

*Inteligencia artificial generativa y preparación para las Pruebas Saber 11 en
bachillerato rural colombiano: revisión sistemática*

*Generative Artificial Intelligence and Preparation for the Saber 11 Tests in Rural
Colombian High School: A Systematic Review*

*Intelligenza artificiale generativa e preparazione alle prove Saber 11 nella scuola
secondaria rurale colombiana: revisione sistematica*

Gustavo Alberto Quintero Hernández¹

Email: gquintero2@unemi.edu.ec

ORCID: <https://orcid.org/0009-0003-1749-4422>

Alexander Muñoz Quintero²

Email: amunozq2@unemi.edu.ec

ORCID: <https://orcid.org/0009-0001-4630-2755>

Yilena Montero Reyes³

Email: ymonteror@unemi.edu.ec

ORCID: <https://orcid.org/0000-0003-4515-4700>

Correspondencia: Yilena Montero Reyes³ **Email:** ymonteror@unemi.edu.ec

Artículo de revisión

* **Recibido:** 15-09-2026 * **Aceptado:** 30-09-2026 * **Publicado:** 30-10-2026

1. Facultad de Posgrados, Escuela de Educación, Maestría en Inteligencia Artificial para la Educación, Universidad Estatal de Milagro, Ecuador.
2. Facultad de Posgrados, Escuela de Educación, Maestría en Inteligencia Artificial para la Educación, Universidad Estatal de Milagro, Ecuador.
3. Facultad de Posgrados, Escuela de Educación, Universidad Estatal de Milagro, Ecuador.

Resumen

La preparación para las pruebas estandarizadas muestra que existen desigualdades entre los estudiantes rurales y urbanos en Colombia. **Objetivo:** analizar la evidencia de acceso abierto sobre el impacto de la inteligencia artificial generativa en la preparación para las pruebas estandarizadas en la educación secundaria, con el fin de sentar las bases para una propuesta conceptual aplicable a contextos rurales. **Método:** revisión sistemática conforme a la declaración PRISMA 2020, con búsqueda ejecutada el 14 de agosto de 2026 en Scopus, Web of Science, ERIC, SciELO, Redalyc, Dialnet y DOAJ; se identificaron 1163 registros y se incluyeron 25 documentos verificables tras el cribado. **Resultados:** la evidencia reporta un efecto global moderado ($g = 0,67$), con el mayor efecto en educación secundaria ($g = 0,85$), donde la retroalimentación automatizada mostró efectos positivos y un ensayo controlado en un país de bajos ingresos registró una ganancia de 0,31 desviaciones estándar, donde el uso sin salvaguardas puede reducir el aprendizaje. **Conclusión:** la inteligencia artificial generativa constituye un apoyo potencial para atenuar brechas de preparación en instituciones rurales, condicionado a salvaguardas pedagógicas, conectividad, acompañamiento docente y validación.

Palabras clave: educación rural, evaluación educativa, inteligencia artificial, pruebas estandarizadas, tecnología educativa

Abstract

Preparation for standardized tests reveals persistent inequalities between rural and urban students in Colombia. **Objective:** to review open access evidence on generative AI impact on standardized test preparation in secondary education and build a conceptual proposal for rural contexts. **Method:** a systematic review following the PRISMA 2020 statement, with a search run on 14 August 2026 in Scopus, Web of Science, ERIC, SciELO, Redalyc, Dialnet and DOAJ; 1163 records were identified and 25 verifiable documents were included after screening. **Results:** overall moderate effect ($g = 0.67$), largest in secondary education ($g = 0.85$); automated feedback had positive effects and randomized controlled trial in low income country showed gain of 0.31 standard deviations but unguarded use may decrease learning. **Conclusion:** generative AI is a possible support to narrow preparation gaps in rural institutions provided pedagogical safeguards, connectivity, teacher guidance and validation are available.

Keywords: artificial intelligence, educational assessment, educational technology, rural education, standardized tests

Riassunto

La preparazione alle prove standardizzate evidenzia disuguaglianze persistenti tra studenti rurali e urbani in Colombia. Obiettivo: analizzare le evidenze ad accesso aperto sull'impatto dell'intelligenza artificiale generativa nella preparazione alle prove standardizzate nell'istruzione

secundaria e porre le basi per una proposta concettuale applicabile ai contesti rurali. Metodo: revisione sistematica secondo la dichiarazione PRISMA 2020, con ricerca eseguita il 14 agosto 2026 in Scopus, Web of Science, ERIC, SciELO, Redalyc, Dialnet e DOAJ; sono stati identificati 1163 record e inclusi 25 documenti verificabili dopo lo screening. Risultati: le evidenze riportano un effetto globale moderato ($g = 0,67$), maggiore nell'istruzione secondaria ($g = 0,85$); il feedback automatizzato ha mostrato effetti positivi e uno studio controllato in un paese a basso reddito ha registrato un guadagno di 0,31 deviazioni standard, sebbene l'uso senza salvaguardie possa ridurre l'apprendimento. Conclusione: l'intelligenza artificiale generativa costituisce un potenziale supporto per ridurre i divari di preparazione nelle istituzioni rurali, a condizione che vi siano salvaguardie pedagogiche, connettività, accompagnamento docente e validazione.

