Tribunal. Revista en Ciencias de la Educación y Ciencias Jurídicas / http://www.revistatribunal.org
Volumen 6 / N° 16 / julio - septiembre 2026 / ISSN: 2959-6513 / pp. 1 – 17
http://doi.org/10.59659/revistatribunal.v6i16.466

Research Article

Tutoría adaptativa con aprendizaje por refuerzo profundo para comprensión lectora en literatura peruana

Adaptive tutoring for reading comprehension in Peruvian literature: A deep reinforcement learning approach

Amit Roy Flores Rivera[1]
Yasmin Mariela Urrego Montoya[2]
Yuli Torres Soldevilla[3]
Uldarico Pillaca Esquivel[4]

Resumen

Introducción: La integración de sistemas inteligentes en educación ha impulsado nuevas formas de personalización del aprendizaje. En este contexto, los modelos de tutoría adaptativa basados en aprendizaje por refuerzo profundo permiten ajustar dinámicamente la instrucción según el desempeño del estudiante. **Objetivo:** Analizar un sistema de tutoría adaptativa basado en aprendizaje por refuerzo profundo para personalizar la comprensión lectora con literatura peruana. **Metodología:** Se desarrolló un estudio cuantitativo de tipo aplicado, con diseño experimental computacional y alcance explicativo. La población estuvo constituida por interacciones simuladas en comprensión lectora, y la muestra por episodios generados durante 500 000 pasos de entrenamiento, utilizando diez preguntas basadas en el texto “Paco Yunque”. Se emplearon simulaciones computacionales, registro de métricas e instrumentos de validación algorítmica. **Resultados:** El modelo basado en aprendizaje por refuerzo profundo (PPO) mostró una precisión promedio de 38,23 %, un retorno medio de 4,35 y una desviación estándar de 3,86, evidenciando estabilidad en el proceso de toma de decisiones pedagógicas del agente, aunque con rendimiento moderado en precisión. **Conclusiones:** El agente PPO demuestra capacidad para generar decisiones pedagógicas consistentes en entornos de tutoría adaptativa. Sin embargo, se requiere ampliar el banco de preguntas y realizar validaciones con estudiantes reales para fortalecer la generalización y aplicabilidad del sistema en contextos educativos auténticos.

Palabras clave: Aprendizaje adaptativo; Aprendizaje por refuerzo; Comprensión lectora; Inteligencia artificial; Literatura peruana; Tutoría inteligente

Abstract

Introducción: La integración de sistemas inteligentes en educación ha impulsado nuevas formas de personalización del aprendizaje. En este contexto, los modelos de tutoría adaptativa basados en aprendizaje por refuerzo profundo permiten ajustar dinámicamente la instrucción según el desempeño del estudiante. **Objetivo:** Analizar un sistema de tutoría adaptativa basado en aprendizaje por refuerzo profundo para personalizar la comprensión lectora con literatura peruana. **Metodología:** Se desarrolló un estudio cuantitativo de tipo aplicado, con diseño experimental computacional y alcance explicativo. La población estuvo constituida por interacciones simuladas en comprensión lectora, y la muestra por episodios generados durante 500 000 pasos de entrenamiento, utilizando diez preguntas basadas en el texto “Paco Yunque”.

[1] Universidad Nacional Autónoma de Huanta. Huanta, Perú, [2] Universidad Nacional de Ingeniería. Lima, Perú, [3-4] Colegio de Contadores Públicos de Ayacucho. Ayacucho, Perú
Artículo recibido 14 de marzo 2026 | Aceptado 18 de mayo 2026 | Publicado 1 de julio 2026
Autor de correspondencia: aflores@unah.edu.pe
Conflicto de interés: Los autores declaran no tener conflicto de intereses.
Como citar: Flores Rivera, A. R., Urrego Montoya, Y. M., Torres Soldevilla, Y., & Pillaca Esquivel, U. (2026). Tutoría adaptativa con aprendizaje por refuerzo profundo para comprensión lectora en literatura peruana. *Tribunal. Revista en Ciencias de la Educación y Ciencias Jurídicas*, 6(16), 1-17. http://doi.org/10.59659/revistatribunal.v6i16.466

Se emplearon simulaciones computacionales, registro de métricas e instrumentos de validación algorítmica. **Resultados:** El modelo basado en aprendizaje por refuerzo profundo (PPO) mostró una precisión promedio de 38,23 %, un retorno medio de 4,35 y una desviación estándar de 3,86, evidenciando estabilidad en el proceso de toma de decisiones pedagógicas del agente, aunque con rendimiento moderado en precisión. **Conclusiones:** El agente PPO demuestra capacidad para generar decisiones pedagógicas consistentes en entornos de tutoría adaptativa. Sin embargo, se requiere ampliar el banco de preguntas y realizar validaciones con estudiantes reales para fortalecer la generalización y aplicabilidad del sistema en contextos educativos auténticos.

Palabras clave: Aprendizaje adaptativo; Aprendizaje por refuerzo; Comprensión lectora; Inteligencia artificial; Literatura peruana; Tutoría inteligente

Introducción

La comprensión lectora constituye una competencia transversal para el aprendizaje escolar, la participación ciudadana y la reducción de desigualdades educativas; no obstante, continúa siendo una de las áreas de mayor rezago en sistemas educativos con brechas socioeconómicas persistentes. En el caso peruano, PISA 2022 informó que solo el 50 % de estudiantes alcanzó al menos el nivel 2 en lectura, mientras que el promedio de la OCDE fue 74 %, y apenas el 1 % llegó a niveles altos de desempeño lector (OECD, 2023). De manera complementaria, esta situación coincide con el indicador de pobreza de aprendizaje, definido como la imposibilidad de leer y comprender un texto simple hacia los diez años, que en Perú fue estimado en 56 % para 2019 (World Bank, 2019). En consecuencia, el problema de investigación se ubica en la necesidad de diseñar mecanismos de personalización que respondan a la heterogeneidad de ritmos, niveles de dominio, motivación, frustración y contexto cultural de los estudiantes, sin asumir que una secuencia fija de actividades produce los mismos efectos en todos los aprendices.

