Sage Sphere in Artificial Intelligence

ISSN: 3103-1307

Vol. 4 Núm.. 1 (2026)

SSAI-Vol.4. N1. 012

Inteligencia artificial aplicada a la evaluación inclusiva del aprendizaje en educación general básica de Ecuador

Artificial intelligence applied to the inclusive assessment of learning in basic general education in Ecuador

Autores:

Virginia Yamilet Mendoza Loor
Investigadora independiente, Quito, Ecuador
jamiletmendoza@outlook.com
https://orcid.org/0009-0003-8178-5267

Fabiana Nairoby Moncayo Quirola
Investigadora independiente, Quito, Ecuador
fimoncayoq@uteq.edu.ec
https://orcid.org/0000-0003-3727-3304

Jessica Viviana Quirola Mera
Investigadora independiente, Quito, Ecuador
jessicaquirola1989@gmail.com
https://orcid.org/0009-0002-8858-2561

Derian Santiago Mendoza Zambrano
Investigador independiente, Quito, Ecuador
dmendozaz4@uteq.edu.ec
https://orcid.org/0000-0003-4880-9098

Emely Jamilet Cárdenas Posligua
Investigadora independiente, Quito, Ecuador
cemely483@gmail.com
https://orcid.org/0000-0001-8682-4147

Autor de correspondencia: *Virginia Yamilet Mendoza Loor*, jamiletmendoza@outlook.com

Enviado: 25-02-2026 **Revisado:** 02-10-2026
Aceptado: 09-11-2026 **Publicado:** 10-11-2026

Cómo citar este artículo:

Mendoza Loor, V. Y., Moncayo Quirola, F. N., Quirola Mera, J. V., Mendoza Zambrano, D. S., & Cárdenas Posligua, E. J. (2026). Artificial intelligence applied to the inclusive assessment of learning in basic general education in Ecuador. *Sage Sphere in Artificial Intelligence*, *4*(1), 1-18. https://doi.org/10.63688/92yma834

© 2026; Los autores. Este es un artículo en acceso abierto, distribuido bajo los términos de una licencia Creative Commons (https://creativecommons.org/licenses/by/4.0) que permite el uso, distribución y reproducción en cualquier medio, siempre que la obra original sea correctamente citada.

Sage Sphere in Artificial Intelligence

ISSN: 3103-1307

Vol. 4 Núm.. 1 (2026)

RESUMEN

La incorporación de inteligencia artificial (IA) en la evaluación escolar plantea oportunidades para personalizar la retroalimentación, analizar respuestas abiertas y adaptar los apoyos a estudiantes con necesidades educativas diversas. El objetivo fue examinar aplicaciones, alcances y riesgos de la IA en la evaluación inclusiva del aprendizaje y discutir su pertinencia para la Educación General Básica (EGB) ecuatoriana. Se efectuó una revisión sistematizada de alcance focalizado, organizada con categorías de PRISMA 2020. Una búsqueda dirigida en SciELO, portales de revistas científicas y registros académicos abiertos permitió conformar un conjunto de 26 publicaciones candidatas; se examinaron títulos, resúmenes y metadatos, se excluyeron 14 registros y se sintetizaron 12 estudios empíricos. Se identificaron aplicaciones de calificación automática, informes formativos, formulación de objetivos individualizados y evaluación de intervenciones adaptativas. Investigaciones sobre respuestas de ciencias mostraron rendimientos variables dependientes de las rúbricas y supervisión humana; otras evidenciaron mejoras en el aprendizaje o en la precisión de objetivos de apoyo, con muestras y diseños heterogéneos. El estudio ecuatoriano identificado ofrece evidencia preliminar de personalización, pero no demuestra una implantación generalizada en EGB. Se concluye que la IA puede complementar una evaluación más accesible y diferenciada, siempre que existan criterios de validez, ajustes razonables, protección de datos y decisiones pedagógicas docentes.

Palabras clave: inteligencia artificial; evaluación inclusiva; educación básica; evaluación formativa; Ecuador.

ABSTRACT

Artificial intelligence (AI) in school assessment creates opportunities to personalize feedback, analyze open-ended responses, and adapt support for learners with diverse educational needs. This paper examines AI applications, benefits, limitations, and their implications for inclusive assessment in Ecuadorian General Basic Education. A focused, structured literature review was conducted using PRISMA 2020 categories as a reporting framework. A targeted search across SciELO, journal platforms, and openly accessible academic records produced a set of 26 candidate publications. Following title, abstract, and publisher-metadata screening, 14 records were excluded and 12 empirical studies were selected for qualitative synthesis. The evidence addressed automated scoring, formative reports, individualized education planning, and outcomes of adaptive interventions. Studies involving written science responses showed that automated scoring performance varied by item complexity, rubric design, and human oversight. Other studies reported improvements in academic measures or individualized goal-writing quality, although sample sizes, contexts, and research designs differed considerably. A study conducted in Ecuador provided preliminary evidence on personalized learning, but does not establish nationwide implementation in basic education. Overall, AI may support more accessible and differentiated assessment when combined with sound validity criteria, reasonable accommodations, data protection, and continued teacher judgment.

Keywords: artificial intelligence; inclusive assessment; basic education; formative assessment; Ecuador.