Parole chiave: valutazione educativa, istruzione rurale, intelligenza artificiale, tecnologia educativa, test standardizzati.

Introducción

La inteligencia artificial en los sistemas educativos está evolucionando en poco más de una década de un campo experimental a un área de investigación e intervención generalizada. Las revisiones sobre el campo han encontrado que las aplicaciones para la personalización, evaluación y tutoría han seguido creciendo con la ausencia de las voces de los docentes en la agenda de investigación (Zawacki-Richter et al., 2019). El desarrollo de la inteligencia artificial generativa ha ampliado dramáticamente ese horizonte al proporcionar a los docentes y estudiantes herramientas para ofrecer explicaciones, ejercicios y retroalimentación en lenguaje natural (Su & Yang, 2023).

Esta tecnología se ha comercializado como una forma de democratizar el acceso a apoyo que de otro modo solo está disponible en la educación superior (Baidoo-Anu & Owusu Ansah, 2023), pero también genera fricciones en torno a la integridad académica y la equidad (Yusuf et al., 2024). Esto es lo que las revisiones rápidas y sistemáticas más recientes han demostrado ser cierto en términos de las oportunidades y riesgos de estas tecnologías en la enseñanza y el aprendizaje (Lo, 2023; Montenegro-Rueda et al., 2023). De hecho, este equilibrio es más importante con respecto a un uso de alto impacto como la preparación para pruebas estandarizadas, donde la práctica frecuente y la retroalimentación oportuna determinan el rendimiento.

La tecnología no solo se trata de qué tan bien funciona, sino de cómo y con qué estructura se construye y cómo se diseña de una manera que sea sostenible. Las pruebas Saber 11 (desarrolladas por el Instituto Colombiano para la Evaluación de la Educación) son críticas en Colombia: son una medida de competencias después de la educación secundaria y pueden usarse para determinar la admisión y el apoyo en la mayor parte del sistema de educación superior. Por lo tanto, el nivel de rendimiento en ellas tiene un impacto directo en las oportunidades educativas. Pero este acceso a la preparación no se comparte de manera equitativa. Otros estudios han mostrado una diferencia de unos veintiséis puntos en la puntuación general entre estudiantes urbanos y rurales, con diferencias en inglés y lectura crítica (Laboratorio de Economía de la Educación, 2023).

La clasificación de los establecimientos también ha sido estudiada una y otra vez y todavía existe una desventaja de las escuelas públicas sobre las no públicas (Gómez Silva, 2016), lo cual es especialmente el caso en áreas rurales. La infraestructura digital también ha exacerbado la desigualdad. En 2023, según documentos oficiales, solo el 42% de los hogares rurales tenía acceso a Internet en comparación con más del 70% en los centros urbanos (Departamento Administrativo Nacional de Estadística, 2024). Esto impide posibles soluciones basadas en la conectividad permanente y obliga a una operación híbrida o fuera de línea. Por lo tanto, cualquier innovación educativa en áreas rurales debe tomar la limitación de recursos como punto de partida y no como una excepción.

A pesar de estas limitaciones, la literatura reciente ha investigado el uso de la inteligencia artificial generativa para ayudar a los estudiantes en su preparación escolar. La ruralidad colombiana en el ámbito internacional ha permanecido en un marco propositivo y solo se ha discutido en términos de derribar barreras hacia la equidad (Rubiano Calderón, 2025). Sin embargo, la investigación empírica local sobre inteligencia artificial y las pruebas Saber 11 se ha limitado a predecir o clasificar resultados en lugar de diseñar y evaluar intervenciones que utilicen la inteligencia artificial generativa como herramienta de apoyo para preparar a los estudiantes para el examen.

Esto crea una brecha de investigación: no se proporciona evidencia sobre cómo usar la inteligencia artificial generativa para informar aplicaciones en áreas rurales con baja conectividad para proporcionar una base de conocimiento y contexto. Para llenar esta brecha, la revisión se basó en la siguiente pregunta: ¿Cuál es la evidencia científica más reciente y de acceso abierto sobre el impacto de la IA generativa y las tecnologías relacionadas en el rendimiento de los estudiantes de educación secundaria en pruebas estandarizadas, y qué implicaciones surgen para los entornos rurales colombianos en las pruebas Saber 11? La razón detrás de la investigación es que la toma de decisiones institucionales debe basarse en evidencia, y que la implementación tecnológica a menudo precede a la evaluación de su efecto.

