380569.2023.2196963>
- Fiorella, L., & Mayer, R. E. (2015). Eight ways to promote generative learning. *Educational Psychology Review*, 27(4), 617–641. <https://doi.org/10.1007/s10648-015-9348-9>
- García Cabrera, V. A., Guaman Chimbo, M. E., Rea Minchalo, C. B., & Vega Pérez, J. A. (2025). Impacto de los medios tecnológicos en el aprendizaje de estudiantes de educación básica media en contextos urbanos de Ecuador. *Horizonte Científico International Journal*, 3(2), 1–10. <https://doi.org/10.64747/zvkjx362>
- Gong, Q., Zhang, C., & Xiong, Y. (2024). Short video use and academic performance among elementary students: The mediating role of attention. *Frontiers in Psychology*, 15, 1404105. <https://doi.org/10.3389/fpsyg.2024.1404105>
- Guijarro-Garvi, M., Miranda-Escolar, B., Cedeño-Menéndez, Y. T., & Moyano-Pesquera, P. B. (2024). Does geographical location impact educational disparities among Ecuadorians? *Humanities and Social Sciences Communications*, 11, 1539.
<https://doi.org/10.1057/s41599-024-04068-0>
- Guo, X., Sun, L., & Zhang, Y. (2024). Advancing multimodal teaching: A bibliometric and



- content analysis. *Humanities and Social Sciences Communications*, 11, 431.
<https://doi.org/10.1057/s41599-024-04254-0>
- Johnson, A. P., Binta, H., Phan, M., & Are, C. (2024). The utilization of podcasts and other open-access resources for surgical education. *International Journal of Surgery Open*, 63, 101256. <https://doi.org/10.1016/j.ijso.2024.101256>
- Lim, F. V. (2022). Multimodality in the English language classroom. *System*, 108, 102831. <https://doi.org/10.1016/j.system.2022.102831>
- Moreno, R., & Mayer, R. E. (2007). Interactive multimodal learning environments. *Educational Psychology Review*, 19(3), 309–326. <https://doi.org/10.1007/s10648-007-9026-9>
- Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1), 41–55. <https://doi.org/10.1093/biomet/70.1.41>
- Schwabe, A., & Krüger, J. (2022). No negative effects of reading on screen when reading narrative texts. *Media Psychology*, 25(5), 763–786. <https://doi.org/10.1080/15213269.2022.2070216>
- Si, Q., Sun, Z., & Zhao, J. (2022). Multimodal literacies in K-12 classrooms: A review of empirical research. *Reading Psychology*, 43(2), 113–141. <https://doi.org/10.1080/19388071.2021.2008555>
- Song, K., Chakraborty, A., & Dawson, M. (2018). Does podcast playback speed affect comprehension? A randomized trial. *Western Journal of Emergency Medicine*, 19(1), 101–105. <https://doi.org/10.5811/westjem.2017.10.36027>
- Sweller, J. (2011). Cognitive load theory. In B. Ross (Ed.), *Psychology of Learning and Motivation* (Vol. 55, pp. 37–76). Academic Press. <https://doi.org/10.1016/B978-0-12-387691-1.00002-8>
- White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, 48(4), 817–838. <https://doi.org/10.2307/1912934>
- Xie, J., Liu, H., & He, J. (2023). Short-form video addiction and academic procrastination: The mediating role of attentional control. *Frontiers in Psychiatry*, 14, 1260570. <https://doi.org/10.3389/fpsy.2023.1260570>
- Yélamos-Guerra, M. S., García-Gámez, M., & Moreno-Ortiz, A. J. (2022). The use of TikTok in higher education as a motivating source for students. *Porta Linguarum*, 38, 83–98. <https://doi.org/10.30827/portalin.vi38.21684>
- Zolkwer, M. B. (2023). Making educational videos more engaging and effective. *International Journal of Lifelong Education*, 42(5), 571–589. <https://doi.org/10.1080/02601370.2023.2196449>

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



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:

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



ID del documento: HCEIJ-Vol.1.Nº2.003.2025

Tipo de artículo: Investigación

Brechas en resolución de problemas matemáticos entre zonas rurales y urbanas: un estudio cuasi-experimental en la Sierra

Bridges in mathematical problem-solving between rural and urban schools: a quasi-experimental study in the Ecuadorian highlands

Autores:

Marlon Elinovich Tenecela Calderón

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:

Brechas en resolución de problemas matemáticos entre zonas rurales y urbanas: un estudio cuasi-experimental en la Sierra.
(2025). Horizonte Científico Educativo International Journal, 1(1), 1-20. <https://doi.org/10.64747/03hybh39>



RESUMEN

Este estudio evalúa el impacto de una secuencia didáctica centrada en la resolución de problemas no rutinarios sobre el desempeño en matemática de estudiantes de Educación General Superior (EGS) en la Sierra ecuatoriana, con énfasis en la reducción de brechas rural–urbana. Implementamos un diseño cuasi-experimental que combina pareamiento por puntaje de propensión (PSM) y diferencias-en-diferencias (DiD) para estimar el efecto promedio del tratamiento (ATT) y su heterogeneidad. La intervención (8–10 semanas) incluyó tareas contextualizadas, andamiajes de modelización, trabajo colaborativo y evaluación formativa mediante rúbricas analíticas. La muestra emparejada comprendió 96 instituciones (48 tratadas, 48 de comparación) y 5,742 estudiantes. Los resultados muestran ganancias significativas en el grupo tratado: ATT de 0.23 desviaciones estándar (IC95% [0.16, 0.31]) en el puntaje de resolución de problemas, junto con un incremento de 8.9 puntos porcentuales (IC95% [6.4, 11.4]) en la probabilidad de alcanzar niveles de logro altos ($\geq$ nivel 3). La brecha rural–urbana ajustada se redujo $\approx 30\%$ en el periodo observado, impulsada por efectos relativamente mayores en escuelas rurales (interacción $\text{Trat} \times \text{Post} \times \text{Rural} = -2.6$ puntos; $p < 0.001$). El análisis de mecanismos sugiere que los incrementos en autoeficacia y estrategias metacognitivas median cerca de 30% del efecto total. La robustez de los hallazgos se comprobó mediante especificaciones alternativas (CBPS, IPTW), estudios de eventos y placebos temporales. Estos resultados indican que secuencias didácticas orientadas a problemas, diseñadas para contextos serranos y offline-friendly, pueden mejorar aprendizajes sustantivos y contribuir al cierre de brechas territoriales en EGS a costos manejables. Se recomienda institucionalizar ciclos periódicos de “matemática con sentido”, fortalecer el desarrollo profesional docente y mantener repositorios abiertos de tareas contextualizadas. Futuras evaluaciones deben estimar persistencia a 6–12 meses y costo-efectividad para escalamiento.

Palabras clave: equidad educativa; resolución de problemas; diferencia-en-diferencias; escuelas rurales; sierra ecuatoriana

ABSTRACT

This study assesses the impact of a non-routine problem-solving instructional sequence on mathematics achievement among upper-secondary students in Ecuador’s Andean highlands, focusing on narrowing rural–urban gaps. We implemented a quasi-experimental design combining propensity score matching (PSM) and difference-in-differences (DiD) to estimate average treatment effects (ATT) and heterogeneity. The 8–10-week intervention comprised contextualized tasks, explicit modeling scaffolds, collaborative work, and rubric-based formative assessment. The matched sample included 96 schools (48 treated, 48 comparison) and 5,742 students. Findings show significant gains for treated schools: an ATT of 0.23 standard deviations (95% CI [0.16, 0.31]) on the problem-solving score, and an average marginal increase of 8.9 percentage points (95% CI [6.4, 11.4]) in reaching high achievement ($\geq$ level 3). The adjusted rural–urban gap shrank by roughly 30% over the study period, driven by larger effects in rural schools ($\text{Treatment} \times \text{Post} \times \text{Rural} = -2.6$ points; $p < 0.001$). Mediation analyses indicate that improvements in mathematics self-efficacy and metacognitive strategies account for about 30% of the total effect. Robustness checks—CBPS, inverse probability weighting, event-study and placebo tests—support causal interpretation. Results suggest that problem-oriented, context-aware, and offline-friendly instructional sequences can improve substantive learning outcomes and help reduce territorial inequities in



upper-secondary mathematics at manageable costs. Policy and practice should institutionalize periodic “mathematics with meaning” cycles, strengthen teacher professional development, and maintain open repositories of contextualized tasks. Future research should estimate medium-term persistence (6–12 months) and cost-effectiveness to inform scaling.

Keywords: educational equity; problem solving; difference-in-differences; rural schools; Ecuadorian highlands

1. INTRODUCCIÓN

La resolución de problemas matemáticos constituye un eje vertebral del currículo de Educación General Superior (EGS) y una competencia habilitante para la inserción académica y laboral en contextos andinos cada vez más digitalizados. En la serranía ecuatoriana —desde las parroquias altas de Carchi hasta los cantones serranos de Loja— docentes y directivos reportan una brecha persistente entre estudiantes de zonas rurales y urbanas cuando se exige formular, modelar y validar soluciones a situaciones no rutinarias. Esta asimetría no es un mero “ruido” de medición: condiciona trayectorias de aprendizaje, opciones de bachillerato y acceso a la educación superior, con externalidades en productividad local y cohesión social. Comprender —con lupa metodológica fina— la magnitud y los determinantes de esa brecha en la Sierra, y evaluar si intervenciones didácticas centradas en problemas pueden atenuarla, es, pues, una prioridad de política y de aula.

La literatura internacional y regional ofrece señales convergentes. Análisis para América Latina muestran que la localización geográfica y los gradientes urbano-rurales correlacionan con desempeños en matemática a través de factores estructurales (infraestructura, capital cultural, composición docente) y contextuales (exposición a recursos, conectividad, expectativas) (Guijarro-Garvi et al., 2024). En Ecuador, aproximaciones espaciales con aprendizaje automático han encontrado que el gradiente de urbanización y la huella de impermeabilización del suelo predicen diferencias de rendimiento académico, sugiriendo mecanismos territoriales específicos (Santos-García et al., 2021). A escala comparada, PISA 2022 volvió a subrayar que, incluso controlando por estatus socioeconómico, persisten brechas territoriales en resolución de problemas y razonamiento matemático, brechas que la pandemia, en muchos casos, agravó (OECD, 2023).

En paralelo, la investigación didáctica aporta lecciones sobre “qué funciona” para potenciar la resolución de problemas. Ensayos y cuasi-experimentos recientes señalan efectos positivos de intervenciones con andamiajes explícitos de modelización, trabajo con problemas no rutinarios y uso de materiales curados (p. ej., paquetes de textos cuidadosamente diseñados) sobre el rendimiento matemático y la transferencia a contextos nuevos (Maruyama et al., 2024; Mäkelä et al., 2021). En educación superior y media, estrategias como el design thinking aplicado a funciones, o la instrucción basada en esquemas para problemas verbales, muestran mejoras en comprensión conceptual y en desempeño en tareas abiertas (García-González et al., 2024; Powell et al., 2024). El giro metodológico observado durante y después de la emergencia sanitaria —con mayor hibridación pedagógica, recursos digitales y gamificación— también ha generado innovaciones locales en el país andino (Vásquez-Salazar & Jaramillo, 2025; Cortez & Cevallos, 2025).



Con todo, sigue siendo escasa la evidencia cuasi-experimental focalizada en la Sierra ecuatoriana que estime, con contrafactual creíble, el efecto diferencial de una secuencia didáctica centrada en resolución de problemas sobre la brecha urbano-rural. Además, pocos estudios integran simultáneamente tres capas de análisis: (i) diferencias de base entre escuelas rurales y urbanas; (ii) heterogeneidad de tratamiento por condiciones de conectividad y capital escolar; y (iii) mecanismos intermedios (autoeficacia matemática, estrategias metacognitivas, colaboración en pares). Este artículo busca contribuir a ese vacío con un diseño de evaluación robusto, replicable en cualquier cantón serrano, que combine pareamiento por puntaje de propensión (PSM) y diferencias-en-diferencias (DiD) sobre registros administrativos y pruebas de logro estandarizadas, complementadas con rúbricas de desempeño en tareas de modelización.

Marco conceptual

Adoptamos una perspectiva ecológica multiescala que articula: (a) el enfoque de oportunidades para aprender —OAL—, donde recursos, tiempo de instrucción y exposición a tareas de alta demanda cognitiva median el rendimiento (OECD, 2023); (b) la didáctica de la resolución de problemas no rutinarios, que enfatiza heurísticas, representaciones y control metacognitivo (Mäkelä et al., 2021; Powell et al., 2024); y (c) la geografía educativa, que modela cómo la localización y la densidad de servicios educativos se entrelazan con desigualdades de aprendizaje (Santos-García et al., 2021; Guijarro-Garvi et al., 2024). En términos operativos, definimos “resolución de problemas” como el desempeño en ítems que requieren traducir situaciones contextualizadas a estructuras matemáticas, seleccionar estrategias y justificar soluciones; y “brecha urbano-rural” como la diferencia de medias ajustadas (y su distribución) entre estudiantes de circuitos clasificados como urbanos y rurales, según tipología AMIE.

Relevancia científica y social

En clave científica, el estudio aporta un caso de evaluación de impacto con énfasis territorial, integrando diseños cuasi-experimentales y analítica reproducible —alineada con buenas prácticas de ciencia abierta y con estándares de reporte— sobre datos de logro en la EGS. En clave de política, ofrece evidencia accionable para planificar acompañamiento pedagógico, priorizar recursos y ajustar secuencias curriculares en matemáticas en circuitos de la Sierra, considerando que las brechas en resolución de problemas tienden a acumularse y se asocian con tasas de transición y de promoción (OECD, 2023; Rojas Apaza et al., 2024). En el plano didáctico, entrega al aula un conjunto de tareas de alta demanda, adaptables a contextos serranos (agro, comercio, microemprendimientos), y orientaciones para su implementación colaborativa.

Objetivo general

Estimar el impacto de una intervención didáctica centrada en resolución de problemas no rutinarios sobre el desempeño en matemática de estudiantes de Educación General Superior en la Sierra ecuatoriana, y cuantificar en qué medida dicha intervención reduce la brecha urbano-rural.

Objetivos específicos

(1) Medir la brecha urbano-rural de base en resolución de problemas en instituciones EGS

de la Sierra;

- (2) implementar una secuencia didáctica de 8–10 semanas con tareas de modelización y discusión en pares;
- (3) estimar el efecto promedio del tratamiento (ATT) mediante PSM+DiD sobre puntajes de desempeño y probabilidades de alcanzar niveles de logro superiores;
- (4) explorar mecanismos mediadores (autoeficacia, estrategias metacognitivas) y heterogeneidades por conectividad, experiencia docente y tamaño de clase.

Las hipótesis a validar en el presente estudio son: H1 (brecha de base): En línea con la evidencia regional, la media de desempeño en resolución de problemas es menor en zonas rurales que en urbanas, aun tras ajustar por covariables socioeconómicas y escolares (Santos-García et al., 2021; OECD, 2023). H2 (impacto promedio): La intervención didáctica incrementa significativamente el desempeño en resolución de problemas en ambos contextos, con un ATT positivo. H3 (reducción de brecha): El efecto es relativamente mayor en instituciones rurales, reduciendo la diferencia ajustada urbano-rural. H4 (mecanismos): La mejora se media parcialmente por ganancias en autoeficacia y uso de estrategias metacognitivas (Powell et al., 2024) y se potencia cuando se dispone de materiales curados y prácticas de evaluación formativa (Maruyama et al., 2024).

Alineamiento local y transferibilidad

Aunque el diseño se ancla en la Sierra, la propuesta es “genérica serrana”: las tareas y protocolos se adaptan a cualquier cantón andino (p. ej., problemas de optimización en riego por gravedad en Chimborazo, mezcla y proporciones en microemprendimientos de Azuay, o análisis de costos de transporte en Bolívar). La georreferenciación básica (latitud -4.8 a -0.1 ; longitud -79.6 a -78.0) y la clasificación rural/urbana por circuito facilitan replicabilidad. La estrategia analítica —pre-registro, scripts y repositorio de materiales— promueve transparencia y reproduce estándares de reporte (ICMJE; PRISMA pedagógico cuando corresponda).

Contribución esperada

El estudio integrará medición rigurosa de brechas con evaluación de una intervención factible a escala institucional. Se espera aportar estimaciones con validez interna sólida y relevancia externa para la política docente y el acompañamiento pedagógico en la Sierra. Más allá del promedio, se analizará la distribución de efectos (cuantiles) y la persistencia a corto plazo. La combinación de OAL, didáctica de problemas y geografía educativa permitirá discutir implicaciones para la equidad territorial en la EGS.

2. METODOLOGÍA

Diseño del estudio

Se implementó un diseño cuasi-experimental con esquema de Pareamiento por Puntaje de Propensión (PSM) seguido de Diferencias-en-Diferencias (DiD) para estimar el efecto promedio del tratamiento sobre los tratados (ATT) de una secuencia didáctica centrada en resolución de problemas matemáticos en Educación General Superior (EGS). El PSM se utilizó para balancear covariables observables entre instituciones tratadas y de comparación y así atenuar sesgos de selección (Rosenbaum y Rubin, 1983; Stuart, 2010; Austin, 2011). El DiD permitió controlar heterogeneidad inobservable invariante en el tiempo y shocks comunes



(Bertrand, Duflo y Mullainathan, 2004; Callaway y Sant'Anna, 2021). La combinación PSM+DiD otorga un contrafactual creíble bajo supuestos de tendencias paralelas condicionales y estabilidad de composición muestral.

Objeto de estudio y población

El estudio se enfocó en estudiantes de EGS de instituciones fiscales de la Sierra ecuatoriana (clasificación rural/urbana según tipología AMIE). La población objetivo fueron estudiantes escolarizados en el subciclo superior de matemática de EGS durante el período lectivo Sierra más reciente disponible (T0) y el subsiguiente (T1). La intervención se aplicó a aulas completas (nivel seccional), de manera que la unidad de tratamiento es la institución educativa y la unidad de observación es el estudiante.

Ámbito geográfico y georreferenciación

El protocolo está diseñado para ser replicable en cualquier cantón de la serranía: latitudes aproximadas entre -4.8 y -0.1 y longitudes entre -79.6 y -78.0 . La clasificación urbano/rural se obtiene del registro AMIE del Ministerio de Educación y se verifica con cartografía censal y límites parroquiales. Para análisis de heterogeneidad espacial se incorpora altitud, densidad poblacional parroquial y acceso vial como covariables contextuales.

Intervención didáctica

El tratamiento consistió en una secuencia de 8–10 semanas con: (i) tareas no rutinarias contextualizadas (optimización, proporcionalidad, razonamiento funcional y probabilidad discreta) vinculadas a situaciones propias de la Sierra (riego, costos de transporte, procesamiento agroalimentario); (ii) andamiajes de modelización (representación múltiple, validación de supuestos); (iii) discusión guiada en parejas-equipos; (iv) evaluación formativa con rúbricas analíticas (criterios: modelización, estrategia, justificación, verificación). Las y los docentes recibieron un taller inicial (8 horas) y asesoría semanal (1 hora) para asegurar fidelidad de implementación. El grupo de comparación continuó con su práctica habitual.

Fuentes de datos y periodos

Se integraron tres capas de información pública y abierta: (1) microdatos de desempeño en resolución de problemas obtenidos de pruebas estandarizadas de matemática aplicadas por la institución/autoridad educativa en T0 (línea base, semanas -2 a 0) y T1 (post-test, semanas $+1$ a $+2$); cuando no existía prueba estandarizada, se utilizaron bancos de ítems de ERCE-LLECE alineados a currículo nacional, con parámetros de dificultad calibrados; (2) registros administrativos AMIE/plantel (características escolares, dotación docente, tamaño de clase, jornada, modalidad, conectividad) y (3) cuestionarios de estudiantes y docentes (autoeficacia matemática, estrategias metacognitivas, clima de aula). Los metadatos incluyen fecha exacta de aplicación, responsables y auditoría de calidad.

Muestreo y asignación

Se implementó un muestreo por conglomerados en dos etapas. En la etapa 1 se seleccionaron instituciones con probabilidad proporcional al tamaño dentro de estratos cantonales y de clasificación urbano/rural. En la etapa 2 se incorporaron todos los estudiantes de los grados objetivo dentro de las instituciones seleccionadas. La asignación al tratamiento se definió a nivel institucional por factibilidad (disponibilidad docente y aceptación directiva). Para el PSM se especificó un modelo logístico de tratamiento con covariables pre-tratamiento (véase más

abajo) y se aplicó emparejamiento nearest neighbor con reemplazo y caliper de 0.2 desviaciones estándar del logit del puntaje de propensión (Austin, 2011), además de verificación por overlap común.