Doi: https://doi.org/10.63688/92yma834

ISSN: 3103-1307
Vol. 4 Núm.. 1 (2026)

1. INTRODUCCIÓN

La evaluación del aprendizaje cumple funciones pedagógicas que exceden la asignación de calificaciones. Permite interpretar el progreso, detectar necesidades de apoyo, ofrecer retroalimentación y modificar la enseñanza. En aulas de Educación General Básica (EGB), donde conviven ritmos, trayectorias, formas de comunicación y condiciones de acceso diversas, evaluar con justicia implica obtener evidencias válidas sin confundir la dificultad del instrumento con el aprendizaje efectivamente alcanzado. La inteligencia artificial (IA) introduce nuevos recursos para analizar respuestas, sugerir actividades y elaborar informes; sin embargo, automatizar una tarea evaluativa no equivale, por sí solo, a hacerla inclusiva.

Las aplicaciones de IA educativa comprenden aprendizaje automático, procesamiento del lenguaje natural, sistemas de tutoría y, de forma más reciente, modelos generativos. González-Calatayud et al. (2021) identificaron en su revisión usos de IA para la evaluación formativa y la calificación automatizada, junto con debilidades en la fundamentación pedagógica de diversos trabajos. Martínez-Comesaña et al. (2023), al revisar investigaciones de primaria y secundaria, describieron aplicaciones en predicción del rendimiento, corrección de respuestas y análisis de procesos educativos. Estas contribuciones demuestran que la evaluación automatizada constituye un campo de estudio concreto, aunque sus resultados no deben extrapolarse sin atender a las condiciones de cada alumnado.

La dimensión inclusiva exige una segunda pregunta: ¿para quién y bajo qué condiciones funcionan las herramientas? Melo-López et al. (2025), en una revisión desarrollada por investigadoras ecuatorianas, y Pagliara et al. (2024) muestran que la accesibilidad, la personalización y los apoyos a estudiantes con discapacidad aparecen como objetivos recurrentes. Li et al. (2025) subrayan que el potencial de la IA inclusiva depende de las tecnologías utilizadas, los resultados efectivamente medidos y los factores que condicionan su adopción docente. La inclusión, por tanto, no puede reducirse a la existencia de una plataforma digital o a una supuesta neutralidad del algoritmo.

La discusión cobra especial importancia en América Latina. Pajuelo-Abanto et al. (2026) examinaron tensiones entre personalización y estandarización, innovación y brechas tecnológicas. Santillán Martínez et al. (2026) argumentaron que los mecanismos algorítmicos pueden reproducir desigualdades sociales si se aplican sin criterios de justicia educativa. En

Doi: https://doi.org/10.63688/92yma834

ISSN: 3103-1307
Vol. 4 Núm.. 1 (2026)

Ecuador, estas advertencias deben relacionarse con la conectividad heterogénea, la disponibilidad de dispositivos, el tiempo docente y la diversidad sociolingüística de las aulas, sin asumir que todas las instituciones enfrentan idénticas condiciones.

Existe, además, una diferencia entre estudios que prueban la precisión de sistemas automáticos, investigaciones que emplean IA para apoyar el diseño de objetivos individualizados y experiencias que comparan resultados de aprendizaje antes y después de una intervención. Los tres tipos de evidencia son relevantes para una evaluación inclusiva, pero no miden el mismo fenómeno. Una herramienta puede clasificar respuestas con precisión y, aun así, producir retroalimentación inadecuada; otra puede facilitar la planificación docente sin haber demostrado mejoras de aprendizaje a largo plazo.

El objetivo de este artículo es analizar las aplicaciones, beneficios y limitaciones de la IA asociadas a la evaluación inclusiva del aprendizaje en la educación escolar, y discutir las condiciones bajo las cuales podrían ser pertinentes para la EGB de Ecuador. Se plantean tres preguntas: (a) ¿qué funciones evaluativas de IA han sido estudiadas empíricamente en niveles escolares comparables?; (b) ¿qué evidencia existe sobre adaptación, retroalimentación y apoyos a la diversidad?; y (c) ¿qué condiciones éticas, técnicas y pedagógicas deben considerarse antes de transferir dichas experiencias al contexto ecuatoriano?

2. METODOLOGÍA

Diseño y alcance

Se realizó una revisión sistematizada de la literatura con síntesis temática cualitativa. La selección y su reporte siguieron las fases de identificación, cribado, elegibilidad e inclusión descritas por PRISMA 2020 (Page et al., 2021), y se tomaron como referencia las recomendaciones de PRISMA-S para describir las fuentes de información (Rethlefsen et al., 2021). La unidad de análisis fue el estudio publicado, con atención a los resultados empíricos relacionados con la evaluación escolar y la inclusión educativa.

La búsqueda tuvo un alcance focalizado y se desarrolló mediante consultas dirigidas en SciELO, plataformas editoriales y registros académicos abiertos, con publicaciones localizadas hasta el 10 de noviembre de 2026. Se organizaron los registros candidatos y se

Doi: https://doi.org/10.63688/92yma834

ISSN: 3103-1307
Vol. 4 Núm.. 1 (2026)

aplicaron criterios de pertinencia temática, nivel educativo y diseño de investigación para seleccionar los estudios de la síntesis.

Pregunta de revisión y estrategia de localización

La pregunta orientadora fue: ¿qué aplicaciones empíricamente evaluadas de IA pueden favorecer, limitar o modificar procesos de evaluación inclusiva en estudiantes de educación primaria y secundaria inferior, y qué implicaciones tienen para EGB de Ecuador? Se consideró la EGB ecuatoriana como referente de transferencia, con prioridad para estudios nacionales y para edades escolares comparables. Las ecuaciones se estructuraron mediante términos para inteligencia artificial (artificial intelligence, machine learning, generative AI), evaluación (assessment, formative feedback, scoring, evaluación formativa), inclusión (inclusive education, disability, special educational needs) y nivel (primary school, middle school, educación básica).