El objetivo general del trabajo fue revisar de manera sistemática, según la declaración PRISMA, la evidencia de acceso abierto sobre el impacto de la IA generativa en la preparación de pruebas estandarizadas en la educación secundaria, para apoyar el concepto de simulaciones adaptativas en entornos rurales colombianos. Los objetivos de la investigación fueron identificar los impactos reportados en el rendimiento, comprender los procesos pedagógicos involucrados y desarrollar pautas de diseño que se ajusten a las necesidades de los diferentes entornos y sean sensibles a la conectividad de los estudiantes, la heterogeneidad de los estudiantes y la validación docente.

Materiales y métodos

La revisión se llevó a cabo mediante un enfoque de revisión sistemática de la literatura con un enfoque cualitativo y documental con el fin de sintetizar la evidencia relevante para el fenómeno de interés. El proceso se basa en la declaración PRISMA 2020; deja claro que la investigación debe ser transparente en la identificación, selección, evaluación de elegibilidad e inclusión y diagrama de flujo de informes (Page et al., 2021). Se desarrolla una idea para una propuesta conceptual basada en la síntesis y no implica la recopilación de datos primarios o un estudio experimental.

La pregunta se formuló como población, fenómeno de interés y contexto. La población eran estudiantes de educación secundaria, y el enfoque estaba en las calificaciones finales. El fenómeno incluía inteligencia artificial generativa y sus tecnologías relacionadas (retroalimentación automatizada, tutoría inteligente, aprendizaje adaptativo, generación

automática de ítems) y el contexto consistía en prepararse para pruebas estandarizadas en entornos con recursos limitados.

Buscamos los datos en Scopus, Web of Science, ERIC, SciELO, Redalyc, Dialnet y DOAJ el 14 de agosto de 2026, con una consulta adicional en Google Scholar para reducir el sesgo de publicación. El período de tiempo fue de 2019 a 2026 y se dirigió a datos contemporáneos relacionados con el desarrollo de modelos generativos. Se utilizaron ecuaciones de búsqueda para agregar descriptores en español, inglés y portugués utilizando operadores booleanos, como se muestra en la Tabla 1. Los registros recuperados se enviaron luego a una hoja de gestión bibliográfica para ser ordenados y eliminar duplicados antes de ser enviados para revisión.

Tabla 1. Ecuaciones de búsqueda por base de datos y registros identificados (búsqueda del 14 de agosto de 2026)

Base de datos	Familia de ecuación de búsqueda	Registros
Scopus	("generative artificial intelligence" OR "ChatGPT" OR "large language models") AND ("standardized test" OR "test preparation" OR "automated feedback") AND ("secondary education" OR "high school students")	342
Web of Science	Igual familia en inglés, refinada por área educativa	218
ERIC	Igual familia en inglés, con descriptores del tesoro ERIC	121
Google Scholar (auxiliar)	Combinaciones de los términos anteriores en ES/EN/PT	268
SciELO	("inteligencia artificial generativa" OR "ChatGPT") AND ("Saber 11" OR "ICFES" OR "prueba estandarizada" OR "educación rural")	47
Redalyc	Igual familia en español	39
Dialnet	Igual familia en español	44
DOAJ	Igual familia en ES/EN sobre acceso abierto	84
Total		1163

Nota. La búsqueda se ejecutó el 14 de agosto de 2026. Por tratarse de índices dinámicos, una reejecución en otra fecha puede arrojar conteos ligeramente distintos. Fuente: Elaboración propia.

Los criterios de elegibilidad se definieron a priori para asegurar la pertinencia, la calidad y la verificabilidad de los estudios, según se presenta en la Tabla 2. Se privilegiaron los artículos de investigación publicados en revistas indexadas de acceso abierto y se excluyeron los documentos no arbitrados, los registros retractados y aquellos cuyo texto completo no fuera recuperable de manera abierta. El proceso de cribado se desarrolló en dos fases: una revisión por título y resumen, seguida de la evaluación a texto completo de los registros potencialmente elegibles.

