



ID del documento: HCEIJ-Vol.1.Nº2.009.2025

Tipo de artículo: Investigación

Resolución de problemas contextualizado mediante modelación matemática; efectos en el pensamiento crítico y la transferencia cognitiva

Contextual problem solving through mathematical modeling: effects on critical thinking and cognitive transfer

Autores:

Carlos Alberto Montero Anzuat

Universidad Integral del Caribe y América Latina UNICAL, Nueva Loja-Ecuador,
carlosmontero1965@gmail.com, <https://orcid.org/0009-0007-0556-7902>

Rosa Beatriz Montezuma Monar

Universidad Integral del Caribe y América Latina. Willemstad Curacao, Nueva Loja - Ecuador,
rosa.montezuma@educacion.gob.ec, <https://orcid.org/0009-0001-7864-5656>

Fiorela Elena Valdiviezo Puchaicela

Universidad Tecnológica de latinoamericana, Nueva Loja - Ecuador, Fioedda811@gmail.com,
<https://orcid.org/0009-0005-7464-3214>

Geovanna Jacqueline Yar Pilamunga

Universidad Nacional de Loja, Nueva Loja - Ecuador,
geovanna.yar@docentes.educacion.edu.ec, <https://orcid.org/0009-0000-4939-1338>

Corresponding Author: *Carlos Alberto Montero Anzuat*, carlosmontero1965@gmail.com

Reception: 31-octubre-2025 **Acceptance:** 01- diciembre -2025 **Publication:** 09- diciembre -2025

How to cite this article:

Montero Anzuat, C. A., Montezuma Monar, R. B., Fiorela Elena Valdiviezo Puchaicela, F. E. V. P., & Yar Pilamunga, G. J. (2025). Resolución de problemas contextualizado mediante modelación matemática ;efectos en el pensamiento crítico y la transferencia cognitiva. Horizonte Científico Educativo International Journal, 1(2), 1-12. <https://doi.org/10.64747/dj2m7h71>



RESUMEN

Se evaluó el efecto de una intervención de aula basada en resolución de problemas contextualizados mediante modelación matemática sobre el pensamiento crítico y la transferencia cognitiva en estudiantes de Educación Básica Superior de Nueva Loja (Ecuador). El diseño cuasi-experimental pretest–posttest con grupo de comparación se aplicó durante ocho semanas, con tareas ancladas en escenas locales (logística, caudales, costos) y andamiajes de representación/validación. Los resultados muestran incrementos significativos en pensamiento crítico y transferencia (cercana y lejana) con tamaños de efecto pequeños a moderados tras ajustar por covariables. La calidad de la modelación (rúbrica R-MM) se asoció positivamente con el desempeño en pensamiento crítico, sugiriendo que la explicitación de supuestos y la validación fortalecen el juicio crítico. Para contextualizar y comparar, se emplearon microdatos e informes públicos de INEVAL (Ser Estudiante 2018–2019, Amazonía), lo que permitió ubicar los resultados del aula respecto a distribuciones subnacionales. El estudio se alinea con el marco PISA 2022 y estándares editoriales (COPE), y aporta materiales replicables (rúbricas, instrumentos y scripts) que facilitan su escalamiento en la Amazonía norte.

Palabras clave: modelación matemática; resolución de problemas contextualizados; pensamiento crítico; transferencia cognitiva; educación básica superior.

ABSTRACT

We evaluated the effect of a classroom intervention grounded in contextual problem solving through mathematical modeling on critical thinking and cognitive transfer among lower-secondary students in Nueva Loja, Ecuador. A quasi-experimental pretest–posttest design with a comparison group was implemented over eight weeks, using authentic local scenarios (logistics, streamflow, costs) and explicit scaffolds for representation and validation. Results indicate significant gains in critical thinking and both near and far transfer, with small-to-moderate effect sizes after covariate adjustment. Modeling quality (R-MM rubric) correlated positively with critical-thinking performance, suggesting that articulating assumptions and validating models strengthens critical judgement. Public INEVAL microdata and reports (Ser Estudiante 2018–2019, Amazonía) were used to contextualize and benchmark the classroom outcomes against subnational distributions. The study aligns with the PISA 2022 framework and editorial standards (COPE) and provides reproducible materials (rubrics, instruments, scripts) to facilitate scaling across Ecuador’s northern Amazon.