Tabla 1.

Matriz de estrategia documental y elegibilidad

Componente	Definición aplicada	Alcance y criterio de registro
Periodo	2020–10 de noviembre de 2026	Antecedentes previos empleados solo para fundamentación.
Fuentes	SciELO; plataformas editoriales; registros académicos públicos	Consultas temáticas y contraste de datos bibliográficos.
Términos	AI OR artificial intelligence OR inteligencia artificial; assessment OR evaluación OR feedback; inclusive OR special needs; primary OR middle school	Expresiones combinadas y adaptadas a las interfaces consultadas.
Incluidos	Investigaciones empíricas de educación escolar o evaluación de instrumentos destinados a población escolar	Aplicación de IA + resultado evaluativo, diagnóstico, retroalimentación o apoyo individualizado.
Excluidos	Revisiones previas, trabajos conceptuales, población universitaria, técnica sin IA aplicada o formato no compatible	Los estudios excluidos del corpus principal pueden citarse como antecedentes.

Doi: https://doi.org/10.63688/92yma834

Componente	Definición aplicada	Alcance y criterio de registro
Síntesis	Análisis narrativo de categorías, resultados, limitaciones y transferibilidad	Síntesis temática cualitativa por heterogeneidad de resultados y diseños.

Nota. La matriz presenta las fuentes, términos y criterios empleados para organizar la revisión documental.

Selección, extracción y valoración de evidencia

Se conformó un registro manual de 26 publicaciones candidatas únicas. Diez se retiraron en el cribado inicial por corresponder a revisiones secundarias, análisis bibliométricos o trabajos eminentemente teóricos. Dieciséis registros continuaron a evaluación de elegibilidad basada en la información metodológica y de resultados disponible; cuatro se excluyeron por tratarse de una prueba adaptativa sin IA aplicada, una contribución de congreso no incorporada al corpus de artículos, un reporte de implementación sin evaluación de resultados comparables y un estudio con población universitaria. Los 12 restantes constituyeron el corpus empírico descrito en resultados.

Para cada estudio se registraron autores, año, país, nivel educativo, población o unidad analítica, tecnología, finalidad evaluativa, diseño, resultados explícitos y limitaciones. Se utilizaron criterios de apreciación crítica inspirados en el Mixed Methods Appraisal Tool (MMAT; Hong et al., 2018), especialmente claridad de la pregunta, pertinencia del diseño, transparencia de la muestra, calidad de las medidas y suficiencia del reporte.

El análisis se organizó en cuatro ejes: evaluación automática de respuestas e instrumentos; retroalimentación formativa y monitoreo; individualización de objetivos y apoyos; y condiciones de inclusión, validez y transferencia. La comparación temática permitió interpretar hallazgos procedentes de diseños heterogéneos y diferenciar las validaciones algorítmicas de las intervenciones educativas.

3. RESULTADOS

Flujo de selección y características de las publicaciones

El cribado del conjunto dirigido comprendió 26 registros únicos. Tras 10 exclusiones durante la primera revisión y cuatro al examinar pertinencia metodológica, se seleccionaron 12 trabajos empíricos. La Figura 1 presenta las fases de identificación, depuración, cribado, elegibilidad e inclusión de las publicaciones analizadas.

ISSN: 3103-1307
Vol. 4 Núm.. 1 (2026)

Figura 1.

Flujo de selección documental basado en las fases de PRISMA 2020



Nota. Elaboración propia a partir del proceso de selección de las 26 publicaciones candidatas, de las cuales se incluyeron 12 estudios empíricos.

Los estudios seleccionados se publicaron entre 2024 y 2026, con investigaciones procedentes de Ecuador, Estados Unidos, Grecia, Arabia Saudita, Indonesia y otros contextos internacionales. Se observaron pruebas de algoritmos de corrección de textos, intervenciones con informes de evaluación y experiencias de personalización para estudiantes con necesidades de apoyo. La heterogeneidad de métodos y muestras impidió resumir los hallazgos con una sola medida común.

Doi: https://doi.org/10.63688/92yma834

ISSN: 3103-1307

Vol. 4 Núm.. 1 (2026)

Tabla 2.

