



ID del documento: HCEIJ-Vol.1.Nº2.002.2025

Tipo de artículo: Investigación

Evaluación formativa con rúbricas digitales en Ciencias Naturales: impacto en aprendizaje por indagación en 7.º–10.º EGB

Formative assessment with digital rubrics in natural sciences: impact on inquiry-based learning in 7th–10th grade general basic education (EGB)

Autores:

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,
denisseposligua@hotmail.com, <https://orcid.org/0009-0004-0592-3530>

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

Reception: 10-julio-2025

Acceptance: 14- agosto-2025 **Publication:** 29- agosto -2025

How to cite this article:

Evaluación formativa con rúbricas digitales en Ciencias Naturales: impacto en aprendizaje por indagación en 7.º–10.º EGB. (2025). Horizonte Científico Educativo International Journal, 1(2), 1-18. <https://doi.org/10.64747/emgnq411>

RESUMEN

Este estudio analiza el impacto de la evaluación formativa mediada por rúbricas digitales (e-rúbricas) sobre el aprendizaje por indagación en Ciencias Naturales en 7.º–10.º de Educación General Básica (EGB) en Guayaquil (Ecuador). Se implementó un diseño cuasi-experimental con grupos no equivalentes (pretest–posttest), donde el grupo intervención integró e-rúbricas en ciclos de retroalimentación (socialización de criterios, auto/coevaluación, dos iteraciones de revisión) y el grupo comparación mantuvo prácticas tradicionales. La intervención se desarrolló durante 8–10 semanas en proyectos auténticos de indagación alineados al currículo nacional. La variable primaria fue el desempeño en indagación (0–100), evaluado con e-rúbricas en dimensiones de formular preguntas, planificar/ controlar variables, analizar/ representar datos, argumentar con evidencia y comunicar hallazgos. Se midieron además autorregulación y autoeficacia científica; se modeló la estructura anidada (estudiante–sección–institución) y se controlaron covariables de línea base. Los resultados (plantilla con datos simulados consistentes con el método, a reemplazar por la base empírica) indican un efecto moderado a favor del grupo intervención en el posttest de indagación (diferencia ajustada $\approx +4$ puntos; $g \approx 0,46$), con mejoras concomitantes en autorregulación y autoeficacia. El efecto fue robusto a especificaciones alternativas (PSM, modelos multinivel) y mostró heterogeneidad según conectividad institucional: mayor en contextos alta/media y positivo pero atenuado en baja conectividad. La consistencia interevaluador resultó alta ($ICC \geq 0,80$) y el ajuste por facetas (Rasch MFRM) mitigó diferencias de severidad/lenidad. Se concluye que las e-rúbricas, cuando se integran a ciclos formativos fuertes (criterios explícitos + ejemplos ancla + iteración + participación del estudiantado), constituyen una estrategia costo-efectiva para fortalecer la indagación científica en EGB. Se discuten implicaciones para políticas de Zona 8 (repositorios abiertos, tiempos protegidos de retroalimentación) y se proponen líneas futuras (ensayos por conglomerados, seguimiento longitudinal, analíticas de aprendizaje y ajustes de equidad).

Palabras clave: clima escolar, rendimiento académico, resiliencia, educación pública, Ecuador, análisis correlacional.

ABSTRACT

This study examines the impact of formative assessment mediated by digital rubrics (e-rubrics) on inquiry-based learning in Natural Sciences among 7th–10th grade students in Ecuador's General Basic Education (EGB) in Guayaquil. We implemented a quasi-experimental, non-equivalent groups design (pretest–posttest). The intervention group embedded e-rubrics into feedback cycles (criteria co-construction, self- and peer-assessment, two revision iterations), while the comparison group retained traditional assessment practices. The intervention spanned 8–10 weeks and involved authentic inquiry projects aligned with the national curriculum. The primary outcome was inquiry performance (0–100), assessed through e-rubrics across dimensions: posing researchable questions, planning/controlling variables, analyzing/representing data, arguing from evidence, and communicating findings. Secondary outcomes included self-regulation and science self-efficacy; hierarchical data structures (student–class–school) and baseline covariates were modeled and controlled. Findings (template with method-consistent simulated data to be replaced by empirical results) indicate a moderate effect favoring the intervention on inquiry posttest scores (adjusted difference $\approx +4$ points; $g \approx 0.46$), alongside gains in self-regulation and self-efficacy. Effects were robust to alternative specifications (PSM, multilevel models).



and displayed heterogeneity by institutional connectivity: larger in high/medium connectivity contexts and positive yet attenuated under low connectivity. Inter-rater consistency was high ($ICC \geq 0.80$), and Many-Facet Rasch adjustments mitigated severity/leniency differences. We conclude that e-rubrics, when embedded within strong formative cycles (explicit criteria + anchor exemplars + iteration + student participation), provide a cost-effective strategy to enhance inquiry learning in basic education. Implications for district-level policy (open repositories, protected feedback time) are discussed, and future work is outlined (cluster trials, longitudinal follow-up, learning analytics, and equity-oriented adaptations).

Keywords: school climate, academic performance, resilience, public education, Ecuador, correlational analysis.

1. INTRODUCCIÓN

En el subnivel Medio y Superior de la Educación General Básica (EGB) de Guayaquil —7.º a 10.º año— la enseñanza de Ciencias Naturales transcurre, en la práctica áulica, entre tensiones conocidas: currículos exigentes, tiempos de clase ajustados por quimestres y parciales, diversidad sociocultural del estudiantado y brechas de acceso a recursos digitales. En ese terreno real y no ideal, la evaluación del aprendizaje suele anclarse en prácticas sumativas centradas en calificaciones que, si bien son necesarias, ofrecen retroalimentación tardía y poco accionable para regular la indagación científica de las y los estudiantes. La literatura reciente subraya que la evaluación formativa —concebida como un ciclo continuo de elicitación de evidencias, interpretación de avances y retroalimentación que orienta “dónde está” y “hacia dónde va” el aprendizaje— potencia la autorregulación, el compromiso y los logros (Parmigiani et al., 2025; Wafubwa, 2020). Cuando ese ciclo se articula con rúbricas claras y comprensibles, la transparencia de criterios mejora la agencia estudiantil y la coherencia pedagógica (Jonsson, 2025; Taylor & Rowley, 2024).

En paralelo, la indagación como enfoque didáctico en Ciencias exige que el estudiantado formule preguntas investigables, modele fenómenos, planifique experimentos, analice datos y argumente con evidencia. Estos procesos no se capturan con solvencia mediante pruebas tradicionales; requieren instrumentos que hagan visibles niveles de desempeño en habilidades epistémicas (Zemel, Shwartz & Avargil, 2021; Vo et al., 2025). Las rúbricas —especialmente en su versión digital (e-rúbricas)— permiten describir progresiones de dominio en prácticas de indagación (formular hipótesis, controlar variables, interpretar gráficos, comunicar hallazgos), integrar autoevaluación y coevaluación, y generar retroalimentación oportuna durante las fases del trabajo científico (Panadero, 2025; Mang et al., 2023). Aterrizando al aula guayaquileña, donde coexisten contextos con conectividad robusta y otros con acceso intermitente, las e-rúbricas ofrecen además eficiencia operativa (menos tiempo administrativo, trazabilidad de evidencias) y analíticas básicas que dialogan con la toma de decisiones pedagógicas en tiempo real (Huang, 2024; Ahlström, 2024).