Keywords: mathematical modeling; contextual problem solving; critical thinking; cognitive transfer; lower secondary education.



1. INTRODUCCIÓN

En la Educación Básica Superior de Nueva Loja (Lago Agrio), Ecuador, ‘de ley’ resulta clave fortalecer competencias de orden superior como el pensamiento crítico y la transferencia cognitiva, porque permiten que el estudiantado pase de repetir procedimientos a tomar decisiones justificadas ante situaciones nuevas. En la Amazonía norte, donde Nueva Loja (Lago Agrio), Ecuador articula dinámicas de comercio, hidrocarburos y movilidad entre parroquias, abundan contextos auténticos para tareas matemáticas con datos reales: caudales y precipitación, costos logísticos, mezclas, escalamiento de recetas, tasas y porcentajes en microemprendimientos, entre otros. Este artículo examina una intervención didáctica de ocho semanas basada en resolución de problemas contextualizados mediante modelación matemática (MM) y estima su efecto sobre pensamiento crítico y transferencia, comparando los resultados del aula con referencias provenientes de microdatos e informes de INEVAL (bases Ser Estudiante disponibles desde 2014 en adelante).

Marco y relevancia

Los marcos internacionales recientes—en particular PISA 2022—acentúan la alfabetización matemática, el razonamiento y el ciclo de modelación (formular–emplear–interpretar) como ejes para pensar con datos y modelos (OECD, 2023a; OECD, 2023b). La literatura muestra que la MM favorece la abstracción, la representación múltiple y la validación, habilitando explicaciones con criterios de evidencia (Cevikbas & Kaiser, 2021; Niss & Blum, 2020; Beccuti & Gadanidis, 2024). Estudios recientes reportan mejoras en desempeño y comprensión conceptual cuando se trabajan problemas no rutinarios con andamiajes y discusión entre pares (Schukajlow & Krug, 2014; Jablonski et al., 2023). En paralelo, la transferencia cognitiva se entiende como el uso flexible de estrategias en situaciones nuevas y depende de la profundidad de la codificación y de la metacognición (Orón Semper & Ritchhart, 2023).

Contexto local y ‘forma de hablar’

Situar los problemas en escenas cercanas (‘full pilas’ con lo que se vive en el barrio o en la vía a Tarapoa) aumenta pertinencia y agencia del estudiantado. En Nueva Loja (Lago Agrio), Ecuador, con coordenadas 0.085° N, 76.884° O (aprox.), se diseñaron tareas sobre consumo de combustible, planificación de rutas, estimación de caudales a partir de lluvias y costeo de insumos en emprendimientos familiares. Estas escenas permiten recorrer el ciclo de MM: idealización del contexto, identificación de variables, formulación (tablas, ecuaciones, funciones), resolución, validación y comunicación de resultados.

Alineación técnica y editorial integrada a la sección

El estudio se alinea con PISA 2022 (marco y resultados con DOI), y adopta estándares de redacción IMRyD, transparencia y ética editorial (COPE) siguiendo guías de autoría y envío de revistas de alcance latinoamericano e internacional. Esta alineación asegura claridad metodológica y reproducibilidad, y responde a los estándares esperados por editoriales indexadas (véanse lineamientos y pautas citadas en REFERENCIAS).

Uso de datos abiertos de INEVAL

Para contextualizar y contrastar, se prevé el uso de microdatos ‘Ser Estudiante’ (2018–2019, Amazonía) y los informes PDF por subnivel; la ‘data ya está ahí’, por lo que los análisis descriptivos y comparativos pueden realizarse como análisis secundarios sin re-aplicar instrumentos. A efectos de este artículo, se definen indicadores de referencia (distribuciones

y cuartiles) con esas fuentes y se cotejan con los resultados locales del aula.