En este marco, el uso de inteligencia artificial en educación se ha propuesto como una vía para ampliar la capacidad de diagnóstico, retroalimentación y adaptación de los entornos de aprendizaje, aunque su incorporación debe responder a criterios de evidencia, equidad y pertinencia pedagógica (UNESCO, 2023). De manera consistente, Zawacki-Richter, Olaf et al. (2019), identificaron que las aplicaciones de IA en educación se concentran en personalización, analítica del aprendizaje y sistemas adaptativos, aunque con limitada evidencia en contextos reales de aula.

En efecto, los sistemas de tutoría inteligente han mostrado efectos positivos en múltiples dominios; en una revisión meta-analítica de 50 evaluaciones controladas, Kulik y Fletcher (2016) reportaron un efecto mediano equivalente a elevar el rendimiento del percentil 50 al 75, mientras que Steenbergen-Hu y Cooper (2013) identificaron beneficios de los tutores inteligentes en matemática escolar. En esta misma línea, VanLehn, (2011) mostró que los sistemas de tutoría inteligente pueden aproximarse en efectividad a la tutoría humana bajo condiciones controladas, lo que refuerza su potencial como alternativa escalable de personalización.

Asimismo, Wang et al. (2024) hallaron que los sistemas adaptativos habilitados por IA tuvieron un efecto positivo medio-alto en resultados cognitivos, y Li et al. (2023) formularon el aprendizaje adaptativo como un proceso de decisión de Markov capaz de

seleccionar materiales según rasgos latentes continuos del estudiante. En conjunto, estos antecedentes demuestran que la personalización computacional puede superar limitaciones de instrucción uniforme, pero también advierten que su eficacia depende de la calidad del modelo del estudiante, de la alineación entre objetivos y evaluación, y de la suficiencia de datos pedagógicamente significativos.

Desde una perspectiva especifica de la comprensión lectora, el antecedente de Wijekumar et al. (2012) es relevante porque evaluó a gran escala la tutoría inteligente de la estrategia de estructura textual en cuarto grado, y Wijekumar et al. (2012) corroboraron, mediante un ensayo aleatorizado multisitio, efectos de la tutoría inteligente para lectores de quinto grado. En una línea complementaria, Abraham (2008) mostró, mediante meta-análisis, que los apoyos mediados por computadora pueden favorecer lectura en segunda lengua y aprendizaje de vocabulario, mientras que McNamara et al. (2004) sostuvieron que los tutores de estrategias de lectura, como iSTART, pueden promover explicaciones, inferencias y monitoreo metacognitivo.

A partir de estos aportes, es posible reconocer que la lectura no solo demanda seleccionar preguntas, sino también ofrecer andamiajes, pistas, retroalimentación y niveles de desafío coherentes con el desempeño. En ese sentido, la teoría sociocultural de Vygotsky (1978) y el problema de las dos sigmas de Bloom (1984) siguen siendo marcos pertinentes: la personalización intenta aproximar, con recursos computacionales, la mediación ajustada que caracteriza a la tutoría individual, aunque sin reemplazar la función pedagógica docente.

Por otra parte, en el campo de los sistemas adaptativos, Anderson et al. (1995) mostraron que los tutores cognitivos se benefician de modelos explícitos del conocimiento, Corbett y Anderson (1994) propusieron el rastreo de conocimiento para estimar adquisición procedimental, y Piech et al. (2015) extendieron esa lógica con rastreo de conocimiento profundo basado en redes neuronales. No obstante, estas aproximaciones tienden a centrarse en la predicción del desempeño y no necesariamente en la decisión pedagógica óptima a largo plazo. En contraste, el aprendizaje por refuerzo ofrece un marco distinto porque permite que un agente aprenda políticas de acción a partir de recompensas acumuladas, como explican Sutton y Barto (2018), y se ha aplicado al aprendizaje de estrategias docentes en entornos adaptativos (Iglesias et al., 2009), a la planificación de rutas de aprendizaje jerárquicas (Li et al., 2021) y a sistemas de recomendación educativa basados en aprendizaje temporal (Lan et al., 2014).

En particular, la optimización de política proximal se considera una alternativa estable para actualizar políticas en entornos con acciones múltiples, debido al objetivo recortado que limita cambios abruptos de la política (Schulman et al., 2017), y el aprendizaje por refuerzo profundo permite integrar representaciones complejas del estado cuando las reglas discretas resultan insuficientes (François-Lavet et al., 2018).

En coherencia con lo anterior, la problemática también requiere atender el contexto cultural del material lector.

En el ámbito regional, en América Latina, Salas-Pilco y Yang (2022) evidenciaron que las aplicaciones de IA educativa se concentran en analítica, predicción y apoyo al aprendizaje, pero todavía requieren mayor contextualización institucional y social. De forma convergente, desde una perspectiva regional ampliada, Saavedra y Molina (2026)

advirtieron que la IA puede personalizar el aprendizaje, aunque también puede ampliar la brecha entre estudiantes empoderados, dependientes o excluidos por conectividad e infraestructura. Por su parte, en Norteamérica, Chen y Perez (2023) destacaron el potencial de la IA para evaluación y aprendizaje personalizado, pero subrayaron la necesidad de principios pedagógicos y éticos.