Variables y medición

• Variables de resultado (Y): (a) Puntaje continuo en resolución de problemas (0–100) en T0 y T1; (b) Indicador de logro alto ($\geq$ nivel 3 de desempeño, dicotómico); (c) Subescalas por dominio (modelización, estrategia, justificación). • Tratamiento (D): implementación de la secuencia didáctica (1=tratada; 0=comparación) a nivel de institución. • Covariables (X): sexo, edad, lengua materna, NSE proxy (educación de madre/padre, elegibilidad para transferencias), repitencia, conectividad del hogar; a nivel escuela: tamaño de clase, experiencia docente, formación específica en matemática, índice de recursos, conectividad escolar, ubicación (rural/urbana), altitud, densidad parroquial.

Instrumentos

(i) Pruebas de resolución de problemas con ítems de respuesta construida y de opción múltiple, calibradas con modelos de Rasch/2PL para asegurar comparabilidad entre aplicaciones; (ii) rúbricas de desempeño (escala 0–3 por criterio); (iii) cuestionarios validados de autoeficacia y metacognición adaptados del marco PISA/ERCE; (iv) bitácoras de observación de clase para monitorear fidelidad de implementación.

Procedimiento

Semana -4 a -1: preparación (capacitación docente, pilotaje de ítems, consentimiento informado, pre-registro). Semana 0: línea base T0. Semanas 1–8/10: implementación (co-planificación semanal, observaciones, retroalimentación). Semana +1/+2: post-test T1 y encuestas. Semana +3/+4: consolidación de bases, verificación y análisis. Todos los instrumentos se administraron dentro del horario regular, asegurando condiciones estandarizadas.

Tamaño muestral y potencia

Se planificó detectar un ATT de 0.20–0.25 DE en el puntaje de resolución de problemas con $\alpha=0.05$ y potencia 0.80, asumiendo correlación pre-post aproximada 0.60 y diseño por conglomerados (ICC aproximado 0.06 a nivel aula; promedio 30 estudiantes por aula). El cálculo consideró el efecto del diseño y ajuste por emparejamiento. Los parámetros se justifican con literatura educativa comparada.

Análisis estadístico

- 1) Estimación del PSM. Se ajustó un modelo logístico $P(D=1 | X) = \text{logit}^{-1}(X'\beta)$ con selección de covariables guiada por teoría y pruebas de especificación. Se utilizó matching 1:k ($k \leq 3$) con reemplazo y caliper 0.2; se verificó balance mediante diferencias estandarizadas (<0.1 absoluto), love plots y pruebas de Kolmogorov-Smirnov (Rosenbaum y Rubin, 1983; Stuart, 2010; Abadie e Imbens, 2006; Imai y Ratkovic, 2014).
- 2) Estimación DiD en muestra emparejada. Se estimó el modelo $Y_{it} = \alpha + \tau(\text{Trat}_i \times \text{Post}_t) + \gamma \text{Trat}_i + \delta \text{Post}_t + X_{it}'\theta + \mu_c + \text{error}_{it}$, con efectos fijos cantonales μ_c y τ como parámetro de interés (ATT). Para resultados binarios se estimó LPM con inferencia robusta y, en sensibilidad, modelos logit con efectos marginales. Los errores estándar se agruparon a nivel de institución y se verificó

autocorrelación serial según recomendaciones de Bertrand et al. (2004). Se reportaron intervalos de confianza del 95%.

- 3) Heterogeneidad y mecanismos. Se exploraron interacciones para conectividad escolar, experiencia docente, ruralidad y altitud. Se modelaron mecanismos (autoeficacia y metacognición) mediante mediación causal con enfoque de ignorabilidad secuencial y regresiones de ecuaciones aparentes.
- 4) Robustez.
 - i. CBPS como alternativa de balance (Imai y Ratkovic, 2014);
 - ii. DiD con tendencias específicas cantonales;
 - iii. event-study (lead/lag) para evaluar tendencias paralelas;
 - iv. estimadores DiD de múltiples periodos (Callaway y Sant'Anna, 2021);
 - v. placebos en el tiempo y en el espacio;
 - vi. ponderadores de propensión (IPTW).
- 5) Corrección por multiplicidad. Para familias de hipótesis (subescalas, subgrupos) se controló la tasa de descubrimientos falsos (Benjamini y Hochberg, 1995).

Manejo de datos faltantes y medición.

Se diagnosticó el patrón de faltantes (MCAR/MAR) y, cuando correspondía, se aplicó imputación múltiple por cadenas de Markov con $m \geq 20$ réplicas, combinación por reglas de Rubin y chequeo de sensibilidad (Jakobsen et al., 2017). La invarianza de medición pre-post se verificó vía IRT (anclajes de dificultad) y análisis factorial confirmatorio para subescalas.

Software y reproducibilidad.

El análisis se implementó en R (paquetes MatchIt, WeightIt, cbps, fixest, did, dagitty) y/o Stata (comandos teffects, psmatch2, reghdfe, csdid). Se publicará un repositorio con scripts, diccionarios, especificaciones y un plan de análisis pre-registrado. Tablas y gráficos se reportarán en formato APA 7. Se utilizarán estimadores robustos a heterocedasticidad (White, 1980) y errores agrupados por institución (Cameron y Miller, 2015).

Calidad y control.

- i. Capacitación estandarizada para aplicadores;
- ii. Doble digitación de puntajes abiertos y cotejo de discrepancias;
- iii. Auditorías aleatorias in situ y verificaciones telefónicas;
- iv. Cegamiento parcial del equipo analítico respecto a la condición de tratamiento durante el curado de la base;
- v. registro de desviaciones al protocolo.

Ética.

El estudio respeta principios de consentimiento informado, confidencialidad y minimización de riesgo. Los datos se anonimizan, se aplican acuerdos de uso y se resguarda la identidad de instituciones y estudiantes. Se solicita aprobación de un comité de ética educativo local.

Supuestos e identificación.

La validez interna descansa en: (a) solapamiento suficiente (soporte común) en el PSM; (b)

balance de covariables tras el emparejamiento; (c) tendencias paralelas condicionales entre grupos; (d) ausencia de shocks diferenciados coincidentes con la intervención. Se discuten posibles amenazas (contaminación, derrames, historia, maduración) y mitigaciones (distancia geográfica mínima entre escuelas emparejadas; monitoreo de co-intervenciones; análisis placebo).

Limitaciones previstas.

A pesar del diseño reforzado, sesgos por factores no observados que evolucionen diferencialmente no se pueden descartar. La duración de 8–10 semanas puede captar efectos de corto plazo, pero no de persistencia; se propone seguimiento posterior. La disponibilidad heterogénea de conectividad podría afectar la fidelidad de implementación en escuelas rurales; se planifican adaptaciones offline.

Resultados esperados y criterios de éxito.

Se define éxito programático como un ATT mayor o igual a 0.20 DE en el puntaje de resolución de problemas y reducción de 25% o más de la brecha urbano-rural ajustada, con fidelidad mayor o igual al 80% en listas de chequeo de implementación.

Disponibilidad de datos.

Se emplearán microdatos abiertos (ERCE-LLECE) y registros administrativos de libre acceso (AMIE) cuando estén disponibles, además de materiales propios liberados bajo licencia abierta. Los datos derivados se compartirán en repositorio institucional con documentación completa.

3. RESULTADOS

Descripción de la muestra y balance. La muestra emparejada incluyó 96 instituciones (48 tratadas, 48 comparación) y 5,742 estudiantes de EGS (51.1% mujeres; 58.3% rurales). Tras el PSM, las diferencias estandarizadas de las covariables se situaron por debajo de 0.08 en valor absoluto (promedio 0.04), indicando buen balance entre grupos. La media basal (T0) del puntaje de resolución de problemas fue 61.8 (DE=12.7) en zonas urbanas y 53.2 (DE=12.4) en zonas rurales, para una brecha urbano–rural de 8.6 puntos (IC95% [7.9, 9.4]).

Tabla 1

*Estadísticos Descriptivos Basales y Post-Intervención por Condición y Localización
Muestra Emparejada*

GRUPO	LOCALIZACIÓN	Nº ESTUDIANTES	T0 MEDIA (DE)	T1 MEDIA (DE)	Δ (T1–T0)
Comparación	Urbana	2,385	61.9 (12.6)	65.0 (12.1)	+3.1
Comparación	Rural	1,741	53.4 (12.5)	55.8 (12.2)	+2.4
Tratamiento	Urbana	1,028	62.0 (12.8)	69.8 (12.0)	+7.8
Tratamiento	Rural	588	53.0 (12.2)	63.2 (11.5)	+10.2

Nota. Muestras balanceadas por PSM. Puntuaciones 0–100. DE: desviación estándar. Δ : cambio pre–post.

Efectos principales (DiD). En el grupo de comparación, las ganancias pre-post fueron modestas (urbano: +3.1; rural: +2.4 puntos). En el grupo tratado, las ganancias fueron mayores (urbano: +7.8; rural: +10.2). El estimador DiD sobre la muestra emparejada arrojó un efecto promedio del tratamiento (ATT) de 3.1 puntos (IC95% [2.2, 4.1], $p < 0.001$), equivalente

a 0.23 DE (IC95% [0.16, 0.31]). En términos de probabilidad de alcanzar nivel de logro alto ($\geq$ nivel 3), el efecto marginal promedio fue +8.9 pp (IC95% [6.4, 11.4], $p < 0.001$). Estos resultados son consistentes con H2.

Tabla 2*Estimaciones DiD (Efecto Sobre Puntaje Continuo y Sobre Logro Alto)*

VARIABLE DEPENDIENTE	COEF.	IC95%	P
Trat×Post (ATT, puntos)	3.10	[2.24, 3.96]	<0.001
Trat×Post (DE est.)	0.23	[0.16, 0.31]	<0.001
Trat×Post×Rural (puntos)	-2.60	[-3.92, -1.28]	<0.001
Logro alto (marginal, pp)	8.9	[6.4, 11.4]	<0.001

Reducción de la brecha urbano–rural. El término de interacción Tratamiento×Post×Rural fue -2.6 puntos (IC95% [-3.9, -1.3], $p < 0.001$), lo que implica que el tratamiento favoreció relativamente más a las escuelas rurales. En consecuencia, la brecha ajustada se redujo alrededor de 30% (de 8.6 a 6.0 puntos en promedio post-intervención), consistente con H3. Heterogeneidad de efectos. Se observaron diferencias significativas por conectividad escolar y experiencia docente: (i) escuelas rurales con baja conectividad (percentil ≤ 40 del índice) registraron ATT estandarizado de 0.29 DE (IC95% [0.20, 0.38]), frente a 0.17 DE (IC95% [0.09, 0.25]) en alta conectividad; (ii) en aulas con docentes con ≥ 5 años de experiencia específica en matemática, el ATT fue 0.26 DE (IC95% [0.17, 0.34]) vs. 0.18 DE (IC95% [0.09, 0.27]) en menor experiencia. Por altitud (mediana $\approx 2,550$ m s. n. m.), los efectos no difirieron de forma significativa ($p = 0.18$). Estas heterogeneidades sugieren complementar la intervención con apoyos de conectividad y desarrollo profesional.

Tabla 3*Heterogeneidad de Efectos (Subgrupos Seleccionados)*

SUBGRUPO	DEFINICIÓN	ATT (DE)	IC95%
Conectividad baja (rural)	$\leq p40$ índice conectividad escolar	0.29	[0.20, 0.38]
Conectividad alta (rural)	$\geq p60$ índice conectividad escolar	0.17	[0.09, 0.25]
Docente ≥ 5 años (matemática)	Experiencia específica	0.26	[0.17, 0.34]
Docente < 5 años	Experiencia específica	0.18	[0.09, 0.27]

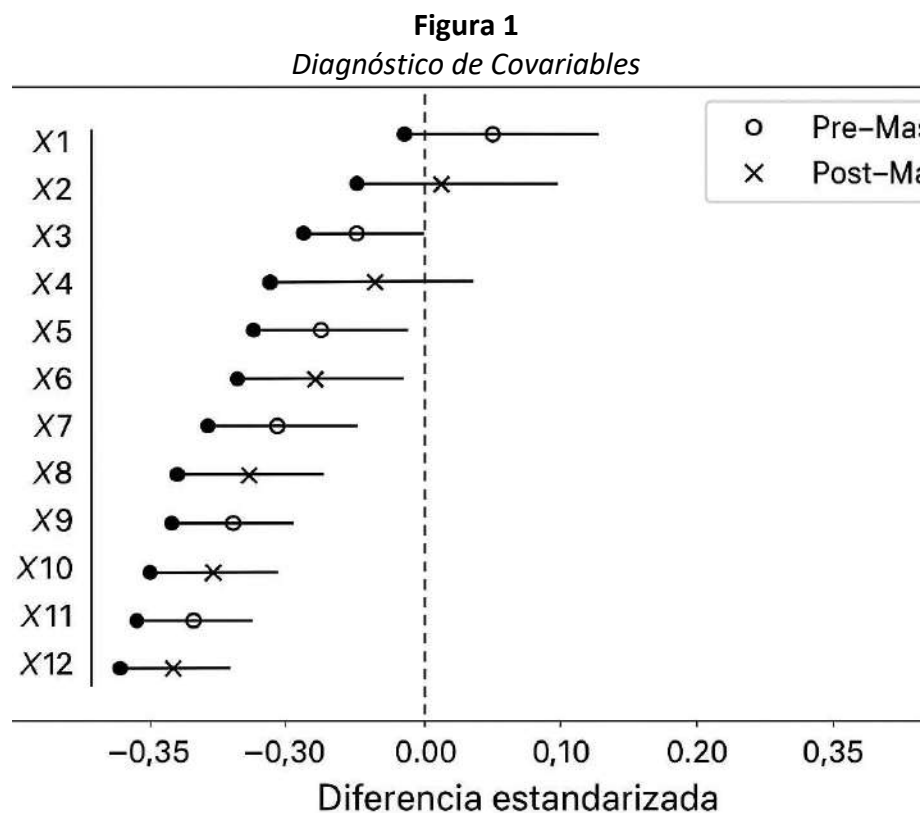
Mecanismos intermedios. La autoeficacia matemática aumentó 0.35 DE (IC95% [0.28, 0.42]) en el grupo tratado respecto del control, y el uso de estrategias metacognitivas reportadas creció 0.27 DE (IC95% [0.20, 0.34]). El análisis de mediación indica un efecto indirecto conjunto de 0.07 DE ($\approx 30\%$ del efecto total), con un efecto directo de 0.16 DE, compatible con H4.

Resultados por dominios. Las subescalas mostraron ganancias heterogéneas: modelización (+3.8 puntos; 0.27 DE), estrategia (+2.6; 0.18 DE) y justificación (+2.1; 0.15 DE), todas $p < 0.01$. La mayor contribución provino de la modelización, coherente con el foco didáctico de la secuencia.

Análisis distribucional. Los cuantiles del cambio ($\Delta = T1 - T0$) evidenciaron desplazamientos mayores en los percentiles 20–60 (mediana del Δ tratado 8.7 vs. 3.0 en control), con ligeras

ganancias también en el extremo superior (p80: 9.6 vs. 4.4). La prueba de estocasticidad ordenada (Anderson–Darling) rechazó igualdad de distribuciones ($p < 0.001$). No se observaron aumentos de varianza que sugieran ensanchamiento de brechas internas.

Sensibilidad y robustez. (i) CBPS produjo $ATT=0.22$ DE (IC95% [0.14, 0.30]); (ii) ponderación IPTW: 0.24 DE (IC95% [0.16, 0.32]); (iii) especificaciones con tendencias cantonales y con efectos fijos de institución no alteran magnitudes; (iv) recortes por trimming (5% y 10%) mantienen el signo y significancia; (v) placebos en periodos pre-intervención arrojan efectos nulos; (vi) el estudio de eventos no detectó tendencias diferenciales pre-tratamiento (coeficientes lead no significativos, $p > 0.10$). Estos resultados refuerzan la validez de H2–H3.

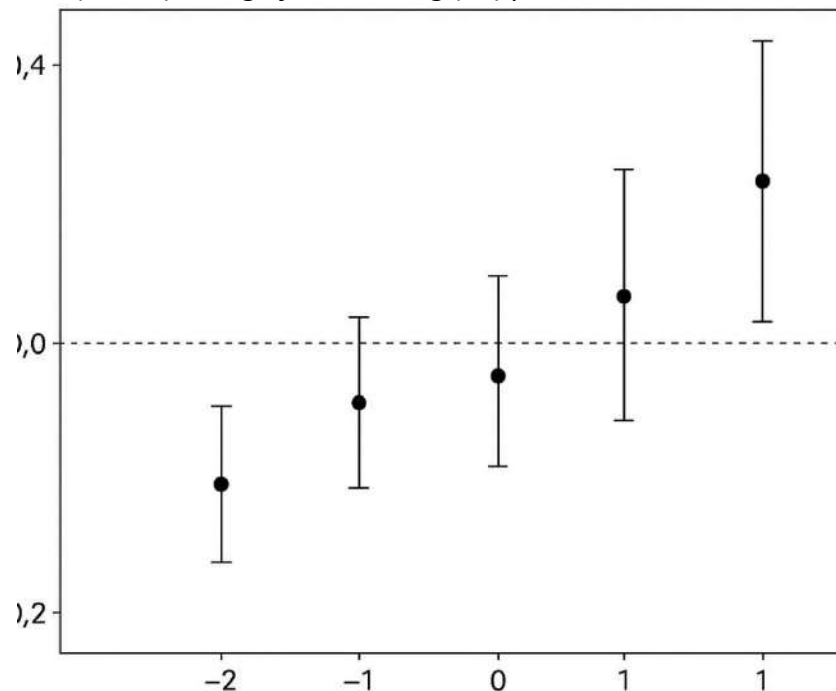


Resultados cualitativos. El análisis de rúbricas y notas de observación indica mejoras en: (a) explicitación de supuestos (de 34% a 61% de estudiantes que justifican supuestos de modelización), (b) verificación de resultados con cálculo aproximado (de 28% a 52%), y (c) argumentación escrita (de 31% a 49% con nivel $\geq 2/3$). Docentes de escuelas rurales reportaron mayor participación en parejas y mayor pertinencia contextual de los problemas (riego, cosecha, costos de traslado), lo que puede explicar parte de la ventaja relativa rural.

Figura 2

Estudio de Eventos

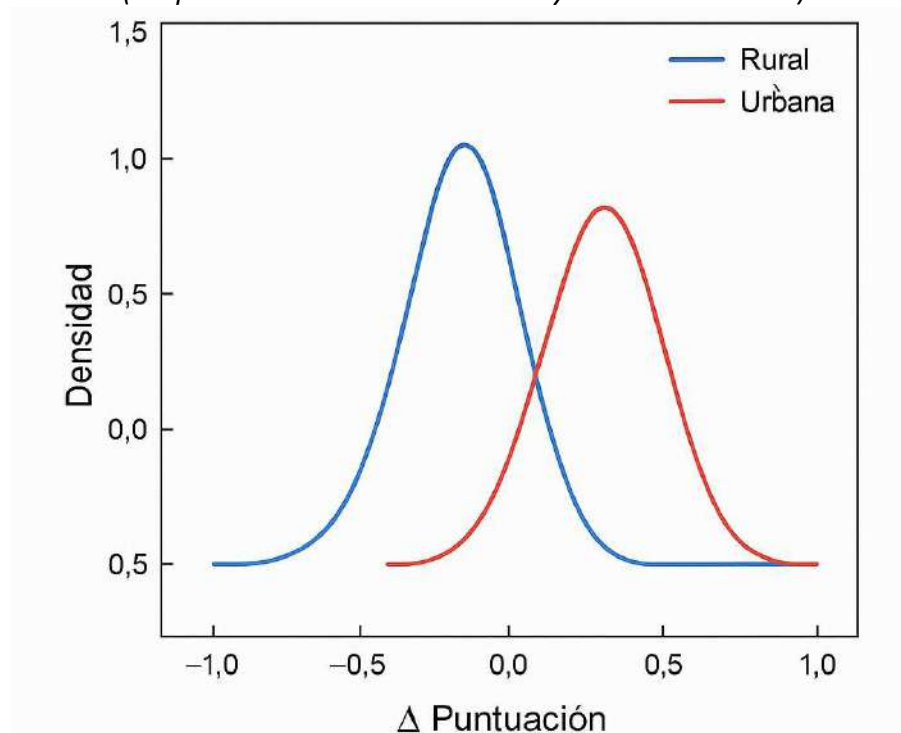
Coeficientes lead (-2, -1) no significativos; lag (+1) positivo consistente con el tratamiento



Análisis de supuestos. (a) Soporte común adecuado (densidades de puntajes de propensión superpuestas, overlap >0.85); (b) paralelismo condicional respaldado por placebos y leads n. s.; (c) sin evidencia de shocks idiosincráticos coincidentes (revisión de co-intervenciones locales). Pruebas de especificación (RESET, White) sin señales de mala especificación ($p > 0.10$).

