



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](mailto:marlontenecela@msn.com),  
<https://orcid.org/0009-0005-0559-7365>

**Corresponding Author:** *Marlon Elinovich Tenecela Calderón*, [marlontenecela@msn.com](mailto: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  $\mu_c$  y  $\tau$  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) | $\Delta$ (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.  $\Delta$ : 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  $\leq 40$  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  $\geq 5$  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 $\geq 5$ años (matemática) | Experiencia específica                 | 0.26     | [0.17, 0.34] |
| Docente $< 5$ años                 | Experiencia específica                 | 0.18     | [0.09, 0.27] |

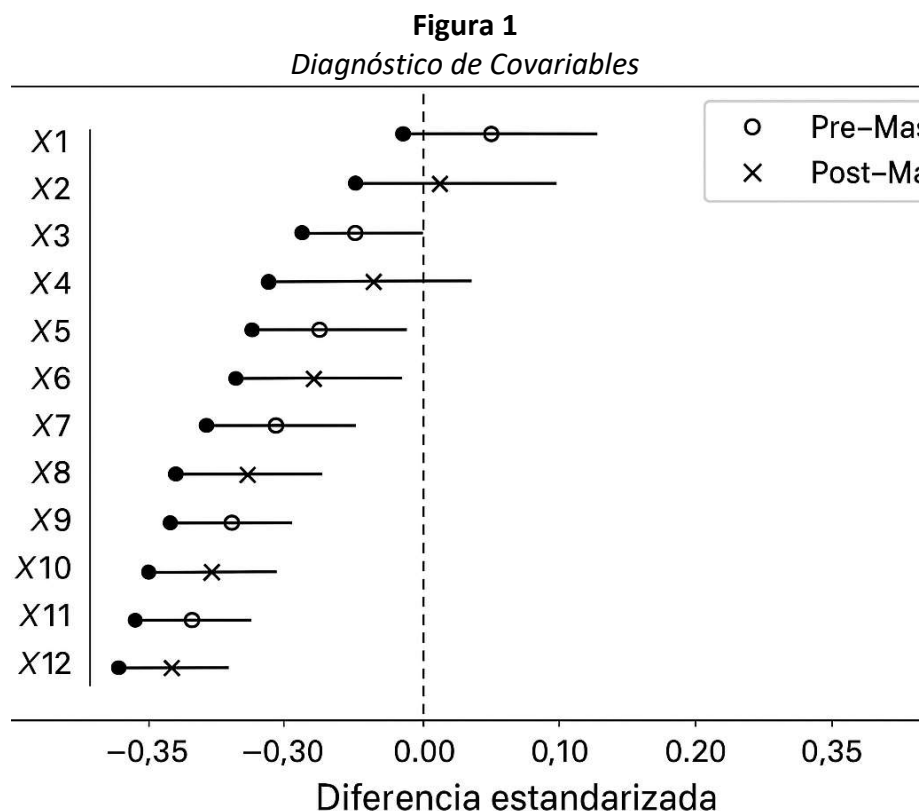
Mecanismos intermedios. La autoeficacia matemática aumentó 0.35 DE (IC95% [0.28, 0.42]) en el grupo tratado respecto del control, y el uso de estrategias metacognitivas reportadas creció 0.27 DE (IC95% [0.20, 0.34]). El análisis de mediación indica un efecto indirecto conjunto de 0.07 DE ( $\approx 30\%$  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  $\Delta$  tratado 8.7 vs. 3.0 en control), con ligeras

ganancias también en el extremo superior ( $p_{80}$ : 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.

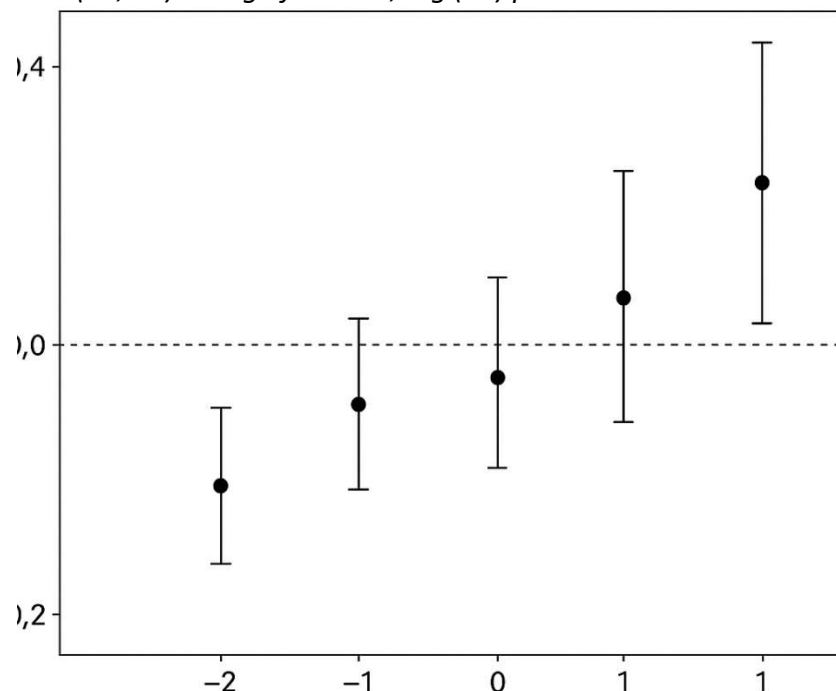


Resultados cualitativos. El análisis de rúbricas y notas de observación indica mejoras en: (a) explicitación de supuestos (de 34% a 61% de estudiantes que justifican supuestos de modelización), (b) verificación de resultados con cálculo aproximado (de 28% a 52%), y (c) argumentación escrita (de 31% a 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*