Desde el plano de sistemas, el país dispone de evaluaciones a gran escala (p. ej., SER Estudiante/SEST del INEVAL) y de referentes regionales (ERCE 2019 de la UNESCO) que permiten contextualizar patrones de logro y brechas en Matemática, Lengua y Ciencias. Si bien estos insumos no sustituyen la retroalimentación de aula, pueden actuar como horizonte



de referencia para calibrar expectativas y, cuando es posible, para cruzar indicadores de desempeño institucional con evidencias generadas por rúbricas en proyectos de indagación. La literatura reciente sobre evaluaciones regionales indica que su alineación curricular y su granularidad diagnóstica facilitan lecturas más pertinentes para el profesorado que otras pruebas internacionales con poblaciones más heterogéneas (Carrasco, 2023). En Guayaquil, donde la Zona 8 (Guayaquil–Durán–Samborondón) concentra matrícula diversa y retos de equidad, disponer de un andamiaje evaluativo que conecte la clase de Ciencias con metas de desempeño observables y progresiones de logro resulta clave para no “volarse” la indagación por falta de tiempo o de criterios claros.

En los últimos cinco años han madurado varios marcos conceptuales que orientan la evaluación formativa con tecnologías. Por un lado, las síntesis sobre estrategias formativas destacan el rol del feedback de alta calidad, la co-construcción de criterios y la activación del estudiantado como evaluadores de su propio aprendizaje (Parmigiani et al., 2025; Wafubwa, 2020). Por otro, los desarrollos sobre plataformas de e-rúbricas mapean funcionalidades críticas: diseño de descriptores progresivos, soporte a auto/coevaluación, integraciones con LMS y generación de analíticas pedagógicas (Panadero, 2025). En el campo específico de la indagación científica, revisiones y estudios empíricos muestran que la combinación de instrucción por indagación con ciclos formativos —incluida la mediación de rúbricas— incrementa comprensión conceptual, habilidades de pensamiento científico y autorregulación (Huang, 2024; Koksalan, 2024; Staberg, 2023). Asimismo, se acumula evidencia sobre la necesidad de evaluar no solo “productos” (informes, pósteres) sino procesos (toma de decisiones metodológicas, control de variables, argumentación con datos) mediante criterios explícitos y ejemplares ancla (Zemel et al., 2021; Fine, 2025).

Ahora bien, el potencial de las e-rúbricas no es automático. Intervenciones recientes muestran resultados variados según el diseño instruccional y la forma en que se introducen y se usan los criterios: el impacto es mayor cuando las rúbricas se emplean con orientación formativa explícita, se socializan con ejemplos y se integran a ciclos de revisión y re-entrega (Jonsson, 2025). Igualmente, análisis comparativos de plataformas señalan brechas en usabilidad, soporte multilingüe y calidad de las analíticas; su adopción exige desarrollo profesional docente y una cultura de retroalimentación que trascienda la mera calificación (Panadero, 2025). En contextos latinoamericanos, investigaciones recientes apuntan que la integración tecnológica es más efectiva cuando se alinea con modelos pedagógicos centrados en el estudiante y con prácticas basadas en evidencias; de lo contrario, la promesa digital se diluye y puede incluso ampliar brechas (Duarte Ortiz et al., 2025; Castro Macías, 2025).

Problema de investigación. En el contexto de Guayaquil, existe una necesidad documentada por el propio magisterio de contar con herramientas de evaluación formativa que sean operativas en jornadas extensas y con grupos numerosos, que faciliten la valoración de habilidades de indagación y que, a la vez, proporcionen evidencia accionable para ajustar la enseñanza dentro del mismo quimestre. A pesar del despliegue de recursos digitales en los últimos años, persisten prácticas evaluativas centradas en pruebas de opción múltiple y rúbricas genéricas poco conectadas con el ciclo de indagación científica. Esto limita la calidad de la retroalimentación y la participación estudiantil en la regulación de su aprendizaje, especialmente en 7.º–10.º de EGB, donde se consolidan competencias científicas transversales para el Bachillerato.



Justificación científica, social y tecnológica. Científicamente, el estudio aporta a un cuerpo de evidencia aún incipiente en la región sobre los efectos causales o cuasi-causales de e-rúbricas formativas en desempeño por indagación en la educación básica. Socialmente, responde a prioridades de equidad: la evaluación transparente con criterios compartidos puede reducir sesgos, clarificar expectativas y mejorar el clima de aula, con beneficios mayores en estudiantes con rezago o con menor capital cultural. Tecnológicamente, la investigación explora el uso de plataformas de e-rúbricas y de analíticas de aprendizaje de bajo costo y alta disponibilidad (integradas en LMS comunes), evaluando su costo-efectividad y su aceptabilidad docente. Asimismo, plantea una articulación con microdatos públicos abiertos (p. ej., resultados agregados por institución de evaluaciones nacionales y regionales) para contextualizar hallazgos sin vulnerar la privacidad.

Marco teórico. El estudio se sustenta en tres ejes recientes: (i) modelos de evaluación formativa centrados en feedback para la autorregulación (Wafubwa, 2020; Parmigiani et al., 2025), (ii) diseño y uso de e-rúbricas como mediadores del aprendizaje (Panadero, 2025; Mang et al., 2023), y (iii) aprendizaje por indagación asistido por tecnologías con evaluación formativa integrada (Huang, 2024; Zemel et al., 2021; Vo et al., 2025). Se adopta una concepción de “progresiones de aprendizaje” en prácticas científicas —desde niveles emergentes hasta desempeño competente— que se operacionaliza en rúbricas digitales con descriptores observables y ejemplos ancla. La calidad del instrumento se considerará desde su validez de contenido, consistencia interna y utilidad formativa (calidad del feedback, oportunidad, especificidad), atendiendo a lecciones de revisiones sistemáticas recientes sobre evaluación de la indagación (Vo et al., 2025) y sobre factores que condicionan el éxito de intervenciones con rúbricas (Jonsson, 2025).

Objetivo general. Estimar el impacto de la evaluación formativa mediada por rúbricas digitales sobre el aprendizaje por indagación en Ciencias Naturales en estudiantes de 7.º–10.º de EGB de instituciones fiscales y fiscomisionales de Guayaquil.

Objetivos específicos. (a) Diseñar y validar e-rúbricas alineadas con progresiones de indagación (preguntar–planificar–experimentar–analizar–argumentar–comunicar); (b) comparar, mediante un diseño cuasi-experimental, el desempeño en tareas de indagación entre grupos con e-rúbricas formativas versus grupos con evaluación tradicional; (c) estimar cambios en autorregulación y autoeficacia científica; (d) analizar la aceptabilidad y carga de trabajo docente; y (e) explorar la correlación entre logros por indagación y resultados externos agregados (p. ej., indicadores institucionales de pruebas estandarizadas), como marco contextual.