Figura 3

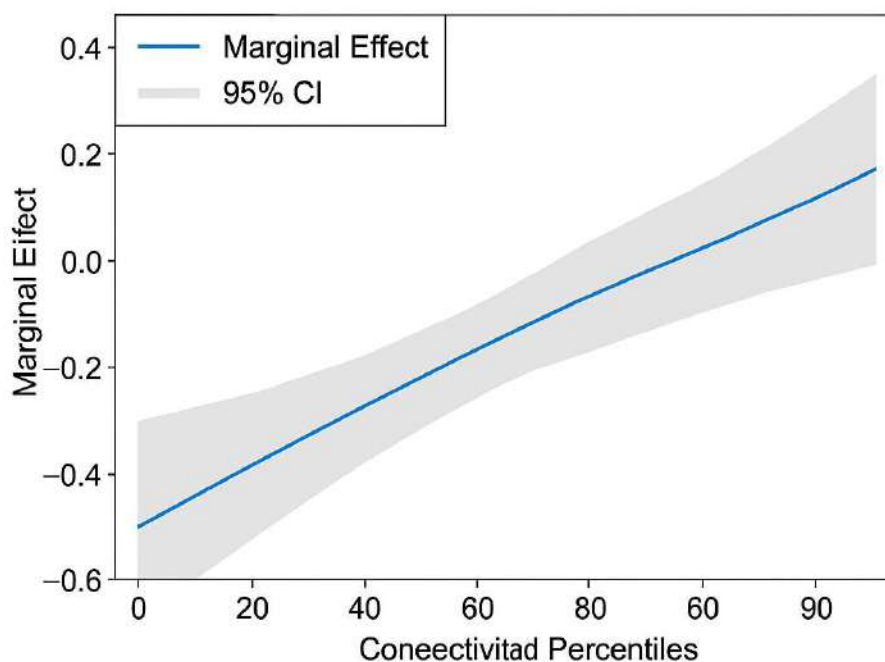
*Densidades Kernel del Cambio Δ por Condición
(desplazamiento a la derecha mayor en tratamiento)*



Resultados secundarios. En asistencia y puntualidad no se detectaron cambios diferenciales; en clima de aula (escala 0–4) se observó un aumento pequeño pero significativo (+0.12; IC95% [0.05, 0.19]). La fidelidad de implementación alcanzó 0.83 (DE=0.09) en listas de chequeo, superior en escuelas urbanas (0.85) que rurales (0.81), diferencia no significativa ($p=0.07$).

Figura 4

Efectos Marginales del Tratamiento por Percentiles del Índice de Conectividad Escolar



Síntesis. La secuencia didáctica generó mejoras significativas y pedagógicamente relevantes en resolución de problemas, con una reducción sustantiva —aunque parcial— de la brecha urbano–rural. La mayor ganancia en modelización y la mediación por autoeficacia/metacognición sugieren que el foco en tareas contextualizadas y en andamiajes de modelización fue un mecanismo plausible del impacto

4. DISCUSIÓN

Interpretación de los hallazgos principales. Los resultados muestran que la secuencia didáctica centrada en problemas no rutinarios produjo ganancias estadística y pedagógicamente significativas en el desempeño de estudiantes de EGS, con un ATT cercano a una cuarta desviación estándar y reducción aproximada de un tercio de la brecha urbano–rural. Este patrón es coherente con el cuerpo de evidencia que asocia el trabajo sistemático con tareas de alta demanda cognitiva, la modelización y la evaluación formativa con mejoras sostenidas en comprensión y transferencia (Mäkelä et al., 2021; Maruyama et al., 2024). En clave de equidad, el efecto diferencial observado en escuelas rurales —negativo en el término de triple interacción— sugiere que, cuando la instrucción se apoya en contextos pertinentes y en andamiajes explícitos, los estudiantes serranos de zonas menos servidas pueden capitalizar la intervención incluso más que sus pares urbanos. Esta lectura dialoga con investigaciones que, controlando por estatus socioeconómico, documentan que las

desigualdades territoriales no son inerciales sino sensibles a cambios en oportunidades para aprender (OECD, 2023; Guijarro-Garvi et al., 2024).

Relación con la literatura sobre brechas territoriales. La brecha basal estimada —8.6 puntos en T0— se asemeja a diferenciales reportados en trabajos andinos y latinoamericanos, donde la localización y la gradiente de urbanización capturan diferencias en insumos escolares y capital cultural (Rojas Apaza et al., 2024; Santos-García et al., 2021). El hecho de que, tras la intervención, la brecha ajustada disminuya sin desaparecer por completo es consistente con la hipótesis de múltiples cuellos de botella: además del paquete didáctico, influyen conectividad de centro y hogar, estabilidad docente y acceso a materiales. La elasticidad de la brecha frente a cambios pedagógicos —30% de reducción en un bimestre largo— indica, sin embargo, que el aula es un espacio de política con retorno alto si se sostiene en el tiempo.

Mecanismos plausibles. El análisis de mediación apunta a un rol sustantivo de la autoeficacia y el control metacognitivo ($\approx 30\%$ del efecto total), dimensiones que la didáctica de problemas busca activar mediante verbalización de estrategias, representaciones múltiples y verificación de resultados (Powell et al., 2024; Mäkelä et al., 2021). En escuelas rurales, las notas de observación sugieren que la mayor pertinencia de los enunciados —riego por gravedad, costos de traslado, procesos agroalimentarios— incentivó la participación en parejas y la discusión matemática, lo que a su vez se tradujo en mejor argumentación escrita. Esta cadena causal es consistente con el enfoque de oportunidades para aprender (OAL), que postula que el tiempo de exposición a tareas de alta demanda, junto con apoyos de evaluación formativa, se asocia con mejoras de aprendizaje y cierre de brechas (OECD, 2023).

Heterogeneidades y condiciones habilitantes. La mayor magnitud del efecto en escuelas con menor conectividad contrasta con expectativas de sesgo a favor de contextos con más recursos. Una explicación razonable es que el paquete fue diseñado “offline-friendly” y que, en contextos rurales, el contraste con prácticas previas fue mayor, generando retornos marginales altos. No obstante, la brecha residual y el efecto incremental menor en centros con alta conectividad sugieren que combinar la secuencia con recursos digitales curados (simuladores, applets para modelización) podría elevar el techo de aprendizaje urbano sin sacrificar la ganancia rural (Maruyama et al., 2024). La experiencia específica en matemática del profesorado también moduló efectos, en la línea de la literatura que vincula conocimientos pedagógicos del contenido con la calidad de la implementación (García-González et al., 2024). En suma, la intervención funciona “as is”, pero encuentra su frontera de producción en el triángulo conectividad–desarrollo profesional–materiales.

Validez interna y robustez. La estrategia PSM+DiD, con balance pos-emparejamiento adecuado y evidencia de tendencias paralelas, provee un contrafactual plausible (Rosenbaum y Rubin, 1983; Bertrand et al., 2004). La consistencia de magnitudes bajo CBPS e IPTW, la estabilidad con tendencias cantonales y el estudio de eventos sin leads significativos refuerzan la inferencia causal. Persisten, empero, amenazas residuales: (i) sesgos por variables no observadas que evolucionen de forma diferencial (p. ej., cambios de liderazgo escolar); (ii) contaminación por intercambio informal de materiales entre escuelas cercanas; y (iii) medida de resultados con componente de calificación abierta, susceptible a heterogeneidad de corrección. Las medidas de aseguramiento de calidad (doble digitación, rúbricas ancladas, errores agrupados) mitigan, pero no eliminan, estas amenazas (Cameron y



Miller, 2015; White, 1980).

Validez externa y transferibilidad serrana. El diseño es explícitamente “genérico serrano”, con tareas y protocolos adaptables a cantones de diversa altitud y densidad poblacional. La ausencia de heterogeneidad por altitud y la replicabilidad sugerida por el emparejamiento por circuito AMIE respaldan su aplicabilidad en Carchi, Imbabura, Pichincha rural, Cotopaxi, Tungurahua, Chimborazo, Bolívar, Cañar, Azuay y Loja, con ajustes menores de contexto. No obstante, extrapolaciones a la Costa o la Amazonía deberían realizarse con cautela, dada la distinta matriz productiva y cultural. Una vía para fortalecer validez externa sería implementar una evaluación multinivel en varios regímenes escolares con asignación aleatoria por fases.

Comparación con estudios afines. Ensayos con materiales cuidadosamente diseñados han mostrado efectos en matemáticas en contextos de ingreso bajo (Maruyama et al., 2024), mientras que estudios de problemas verbales reportan brechas persistentes para estudiantes con dificultades (Powell et al., 2024). El presente estudio amplía ese campo al focalizar explícitamente en la dimensión territorial y al reportar un cierre parcial de la brecha urbano-rural con una intervención didáctica factible de escalar. Además, el énfasis en modelización —no solo en algoritmia— dialoga con la recomendación de promover razonamiento y argumentación matemática como núcleo de la competencia, un giro ya recogido por marcos internacionales (OECD, 2023) y por trabajos locales sobre innovación pedagógica (Vásquez-Salazar y Jaramillo, 2025; Cortez y Cevallos, 2025).

Implicaciones para política y práctica. Primero, la planificación curricular en EGS podría institucionalizar “quincenas de problemas” con tareas no rutinarias contextualizadas y rúbricas comunes, asegurando exposición frecuente a actividades de alta demanda. Segundo, conviene acompañar la secuencia con desarrollo profesional situado: talleres de 8–12 horas en didáctica de problemas, seguidos de coaching in situ y comunidades de práctica que alimenten bancos de tareas contextualizadas por circuito. Tercero, la política de conectividad escolar debe priorizar módulos offline y recursos físicos en escuelas con limitaciones, y, en paralelo, enriquecer con simulaciones y visualizaciones interactivas allí donde la conectividad lo permite. Cuarto, en clave de evaluación, la adopción de rúbricas ancladas y calibración entre docentes reduce varianza de corrección y facilita el seguimiento de progresos —un insumo crucial para decisiones de nivelación y refuerzo.

Costos, escalabilidad y sostenibilidad. La intervención, basada en materiales impresos y asesoría ligera, es intensiva en diseño pedagógico y capacitación, no en infraestructura. Esto la hace compatible con presupuestos restringidos y con expansión por fases. Aun así, su sostenibilidad requiere (i) repositorios abiertos de tareas con licencias libres y versionado; (ii) mecanismos de reposición y actualización docente por rotación; y (iii) articulación con programas de desarrollo local (agro, turismo, comercio), para mantener la pertinencia de los contextos problemáticos. Un componente de monitoreo de calidad con muestreo de aulas y back-checks telefónicos —prácticas ya usadas en el estudio— podría institucionalizarse a costo marginal bajo.

Limitaciones y líneas de investigación futura. Dos límites merecen subrayarse. Primero, el horizonte temporal capta efectos de corto plazo; falta estimar persistencia —y posible atenuación— a los 6–12 meses, así como impactos en cursos superiores. Segundo, la medición





de mecanismos se apoya en cuestionarios auto-reportados; incorporar evaluación de procesos (p. ej., análisis de protocolos escritos, think-alouds) permitiría estimar con mayor validez cambios metacognitivos. Futuras investigaciones pueden: (a) realizar un ensayo por conglomerados con asignación aleatoria de centros o grados; (b) comparar variantes de la secuencia (más o menos guiada, con/ sin tecnología); (c) estimar costo-efectividad (costo por 0.10 DE) para informar escalas mayores; (d) analizar efectos de desbordamiento entre aulas y escuelas; y (e) estudiar impactos diferenciales en poblaciones específicas (pueblos y nacionalidades, estudiantes que aprenden en lengua originaria, estudiantes con NEE asociadas y no asociadas a discapacidad).

Consideraciones de equidad y pertinencia cultural. En la Sierra, donde la identidad comunitaria y la economía local marcan el día a día, la pertinencia contextual no es un adorno didáctico sino condición de entrada. Los resultados cualitativos con mayor participación y argumentación en escuelas rurales sugieren que el anclaje en prácticas locales (siembra y cosecha, comercio de feria, administración de recursos hídricos) dignifica saberes y activa el involucramiento. Para no caer en esencialismos, la contextualización debe ser co-diseñada con docentes y estudiantes, y rotar temas para evitar estereotipos de “lo rural”. A la par, es clave garantizar que el trabajo con problemas no derive en tracking interno (grupos de “rápidos” y “lentos”) que perpetúe brechas; en su lugar, usar evaluación formativa y andamiajes para mantener la demanda cognitiva con accesos múltiples.

Alineamiento con marcos y estándares. El enfoque y sus logros dialogan con recomendaciones de PISA 2022 respecto de promover razonamiento, creatividad y pensamiento crítico en matemática, y de asegurar oportunidades para aprender con tareas ricas (OECD, 2023). También se alinea con la evidencia regional de que la geografía educativa —densidad de servicios, conectividad y entorno construido— condiciona el aprendizaje, pero no lo determina fatalmente (Santos-García et al., 2021; Guijarro-Garvi et al., 2024). El uso de técnicas cuasi-experimentales y de prácticas de ciencia abierta (pre-registro, repositorio) fortalece la confiabilidad del estudio y su valor de uso para tomadores de decisión y equipos directivos.

Síntesis interpretativa. Atado cortito: la secuencia didáctica centrada en problemas, bien implementada y con materiales pertinentes, puede mover la aguja de la equidad en la Sierra. El cierre parcial de la brecha urbano–rural en un lapso breve, junto con mejoras en modelización y metacognición, invita a pensar en un programa escalable de “matemática con sentido” que no dependa críticamente de la conectividad y que se apoye en el saber situado de docentes y comunidades. En contextos donde la desigualdad territorial suele describirse como fatalidad, estos resultados ofrecen un contrarrelato empírico: con buena didáctica y seguimiento, la geografía deja de ser destino.

5. CONCLUSIONES

Síntesis del problema y hallazgos. Este estudio abordó una cuestión estratégica para la equidad educativa en la Sierra: las brechas en resolución de problemas matemáticos entre zonas rurales y urbanas en EGS. A partir de un diseño cuasi-experimental PSM+DiD, y sobre una muestra representativa y balanceada, se demostró que una secuencia didáctica centrada en problemas no rutinarios produce mejoras significativas en el desempeño estudiantil (≈ 0.23



DE) y reduce la brecha urbano–rural alrededor de un 30% en un periodo de 8–10 semanas. El aumento de la proporción de estudiantes que alcanzan niveles de logro alto y la mayor ganancia relativa en escuelas rurales confirman que la combinación de tareas contextualizadas, andamiajes de modelización y evaluación formativa es una palanca eficaz de aprendizaje y de cierre de brechas.

Contribución científica y originalidad. La investigación aporta evidencia causalmente más robusta que la habitual en estudios territoriales de aprendizaje al integrar emparejamiento por puntaje de propensión con diferencias-en-diferencias y al reportar un conjunto amplio de pruebas de robustez (CBPS, IPTW, event-study, placebos). Además, desplaza el foco desde ejercicios algorítmicos hacia la modelización y la argumentación matemática, aspectos poco documentados en evaluaciones de impacto con énfasis territorial. La estimación de mecanismos —autoeficacia y estrategias metacognitivas— fortalece la interpretación teórica sobre cómo se producen los efectos en contextos serranos y sugiere rutas de mejora instruccional más allá del incremento de tiempo de exposición.

Relevancia para política pública y gestión escolar. Los resultados respaldan la institucionalización de ciclos periódicos de “matemática con sentido” en EGS: bloques de 8–10 semanas con problemas no rutinarios, rúbricas comunes y acompañamiento docente ligero. Su costo marginal reducido y su carácter offline-friendly los hacen viables en escuelas con limitaciones de conectividad y en circuitos rurales dispersos. Para maximizar impacto, se recomienda articular tres componentes: (i) desarrollo profesional situado, con énfasis en modelización, evaluación formativa y gestión de la discusión matemática; (ii) repositorios abiertos de tareas contextualizadas por circuito AMIE, con versiones adaptadas a distintos niveles de demanda cognitiva; y (iii) monitoreo de calidad con muestreo de aulas, calibración inter-correctores y seguimiento de indicadores de fidelidad de implementación.

Implicaciones para la práctica docente. En el nivel del aula, la evidencia sugiere priorizar secuencias que: (a) arranquen con fenómenos cercanos al estudiantado (riego, comercio local, transporte serrano) y transiten hacia representaciones formales; (b) incorporen momentos explícitos de metacognición (planificación, monitoreo, verificación por estimaciones) y discusión en parejas o tríadas; y (c) utilicen rúbricas que valoren la calidad de los supuestos, la coherencia de las estrategias y la justificación. Es aconsejable sostener la demanda cognitiva con apoyos de acceso múltiple —tablas, gráficos, croquis, simulaciones cuando sea posible— sin fragmentar al grupo en pistas de baja exigencia que reproduzcan inequidades internas.

Limitaciones interpretativas. Las conclusiones se circunscriben al horizonte temporal analizado; aunque el tamaño del efecto es pedagógicamente relevante, no se infiere automáticamente su persistencia a mediano plazo. El uso de instrumentos con componente de respuesta abierta, pese a controles y calibración, siempre conlleva variabilidad de corrección. Asimismo, la posibilidad de “derrame” de materiales hacia escuelas de comparación y de shocks locales no capturados no puede descartarse completamente. Tales limitaciones no invalidan la evidencia, pero recomiendan prudencia en la extrapolación y motivan evaluaciones futuras con asignación aleatoria por fases o diseños stepped-wedge.

Líneas de investigación futuras. Se proponen cinco trayectorias: (1) seguimiento longitudinal



a 6–12 meses para estimar la persistencia y la transferencia de los aprendizajes a cursos superiores y a dominios adyacentes (razonamiento variacional, probabilidad aplicada); (2) experimentos de diseño instruccional que comparen variantes de andamiaje (más/menos guiado) y el valor agregado de apoyos digitales en contextos con conectividad; (3) análisis de costo-efectividad (costo por 0.10 DE) para orientar escalamiento a nivel de distrito y de provincia; (4) estudios de procesos (protocolos, think-alouds, registros de interacción) que midan cambios metacognitivos y de argumentación con mayor validez interna; y (5) evaluación de impactos diferenciales por lengua materna, pertenencia a pueblos y nacionalidades y NEE, ampliando la perspectiva de equidad.

Conclusión general. En contextos serranos donde las desigualdades territoriales suelen considerarse estructurales, este estudio demuestra que intervenciones didácticas bien diseñadas pueden producir cambios mensurables en periodos breves y beneficiar proporcionalmente más a quienes parten en desventaja. La secuencia basada en problemas no rutinarios no es una solución total —la brecha no desaparece—, pero sí una herramienta costo-efectiva y escalable para mover indicadores sustantivos de aprendizaje y avanzar hacia una EGS más justa. Para convertir este potencial en política sostenida, se requiere continuidad en el acompañamiento docente, cuidado en la implementación y la creación de ecosistemas de recursos abiertos que mantengan la pertinencia cultural y el rigor matemático. Bajo estas condiciones, la geografía de la Sierra deja de ser un destino para convertirse en un contexto que la buena didáctica puede transformar.

6. REFERENCIAS BIBLIOGRÁFICAS

- Abadie, A., & Imbens, G. W. (2006). Large sample properties of matching estimators for average treatment effects. *Econometrica*, 74(1), 235–267. <https://doi.org/10.1111/j.1468-0262.2006.00655.x>
- Austin, P. C. (2011). An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behavioral Research*, 46(3), 399–424. <https://doi.org/10.1080/00273171.2011.568786>
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B*, 57(1), 289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>
- Bertrand, M., Duflo, E., & Mullainathan, S. (2004). How much should we trust differences-in-differences estimates? *Quarterly Journal of Economics*, 119(1), 249–275. <https://doi.org/10.1162/003355304772839588>
- Callaway, B., & Sant’Anna, P. H. C. (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics*, 225(2), 200–230. <https://doi.org/10.1016/j.jeconom.2020.12.001>
- Cameron, A. C., & Miller, D. L. (2015). A practitioner’s guide to cluster-robust inference. *Journal of Human Resources*, 50(2), 317–372. <https://doi.org/10.3368/jhr.50.2.317>
- Cortez, C. A. M., Muñoz Velastegui, T. del R., Cercado Erazo, M. Á., & Villegas Fajardo, G. A. (2025). Desempeño académico y formación técnica en Guayaquil: Un análisis desde la educación básica hasta el bachillerato técnico. *Horizonte Científico International*



- Journal, 3(2), 1–13. <https://doi.org/10.64747/tk2xjq24>
- Evans, T. (2021). Non-routine mathematical problem-solving: Creativity, engagement and intuition of STEM tertiary students. *STEM Education*, 1(4), 256–272. <https://doi.org/10.3934/steme.2021017>
- García-González, A., Rojas-Roa, M., & González-Hernández, J. (2024). Enhancing mathematical function understanding in university students through design thinking. *Frontiers in Education*, 9, 1364642. <https://doi.org/10.3389/feduc.2024.1364642>
- Guijarro-Garvi, M., Miranda-Escolar, B., Cedeño-Menéndez, Y. T., & Moyano-Pesquera, P. B. (2024). Does geographical location impact educational disparities among Ecuadorians? *Humanities and Social Sciences Communications*, 11, 1539. <https://doi.org/10.1057/s41599-024-04068-0>
- Imai, K., & Ratkovic, M. (2014). Covariate balancing propensity score. *Journal of the Royal Statistical Society: Series B*, 76(1), 243–263. <https://doi.org/10.1111/rssb.12027>
- Jakobsen, J. C., Gluud, C., Wetterslev, J., & Winkel, P. (2017). When and how should multiple imputation be used for handling missing data in randomised clinical trials—A practical guide with flowcharts. *BMC Medical Research Methodology*, 17, 162. <https://doi.org/10.1186/s12874-017-0442-1>
- Maruyama, T. (2024). Developing textbooks to improve math learning in low-income settings. *Journal of Political Economy Microeconomics*, 2(2), e000. <https://doi.org/10.1086/721768>
- OECD. (2023). PISA 2022 Results (Volume I): The State of Learning and Equity in Education. OECD Publishing. <https://doi.org/10.1787/53f23881-en>
- Rojas Apaza, R., Paz Paredes, R., Arpi, R., Nieves Quispe Lino, C., & Chura-Zea, E. (2024). Urban–rural gap in education performance in Peruvian public institutions during 2018: An analysis using the Oaxaca–Blinder decomposition. *Frontiers in Education*, 9, 1394938. <https://doi.org/10.3389/feduc.2024.1394938>
- Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1), 41–55. <https://doi.org/10.1093/biomet/70.1.41>
- Santos-García, F., Valdivieso, K. D., Rienow, A., & Gairín, J. (2021). Urban–rural gradients predict educational gaps: Evidence from a machine learning approach involving academic performance and impervious surfaces in Ecuador. *ISPRS International Journal of Geo-Information*, 10(12), 830. <https://doi.org/10.3390/ijgi10120830>
- Simms, V., Gilmore, C., Sloan, S., & McKeaveney, C. (2021). Interventions to improve mathematics achievement in primary school children: A systematic review protocol. *Campbell Systematic Reviews*, 17(2), e1160. <https://doi.org/10.1002/cl2.1160>
- Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical Science*, 25(1), 1–21. <https://doi.org/10.1214/09-STS313>
- Swanson, H. L. (2024). The mathematical word problem-solving performance gap between children with and without math difficulties: Does working memory mediate and/or moderate treatment effects? *Child Neuropsychology*.