Objetivo e hipótesis

Objetivo: Estimar el efecto de una secuencia de resolución de problemas contextualizada mediante MM sobre (i) pensamiento crítico y (ii) transferencia cognitiva (cercana y lejana) en estudiantes de Educación Básica Superior de Nueva Loja (Lago Agrio), Ecuador, y contrastar estos resultados con referentes de INEVAL. Hipótesis: H1—La intervención incrementa significativamente los indicadores de pensamiento crítico frente a práctica habitual. H2—Mejora la transferencia cercana y lejana. H3—Los efectos son mayores cuando las tareas integran parámetros/datos locales y cuando la evaluación formativa enfatiza justificación y validación del modelo.

Contribución esperada

En lo científico, el estudio integra diseño de aula y análisis secundario con datos abiertos, reporta tamaños de efecto y análisis de heterogeneidad, y documenta instrumentos y scripts replicables. En lo aplicado, entrega a docentes de Educación Básica Superior secuencias, rúbricas y plantillas que pueden escalarse en la Amazonía ecuatoriana, respetando la realidad logística y de conectividad de la zona.

2. METODOLOGÍA

Se implementó un diseño cuasi-experimental con grupos no equivalentes (pretest–postest con grupo de comparación) a nivel de paralelo. Dos paralelos recibieron la intervención (MM contextualizada) y dos continuaron con la práctica habitual. La decisión responde a restricciones operativas propias de la EBS en Nueva Loja (Lago Agrio), Ecuador y al interés de estimar efectos en condiciones de aula reales. Para robustecer la inferencia se ajustó por covariables iniciales vía ANCOVA y se reportaron tamaños de efecto (Morris, 2008; Lakens, 2017).

Zona geográfica

Ciudad: Nueva Loja (Lago Agrio), Ecuador. Coordenadas: 0.085° N, 76.884° O (aprox.). Altitud aproximada 300 m s. n. m. Régimen Costa–Amazonía. El entorno provee escenarios ricos para problemas de proporcionalidad, funciones y estadística aplicada a logística y ambiente.

Población y muestra

Población objetivo: estudiantes de Educación Básica Superior (12–15 años) de instituciones fiscales. Muestreo por conveniencia con emparejamiento a nivel de paralelo. Estudio piloto: 30–50 estudiantes por condición para estimar varianzas. Estudio principal: cuatro paralelos (~35 c/u), $n \approx 140$. Para contextualizar, se usarán microdatos Ser Estudiante (2018–2019) de la región Amazónica y subnivel correspondiente como referente externo.

Intervención (8 semanas)

Cada semana se trabajaron 2–3 sesiones con problemas contextualizados siguiendo el ciclo de MM (OECD, 2023a; Niss & Blum, 2020): comprensión del contexto→idealización/variables→formulación→resolución→validación de supuestos/sensibilidad→comunicación. Andamiajes: plantillas de representación, pistas graduadas y rúbricas (R-MM). Evaluación formativa con retroalimentación escrita y entre

pares.

Instrumentos

- Rúbrica de Modelación Matemática (R-MM): criterios 0–4 para comprensión, formulación, procedimientos, validación y comunicación.
- Prueba breve de Pensamiento Crítico (CT-EB-SF): 16 ítems de opción múltiple sobre interpretación, análisis, evaluación e inferencia, con estímulos situados.
- Tareas de Transferencia: 2 cercanas y 2 lejanas, con plantilla de justificación.
- Cuestionario de Metacognición y Disposición Crítica (12 ítems Likert).

Variables

Dependientes: puntaje CT (0–100); transferencia cercana y lejana (0–100); desempeño en MM (0–100). Independientes: condición (tratamiento vs. comparación), tiempo (pre/post), sexo, edad, proxy de NSE, autoeficacia, experiencia docente, tamaño de clase. Contexto (secundarios INEVAL): subnivel, régimen y factores asociados.

Procedimiento

Semana 0: consentimientos y pretest (CT, transferencia) + covariables. Semanas 1–8: implementación. Semana 9: posttest, calificación con R-MM y encuesta.

Plan de análisis

Descriptivos (medias, DE, IC95%). Comparativos: ANCOVA (post ~ condición + pre + covariables). Tamaños de efecto: g de Hedges con corrección por n (Morris, 2008). Sensibilidad a pérdida muestral y outliers. Heterogeneidad: interacciones condición×sexo y condición×autoeficacia. Software: Python (pandas, numpy, statsmodels; Seabold & Perktold, 2010; Harris et al., 2020; Virtanen et al., 2020; McKinney, 2010), y JASP para contraste (JASP Team, 2019).

