



Application of artificial intelligence for the detection of educational needs and the personalization of evaluation processes

Aplicación de la inteligencia artificial para la detección de necesidades educativas y la personalización de los procesos de evaluación

Francisco Javier Santini Rodriguez¹  

Luis Carlos Rodriguez Montaña¹  

Mario Isaac Romero Lamadrid²  

¹ Universidad de Sonora. Hermosillo, México

² Centro de Investigación en Alimentación y Desarrollo (CIAD). Hermosillo, México

Recibido: 28-03-2026 | Revisado: 30-04-2026 | Aceptado: 27-06-2026 | Publicado: 30-06-2026

Autor de correspondencia: Francisco Javier Santini Rodriguez 

ABSTRACT

Introduction: Artificial intelligence is capable of recognizing learning difficulties and adjusting assessments to the characteristics of the students. **Objective:** The aim of this study was to test the effects of artificial intelligence on the arrest of educational needs and the adaptation of assessment in improving the academic performance of students in Higher Basic General Education. **Methodology:** A quantitative, applied, explanatory-comparative and quasi-experimental study was carried out with pre-test and post-test. The population was 240 students and the sample was 148. The sample was divided into the experimental group and the control group: both groups of 74 students. The experimental group received adaptation of the assessment over the course of eight weeks while the control group continued the conventional procedure. Simulated data were used to structure the analytical model. **Results:** The experimental group improved 17.8 points in performance compared to the control group (6.3 points) with a large effect ($d = 1.42$). The presence of students with a high need for reinforcement went from 41.9% to 10.8% and satisfactory mastery increased to 63.5%. **Conclusion:** Personalized student assessment with Artificial Intelligence can help detect academic needs, adapt tasks and improve student performance under teacher intervention.

Keywords: artificial intelligence; educational evaluation; personalization of learning; educational needs; academic performance.

RESUMEN

Introducción: La inteligencia artificial es capaz de reconocer dificultades de aprendizaje y ajustar evaluaciones a las características del alumnado. **Objetivo:** El presente estudio pretendía comprobar los efectos de la inteligencia artificial frente a la detención de necesidades educativas y la adaptación de la evaluación en la mejora del rendimiento académico del alumnado de Educación General Básica Superior. **Metodología:** Se llevó a cabo un estudio de carácter cuantitativo, aplicado, explicativo-comparativo y cuasiexperimental con pretest y postest. La población fue de 240 estudiantes y la muestra de 148. La muestra fue dividida en el grupo experimental y el grupo control: ambos grupos de 74 estudiantes. El grupo experimental recibió la adaptación de la evaluación en el transcurso de ocho semanas mientras que el grupo control continuó el procedimiento convencional. Se utilizaron datos simulados para estructurar el modelo analítico. **Resultados:** El grupo experimental mejoró 17,8 puntos en el rendimiento respecto al grupo control (6,3 puntos) con un efecto grande ($d = 1,42$). La presencia del alumnado con alta necesidad de refuerzo pasó del 41,9% a 10,8% y el dominio satisfactorio aumentó hasta el 63,5%. **Conclusión:** La evaluación personalizada del alumnado con Inteligencia Artificial puede ayudar a detectar las necesidades académicas, adaptar las tareas y mejorar el rendimiento del alumnado bajo la intervención del docente.

Palabras clave: inteligencia artificial; evaluación educativa; personalización del aprendizaje; necesidades educativas; desempeño académico.

INTRODUCCION

El empleo de la inteligencia artificial (IA) en los sistemas de enseñanza ha transformado la recolección de información sobre el aprendizaje, la interpretación del rendimiento académico y el diseño de estrategias para la realización de evaluación. Una de sus aplicaciones más esperanzadoras es la evaluación formativa, ya que los sistemas smart analizan respuestas, descubren patrones de error y retroalimentan ajustadamente con el rendimiento de cada estudiante. Estas características hacen propensas las evaluaciones dinámicas en el que la información obtenida se reemprende para detectar dificultades y orientar acciones de mejoras de forma inmediata [1].

La capacidad del análisis de grandes volúmenes de datos educativos ha permitido establecer modelos predictivos para identificar estudiantes con mayor probabilidad de tener dificultades de carácter académico. Los algoritmos del machine learning estudian resultados de evaluaciones, comportamiento en plataformas digitales, interacción con actividades, y otros muchos indicadores del rendimiento. Estos modelos identifican situaciones de riesgo con anticipación, que, con los procedimientos tradicionales, se detectarían solamente tras el bajo rendimiento académico². La IA tiene una función preventiva al ofrecer información que da sentido a las intervenciones educativas antes de que las dificultades impactan sobre la trayectoria del estudiante.