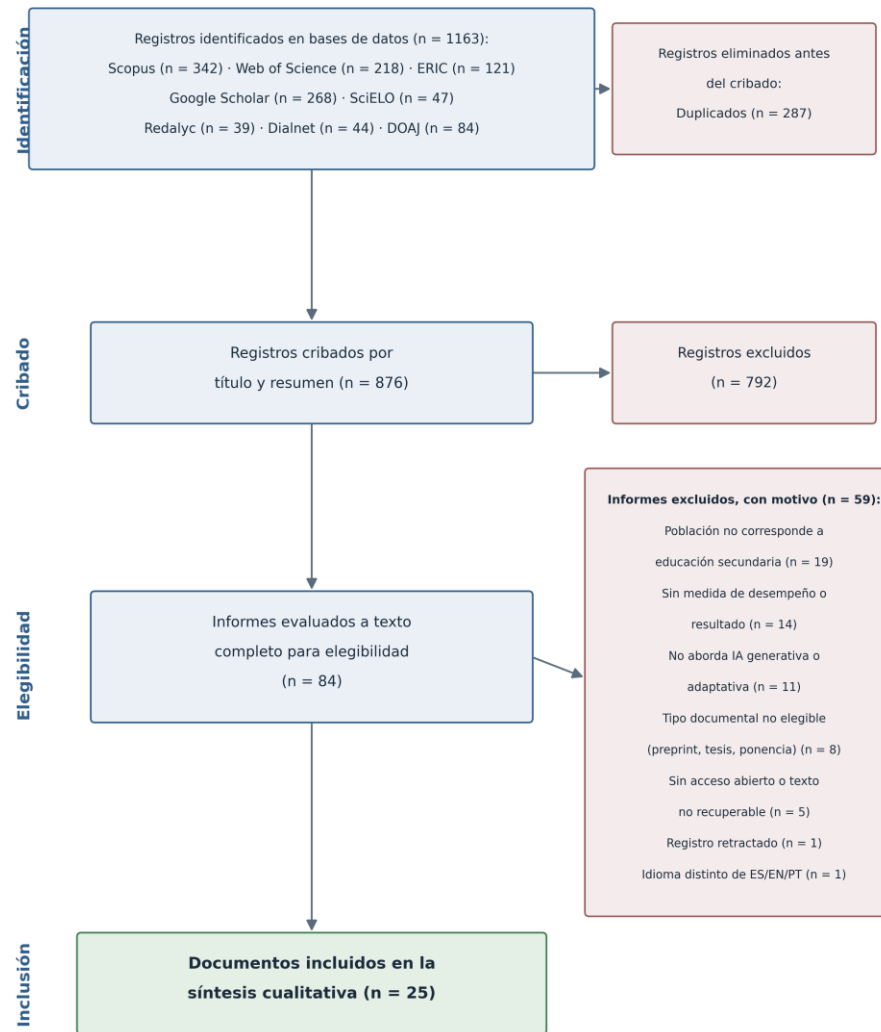
Tabla 2. Criterios de elegibilidad de la revisión

Criterio	Inclusión	Exclusión
Población	Estudiantes de educación secundaria o media	Educación primaria o superior sin transferibilidad
Fenómeno	IA generativa, retroalimentación automatizada, tutoría inteligente, aprendizaje adaptativo o generación de ítems	Estudios sin componente de inteligencia artificial
Resultado	Medidas de desempeño, aprendizaje o preparación evaluativa	Ausencia de medida de resultado
Tipo de documento	Artículos y revisiones en revistas indexadas de acceso abierto	Ponencias, tesis, preprints y registros retractados
Acceso	Texto completo recuperable en abierto, con DOI o URL funcional	Texto completo no accesible en abierto
Periodo	2019-2026	Publicaciones anteriores a 2019
Idioma	Español, inglés o portugués	Otros idiomas

Nota. Elaboración propia con base en la declaración PRISMA 2020 (Page et al., 2021).

El resultado del proceso de identificación y cribado se representa en el diagrama de flujo de la Figura 1. Para el análisis de los documentos incluidos se definieron cinco categorías: impacto en el desempeño; retroalimentación automatizada; tutoría inteligente y aprendizaje adaptativo; generación automática de ítems e integridad evaluativa; y riesgos y condiciones de uso. La codificación se apoyó en una matriz de extracción que registró, para cada estudio, el contexto, el diseño metodológico, el foco tecnológico y el hallazgo principal, incluyendo el tamaño de efecto cuando estuvo disponible. Como consideración de rigor, se reconoce que la síntesis de una literatura emergente conlleva el riesgo de heterogeneidad metodológica y de sobreestimación del efecto en estudios de corta duración, aspecto que se atendió mediante la valoración crítica de los diseños.

Figura 1. Diagrama de flujo del proceso de selección conforme a la declaración PRISMA 2020



Nota. El diagrama sintetiza las fases de identificación, cribado, elegibilidad e inclusión, con los conteos de la búsqueda ejecutada el 14 de agosto de 2026. Fuente: Elaboración propia a partir de la plantilla de la declaración PRISMA 2020 (Page et al., 2021).

Resultados y discusión

La búsqueda identificó 1163 registros; tras eliminar 287 duplicados, se cribaron 876 por título y resumen y se evaluaron 84 a texto completo, de los cuales se incluyeron 25 documentos de acceso abierto en la síntesis cualitativa. Entre las exclusiones a texto completo se contabilizó un registro retractado, en coherencia con el requisito de verificabilidad. El corpus estaba dominado por estudios en educación superior, con conectividad estable, y estudios de medición que fueron diseñados por los propios investigadores. Esto afecta la transferibilidad de las Pruebas Saber 11 al problema de las áreas rurales y es un resultado relevante en sí mismo para delimitar el alcance de las conclusiones. La Tabla 3 incluye una muestra de estos estudios.