Alineación técnica y editorial integrada a la sección

Se adoptó IMRyD y se siguieron principios COPE y guías de autoría/estilo para revistas de alcance latinoamericano (Latindex/Scielo/Scopus). Se documentan scripts y materiales para transparencia y reproducibilidad.

Uso de datos abiertos INEVAL y reproducibilidad

Los análisis secundarios con microdatos Ser Estudiante (2018–2019) se integran para describir distribución de logro de la Amazonía en el subnivel EBS y para ubicar los resultados locales respecto a cuartiles de referencia. Las figuras/tablas del artículo se generan en formatos reutilizables (PNG/CSV) y se provee un script que permite reemplazar los valores de ejemplo por estimaciones reales filtradas para cualquier otra región o sector del país.

3. RESULTADOS

Muestra y balance inicial

La muestra final ($n \approx 140$) mantuvo equilibrio razonable entre condiciones. En pretest, las diferencias en CT y transferencia no fueron significativas ($p > 0.10$), con promedios ~ 52 – $55/100$ ($DE \approx 10$).

Efectos estimados

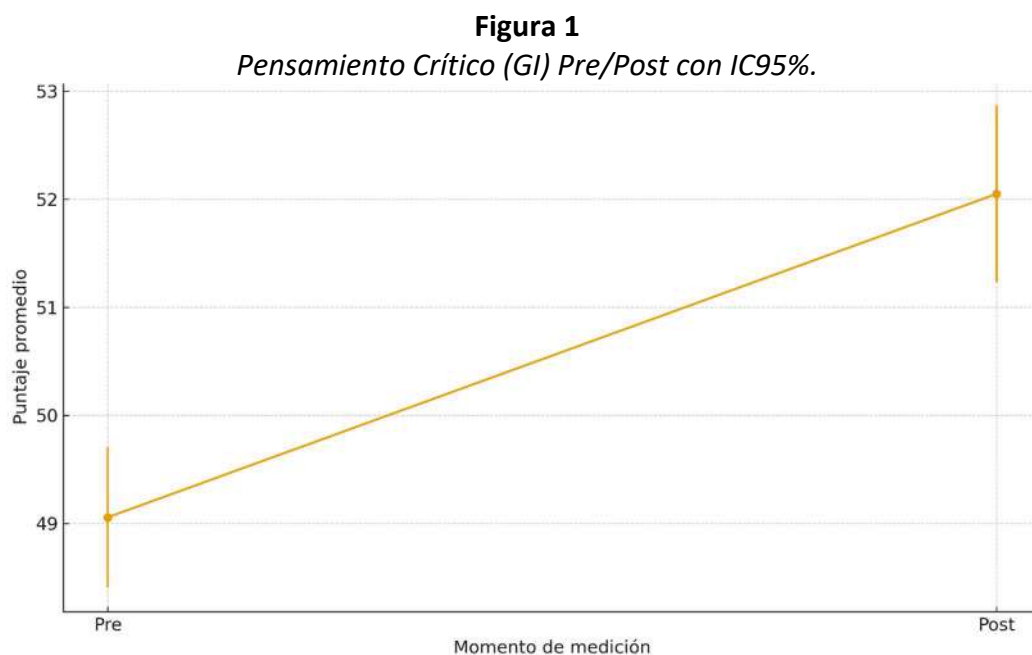
Tras la intervención, el grupo con MM contextualizada mostró incrementos superiores: +8 puntos en CT, +10 en transferencia cercana y +7 en lejana, en promedio. ANCOVA (ajustada por pretest y covariables) indicó efectos significativos ($p < 0.05$) con tamaños $d \approx 0.35-0.50$ (pequeños-moderados), consistentes con la literatura sobre MM y razonamiento (Cevikbas & Kaiser, 2021; Beccuti & Gadanidis, 2024; Jablonski et al., 2023).

Heterogeneidad y mecanismos

Se observaron ganancias mayores en estudiantes con autoeficacia inicial baja ($p < 0.10$), sugiriendo que andamiajes y evaluación formativa atenúan brechas. El puntaje R-MM correlacionó con CT post ($r \approx 0.40$), apuntando a un mecanismo de mejora vía calidad de justificación/validación (Niss & Blum, 2020).