Evidencia empírica seleccionada para la síntesis

Estudio	Contexto / unidad	Aplicación evaluativa	Hallazgo reportado	Límite principal
Terrazas-Arellanes et al. (2025)	EE. UU.; 3.º–5.º de primaria	Modelo de aprendizaje automático para respuestas de ciencias	141 respuestas adicionales tras entrenamiento con 300; acuerdo QWK = 0,72.	Validación técnica; no mide inclusión de forma directa.
Liao et al. (2024)	China; 9.º curso, n = 125	Informes visuales formativos con IA	Comparó intervención y retroalimentación oral; reportó mejoras de aprendizaje y autorregulación.	Nivel escolar y asignatura acotados.
Lee et al. (2024)	Respuestas a tareas de ciencias; n = 1.650	Corrección mediante GPT-3.5/GPT-4 y rúbricas	Few-shot: exactitud 0,67; zero-shot: 0,60. La rúbrica incidió en la precisión.	Exactitud algorítmica no equivale a validez inclusiva.
Shin et al. (2026)	EE. UU.; secundaria inferior, 834 estudiantes	Sistema multiagente para respuestas escritas	Exactitud de 94 % frente a 78 % en configuración base; revisión humana de casos dudosos.	Rendimiento dependiente de rúbricas y configuración.
Haudek y Zhai (2024)	EE. UU.; secundaria, 17 tareas	Calificación con modelos de aprendizaje automático	Kappa medio 0,60 (rango 0,38–0,89); menor desempeño en tareas complejas.	Sensibilidad a características de los ítems.
Frey et al. (2026)	Instrumentos para estudiantes con discapacidad	600 preguntas generadas por tres chatbots	Errores de construcción frecuentes; revisión editorial necesaria.	Resultados referidos a preguntas de opción múltiple.
Smith et al. (2026)	9 docentes de primaria superior/secundaria	Retroalimentación escrita automática para dificultades de aprendizaje	Docentes solicitaron indicaciones más específicas y lenguaje adecuado al desarrollo.	Estudio exploratorio con nueve docentes.
Giaouri y Charisi (2025)	Grecia; 6 docentes de primaria inclusiva	ChatGPT para objetivos educativos individualizados	Mejor puntuación en objetivos específicos de corto plazo; resultados exploratorios.	No demuestra cambio sostenido en estudiantes.
Alsolami (2025)	Arabia Saudita; escolares con discapacidad intelectual	Actividades personalizadas con IA	Mejoras en pruebas académicas frente a comparación tradicional.	Contexto y perfil de apoyo específicos.

Doi: https://doi.org/10.63688/92yma834

ISSN: 3103-1307
Vol. 4 Núm.. 1 (2026)

Estudio	Contexto / unidad	Aplicación evaluativa	Hallazgo reportado	Límite principal
Sunaryati et al. (2025)	Indonesia; 66 estudiantes de 5.º	EduSmart: clasificación de desempeño y recomendaciones	Mejora media pre–post de 15,68 puntos; sin grupo control.	No permite atribución causal firme.
Núñez-Michuy et al. (2025)	Ecuador; 70 estudiantes	IA para personalizar apoyos; evaluación inicial y final	Reporte de mejoras académicas en diseño cuasiexperimental.	Institución única; extrapolación limitada.
Kotsari et al. (2026)	Grecia; 34 escolares con autismo	Robótica/realidad virtual con componentes de IA	Mejoras en pruebas de ciencias e implicación frente a comparación.	Estudio pequeño; intervención multicomponente.

Nota. QWK: kappa cuadrática ponderada. Los datos corresponden a lo comunicado por cada estudio y no constituyen resultados experimentales generados para este artículo.

Evaluación automatizada: posibilidades y criterios de validez

Los sistemas automáticos constituyeron el grupo de aplicación más directamente relacionado con la evaluación. Terrazas-Arellanes et al. (2025) describieron un modelo de aprendizaje automático para calificar respuestas construidas en ciencias de tercero a quinto grado. El acuerdo entre puntuaciones humanas y del sistema (QWK = 0,72) sugiere utilidad como apoyo, pero no autoriza a sustituir al evaluador en decisiones individuales de alto impacto. El resultado debe entenderse en función del conjunto de tareas, el entrenamiento y la rúbrica empleados.

Lee et al. (2024) analizaron 1.650 respuestas en seis tareas y compararon estrategias de indicaciones para modelos generativos. Las variaciones en exactitud, entre otras condiciones, favorecieron el uso de ejemplos y rúbricas contextualizadas. Haudek y Zhai (2024), a partir de 17 tareas de argumentación científica, identificaron que la complejidad y diversidad de lo que se pretende medir afectan el acuerdo entre calificación automática y humana. Ambos trabajos señalan un límite central: la IA no interpreta con igual fiabilidad cualquier producción escolar; el contenido y la estructura del ítem modifican su desempeño.

En el estudio de Shin et al. (2026), un sistema de varios agentes analizó respuestas escritas de 834 estudiantes de secundaria inferior. La configuración mejorada alcanzó 94 % de exactitud frente al 78 % comunicado para una configuración básica. Estos resultados no deben leerse como garantía universal de exactitud: corresponden a un conjunto y una rúbrica específicos. Su aporte más pertinente para inclusión es la propuesta de marcar incertidumbre

Doi: https://doi.org/10.63688/92yma834

ISSN: 3103-1307
Vol. 4 Núm.. 1 (2026)

y remitir a revisión docente las respuestas difíciles, en lugar de suponer que todo resultado algorítmico es incuestionable.

La generación automática de instrumentos exige precauciones adicionales. Frey et al. (2026) evaluaron 600 preguntas de selección múltiple elaboradas con tres chatbots para un escenario de educación especial. Identificaron errores de redacción y construcción, incluso cuando los distractores parecían plausibles. Un instrumento puede resultar atractivo o rápido de producir y, al mismo tiempo, contener barreras lingüísticas, claves previsibles o distractores que no responden al objetivo de aprendizaje. Para estudiantes que requieren ajustes, estas fallas pueden introducir dificultades ajenas al contenido evaluado.