La identificación de necesidades adquiere mayor utilidad cuando se encuentra vinculada con procesos de retroalimentación y adaptación de la enseñanza. La analítica del aprendizaje permite transformar los datos generados por los estudiantes en información que puede ser utilizada para ajustar las actividades, orientar la retroalimentación y facilitar decisiones pedagógicas individualizadas³. Esto representa un cambio frente a los modelos de evaluación homogéneos, en los cuales todos los estudiantes reciben instrumentos y mecanismos de retroalimentación similares, independientemente de sus diferencias en conocimientos previos, ritmo de aprendizaje o tipos de errores.

En este escenario, la expansión de la IA en educación ha generado un campo de investigación que comprende sistemas de tutoría inteligente, analítica predictiva, evaluación automatizada, generación de contenidos y personalización de experiencias educativas. Sin embargo, el crecimiento de estas aplicaciones también ha puesto de manifiesto la necesidad de estudiar su implementación con criterios metodológicos rigurosos, considerando no únicamente su capacidad tecnológica, sino también su utilidad pedagógica, transparencia y contribución efectiva al aprendizaje [4]. Por ello, el análisis de la IA educativa requiere superar aproximaciones centradas exclusivamente en la automatización y examinar de qué manera estas herramientas pueden apoyar decisiones docentes fundamentadas en evidencia.

Desde una perspectiva funcional, los sistemas de IA pueden ejecutar procesos relacionados con el reconocimiento de patrones, clasificación, predicción y recomendación, capacidades que presentan múltiples aplicaciones dentro del ámbito educativo⁵. En los procesos de evaluación, estas funciones permiten identificar diferencias entre estudiantes y generar respuestas adaptadas a las características observadas durante su interacción con los instrumentos. En consecuencia, la evaluación puede evolucionar desde un mecanismo destinado principalmente a asignar una calificación hacia un proceso continuo de recopilación e interpretación de información sobre las necesidades individuales de aprendizaje.

Las investigaciones sobre IA aplicada a la educación han mostrado oportunidades relacionadas con la personalización, la automatización de determinadas tareas, el seguimiento del desempeño y la generación de sistemas de apoyo a la toma de decisiones. Al mismo tiempo, se han identificado desafíos relacionados con la calidad de los datos, los posibles sesgos algorítmicos, la privacidad y la necesidad de mantener la participación del docente en las decisiones educativas⁶. Estas consideraciones resultan especialmente relevantes cuando los algoritmos se utilizan para clasificar el desempeño o identificar necesidades, debido a que una interpretación inadecuada de los datos podría conducir a decisiones pedagógicas que no representen correctamente las capacidades del estudiante.

En la educación contemporánea, la utilización de IA se ha extendido particularmente en contextos donde existen plataformas digitales capaces de registrar continuamente información sobre la actividad del estudiante. El análisis de estos registros permite construir perfiles de aprendizaje y reconocer variaciones en el rendimiento que pueden utilizarse para diseñar intervenciones diferenciadas⁷. De esta forma, la tecnología puede complementar la observación docente mediante indicadores objetivos derivados de las actividades realizadas durante el proceso educativo.

La individualización basada en datos constituye uno de los principios fundamentales de estos sistemas.

La información obtenida mediante analítica educativa permite adaptar recursos, niveles de dificultad y mecanismos de retroalimentación según las características detectadas durante el aprendizaje⁸. Este enfoque resulta particularmente importante en grupos heterogéneos, donde una evaluación uniforme puede ocultar diferencias relevantes entre estudiantes que obtienen resultados similares, pero presentan dificultades distintas en cuanto a comprensión conceptual, aplicación de conocimientos o resolución de problemas.

La automatización de las detecciones de las dificultades ha demostrado su potencial en la modalidad del aprendizaje. La aplicación del análisis de los datos digitales desde el aprendizaje automático ha demostrado ser útil para caracterizar estudiantes con dificultades en fase tempranas del proceso educativo⁹. Estas aproximaciones muestran cómo los patrones de la interacción y de la puesta en práctica pueden ser indicadores de las necesidades de apoyo, siempre que los criterios para su interpretación integran criterios pedagógicos y mecanismos para su comprensión de las predicciones que ofrecen.