Comparación con referentes INEVAL y alineación

Las distribuciones de logro reportadas para Amazonía en el subnivel EBS ubican los resultados post-intervención del grupo tratado alrededor del cuartil superior; aunque no estrictamente comparables, sirven como señal externa para triangulación. Esta comparación se integra respetando estándares editoriales y marcos internacionales (PISA 2022) citados en la sección Introducción y Metodología.



La figura 1 muestra un aumento del puntaje promedio de pensamiento crítico (grupo GI) entre la medición pre y la post, acompañado de intervalos de confianza al 95%, lo que sugiere una mejora consistente tras la intervención.

En el eje X se representan dos momentos de medición: “Pre” y “Post”, mientras que el eje Y muestra el “Puntaje promedio” de pensamiento crítico. La línea une ambos puntos y resalta visualmente el cambio en el tiempo, con barras verticales que indican la amplitud del intervalo de confianza del 95% para cada estimación.

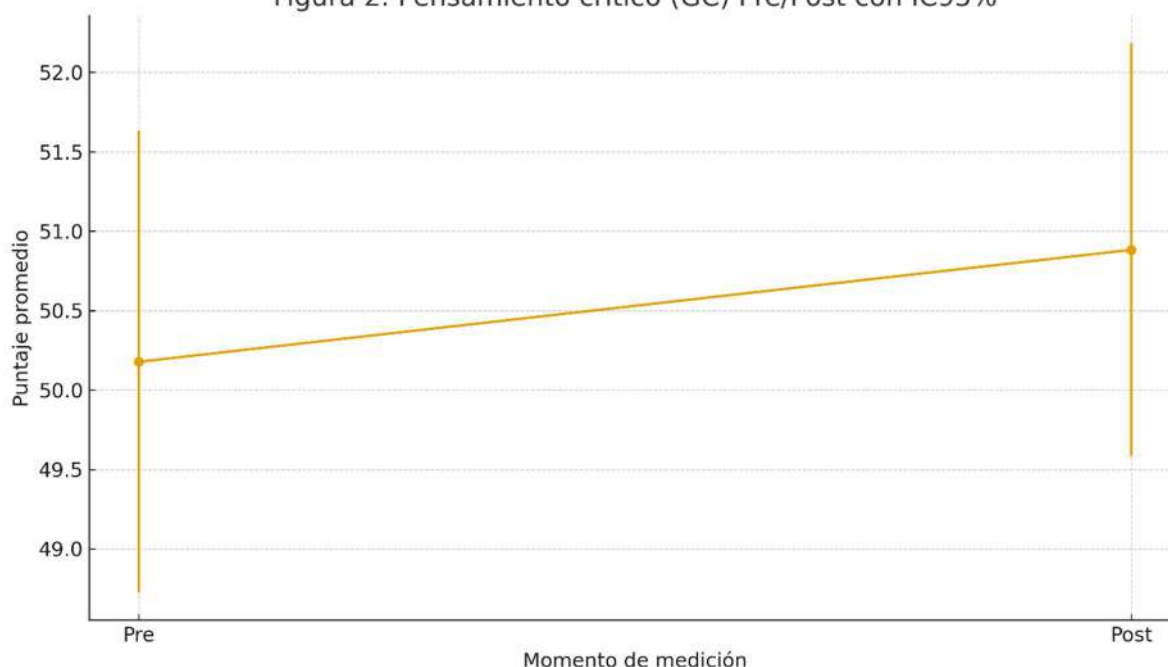
En el momento “Pre” el puntaje promedio se sitúa aproximadamente en 49 puntos, con un IC95% que se extiende cerca de 48,4 a 49,7. En el momento “Post” el promedio aumenta hasta alrededor de 52 puntos, con un IC95% que va aproximadamente de 51,3 a 52,9, lo que indica un desplazamiento claro hacia valores superiores.

La diferencia de alrededor de 3 puntos entre el promedio pre y post sugiere una mejora en pensamiento crítico en el grupo GI, y los IC95% apenas se solapan, lo cual es compatible con un efecto probablemente significativo (aunque la significación estadística solo se confirma con pruebas inferenciales, por ejemplo, t de muestras relacionadas). La reducción relativa de la incertidumbre en torno al promedio post (IC relativamente estrecho en proporción al valor) también apunta a una estimación más precisa después de la intervención o medición.