En contraste geográfico, en Asia y otros sistemas comparados, UNESCO (2022) documentó currículos nacionales de IA que muestran una incorporación progresiva de competencias digitales desde la educación básica. Estas perspectivas geográficas convergen en una idea central: la innovación tecnológica solo adquiere sentido educativo cuando se articula con equidad, contenido local, transparencia y mediación docente.

De manera integradora, la literatura especializada ha señalado que los sistemas adaptativos deben combinar evidencia de aprendizaje, decisiones pedagógicas interpretables y evaluación responsable. Kulik y Fletcher (2016), Steenbergen-Hu y Cooper (2013) y Wang et al. (2024) respaldan el potencial de la tutoría inteligente; Anderson et al. (1995), Corbett y Anderson (1994) y Piech et al. (2015) muestran la importancia de modelar conocimiento; e Iglesias et al. (2009), Li et al. (2021), Li et al. (2023), Lan et al. (2014), Schulman et al. (2017), Sutton y Barto (2018) y François-Lavet et al. (2018) fundamentan la decisión adaptativa. En el dominio especifico de la lectura, Wijekumar et al. (2012), Wijekumar et al. (2014), McNamara et al. (2004) y Abraham (2008) evidencian que la retroalimentación computacional puede apoyar la comprensión. Por consiguiente, el problema no consiste únicamente en incorporar IA, sino en diseñarla con pertinencia cultural, validez pedagógica y evaluación verificable.

En esta línea, la elección de *Paco Yunque*, de César Vallejo, responde a esa necesidad de contextualización porque la obra permite trabajar comprensión literal, inferencial y crítica a partir de temas de desigualdad, abuso de poder, discriminación y escuela. Así, a diferencia de plataformas globales con contenidos estandarizados, un sistema adaptativo basado en literatura peruana puede fortalecer la comprensión lectora y sostener una experiencia culturalmente significativa.

Finalmente, el estudio se justifica porque integra un algoritmo de aprendizaje por refuerzo profundo con un banco de preguntas literarias, un modelo de estado que incorpora dimensiones cognitivas, emocionales y temporales, y una función de recompensa multiobjetivo orientada a balancear exactitud, progreso, compromiso, eficiencia y ajuste de dificultad. En este sentido, esta contribución resulta pertinente para contextos de bajos recursos, pues el prototipo fue entrenado en CPU y se orienta a escenarios donde la personalización debe ser eficiente, replicable y culturalmente situada. Por tanto, el objetivo del estudio fue analizar un sistema de tutoría adaptativa basado en aprendizaje por refuerzo profundo para personalizar la comprensión lectora mediante literatura peruana, evaluando su estabilidad, precisión y comportamiento pedagógico en interacciones simuladas.

Método

El estudio se desarrolló con enfoque cuantitativo, de tipo aplicado, diseño experimental computacional y alcance explicativo, debido a que se implementó y evaluó un agente de

aprendizaje por refuerzo profundo en un entorno controlado de simulación para estimar su capacidad de tomar decisiones pedagógicas adaptativas. La elección metodológica se justificó porque el objetivo no fue describir percepciones ni explorar significados subjetivos, sino medir el comportamiento de una política algorítmica ante estados de estudiante definidos formalmente, recompensas observables y episodios repetibles de interacción.

En esta lógica, el problema de tutoría se formuló como un proceso de decisión de Markov compuesto por estados del estudiante, acciones pedagógicas, transición de desempeño, recompensa multiobjetivo y factor de descuento, coherente con la teoría del aprendizaje por refuerzo (Sutton & Barto, 2018) y con aplicaciones recientes de aprendizaje adaptativo mediante decisiones secuenciales (Li et al., 2023). El contexto de estudio fue un entorno computacional denominado *PacoYunqueTutoringEnv*, construido con Gymnasium y orientado a simular una sesión de tutoría para comprensión lectora de *Paco Yunque*, cuento de César Vallejo.

La población de análisis estuvo constituida por las interacciones posibles entre un tutor adaptativo y estudiantes simulados durante sesiones de comprensión lectora; la muestra fue no probabilística, intencional y algorítmica, conformada por los episodios generados durante 500 000 pasos de entrenamiento y validación. Este tipo de muestra fue pertinente porque la unidad de análisis no correspondió a estudiantes humanos, sino a trayectorias de decisión y respuesta simulada, lo que permitió evaluar estabilidad, retorno, precisión y variación por dificultad antes de una validación educativa presencial. La muestra incluyó interacciones con diez preguntas de opción múltiple, distribuidas en cuatro preguntas básicas, tres intermedias y tres avanzadas, relacionadas con personajes, injusticia, educación, discriminación y violencia.

El instrumento principal fue el entorno de simulación con un vector de estado de 16 dimensiones. Dicho vector incorporó el nivel del estudiante, cinco puntajes de dominio temático, precisión reciente, precisión total, progreso normalizado, compromiso, frustración, confianza, duración de sesión, tiempo de respuesta y rachas de respuestas correctas e incorrectas. El espacio de acción fue multi-discreto y permitió optimizar de manera simultánea cuatro dimensiones: selección de tema con cinco opciones, ajuste de dificultad con tres niveles, nivel de pistas con tres estrategias y metodología de enseñanza con cuatro enfoques, lo que produjo 180 acciones conjuntas posibles. La técnica de recolección fue el registro automatizado de métricas del entrenamiento y la validación; los instrumentos fueron archivos de registro, puntos de control del modelo, métricas de precisión, retorno, entropía, pérdida de política, pérdida de valor, distribución de dificultad y trayectorias de estudiante.