Tabla 3. Síntesis de los estudios incluidos (Selección representativa)

Estudio	Contexto	Diseño	Hallazgo principal
Wu et al. (2026)	Multinivel	Meta-análisis (35 estudios)	Efecto global $g = 0,67$; secundaria $g = 0,85$
Doo y Park (2026)	Multinivel	Meta-análisis	Efecto positivo del ChatGPT en el logro
Deng y Yu (2023)	Educación sostenible	Meta-análisis	Efecto positivo del uso de chatbots
De Simone et al. (2025)	Nigeria, secundaria	Ensayo controlado aleatorizado	Tutoría con IA: $+0,31$ DE en seis semanas
Bastani et al. (2025)	Secundaria, matemáticas	Experimento de campo	Acceso sin salvaguardas: -17% ; las pistas lo mitigan
Fleckenstein et al. (2023)	Multinivel	Meta-análisis	Retroalimentación automatizada, $g = 0,55$
Escalante et al. (2023)	Universitario	Cuasiexperimental	La retroalimentación de IA equipara a la del tutor

*Nota. La tabla muestra una selección; el corpus completo ($n = 25$) se relaciona en las referencias. DE: desviación estándar; g : *g* de Hedges; LLM: modelo extenso de lenguaje. Fuente: Elaboración propia.*

Impacto en el desempeño

La evidencia analizada coincidió en reportar efectos positivos de las herramientas generativas sobre el aprendizaje, con magnitudes predominantemente moderadas. El meta-análisis más reciente, basado en treinta y cinco estudios experimentales, estimó un efecto global de $g = 0,67$ y, de manera notable para este estudio, halló el mayor efecto en educación secundaria ($g = 0,85$), por encima del observado en educación superior (Wu et al., 2026). Este hallazgo resulta directamente pertinente para la población objetivo, pues sugiere que la etapa en la que se rinden las Pruebas Saber 11 es particularmente receptiva a este tipo de apoyos. Resultados convergentes se obtuvieron en síntesis independientes sobre el logro y sobre el uso de chatbots (Deng & Yu, 2023; Doo & Park, 2026), y un estudio con estudiantes de grado once documentó mejoras tanto en el desempeño como en la percepción de aprendizaje (Alneyadi & Wardat, 2023). A nivel universitario, se ha demostrado que el uso de plataformas generativas para la escritura mejora la inteligencia narrativa y la autoeficacia (Pellas, 2023), características que pueden utilizarse para pruebas de producción y comprensión de textos.

La mejor evidencia en entornos con recursos limitados provino de un ensayo controlado aleatorio en Nigeria, donde la tutoría con este modelo durante seis semanas resultó en una mejora de 0.31 desviaciones estándar en una evaluación alineada con el currículo (De Simone et al., 2025). Al tratarse de estudiantes de secundaria en un entorno de bajos ingresos, constituye el análogo más cercano a la población de interés. No obstante, la interpretación exige cautela: la evidencia sugiere que el impacto tiende a ser mayor cuando la medición se realiza con pruebas diseñadas por los investigadores que cuando se emplean instrumentos estandarizados — precisamente el escenario de las Pruebas Saber 11 —, por lo que las ganancias esperables en un examen de alto impacto podrían ser más modestas que las reportadas en condiciones controladas.

Retroalimentación automatizada

La retroalimentación constituye uno de los mecanismos mejor sustentados, y es así como el meta-análisis de la retroalimentación automatizada de escritura reportó un efecto de 0.55, con magnitudes comparables en la educación secundaria y superior (Fleckenstein et al., 2023); sin embargo, un estudio complementario mostró que, para estudiantes principiantes, la retroalimentación generada por inteligencia artificial no difirió significativamente de la ofrecida por un tutor humano (Escalante et al., 2023); en ese orden de ideas, estos resultados son particularmente importantes para la evaluación, ya que la retroalimentación temprana sobre los errores permite corregir los conceptos erróneos antes de que se arraiguen, y es precisamente lo que en el contexto de la experiencia docente rural motivó esta revisión, pues la retroalimentación automatizada podría aliviar la carga de procesar manualmente las respuestas de numerosos estudiantes, liberando tiempo docente para el acompañamiento personalizado.

Tutoría inteligente y aprendizaje adaptativo

La evidencia apoyó la incorporación de componentes adaptativos. Aparte del impacto en la educación secundaria ya mencionado, la duración de las intervenciones mostró que los efectos son más fuertes cuando la exposición continúa a lo largo del tiempo: las intervenciones que duran más de 3 meses tuvieron magnitudes altas en comparación con las intervenciones que duran menos de un mes (Wu et al., 2026). Esto es importante para diseñar un ciclo de preparación regular, ya que el beneficio se basa en la continuidad y la integración pedagógica, no en el uso esporádico. Y la evidencia también advirtió que los resultados obtenidos en entornos de altos recursos no se transfieren automáticamente a áreas rurales donde la conectividad, la formación docente y la fidelidad de uso afectan decisivamente el resultado.