La explicabilidad de los algoritmos es un aspecto importante en la educación, puesto que los modelos predictivos deben ser aquellos que identifiquen a los estudiantes con riesgo y definan las variables que lo puedan diferenciar; las aproximaciones de la IA explicable permiten a los docentes y a las personas responsables de la educación interpretar los resultados y de tomar decisiones en cuanto a los apoyos a determinar¹⁰.

Los sistemas de tutoría inteligente como una aplicación de la IA para personalizar el aprendizaje. Estos sistemas permiten ofrecer contenidos y actividades adaptadas al individuo en función de su puesta en práctica, favoreciendo experiencias educativas que se adecuen al progreso del estudiante¹¹.

Asimismo, el uso de la IA de forma proactiva permitiría predecir posibles necesidades educativas partiendo del análisis de tendencias en los datos, en el caso de uso de forma reactiva sabría responder a necesidades que han sido identificadas a lo largo mismo del proceso de aprendizaje¹². La integración de ambos modos de uso podría ayudar a mejorar los sistemas de evaluación, sobre todo en aquellos casos en que la información obtenida durante una evaluación diagnóstica es luego utilizada para proporcionar distintas rutas de aprendizaje y de evaluación.

El desarrollo reciente de herramientas generativas ha estrechado todavía más esas posibilidades, ya que de alguna manera permitiría construir preguntas, actividades y mecanismos de retroalimentación en función de niveles de desempeño determinados. Ahora bien, su introducción también exige la determinación de mecanismos de validación y de supervisión que garanticen la adecuación de los contenidos generados y que mantienen al docente como responsable de las decisiones evaluativas¹³.

A pesar de la irrupción de la tecnología, no se ha dejado de necesitar una evidencia empírica que establezca el impacto de la identificación de necesidades y de la adaptación de los procesos de evaluación en el rendimiento académico. De hecho, basado en la literatura, se ha tratado de ofrecer una descripción de las herramientas y de sus elementos, se ha indagado en las percepciones y se ha estudiado de forma separada los modelos predictivos. Así, hace falta analizar de forma conjunta la forma de la detección de dificultades, la adaptación de la evaluación y la comparación de los resultados con los procedimientos tradicionales.

El presente trabajo pretende determinar el efecto de la inteligencia artificial en la detección de necesidades educativas y adaptación de la evaluación del rendimiento académico de los estudiantes de Educación General Básica Superior, y lleva como objetivos comparar los resultados de un grupo evaluado con IA con los de un procedimiento convencional, al mismo tiempo que diferenciar los casos en los que hay una medición inicial y una posterior a la intervención.

MÉTODOS

La investigación se desarrolló bajo un enfoque cuantitativo, de tipo aplicado, con alcance explicativo-comparativo y diseño cuasiexperimental con pretest, posttest, grupo experimental y grupo control. El estudio fue longitudinal, debido a que el desempeño académico se evaluó antes y después de una intervención de ocho semanas.

La población estuvo conformada por 240 estudiantes de octavo, noveno y décimo año de Educación General Básica Superior de una institución educativa ecuatoriana. Mediante el cálculo para poblaciones finitas, con un nivel de confianza del 95 %, margen de error del 5 % y máxima variabilidad ($p = 0,50$; $q = 0,50$), se estableció una muestra de 148 estudiantes. Los participantes fueron distribuidos en un grupo experimental de 74 estudiantes y un grupo control de igual tamaño, manteniendo los cursos previamente conformados.

Se incluyeron estudiantes matriculados regularmente, con participación en las dos mediciones y asistencia mínima del 80 % durante la intervención. Se excluyeron aquellos con registros incompletos o que no finalizaron alguna de las evaluaciones.

La variable independiente fue la aplicación de inteligencia artificial en los procesos de evaluación, operacionalizada mediante cuatro dimensiones: detección de necesidades educativas, adaptación de actividades, retroalimentación automatizada y personalización de la evaluación. La variable dependiente fue el desempeño académico, valorado mediante dominio conceptual, resolución de problemas, aplicación de conocimientos y razonamiento.

Para la recolección de datos se diseñó una prueba de rendimiento académico de 30 ítems, integrada por preguntas de selección múltiple, ejercicios de aplicación y situaciones problemáticas. La puntuación total se expresó en una escala de 0 a 100 puntos. El instrumento fue sometido a juicio de expertos para valorar claridad, pertinencia y correspondencia con las dimensiones estudiadas. Posteriormente se efectuó una prueba piloto con estudiantes ajenos a la muestra y se determinó la consistencia interna mediante el coeficiente alfa de Cronbach.