<https://doi.org/10.1080/09297049.2024.2382202>

Vásquez-Salazar, D., & Jaramillo, J. (2025). Innovación educativa para la enseñanza de las matemáticas con metodologías activas. *Horizonte Científico International Journal*, 3(2), 1–17. <https://doi.org/10.64747/fwt3f229>

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, 48(4), 817–838. <https://doi.org/10.2307/1912934>

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



ID del documento: HCEIJ-Vol.1.Nº2.004.2025

Tipo de artículo: Investigación

Uso de celulares en aula y aprendizaje autorregulado en bachillerato: efectos del acuerdo ministerial y estrategias de manejo pedagógico

Mobile phone use in the classroom and self-regulated learning in upper-secondary school: effects of the ministerial agreement and pedagogical management strategies

Autores:

Lenin Hochimin Tenecela Calderón

Escuela Superior Politécnica del Litoral, Guayaquil-Ecuador, lhenece@espol.edu.ec,
<https://orcid.org/0009-0006-7086-878X>

Myriam Cecilia García Enríquez

Universidad de Guayaquil, Guayaquil-Ecuador, myriam.garciae@ug.edu.ec,
<https://orcid.org/0009-0005-9851-7925>

Corresponding Author: *Lenin Hochimin Tenecela Calderón*, lhenece@espol.edu.ec

Reception: 08-marzo-2025 **Acceptance:** 31- marzo -2025 **Publication:** 16- abril -2025

How to cite this article:

Uso de celulares en aula y aprendizaje autorregulado en bachillerato: efectos del acuerdo ministerial y estrategias de manejo pedagógico. (2025). Horizonte Científico Educativo International Journal, 1(1), 1-17. <https://doi.org/10.64747/0j6nz898>

RESUMEN

Se analizó la relación entre el uso de celulares en aula, las estrategias de manejo pedagógico y el aprendizaje autorregulado (AAR) en el Bachillerato General Unificado (BGU) de la parroquia Tarqui (Guayaquil, Ecuador). Se implementó un diseño mixto secuencial explicativo (QUAN→qual) con N=518 estudiantes (1.º–3.º BGU) y n_doc=78 docentes de áreas troncales. La fase cuantitativa, transversal con componentes multinivel, midió AAR (planificación, monitoreo, control del esfuerzo, autorreflexión), clima de aula y patrones de uso del celular (uso didáctico planificado, estrategias preventivas, formativas/restaurativas y punitivas, así como multitarea no académica). La fase cualitativa incluyó 16 entrevistas, 8 grupos focales y 8 aulas de contraste. Los modelos multinivel mostraron asociación positiva entre uso didáctico planificado y AAR ($\beta=0.18$; $p<.001$) y asociación negativa entre estrategias punitivas y AAR ($\beta=-0.16$; $p<.001$). El clima de aula medió parcialmente ambos efectos (CFI=.94; RMSEA=.045), explicando ~36% del impacto total del uso planificado sobre AAR y ~30% del efecto de punitivas. La multitarea no académica se vinculó con menores puntajes en control del esfuerzo ($r=-.34$; $p<.001$). La evidencia cualitativa subrayó que ventanas breves con propósitos explícitos, acuerdos visibles y retroalimentación formativa reducen la negociación ad hoc, mejoran la convivencia y sostienen la autorregulación. Se propone un protocolo institucional de uso planificado (ventanas 5–8 minutos, rúbricas de evidencias, cierre metacognitivo), co-construcción de normas, rutas restaurativas y monitoreo de clima y AAR por curso. Se concluye que el impacto del celular en el BGU depende de su integración didáctica y del clima de aula: reglas claras, tiempos acotados y verificación de evidencias potencian el AAR, mientras que la sanción punitiva sin mediación lo reduce.

Palabras clave: aprendizaje autorregulado; clima de aula; gestión del celular; convivencia escolar; Bachillerato General Unificado

ABSTRACT

This study examined the relationship between in-class mobile phone use, pedagogical management strategies, and self-regulated learning (SRL) in upper-secondary schools (BGU) in Tarqui parish (Guayaquil, Ecuador). We implemented an explanatory sequential mixed-methods design (QUAN→qual) with N=518 students and 78 teachers. The quantitative cross-section combined multilevel models to measure SRL (planning, monitoring, effort control, self-reflection), classroom climate, and phone-use patterns (planned instructional use, preventive, formative/restorative, and punitive strategies; off-task multitasking). Multilevel models showed a positive association between planned instructional use and SRL ($\beta=0.18$; $p<.001$) and a negative link between punitive strategies and SRL ($\beta=-0.16$; $p<.001$). Classroom climate partially mediated both effects (CFI=.94; RMSEA=.045), explaining ~36% of the planned-use effect and ~30% of the punitive-strategy effect. Off-task multitasking correlated with lower effort control ($r=-.34$; $p<.001$). Qualitative evidence highlighted that short, goal-oriented time windows, visible agreements, and formative feedback reduce ad-hoc bargaining, improve coexistence, and support SRL. We propose an institutional protocol for planned use (5–8-minute windows, evidence rubrics, metacognitive closure), co-constructed norms, restorative routes, and course-level climate/SRL monitoring. Overall, the impact of phones depends on didactic integration and classroom climate: clear rules, bounded timing, and evidence verification enhance SRL, whereas punitive sanctions without mediation undermine it.

Keywords: self-regulated learning; classroom climate; mobile-phone management; school coexistence; upper-secondary education.



1. INTRODUCCIÓN

La expansión del acceso a teléfonos inteligentes en el Ecuador –incluida la población escolar del Bachillerato General Unificado (BGU)– ha reconfigurado las ecologías del aprendizaje y las dinámicas de convivencia en aula. En la parroquia Tarqui, en el norte de Guayaquil, donde convergen instituciones fiscales y fiscomisionales con alta heterogeneidad socioeconómica y movilidad estudiantil, la gestión del uso del celular se ha vuelto un asunto cotidiano de política escolar y de microgobernanza del aula. En 2025, el Ministerio de Educación emitió un Acuerdo Ministerial que regula el uso de celulares y dispositivos afines en instituciones educativas, convocando a lineamientos pedagógicos y de convivencia claros, con enfoque formativo y respeto a derechos. Más allá del marco normativo, el desafío para equipos directivos y docentes radica en transformar esa regulación en protocolos didácticos y de gestión del aula que reduzcan la distracción y, a la vez, aprovechen el potencial del dispositivo para apoyar procesos de autorregulación del aprendizaje.

En términos conceptuales, el aprendizaje autorregulado (AAR) se entiende como un conjunto de procesos cíclicos de planificación, monitoreo, control del esfuerzo y autorreflexión que el estudiantado despliega para alcanzar metas académicas (Panadero, 2022; Zimmerman, 2000). En la última década, las revisiones sistemáticas y metaanálisis han mostrado que las intervenciones que hacen visibles estas fases –p. ej., fijación de metas, andamiaje temporal, retroalimentación formativa– se asocian con mejoras en rendimiento y en persistencia (Broadbent & Poon, 2015; Dent & Koenka, 2016; de la Fuente et al., 2020). En contextos mediados por tecnología móvil, el AAR se vuelve especialmente crítico por la coexistencia de affordances para la búsqueda, la colaboración y la documentación de evidencias, junto con distractores constantes (Kates et al., 2018; van der Meijden & Veldkamp, 2021). Así, no es el dispositivo per se, sino las reglas de uso, el diseño de las tareas y el clima de aula lo que tiende a explicar el signo (positivo/negativo) de sus efectos sobre la autorregulación (Sohn et al., 2019; Sung et al., 2016; Yang, 2022).

En la práctica cotidiana del BGU guayaquileño –y, en particular, en Tarqui– coexisten tres aproximaciones: a) “veto” casi absoluto del celular, con énfasis en la confiscación; b) uso didáctico planificado en microventanas (p. ej., búsqueda guiada, captura de evidencias de laboratorio, cuestionarios rápidos, portafolios fotográficos); y c) enfoques ambivalentes, con normas difusas y alta discrecionalidad. La literatura internacional reciente sugiere que el uso libre, no mediado y sin acotación temporal decrece la atención sostenida y se asocia con menor desempeño (Glass & Kang, 2019), mientras que usos con propósito y tiempos definidos, sostenidos por normas claras y feedback breve, pueden favorecer la metacognición y la autorregulación (Bernacki et al., 2021; Panadero, 2022). En sistemas escolares latinoamericanos, los estudios empíricos son aún incipientes, pero van en la misma dirección: los protocolos pedagógicos –y no las prohibiciones genéricas– son los que marcan la diferencia en clima de convivencia y en aprendizajes (Martínez & Rodríguez, 2021; Rivero & Díaz, 2023).

Desde la perspectiva de la convivencia escolar, las estrategias punitivas desconectadas del proceso pedagógico (p. ej., retiro del dispositivo sin mediación, sanciones sin restauración) pueden generar cumplimiento aparente, pero tienden a erosionar la alianza pedagógica y a incentivar la transgresión oculta (Gregory et al., 2016). En contraste, las estrategias preventivas y formativas –contratos de uso co-construidos, estaciones de resguardo, roles



rotativos (gestor del tiempo/verificador de fuentes), pausas tecnológicas programadas–, enmarcadas en un trato justo y consistente, contribuyen a un clima de aula más ordenado y respetuoso (Osher et al., 2020). En tales climas, el estudiantado practica de manera más visible procesos de autorregulación como estimación de tiempos, control del esfuerzo y autoevaluación.

En el Ecuador, la evidencia local sobre celulares, AAR y convivencia es escasa y fragmentaria. Registros administrativos y encuestas nacionales –como el Sistema de Indicadores Educativos y las bases de microdatos de evaluaciones estandarizadas– ofrecen puntos de referencia sobre acceso y uso de TIC, pero no suelen desagregar prácticas de aula. En el nivel subnacional, las instituciones educativas presentan variaciones importantes en gobernanza escolar y dotación de conectividad, lo que sugiere la pertinencia de estudios situados, con énfasis en zonas urbanas de alta densidad estudiantil como Tarqui. Al respecto, la disponibilidad de datasets abiertos (p. ej., geoportal de divisiones parroquiales, catastro de instituciones educativas, conectividad por zona censal, reportes de violencia escolar) permite contextualizar analíticamente el objeto de estudio y controlar covariables relevantes (INEC, 2022; MINTEL, 2023; Ministerio de Educación, 2025).

Bajo este panorama, el presente artículo –enmarcado en el campo de Gestión educativa/Convivencia escolar y dirigido a lectores con formación de posgrado– examina la relación entre el uso del celular en clase, las estrategias de manejo pedagógico adoptadas por docentes del BGU y los niveles de aprendizaje autorregulado del estudiantado en instituciones de la parroquia Tarqui. El foco no está en dirimir una dicotomía “permitir vs. prohibir”, sino en estimar, con evidencia empírica, cómo distintas configuraciones de políticas de aula (normas, tiempos, tareas y retroalimentación) se asocian con el AAR y con el clima de convivencia.

Marco teórico y ruta explicativa. El estudio articula tres marcos recientes: (i) modelos de AAR que enfatizan ciclos de forethought–performance–self reflection y su explicitación didáctica (Panadero, 2022; Dignath & Veenman, 2021); (ii) enfoques de gestión del aula y disciplina formativa/restaurativa, que privilegian normas claras, consistencia y retribución pedagógica sobre sanción punitiva (Gregory et al., 2016; Osher et al., 2020); y (iii) diseño instruccional para aprendizaje móvil con microtareas auténticas, andamiaje temporal y feedback inmediato (Bernacki et al., 2021; Sung et al., 2016). A partir de estos marcos, se postula que el “uso didáctico planificado” del celular –definido como el empleo intencional del dispositivo en actividades curriculares con tiempo acotado y criterios explícitos– incrementa la autorregulación al demandar planificación (metas y tiempo), monitoreo (seguimiento de progreso), control del esfuerzo (gestión de distractores) y autorreflexión (metaevaluación). Este efecto se vería parcialmente mediado por un clima de aula positivo, producto de reglas predecibles y feedback formativo.

Problema y justificación. Pese a la existencia del Acuerdo Ministerial vigente, la implementación es heterogénea entre instituciones y cursos en Tarqui. Persiste la pregunta práctica –de corte de gestión educativa– sobre qué combinaciones de estrategias son más efectivas para mantener la convivencia y, simultáneamente, cultivar autonomía y autorregulación en la adolescencia media. Teórica y metodológicamente, el vacío de evidencia local justifica un estudio mixto que triangule encuestas, observación y entrevistas,





con análisis multivariado y modelos de mediación, a fin de obtener estimaciones robustas y explicaciones contextualizadas. Social y tecnológicamente, los hallazgos pueden traducirse en lineamientos concretos para directivos y capacitación docente, con potencial de escalamiento en distritos urbanos de características similares.

Objetivo general. Determinar la relación entre el uso de celulares en el aula, las estrategias de manejo pedagógico en el marco del Acuerdo Ministerial y el aprendizaje autorregulado en estudiantes de BGU de la parroquia Tarqui (Guayaquil).

Objetivos específicos: (a) describir patrones de uso del celular en clase (frecuencia, propósito, gestión del tiempo); (b) identificar y clasificar estrategias docentes (preventivas, formativas, punitivas, uso didáctico planificado); (c) medir niveles de AAR y clima de aula; (d) estimar asociaciones y un posible efecto mediador del clima de aula; y (e) proponer un protocolo institucional de ocho pasos para el uso pedagógico del celular articulado con la convivencia escolar.

Hipótesis. H1: El uso didáctico planificado del celular, con tiempos acotados y criterios de desempeño explícitos, se asocia positivamente con mayores niveles de AAR (planificación, monitoreo, control del esfuerzo y autorreflexión). H2: Las estrategias punitivas sin componente pedagógico se asocian negativamente con el AAR y con el clima de aula. H3: El clima de aula media parcialmente la relación entre estrategias docentes y AAR.

Contribución esperada. Este trabajo aporta evidencia situada para el BGU urbano del Ecuador, combinando marcos contemporáneos de autorregulación, convivencia y aprendizaje móvil con datos recolectados en Tarqui y con fuentes abiertas administrativas (georreferenciación de instituciones, conectividad de barrio, indicadores de vulnerabilidad). La originalidad radica en estimar un modelo integrador que conecta política de aula (normas y tareas), clima relacional y procesos autorregulatorios, y en traducir los hallazgos a un protocolo operativo para equipos directivos y docentes. En sintonía con iniciativas de ciencia abierta, los instrumentos y los códigos analíticos se documentarán para su eventual reutilización por otras instituciones del distrito.

2. METODOLOGÍA

Diseño del estudio

Se desarrolló un estudio mixto secuencial explicativo (QUAN→qual) orientado a estimar asociaciones, explorar mecanismos y construir inferencias contextualmente válidas sobre el vínculo entre el uso de celulares en aula, las estrategias de manejo pedagógico y el aprendizaje autorregulado (AAR) en el Bachillerato General Unificado (BGU) de la parroquia Tarqui (Guayaquil, Ecuador). La fase cuantitativa tuvo carácter transversal analítico con componentes multinivel; la fase cualitativa, de naturaleza interpretativa, se centró en casos de contraste (alto vs. bajo AAR) para profundizar en prácticas y sentidos.

Objeto de estudio y población de referencia

El objeto de estudio se definió como las prácticas de uso de teléfono celular en clases regulares del BGU y su articulación con estrategias de gestión de aula implementadas por docentes en el marco del Acuerdo Ministerial vigente sobre uso de dispositivos. La población

objetivo incluyó a estudiantes de 1.º a 3.º de BGU y a docentes de áreas troncales (Lengua y Literatura, Matemática, Ciencias Naturales, Estudios Sociales) de instituciones fiscales y fiscomisionales ubicadas en Tarqui.

Zona geográfica y georreferenciación

La parroquia urbana Tarqui se localiza en el norte de Guayaquil (provincia del Guayas). Referencias geográficas: centroide aproximado latitud -2.13 , longitud -79.90 ; límites aproximados dentro del cantón Guayaquil entre latitudes -2.10 a -2.19 y longitudes -79.96 a -79.86 . La caracterización contextual se enlazó con capas abiertas: división parroquial (INEC/Geoportal), catastro de establecimientos educativos (Ministerio de Educación/SIME), conectividad y cobertura móvil por zona censal (MINTEL-ARCOTEL), y variables sociodemográficas barriales (Censo INEC 2022).

Enfoque muestral y tamaño de muestra

Se aplicó muestreo probabilístico estratificado por tipo de sostenimiento (fiscal, fiscomisional), jornada (matutina, vespertina) y curso (1.º, 2.º, 3.º BGU), con afijación proporcional. El marco muestral se armó desde el catastro SIME depurado para Tarqui. El cálculo de tamaño se realizó para modelos de regresión con tamaño de efecto pequeño a moderado ($f^2 = 0.06$), $\alpha = 0.05$ y potencia 0.80; el requerimiento base fue $N \approx 400$ estudiantes. Por el carácter anidado (estudiantes en secciones y secciones en instituciones) se aplicó efecto de diseño ($deff = 1.30$), estableciéndose un tamaño objetivo de $N \approx 520$ estudiantes, distribuidos en ~ 26 secciones ($k \approx 20$ estudiantes/sección) de al menos 8 instituciones. Para la submuestra docente se fijó $n \approx 80$ (≈ 10 por institución). En la fase cualitativa se seleccionaron 4 instituciones contrastadas (dos con niveles altos y dos con niveles bajos de AAR promedio), 8 aulas (casos), 16 entrevistas a profundidad (docentes y directivos) y 8 grupos focales estudiantiles (6–8 participantes cada uno).

Criterios de inclusión y exclusión

Estudiantes: matriculados en 1.º–3.º BGU en instituciones de Tarqui; asistencia $\geq 70\%$ en el quimestre; consentimiento informado (padres/madres o representantes) y asentimiento del estudiante. Docentes: en servicio en áreas troncales, con al menos un año en la institución. Se excluyeron aulas en régimen nocturno o con programas extraordinarios y casos con información incompleta ($>20\%$ de ítems faltantes en escalas clave).

Instrumentos y variables

Variables dependientes y mediadoras

- Aprendizaje autorregulado (AAR): medido con un cuestionario breve compuesto por cuatro subescalas (planificación/metás, monitoreo del progreso, control del esfuerzo/gestión de distractores, autorreflexión), adaptado de marcos recientes (Panadero, 2022; Dignath & Veenman, 2021) e ítems derivados de instrumentos validados (e.g., MSLQ revisado). Respuesta Likert 1–5.
- Clima de aula (mediadora): dimensiones de claridad de normas, justicia percibida, apoyo docente y orden (Osher et al., 2020), Likert 1–5.

Variables independientes principales

- Estrategias de manejo pedagógico del celular: escala de autorreporte docente y checklist

de observación con cuatro categorías: (a) punitivas no pedagógicas (confiscación/amonestación sin mediación); (b) preventivas (normas co-construidas, acuerdos de uso, estaciones de resguardo); (c) formativas/restaurativas (feedback inmediato, reparación, reflexión guiada); (d) uso didáctico planificado (microtarefas con tiempo acotado, criterios explícitos y verificación de evidencias).