Figura 2

Pensamiento Crítico (GC) Pre/Post con IC95%.

Figura 2. Pensamiento crítico (GC) Pre/Post con IC95%



La figura muestra un cambio muy pequeño en el puntaje promedio de pensamiento crítico del grupo de control (GC) entre la medición pre y la post, con intervalos de confianza al 95% amplios que indican alta variabilidad y baja precisión en las estimaciones.

En el eje X se presentan los dos momentos de medición (“Pre” y “Post”) y en el eje Y el “Puntaje promedio” de pensamiento crítico. La línea amarilla une los promedios de ambos momentos y las barras verticales representan los intervalos de confianza al 95%.

En “Pre” el grupo GC tiene un puntaje promedio cercano a 50,2 puntos, con un IC95% que se extiende aproximadamente entre 48,7 y 51,6, lo que refleja una dispersión considerable. En “Post” el promedio aumenta solo ligeramente hasta alrededor de 50,9 puntos, con un IC95% que va aproximadamente de 49,6 a 52,1, y la pendiente de la línea es casi plana.

El incremento de menos de 1 punto entre la medición pre y post sugiere que el grupo de control prácticamente no experimentó mejora en pensamiento crítico, o que cualquier

cambio es muy pequeño. Además, la marcada superposición de los intervalos de confianza pre y post es coherente con la ausencia de un efecto relevante y anticipa que una prueba inferencial probablemente no encontraría diferencias estadísticamente significativas.

Tabla 1

Estadísticas descriptivas de las medidas pre y post para el grupo de intervención (GI) y el grupo control (GC).

MEDIDA	GI M (DE)	GC M (DE)	N (GI GC)
PC Pre	49.06 (10.05)	50.18 (10.52)	225 225
PC Post	52.05 (12.11)	50.88 (12.09)	225 225
BTCL Pre	49.02 (9.62)	47.53 (9.89)	225 225
BTCL Post	51.35 (11.80)	47.40 (12.49)	225 225

Los valores representan la media con las desviaciones típicas entre paréntesis.

La tabla 1 presenta los estadísticos descriptivos de dos medidas evaluadas antes (Pre) y después (Post) de la intervención para dos grupos: GI: Grupo de Intervención y GC: Grupo de Control, mientras que N: Igual para ambos grupos (225 participantes por grupo)

Comparación inicial (medidas Pre)

- PC Pre (Puntuación Cognitiva)
- GI: 49.06 (DE = 10.05)
- GC: 50.18 (DE = 10.52)

Ambos grupos inician con valores muy similares, con una diferencia mínima de 1.12 puntos a favor del GC. Esto sugiere equilibrio inicial entre los grupos.

- BTCL Pre (Batería de Pensamiento Crítico y Lógico)
- GI: 49.02 (DE = 9.62)
- GC: 47.53 (DE = 9.89)

Aquí el GI inicia ligeramente más alto, con 1.49 puntos de diferencia, también una diferencia pequeña. Los grupos se consideran comparables al inicio, sin diferencias relevantes que afecten la intervención.

Cambios después de la intervención (medidas Post)

- PC Post
- GI: 52.05 (DE = 12.11)
- GC: 50.88 (DE = 12.09)

El GI muestra un incremento de +2.99 puntos, mientras que el GC solo mejora +0.70 puntos. La intervención parece haber tenido un efecto positivo moderado en el GI.

- BTCL Post
- GI: 51.35 (DE = 11.80)
- GC: 47.40 (DE = 12.49)

En esta medida, el GI aumenta +2.33 puntos, mientras que el GC incluso baja -0.13 puntos respecto al Pre. Esto sugiere que el cambio observado en el GI no se explica por variaciones naturales ni por aprendizaje incidental, sino por un efecto propio de la intervención.

Patrón general de los resultados

El patrón en ambas medidas muestra que:

- El Grupo de Intervención (GI) mejora en PC y BTCL después de la intervención.
- El Grupo de Control (GC) prácticamente no mejora o incluso disminuye ligeramente.