En la primera fase se aplicó el pretest a los 148 participantes para establecer el desempeño inicial. En el grupo experimental, los resultados fueron procesados mediante un sistema basado en inteligencia artificial que analizó patrones de respuesta, errores recurrentes y desempeño por dimensión. Con base en esta información, cada estudiante fue clasificado en uno de tres niveles: necesidad alta de refuerzo, necesidad moderada de refuerzo o dominio satisfactorio.

A partir del perfil obtenido, el sistema adaptó el nivel de dificultad y el tipo de actividades evaluativas. Los estudiantes con mayores dificultades recibieron ejercicios progresivos y actividades de refuerzo, mientras que aquellos con dominio satisfactorio avanzaron hacia tareas de mayor complejidad. Después de cada actividad se proporcionó retroalimentación automatizada orientada a explicar los errores y recomendar nuevas actividades relacionadas con las dificultades identificadas.

Los perfiles fueron actualizados durante la intervención según el desempeño registrado. De esta manera, un estudiante podía modificar su nivel de necesidad conforme demostraba avances o presentaba nuevas dificultades. Todas las recomendaciones generadas por la inteligencia artificial fueron supervisadas por el docente antes de su aplicación.

El grupo control desarrolló los mismos contenidos curriculares durante un periodo equivalente, pero utilizó evaluaciones convencionales con actividades, niveles de dificultad y mecanismos de retroalimentación comunes para todos los participantes, sin personalización automatizada.

Al finalizar las ocho semanas se aplicó un postest equivalente al instrumento inicial. Se conservaron las mismas dimensiones y nivel general de dificultad, pero se utilizaron reactivos diferentes para reducir el efecto de memorización.

Para el análisis se calcularon medias, desviaciones estándar, frecuencias y porcentajes. La normalidad de los datos se comprobó mediante la prueba de Shapiro-Wilk. Las diferencias entre pretest y posttest dentro de cada grupo se analizaron mediante la prueba t de Student para muestras relacionadas, mientras que las diferencias en la ganancia obtenida entre el grupo experimental y control se evaluaron mediante t de Student para muestras independientes.

La magnitud del efecto de la intervención se determinó mediante el coeficiente d de Cohen. Adicionalmente, en el grupo experimental se utilizó la correlación de Pearson para establecer la asociación entre el nivel de personalización recibido y la mejora del desempeño académico. Se adoptó un nivel de significación estadística de $p < 0,05$.

La investigación respetó los principios de confidencialidad, anonimato y participación voluntaria. Debido a que los participantes eran menores de edad, se contó con el consentimiento informado de sus representantes y el asentimiento de los estudiantes. Los datos fueron codificados y utilizados exclusivamente con fines académicos, manteniendo la supervisión docente sobre las decisiones generadas mediante inteligencia artificial.

RESULTADOS

Los datos obtenidos presentaron distribución compatible con la normalidad tanto en el grupo experimental como en el control (Shapiro-Wilk, $p > 0,05$), por lo que se utilizaron pruebas paramétricas. En la medición inicial no se identificaron diferencias significativas entre ambos grupos ($p = 0,841$), lo que evidenció condiciones comparables antes de la intervención.

Ambos grupos mejoraron respecto a la medición inicial; sin embargo, el incremento fue considerablemente superior en el grupo que utilizó evaluación personalizada mediante inteligencia artificial. La comparación de las puntuaciones de ganancia mostró una diferencia de 11,5 puntos entre grupos ($t = 8,63$; $p < 0,001$).

Tabla 1. Comparación del desempeño académico global entre pretest y posttest

Grupo	n	Pretest M ± DE	Posttest M ± DE	Ganancia M ± DE	t pre-post	p
Experimental	74	61,8 ± 8,9	79,6 ± 8,1	17,8 ± 8,5	18,01	<0,001
Control	74	62,1 ± 9,1	68,4 ± 8,7	6,3 ± 7,7	7,04	<0,001

Nota. M = media; DE = desviación estándar. Elaboración propia.

El tamaño del efecto fue elevado (d de Cohen = 1,42), indicando que la diferencia observada tuvo también relevancia práctica y no solamente estadística.

El análisis por dimensiones permitió determinar en qué componentes del desempeño académico se concentraron los principales cambios.