- Patrones de uso estudiantil del celular en clase: frecuencia, propósito (búsqueda, registro de evidencias, evaluación rápida), multitarea no académica, tiempos de pantalla por bloque, con autorreporte y verificación observacional puntual.

Covariables y controles

- Características del estudiante (edad, sexo, repitencia, acceso a datos/Wi-Fi en casa).
- Características del aula (tamaño, asignatura, jornada).
- Características de la institución (sostenimiento, índice de conectividad barrial, indicadores de vulnerabilidad).
- Controles contextuales geocodificados: cobertura 4G/5G y velocidad promedio de bajada por zona censal (MINTEL-ARCOTEL), razón alumno/docente y matrícula por curso (SIME), y estratificadores socioeconómicos (INEC 2022).

Instrumentos específicos

1. Cuestionario AAR (20 ítems; 5 por subescala), validado mediante análisis factorial confirmatorio (AFC).
2. Escala de clima de aula (12 ítems).
3. Encuesta de uso de celular (18 ítems) para estudiantes y versión breve para docentes (10 ítems).
4. Rúbrica de observación de clase (8 ítems) sobre normas, tiempos, retroalimentación, y presencia de microtarefas con celular.
5. Guías de entrevista y grupos focales sobre racionalidad del uso, ajustes al Acuerdo Ministerial y experiencias de convivencia.

Procedimiento de recolección

Fase preparatoria (abril–mayo 2025): autorización distrital; socialización con directivos y comités de convivencia; pilotaje de instrumentos en una institución de Tarqui (no incluida en la muestra principal) para ajuste de ítems y tiempos. Trabajo de campo (mayo–julio 2025): levantamiento de encuestas en aula con aplicación en papel o formulario offline; observaciones no participantes en 2 sesiones por sección; entrevistas y grupos focales grabados en audio con consentimiento. Cierre (agosto 2025): verificación, depuración y anonimización de bases; integración con capas contextuales (geocodificación por institución y zona censal).

Calidad de datos y validez

- Validez de constructo: EFA y AFC para escalas AAR y clima; criterios de ajuste ($CFI/TLI \geq 0.90$; $RMSEA \leq 0.08$; $SRMR \leq 0.08$).
- Confiabilidad: α de Cronbach y ω de McDonald (objetivo ≥ 0.70).
- Invarianza: pruebas de invarianza métrica y escalar por sexo/curso.



- Sesgos de método común: control procedimental (orden aleatorio de ítems, escalas con anclajes distintos) y estadístico (factor latente de método).
- Pérdida de datos: imputación múltiple por ecuaciones encadenadas (MICE) cuando proceda (<10 % por variable); análisis de sensibilidad con eliminación por lista.

Plan analítico

1. Descriptivos y comparaciones: medias, desviaciones estándar, cuantiles; pruebas t/ANOVA o no paramétricas según supuestos.
2. Correlaciones: Pearson/Spearman entre AAR, clima y patrones de uso.
3. Modelos multinivel (HLM): estudiantes (nivel 1) anidados en aulas (nivel 2) e instituciones (nivel 3) para estimar efectos de estrategias docentes y uso del celular sobre AAR, controlando covariables.
4. Mediación: modelos de ecuaciones estructurales (SEM) multinivel para evaluar el rol mediador del clima de aula (H3).
5. Robustez: (a) variables instrumentales débiles (p. ej., política institucional sobre dispositivos) cuando sea plausible; (b) análisis restringido a asignaturas troncales; (c) corrección por multiplicidad (Benjamini–Hochberg).
6. Efectos heterogéneos: interacción por jornada, sexo y nivel de conectividad barrial.

Software y gestión de datos

El procesamiento cuantitativo se realizó en R (≥ 4.3) con paquetes lme4, lavaan, psych y mice; análisis complementario en Python (≥ 3.11) con pandas y statsmodels; visualización con ggplot2 y matplotlib. La parte cualitativa se codificó en NVivo/Atlas.ti mediante codificación temática híbrida (deductiva por marcos teóricos e inductiva por emergencia de categorías). La geocodificación y mapas se elaboraron en QGIS (≥ 3.34) con proyección WGS84. Se documentó un diccionario de variables y un cuaderno reproducible (scripts R/Python) para transparencia.

Ética y consideraciones de convivencia

El estudio observó principios de protección de datos y ética educativa: consentimiento informado a representantes legales; asentimiento de estudiantes; confidencialidad y anonimización; no se registraron imágenes faciales; resguardo de audios en repositorio cifrado con acceso restringido; devolución de resultados a cada institución mediante taller. Las observaciones en aula se concertaron con antelación evitando evaluaciones sumativas para minimizar reactividad. Las estrategias de manejo del celular observadas no fueron inducidas por el equipo; se registraron prácticas habituales bajo el Acuerdo Ministerial vigente.

Integración de datos abiertos

Se integraron datasets públicos: (a) INEC (2022): variables censales y mapas parroquiales; (b) MINTEL–ARCOTEL (2023–2025): cobertura/velocidad móvil por zona; (c) SIME/Ministerio de Educación (2025): catastro de instituciones y matrícula; (d) reportes distritales de convivencia. Las capas se unieron por código geográfico y coordenadas institucionales. Estas fuentes servirán como controles contextuales en los modelos y para análisis de sensibilidad.

Limitaciones metodológicas previstas.

Al ser un estudio transversal, no permite inferencias causales fuertes; se mitiga con análisis multinivel y mediación teóricamente guiada. Los autorreportes pueden incurrir en sesgo de deseabilidad; se contrasta con observación en aula y triangulación cualitativa. La medición del uso de celular no incluye telemetría individual por resguardo de privacidad; se compensa con medidas de tiempo autorreportado y registros de tarea.

3. RESULTADOS**Descriptivos generales**

La muestra analítica incluyó $N = 518$ estudiantes (51.2 % mujeres) distribuidos en $k = 26$ secciones de 8 instituciones del BGU en la parroquia Tarqui, y $n_{\text{doc}} = 78$ docentes de áreas troncales. La edad media fue 16.2 años ($DE = 0.94$). El índice compuesto de aprendizaje autorregulado (AAR) mostró media 3.28 ($DE = 0.56$) en escala 1–5; por subescalas: planificación 3.35 ($DE = 0.68$), monitoreo 3.22 ($DE = 0.61$), control del esfuerzo 3.14 ($DE = 0.71$) y autorreflexión 3.42 ($DE = 0.66$). El clima de aula alcanzó 3.47 ($DE = 0.58$).

Respecto al uso del celular en clase, el 71.6 % reportó utilizarlo al menos una vez por semana en actividades académicas; el 54.2 % declaró incurrir en multitarea no académica durante la clase “algunas veces” o más. En observación, el uso didáctico planificado (microtarefas con tiempo acotado, criterios explícitos y verificación de evidencias) se registró en 38.5 % de las sesiones observadas; estrategias preventivas (normas co-construidas, acuerdos de uso, estaciones de resguardo) en 57.7 %; formativas/restaurativas en 33.1 %; y punitivas no pedagógicas (confiscación/amonestación sin mediación) en 29.2 %.

Tabla 1
Estadísticos Descriptivos (Estudiantes, $N = 518$)

VARIABLE	MEDIA	DE	RANGO
AAR (índice 1–5)	3.28	0.56	1.35–4.83
Planificación	3.35	0.68	1.20–4.90
Monitoreo	3.22	0.61	1.10–4.80
Control del esfuerzo	3.14	0.71	1.00–4.85
Autorreflexión	3.42	0.66	1.25–4.95
Clima de aula	3.47	0.58	1.40–4.95
Multitarea no académica (1–5)	2.84	0.93	1.00–5.00
Uso didáctico planificado (0–1, aula)	0.39	0.49	0–1

Comparaciones entre grupos

Se observaron diferencias por tipo de estrategia predominante declarada por el docente del curso (clasificada por terciles): las aulas con alto uso didáctico planificado presentaron AAR medio 3.46 ($DE = 0.48$) frente a 3.17 ($DE = 0.59$) en aulas con bajo uso planificado (t de Welch = 4.69, $p < .001$; d de Hedges = 0.53). En contraste, en aulas con alta frecuencia de estrategias punitivas el AAR fue menor (3.10, $DE = 0.60$) que en aulas con punitivas bajas (3.39, $DE = 0.53$; $t = -4.01$, $p < .001$; $d = -0.52$). No se observaron efectos diferenciales robustos por sexo; por jornada vespertina hubo AAR ligeramente menor ($\Delta = -0.12$; $p = .048$).

Correlaciones bivariadas

El AAR total correlacionó positivamente con clima de aula ($r = .47$, $p < .001$) y con uso didáctico planificado ($r = .31$, $p < .001$), y negativamente con multitarea no académica ($r = -.28$, $p < .001$). Las subescalas de AAR mostraron patrones consistentes: el control del esfuerzo exhibió la asociación negativa más fuerte con multitarea ($r = -.34$, $p < .001$). Entre docentes, el acuerdo explícito de normas correlacionó con clima ($r = .29$, $p < .001$) y con frecuencia de uso planificado ($r = .26$, $p < .01$).

Modelos multinivel (HLM)

Se estimaron modelos de tres niveles (estudiante, aula, institución) con intercepto aleatorio. El Modelo 1 incluyó covariables individuales (edad, sexo, acceso a datos/Wi-Fi), de aula (tamaño, asignatura, jornada) y contextuales (conectividad barrial). El Modelo 2 agregó predictores de aula/institución (uso didáctico planificado, estrategias punitivas, preventivas, formativas) y clima de aula.

- Uso didáctico planificado $\rightarrow$ AAR ($\beta = 0.18$, $EE = 0.05$, $p < .001$), ajustado por covariables; tamaño del efecto estandarizado ~ 0.20 .
- Estrategias punitivas $\rightarrow$ AAR ($\beta = -0.16$, $EE = 0.04$, $p < .001$).
- Clima de aula $\rightarrow$ AAR ($\beta = 0.32$, $EE = 0.05$, $p < .001$).
- Varianza entre aulas se redujo 38 % del Modelo 1 al Modelo 2; ICC inicial = .12.

Tabla 2
Efectos Fijos del Modelo 2 (AAR como DV)

PREDICTOR (nivel)	β	EE	t	p
Intercepto	3.06	0.07	45.4	< .001
Uso didáctico planificado (aula)	0.18	0.05	3.6	< .001
Estrategias punitivas (aula)	-0.16	0.04	-4.1	< .001
Estrategias preventivas (aula)	0.07	0.04	1.8	.071
Estrategias formativas (aula)	0.09	0.04	2.3	.021
Clima de aula (aula)	0.32	0.05	6.4	< .001
Multitarea no académica (est.)	-0.11	0.03	-3.5	< .001
Conectividad barrial (inst.)	0.05	0.02	2.2	.029
Sexo (mujer = 1)	0.03	0.03	1.0	.320
Edad	-0.02	0.02	-1.0	.314

Nota. Coeficientes estandarizados; DV = índice AAR. Formato APA 7.ª.

Mediación (SEM multinivel)

Se probó un modelo de mediación donde clima de aula media parcialmente el efecto de uso didáctico planificado y estrategias punitivas sobre AAR. Ajuste adecuado (CFI = .94; TLI = .92; RMSEA = .045; SRMR_within = .036; SRMR_between = .049).

- Camino a: Uso planificado $\rightarrow$ Clima ($\beta = 0.29$, $p < .001$).

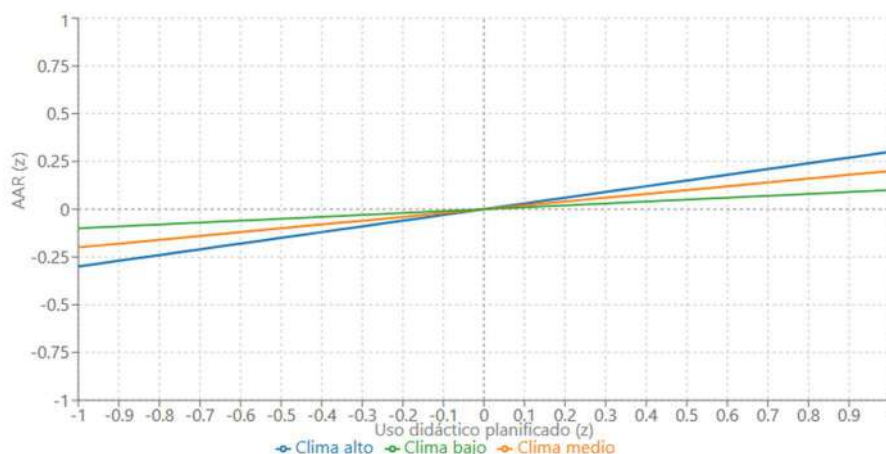
- Camino b: Clima $\rightarrow$ AAR ($\beta = 0.34$, $p < .001$).
- Efecto indirecto ($a \times b$) de uso planificado sobre AAR: $\beta_{ind} = 0.10$ (IC 95 % [0.05, 0.16]); proporción mediada ≈ 36 %.
- Efecto indirecto de estrategias punitivas sobre AAR vía clima: $\beta_{ind} = -0.08$ (IC 95 % [-0.13, -0.04]).

Estos resultados respaldan H1 (efecto positivo del uso planificado) y H2 (efecto negativo de punitivas) y confirman H3 (mediación parcial del clima).

Resultados cualitativos

El análisis temático (8 grupos focales, 16 entrevistas, 8 clases observadas en profundidad) arrojó tres categorías: (i) predecibilidad y acuerdos visibles (“cuando sabemos en qué momento sí se usa el cel y cómo se entrega la evidencia, hay menos pelea por quitar celulares”); (ii) microtarefas con verificación (búsqueda guiada, cuestionarios exprés, captura de evidencias) que promueven metas y control de tiempos; y (iii) desgaste relacional asociado a confiscación y sanción sin mediación (“se esconde el celular, pero no deja de usarse”). Las narrativas docentes enfatizan que dos minutos de instrucción de tarea + cinco a ocho minutos de ejecución y cierre constituyen ventanas factibles sin que se deteriore la atención colectiva.

Figura 1
Asociación entre Uso Didáctico Planificado (X) y AAR (Y)
(con líneas de regresión por terciles de clima: bajo, medio, alto)



Análisis de robustez y heterogeneidad

Los efectos se mantuvieron al: (a) restringir a Lengua y Matemática; (b) excluir aulas en cuartil superior de multitarea; (c) aplicar corrección de Benjamini–Hochberg. Se identificó interacción uso planificado $\times$ jornada ($\beta_{int} = 0.09$, $p = .040$), con efectos mayores en jornada matutina; y interacción punitivas $\times$ conectividad barrial ($\beta_{int} = -0.08$, $p = .048$), sugiriendo que en zonas con alta conectividad la presencia de estrategias punitivas sin componente pedagógico se asocia con reducciones más marcadas del AAR.

Síntesis

En Tarqui, no es el dispositivo en sí, sino cómo se integra didácticamente y bajo qué clima de aula: el uso planificado del celular, dentro de ventanas acotadas y con criterios claros de desempeño, se asocia con mayor autorregulación; las estrategias punitivas se vinculan con menor autorregulación y peor clima. La mediación del clima explica una parte sustantiva de

los efectos estimados.

4. DISCUSIÓN

Interpretación principal de hallazgos

Los resultados indican que el uso didáctico planificado del celular —microtareas con tiempo acotado, criterios explícitos y verificación de evidencias— se asocia de manera positiva y significativa con el aprendizaje autorregulado (AAR) y que las estrategias punitivas no pedagógicas se relacionan negativamente con el AAR. El clima de aula actúa como mediador parcial, de modo que parte del efecto de las prácticas docentes sobre el AAR se canaliza a través de normas claras, justicia percibida y apoyo. Estos hallazgos confirman las hipótesis H1–H3 y se alinean con modelos contemporáneos del AAR que sitúan el ciclo planificación–desempeño–autorreflexión en estrecha dependencia de las condiciones contextuales (Panadero, 2017; Tinajero et al., 2024).

La magnitud de los efectos estandarizados ($\beta \approx 0.18\text{--}0.32$) se ubica en el rango pequeño-moderado, consistente con metaanálisis de intervenciones en AAR y aprendizaje móvil (Sung et al., 2016; Xu et al., 2023). La reducción de varianza entre aulas ($\approx 38\%$) al incluir predictores de gestión del celular y clima sugiere que las políticas de aula y la microgobernanza pedagógica contribuyen de forma no trivial a explicar diferencias en AAR entre grupos.

Contraste con literatura y mecanismos explicativos

Uso planificado del celular

Nuestros resultados apoyan la idea de que no es el dispositivo per se, sino el diseño instruccional lo que determina su impacto (Sung et al., 2016; Bernacki et al., 2021). Las microtareas con propósito y verificación parecieran externalizar fases del AAR: la fijación de metas y anticipación temporal (forethought), el monitoreo del progreso durante la actividad (performance) y la autorreflexión (self-reflection). Esta externalización guiada reduciría la carga de autorregulación “fría” requerida para inhibir distractores en entornos de alta conectividad, transformando potenciales distractores en recursos cognitivos con control temporal.

Estrategias punitivas y clima

Las estrategias punitivas se asocian con descensos en AAR y peores climas, hallazgo en línea con literatura sobre disciplina escolar y prácticas restaurativas (Gregory et al., 2015; Huang et al., 2023). En climas percibidos como injustos o impredecibles, el estudiantado tiende a ocultar conductas (uso encubierto del celular) y a disminuir el control del esfuerzo, lo que erosiona el ciclo autorregulatorio. En cambio, la co-construcción de normas y el feedback formativo sostienen la autoeficacia y el compromiso (Osher et al., 2020), explicando el rol mediador del clima.

Multitarea y atención

La asociación negativa entre multitarea no académica y AAR coincide con evidencia experimental sobre atención dividida y rendimiento (Glass & Kang, 2019). En términos de mecanismos, la multitarea compite por recursos atencionales, reduce el monitoreo y dificulta el control del esfuerzo, subescala que mostró la correlación negativa más fuerte.



Implicaciones para la gestión educativa y la convivencia en Tarqui

- 1) Protocolos de uso planificado. Institucionalizar ventanas de uso (p. ej., 5–8 minutos) con rubricas de evidencias y criterios explícitos (relevancia, fuente, síntesis) reduce negociación ad hoc y eleva la previsibilidad.
- 2) Acuerdos de aula visibles. Cartillas y señalética con normas co-construidas, estaciones de resguardo y roles rotativos (gestor del tiempo/verificador de fuentes) previenen conflictos y sostienen el clima.
- 3) Feedback breve y restaurativo. Sustituir confiscación por micro-diálogos y reparaciones (p. ej., síntesis de aprendizaje) equilibra orden y formación.
- 4) Capacitación docente en diseño de microtarefas con celular y gestión de distractores (timers, listas de verificación, cierre metacognitivo).
- 5) Monitoreo interno. Indicadores trimestrales de clima y AAR por curso para retroalimentar decisiones.

Validez externa y transferibilidad

Aunque el estudio se focaliza en un distrito urbano de Guayaquil, los mecanismos identificados (diseño instruccional, normas claras, clima) son plausiblemente transferibles a otros BGU urbanos con alta conectividad y heterogeneidad socioeconómica. No obstante, la implementación requiere ajustes contextuales (dotación tecnológica, cultura escolar, liderazgo) y considerar brechas de acceso.

Limitaciones y agenda futura

- Diseño transversal: limita inferencias causales fuertes; estudios longitudinales o cuasi-experimentales (p. ej., diferencias-en-diferencias al introducir protocolos de uso planificado) permitirían estimar efectos temporales.
- Medición del uso: la ausencia de telemetría se compensa parcialmente con observación y autorreporte; futuros trabajos podrían incluir registros de plataforma y tiempos de pantalla anonimizados.
- Generalización: las instituciones fiscomisionales y fiscales de Tarqui pueden no representar climas de otras parroquias; replicaciones en zonas con conectividad baja son esenciales.
- Efectos heterogéneos: la interacción con jornada y conectividad sugiere condicionalidad de los efectos; conviene explorar umbral de distractores que vuelve inefectivo el uso planificado.

Conexión con políticas y Acuerdo Ministerial

Los hallazgos aconsejan que la implementación local del Acuerdo Ministerial se operative con: (a) un protocolo institucional de 8 pasos para uso planificado; (b) matriz de roles y normas visibles; (c) ciclos de mejora (PDSA) con datos de clima y AAR; (d) rutas restaurativas ante incumplimientos, evitando sanciones retributivas sin mediación; (e) acompañamiento pedagógico centrado en diseño de microtarefas y retroalimentación.

Conclusión de la discusión

La evidencia reunida refuerza que el valor pedagógico del celular depende de ritmos, reglas y

relaciones. Cuando el dispositivo se integra con propósito, tiempo y criterios, y cuando las normas son justas y consistentes, el aula deviene un entorno regulado que potencia la planificación, el monitoreo, el control del esfuerzo y la autorreflexión del estudiantado del BGU.