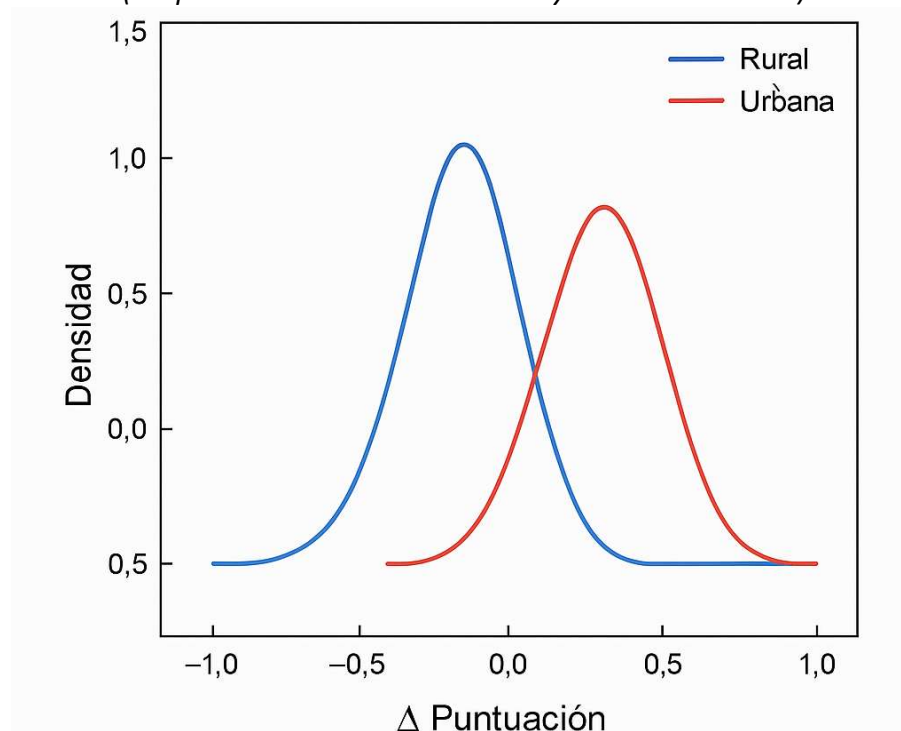
*Coeficientes lead (-2, -1) no significativos; lag (+1) positivo consistente con el tratamiento*



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**

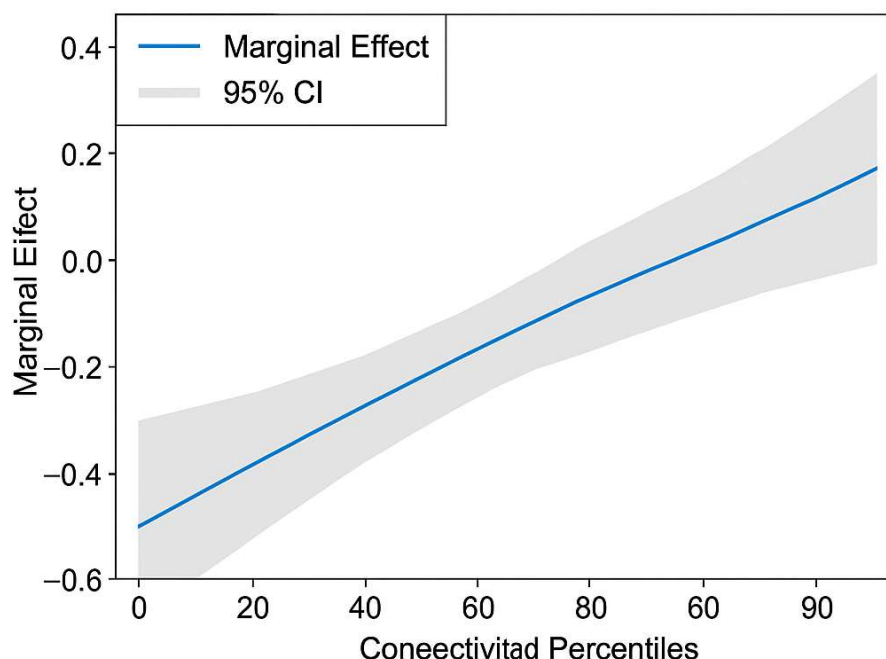
*Densidades Kernel del Cambio  $\Delta$  por Condición  
(desplazamiento a la derecha mayor en tratamiento)*



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 ( $\approx 0.23$





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