La validez de contenido del banco de preguntas se sostuvo mediante la correspondencia entre cada ítem y los niveles de comprensión literal, inferencial y crítica de la obra literaria seleccionada. La validez algorítmica se verificó por consistencia interna de la arquitectura, separación entre entrenamiento y validación determinística, definición explícita de la función de recompensa y seguimiento de métricas de estabilidad. La confiabilidad se estimó a través de la repetición de episodios de validación cada 5 000 pasos, con veinte episodios por punto de control, así como mediante la desviación estándar de la precisión final. Aunque no se trató de confiabilidad psicométrica aplicada a respuestas humanas, el procedimiento permitió examinar la estabilidad del agente bajo

condiciones replicables de simulación, lo cual resulta adecuado para una fase de prototipado computacional.

El algoritmo utilizado fue optimización de política proximal, con arquitectura actor-crítico y codificador compartido. La red recibió un vector de entrada de 16 dimensiones y lo procesó mediante capas densas de 512, 256 y 128 unidades, normalización por capa, activación GELU y abandono de 0,10. La salida del actor incluyó cuatro cabezas categóricas para tema, dificultad, pistas y estrategia; la salida del crítico estimó el valor del estado mediante una cabeza de regresión. La función de recompensa integró cinco componentes ponderados: corrección de respuesta con peso de 35 %, progreso de aprendizaje con 25 %, compromiso con 20 %, eficiencia temporal con 10 % y emparejamiento de dificultad con 10 %. Esta estructura permitió evitar una optimización centrada únicamente en exactitud y favoreció decisiones pedagógicas que también consideraran frustración, confianza, ritmo y ajuste de desafío.

El entrenamiento se ejecutó con 500 000 pasos totales, longitud de despliegue de 2 048, tamaño de lote de 64, mini-lote de 32, diez épocas de optimización, tasa de aprendizaje de 2,5 × 10^-4, descuento $\gamma = 0,99$, parámetro GAE $\lambda = 0,95$, epsilon de recorte de 0,20, coeficiente de valor de 0,50, coeficiente de entropía de 0,01 y recorte máximo de gradiente de 0,50. La evaluación fue determinística, mediante selección argmax de las acciones, para reducir variabilidad de muestreo y observar la política aprendida. El procesamiento se realizó con Python, PyTorch, Stable-Baselines3, Gymnasium, NumPy y Pandas en una infraestructura CPU de 32 núcleos, 32 GB de RAM y almacenamiento SSD, lo que permitió medir la factibilidad del enfoque para entornos con recursos limitados. El análisis de la información se realizó mediante estadística descriptiva, comparación por dificultad, análisis de progresión por fases, correlación entre retorno y precisión, y evaluación de estabilidad mediante desviación estándar.

Cada episodio inició con un estudiante simulado caracterizado por nivel base, dominio temático y variables afectivas normalizadas. El agente seleccionó una acción compuesta, el entorno calculó la probabilidad de respuesta correcta y luego actualizó dominio, precisión, compromiso, frustración, confianza y rachas de desempeño. Las respuestas correctas incrementaron el dominio del tema en función de dificultad y pistas empleadas, mientras que las respuestas incorrectas redujeron el dominio y aumentaron frustración cuando se acumulaban errores consecutivos. Esta dinámica buscó aproximar el aprendizaje gradual sin atribuir al simulador propiedades psicológicas no observadas. Por ello, los resultados se interpretaron como desempeño del sistema en un entorno artificial controlado y no como evidencia directa de aprendizaje humano.

El procedimiento de validación se estructuró en tres momentos. Primero, se revisó la coherencia del banco de preguntas con los niveles de comprensión definidos para el cuento. Segundo, se entrenó el agente y se almacenaron métricas por intervalo de validación, de modo que fuera posible observar progresión, oscilaciones y estabilidad. Tercero, se compararon las métricas finales por dificultad y se examinó la relación entre retorno, precisión y entropía. Este proceso permitió valorar si la recompensa multiobjetivo conducía a políticas pedagógicas razonables. La decisión de usar simulación se justificó por razones éticas y técnicas, pues un prototipo inicial no debe aplicarse directamente en estudiantes sin verificar previamente seguridad, consistencia y

ausencia de comportamientos pedagógicos extremos. En cuanto al entrenamiento del agente basado en el algoritmo PPO (Proximal Policy Optimization) se configuró de la siguiente manera:

Pasos totales planificados: 500 000

Pasos efectivos reportados: 491 520

Longitud de rollout: 2 048

Tamaño de lote: 64

Tamaño de mini-lote: 32

Épocas de optimización: 10

Tasa de aprendizaje: 2,5 × 10^-4

Factor de descuento γ : 0,99

Parámetro GAE λ :0,95

Epsilon de recorte : 0,20

Coeficiente de valor: 0,50

Coeficiente de entropía : 0,01

Norma máxima de gradiente: 0,50

Por último, en cuanto a las consideraciones éticas del estudio, se estableció que la investigación no involucró la participación directa de seres humanos ni el uso de datos personales identificables, dado que el sistema fue evaluado exclusivamente en un entorno de simulación computacional. En consecuencia, no fue necesario gestionar consentimiento informado ni aprobación por un comité de ética en investigación con sujetos humanos, aunque se adoptaron principios internacionales de investigación responsable en inteligencia artificial aplicada a la educación (UNESCO, 2023).

Resultados

Los resultados se organizaron a partir de las métricas finales del entrenamiento, la progresión de precisión, el desempeño por dificultad y la interpretación pedagógica de las decisiones del agente. En términos generales, los hallazgos muestran que el sistema logró una política estable, aunque con precisión absoluta moderada, por consiguiente, el aporte principal se ubicó en la consistencia del comportamiento adaptativo y no en la obtención de un desempeño final suficiente para despliegue directo en aula.