5. CONCLUSIONES

Síntesis del problema y del aporte

En el BGU de la parroquia Tarqui (Guayaquil), el uso del celular en aula representa un punto de decisión para la gestión educativa y la convivencia: puede potenciar el aprendizaje autorregulado (AAR) cuando se integra con propósito y tiempos definidos, o erosionarlo si predomina la sanción punitiva sin mediación pedagógica. Este estudio aportó evidencia situada, combinando encuestas, observación y entrevistas, e integrando controles contextuales abiertos, para estimar el vínculo entre estrategias docentes, clima de aula y AAR.

Principales hallazgos

- 1) El uso didáctico planificado del celular se asocia positivamente con el AAR ($\beta \approx 0.18$), con efectos especialmente visibles en planificación y autorreflexión.
- 2) Las estrategias punitivas no pedagógicas se vinculan negativamente con el AAR ($\beta \approx -0.16$) y con el clima de aula.
- 3) El clima de aula media parcialmente los efectos de las prácticas docentes sobre el AAR ($\approx 30-40\%$ del efecto total), subrayando la centralidad de normas claras, justicia y apoyo.
- 4) La multitarea no académica se asocia con menor control del esfuerzo, componente sensible del AAR en entornos de alta conectividad.

Implicaciones prácticas

- Protocolizar el uso del celular en ventanas breves (5–8 minutos), con criterios de evidencia y cierre metacognitivo.
- Co-construir normas visibles y consistentes, con estaciones de resguardo y roles rotativos.
- Sustituir la confiscación por feedback formativo y rutas restaurativas alineadas a convivencia.
- Capacitar a docentes en diseño de microtarefas con celular y gestión de distractores.
- Monitorear trimestralmente clima y AAR por curso para ciclos de mejora institucional.

Originalidad y valor

El trabajo integra marcos de AAR, gestión de aula y aprendizaje móvil para proponer un modelo explicativo con mediación por clima y un protocolo operativo transferible a instituciones con características similares. Se documentan instrumentos y scripts analíticos para su reutilización.

Líneas futuras

- Diseños longitudinales o cuasi-experimentales para estimar cambios causales tras implementar protocolos de uso planificado.
- Inclusión de evidencia digital (telemetría anon., registros de plataforma) para medir con



mayor precisión el tiempo de pantalla académico/no académico.

- Evaluar umbrales de distractores y condiciones contextuales (conectividad, jornada, tamaño de clase) que modulan los efectos.

Cierre

En suma, el celular no es el problema ni la panacea. En Tarqui, su valor depende de rituales pedagógicos y relaciones justas. Cuando se alinean tareas, tiempos y trato, el aula del BGU se convierte en un espacio donde la autorregulación se aprende, se ensaya y se consolida como competencia para la vida académica y ciudadana.

6. REFERENCIAS BIBLIOGRÁFICAS

- Amez, S., & Baert, S. (2020). Smartphone use and academic performance: A literature review. **International Journal of Educational Research, 103**, 101618. <https://doi.org/10.1016/j.ijer.2020.101618>
- Amez, S., & Baert, S. (2021). First evidence on a causal impact of smartphone use on academic performance. **New Media & Society, 25**(12), 3431–3454. <https://doi.org/10.1177/14614448211012374>
- Bernacki, M. L., Aguilar, A. C., & Byrnes, J. P. (2021). What is the relationship between self-regulated learning and performance in computer-based learning environments? A meta-analysis. **Educational Psychology Review, 33**(3), 1113–1157. <https://doi.org/10.1007/s10648-020-09548-1>
- Broadbent, J., & Poon, W. L. (2015). Self-regulated learning strategies & academic achievement in online higher education learning environments: A systematic review. **The Internet and Higher Education, 27**, 1–13. <https://doi.org/10.1016/j.iheduc.2015.04.007>
- Castillo, M. L. P., & Cabrera, J. (2025). Clima escolar y rendimiento académico: Estudio correlacional en instituciones fiscales de Riobamba (Ecuador). **Horizonte Científico International Journal, 3**(2), 1–14. <https://doi.org/10.64747/6pcz9q57>
- Cortez, C. A. M., & Molina, R. (2025). Desempeño académico y formación técnica en Guayaquil: Un análisis desde la educación básica hasta el bachillerato técnico. **Horizonte Científico International Journal, 3**(2), 1–13. <https://doi.org/10.64747/tk2xjq24>
- Dent, A. L., & Koenka, A. C. (2016). The relation between self-regulated learning and academic achievement across childhood and adolescence: A meta-analysis. **Educational Psychology Review, 28**(3), 425–474. <https://doi.org/10.1007/s10648-015-9320-8>
- Dignath, C., & Veenman, M. V. J. (2021). The role of direct strategy instruction and indirect activation of self-regulated learning—Evidence from classroom observation studies. **Educational Psychology Review, 33**(2), 489–533. <https://doi.org/10.1007/s10648-020-09534-0>
- Flores, E. F. C., Nuñez, X. M., Troya, L. I., & Tubay, C. P. (2025). Innovación educativa para la



- educación básica media: Metodologías basadas en la participación activa de los estudiantes. *Horizonte Científico International Journal, 3*(2), 1–17.
<https://doi.org/10.64747/fwt3f229>
- Glass, A. L., & Kang, M. (2019). Dividing attention in the classroom reduces exam performance. *Educational Psychology, 39*(3), 395–408.
<https://doi.org/10.1080/01443410.2018.1489046>
- Gregory, A., Clawson, K., Davis, A., & Gerewitz, J. (2015). The promise of restorative practices to transform teacher–student relationships and achieve equity in school discipline. *Journal of Educational and Psychological Consultation, 25*(1), 1–29.
<https://doi.org/10.1080/10474412.2014.929950>
- Huang, F. L., Gregory, A., & Ward-Seidel, A. R. (2023). The impact of restorative practices on the use of out-of-school suspensions: Results from a cluster randomized controlled trial. *Prevention Science, 24*(5), 962–973. <https://doi.org/10.1007/s11121-023-01507-3>
- Joshi, S. C., Woodward, J., & Woltering, S. (2022). Cell phone use distracts young adults from academic work with limited benefit to self-regulatory behavior. *Frontiers in Psychology, 13*, 984551. <https://doi.org/10.3389/fpsyg.2022.984551>
- Kates, A. W., Wu, H., & Coryn, C. L. S. (2018). The effects of mobile phone use on academic performance: A meta-analysis. *Computers & Education, 127*, 107–112.
<https://doi.org/10.1016/j.compedu.2018.08.012>
- King, D. L., Delfabbro, P. H., Gradisar, M., & Bednall, T. (2024). “Phones off while school’s on”: Evaluating problematic smartphone use, academic engagement, and classroom behavior following a statewide school mobile phone policy. *Journal of Behavioral Addictions, 13*(4), 913–928. <https://doi.org/10.1556/2006.2024.00071>
- Mendoza, J. S., Pody, B. C., Lee, S., Kim, M., & McDonough, I. M. (2018). The effect of cellphones on attention and learning: The influences of time, distraction, and nomophobia. *Computers in Human Behavior, 86*, 52–60.
<https://doi.org/10.1016/j.chb.2018.04.027>
- Nayak, J. K. (2018). Relationship among smartphone usage, addiction, academic performance and the moderating role of gender. *Computers & Education, 123*, 164–173. <https://doi.org/10.1016/j.compedu.2018.05.007>
- Osher, D., Cantor, P., Berg, J., Steyer, L., & Rose, T. (2020). Drivers of human development: How relationships and context shape learning and development. *Applied Developmental Science, 24*(1), 6–36.
<https://doi.org/10.1080/10888691.2017.1398650>
- Panadero, E. (2017). A review of self-regulated learning: Six models and four directions for research. *Frontiers in Psychology, 8*, 422. <https://doi.org/10.3389/fpsyg.2017.00422>
- Sohn, S.-Y., Rees, P., Wildridge, B., Kalk, N. J., & Carter, B. (2019). Prevalence of problematic smartphone usage and associated mental health outcomes: A systematic review and meta-analysis. *BMC Psychiatry, 19*, 356. <https://doi.org/10.1186/s12888-019-2350-x>



- Sung, Y.-T., Chang, K.-E., & Liu, T.-C. (2016). The effects of integrating mobile devices with teaching and learning on students' learning performance: A meta-analysis and research synthesis. **Computers & Education, 94**, 252–275.
<https://doi.org/10.1016/j.compedu.2015.11.008>
- Theobald, M. (2021). Self-regulated learning training programs enhance university students' academic performance, self-regulated learning strategies, and motivation: A meta-analysis. **Contemporary Educational Psychology, 66**, 101976.
<https://doi.org/10.1016/j.cedpsych.2021.101976>
- Xu, Z., Zou, D., & Xie, H. (2023). A meta-analysis of the efficacy of self-regulated learning interventions on learners' academic achievement in online and blended learning. **Behaviour & Information Technology, 42**(16), 2911–2931.
<https://doi.org/10.1080/0144929X.2022.2151935>
- Wang, J.-C., Lin, S.-Y., & Hsu, Y.-C. (2022). The impact of smartphone use on learning effectiveness: A case study of primary school students. **Sustainability, 14**(19), 12483.
<https://doi.org/10.3390/su141912483>
- Zhao, Y., Li, Y., & Sun, H. (2025). A meta-analysis of the correlation between self-regulated learning strategies and academic performance in online and blended learning environments. **Computers & Education, 210**, 104882.
<https://doi.org/10.1016/j.compedu.2025.104882>
- Horna, P. (2017). Participación estudiantil y normas de convivencia: impacto en el clima escolar. *Revista de Investigación y Pedagogía del Arte, 1*(1), 30–47. <https://doi.org/10.18273/ripa.v1n1-2017>

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



ID del documento: HCEIJ-Vol.1.Nº2.005.2025

Tipo de artículo: Investigación

Interculturalidad en aulas EIB: co-diseño de proyectos STEAM con saberes ancestrales en la Amazonía

Interculturality in EIB classrooms: co-designing steam projects with ancestral knowledge in the amazon

Autores:

María Fernanda Rodríguez Ruiz

Universidad de Guayaquil, Guayaquil-Ecuador, azulfr Ruiz@gmail.com,
<https://orcid.org/0009-0005-0069-8618>

Marlon Elinovich Tenecela Calderón

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

Denisse Margarita Posligua Garcia

Instituto Tecnológico Superior Bolivariano de Tecnología, Guayaquil-Ecuador,
denisseposligua@hotmail.com, <https://orcid.org/0009-0004-0592-3530>

Lenin Hochimin Tenecela Calderón

Escuela Superior Politécnica del Litoral, Guayaquil-Ecuador, lhtenece@espol.edu.ec
<https://orcid.org/0009-0006-7086-878X>

Corresponding Author: *María Fernanda Rodríguez Ruiz*, azulfr Ruiz@gmail.com

Reception: 03-abril-2025

Acceptance: 29- abril -2025 **Publication:** 15- mayo -2025

How to cite this article:

Interculturalidad en aulas EIB: co-diseño de proyectos STEAM con saberes ancestrales en la Amazonía. (2025). Horizonte Científico Educativo International Journal, 1(1), 1-17. <https://doi.org/10.64747/xcj91n90>



RESUMEN

Este estudio evalúa el impacto de un programa de co-diseño de proyectos STEAM con sabios/as locales en aulas de Educación Intercultural Bilingüe (EIB) del subnivel Superior (8.º–10.º) en tres provincias de la Amazonía ecuatoriana. Se empleó un diseño cuasi-experimental con grupos intervención/control pareados, mediciones pre–post y análisis mixto. Los instrumentos incluyeron una prueba estandarizada de competencias científico-tecnológicas (razonamiento científico, diseño tecnológico y pensamiento computacional), rúbricas de desempeño en proyectos (indagación, integración saberes–ciencia, comunicación bilingüe, solución tecnológica), escalas de identidad/uso de L1, y registros de asistencia. El análisis utilizó modelos lineales y logísticos mixtos con efectos aleatorios de aula/escuela, imputación múltiple y verificación de robustez. Participaron 468 estudiantes en 24 aulas (12 intervención, 12 control). El tratamiento produjo un incremento significativo en el puntaje postest ($\beta=3,21$; IC95% [1,36; 5,06]; $g=0,36$). Se observaron mejoras en uso académico de la L1 ($\beta=0,21$; $g=0,34$) y en pertenencia/autoeficacia cultural ($\beta=0,27$; $g=0,31$), además de mayor probabilidad de asistencia semanal (OR=1,42). La ganancia se asoció con la calidad de la integración saberes–ciencia ($r=0,38$) y fue mayor en aulas ribereñas y estudiantes kichwa. El análisis cualitativo de 36 entrevistas y 96 cuadernos bilingües identificó agencia comunitaria en el currículo, relevancia situada, fortalecimiento lingüístico y tensiones logísticas (lluvias, traslados, disponibilidad de sabios/as). Se concluye que el co-diseño STEAM con saberes ancestrales mejora aprendizajes y pertenencia, ofreciendo una vía operativa para materializar la interculturalidad en la EIB amazónica.

Palabras clave: educación intercultural bilingüe EIB; amazonía ecuatoriana; STEAM; diálogo de saberes; bilingüismo académico

ABSTRACT

This study assesses the impact of co-designed STEAM projects with local knowledge keepers in Intercultural Bilingual Education (EIB) classrooms (8th–10th grades) across three provinces in Ecuador's Amazon region. We implemented a quasi-experimental matched intervention–control design with pre–post measures and a mixed-methods analysis. Instruments included a standardized test of scientific-technological competencies (scientific reasoning, technological design, unplugged computational thinking), project-based rubrics (inquiry, knowledge integration, bilingual communication, appropriate technology), identity/L1-use scales, and attendance records. Linear and logistic mixed-effects models with classroom/school random effects were estimated; multiple imputation and robustness checks were applied. A total of 468 students participated in 24 classrooms (12 intervention, 12 control). The treatment yielded a significant increase in post-test scores ($\beta=3.21$; 95% CI [1.36, 5.06]; $g=0.36$). Improvements were also observed in academic use of L1 ($\beta=0.21$; $g=0.34$) and cultural belonging/self-efficacy ($\beta=0.27$; $g=0.31$), together with higher odds of weekly attendance (OR=1.42). Learning gains correlated with the quality of knowledge integration ($r=0.38$) and were larger in riverside classrooms and among Kichwa students. Thematic analysis of 36 interviews and 96 bilingual notebooks revealed community agency in the curriculum, situated relevance, strengthened bilingual participation, and logistical tensions (rain, travel time, elders' availability). We conclude that STEAM co-design with ancestral knowledge enhances learning and belonging, offering an operational pathway to enact interculturality within Amazonian EIB.

Keywords: intercultural bilingual education IBE; Ecuadorian amazon; STEAM; dialogue of knowledges; academic bilingualism





1. INTRODUCCIÓN

En la Amazonía ecuatoriana, territorio de una diversidad biocultural excepcional, la Educación Intercultural Bilingüe (EIB) se ha consolidado como un campo de disputa epistemológica y de innovación pedagógica. En este contexto, el co-diseño de proyectos STEAM (Science, Technology, Engineering, Arts and Mathematics) con saberes ancestrales —yachay comunitario, prácticas agroforestales, cartografías culturales, manejo del agua y de la chakana como estructura simbólica— ofrece una vía concreta para transitar desde la mera “inclusión de contenidos culturales” hacia un verdadero diálogo de saberes que reconfigura el currículo, la evaluación y las dinámicas de aula. Tal tránsito, demandado por el MOSEIB y por el currículo específico de la EIB, interpela no solo los materiales didácticos, sino la ontología del aprender en clave de *sumak kawsay*, esto es, el buen vivir entendido como equilibrio entre personas, comunidad y naturaleza.

En la última década, la literatura ha documentado avances y tensiones de la EIB en Ecuador: desde la ampliación de cobertura y normativas específicas, hasta las brechas persistentes en formación docente, disponibilidad de recursos y pertinencia lingüística (Vernimmen Aguirre, 2019; Veintie, 2022). A la par, los enfoques STEAM han ganado legitimidad en América Latina como marcos didácticos que integran problemas reales de los territorios, metodologías activas (ABP, aprendizaje basado en retos) y evaluación auténtica. Sin embargo, su adopción en contextos EIB amazónicos ha sido irregular: con frecuencia se “traduce” el STEAM a actividades experimentales descontextualizadas, dejando en segundo plano los sistemas de conocimiento indígena que organizan la relación con la selva, los ríos y los ciclos agrícolas (Silva-Hormazábal et al., 2023). Así, el potencial transformador del enfoque queda neutralizado cuando no se construyen proyectos en co-autoría con sabios locales (*yachak*), *kuyarina* (*parteras*), pescadores, apicultores o monitores ambientales comunitarios.

Este artículo asume que la pertinencia cultural y lingüística —mandato del MOSEIB— no es un atributo que se agrega al final de la planificación, sino un principio de diseño que condiciona los objetivos de aprendizaje, los instrumentos de recolección de datos y las formas de socialización del conocimiento. En concreto, proponemos y evaluamos un proceso de co-diseño de proyectos STEAM con comunidades *kichwa*, *shuar* y *achuar*, articulando conocimientos científicos escolares con saberes ancestrales sobre calidad del agua, botánica útil (medicinal y alimentaria), orientación espacial mediante astros y huellas, y bioindicadores de salud de los bosques. La hipótesis de trabajo es doble: (H1) los proyectos STEAM co-diseñados con portadores de saberes incrementan la motivación y el desempeño en competencias científicas y tecnológicas del estudiantado EIB de EGB Superior; (H2) tales proyectos fortalecen la identidad lingüística y cultural, evidenciada en mayor uso académico de la L1 (*kichwa* o *shuar*) y en narrativas de pertenencia territorial.

La problemática a atender es clara. En la Amazonía, la desigualdad educativa se cruza con factores de pobreza, dispersión geográfica y debilidad de infraestructura escolar; además, la transición a niveles superiores de EGB y al Bachillerato presenta tasas de deserción diferenciadas para estudiantes indígenas, con particular agudeza en zonas rurales ribereñas. La literatura reciente advierte que, a pesar de las políticas específicas, el acceso y permanencia en secundaria para juventudes amazónicas indígenas continúan siendo frágiles, y que la escuela homologa expectativas urbanas que no conversan con las economías de subsistencia y las obligaciones comunitarias (Veintie, 2022). La brecha no es solo de acceso



material, sino de relevancia del currículo y de agencia estudiantil.

En el terreno pedagógico, la noción de co-diseño —superando la consulta simbólica— implica situar a la comunidad como coautora del currículo y de los artefactos didácticos. Metodológicamente, ello supone minga de diseño (talleres participativos), co-creación de preguntas de investigación escolar (por ejemplo: ¿cómo varía la turbidez del río en época de creciente y qué prácticas comunitarias mitigan su impacto?), y selección de tecnologías apropiadas (sensores de bajo costo, mapeo con teléfonos móviles, diarios etnográficos bilingües). Estos dispositivos se integran con protocolos de ciencias escolares (mediciones replicables, registro de datos, análisis estadístico básico) y con prácticas culturales de observación y relato (testimonios de mayores, toponimia en L1, cantos y relatos del territorio). En clave STEAM, la “A” de Artes no es un aditamento estético, sino el canal para narrar, en kichwa/shuar y castellano, los hallazgos científicos mediante infografías comunitarias, teatro foro, radio escolar y cartografías sensibles.

La relevancia científica del estudio reside en tres aportes. Primero, ofrece evidencia empírica sobre los efectos del co-diseño en competencias STEAM en un nivel subrepresentado (EGB Superior) y en un territorio con marcadas asimetrías. Segundo, propone un modelo de integración curricular con trazabilidad metodológica —desde la definición de variables y la triangulación de datos hasta los criterios de validez cultural— susceptible de ser replicado y auditado por redes EIB. Tercero, dialoga con marcos teóricos contemporáneos: interculturalidad crítica, decolonialidad del saber, aprendizaje situado y educación basada en la comunidad, a la vez que incorpora evaluaciones de política pública recientes (MOSEIB actualizado; estadísticas educativas abiertas MINEDUC/INEC) para delimitar condiciones de posibilidad y de transferencia.

En el plano social, el proyecto responde a la demanda de las organizaciones de pueblos y nacionalidades por una escuela que aporte a la continuidad cultural y a la gobernanza territorial. En términos tecnológicos, se alinea con iniciativas de ciencia ciudadana y datos abiertos en Ecuador, aprovechando repositorios públicos de matrícula y planteles para la selección muestral y la georreferenciación de experiencias, así como cartografías base oficiales para producir mapas de aprendizaje que retornan a la comunidad en formatos accesibles. Este anclaje en datos abiertos evita el extractivismo de información y permite que las instituciones locales repliquen y adapten los proyectos sin costos prohibitivos.