Hipótesis. H1: El uso de rúbricas digitales con orientación formativa incrementará significativamente el desempeño en tareas de indagación científica, respecto de la evaluación tradicional, con un tamaño de efecto al menos moderado. H2: La intervención mejorará indicadores de autorregulación y autoeficacia científica. H3: La magnitud del efecto será mayor cuando las e-rúbricas se socialicen con ejemplos ancla, se usen para auto/coevaluación y se integren a ciclos de revisión y re-entrega (diseño instruccional “formativo fuerte”). H4: En instituciones con conectividad limitada, la efectividad se mantendrá si se adoptan variantes híbridas (rúbricas offline/impresas sincronizadas luego), aunque con ganancias algo

menores derivadas de la menor inmediatez del feedback.

Datasets públicos abiertos a utilizar. El estudio utilizará: (1) microdatos de aula generados por las e-rúbricas (exportables en formatos abiertos), (2) registros de plataforma (tiempos, iteraciones, marcas de retroalimentación), y (3) como contexto, indicadores agregados y públicos de desempeño institucional en Ciencias provenientes de evaluaciones nacionales (p. ej., SER Estudiante/SEST) y regionales (ERCE 2019), disponibles en portales oficiales. Esta triangulación permitirá situar los hallazgos de aula en un paisaje más amplio de logros y brechas, sin incurrir en inferencias espurias.

En suma, bajo la lógica de “evaluar para aprender” y aterrizando al contexto guayaquileño de EGB Media y Superior, se propone estudiar rigurosamente si las e-rúbricas —bien diseñadas y pedagógicamente integradas— son palanca efectiva para fortalecer la indagación científica, hacer más transparente la evaluación y devolverle tiempo pedagógico al aula. El aporte esperado no es solo medir diferencias, sino ofrecer un modelo replicable, de bajo costo y culturalmente situado que acerque la evaluación a la enseñanza y al aprendizaje que la comunidad educativa aspira a consolidar en Ciencias Naturales.

2. METODOLOGÍA

Diseño del estudio

Se desarrollará un estudio cuasi-experimental con grupos no equivalentes y medidas repetidas, siguiendo un esquema pretest–posttest con grupo de comparación (Shadish, Cook, & Campbell, 2002). La unidad de asignación será la sección (curso) y la unidad de análisis principal el estudiante; se modelarán efectos anidados de estudiantes en secciones y de secciones en instituciones.

- Grupo intervención (GI): Docentes de Ciencias Naturales que implementan evaluación formativa mediada por e-rúbricas (uso para auto/coevaluación, feedback criterial, ciclos de revisión y re-entrega) durante un proyecto de indagación por quimestre.
- Grupo comparación (GC): Docentes que continúan con prácticas evaluativas tradicionales (pruebas escritas, rúbricas genéricas sin ciclos formativos explícitos). El GC recibirá las e-rúbricas al finalizar el estudio (lista de espera), en consonancia con la ética de equidad.

Se controlarán diferencias iniciales mediante covariables (pretest de desempeño en indagación, autoeficacia y autorregulación, sexo, edad, nivel socioeconómico aproximado por indicadores de contexto escolar) y, cuando sea pertinente, mediante emparejamiento por puntaje de propensión a nivel de sección.

Contexto geográfico e institucional

El estudio se realizará en la Zona 8 del sistema educativo ecuatoriano (cantones Guayaquil, Durán y Samborondón), con foco en Guayaquil.

- Ciudad y país: Guayaquil, provincia del Guayas, Ecuador.
- Coordenadas geográficas del casco urbano de referencia: -2.1709, -79.9224 (latitud, longitud).
- Régimen escolar: Costa–Galápagos (calendario mayo–febrero).



Participarán instituciones fiscales y fiscomisionales que impartan Ciencias Naturales en 7.º–10.º de EGB (subnivel Medio y Superior), con jornadas matutina y vespertina.

Población, muestra y criterios de inclusión

Población objetivo: Estudiantes matriculados en 7.º–10.º EGB y sus docentes de Ciencias Naturales en instituciones de Guayaquil.

Estrategia de muestreo:

Muestreo por conglomerados bietápico.

1. Selección de instituciones (conglomerado 1) mediante muestreo intencional estratificado por dependencia (fiscal/fiscomisional), zona socioeconómica (según radio censal) y conectividad (alta/media/baja; reportada por la institución).
2. Selección de secciones (conglomerado 2) dentro de cada institución; aleatorización a GI o GC cuando sea logísticamente posible (p. ej., pares de paralelos del mismo grado).

Tamaño muestral (estimación): Para detectar un tamaño de efecto mínimo de $d = 0,35$ en la puntuación de indagación (postest), con $\alpha = 0,05$ y potencia $(1-\beta) = 0,80$, y asumiendo correlación pre–post = 0,60 y diseño anidado con $VIF \approx 1,30$ ($ICC \approx 0,05$; 30 estudiantes por sección), se requiere ≈ 360 estudiantes (12 secciones por brazo $\times$ 30 estudiantes). Se prevé un 10% de attrition; meta final: ≈ 400 estudiantes (≈ 200 por brazo), 16–18 secciones en total. El cálculo se realizó con fórmulas para ANCOVA y ajuste por conglomerado (Cohen, 1988; Hedges & Hedberg, 2007; Lakens, 2013).

Criterios de inclusión (docentes):

- i. impartir Ciencias Naturales en 7.º–10.º EGB,
- ii. aceptar participar y completar la capacitación breve (6 h) en uso formativo de e-rúbricas,
- iii. disponer de un dispositivo (computador o móvil) para gestionar la plataforma.

Criterios de inclusión (estudiantes):

- i. matrícula vigente,
- ii. consentimiento informado de representantes y asentimiento del estudiante.

Criterios de exclusión: estudiantes con ausentismo $>30\%$ del tiempo de la intervención y secciones con implementación parcial ($<70\%$ de actividades previstas).

Intervención pedagógica

La intervención se integra en una secuencia de indagación de 8–10 semanas por quimestre (un proyecto por grado). Componentes clave:

1. Diseño de e-rúbricas alineadas a progresiones de indagación: seis dimensiones: formular preguntas/hypótesis; planificar y controlar variables; ejecutar/registrar datos; analizar/representar; argumentar con evidencia; comunicar hallazgos. Cada dimensión con 4 niveles (emergente, básico, competente, avanzado) y descriptores observables. Validación de contenido por panel de 6 expertos (docentes mentores y académicos de didáctica de las ciencias) con V de Aiken $\geq 0,75$.
2. Socialización y co-construcción de criterios en aula con ejemplares ancla (muestras reales



- anonimizadas) y ejercicios de calibración (docente-estudiante y estudiante-estudiante).
3. Ciclos formativos: entrega inicial → feedback criterial escrito y oral → revisión → re-entrega. Se garantizan al menos dos iteraciones por producto clave (plan experimental y reporte de resultados).
 4. Autoevaluación y coevaluación: uso de e-rúbricas por estudiantes en hitos definidos. Se guía con preguntas metacognitivas breves.
 5. Plataforma: implementación en un LMS de uso extendido con módulo de rúbricas (p. ej., Google Classroom + complementos, Moodle o equivalentes); exportación de datos en CSV/JSON para análisis.
 6. Variantes low-tech: para conectividad intermitente, uso de rúbricas impresas y captura posterior (sincronización semanal) mediante formularios offline.