Desempeño general del agente

La evaluación final evidenció una tasa de completación de 98,3 % sobre los pasos planificados, retorno promedio de 4,35 y precisión promedio de 38,23 %. La precisión máxima alcanzó 45,08 %, mientras que la desviación estándar fue 3,86 %, lo que indica estabilidad relativa de las decisiones aprendidas. Este resultado adquiere relevancia porque la política no se comportó de manera errática, aunque tampoco alcanzó el umbral

pedagógico sugerido para un sistema listo para producción. En este sentido, la ausencia de convergencia clara al cierre de 500 000 pasos sugiere que el entrenamiento requeriría más episodios, ampliación de contenidos y ajustes de recompensa para mejorar exactitud sin sacrificar compromiso.

Tabla 1. *Métricas finales del entrenamiento*

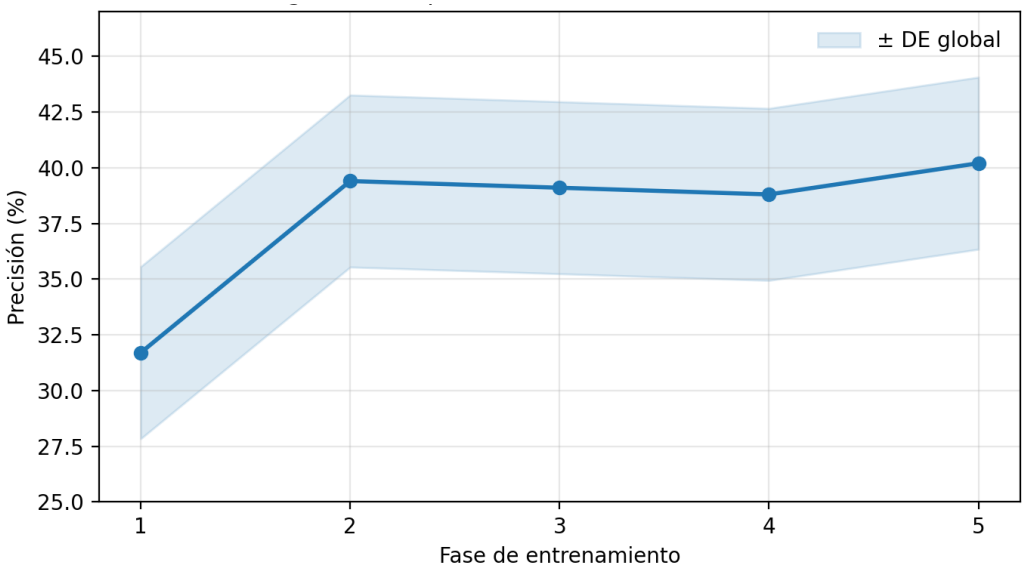
Métrica	Valor	Desviación estándar
Pasos efectivos	491 520	—
Tasa de completación	98,3 %	—
Retorno promedio	4,35	0,78
Retorno máximo	5,57	—
Precisión promedio	38,23 %	3,86 %
Precisión máxima	45,08 %	—
Eficiencia de aprendizaje	1,00 %/paso	—
Convergencia detectada	No	—

Nota. *Los valores proceden del entrenamiento reportado en el prototipo original. Fuente: elaboración propia con base en los registros experimentales.*

Progresión de precisión durante el entrenamiento

La progresión por fases mostró una mejora desde 31,7 % en la primera fase hasta 40,2 % en la quinta fase. Aunque se observaron pequeñas oscilaciones intermedias, la tendencia general fue ascendente y compatible con un proceso de aprendizaje de política. De hecho, la precisión se estabilizó alrededor del rango de 38 % a 45 %, lo que indica que el agente aprendió regularidades del entorno, aunque enfrentó límites por el tamaño reducido del banco de preguntas y por la naturaleza simulada de las respuestas. Desde una perspectiva pedagógica, esta progresión sugiere que el agente fue capaz de ajustar el nivel de desafío y la provisión de pistas, no obstante, aún requiere calibración para elevar la probabilidad de respuestas correctas.

Figura 1. *Progresión de precisión durante el entrenamiento*



Nota. *La figura representa la precisión por fases y una banda aproximada de variabilidad construida con la desviación estándar global reportada. Fuente: elaboración propia con base en los resultados del prototipo.*

Desempeño por nivel de dificultad

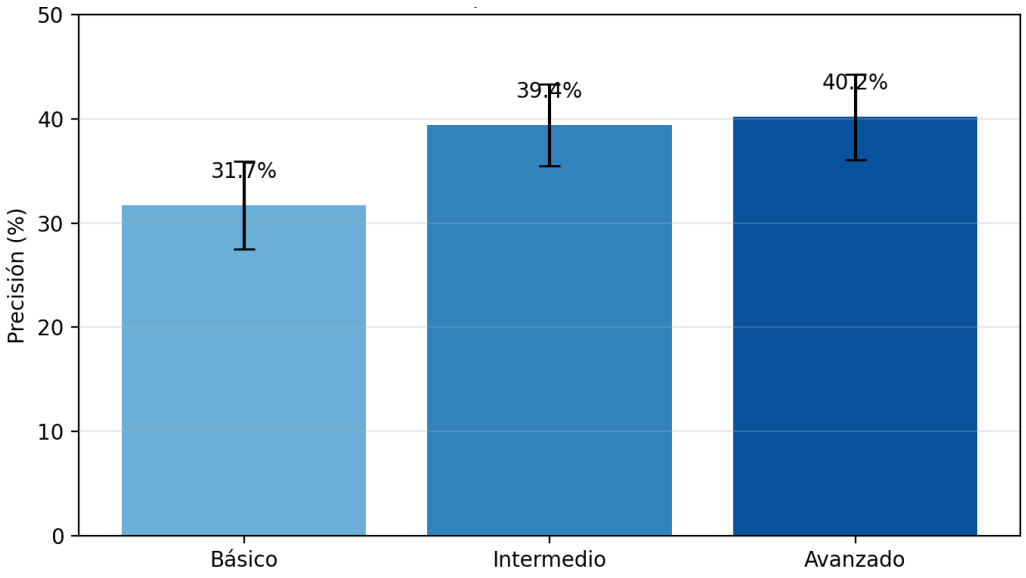
El desempeño por dificultad mostró un patrón relevante: la precisión en nivel avanzado fue 40,2 %, superior a la básica de 31,7 %, mientras que el nivel intermedio alcanzó 39,4 %. En principio, este comportamiento puede interpretarse como evidencia de que el agente aprendió a utilizar mejor las estrategias de andamiaje en preguntas complejas que en ítems básicos. Sin embargo, también puede reflejar un sesgo del banco de preguntas, pues diez ítems no son suficientes para garantizar equivalencia entre dificultad nominal y dificultad empírica. Por tanto, el hallazgo se considera prometedor pero preliminar, y se recomienda ampliar el banco a más de cien preguntas con validación por expertos y pilotaje psicométrico.