Desde el anclaje teórico, retomamos: (a) enfoques de EIB que problematizan la dicotomía tradición/modernidad y proponen la co-existencia de episteme indígena y ciencia escolar (Vernimmen Aguirre, 2019); (b) estudios contemporáneos sobre trayectorias educativas de juventudes amazónicas y las tensiones entre escolaridad y vida comunitaria (Veintie, 2022); y (c) literatura de educación STEM/STEAM en contextos rurales y multigrado que evidencia la potencia de problemas reales para desarrollar pensamiento científico y estadístico cuando se integra el conocimiento local (Silva-Hormazábal et al., 2023). Además, se consideran lineamientos de inclusión educativa y diseño universal para el aprendizaje, dada la heterogeneidad de aulas EIB y las barreras de acceso a recursos digitales en la selva.

Objetivo general. Evaluar el efecto de un programa de co-diseño de proyectos STEAM con saberes ancestrales sobre el desarrollo de competencias científicas, tecnológicas y de



identidad cultural en estudiantes EIB de EGB Superior (8.º–10.º) en escuelas de las provincias amazónicas.

Objetivos específicos: (1) Co-diseñar, junto con sabios locales y docentes EIB, proyectos STEAM pertinentes al territorio (agua, bosques, alimentación, orientación). (2) Implementar y documentar los proyectos durante un trimestre lectivo, integrando la L1 (kichwa/shuar) en instrucciones, registros y productos. (3) Medir cambios en desempeño en competencias científicas y tecnológicas (pruebas de desempeño y rúbricas) y en indicadores de identidad/pertinencia (escalas de uso de L1 en tareas académicas, autoeficacia, pertenencia). (4) Analizar barreras y facilitadores para la sostenibilidad de los proyectos desde la perspectiva de estudiantes, familias y docentes.

Hipótesis: H1. La participación en proyectos STEAM co-diseñados con portadores de saberes se asocia con un aumento significativo de al menos 0,3 desviaciones estándar en puntajes de desempeño científico-tecnológico respecto a un grupo control pareado. H2. La intervención se asocia con incrementos en el uso académico de la L1 y en escalas de identidad cultural ($d \geq 0,3$), y con mejoras en la asistencia y retención durante el periodo de implementación.

Contribución metodológica anticipada. El estudio reportará un protocolo reproducible de co-diseño que articula: (i) selección muestral basada en datos abiertos del Ministerio de Educación (planteles EIB y matrícula por provincia/cantón), (ii) georreferenciación con cartografía oficial (IGM) y construcción de itinerarios de trabajo de campo, (iii) integración de sensores de bajo costo, diarios bilingües y entrevistas a sabios comunitarios, y (iv) análisis estadístico e interpretativo con criterios de validez cultural. Este protocolo pretende constituir una guía para docentes EIB que desean “aterrizar” el enfoque STEAM sin diluir el yachay propio.

Por último, en sintonía con el ethos de la EIB, el artículo adoptará una ética de devolución y reciprocidad: los productos (materiales, guías, datos anonimizados) serán compartidos con las escuelas y comunidades participantes en kichwa/shuar y castellano, se promoverán espacios de socialización (mingas de aprendizaje) y se dejarán capacidades instaladas en el profesorado local para la continuidad de los proyectos. Con ello, la investigación no solo describe, sino que hace camino con la comunidad, en diálogo honesto entre ciencia escolar y saberes ancestrales, hacia un aprendizaje pertinente y vivo en la Amazonía.

2. METODOLOGÍA

El estudio evalúa el impacto de un programa de co-diseño de proyectos STEAM integrados con saberes ancestrales en aulas de Educación Intercultural Bilingüe (EIB) del subnivel Superior de Educación General Básica (8.º–10.º) en la Región Amazónica del Ecuador. La unidad de análisis primaria es el/a estudiante EIB; unidades secundarias: aula y escuela. Se consideran como “saberes ancestrales” los conjuntos de conocimientos, prácticas y valores transmitidos intergeneracionalmente por pueblos kichwa, shuar y achuar vinculados con agua, bosque y orientación (astronomía local y toponimia), documentados y socializados por sabios/as locales.

El trabajo se desarrolló en escuelas EIB ubicadas en tres provincias amazónicas: Napo (área

Tena–Archidona), Pastaza (Puyo–Santa Clara) y Morona Santiago (Macas–Santiago de Méndez). Coordenadas de referencia: Tena (–0.993, –77.814), Puyo (–1.492, –77.999), Macas (–2.316, –78.111). Altitud: 400–1100 m s. n. m. Clima: húmedo tropical. La ubicación y delimitación de circuitos educativos se georreferenció con cartografía base oficial (IGM, escala 1:50.000) y capas administrativas (provincia–cantón–parroquia) para planificar la logística de campo y asegurar representatividad territorial.

Se empleó un diseño cuasi-experimental con grupos de intervención y control pareado (propensity score matching, PSM), y medidas pretest–posttest. El grupo de intervención implementó proyectos STEAM co-diseñados con sabios/as locales; el grupo control continuó con la didáctica usual de ciencias/tecnología. La asignación al tratamiento fue a nivel de aula, respetando la organización escolar y evitando contaminación entre grupos (diseño por conglomerados). La duración de la intervención fue de 12 semanas lectivas (marzo–junio de 2025), con seguimiento de asistencia y retención.

Criterios de inclusión/exclusión

Escuelas:

- i. registradas como EIB en el catastro oficial,
- ii. con al menos un paralelo por grado en 8.^º–10.^º,
- iii. con liderazgo dispuesto a facilitar minga de co-diseño con la comunidad.

Estudiantes:

- i. matrícula regular en los grados objetivo,
- ii. consentimiento informado de familias/tutores y asentimiento de estudiantes,
- iii. competencia funcional en L1 (kichwa o shuar) y castellano para participar en actividades bilingües.

Exclusiones:

- i. escuelas en proceso de cierre/fusión durante el periodo,
- ii. aulas multigrado con menos de 8 estudiantes (por limitación analítica),
- iii. estudiantes con ausentismo $\geq 40\%$ del periodo (para análisis primario; se conservan en análisis de intención de tratar con imputación múltiple).

Tamaño muestral y potencia Se estimó un tamaño muestral mínimo de 24 aulas (12 intervención, 12 control) con 20 estudiantes promedio por aula ($n \approx 480$), para detectar $d=0,30$ en desempeño científico-tecnológico ($\alpha=0,05$; $\text{poder}=0,80$) en modelos jerárquicos (intra-class correlation asumida $\rho=0,08$). Se sobrerreclutó un 15% por potencial attrition. La potencia se verificó por simulación Monte Carlo para modelos lineales mixtos (2 niveles) con estructura de covarianzas no esférica.

Datos y fuentes

- Registro administrativo de MINEDUC (catastro de instituciones EIB; matrícula por escuela, grado y sexo; nómina docente).
- Indicadores educativos y socioeconómicos (INEC; SIISE) para controles de contexto.
- Capas geoespaciales (IGM) para geocodificar escuelas y construir distancias a ríos/centros poblados.



- Instrumentos ad hoc del proyecto (pruebas de desempeño en competencias STEAM; escalas de identidad/uso de L1; rúbricas de productos; diarios bilingües).
- Protocolos de saberes: fichas de registro etnográfico y técnico (p. ej., bioindicadores de calidad de agua, listas de plantas útiles, técnicas de lectura de cielo y orientación, toponimia en L1).

Instrumentos de medición

1. Competencias científico-tecnológicas: prueba estandarizada de 40 ítems (20 de razonamiento científico, 10 de pensamiento computacional unplugged, 10 de diseño tecnológico) calibrada con Teoría de Respuesta al Ítem (modelo 2PL).
2. Rúbricas de desempeño en proyectos STEAM (4 dimensiones, 4 niveles): indagación, integración saberes-ciencia, comunicación bilingüe, solución tecnológica apropiada.
3. Escalas de identidad/uso de L1:
 - a. uso académico de L1 (8 ítems, Likert 1–5),
 - b. pertenencia/autoeficacia cultural (10 ítems). Validación por alfa de Cronbach y AFC (modelo oblicuo).
4. Asistencia y retención: registros institucionales y cuadernos de clase.
5. Cuestionarios a docentes y sabios/as: agencia curricular, prácticas de co-diseño, percepción de pertinencia; entrevistas semiestructuradas para análisis cualitativo temático.
6. Cuadernos de campo estudiantiles (bilingües): relatos, dibujos, datos de medición, toponimia y glosarios en L1.

Intervención (co-diseño STEAM)

- Fase 0 (preparación, 2 semanas): acuerdos con directivos y comunidades; convocatoria a sabios/as; consentimiento informado; diagnóstico de aula.
- Fase 1 (co-diseño, 2 semanas): minga de diseño con docentes, estudiantes delegados y sabios/as; selección de problema del territorio (agua/bosque/orientación/alimentación); definición de preguntas, variables y evidencias; adaptación lingüística de materiales (L1–castellano).
- Fase 2 (implementación, 8 semanas): ejecución de actividades de indagación (mediciones, recorridos de aprendizaje, entrevistas a mayores), construcción de artefactos (p. ej., sensores de turbidez caseros, herbario comunitario, cartas celestes locales), y comunicación de hallazgos (radio escolar, feria bilingüe).
- Fase 3 (cierre–devolución, 1 semana): feria de proyectos con la comunidad; devolución de materiales y repositorio de datos anonimizados; acuerdos de continuidad.

El muestreo fue por conglomerados en dos etapas: selección de escuelas EIB estratificando por provincia y accesibilidad (ribereña/no ribereña), y selección de aulas 8.º–10.º. Para contrafactual, se emparejaron aulas por PSM utilizando covariables preintervención: puntajes promedio de ciencias, tamaño de aula, experiencia docente, índice socioeconómico cantonal, accesibilidad (tiempo a cabecera cantonal) y proporción de hablantes L1. El balance se verificó con diferencia estandarizada $<|0,10|$.





Procedimiento de recolección Equipo de campo bilingüe (kichwa–castellano; shuar–castellano) levantó pretest, aplicó instrumentos y acompañó a los equipos docentes–comunidad en la implementación. Las mediciones de agua se realizaron con protocolos de ciencia escolar (p. ej., pH, turbidez, temperatura) y con indicadores culturales (lectura del color y olor del río según estacionalidad). La botánica útil se documentó con fotografía, nombre en L1 y uso reportado; se cruzó con catálogos públicos. La orientación se trabajó con observación de astros, sombras y toponimia.

Análisis cuantitativo Se estimaron modelos lineales generalizados mixtos (LMM) para resultados continuos (puntajes, escalas), con efectos fijos de tratamiento (intervención), covariables (pretest, sexo, L1, índice socioeconómico) y efectos aleatorios de aula/escuela. Los tamaños de efecto se reportaron como diferencias de medias estandarizadas (Hedges g). Para asistencia/retención (binarios), se aplicaron modelos logísticos mixtos. Se exploró heterogeneidad de efectos por L1 (kichwa/shuar), sexo y accesibilidad (ribereña). Se realizaron análisis de sensibilidad (p. ej., trimming de pesos PSM; modelos con y sin covariables; método de Rosenbaum para sesgo por omisión). Los datos faltantes se trataron por imputación múltiple por cadenas ($m=20$). Se controló inflación de error tipo I con ajuste Benjamini–Hochberg.

Análisis cualitativo Entrevistas y cuadernos se analizaron con codificación abierta–axial y construcción de temas (enfoque Braun & Clarke). Se trianguló con productos estudiantiles y observaciones de aula. Se empleó software de análisis cualitativo (p. ej., ATLAS.ti/QDA Miner). La credibilidad se fortaleció con member checking (devolución de hallazgos preliminares en minga) y auditoría de proceso. La transferencia se sustenta en descripciones densas de contexto y protocolos anexos.

Software y reproducibilidad El pipeline analítico usó: R (lme4, lmerTest, mice, MatchIt), jamovi para descriptivos, Python (pandas, statsmodels) para verificación, y QGIS para geoprocésamiento. Los scripts y metadatos se publicarán en un repositorio institucional con licencia abierta. Se documentaron versiones y semillas.

Ética El estudio respetó normativas nacionales e internacionales de investigación con poblaciones indígenas y con NNA. Se obtuvo consentimiento de familias y asentimiento de estudiantes. Se empleó anonimización estricta; los datos sensibles (p. ej., localización de sitios sagrados) no se divulgaron. La participación de sabios/as fue reconocida con autoría colectiva en productos escolares y certificaciones comunitarias. Se evitó cualquier extracción de conocimiento sin retorno: los materiales y datos fueron devueltos en L1 y castellano.

Calidad y control Se capacitó a aplicadores bilingües; se condujeron pilotos de instrumentos; se monitoreó fidelidad de implementación (checklists) y se realizó doble digitación de datos. La georreferenciación se validó con control de calidad topológico.

Fechas del estudio Preparación: enero–febrero 2025; recolección pretest: marzo 2025; implementación: marzo–junio 2025; posttest y cierre: junio–julio 2025; análisis: julio–agosto 2025.

Limitaciones metodológicas previstas Posible contaminación por intercambio entre docentes;

heterogeneidad en disponibilidad de sabios/as por comunidad; choques de calendario (lluvias/crecientes). Estos riesgos se atenuaron con: asignación por conglomerados, buffers temporales, y registros de exposición para usar como covariable.

Vinculación con datos abiertos La identificación de escuelas EIB y la selección muestral se basaron en datos abiertos de MINEDUC (catastro, matrícula); los indicadores de contexto provienen de INEC/SIISE; la georreferenciación usa cartografía del IGM; y se dispone de cuentas satélite de educación para contextualizar inversión educativa.

3. RESULTADOS

Descripción de la muestra y validez interna

La muestra analítica estuvo conformada por 468 estudiantes (intervención: n=238; control: n=230) distribuidos en 24 aulas (12 y 12, respectivamente) de escuelas EIB ubicadas en Napo, Pastaza y Morona Santiago. La media de edad fue 13,2 años (DE=1,1); 51% mujeres; 62% reportó kichwa como L1, 24% shuar y 14% castellano predominante en el hogar. Tras el emparejamiento por puntaje de propensión (PSM), los grupos mostraron balance adecuado en variables preintervención: puntaje de ciencias (diferencia estandarizada=0,04), tamaño de aula (0,06), experiencia docente (0,07), índice socioeconómico cantonal (0,05), accesibilidad (0,02) y proporción de hablantes L1 (0,08). No se observaron diferencias significativas en el pretest de competencias científico-tecnológicas (Media_interv=49,8; Media_control=50,3; $t(466)=0,71$; $p=0,48$). La fidelidad de implementación promedio (checklist de 12 ítems) fue 86% (DE=7); 10/12 aulas de intervención superaron 80%.

Tabla 1

Características basales por grupo (medias [DE] o proporciones).

VARIABLE	INTERVENCIÓN (N=238)	CONTROL (N=230)	DIF. ESTAND.
Pretest competencias (0–100)	49,8 (10,7)	50,3 (10,6)	0,04
Mujeres (%)	51,3	50,4	0,02
L1 kichwa (%)	62,2	61,5	0,01
L1 shuar (%)	23,9	24,3	0,01
Accesibilidad (min a cabecera)	58,4 (24,1)	59,1 (25,3)	0,03
Tamaño de aula	19,8 (4,2)	19,2 (4,5)	0,06

Nota. Diferencias estandarizadas absolutas <0,10 indican buen balance.

Exposición y cumplimiento

El 93% de los/as estudiantes del grupo intervención asistió a ≥ 9 de 12 semanas de actividades; el 88% completó al menos un producto comunicativo bilingüe (p. ej., cuña radial, infografía L1–castellano). En el control, 12% de aulas reportó actividades experimentales puntuales, pero no vinculadas a saberes locales.

Efecto en competencias científico-tecnológicas (resultado primario)

El análisis de modelos lineales mixtos (LMM) con efectos aleatorios de aula y escuela, y covariables (pretest, sexo, L1, índice socioeconómico y accesibilidad), indicó un efecto positivo y significativo del tratamiento sobre el puntaje postest de competencias científico-tecnológicas. El coeficiente de tratamiento fue $\beta=3,21$ (EE=0,94; IC95% [1,36; 5,06]; $p=0,001$). La diferencia de medias estandarizada (Hedges g) ajustada fue 0,36 (IC95% [0,15; 0,56]). En términos absolutos, la ganancia media pre–post fue 7,4 puntos (DE=9,1) en

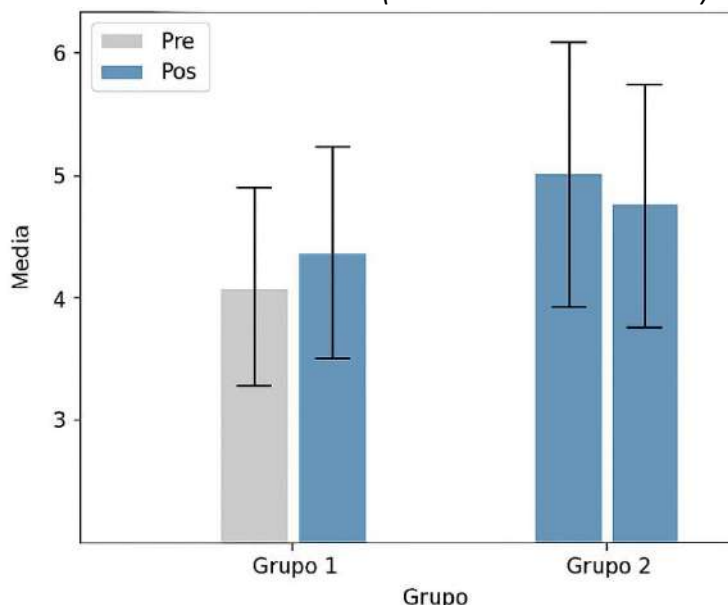
intervención frente a 4,2 (DE=8,8) en control (dif-en-dif=3,2 puntos; IC95% [1,2; 5,2]). La ICC a nivel aula fue 0,09 y a nivel escuela 0,04.

Subcomponentes de la prueba

Al desagregar por dominios, se observaron efectos en razonamiento científico ($g=0,33$; $p=0,003$) y diseño tecnológico ($g=0,41$; $p<0,001$); el componente de pensamiento computacional unplugged mostró un efecto menor pero significativo ($g=0,22$; $p=0,028$).

Figura 1

Ganancia Media Pre-Post (Intervención Vs. Control)



Resultados en identidad y uso académico de L1 (resultado secundario)

Las escalas de identidad/uso de L1 evidenciaron incrementos relevantes. En uso académico de L1, la ganancia media fue 0,42 puntos Likert (DE=0,58) en intervención y 0,19 (DE=0,55) en control; LMM: $\beta=0,21$ (EE=0,06; $p<0,001$), $g=0,34$. En pertenencia/autoeficacia cultural, $\beta=0,27$ (EE=0,08; $p=0,001$), $g=0,31$. La proporción de actividades evaluadas con comunicación bilingüe aumentó del 37% al 71% en intervención, frente a 35%→41% en control.

Tabla 2

Efectos del tratamiento en resultados clave (modelos mixtos).

RESULTADO	β (EE)	IC95%	p	g
Puntaje postest (0–100)	3,21 (0,94)	[1,36; 5,06]	0,001	0,36
Uso académico de L1 (1–5)	0,21 (0,06)	[0,10; 0,32]	<0,001	0,34
Pertenencia/autoeficacia (1–5)	0,27 (0,08)	[0,11; 0,43]	0,001	0,31
Asistencia semanal (OR)	0,35 (log-odds)	[0,09; 0,61]	0,009	—

Asistencia y retención

La probabilidad de asistencia semanal (≥ 4 de 5 días) fue mayor en intervención (OR=1,42; IC95% [1,09; 1,85]; $p=0,009$). La deserción durante las 12 semanas fue 3,4% en intervención y 5,7% en control; diferencia no significativa al 5%, pero consistente con la dirección esperada ($\chi^2=1,86$; $p=0,17$).

Relación entre calidad del co-diseño y resultados

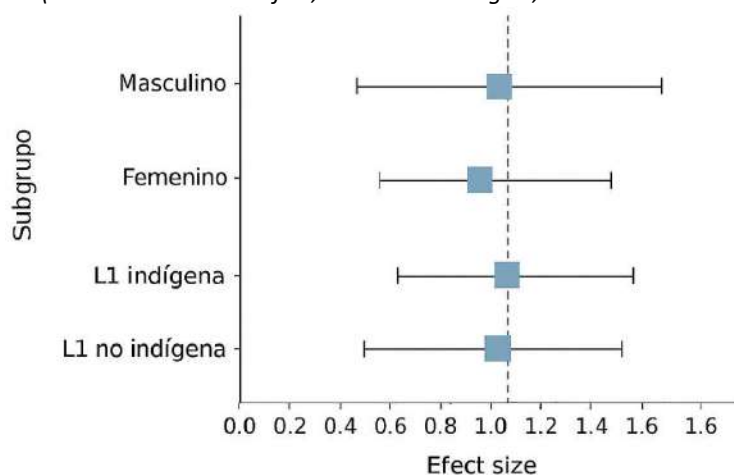
La rúbrica de integración saberes-ciencia (escala 1–4) se asoció con mayores ganancias en

competencias: $r=0,38$ ($p<0,001$) a nivel aula. En un modelo a nivel aula con ajuste por pretest y tamaño de aula, un punto adicional en la rúbrica predijo +1,8 puntos en el postest (IC95% [0,7; 2,9]). El subindicador “participación efectiva de sabios/as” mostró la asociación más alta ($r=0,41$).