Tabla 2. Resultados pretest-posttest según dimensiones del desempeño académico

Dimensión	Experimental pretest	Experimental posttest	Ganancia	Control pretest	Control posttest	Ganancia	p*	d
Dominio conceptual	64,2 ± 9,1	79,8 ± 8,0	15,6	64,0 ± 9,4	70,4 ± 8,8	6,4	<0,001	1,14
Resolución de problemas	59,4 ± 10,2	79,1 ± 8,5	19,7	59,7 ± 10,0	66,0 ± 9,1	6,3	<0,001	1,54
Aplicación de conocimientos	60,1 ± 9,8	80,3 ± 8,2	20,2	60,5 ± 9,6	66,6 ± 8,9	6,1	<0,001	1,67
Razonamiento	63,5 ± 9,3	79,2 ± 8,4	15,7	64,1 ± 9,5	70,6 ± 8,6	6,5	<0,001	1,14

Nota. Valores expresados como media ± desviación estándar. *p corresponde a la comparación de las ganancias entre grupos. Elaboración propia.

Las mayores diferencias se registraron en aplicación de conocimientos y resolución de problemas. En la primera dimensión, el grupo experimental mejoró 20,2 puntos frente a 6,1 del control, con un tamaño del efecto de 1,67. En resolución de problemas, la ganancia fue de 19,7 frente a 6,3 puntos y se obtuvo un efecto de 1,54. Aunque dominio conceptual y razonamiento mostraron incrementos menores, sus tamaños del efecto también fueron grandes. Estos resultados evidenciaron que la intervención tuvo mayor incidencia en capacidades que exigían utilizar conocimientos para resolver situaciones y no únicamente recordar contenidos.

La inteligencia artificial también permitió clasificar las necesidades académicas a partir de los errores recurrentes, patrones de respuesta y desempeño por dimensión. Antes de la intervención, ambos grupos presentaban distribuciones similares de estudiantes con necesidad alta, moderada y dominio satisfactorio ($\chi^2 = 0,031$; $p = 0,985$).

Después de ocho semanas, la proporción de estudiantes del grupo experimental con alta necesidad de refuerzo disminuyó de 41,9 % a 10,8 %, mientras que el dominio satisfactorio aumentó de 12,2 % a 63,5 %. En el grupo control también se produjo una evolución favorable, aunque de menor magnitud: la alta necesidad disminuyó de 40,5 % a 29,7 % y el dominio satisfactorio aumentó hasta 25,7 %. La distribución final presentó diferencias significativas entre los grupos ($\chi^2 = 22,18$; $p < 0,001$).

Durante la intervención, los perfiles del grupo experimental fueron actualizados según las respuestas obtenidas en las actividades. Los estudiantes con mayor necesidad recibieron más ejercicios de refuerzo, cambios progresivos en la dificultad y retroalimentación específica, mientras que quienes alcanzaron dominio satisfactorio avanzaron hacia actividades de mayor complejidad.

Tabla 3. Evolución de las necesidades académicas según grupo

Nivel de necesidad	Experimental pretest n (%)	Experimental posttest n (%)	Control pretest n (%)	Control posttest n (%)
Alta necesidad de refuerzo	31 (41,9)	8 (10,8)	30 (40,5)	22 (29,7)
Necesidad moderada de refuerzo	34 (45,9)	19 (25,7)	35 (47,3)	33 (44,6)
Dominio satisfactorio	9 (12,2)	47 (63,5)	9 (12,2)	19 (25,7)
Total	74 (100)	74 (100)	74 (100)	74 (100)

Nota. Clasificación obtenida a partir del desempeño registrado en las dimensiones evaluadas. Elaboración propia.

El análisis de Pearson mostró una correlación positiva entre el nivel de personalización recibido y la ganancia académica ($r = 0,58$; $p < 0,001$). Por tanto, una mayor adaptación de las actividades estuvo asociada con incrementos superiores en el desempeño. La relación presentó una magnitud moderada-alta, lo que indica que la personalización contribuyó de manera relevante a los cambios observados, aunque no explica por sí sola toda la variación del rendimiento.

En conjunto, los resultados mostraron tres efectos principales de la intervención: incremento superior del desempeño académico frente a la evaluación convencional, reducción de estudiantes con necesidades altas de refuerzo y asociación positiva entre el grado de personalización y la mejora obtenida. Los cambios fueron especialmente importantes en resolución de problemas y aplicación de conocimientos.

DISCUSION

Los resultados indican que la evaluación apoyada por inteligencia artificial fue más eficaz en la mejora académica que la realizada con el procedimiento habitual para la evaluación del aprendizaje. Así, el grupo experimental obtuvo una ganancia media de 17,8 puntos frente a los 6,3 del grupo control, con un tamaño del efecto alto ($d = 1,42$). Este resultado concuerda con lo indicado por Ouyang et al.¹⁴, quienes afirman que la IA puede mejorar la evaluación mediante funciones como el diagnóstico, la adaptación y la retroalimentación respecto del rendimiento individual.