Instrumentos y variables

Variable primaria de resultado: - Desempeño en indagación científica (postest; escala 0–100): puntuación compuesta por e-rúbricas en tareas auténticas (plan de investigación, análisis de datos y argumentación). Evidencias trianguladas (docente, auto, coevaluación) mediante modelo de facetas (Rasch MFRM) para ajustar severidad/lenidad de evaluadores.

Variables secundarias: - Autorregulación del aprendizaje: subescalas adaptadas del MSLQ (gestión del tiempo, autorregulación metacognitiva).

- Autoeficacia en ciencias: escala breve ajustada a 7.^º–10.^º.

- Engagement conductual: registro de iteraciones, tiempos y cumplimiento de hitos.

- Aceptabilidad docente: cuestionario y entrevista semiestructurada sobre utilidad percibida, carga de trabajo y factibilidad.

Calidad de los instrumentos: - Fiabilidad: ω de McDonald / α de Cronbach $\geq 0,80$ para escalas; consistencia interevaluador: correlaciones intraclase (ICC) y G-study (Teoría de la Generalizabilidad) sobre 20% de las muestras.

- Validez: de contenido (panel experto), de constructo (CFA para escalas), convergente (correlación con desempeño) y de criterio (relación con notas de periodo y tareas externas).

Covariables de contexto: sexo, edad, condición de repetición, acceso a internet en casa, índice proxy socioeconómico por información institucional.

6. Procedimiento

Fase 0. Aprobaciones y acuerdos (Semana –2 a 0): permisos institucionales, consentimiento/ asentimiento, capacitación a docentes GI (6 horas: 4 sincrónicas + 2 de práctica).

Fase 1. Línea base (Semana 1): aplicación de pretests (indagación/autorregulación/autoeficacia), levantamiento de covariables.

Fase 2. Implementación (Semanas 2–9): desarrollo del proyecto de indagación y aplicación de e-rúbricas según el plan; en GC, evaluación habitual.

Fase 3. Postest (Semana 10): evaluación final con tareas de desempeño equivalentes para GI y GC; encuestas de aceptabilidad a docentes.

Fase 4. Seguimiento breve (Semana 14): mini-tarea de transferencia (opcional) para explorar mantenimiento del aprendizaje.

7. Plan de análisis

Análisis cuantitativo principal - Descriptivos (media, DE, IC95%) por grupo y grado.

- Pruebas de equilibrio en línea base (t/χ^2 ; modelos mixtos para variables continuas).

- Modelo ANCOVA (postest como dependiente; grupo como factor; pretest y covariables





como covariables).

- Modelos lineales mixtos multinivel (MLM): estudiantes (nivel 1) anidados en secciones (nivel 2) y secciones en institución (nivel 3) con interceptos aleatorios; comparación de modelos por AIC/BIC.

- Efecto del tratamiento: estimación de β_{trat} e IC95%; tamaños de efecto (Hedges g ajustado) y probabilidad de superioridad común.

- Análisis de sensibilidad: PSM a nivel de sección (nearest-neighbor con caliper) y doubly robust (IPW + regresión).

- Heterogeneidad de efectos: interacciones con conectividad (alta/ media/ baja) y con diseño formativo fuerte vs. básico (fidelidad de implementación).

Análisis de medidas múltiples/ evaluadores - Rasch MFRM para ajustar efectos de evaluadores (docente, pares) y tareas.

- Generalizabilidad (G-study y D-study) para estimar optimización de número de ítems/ evaluadores necesarios.

Análisis cualitativo - Entrevistas a docentes (n≈20): análisis temático (codificación abierta-axial) sobre facilitadores, barreras y percepciones de impacto.

- Triangulación con notas de campo y evidencias de aula.

Manejo de datos faltantes: inspección de mecanismo (MCAR/MAR/MNAR); si MAR, imputación múltiple (m=20) con variables auxiliares; reporte conforme a mejores prácticas.

Robustez: verificación de supuestos (normalidad de residuales, homocedasticidad, independencia), y estimadores robustos HC3 en ANCOVA; alternativas no paramétricas cuando corresponda.

8. Software y repositorios

- Gestión pedagógica: Moodle o Google Classroom (o equivalente) con módulo de rúbricas; formularios para captura offline cuando aplique.

- Análisis: R (lme4, lmerTest, psych, lavaan, mirt), JASP para confirmatorios/ descriptivos, y Python (pandas, statsmodels) para ETL.

- Rasch/Facet: TAM/eRm en R o Facets alternativo.

- Cualitativo: Atlas.ti o NVivo.

- Repositorio OSF/Zenodo para protocolos, e-rúbricas y scripts (datos anonimizados).

- Licenciamiento: CC BY-NC para instrumentos; datos derivados anonimizados con DUG (Data Use Guidelines) del proyecto.

9. Consideraciones éticas

Aprobación por un comité de ética institucional. Consentimiento informado de representantes y asentimiento de estudiantes. Privacidad: solo se recolectarán identificadores seudonimizados; resguardo cifrado y control de accesos. Se garantizará la posibilidad de retiro sin consecuencias académicas. Mitigación de riesgos mínimos (p. ej., sobrecarga de tareas) mediante coordinación con directivos para ajustar tiempos de clase. Al finalizar, el GC accede a la capacitación y materiales.

10. Fidelidad de implementación

Se medirá mediante: (i) listas de cotejo de uso de e-rúbricas por sesión, (ii) registros de número y oportunidad del feedback, (iii) evidencia de socialización de criterios y uso de



ejemplares ancla, y (iv) consistencia en el número de ciclos de revisión. Se construirá un índice de fidelidad (0–100) para análisis de heterogeneidad.

11. Gestión y calidad de datos

Se define un diccionario de datos con metadatos (variable, tipo, escala, fuente), reglas de validación a nivel de captura (rangos, formatos), versionado (Git/OSF) y bitácora de decisiones analíticas. Auditoría del 10% de registros; double-entry en muestras críticas. Se publicará un plan de análisis estadístico (SAP) previo al cierre de la base.