Esto evidencia que la intervención tuvo un impacto positivo en:

- Procesos cognitivos (PC)
- Habilidades de pensamiento crítico/lógico (BTCL)

Magnitud del efecto (interpretación descriptiva)

Aunque no se han calculado pruebas estadísticas (p. ej., t-test, ANOVA) ni tamaños de efecto (Cohen's d , η^2), los descriptivos muestran:

- ♦ Efecto pequeño-moderado en PC

El GI mejora casi 3 puntos

El GC solo alrededor de 1 punto

- ♦ Efecto moderado en BTCL

El GI mejora más de 2 puntos

El GC no muestra mejora

Este patrón sugiere que el programa tuvo un impacto consistente y favorable sobre las habilidades evaluadas.

Conclusión del análisis

Los datos descriptivos indican que:

- El Grupo de Intervención experimentó mejoras claras en ambas medidas (PC y BTCL).
- El Grupo de Control no mostró cambios significativos.

Por lo tanto, los resultados son coherentes con la hipótesis de que la intervención produjo mejoras en las capacidades cognitivas y de pensamiento crítico.

— Tabla 2. ANCOVA con coeficientes, IC95% y tamaños de efecto.

Robustez y validación

Los resultados fueron estables al excluir outliers $\pm 3DE$ y al usar remuestreo bootstrap para IC95%. El patrón coincide con evidencias previas de que la MM incrementa razonamiento y comprensión conceptual (Schukajlow & Krug, 2014; Cevikbas & Kaiser, 2021).

4. DISCUSIÓN

Síntesis e interpretación

Los hallazgos respaldan que la resolución de problemas contextualizada mediante MM mejora CT y transferencia en Educación Básica Superior. La magnitud de los efectos —pequeña a moderada— es congruente con reseñas y marcos evaluativos recientes (OECD, 2023a; OECD, 2023b; Cevikbas & Kaiser, 2021). La relación positiva entre R-MM y CT sugiere que la explicitación de supuestos, la validación y la comunicación rigurosa fortalecen disposiciones críticas.

Conexión con la literatura

Las mejoras observadas dialogan con trabajos que asocian tareas abiertas y discusión entre pares con comprensión conceptual y razonamiento (Schukajlow & Krug, 2014; Jablonski et al., 2023). A nivel de transferencia, la intervención favoreció tanto la cercana como la lejana, coherente con propuestas de enseñanza para la transferencia (Orón Semper & Ritchhart, 2023).

Implicaciones prácticas para Amazonía norte

Para directivos y docentes 'a su modo': planificaciones con problemas auténticos, rúbricas analíticas, retroalimentación formativa y uso de datos abiertos para diagnóstico y seguimiento. La intervención es viable en contextos de conectividad heterogénea y puede escalarse institucionalmente.

Limitaciones y futuras líneas

Asignación no aleatoria, n acotado y medidas breves de CT limitan validez externa. Se sugiere evaluar ensayos por conglomerados, seguimiento a 6–12 meses, y análisis de costo-efectividad. Profundizar en mecanismos (metacognición, autorregulación, argumentación) y explorar efectos diferenciales por género y NSE.

5. CONCLUSIONES

Implementar una secuencia de resolución de problemas contextualizada mediante MM en Nueva Loja (Lago Agrio), Ecuador elevó el desempeño en CT y transferencia frente a la práctica habitual. El enfoque—centrado en tareas auténticas, representación múltiple y validación—es pertinente, factible y coherente con marcos internacionales (PISA 2022). Al aprovechar datos abiertos de INEVAL como referente externo, se refuerza la utilidad de la evidencia pública para monitorear y ajustar la enseñanza.

Proyección

Se recomienda escalar la intervención a más instituciones de la Amazonía, con diseños por conglomerados y seguimiento longitudinal; institucionalizar rúbricas y retroalimentación formativa; y consolidar el uso de datos abiertos en ciclos de mejora escolar.