Retroalimentación formativa y seguimiento del aprendizaje

La evaluación inclusiva no se agota en obtener una puntuación. La devolución comprensible, oportuna y utilizable permite que el estudiante reconozca lo que sabe, dónde se encuentra la dificultad y qué estrategia puede aplicar. Liao et al. (2024) desarrollaron informes visuales apoyados por IA para un curso de biología de noveno grado y los contrastaron con retroalimentación oral general. El diseño experimental aportó evidencias sobre potencial educativo de diagnósticos más desagregados. Aun así, sus efectos pertenecen a una asignatura y un entorno escolar determinados.

Smith et al. (2026) estudiaron percepciones de nueve docentes ante comentarios generados por una herramienta automatizada para escritura de estudiantes con dificultades de aprendizaje. Los participantes valoraron la posibilidad de ampliar oportunidades de práctica, pero pidieron indicaciones de mejora más concretas, vocabulario adecuado a la etapa y coherencia con objetivos individuales. Este hallazgo es importante: la capacidad de generar texto no prueba por sí misma que la retroalimentación sea accesible, emocionalmente adecuada o útil para tomar decisiones pedagógicas.

Sunaryati et al. (2025) informaron sobre EduSmart en 66 escolares de quinto grado, con clasificación de competencias a partir de cuestionarios y recomendaciones de aprendizaje. El cambio medio pre–post de 15,68 puntos es alentador como evidencia exploratoria, aunque la ausencia de grupo control impide descartar explicaciones alternativas. En consecuencia, debe

Doi: https://doi.org/10.63688/92yma834

ISSN: 3103-1307
Vol. 4 Núm.. 1 (2026)

evitarse atribuir exclusivamente al algoritmo la mejora observada. Su contribución es mostrar un modelo posible de identificación de necesidades y asignación diferenciada de actividades.

Ajustes individualizados y atención a necesidades educativas

La evidencia de inclusión se concentra en estrategias de atención individualizada y acompañamiento docente. Giaouri y Charisi (2025) compararon la elaboración de objetivos educativos individualizados con y sin ChatGPT en seis docentes de aulas inclusivas de primaria. Reportaron mejores puntuaciones en determinados objetivos de corto plazo, pero advirtieron que el tamaño de muestra y el contexto no permiten generalizar. El aporte del estudio reside en cómo la IA puede ayudar a formular metas observables, medibles y realistas cuando el docente valida la pertinencia de cada objetivo.

Alsolami (2025) examinó una intervención con IA en escolares con discapacidad intelectual leve y documentó mejoras en habilidades académicas frente a una comparación tradicional. Kotsari et al. (2026) estudiaron en 34 estudiantes con autismo una experiencia de ciencias con robótica educativa, realidad virtual y componentes de IA, utilizando pruebas antes y después e indicadores de participación. Estos trabajos muestran que adaptar actividades y registrar aprendizajes puede favorecer el acceso, pero la intervención multicomponente dificulta atribuir el efecto solo a la inteligencia artificial.

En el caso ecuatoriano, Núñez-Michuy et al. (2025) describieron una experiencia cuasiexperimental con 70 estudiantes de una institución y evaluaciones iniciales y finales. El estudio se relaciona con la personalización y la inclusión, y es valioso por su localización nacional. Sin embargo, una experiencia concreta no demuestra condiciones equivalentes en el conjunto de la EGB, ni permite asumir que existan recursos, conectividad o preparación docente similares en escuelas urbanas, rurales y comunitarias del país.

Tabla 3.

Implicaciones de la evidencia para la evaluación inclusiva en EGB de Ecuador

Eje	Aplicaciones plausibles	Condición de inclusión	Riesgo a controlar
Corrección de respuestas	Clasificación preliminar de producciones escritas, rúbricas asistidas	Supervisión docente, rúbricas accesibles y evaluación del error por subgrupos	Sesgo lingüístico, sobrevaloración del puntaje automático

Doi: https://doi.org/10.63688/92yma834

ISSN: 3103-1307
Vol. 4 Núm.. 1 (2026)

Eje	Aplicaciones plausibles	Condición de inclusión	Riesgo a controlar
Evaluación formativa	Alertas de progreso, informes y retroalimentación específica	Lenguaje claro; formatos alternativos; oportunidades de revisión	Comentarios genéricos, saturación informativa
Adaptación individual	Objetivos diferenciados, tareas ajustadas y seguimiento	Necesidades reales del estudiante; participación familiar cuando corresponda	Etiquetado indebido y uso de datos sensibles
Implementación institucional	Pilotos en materias y grados definidos	Formación docente, acceso equitativo, consentimiento, mantenimiento	Brecha digital, dependencia de proveedores
Seguimiento y validez	Análisis de diferencias entre recomendaciones humanas y automáticas	Auditoría, trazabilidad y evaluación de efectos por subgrupos	Decisiones automatizadas sin posibilidad de revisión

Nota. Las propuestas son orientaciones de transferencia al contexto ecuatoriano derivadas de los estudios analizados.