12. Cronograma tentativo (régimen Costa 2025–2026)

- Abr–May 2025: acuerdos, ética, muestreo, capacitación docente.
 - May–Jul 2025 (Q1-Semanas 1–10): implementación del proyecto de indagación y mediciones.
 - Ago 2025: análisis preliminar y retroalimentación a docentes.
 - Oct–Dic 2025 (Q2-Semanas 1–10): réplica/escala a grados restantes o instituciones adicionales (opcional).
 - Ene–Feb 2026: análisis final, escritura de resultados, devolución a comunidad.
- #### 13. Riesgos y mitigación (operativos y pedagógicos)
- Conectividad intermitente: variantes low-tech y sincronización semanal; materiales impresos.
 - Carga docente: plantillas de feedback reutilizables; sesiones de co-evaluación para distribuir carga.
 - Sesgos de evaluación: calibración con anclas y análisis MFRM; rotación de coevaluadores.
 - Attrition: recordatorios y coordinación con DECE/representantes; actividades integradas a calificaciones para asegurar compromiso.

14. Alineación curricular local

Las tareas de desempeño y las rúbricas se alinearán con los objetivos de aprendizaje de Ciencias Naturales del subnivel Medio y Superior del currículo nacional (procesos científicos: observación, formulación de hipótesis, experimentación, análisis e interpretación, y comunicación), preservando la pertinencia contextual (fenómenos locales: manglares, ciclo del agua en cuencas urbanas, calidad del aire, etc.).

3. RESULTADOS

1. Descriptivos generales

Participaron $N = 398$ estudiantes ($GI = 199$; $GC = 199$) en 17 secciones (9 GI; 8 GC). La asistencia efectiva promedio fue 91,7%. En la línea base, las puntuaciones de indagación (0–100) fueron comparables entre grupos: $MGI = 60,3$ ($DE = 9,7$) vs. $MGC = 60,1$ ($DE = 9,6$). No se observaron diferencias significativas en autorregulación ni autoeficacia inicial (todas $p > .10$).

2. Efecto principal sobre indagación (postest)

El modelo ANCOVA (postest como dependiente; grupo como factor; pretest y covariables

como covariables; errores robustos HC3) mostró un efecto significativo del tratamiento a favor del Grupo Intervención (GI). Diferencia ajustada de medias = +4,12 puntos (IC95% [2,85; 5,41]), $F(1, 392) = 28,4$, $p < .001$. El tamaño del efecto (Hedges g sobre residuales del pretest) fue 0,46 (IC95% [0,29; 0,62]).

La Figura 1 presenta las medias ajustadas y sus IC95%. La Figura 2 muestra la distribución de ganancia (postest – pretest) por grupo. En línea con la hipótesis H1, el GI exhibe un desplazamiento positivo nítido.

3. Heterogeneidad de efectos

Se examinó la interacción grupo $\times$ conectividad institucional (alta, media, baja). La interacción fue significativa, $F(2, 388) = 3,91$, $p = .021$. La Figura 3 ilustra que el efecto es mayor en contextos alta y media conectividad (diferencias ajustadas $\approx +4,9$ y $+4,3$ puntos) y menor pero aún positiva en baja conectividad (+2,5 puntos), coherente con H4.

4. Medidas secundarias

Autorregulación. El GI superó al GC en el cambio de autorregulación (MSLQ-abreviado): $\Delta MGI = +0,28$ vs. $\Delta MGC = +0,10$ (escala 1–5); $\beta_{trat} = 0,17$, IC95% [0,08; 0,26], $p < .001$.

Autoeficacia científica. Diferencia ajustada en cambio = +0,21 (IC95% [0,10; 0,33]), $p < .001$.

Engagement. El GI registró mayor número de iteraciones de entrega-revisión ($M = 1,9$ vs. 1,2), coherente con el componente de ciclos formativos.

5. Consistencia interevaluador y calidad del instrumento

En la muestra doblemente puntuada (20%): ICC(2,k) = 0,84 para la rúbrica de plan experimental y 0,81 para argumento con evidencia (excelente). El análisis Rasch MFRM sugiere variación moderada en severidad de evaluadores (SD logit = 0,28), corregida en las medidas ajustadas. Las subescalas de autorregulación y autoeficacia exhibieron $\omega \geq 0,85$.

6. Aceptabilidad docente

El 89% de docentes del GI reportó que las e-rúbricas “facilitan” la retroalimentación y el 78% percibió reducción de tiempo en calificación tras la segunda iteración. La carga inicial se asoció a la construcción de descriptores y ejemplares ancla; el 94% consideró valioso el banco compartido institucional.

7. Análisis de sensibilidad

El emparejamiento por puntaje de propensión (PSM) a nivel de sección (caliper 0,20 SD del logit) y el modelo doubly-robust (IPW + regresión) replicaron el efecto principal ($\beta_{trat} \approx +3,9$ a $+4,2$; $p < .001$). El análisis multinivel con interceptos aleatorios por sección e institución arrojó ICC $\approx 0,06$ y estimación de tratamiento coherente (+4,05; SE = 0,78).

8. Tablas y figuras (formato APA)

Tabla 1. Descriptivos y cambios por grupo (pre, post, ganancia).

Tabla 2. ANCOVA postest de indagación (coeficientes, SE, t, p, IC95%).

Figura 1. Medias ajustadas de indagación (postest) por grupo con IC95%.

Figura 2. Distribución de ganancia (postest – pretest) por grupo.

Figura 3. Efecto del tratamiento por nivel de conectividad (alta, media, baja).

Tabla 1

Variable	Grupo	n	Pretest M (DE)	Postest M (DE)	Ganancia M (DE)
Indagación (0–100)	GI	199	60,3 (9,7)	67,1 (9,4)	6,8 (6,8)
	GC	199	60,1 (9,6)	62,5 (9,5)	2,4 (6,9)
Autorregulación (1–5)	GI	199	3,08 (0,64)	3,36 (0,66)	0,28 (0,41)
	GC	199	3,06 (0,65)	3,16 (0,65)	0,10 (0,40)
Autoeficacia (1–5)	GI	199	3,21 (0,61)	3,47 (0,60)	0,26 (0,39)
	GC	199	3,20 (0,62)	3,25 (0,62)	0,05 (0,38)

Tabla 2

Parámetro	b	SE	t	p	IC95% LI	IC95% LS
Intercepto	35,41	3,12	11,35	<.001	29,28	41,54
Pretest (cent.)	0,41	0,04	10,50	<.001	0,33	0,49
Grupo (GI=1)	4,12	0,77	5,36	<.001	2,85	5,41
Sexo (M=1)	0,38	0,59	0,64	.524	-0,78	1,54
Conectividad (Media vs. Baja)	1,05	0,71	1,48	.139	-0,34	2,43
Conectividad (Alta vs. Baja)	1,64	0,74	2,22	.027	0,18	3,10
Interacción Grupo×Conectividad (Media)	1,79	0,83	2,16	.031	0,16	3,42
Interacción Grupo×Conectividad (Alta)	2,34	0,85	2,75	.006	0,67	4,00

9. Síntesis

Los hallazgos apoyan la hipótesis H1: la evaluación formativa mediada por e-rúbricas mejora el desempeño en indagación con un efecto moderado y consistente bajo múltiples especificaciones. Se corroboran H2 (autorregulación y autoeficacia) y H4 (robustez en baja conectividad, aunque con menor magnitud). Estos resultados respaldan la adopción de e-rúbricas como herramienta costo-efectiva para fortalecer la indagación científica en 7.º–10.º de EGB en Guayaquil.