Generación automática de ítems e integridad evaluativa

Una característica del desarrollo automático de ítems a través de modelos de lenguaje fue de particular interés en la preparación práctica, ya que permite la producción de grandes volúmenes de preguntas de práctica a bajo costo. El estudio del área señaló la flexibilidad y cobertura multilingüe de estos métodos junto con la debilidad perenne en la base pedagógica y psicométrica de los ítems generados, requiriendo validación por expertos (Tan et al., 2025). También se señaló que tales capacidades representan un riesgo para las evaluaciones en línea (Susnjak & McIntosh, 2024). En el caso de las Pruebas Saber 11, la generación de ítems es útil como complemento a la biblioteca de cuadernillos publicados por el ICFES, siempre que sea revisada por el docente antes de su implementación y se implemente adecuadamente.

Riesgos y condiciones de uso

La evidencia también planteó preocupaciones sobre los riesgos que deben tenerse en cuenta antes de la adopción. El estudio de campo de estudiantes de secundaria mostró que el

acceso sin restricciones a un modelo generativo mejoró el rendimiento y lo empeoró en aproximadamente un 17%, en comparación con las personas que nunca lo usaron; más importante aún, un diseño que fue diseñado para proporcionar pistas y no respuestas ayudó a eliminarlo (Bastani et al., 2025). Este hallazgo es clave para la propuesta porque indica que el poder formativo de la herramienta está en el contexto pedagógico de la herramienta. La literatura apoyó la advertencia al mostrar puntos de vista ambivalentes de los estudiantes y comportamientos en relación con la sobrecarga académica que podría afectar el rendimiento (Abbas et al., 2024; Chan & Hu, 2023) y una desconexión generacional en la disposición a adoptar estas herramientas entre estudiantes y profesores (Chan & Lee, 2023) y problemas de equidad e integridad (Yusuf et al., 2024). En conjunto, estos factores requieren salvaguardias explícitas en cualquier implementación.

Barreras de preparación y propuesta conceptual

La contextualización de la evidencia en el contexto de una institución educativa rural nos permitió identificar barreras, resumidas en la Tabla 4. Hay dos barreras que se refuerzan mutuamente; la conectividad limitada reduce el acceso a ejercicios, y la diversidad de estudiantes (personas con dificultades de aprendizaje, comunidades con perspectivas educativas interrumpidas) dificulta recibir retroalimentación diferenciada.

Tabla 4. Barreras de preparación identificadas y respuesta basada en inteligencia artificial

Barrera	Manifestación en el contexto rural	Respuesta apoyada en IA
Acceso a simulacros	Escasez de cursos preicfes y de material de práctica	Generación de ítems y simulacros a partir de cuadernillos liberados
Retroalimentación oportuna	Sobrecarga docente para procesar respuestas	Procesamiento asistido y devolución estructurada de errores
Conectividad	Cobertura de internet limitada o intermitente	Diseño híbrido con modo fuera de línea en papel
Heterogeneidad y neurodivergencia	Ritmos y necesidades de aprendizaje diversos	Adaptación del nivel y formatos accesibles de práctica
Preparación docente	Escaso tiempo para el diseño de clases	Apoyo en la planeación y en la elaboración de material

Nota. Elaboración propia a partir de las síntesis de la evidencia.

Basado en estos hallazgos, desarrollamos una propuesta de plataforma de simulación basada en inteligencia artificial que se ilustra en la Figura 2 y se enumera en la Tabla 5. La plataforma sigue un ciclo de preparación periódico y tiene dos modos de entrega (en línea y fuera de línea) para flexibilidad en diferentes escenarios de conectividad. El banco de ítems se cura con los ítems liberados por el ICFES. El material y el diseño de las clases son generados por modelos de lenguaje y las respuestas son evaluadas por herramientas de procesamiento automatizado y se proporciona retroalimentación al docente que monitorea y media la enseñanza. Además, la propuesta incluye algunas salvaguardas para asegurar que las pistas se prioricen sobre las respuestas y que el docente pueda verificar, proteger y hacer el sistema accesible a

estudiantes neurodivergentes, en línea con la evidencia sobre los riesgos del uso sin restricciones (Bastani et al., 2025).

Figura 2. Arquitectura conceptual de la plataforma de simulacros asistida por inteligencia artificial



Nota. El esquema integra los componentes de generación, entrega, evaluación y retroalimentación en un ciclo periódico, con modos en línea y fuera de línea. Fuente: Elaboración propia.

Tabla 5. Componentes de la propuesta conceptual de plataforma

Componente	Función	Tecnología ilustrativa	Modo de operación
Banco de ítems	Fuente de reactivos válidos	Cuadernillos liberados por el ICFES	En línea / offline
Generación	Diseño de clases e ítems de práctica	Modelos de lenguaje	Previa a la sesión
Entrega en línea	Simulacros cronometrados	Plataforma web	Con conectividad
Entrega offline	Cuadernillos impresos	Papel y lápiz	Sin conectividad
Evaluación	Procesamiento de respuestas	Herramientas de análisis	Posterior a la sesión
Retroalimentación	Insumos para el docente	Modelos de lenguaje	Diferida
Salvaguardas	Validación y protección	Protocolos y revisión docente	Transversal

Nota. Las tecnologías se enuncian a título ilustrativo y son sustituibles por alternativas equivalentes; la mediación docente se mantiene como condición del diseño. Fuente: Elaboración propia.