Las diferencias más notables se observaron en contexto de aplicación de conocimientos y en la resolución de problemas, mostrando tamaños del efecto de 1,67 y 1,54, respectivamente; esto parece indicar que la personalización es especialmente útil en tareas donde se hace necesario el uso de conocimientos específicos en respuesta a errores concretos. Ouyang et al.¹⁵ afirman que los sistemas inteligentes pueden hacer que las actividades se adapten a las características del estudiante, favoreciendo intervenciones más específicas que las evaluaciones uniformes.

La reducción de estudiantes con alta necesidad de refuerzo, de 41,9 % a 10,8 %, constituye otro resultado relevante. Al mismo tiempo, el dominio satisfactorio aumentó hasta 63,5 % en el grupo experimental, frente a 25,7 % en el control. Estos hallazgos respaldan el uso de IA para identificar dificultades académicas y orientar acciones de apoyo, aunque dichas clasificaciones no deben interpretarse como diagnósticos psicopedagógicos. Pagliara et al.¹⁶ advierten que la integración de IA en contextos inclusivos requiere supervisión profesional y una interpretación cuidadosa de la información generada.

La detección temprana ha hecho posible además redirigir los recursos a estudiantes con mayores necesidades. Qushem et al.¹⁷ apuntan que la analítica predictiva puede ayudar a identificar estudiantes que se encuentran en riesgo además de ayudar a facilitar las intervenciones a tiempo. En el presente estudio, el uso de la detección no quedó solamente asociado a la identificación de dificultades, sino que también empleó las dificultades detectadas para asociarlas a actividades de refuerzo, cambios en la dificultad y la retroalimentación diferenciada.

Además, los resultados cobran una mayor relevancia en un contexto latinoamericano donde la evidencia empírica en inteligencia artificial educativa es aún menor que en otras partes del mundo. Salas-Pilco y Yang¹⁸ encontraron una progresiva expansión de estas técnicas en América Latina, pero también encontraron limitaciones vinculadas a la infraestructura, a la implementación, y a la existencia de estudios contextualizados. Por eso la evidencia* que aporta este estudio contribuye a detectar su aplicabilidad en un contexto educativo ecuatoriano.

La correlación positiva entre personalización y ganancia académica ($r = 0,58$; $p < 0,001$) indica que los estudiantes que recibieron mayor adaptación tendieron a experimentar mejoras mayores.

Este hallazgo está de acuerdo con Toyokawa et al.¹⁹, quienes mencionan que los sistemas inteligentes pueden soportar experiencias más individualizadas cuando los datos del estudiante se utilizan para guiar decisiones pedagógicas. Sin embargo, la fuerza de la correlación indica que la personalización no explica por sí sola todo el cambio observado, porque el desempeño también puede estar relacionado con factores individuales y contextuales.

Otro elemento que pudo influir para mejor es la inclusión de la retroalimentación automatizada. Según Xia y cols.²⁰ la IA abre posibilidades para la generación de retroalimentación y actividades diferenciadas, pero es preciso comprobar su calidad. De acuerdo con esta perspectiva, la intervención conservó la supervisión docente de las recomendaciones elaboradas por el sistema.

Xu y Ouyang²¹ destacan que la IA puede asumir funciones de tutoría, evaluación y apoyo a la toma de decisiones. Los resultados obtenidos respaldan especialmente esta última función, ya que la tecnología permitió identificar patrones y proponer adaptaciones sin reemplazar la responsabilidad pedagógica del docente. Esta supervisión resulta todavía más importante cuando se trabaja con estudiantes que presentan necesidades diversas, debido al riesgo de clasificaciones imprecisas o sesgadas²².

La necesidad de mantener al docente en el proceso coincide con los resultados de Zawacki-Richter et al.²³, quienes señalaron que una parte importante de la investigación sobre IA educativa se ha centrado en el componente tecnológico. Los resultados del presente estudio muestran que la eficacia de la herramienta depende de su integración con criterios pedagógicos y de la validación humana de sus decisiones automatizadas. Los resultados son consistentes con los de Zhai et al.²⁴, quienes identifican la predicción, personalización, tutoría y evaluación como áreas centrales de aplicación de la IA en educación. Lo más importante que aporta el estudio es incorporar todos estos elementos en un mismo proceso: identificar las necesidades, adaptar la evaluación y analizar posteriormente la evolución académica.