Tabla 2*. Precisión por nivel de dificultad*

Dificultad	Precisión	Desviación estándar	Tamaño de muestra
Básico	31,7 %	4,2 %	1 847
Intermedio	39,4 %	3,9 %	2 103
Avanzado	40,2 %	4,1 %	1 972

Nota. *La distribución por dificultad se adaptó del artículo original. Fuente: elaboración propia con base en registros de validación.*

Figura 2*. Precisión por nivel de dificultad*



Nota. *Las barras muestran precisión media y barras de error de desviación estándar por dificultad. Fuente: elaboración propia con base en los resultados del prototipo.*

Comportamiento de la política adaptativa

El análisis del comportamiento de la política mostró que la función de recompensa alineó

fuertemente retorno y precisión, con una correlación reportada de $\rho = 0,98$ y un coeficiente de determinación de $R^2 = 0,9547$ para la relación lineal entre retorno y precisión. Adicionalmente, la relación negativa entre entropía y retorno, reportada como $\rho = -0,83$, sugiere que la política se volvió más determinística a medida que mejoraba el desempeño. Este patrón resulta consistente con el comportamiento esperado en PPO, siempre que no exista colapso prematuro de exploración. En este caso, la entropía conservó valores aproximados de 3,5, lo que indica que el agente conservó diversidad de acciones y no se limitó a repetir una única estrategia pedagógica.

Desde una interpretación educativa, la política adaptativa integró selección de tema, dificultad, pistas y estrategia de enseñanza. Esta multidimensionalidad resulta más rica que una recomendación lineal de preguntas porque permite modificar simultáneamente qué contenido se trabaja, cuán difícil es, qué apoyo se ofrece y bajo qué metodología se presenta. No obstante, esta interpretación debe asumirse con cautela, puesto que la simulación no incorpora lenguaje natural abierto, comprensión real de explicaciones ni emociones observadas, sino variables programadas para aproximar dimensiones cognitivas, afectivas y temporales. En consecuencia, los resultados validan la factibilidad computacional del enfoque, pero no demuestran todavía impacto educativo en estudiantes reales.

En relación con el propósito experimental, los resultados adquieren mayor claridad al considerar el propósito de la fase experimental. El prototipo no buscó reemplazar una evaluación de aula, sino establecer si una arquitectura actor-crítico podía sostener decisiones multi-discretas y recompensas pedagógicas compuestas sin inestabilidad severa. Bajo esta premisa, el retorno promedio de 4,35 y la alta correlación con la precisión sugieren coherencia interna de la función de recompensa. Por el contrario, la precisión moderada debe interpretarse como una señal diagnóstica: el sistema aprendió una política, pero todavía no aprendió una política suficientemente eficaz. Esta distinción es crítica, ya que evita sobredimensionar los hallazgos y orienta mejoras específicas, como la ampliación del banco de preguntas, el ajuste de pesos de recompensa, la incorporación de aprendizaje curricular y la comparación con líneas base estáticas.

Finalmente, el análisis de la distribución por dificultad también aporta una lectura de diseño instruccional. Si el agente obtiene mejores resultados en niveles intermedios y avanzados, es posible que el sistema esté capitalizando contextos donde las pistas y estrategias generan mayor valor marginal. En contraste, en preguntas básicas, en cambio, la ganancia por andamiaje puede ser menor o incluso generar interferencia si el estudiante simulado ya posee suficiente dominio. Esta hipótesis deberá contrastarse con datos empíricos en población real, considerando que en comprensión lectora el exceso de apoyo puede disminuir el esfuerzo cognitivo, mientras que su ausencia puede incrementar la frustración. Por tanto, un tutor adaptativo eficaz, no solo debe aprender no solo a proporcionar, sino a regularla y retirarla oportunamente.

Discusión

Los resultados del estudio evidencian que el aprendizaje por refuerzo profundo puede emplearse para organizar decisiones pedagógicas simultáneas en comprensión lectora, particularmente cuando el sistema considera precisión, progreso, compromiso,

frustración y ajuste de dificultad. Este hallazgo coincide con Li et al. (2023), quienes plantearon que el aprendizaje adaptativo puede modelarse como un proceso de decisión de Markov cuando el sistema debe seleccionar materiales en función de rasgos latentes del estudiante. Sin embargo, el desempeño absoluto de 38,23 % también indica que la formulación algorítmica no garantiza, por sí sola, el aprendizaje efectivo; por el contrario, la calidad del banco de preguntas, la validez del simulador y el diseño de recompensa determinan la utilidad pedagógica de la política aprendida.