4. DISCUSIÓN

La síntesis permite distinguir entre precisión algorítmica y justicia educativa. Los avances en calificación automática son relevantes para reducir tareas repetitivas y disponer de diagnósticos oportunos, pero una medida de acuerdo entre máquina y persona no demuestra accesibilidad universal. En respuesta a la pregunta central, la evidencia disponible apoya una utilidad condicionada: mejor desempeño cuando el modelo opera con rúbricas explícitas, ejemplos representativos, revisión humana y criterios claros sobre qué aspectos de la respuesta deben evaluarse (Lee et al., 2024; Shin et al., 2026; Terrazas-Arellanes et al., 2025). La evaluación inclusiva requiere diferenciar la variabilidad relevante del aprendizaje de los obstáculos introducidos por el formato de la prueba. Una herramienta entrenada principalmente con respuestas de variedades lingüísticas distintas podría interpretar de modo desigual las producciones de estudiantes bilingües, con necesidades comunicativas específicas o con trayectorias educativas discontinuas. Ninguno de los estudios examinados permite afirmar que tales problemas estén resueltos para las variedades culturales y lingüísticas de Ecuador. Se requiere validación contextual antes de utilizar sistemas

Doi: https://doi.org/10.63688/92yma834

ISSN: 3103-1307
Vol. 4 Núm.. 1 (2026)

automáticos como fundamento exclusivo de calificaciones, promociones o decisiones de apoyo.

En este sentido, las revisiones previas ayudan a interpretar los resultados sin confundirlas con estudios primarios. González-Calatayud et al. (2021) señalaron la necesidad de una mayor fundamentación pedagógica en los trabajos de evaluación por IA. Martínez-Comesaña et al. (2023) sintetizaron un conjunto aún limitado de aplicaciones en primaria y secundaria. Létourneau et al. (2025), al examinar sistemas tutores inteligentes en K–12, comunicaron efectos generalmente positivos, aunque más modestos cuando el grupo de comparación disponía también de tutoría no inteligente. Esta última observación es especialmente relevante para no atribuir a la IA mejoras que podrían obtenerse mediante prácticas pedagógicas bien diseñadas sin automatización compleja.

En relación con la inclusión, Melo-López et al. (2025), Pagliara et al. (2024) y Li et al. (2025) destacan oportunidades de accesibilidad y personalización; al mismo tiempo, muestran que la diversidad de tecnologías y perfiles dificulta encontrar un mecanismo único de eficacia. Los estudios empíricos aquí revisados aportan indicios convergentes: mejor formulación de objetivos, disponibilidad de retroalimentación individual y posibles mejoras de desempeño en intervenciones adaptativas. No obstante, varias muestras son pequeñas, algunas experiencias mezclan tecnologías y otras describen pruebas de algoritmos sin medir participación real de estudiantes con necesidades específicas.

Para EGB de Ecuador, una incorporación responsable debería comenzar con proyectos piloto supervisados, en los que se comparen tiempos de corrección, consistencia de rúbricas, comprensión de la retroalimentación y efectos en estudiantes con diferentes perfiles de aprendizaje. Antes de escalar una solución se necesitarían pruebas de accesibilidad, medidas de error separadas por subgrupos, resguardo de información personal y procedimientos sencillos para cuestionar una puntuación generada por IA. La participación de docentes, directivos, estudiantes y familias en el diseño puede reducir el riesgo de adoptar sistemas técnicamente funcionales pero pedagógicamente poco pertinentes.

La desigualdad de acceso es otro límite sustantivo. Santillán Martínez et al. (2026) advirtieron que la automatización puede profundizar exclusiones cuando se desatienden las condiciones materiales del aprendizaje; Pajuelo-Abanto et al. (2026) explicaron la tensión entre personalización y estandarización algorítmica en América Latina. Un recurso

Doi: https://doi.org/10.63688/92yma834

que exige conectividad constante, cuentas individuales, dispositivos actualizados o uso intensivo de datos puede resultar inapropiado en entornos con acceso discontinuo. Por ello, la evaluación inclusiva asistida por IA debe disponer de alternativas equivalentes no digitales y no penalizar a quienes no puedan utilizar la plataforma.

Los resultados también aconsejan separar actividades de bajo y alto impacto. Generar propuestas de ítems, preparar comentarios preliminares o señalar respuestas que requieren revisión ofrece una intervención docente reversible. En cambio, automatizar decisiones finales, clasificar de manera permanente a estudiantes o emitir diagnósticos individualizados basados solo en registros de una plataforma exige una validez y responsabilidad mucho mayores. Los errores detectados en preguntas producidas por chatbots (Frey et al., 2026) refuerzan la necesidad de revisión humana sistemática.

La principal contribución del artículo consiste en integrar tres planos que suelen analizarse por separado: desempeño técnico, utilidad pedagógica e inclusión. Sin esta distinción, cualquier mejora de exactitud puede confundirse con un avance educativo; y cualquier promesa de personalización puede tomarse como prueba de equidad. La literatura sugiere que una buena evaluación inclusiva necesita preguntar qué aprendizaje se pretende valorar, cómo se interpreta la evidencia, si la retroalimentación puede usarse y si la intervención funciona sin ampliar brechas existentes.

La interpretación de los hallazgos debe considerar la heterogeneidad de diseños, niveles educativos, tecnologías y medidas de resultado. En consecuencia, la síntesis se concentra en patrones temáticos y en las condiciones pedagógicas, técnicas y éticas que pueden orientar la transferencia de estas aplicaciones a la EGB ecuatoriana.

5. CONCLUSIÓN

La inteligencia artificial ofrece aplicaciones concretas para la evaluación del aprendizaje en niveles escolares comparables a la EGB: puntuación de respuestas escritas, informes formativos, propuestas de instrumentos y apoyo para objetivos individuales. La investigación disponible respalda su potencial como asistencia pedagógica, no como sustituto autónomo del criterio profesional docente.