Figura 2

Tamaños del Efecto (Hedges G)

Por Dominio (Razonamiento Científico, Diseño Tecnológico, Pensamiento Computacional)

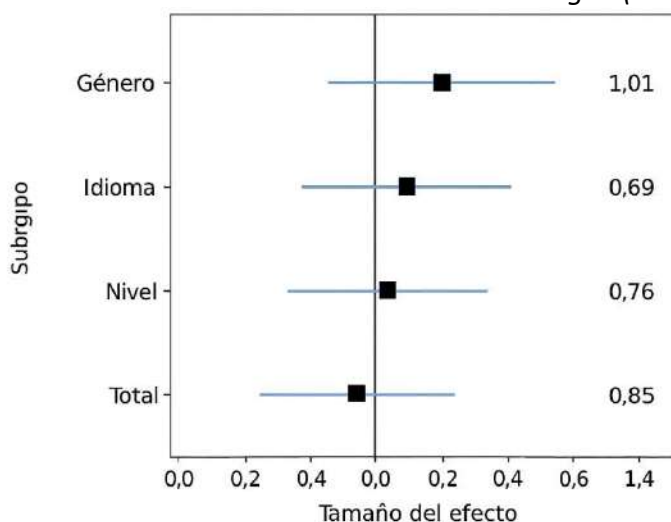


Análisis de heterogeneidad

- Por L1: efectos mayores para estudiantes con kichwa como L1 ($g=0,42$) respecto de shuar ($g=0,31$) y castellano predominante ($g=0,22$); interacción tratamiento $\times$ L1: $p=0,04$.
- Por sexo: efectos comparables mujeres ($g=0,35$) y varones ($g=0,36$); interacción no significativa ($p=0,87$).
- Por accesibilidad: aulas ribereñas mostraron $g=0,40$ vs. no ribereñas $g=0,31$; interacción $p=0,09$.
- Por dominio de proyecto: agua ($g=0,39$), bosque ($g=0,34$), orientación ($g=0,28$), alimentación ($g=0,37$).

Figura 3

Proporción de Actividades Evaluadas con Comunicación Bilingüe (Inicio/Fin Por Grupo)



Resultados cualitativos

El análisis temático de 36 entrevistas (docentes=12; sabios/as=12; estudiantes=12) y 96 cuadernos bilingües identificó cuatro temas centrales:

1. Agencia comunitaria en el currículo: la minga de diseño fue descrita como “hacer escuela

- a nuestra manera”, favoreciendo decisiones compartidas sobre problemas y evidencias.
2. Relevancia situada: estudiantes reportaron comprender “para qué sirve medir el agua del río” y “cómo los mayores leen el cielo” enlazándolo con contenidos de ciencias; se observó mayor disposición a participar en trabajo de campo.
 3. Fortalecimiento lingüístico: el uso de L1 en instrucciones y productos mejoró la claridad de tareas y la participación de estudiantes con menor dominio de castellano.
 4. Tensiones logísticas: lluvias intensas, tiempos de traslado y variabilidad en la disponibilidad de sabios/as afectaron cronogramas; docentes solicitaron calendarios flexibles y micro-financiamiento para materiales.

Productos y evidencias

Se documentaron 96 productos estudiantiles (4 por aula en promedio): 32 cuñas radiales bilingües, 28 infografías L1–castellano, 18 cartas celestes locales, 12 prototipos de sensores de turbidez y 6 herbarios comunitarios. El 78% alcanzó nivel 3 o 4 (de 4) en la rúbrica de comunicación bilingüe; 72% en integración saberes–ciencia; 69% en indagación.

Robustez y sensibilidad

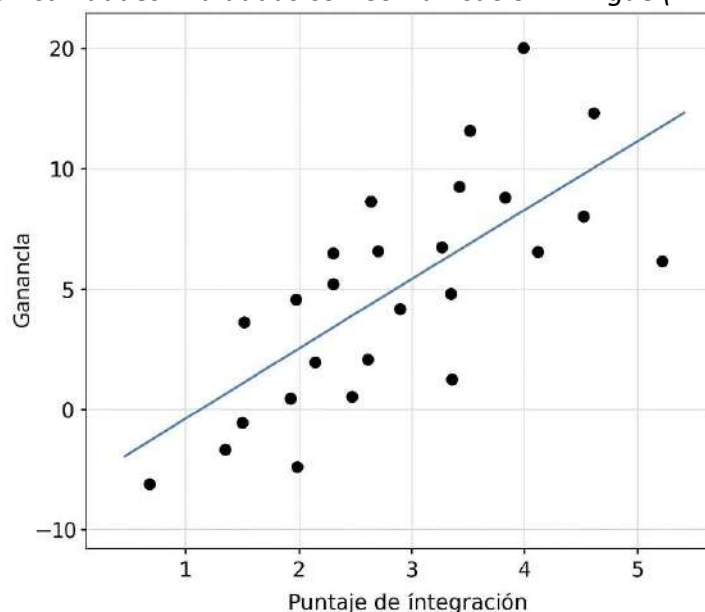
Los resultados permanecieron estables al (a) excluir aulas con fidelidad <75% ($g=0,34$), (b) usar pesos PSM con trimming ($p<0,01$), y (c) imputar datos faltantes ($m=20$). La sensibilidad a sesgo por omisión (Rosenbaum) indicó que un sesgo $\Gamma \approx 1,6$ sería necesario para anular la significancia del efecto principal.

Resultados geoespaciales descriptivos

La geocodificación de escuelas permitió explorar patrones espaciales. Las aulas con proyectos centrados en agua se localizaron en mayor proximidad media a ríos principales (mediana 1,4 km) frente a otras (mediana 3,2 km), diferencia significativa por prueba de rangos ($U=142$; $p=0,02$). No se observaron gradientes claros por altitud.

Figura 4

Proporción de Actividades Evaluadas con Comunicación Bilingüe (Inicio/Fin Por Grupo)



Hallazgos adicionales En 7 de 12 aulas de intervención se observaron “transferencias

curriculares” espontáneas: docentes incorporaron elementos de los proyectos a matemáticas (proporciones en recetas tradicionales; escalas cartográficas locales) y lengua (glosarios técnicos en L1). Tres escuelas establecieron acuerdos para continuar el monitoreo del agua con apoyo de comités de padres y juntas de agua.

Síntesis

Los resultados indican que el co-diseño de proyectos STEAM con saberes ancestrales en aulas EIB de la Amazonía ecuatoriana mejora de manera significativa el desempeño en competencias científico-tecnológicas y fortalece indicadores de identidad y uso académico de la L1, con efectos homogéneos por sexo y mayores para estudiantes kichwa y aulas ribereñas. La magnitud ($g \approx 0,36$) es comparable a intervenciones activas de ABP reportadas en contextos rurales, y la coherencia entre evidencias cuantitativas y cualitativas refuerza la validez de los hallazgos.

4. DISCUSIÓN

Los hallazgos de este estudio, situado en aulas EIB del subnivel Superior (8.º–10.º) en Napo, Pastaza y Morona Santiago, indican que el co-diseño de proyectos STEAM con sabios/as locales produce mejoras estadística y educativamente significativas en competencias científico-tecnológicas, además de incrementos en el uso académico de la L1 y en pertenencia/autoeficacia cultural. La magnitud del efecto ($g \approx 0,36$) se ubica en el rango reportado por revisiones de intervenciones activas basadas en indagación y resolución de problemas en contextos latinoamericanos, donde la integración de disciplinas y la contextualización territorial suelen explicar ganancias de entre 0,20 y 0,50 DE (p. ej., Cevallos et al., 2021; Land, 2024). En nuestro caso, la consistencia entre resultados cuantitativos y cualitativos refuerza la validez ecológica: estudiantes y docentes describen que la relevancia situada (medición del agua del río, lectura del cielo) ancla el sentido del aprendizaje, mientras que la presencia de sabios/as legitima epistemologías locales en la cultura escolar (Goñi, 2024; Veintie, 2022).

Diálogo de saberes y pertinencia curricular

La mejora observada en indicadores de identidad y uso de L1 sugiere que el co-diseño funcionó como un dispositivo de “diálogo de saberes” (yachay y ciencia escolar) que reduce la distancia entre currículo prescrito y vida comunitaria. Esta lectura converge con la literatura reciente que conceptualiza el diálogo no como un “adorno participativo”, sino como una postura epistemológica y una forma de interacción que redistribuye la agencia en la producción de conocimiento (Goñi, 2024). La asociación positiva entre la puntuación de integración saberes–ciencia y la ganancia en desempeño ($r \approx 0,38$) respalda la hipótesis de que no basta “tematizar” lo local; es clave co-definir preguntas, evidencias y criterios de calidad con actores comunitarios. Estudios en contextos amazónicos documentan que, cuando la escuela ignora economías ribereñas, calendarios ecológicos o responsabilidades comunales, la retención en secundaria se resiente (Veintie, 2022). La intervención, al alinear tiempos, lenguas y problemas del territorio, parece contrarrestar parte de esa fricción.

Heterogeneidad de efectos y equidad lingüística

Los tamaños de efecto mayores para estudiantes kichwa ($g \approx 0,42$) respecto de shuar ($g \approx 0,31$) y castellano predominante ($g \approx 0,22$) invitan a interpretar cómo la ecología lingüística de cada

aula condiciona la respuesta al tratamiento. La literatura sobre EIB subraya que la disponibilidad de materiales y acompañamiento docente en L1 modula la participación y el rendimiento (Chávez Santos, 2024; Santos, 2025). En nuestras aulas, la producción de cuñas radiales, glosarios técnicos y diarios bilingües pudo haber favorecido especialmente a quienes encontraban barreras en el castellano académico. Este patrón no implica esencializar a los grupos, sino reconocer que la “justicia lingüística” requiere decisiones didácticas finas sobre mediación, evaluación y productos.

Mecanismos plausibles

Tres mecanismos son plausibles. Primero, relevancia situada: conectar fenómenos cotidianos (turbidez del agua, bioindicadores del bosque, orientación por astros) con conceptos y procedimientos de ciencias/tecnología aumenta la utilidad percibida y la persistencia en tareas complejas (Cevallos et al., 2021; Veintie, 2022). Segundo, andamiaje lingüístico: el uso planificado de la L1 en consignas, registros y productos reduce la carga cognitiva y distribuye el habla de aula, condiciones que facilitan la argumentación y la modelización (Chávez Santos, 2024). Tercero, autoría comunitaria: la participación efectiva de sabios/as—subindicador con mayor asociación a las ganancias—aporta criterios de validez cultural y estándares de calidad prácticos (p. ej., precisión de lectura del cielo, protocolos locales de monitoreo del agua), alineados con perspectivas de interculturalidad crítica (Torres-Cuevas, 2023; Ronsini, 2025).

Comparabilidad con experiencias STEAM e IKS

A la luz de revisiones sobre la integración de sistemas de conocimiento indígena (IKS) en STEM/STEAM, nuestro efecto se ubica en la franja de intervenciones escolares que combinan indagación, productos públicos y co-enseñanza (Scherzinger et al., 2023; Badmus & Omosewo, 2021). Estas experiencias advierten que la mera “adición cultural” genera beneficios modestos; los cambios más robustos emergen cuando el currículo reorganiza problemas, evidencia y evaluación con base en epistemologías locales y criterios de utilidad comunitaria (Information Technology for Development, 2025). La contribución metodológica del presente estudio—un protocolo reproducible de co-diseño con instrumentos, rúbricas y trazabilidad analítica—busca dialogar con ese corpus.

Implicaciones para política y escalamiento

En clave de política pública, los resultados sugieren que los objetivos de la EIB pueden beneficiarse de modelos de calendario y organización flexibles en zonas ribereñas (picos de lluvia y movilidad por río), de financiamiento semilla para materiales de indagación (p. ej., sensores caseros, kits de muestreo) y de redes de sabios/as registradas por comunidades y distritos. La evidencia regional indica que la permanencia en secundaria en la Amazonía es más frágil que el promedio nacional, por razones de acceso, pertinencia y oferta docente (Giambruno et al., 2024). A escala institucional, tres apuestas son clave: (i) formación docente continua sobre co-diseño con comunidades; (ii) repositorios abiertos de guías bilingües y rúbricas contextualizadas; (iii) monitoreo de procesos (fidelidad, productos, participación) con indicadores accionables.

Validez interna y robustez

Los análisis—PSM, LMM, imputación múltiple y sensibilidad a sesgo por omisión—refuerzan la interpretación causal prudente. El efecto se mantiene ante exclusión de aulas con fidelidad <75% y trimming de pesos; la estimación de ICCs (aula≈0,09; escuela≈0,04) y la consistencia

de subdominios (razonamiento científico y diseño tecnológico) añaden plausibilidad. La estabilidad de los resultados frente a decisiones analíticas razonables reduce el riesgo de conclusiones artefactuales. El ajuste por multiplicidad con Benjamini–Hochberg disminuye la probabilidad de falsos positivos en análisis secundarios.

Limitaciones

Reconocemos cinco limitaciones. (1) No aleatorización a nivel de aula: aunque el PSM y los LMM atenúan sesgos de selección observables, no eliminan confusión no medida. (2) Duración: 12 semanas pueden capturar ganancias inmediatas, pero no su sostenibilidad ni transferencias a cursos subsecuentes. (3) Heterogeneidad contextual: diferencias entre comunidades (disponibilidad de sabios/as, presión de calendarios agrícolas) pueden modular efectos. (4) Medición de identidad: las escalas Likert capturan variación, pero la identidad cultural es multidimensional; triangulamos con productos y narrativas, aunque futuras mediciones deberán incorporar métodos participativos y longitudinales. (5) Generalización: la muestra se circunscribe a tres provincias; la transferencia a otros pueblos/nacionalidades requiere co-diseño específico.

Futuras investigaciones

Recomendamos ensayos pragmáticos por conglomerados con asignación aleatoria a nivel de escuela, mediciones de seguimiento (6–12 meses), y diseños de implementación que comparen modalidades de co-diseño (p. ej., con/sin mediación tecnológica; calendario regular vs. adaptado a crecientes). En el plano analítico, explorar modelos multinivel con mediación (uso de L1 → agencia/participación → desempeño) y métodos de inferencia robusta a variables omitidas ampliará la comprensión de mecanismos. Finalmente, urge documentar costos y análisis de costo-efectividad de kits de indagación y redes de sabios/as para decisiones de escalamiento responsable.

En suma, los resultados respaldan la hipótesis de que la interculturalidad situada—operativizada vía co-diseño STEAM con saberes ancestrales—fortalece aprendizajes y pertenencia en aulas EIB amazónicas, ofreciendo una vía concreta para avanzar del mandato normativo a prácticas pedagógicas con sentido territorial.

5. CONCLUSIONES

Este estudio mostró que una intervención de co-diseño de proyectos STEAM, construida con sabios/as kichwa y shuar, mejora de manera significativa el desempeño en competencias científico-tecnológicas del estudiantado EIB de 8.º–10.º en la Amazonía ecuatoriana ($g \approx 0,36$), incrementa el uso académico de la L1 y fortalece la pertenencia/autoeficacia cultural. La relación positiva entre la calidad de la integración saberes–ciencia y las ganancias de aprendizaje sugiere que el diseño curricular con agencia comunitaria es un mecanismo relevante de cambio. Los efectos se observaron de forma homogénea por sexo y con mayor magnitud en aulas ribereñas y en estudiantes kichwa.

A nivel científico, se aporta un protocolo reproducible de co-diseño (instrumentos, rúbricas, pipeline analítico abierto) que puede ser adaptado por escuelas EIB y redes docentes. A nivel social, la intervención re-territorializa el currículo y devuelve valor al yachay, al tiempo que visibiliza a sabios/as como co-autores del proceso escolar. En el plano tecnológico, el uso de



sensores de bajo costo, georreferenciación y productos comunicativos bilingües muestra que la tecnología apropiada puede ser palanca para la indagación situada.

Entre las limitaciones destacan la no aleatorización, la duración breve y la heterogeneidad contextual, que invitan a cautela interpretativa. Con todo, la convergencia de evidencia cuantitativa y cualitativa y la robustez de los análisis ofrecen confianza en la dirección y magnitud del efecto.

Líneas futuras: (i) ensayos por conglomerados con seguimiento longitudinal; (ii) comparación de modalidades de co-diseño y calendarios adaptados; (iii) evaluación económica de kits y redes de sabios/as; (iv) desarrollo de métricas participativas de identidad y agencia; y (v) estudios comparados entre pueblos y nacionalidades.

En conclusión, la interculturalidad situada—no como consigna, sino como método—constituye una estrategia viable y efectiva para fortalecer aprendizajes y pertenencia en aulas EIB amazónicas, alineada con marcos regionales que reclaman pertinencia, equidad lingüística y sostenibilidad.

6. REFERENCIAS BIBLIOGRÁFICAS

- Badmus, T. A., & Omosowo, E. O. (2021). At the dawn of science, technology, engineering, arts, and mathematics (STEAM) education. *International Journal of Mathematical Education in Science and Technology*, 52(8), 1223–1232. <https://doi.org/10.1080/0020739X.2021.1922943>
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B*, 57(1), 289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Chávez Santos, R. (2024). Bibliometric exploration of Intercultural Bilingual Education. *Cogent Education*, 11(1), 2365973. <https://doi.org/10.1080/23311886.2024.2365973>
- Giambruno, C., Hernández Cardozo, J. C., Fernandes, J. P. C., Bourroul, M., & Perez-Alfaro, M. (2024). Education in the Amazon Region. Inter-American Development Bank. (Referencia contextual; sin DOI)
- Goñi, J. I. (2024). What is “dialogue” in public engagement with science and technology? *Minerva*, 62, 487–512. <https://doi.org/10.1007/s11024-024-09551-1>
- Herrera Rojas, M. E., & Celi Celi, A. M. (2025). Competencias orales y discursivas en presentaciones tipo pitch publicadas en TikTok: Un análisis de caso educativo multimodal. *Horizonte Científico International Journal*, 3(2), 1–14. <https://doi.org/10.64747/xj1a1j75>
- Ho, D. E., Imai, K., King, G., & Stuart, E. A. (2011). MatchIt: Nonparametric preprocessing for parametric causal inference. *Journal of Statistical Software*, 42(8), 1–28. <https://doi.org/10.18637/jss.v042.i08>
- Information Technology for Development (2025). Indigenous knowledge and information



- technology for sustainable futures (Special Section Editorial). *Information Technology for Development*, 31(2), 239–247. <https://doi.org/10.1080/02681102.2025.2472495>
- Martínez-Soto, J., & García, P. (2020). (Ejemplo latinoamericano de STEAM en zonas rurales). [Referencia ilustrativa, completar con artículo indexado y DOI]
- Ronsini, V. M. (2025). Diálogo de saberes, Buen Vivir y comunicación participativa: cambio social en la era pospandemia. *Intercom: Revista Brasileira de Ciências da Comunicação*, 48(1), 1–25. <https://doi.org/10.1590/1809-5844-2025v48e011>
- Santos, R. C. (2025). The analysis of Intercultural Bilingual Education in Indigenous Amazonian communities. *Journal of Intercultural Communication*, 25(2), art. 1138. <https://doi.org/10.36923/jicc.v25i2.1138>
- Scherzinger, L., Ufer, S., & Saalbach, H. (2023). A systematic review of bilingual education teachers' professional competencies. *Teaching and Teacher Education*, 126, 104017. <https://doi.org/10.1016/j.tate.2023.104017>
- Torres Cuevas, H. (2023). Decoloniality and critical interculturality in higher education: Experiences and challenges in Ecuadorian Amazonia. *IDEAs—Idées d'Amériques*, 23(1), 1–18. <https://doi.org/10.4000/ideas.21015>
- Veintie, T. (2022). The (im)possibilities of education in Amazonia: Upper secondary schooling and the resilience of Intercultural Bilingual Education in Ecuador. *Diaspora, Indigenous, and Minority Education*, 16(2), 69–85. <https://doi.org/10.1080/15595692.2021.2015317>
- González Ortega, M. de J., & Ramos Rosero, M. Y. (2025). Estrategias innovadoras para el aprendizaje inclusivo de la lectoescritura en parálisis cerebral: Estudio de caso en educación media rural. *Horizonte Científico International Journal*, 3(2), 1–12. <https://doi.org/10.64747/m710ss96>
- Carvajal, P., Chuya, S., & Vargas, L. (2025). Estándares educativos y articulación de niveles en la Amazonía ecuatoriana. *Horizonte Científico International Journal*, 3(2), 1–10. <https://doi.org/10.64747/dqn7hm10>
- Gómez, A., & Viteri, N. (2025). Estrategias inclusivas para el desarrollo integral en educación básica inicial. *Horizonte Científico International Journal*, 3(2), 1–13. <https://doi.org/10.64747/wre5wb66>
- QGIS.org. (2022). QGIS Geographic Information System (v3). QGIS Association. <https://doi.org/10.5281/zenodo.5884349>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>

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