El resultado de estabilidad, expresado en una desviación estándar de 3,86 %, es consistente con la intención de PPO de limitar actualizaciones abruptas de la política y preservar seguridad de aprendizaje (Schulman et al., 2017). Desde una perspectiva aplicada, la estabilidad es valiosa porque un tutor educativo no debe modificar de forma impredecible la dificultad o los apoyos durante una sesión. No obstante, la estabilidad debe diferenciarse de eficacia. Un agente puede ser estable y, aun así, mantener una precisión insuficiente; por tanto, el hallazgo debe interpretarse como validación de factibilidad computacional y no como evidencia concluyente de mejora lectora.

En comparación con la literatura previa, los sistemas de tutoría inteligente revisados por Kulik y Fletcher (2016), el prototipo presenta una ventaja conceptual al integrar indicadores socioemocionales en la recompensa, pero todavía se encuentra lejos de la madurez empírica de evaluaciones controladas con estudiantes. Steenbergen-Hu y Cooper (2013) mostraron que los efectos de los tutores inteligentes en K-12 dependen de diseños rigurosos, muestras reales y medidas de aprendizaje. En el presente caso, la simulación permitió explorar la política sin exponer estudiantes a decisiones experimentales no validadas; sin embargo, la fase siguiente debe incorporar pilotos con medición pretest-postest, grupo de comparación y control de variables docentes.

La comparación con estudios de comprensión lectora muestra una coincidencia relevante: la tutoría computacional funciona mejor cuando enseña estrategias explícitas y ofrece retroalimentación estructurada. Wijekumar et al. (2012) y Wijekumar et al. (2014) demostraron efectos de la tutoría inteligente de estructura textual en comprensión de textos informativos, mientras que McNamara et al. (2004) documentaron el potencial de iSTART para enseñar estrategias de lectura activa. En contraste, el prototipo basado en *Paco Yunque* avanza, no solo implementa estrategias de comprensión, sino que decide simultáneamente dificultad, pista, tema y metodología. Esta ampliación del espacio de decisión, aunque prometedora, también introduce más fuentes de error: si el agente selecciona una estrategia inadecuada, el estudiante puede recibir ayuda insuficiente, excesiva o poco relevante.

El resultado por dificultad merece una interpretación especifica. Que el nivel avanzado presente mayor precisión que el básico no implica necesariamente una mejor comprensión en tareas complejas. Alternativamente, este patrón puede derivarse de las reglas del simulador, la estructura del banco de ítems o la interacción entre dificultad y pistas. Desde la teoría del flujo de Csikszentmihalyi (1990), el aprendizaje se favorece cuando el desafío se equilibra con la habilidad; sin embargo, esa explicación solo sería válida si la dificultad fue calibrada empíricamente. Por ello, se requiere una validación del banco de preguntas mediante juicio de expertos, análisis de dificultad, discriminación e invariancia, antes de atribuir valor pedagógico pleno a la progresión observada.

En cuanto a las implicaciones prácticas, los resultados sugieren la viabilidad de construir sistemas adaptativos culturalmente situados. Frente a plataformas globales que estandarizan contenidos, el uso de literatura peruana permite articular comprensión lectora con identidad, desigualdad social y reflexión crítica. De este modo, se responde a la necesidad de contextualización señalada por Salas-Pilco y Yang (2022) en América Latina. Asimismo, esta perspectiva se alinea con la UNESCO (2023), que enfatiza la evaluación de la tecnología educativa en función de su pertinencia pedagógica y no únicamente de su novedad. En consecuencia, el valor del prototipo reside en demostrar la aplicabilidad de arquitecturas avanzadas a contenidos locales, aunque su impacto dependerá de factores como accesibilidad, formación docente y gobernanza de datos.

Desde una perspectiva teórica, el estudio integra tres tradiciones: tutoría inteligente, aprendizaje por refuerzo y pedagogía sociocultural. Anderson et al. (1995) mostraron que los tutores cognitivos requieren modelos del conocimiento; Corbett y Anderson (1994) explicaron cómo estimar adquisición procedimental; y Sutton y Barto (2018) fundamentaron la optimización de decisiones secuenciales mediante recompensa acumulada. Sobre esta base, el presente enfoque combina estas bases con variables afectivas inspiradas en la necesidad de sostener compromiso y reducir frustración. Desde Vygotsky (1978), el andamiaje debe adaptarse a la zona de desarrollo próximo; desde Bloom (1984), la tutoría individual produce beneficios difíciles de escalar. El sistema propuesto intenta aproximar computacionalmente esa mediación, aunque todavía no sustituye el juicio docente ni la interacción humana.

En relación con las limitaciones se identifican varios aspectos críticos. En primer lugar, el banco de preguntas fue reducido y no permite generalizar resultados a comprensión lectora amplia. En segundo lugar, los estudiantes fueron simulados, por lo que las emociones y respuestas no proceden de observación humana. En tercer lugar, la precisión alcanzada fue moderada y requiere optimización adicional antes de uso educativo real. En cuarto lugar, el entorno no incorpora comprensión abierta, producción escrita ni diálogo natural, dimensiones centrales de la lectura crítica. Adicionalmente, la ausencia de una línea base empírica en el mismo entorno restringe la interpretación comparativa de los resultados.

En efecto, aunque el prototipo se contrastó conceptualmente con enfoques como Q-learning, POMDP, bandits y rastreo de conocimiento, sería metodológicamente más robusto incluir comparaciones directas con políticas heurísticas, secuencias fijas o algoritmos alternativos como DQN o SAC. Del mismo modo, resulta necesario aplicar análisis de sensibilidad sobre pesos de recompensa, tasa de aprendizaje, tamaño de red y coeficiente de entropía, pues pequeñas variaciones pueden cambiar la política pedagógica aprendida. Esta precaución es relevante en educación porque una política que optimiza retorno puede priorizar métricas internas sin maximizar comprensión profunda.