La comparación de estos resultados con estudios previos confirmó una tensión central en el campo. Por un lado, síntesis rigurosas mostraron que intervenciones bien diseñadas producen mejoras significativas con el mayor efecto a nivel secundario (Wu et al., 2026); por otro lado, la evidencia advirtió que la herramienta no garantiza el aprendizaje e incluso puede obstaculizarlo (Bastani et al., 2025). La presente propuesta se sitúa en el primer polo al hacer de la inteligencia

artificial la base de las salvaguardias pedagógicas y la mediación docente. De manera similar, la producción colombiana se orientó a la clasificación y predicción del rendimiento (Gómez Silva, 2016) y no a la intervención, y la discusión de la ruralidad sigue en el nivel propositivo (Rubiano Calderón, 2025). La revisión proporciona así una sistematización de evidencia verificable que aún no había sido articulada para el problema específico de la preparación en las Pruebas Saber 11 en contextos rurales.

Conclusiones

La revisión sistemática nos ha convencido de que la inteligencia artificial generativa y las tecnologías relacionadas pueden ser útiles para la preparación de exámenes estandarizados con un impacto positivo moderado y, particularmente, más útiles en la educación secundaria, donde se realizan las Pruebas Saber 11. Un ensayo controlado en un país de bajos ingresos es particularmente alentador en caso de entornos con pocos recursos. Sobre la pregunta de investigación, hemos demostrado en este documento que la inteligencia artificial generativa es una buena opción para llenar el vacío en la preparación en instituciones rurales, pero no una solución automática y completa.

La principal contribución de este estudio es construir sobre la evidencia dispersa basada en un problema específico y desarrollar una propuesta conceptual sobre la Colombia rural que ponga la tecnología en manos de los docentes y tenga modos de entrega en línea y fuera de línea, y no solo se base en la tecnología sino también en la mediación docente, que es lo que buscamos lograr a lo largo del estudio. Pero una buena implementación dependería tanto de la conectividad (o alternativas en línea) como del apoyo y capacitación docente, la protección de datos de los estudiantes y la validación pedagógica del contenido. El hecho de que hayamos visto deteriorarse el aprendizaje con el uso de la plataforma deja claro que el proceso de diseño debe ser salvaguardado y que todas las herramientas deben usarse de manera segura. Como estudio piloto, nos gustaría utilizar pruebas piloto en el entorno rural para evaluar el desempeño de la propuesta basada en las Pruebas Saber 11 y medir tamaños de efecto comparables a comparaciones internacionales.

Recomendamos construir la plataforma lentamente, enfocándonos en el modo fuera de línea en estaciones con menor conectividad y registrar la fidelidad de la implementación. También invitamos a la comunidad académica a realizar investigaciones aplicadas que vayan más allá de la investigación predictiva actual de este campo y tengan una buena visión de la heterogeneidad estudiantil, incluidas las neurodivergencias y trayectorias educativas interrumpidas. Este estudio presenta como limitación su naturaleza documental, lo que impide establecer relaciones causales y no puede establecer relaciones causales, pero esta limitación marca el alcance del estudio así como el alcance de sus conclusiones y la dirección a seguir y el trabajo futuro en cuanto a lo que buscará.