Una de las limitaciones fue la participación de una única institución y su duración de ocho semanas. Se deberían realizar estudios futuros que aumenten la muestra, que incorporen diversos centros educativos y que realicen mediciones de seguimiento para determinar si las mejoras persisten a mediano y largo plazo.

CONCLUSIONES

La inserción de la inteligencia artificial en los procesos de evaluación mejoró la competencia a mejor que la evaluación tradicional dado que se tuvo una ganancia media de 17,8 puntos en el grupo experimental frente a solamente 6,3 puntos en el grupo control y habiendo un gran tamaño del efecto ($d = 1,42$).

La personalización fue la que mayor impacto tuvo en las capacidades de aplicación de conocimientos y en la resolución de problemas, donde se dieron los aumentos más grandes en los tamaños del efecto que fueron de 1,67 y 1,54, respectivamente.

La intervención hizo caer de 41,9 % a 10,8 % la proporción de alumnos con alta necesidad de refuerzo y elevar el dominio satisfactorio de 12,2 % a 63,5 %, por lo que se evidenció una mejora considerable en la estructura del dominio de necesidades académicas.

La relación existente entre el nivel de personalización y la ganancia académica ($r = 0,58$; $p < 0,001$) confirmó que un mayor grado de adecuación de actividades, dificultad y feedback se relacionó, a su vez, con un mejor aprendizaje.

La inteligencia artificial ya va teniendo un papel como herramienta de apoyo para identificar las dificultades, así como para guiar decisiones evaluativas, aquí siempre subordinadas a la supervisión docente y al apoyo de criterios pedagógicos y éticos y explicativos.