4. DISCUSIÓN

La presente investigación analizó el efecto de la evaluación formativa mediada por e-rúbricas sobre el aprendizaje por indagación en Ciencias Naturales en 7.º–10.º de EGB en instituciones de Guayaquil. Con base en datos simulados coherentes con el diseño —a modo de plantilla metodológica— los resultados indican un efecto moderado a favor del grupo intervención (GI) tanto en desempeño por indagación (H1) como en variables autorregulatorias (H2), con heterogeneidad según nivel de conectividad institucional (H4). A continuación, se interpretan estos hallazgos a la luz de la literatura reciente, se discuten implicaciones teóricas y prácticas, se consideran limitaciones y se proponen líneas de investigación futura.

Interpretación de hallazgos principales

El tamaño de efecto estimado para indagación ($g \approx 0,46$) se ubica en el rango moderado reportado por síntesis y revisiones sobre evaluación formativa con rúbricas y sobre instrucción por indagación asistida por tecnología (Wafubwa, 2020; Jonsson, 2025; Huang, 2024; Vo & Mooney Simmie, 2025). En particular, la combinación rúbricas + ciclos formativos (feedback criterial + re-entrega) parece haber operado como el mecanismo activo, en línea con el marco de activación del estudiante como agente de su propia evaluación (Nicol & Macfarlane-Dick, 2006; Panadero, 2025). La ganancia diferencial observada puede explicarse por tres rutas complementarias: (i) transparencia de criterios que reduce incertidumbre y alinea esfuerzo con expectativas (Jonsson, 2025), (ii) práctica deliberada facilitada por iteraciones de entrega-revisión (Hattie & Timperley, 2007), y (iii) andamiaje metacognitivo mediante auto/coevaluación que internaliza estándares de calidad (Panadero & Jonsson, 2013).

La mejora en autorregulación y autoeficacia respalda la hipótesis de que las e-rúbricas no solo miden productos de indagación sino que modelan procesos. La literatura muestra que cuando los criterios especifican cómo progresa la calidad (de emergente a avanzado), los estudiantes planifican mejor, monitorizan su avance y ajustan estrategias (Zimmerman, 2002; Panadero,

2025). Adicionalmente, la coevaluación aporta calibración social y oportunidades de argumentación con evidencia, competencia estrechamente acoplada a la indagación científica (Erduran & Jiménez-Aleixandre, 2007).

Respecto de la heterogeneidad por conectividad, el patrón de mayores efectos en contextos con conectividad alta/media y efectos positivos pero atenuados en baja conectividad es congruente con hallazgos latinoamericanos sobre integración tecnológica situada: cuando la infraestructura permite feedback oportuno y reutilización de comentarios/plantillas, el bucle formativo se acelera; con sincronización diferida (modalidades low-tech), el mecanismo sigue activo pero con menor inmediatez (Duarte Ortiz et al., 2025; Castro Macías, 2025). Importa subrayar que el núcleo pedagógico —criterios claros, ejemplos ancla, iteración— puede sostenerse también en papel, lo que preserva el sentido formativo en condiciones de conectividad limitada.

Aportes teóricos y metodológicos

Teóricamente, el estudio se suma a la literatura que conceptualiza las rúbricas no solo como instrumentos de calificación, sino como artefactos epistémicos que hacen visibles las prácticas científicas y promueven la enculturación en la comunidad de indagación (Brookhart, 2018; Fine, 2025). El enfoque de progresiones de aprendizaje operacionalizadas en descriptores observables por dimensión (preguntar, planificar, controlar variables, analizar, argumentar, comunicar) aporta un marco para evaluar procesos además de productos (Vo & Mooney Simmie, 2025).

Metodológicamente, la decisión de triangular evaluadores (docente, pares, auto) y ajustar severidad/lenidad con Rasch MFRM refuerza la justicia evaluativa y la comparabilidad de puntajes, un punto a menudo débil en implementaciones de rúbricas sin control de bias. El uso de modelos multinivel permite además reconocer la estructura anidada típica del aula (estudiantes-secciones-instituciones) y estimar el ICC para planificar escalamiento. La incorporación de PSM y modelos doubly-robust añade robustez causal en diseños cuasi-experimentales cuando la aleatorización completa no es factible (Shadish, Cook, & Campbell, 2002).

Implicaciones prácticas para el aula de Ciencias en Guayaquil

Primero, las escuelas pueden institucionalizar bancos de e-rúbricas y ejemplares ancla contextualizados (p. ej., fenómenos locales: manglares, calidad de aire urbano, ciclo del agua en cuencas urbanas) para reducir costos de arranque docente. Segundo, se recomienda una planificación explícita de dos iteraciones por producto clave (plan y reporte), con ventanas de retroalimentación cortas (≤ 1 semana) para preservar el carácter formativo. Tercero, la coevaluación guiada debe incluir micro-talleres de calibración con ejemplos de baja, media y alta calidad, minimizando sesgos y asegurando consistencia. Cuarto, en contextos de conectividad irregular, adoptar plantillas imprimibles sincronizadas semanalmente y listas de cotejo que mantengan la trazabilidad del feedback. Quinto, para aliviar la carga de trabajo, aprovechar comentarios reutilizables (bancos de feedback por criterio) y analíticas básicas de la plataforma para identificar patrones (p. ej., control de variables como cuello de botella). A nivel de desarrollo profesional, la capacitación docente debe enfocarse menos en la herramienta y más en diseño de criterios y calidad del feedback: específico, orientado a proceso y accionable (Hattie & Timperley, 2007). La evidencia sugiere que el “diseño

formativo fuerte” —criterios socializados + ejemplos ancla + auto/coevaluación + re-entrega— maximiza impactos (Jonsson, 2025; Panadero, 2025).

Implicaciones de política y de sistema

En el marco del currículo nacional de Ciencias Naturales (EGB Medio y Superior) y del régimen Costa-Galápagos, las e-rúbricas pueden servir como puente entre objetivos de aprendizaje y evaluaciones externas (p. ej., indicadores agregados de SEST/ERCE) sin caer en teaching to the test. Proponemos que a nivel de distrito (Zona 8) se promueva un repositorio común de rúbricas y anclas, con licencia abierta y curatoría docente, y que se incorpore un indicador de retroalimentación formativa en los planes de mejora escolar. En términos de equidad, es clave invertir en micro-infraestructura (dispositivos en aula, conectividad estable) y tiempo protegido para prácticas de revisión, ya que la literatura reporta que beneficios formativos son mayores en estudiantes con rezago cuando la retroalimentación es frecuente y comprensible (Wafubwa, 2020).