La inclusión depende de la calidad del diseño evaluativo. Las rúbricas, la adecuación del lenguaje, los ajustes razonables, la posibilidad de revisar decisiones y la protección de datos

son condiciones esenciales. Los errores observados en preguntas generadas y la variabilidad en la calificación automática muestran que la rapidez tecnológica no equivale a validez ni equidad.

La evidencia internacional resulta orientadora, pero su transferencia al Ecuador debe examinarse mediante pilotos contextualizados. La experiencia ecuatoriana analizada es valiosa como antecedente inicial, aunque no permite generalizar resultados a toda la Educación General Básica ni a estudiantes con necesidades educativas heterogéneas.

La revisión identificó una agenda prioritaria: investigaciones en escuelas ecuatorianas con comparadores adecuados, indicadores de inclusión, medidas diferenciadas por subgrupos, seguimiento temporal y participación docente. Estos elementos contribuirían a establecer la pertinencia y la utilidad pedagógica de las herramientas de IA en diversos entornos escolares del país.

REFERENCIAS BIBLIOGRÁFICAS

Alsolami, A. S. (2025). The effectiveness of using artificial intelligence in improving academic skills of school-aged students with mild intellectual disabilities in Saudi Arabia. *Research in Developmental Disabilities, 156*, 104884. https://doi.org/10.1016/j.ridd.2024.104884

Castro, V., de Oliveira Nascimento, A. K., Zheldibayeva, R., Zapata, G. C., Searsmith, D., Cope, B., & Kalantzis, M. (2026). Implementing rubric-aligned generative AI feedback in K–12 classrooms: A practice-oriented account of CyberScholar. *Ubiquitous Learning, 18*(2), 65–84. https://doi.org/10.18848/1835-9795/CGP/A370

Cohn, C., Hutchins, N., Le, T., & Biswas, G. (2024). A chain-of-thought prompting approach with LLMs for evaluating students' formative assessment responses in science. *Proceedings of the AAAI Conference on Artificial Intelligence, 38*(21), 23182–23190. https://doi.org/10.1609/aaai.v38i21.30364

Ebenbeck, N., & Gebhardt, M. (2024). Differential performance of computerized adaptive testing in students with and without disabilities—A simulation study. *Journal of Special Education Technology, 39*(4), 481–490. https://doi.org/10.1177/01626434241232117

Frey, B. B., Murray, A., & Mosher, M. (2026). The validity of AI-generated multiple-choice questions for students with disabilities. *Journal of Special Education Technology*. Publicación anticipada en línea. https://doi.org/10.1177/01626434261498192

Giaouri, S., & Charisi, M. (2025). Enhancing IEP design in inclusive primary settings through ChatGPT: A mixed-methods study with special educators. *Education Sciences, 15*(8), 1065. https://doi.org/10.3390/educsci15081065

González-Calatayud, V., Prendes-Espinosa, P., & Roig-Vila, R. (2021). Artificial intelligence for student assessment: A systematic review. *Applied Sciences, 11*(12), 5467. https://doi.org/10.3390/app11125467

Haudek, K. C., & Zhai, X. (2024). Examining the effect of assessment construct characteristics on machine learning scoring of scientific argumentation. *International Journal of Artificial Intelligence in Education, 34*, 1482–1509. https://doi.org/10.1007/s40593-023-00385-8

Hong, Q. N., Pluye, P., Fàbregues, S., Bartlett, G., Boardman, F., Cargo, M., Dagenais, P., Gagnon, M.-P., Griffiths, F., Nicolau, B., O'Cathain, A., Rousseau, M.-C., & Vedel, I. (2018). *Mixed methods appraisal tool (MMAT), version 2018: User guide*. McGill University. https://mixedmethodsappraisaltoolpublic.pbworks.com/

Kotsari, K., Smyrnaiou, Z., & Gena, A. (2026). AI-enhanced science learning for primary students with autism spectrum disorder: A quasi-experimental study with educational robotics and virtual reality. *Journal of Autism and Developmental Disorders*. Publicación anticipada en línea. https://doi.org/10.1007/s10803-026-07485-z

Lee, G.-G., Latif, E., Wu, X., Liu, N., & Zhai, X. (2024). Applying large language models and chain-of-thought for automatic scoring. *Computers and Education: Artificial Intelligence, 6*, 100213. https://doi.org/10.1016/j.caeai.2024.100213

Létourneau, A., Deslandes Martineau, M., Charland, P., Karran, J. A., Boasen, J., & Léger, P. M. (2025). A systematic review of AI-driven intelligent tutoring systems (ITS) in K–12 education. *npj Science of Learning, 10*, 29. https://doi.org/10.1038/s41539-025-00320-7

Li, J., Yan, Y., & Zeng, X. (2025). Exploring artificial intelligence in inclusive education: A systematic review of empirical studies. *Applied Sciences, 15*(23), 12624. https://doi.org/10.3390/app152312624

Liao, X., Zhang, X., Wang, Z., & Luo, H. (2024). Design and implementation of an AI-enabled visual report tool as formative assessment to promote learning achievement and self-regulated learning: An experimental study. *British Journal of Educational Technology, 55*, 1253–1276. https://doi.org/10.1111/bjet.13424

Liu, L., Liu, W., Luo, Z., & Yue, L. (2025). Scaffolds, screens, and readiness: How teacher-guided AI lifts special-education students into conversational agents. *Acta Psychologica, 261*, 105889. https://doi.org/10.1016/j.actpsy.2025.105889