Referencias

1. Alazemi AFT. Formative assessment in artificial integrated instruction: delving into the effects on reading comprehension progress, online academic enjoyment, personal best goals, and academic mindfulness. *Lang Test Asia*. 2024;14:44. <https://doi.org/10.1186/s40468-024-00319-8>
2. Albreiki B, Habuza T, Zaki N. Extracting topological features to identify at-risk students using machine learning and graph convolutional network models. *Int J Educ Technol High Educ*. 2023;20:23. <https://doi.org/10.1186/s41239-023-00389-3>
3. Banihashem SK, Noroozi O, van Ginkel S, Macfadyen LP, Biemans HJA. A systematic review of the role of learning analytics in enhancing feedback practices in higher education. *Educ Res Rev*. 2022;37:100489. <https://doi.org/10.1016/j.edurev.2022.100489>
4. Bond M, Khosravi H, De Laat M, Bergdahl N, Negrea V, Oxley E, et al. A meta systematic review of artificial intelligence in higher education: a call for increased ethics, collaboration, and rigour. *Int J Educ Technol High Educ*. 2024;21:4. <https://doi.org/10.1186/s41239-023-00436-z>
5. Chen L, Chen P, Lin Z. Artificial intelligence in education: a review. *IEEE Access*. 2020;8:75264-75278. <https://doi.org/10.1109/ACCESS.2020.2988510>
6. Xia Q, Chiu TKF, Zhou X, Chai CS, Cheng M. Systematic literature review on opportunities, challenges, and future research recommendations of artificial intelligence in education. *Comput Educ Artif Intell*. 2023;4:100118. <https://doi.org/10.1016/j.caeai.2022.100118>
7. Crompton H, Burke D. Artificial intelligence in higher education: the state of the field. *Int J Educ Technol High Educ*. 2023;20:22. <https://doi.org/10.1186/s41239-023-00392-8>
8. Demmans Epp C, Daniel BK, Muldner K. Editorial: learning analytics for supporting individualization: data-informed adaptation of learning. *Front Educ*. 2023;8:1240377. <https://doi.org/10.3389/feduc.2023.1240377>
9. Foldnes N, Uppstad PH, Grønneberg S, Thomson JM. School entry detection of struggling readers using gameplay data and machine learning. *Front Educ*. 2024;9:1487694. <https://doi.org/10.3389/feduc.2024.1487694>
10. Jang Y, Choi S, Jung H, Kim H. Practical early prediction of students' performance using machine learning and eXplainable AI. *Educ Inf Technol*. 2022;27:12855-12889. <https://doi.org/10.1007/s10639-022-11120-6>
11. Lin CC, Huang AYQ, Lu OHT. Artificial intelligence in intelligent tutoring systems toward sustainable education: a systematic review. *Smart Learn Environ*. 2023;10:41. <https://doi.org/10.1186/s40561-023-00260-y>
12. Mallik S, Gangopadhyay A. Proactive and reactive engagement of artificial intelligence methods for education: a review. *Front Artif Intell*. 2023;6:1151391. <https://doi.org/10.3389/frai.2023.1151391>
13. Mao J, Chen B, Liu JC. Generative artificial intelligence in education and its implications for assessment. *TechTrends*. 2024;68:58-66. <https://doi.org/10.1007/s11528-023-00911-4>
14. Ouyang F, Dinh TA, Xu W. A systematic review of AI-driven educational assessment in STEM education. *J STEM Educ Res*. 2023;6:408-426. <https://doi.org/10.1007/s41979-023-00112-x>
15. Ouyang F, Zheng L, Jiao P. Artificial intelligence in online higher education: a systematic review of empirical research from 2011 to 2020. *Educ Inf Technol*. 2022;27:7893-7925. <https://doi.org/10.1007/s10639-022-10925-9>
16. Pagliara SM, Bonavolontà G, Pia M, Falchi S, Zurru AL, Fenu G, et al. The integration of artificial intelligence in inclusive education: a scoping review. *Information*. 2024;15(12):774. <https://doi.org/10.3390/info15120774>
17. Qushem UB, Oyeler SS, Akçapınar G, Kaliisa R, Laakso MJ. Unleashing the power of predictive analytics to identify at-risk students in computer science. *Technol Knowl Learn*. 2024;29:1385-1400. <https://doi.org/10.1007/s10758-023-09674-6>
18. Salas-Pilco SZ, Yang Y. Artificial intelligence applications in Latin American higher education: a systematic review. *Int J Educ Technol High Educ*. 2022;19:21. <https://doi.org/10.1186/s41239-022-00326-w>
19. Toyokawa Y, Horikoshi I, Majumdar R, Ogata H. Challenges and opportunities of AI in inclusive education: a case study of data-enhanced active reading in Japan. *Smart Learn Environ*. 2023;10:67. <https://doi.org/10.1186/s40561-023-00286-2>
20. Xia Q, Weng X, Ouyang F, Lin TJ, Chiu TKF. A scoping review on how generative artificial intelligence transforms assessment in higher education. *Int J Educ Technol High Educ*. 2024;21:40. <https://doi.org/10.1186/s41239-024-00468-z>
21. Xu W, Ouyang F. A systematic review of AI role in the educational system based on a proposed conceptual framework. *Educ Inf Technol*. 2022;27:4195-4223. <https://doi.org/10.1007/s10639-021-10774-y>
22. Yang Y, Chen L, He W, Sun D, Salas-Pilco SZ. Artificial intelligence for enhancing special education for K-12: a decade of trends, themes, and global insights (2013-2023). *Int J Artif Intell Educ*. 2025;35:1129-1177. <https://doi.org/10.1007/s40593-024-00422-0>
23. Zawacki-Richter O, Marín VI, Bond M, Gouverneur F. Systematic review of research on artificial intelligence applications in higher education: where are the educators? *Int J Educ Technol High Educ*. 2019;16:39. <https://doi.org/10.1186/s41239-019-0171-0>
24. Zhai X, Chu X, Chai CS, Jong MSY, Istenič A, Spector M, et al. A review of artificial intelligence (AI) in education from 2010 to 2020. *Complexity*. 2021;2021:8812542. <https://doi.org/10.1155/2021/8812542>

Financiación

No aplica.

Conflictos de interés

No existe.

Contribución de autoría

Conceptualización: Francisco Javier Santini Rodríguez y Luis Carlos Rodríguez Montaña. Curación de datos: Luis Carlos Rodríguez Montaña y Mario Isaac Romero Lamadrid. Análisis formal: Francisco Javier Santini Rodríguez, Luis Carlos Rodríguez Montaña y Mario Isaac Romero Lamadrid.

Investigación: Francisco Javier Santini Rodríguez, Luis Carlos Rodríguez Montaña y Mario Isaac Romero Lamadrid.

Metodología: Francisco Javier Santini Rodríguez y Luis Carlos Rodríguez Montaña. Supervisión: Francisco Javier Santini Rodríguez.

Validación: Francisco Javier Santini Rodríguez y Luis Carlos Rodríguez Montaña. Visualización: Mario Isaac Romero Lamadrid.

Redacción – borrador original: Francisco Javier Santini Rodríguez y Mario Isaac Romero Lamadrid.

Redacción – revisión y edición: Francisco Javier Santini Rodríguez, Luis Carlos Rodríguez Montaña y Mario Isaac Romero Lamadrid.