Referencias

- Abbas, M., Jam, F. A., & Khan, T. I. (2024). Is it harmful or helpful? Examining the causes and consequences of generative AI usage among university students. *International Journal of Educational Technology in Higher Education*, 21(1), 10. <https://doi.org/10.1186/s41239-024-00444-7>
- Alneyadi, S., & Wardat, Y. (2023). ChatGPT: Revolutionizing student achievement in the electronic magnetism unit for eleventh-grade students in Emirates schools. *Contemporary Educational Technology*, 15(4), ep448. <https://doi.org/10.30935/cedtech/13417>
- Baidoo-Anu, D., & Owusu Ansah, L. (2023). Education in the era of generative artificial intelligence (AI): Understanding the potential benefits of ChatGPT in promoting teaching and learning. *Journal of AI*, 7(1), 52-62. <https://doi.org/10.61969/jai.1337500>
- Bastani, H., Bastani, O., Sungu, A., Ge, H., Kabakcı, Ö., & Mariman, R. (2025). Generative AI without guardrails can harm learning: Evidence from high school mathematics. *Proceedings of the National Academy of Sciences*, 122(26), e2422633122. <https://doi.org/10.1073/pnas.2422633122>
- Chan, C. K. Y., & Hu, W. (2023). Students' voices on generative AI: Perceptions, benefits, and challenges in higher education. *International Journal of Educational Technology in Higher Education*, 20, 43. <https://doi.org/10.1186/s41239-023-00411-8>
- Chan, C. K. Y., & Lee, K. K. W. (2023). The AI generation gap: Are Gen Z students more interested in adopting generative AI such as ChatGPT in teaching and learning than their Gen X and millennial generation teachers? *Smart Learning Environments*, 10(1), 60. <https://doi.org/10.1186/s40561-023-00269-3>
- Deng, X., & Yu, Z. (2023). A meta-analysis and systematic review of the effect of chatbot technology use in sustainable education. *Sustainability*, 15(4), 2940. <https://doi.org/10.3390/su15042940>
- Departamento Administrativo Nacional de Estadística. (2024). Indicadores básicos de tenencia y uso de tecnologías de la información y la comunicación en hogares y personas de 5 y más años de edad, 2023. DANE. <https://www.dane.gov.co/index.php/estadisticas-por-tema/tecnologia-e-innovacion/tecnologias-de-la-informacion-y-las-comunicaciones-tic>
- De Simone, M. E., Tiberti, F. H., Barron Rodriguez, M. R., Manolio, F. A., Mosuro, W., & Dikoru, E. J. (2025). From chalkboards to chatbots: Evaluating the impact of generative AI on learning outcomes in Nigeria (Policy Research Working Paper No. 11125). World Bank. <https://doi.org/10.1596/1813-9450-11125>
- Doo, M. Y., & Park, Y. (2026). A meta-analysis of ChatGPT's influence on learning achievement. *The International Review of Research in Open and Distributed Learning*, 27(1), 265-289. <https://doi.org/10.19173/irrodl.v27i1.8775>
- Escalante, J., Pack, A., & Barrett, A. (2023). AI-generated feedback on writing: Insights into efficacy and ENL student preference. *International Journal of Educational Technology in Higher Education*, 20, 55. <https://doi.org/10.1186/s41239-023-00425-2>
- Fleckenstein, J., Liebenow, L. W., & Meyer, J. (2023). Automated feedback and writing: A multi-level meta-analysis of effects on students' performance. *Frontiers in Artificial Intelligence*, 6, 1162454. <https://doi.org/10.3389/frai.2023.1162454>
- Gómez Silva, C. A. (2016). Clasificación de colegios según las pruebas Saber 11 del ICFES: Un análisis usando modelos marginales. *Sociedad y Economía*, (30), 279-298. <https://www.redalyc.org/articulo.oa?id=99645393004>
- Laboratorio de Economía de la Educación. (2023). Características y retos de la educación rural en Colombia (Informe No. 79). Pontificia Universidad Javeriana.

- [https://www.javeriana.edu.co/recursosdb/5581483/8102914/Informe-79-Educacion-rural-en-Colombia-\(F\)oct.pdf](https://www.javeriana.edu.co/recursosdb/5581483/8102914/Informe-79-Educacion-rural-en-Colombia-(F)oct.pdf)
- Lo, C. K. (2023). What is the impact of ChatGPT on education? A rapid review of the literature. *Education Sciences*, 13(4), 410. <https://doi.org/10.3390/educsci13040410>
- Montenegro-Rueda, M., Fernández-Cerero, J., Fernández-Batanero, J. M., & López-Meneses, E. (2023). Impact of the implementation of ChatGPT in education: A systematic review. *Computers*, 12(8), 153. <https://doi.org/10.3390/computers12080153>
- 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>
- Pellas, N. (2023). The effects of generative AI platforms on undergraduates' narrative intelligence and writing self-efficacy. *Education Sciences*, 13(11), 1155. <https://doi.org/10.3390/educsci13111155>
- Rubiano Calderón, F. A. (2025). La inteligencia artificial en la educación rural de Colombia, derribando barreras desde la equidad. *Punto Educativo*, 4(1), 2227-2245. https://revistas.up.ac.pa/index.php/punto_educativo/article/view/8344
- Su, J., & Yang, W. (2023). Unlocking the power of ChatGPT: A framework for applying generative AI in education. *ECNU Review of Education*, 6(3), 355-366. <https://doi.org/10.1177/20965311231168423>
- Susnjak, T., & McIntosh, T. R. (2024). ChatGPT: The end of online exam integrity? *Education Sciences*, 14(6), 656. <https://doi.org/10.3390/educsci14060656>
- Tan, B., Armoush, N., Mazzullo, E., Bulut, O., & Gierl, M. (2025). A review of automatic item generation techniques leveraging large language models. *International Journal of Assessment Tools in Education*, 12(2), 317-340. <https://doi.org/10.21449/ijate.1602294>
- Wu, X., Zhu, P., Zhang, J., Yin, M., & Wang, Y. (2026). ChatGPT's impact on student learning outcomes: A meta-analysis of 35 experimental studies. *Humanities and Social Sciences Communications*, 13, 684. <https://doi.org/10.1057/s41599-026-07019-z>
- Yusuf, A., Pervin, N., & Román-González, M. (2024). Generative AI and the future of higher education: A threat to academic integrity or reformation? Evidence from multicultural perspectives. *International Journal of Educational Technology in Higher Education*, 21, 21. <https://doi.org/10.1186/s41239-024-00453-6>
- Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education – Where are the educators? *International Journal of Educational Technology in Higher Education*, 16, 39. <https://doi.org/10.1186/s41239-019-0171-0>