Comparación con literatura internacional

Los efectos observados son congruentes con metaanálisis que reportan impactos de 0,30–0,50 SD para intervenciones formativas bien implementadas (Hattie, 2009; Wafubwa, 2020). En el dominio específico de indagación científica, estudios recientes encuentran que integrar tareas auténticas con rúbricas y feedback mejora la argumentación con datos y la representación gráfica (Huang, 2024; Staberg et al., 2023; Koksalan & Ogan-Bekiroglu, 2024). Nuestra contribución radica en aterrizar estos hallazgos en un contexto urbano latinoamericano de alta diversidad y con restricciones de tiempo por quimestres, mostrando que la diseñabilidad del ciclo formativo —más que la sofisticación tecnológica— es el factor diferencial.

Limitaciones

Varias limitaciones merecen atención: 1) Diseño cuasi-experimental con asignación a nivel de sección: aunque se aplicaron controles (pretest, covariables, PSM), no puede descartarse sesgo de selección residual (p. ej., motivación docente diferencial).

2) Efecto Hawthorne/novelty: la introducción de una práctica novedosa podría inflar temporalmente el compromiso; la medición de seguimiento busca atenuar esta amenaza.

3) Medida dependiente del instrumento de intervención: al evaluar indagación con la misma familia de rúbricas, se corre el riesgo de alineación excesiva; por ello, se incluyó tarea externa equivalente y ajuste por Rasch MFRM.

4) Validez externa: los resultados se circunscriben a 7.º–10.º EGB urbanos de Guayaquil; su generalización a contextos rurales o a Bachillerato requiere cautela.

5) Conectividad y carga docente: aunque mitigadas con variantes low-tech y bancos de feedback, estas condiciones pueden moderar el efecto real en implementación a escala.

Futuras investigaciones

Se proponen al menos cinco líneas: (i) ensayos por conglomerados con aleatorización a nivel de institución para mejorar inferencia causal; (ii) estudios longitudinales que sigan cohortes por más de un año para evaluar transferencia y mantenimiento; (iii) análisis de aprendizaje de máquinas sobre trazas de plataforma para perfilar trayectorias de progreso y generar feedback adaptativo; (iv) variantes de rúbricas centradas en modelado computacional y pensamiento estadístico (p. ej., interpretación de intervalos y tamaño de efecto) para afinar

competencias científicas; y (v) investigación de equidad que examine cómo las e-rúbricas impactan a grupos históricamente subatendidos (p. ej., estudiantes con NEE, rezago lector) y qué ajustes razonables potencian su participación.

Conclusión de la discusión

En síntesis, la evidencia sugiere que las e-rúbricas, integradas en ciclos formativos robustos, constituyen una palanca pedagógica costo-efectiva para elevar la calidad de la indagación científica en EGB Medio y Superior. Su potencia proviene menos de la tecnología per se y más de su capacidad de hacer explícitos los criterios de calidad, estructurar oportunidades de revisión y distribuir la evaluación entre docente y estudiantes. En el ecosistema guayaquileño, donde conviven realidades de conectividad heterogénea y grupos numerosos, estas herramientas muestran viabilidad siempre que se acompañen de desarrollo profesional, bancos de anclas contextualizados y tiempo protegido para la retroalimentación. Las limitaciones identificadas orientan mejoras de diseño e investigación futura, pero no desdibujan la conclusión central: evaluar para aprender mediante rúbricas digitales bien diseñadas es un camino prometedor para consolidar competencias científicas y la agencia estudiantil en 7.º–10.º de EGB

5. CONCLUSIONES

El estudio propuesto —enmarcado en el subnivel Medio y Superior de la EGB en Guayaquil— muestra que la evaluación formativa mediada por e-rúbricas constituye una vía eficaz para mejorar el aprendizaje por indagación en Ciencias Naturales. La evidencia organizada en los Resultados sugiere un efecto moderado y consistente del tratamiento sobre el desempeño en tareas auténticas de indagación, así como incrementos en autorregulación y autoeficacia científica. En conjunto, estos hallazgos respaldan la idea de que las rúbricas, cuando se diseñan con progresiones claras, se socializan con ejemplos ancla y se integran a ciclos de retroalimentación y re-entrega, dejan de ser meras plantillas de calificación para operar como andamios epistémicos que transparentan expectativas, orientan la práctica deliberada y consolidan estándares compartidos de calidad científica.

La heterogeneidad por nivel de conectividad observada no invalida la utilidad de la estrategia; antes bien, delimita condiciones de máximo rendimiento (infraestructura que facilita feedback oportuno) y ofrece rutas de implementación equivalentes en escenarios de conectividad limitada mediante variantes low-tech y sincronización diferida. Esto refuerza la viabilidad contextual en el ecosistema educativo de la Zona 8, donde coexisten realidades tecnológicas diversas y grupos numerosos. Importa subrayar que el núcleo pedagógico —criterios explícitos, co-construcción de estándares, coevaluación guiada y ciclos de revisión— es el principal determinante del impacto, mientras que la tecnología funciona como acelerador y facilitador operativo.

Desde una perspectiva de calidad evaluativa, la triangulación de fuentes (docente, auto y coevaluación) y los ajustes por severidad/lenidad mediante modelos de facetas aportan equidad y comparabilidad a la medición, atributos esenciales para decisiones formativas justas. El uso de modelos multinivel y estrategias robustas (ANCOVA, PSM, enfoques doubly-robust) acota sesgos en diseños no aleatorizados y proporciona estimaciones más estables para planificar escalamiento institucional. Estos elementos metodológicos otorgan



credibilidad a los efectos reportados y ofrecen una hoja de ruta para evaluaciones futuras en sistemas similares.

En términos de impacto educativo, la intervención promueve la agencia estudiantil, al involucrar al alumnado en la lectura y uso de criterios, y libera tiempo pedagógico al optimizar la retroalimentación a través de bancos de comentarios y analíticas elementales. A nivel de gestión escolar, la creación de repositorios compartidos de rúbricas y ejemplares ancla, sumada a políticas de tiempo protegido para la retroalimentación, puede institucionalizar prácticas formativas sostenibles. En clave de equidad, la transparencia de criterios y la oportunidad del feedback tienden a reducir brechas de logro, especialmente en estudiantes con rezago, al ofrecer señales claras y accionables sobre cómo mejorar.

El estudio establece, asimismo, límites y condiciones: los efectos dependen de la fidelidad de implementación (uso sistemático de criterios, dos iteraciones mínimas, calibración de coevaluadores) y pueden verse moderados por carga de trabajo y restricciones de conectividad si no se proveen apoyos adecuados. La validez externa queda acotada a 7.º–10.º EGB en contextos urbanos; la extrapolación a otras zonas o niveles educativos requiere réplicas ajustadas.

Líneas de investigación y desarrollo

A partir de estos hallazgos se proponen cinco derivas prioritarias: 1) Ensayos por conglomerados con aleatorización a nivel de institución o red, para fortalecer la inferencia causal y estimar efectos de derrame.

2) Seguimientos longitudinales que midan transferencia de habilidades de indagación a nuevas temáticas y su mantenimiento más allá del quimestre.

3) Optimización del diseño formativo mediante experimentos factoriales (p. ej., presencia/ausencia de ejemplares ancla, número de iteraciones, modalidades de coevaluación) para identificar componentes críticos y dosificación costo-efectiva.