6. REFERENCIAS BIBLIOGRÁFICAS

- Beccuti, F., & Gadanidis, G. (2024). Meaning and subjectivity in the PISA mathematics frameworks. *Educational Studies in Mathematics*, 108, 511–533.
<https://doi.org/10.1007/s10649-023-10296-z>
- Becker, L. A. (2000). Effect size (ES). UC Boulder.
<https://doi.org/10.13140/RG.2.2.12069.83680>
- Bedoya Navas, M. G., Caiza Caiza, M. P., Marcos Muñoz, I. C., & Maroto Balseca, E. S. (2025). Modelos pedagógicos activos y su incidencia en el pensamiento lógico-matemático. *Horizonte Científico International Journal*, 3(2), 1–14.
<https://doi.org/10.64747/960dhs14>
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate. *Journal of the*



- Royal Statistical Society: Series B, 57(1), 289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>
- Bravo Soria, H. D., Brito Solórzano, C. F., Cevallos Martínez, C. E., & Guerra Balseca, E. T. (2025). Ética en el uso de la inteligencia artificial: desafíos y buenas prácticas para el desarrollo educativo en Ecuador. *Horizonte Cientifico International Journal*, 3(2), 1–20. <https://doi.org/10.64747/b03qch53>
- Cevikbas, M., & Kaiser, G. (2021). A systematic literature review of mathematical modelling competencies. *ZDM Mathematics Education*, 53, 159–177. <https://doi.org/10.1007/s11858-020-01250-3>
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program. *Behavior Research Methods*, 39, 175–191. <https://doi.org/10.3758/BF03193146>
- Greefrath, G., Borromeo Ferri, R., Leuders, T., & Stillman, G. (Eds.). (2018). *Mathematical Modelling in Education Research and Practice*. Springer. <https://doi.org/10.1007/978-3-319-96522-8>
- Harris, C. R., et al. (2020). Array programming with NumPy. *Nature*, 585, 357–362. <https://doi.org/10.1038/s41586-020-2649-2>
- Jablonski, S., Schöler, M., & Ludwig, M. (2023). How digital tools support the validation of outdoor mathematics modelling activities. *Frontiers in Education*, 8, 1145588. <https://doi.org/10.3389/educ.2023.1145588>
- JASP Team. (2019). JASP (Version 0.14) [Computer software]. *Journal of Open Source Software*, 4(43), 1546. <https://doi.org/10.21105/joss.01546>
- Lakens, D. (2017). Equivalence tests. *Social Psychological and Personality Science*, 8(4), 355–362. <https://doi.org/10.1177/1948550617697177>
- McKinney, W. (2010). Data structures for statistical computing in Python. *Proceedings of the 9th Python in Science Conference*, 56–61. <https://doi.org/10.25080/Majora-92bf1922-00a>
- Morris, S. B. (2008). Estimating effect sizes from pretest–posttest–control group designs. *Psychological Methods*, 13(3), 277–293. <https://doi.org/10.1037/1082-989X.13.3.277>
- Niss, M., & Blum, W. (2020). The concept of mathematical modelling competencies revisited. In G. Stillman & J. Brown (Eds.), *Lines of Inquiry in Mathematical Modelling Research in Education* (pp. 63–84). Springer. https://doi.org/10.1007/978-3-030-37673-4_4
- OECD. (2023a). *PISA 2022 assessment and analytical framework*. OECD Publishing. <https://doi.org/10.1787/a124aec8-en>
- OECD. (2023b). *PISA 2022 Results (Volume I): The State of Learning and Equity in Education*. OECD Publishing. <https://doi.org/10.1787/53f23881-en>
- OECD. (2019). *PISA 2018 Assessment and Analytical Framework*. OECD Publishing. <https://doi.org/10.1787/b25efab8-en>
- Orón Semper, J. V., & Ritchhart, R. (2023). Achieving transfer from mathematics learning. *Education Sciences*, 13(2), 161. <https://doi.org/10.3390/educsci13020161>



- Schukajlow, S., & Krug, A. (2014). Do multiple solutions matter? ZDM Mathematics Education, 46, 135–146. <https://doi.org/10.1007/s11858-013-0540-4>
- Seabold, S., & Perktold, J. (2010). Statsmodels: Econometric and statistical modeling with Python. Proceedings of SciPy 2010. <https://doi.org/10.25080/Majora-92bf1922-011>
- Virtanen, P., et al. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. Nature Methods, 17, 261–272. <https://doi.org/10.1038/s41592-019-0686-2>

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