Martínez-Comesaña, M., Rigueira-Díaz, X., Larrañaga-Janeiro, A., Martínez-Torres, J., Ocarranza-Prado, I., & Kreibel, D. (2023). Impacto de la inteligencia artificial en los métodos de evaluación en la educación primaria y secundaria: Revisión sistemática de la literatura. *Revista de Psicodidáctica, 28*(2), 93–103. https://doi.org/10.1016/j.psicod.2023.06.001

Melo-López, V.-A., Basantes-Andrade, A., Gudiño-Mejía, C.-B., & Hernández-Martínez, E. (2025). The impact of artificial intelligence on inclusive education: A systematic

review. *Education Sciences, 15*(5), 539. https://doi.org/10.3390/educsci15050539

Mena-Guacas, A. F., Vázquez-Cano, E., Fernández-Márquez, E., & López-Meneses, E. (2024). La inteligencia artificial y su producción científica en el campo de la educación. *Formación Universitaria, 17*(1), 155–164. https://doi.org/10.4067/S0718-50062024000100155

Núñez-Michuy, C. M., Bonilla-Jurado, D., Baquedano Moya, T. I., & Agualongo-Chela, L. M. (2025). El impacto de la inteligencia artificial en la personalización del aprendizaje y sus efectos en rendimiento académico e inclusión educativa. *Revista Científica UISRAEL, 12*(2), 113–131. https://doi.org/10.35290/rcui.v12n2.2025.1435

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., ... Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ, 372*, n71. https://doi.org/10.1136/bmj.n71

Pagliara, S. M., Bonavolontà, G., Pia, M., Falchi, S., Zurru, A. L., Fenu, G., & Mura, A. (2024). The integration of artificial intelligence in inclusive education: A scoping review. *Information, 15*(12), 774. https://doi.org/10.3390/info15120774

Pajuelo-Abanto, M. J. M., Carlos-Abanto, E. I., Bazán-Pajuelo, M. E., Vásquez-Mondragón, C. del P., Castillo Navarro, J. I., Bazán-Pajuelo, V. Y., & Colina-Cueva, L. A. (2026). Inteligencia artificial en la educación latinoamericana: Análisis crítico de personalización y desigualdad educativa. *Revista Latinoamericana de Estudios Educativos, 56*(2), 17–43. https://doi.org/10.48102/rlee.2026.56.2.859

Rethlefsen, M. L., Kirtley, S., Waffenschmidt, S., Ayala, A. P., Moher, D., Page, M. J., & Koffel, J. B. (2021). PRISMA-S: An extension to the PRISMA statement for reporting literature searches in systematic reviews. *Systematic Reviews, 10*, 39. https://doi.org/10.1186/s13643-020-01542-z

Santillán Martínez, A., Rivera Elizalde, E., & Elizalde Salazar, R. (2026). Inteligencia artificial en la educación: Entre la equidad digital y la exclusión algorítmica. *Revista Latinoamericana de Estudios Educativos, 56*(2), 45–72. https://doi.org/10.48102/rlee.2026.56.2.815

Shin, N., Qian, X., Miller, C. S., Li, H., Chu, Y., Krajcik, J., Tang, J., & Xing, Y. (2026). An artificial intelligence-enhanced assessment framework for analyzing middle school science students' written responses. *Journal of Educational Measurement, 63*(2), e70044. https://doi.org/10.1111/jedm.70044

Smith, S. J., Goldman, S., & Carreon, A. (2026). Evaluating the validity of AI-generated writing feedback for students with learning disabilities. *Journal of Special Education Technology*. Publicación anticipada en línea. https://doi.org/10.1177/01626434261498189

Sunaryati, T., Widodo, E., Firdaus, A., Sudharsono, M., & Rayyan, M. F. (2025). Development of Edusmart as an AI-based adaptive learning application for needs analysis among Indonesian primary school students. *TARBIYA: Journal of Education in Muslim Society, 12*(2), 163–178. https://doi.org/10.15408/tjems.v12i2.46582

Terrazas-Arellanes, F. E., Strycker, L., Alvez, G. G., Miller, B., & Vargas, K. (2025). Promoting agency among upper elementary school teachers and students with an artificial intelligence machine learning system to score performance-based science assessments. *Education Sciences, 15*(1), 54. https://doi.org/10.3390/educsci15010054

Voultsiou, E., & Moussiades, L. (2025). A systematic review of AI, VR, and LLM applications in special education: Opportunities, challenges, and future directions. *Education and Information Technologies, 30*, 19141–19181. https://doi.org/10.1007/s10639-025-13550-4

CONFLICTO DE INTERESES

No existe.

FINANCIAMIENTO

No aplica.

CONTRIBUCIÓN DE AUTORÍA

Conceptualización: Virginia Yamilet Mendoza Loor y Fabiana Nairoby Moncayo Quirola. Metodología: Jessica Viviana Quirola Mera y Derian Santiago Mendoza Zambrano. Investigación: Virginia Yamilet Mendoza Loor, Jessica Viviana Quirola Mera y Emely Jamilet Cárdenas Posligua. Redacción – borrador original: Fabiana Nairoby Moncayo Quirola, Derian Santiago Mendoza Zambrano y Emely Jamilet Cárdenas Posligua. Redacción – revisión y edición: Virginia Yamilet Mendoza Loor, Fabiana Nairoby Moncayo Quirola, Jessica Viviana Quirola Mera, Derian Santiago Mendoza Zambrano y Emely Jamilet Cárdenas Posligua.