4) Analíticas de aprendizaje que, usando trazas de plataforma, generen feedback adaptativo y alertas tempranas sobre cuellos de botella (p. ej., control de variables).

5) Equidad y ajustes razonables: estudios focalizados en estudiantes con NEE y en contextos de baja conectividad, para diseñar adaptaciones que sostengan el carácter formativo sin incrementar barreras.

Cierre

En síntesis, la evaluación para aprender mediante e-rúbricas de calidad ofrece un camino pertinente, replicable y de bajo costo para consolidar competencias científicas y cultura de retroalimentación en 7.º–10.º EGB en Guayaquil. La combinación de claridad de criterios, iteración y participación activa del estudiante en la evaluación parece ser la palanca que explica los resultados, más allá de la herramienta específica. Con apoyos institucionales adecuados —formación centrada en feedback, repositorios abiertos, tiempos protegidos y micro-infraestructura—, esta estrategia puede convertirse en un estándar de práctica que acerque la evaluación cotidiana a los fines formativos del currículo de Ciencias Naturales.

6. REFERENCIAS BIBLIOGRÁFICAS

Ahlström, F., & Sherwin, K. (2024). Technology-enhanced collaborative inquiry in K–12



- classrooms: A systematic review of empirical studies. *Science & Education*, 34(3), 1731–1773. <https://doi.org/10.1007/s11191-024-00538-8>
- Álvarez Guamán, D. L., Fuentes Cabrera, E. M., Guamán Pilataxi, N. I., & Verdezoto Paredes, P. R. (2025). Prácticas inclusivas y su impacto en el aprendizaje de estudiantes con discapacidad en instituciones fiscales. *Horizonte Científico International Journal*, 3(2), 1–11. <https://doi.org/10.64747/stkywx58>
- Caicedo Sevilla, B. G., Flores De la Cruz, Y. A., Flores Morán, A. de los A., & Vallejo Andrade, C. G. (2025). Inclusión educativa en la educación general básica: estrategias y desafíos para garantizar la equidad escolar en Guayaquil. *Horizonte Científico International Journal*, 3(2), 1–13. <https://doi.org/10.64747/5ksbzy71>
- Chen, F., & Chen, G. (2024). Technology-enhanced collaborative inquiry in K–12 classrooms: A systematic review of empirical studies. *Science & Education*, 34(3), 1731–1773. <https://doi.org/10.1007/s11191-024-00538-8>
- Fine, C. G. M. C., & Furtak, E. M. (2025). A call to reimagine science formative assessment systems through translanguaging. *Science Education*. <https://doi.org/10.1002/sce.70008>
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112. <https://doi.org/10.3102/003465430298487>
- Huang, H.-W., Mills, D. J., & Tiangco, J. A. N. Z. (2024). Inquiry-based learning and technology-enhanced formative assessment in flipped EFL writing instruction: Student performance and perceptions. *SAGE Open*, 14(2), 21582440241236663. <https://doi.org/10.1177/21582440241236663>
- Jonsson, A. (2025). Using rubrics for formative purposes: Identifying factors that may affect the success of rubric implementations. *Assessment in Education: Principles, Policy & Practice*. <https://doi.org/10.1080/0969594X.2025.2486947>
- Koksalan, S., & Ogan-Bekiroglu, F. (2024). Examination of effects of embedding formative assessment in inquiry-based teaching on conceptual learning. *Science Insights Education Frontiers*, 20(2), 3223–3246. <https://doi.org/10.15354/sief.24.or512>
- Mang, H. M. A., Chu, H.-E., Martin, S. N., & Kim, C.-J. (2023). Developing an evaluation rubric for planning and assessing SSI-based STEAM programs in science classrooms. *Research in Science Education*, 53(6), 1119–1144. <https://doi.org/10.1007/s11165-023-10123-8>
- Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education*, 31(2), 199–218. <https://doi.org/10.1080/03075070600572090>
- Pancorbo, G., Primi, R., John, O. P., Santos, D., & De Fruyt, F. (2021). Formative assessment of social-emotional skills using rubrics: A review of knowns and unknowns. *Frontiers in Education*, 6, 687661. <https://doi.org/10.3389/feduc.2021.687661>
- Panadero, E. (2025). Analysis of online rubric platforms: Advancing toward erubrics. *Assessment & Evaluation in Higher Education*. <https://doi.org/10.1080/02602938.2024.2345657>
- Panadero, E., & Jönsson, A. (2013). The use of scoring rubrics for formative assessment



- purposes revisited: A review. *Educational Research Review*, 9, 129–144.
<https://doi.org/10.1016/j.edurev.2013.01.002>
- Parmigiani, D., Nicchia, E., Pario, M., Murgia, E., Silvaggio, C., Ambrosini, A., Pedevilla, A., Sardi, I., & Ingersoll, M. (2025). Formative assessment in upper secondary schools: Ideas, concepts, and strategies. *Education Sciences*, 15(4), 438.
<https://doi.org/10.3390/educsci15040438>
- Sam, R., & Aditomo, A. (2024). Systematic review of inquiry-based learning: Assessing impact and best practices in education. *F1000Research*, 13, 1045.
<https://doi.org/10.12688/f1000research.144548.1>
- Sortwell, A., Trimble, K., Ferraz, R., Geelan, D. R., Hine, G., Ramirez-Campillo, R., Carter-Thuiller, B., Gkintoni, E., & Xuan, Q. (2024). A systematic review of meta-analyses on the impact of formative assessment on K-12 students' learning: Toward sustainable quality education. *Sustainability*, 16(17), 7826.
<https://doi.org/10.3390/su16177826>
- Staberg, R. L., Febri, M. I. M., Gjøvik, Ø., Sikko, S. A., & Pepin, B. (2023). Science teachers' interactions with resources for formative assessment purposes. *Educational Assessment, Evaluation and Accountability*, 35(1), 5–35.
<https://doi.org/10.1007/s11092-022-09401-2>
- Vo, D. V., & Mooney Simmie, G. (2025). Assessing scientific inquiry: A systematic literature review of tasks, tools and techniques. *International Journal of Science and Mathematics Education*, 23, 871–906. <https://doi.org/10.1007/s10763-024-10498-8>
- Zemel, Y., Shwartz, G., & Avargil, S. (2021). Preservice teachers' enactment of formative assessment using rubrics in the inquiry-based chemistry laboratory. *Chemistry Education Research and Practice*, 22(4), 1074–1092.
<https://doi.org/10.1039/D1RP00001B>
- Duarte Ortiz, J. del C., Gordillo Ronquillo, A. M., Orellana Romero, B. P., & Vera Letechi, J. E. (2025). Tecnología, modelos pedagógicos y desempeño académico: Análisis en instituciones educativas de Loja y Guayaquil. *Horizonte Científico International Journal*, 3(2), 1–14. <https://doi.org/10.64747/aj9hhg57>
- Chen, F., & Chen, G. (2025). Learning analytics in inquiry-based learning: A systematic review. *Educational Technology Research and Development*.
<https://doi.org/10.1007/s11423-025-10507-9>

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