Desde el punto de vista ético, el prototipo plantea una oportunidad y una advertencia. La oportunidad consiste en diseñar herramientas abiertas, de bajo costo y culturalmente pertinentes para apoyar comprensión lectora en contextos con escasa personalización. La advertencia radica en que la automatización pedagógica puede introducir sesgos si los textos, preguntas, niveles de dificultad o reglas de recompensa reproducen supuestos no revisados. Por ello, cualquier implementación futura debería incluir revisión docente,

participación de especialistas en literatura peruana, consentimiento informado, resguardo de datos y explicaciones comprensibles para estudiantes y familias.

En este contexto, la adaptatividad no debe entenderse como delegación total al algoritmo, sino como apoyo para decisiones pedagógicas supervisadas. Esta condición cobra mayor importancia cuando el sistema se orienta a niños o adolescentes, porque la recomendación automática puede modificar oportunidades de aprendizaje, autoestima académica y relación con el texto. Por ello, las métricas de precisión deben complementarse con indicadores de comprensión profunda, participación, satisfacción, equidad de acceso y apropiación docente. La evaluación futura también debería registrar errores del sistema, decisiones inesperadas y casos donde la intervención humana corrige la ruta sugerida por el agente. De este modo, el tutor puede evolucionar como herramienta de apoyo pedagógico y no como mecanismo opaco de clasificación estudiantil.

Finalmente, en términos de implementación, el bajo costo computacional constituye una ventaja para instituciones educativas con infraestructura limitada. El entrenamiento en CPU y el tamaño reducido del modelo sugieren que una versión optimizada podría operar en laboratorios escolares modestos o servidores institucionales sin dependencia permanente de servicios externos. No obstante, esa ventaja técnica debe acompañarse de gobernanza de datos, trazabilidad de decisiones y participación docente. Un sistema que adapta preguntas y pistas puede influir en trayectorias de aprendizaje, por lo que su uso exige transparencia sobre criterios de recomendación, posibilidad de supervisión humana y mecanismos para corregir sesgos del contenido o del modelo. La equidad no se garantiza por liberar una herramienta; se construye mediante accesibilidad, formación, evaluación continua y adecuación al contexto escolar.

En síntesis, la discusión permite afirmar que el sistema constituye una prueba de concepto técnicamente viable, culturalmente pertinente y metodológicamente replicable. Su principal aporte no es haber alcanzado una precisión alta, sino demostrar que una política PPO puede tomar decisiones pedagógicas multi-dimensionales con estabilidad y bajo costo computacional. La investigación futura debe ampliar el corpus, incorporar estudiantes reales, comparar contra líneas base no adaptativas y evaluar si las decisiones del agente producen mejoras significativas en comprensión literal, inferencial y crítica. Solo con esa evidencia será posible pasar de la promesa algorítmica a una innovación educativa validada, sostenible y pedagógicamente responsable.

Conclusión

El estudio permitió analizar un sistema de tutoría adaptativa basado en aprendizaje por refuerzo profundo para personalizar la comprensión lectora mediante literatura peruana. Los hallazgos evidenciaron que el agente PPO logró decisiones pedagógicas estables en un entorno simulado, con precisión promedio moderada, retorno consistente y capacidad para combinar selección de tema, ajuste de dificultad, provisión de pistas y estrategia de enseñanza. Esta integración muestra que la personalización lectora puede formularse como un problema de decisión secuencial, especialmente cuando la función de recompensa no se limita a la respuesta correcta, sino que incorpora progreso, compromiso, eficiencia y ajuste de desafío.

La principal contribución del trabajo consiste en articular un modelo computacional avanzado con contenido literario culturalmente situado. El uso de *Paco Yunque* permitió vincular comprensión lectora con reflexión social, lo cual abre una ruta para desarrollar tutores adaptativos que no dependan exclusivamente de contenidos genéricos. Sin embargo, el sistema aún se encuentra en fase de prototipo: requiere mayor banco de preguntas, calibración por expertos, entrenamiento extendido, comparación con líneas base, análisis de sesgos y validación con estudiantes reales. De ese modo, los resultados deben comprenderse como una base experimental para construir evidencia posterior, no como recomendación inmediata de implementación escolar.

Como líneas de investigación futura, se propone ampliar el corpus de literatura peruana, incorporar respuestas abiertas, integrar analítica de aprendizaje explicable y realizar estudios cuasiexperimentales o experimentales en instituciones educativas. También se recomienda estudiar el papel docente en la supervisión del tutor, así como la percepción estudiantil sobre ayuda, dificultad y motivación.

En definitiva, el estudio enfatiza que la inteligencia artificial educativa debe evaluarse por su capacidad de fortalecer aprendizajes significativos, reducir brechas y apoyar la mediación humana que sostiene la comprensión crítica. Su valor dependerá de convertir la innovación algorítmica en una práctica pedagógica ética, transparente y culturalmente pertinente.

Acerca de

Contribución de los autores: Los autores contribuyeron a la conceptualización del estudio, desarrollo metodológico, análisis e interpretación de los datos, redacción del manuscrito y revisión crítica de su contenido intelectual.

Financiamiento: Los autores declaran que no recibieron financiamiento para esta investigación.

Certificación ética: El protocolo del presente estudio fue sometido a revisión y aprobado por el Comité de Ética en Investigación de la Universidad, en cumplimiento de los principios éticos y normativas institucionales aplicables.

Referencias

Abraham, L. B. (2008). Computer-mediated glosses in second language reading comprehension and vocabulary learning: A meta-analysis. *Computer Assisted Language Learning, 21*(3), 199–226. https://doi.org/10.1080/09588220802090246

Anderson, J. R., Corbett, A. T., Koedinger, K. R., & Pelletier, R. (1995). Cognitive tutors: Lessons learned. *Journal of the Learning Sciences, 4*(2), 167–207. https://doi.org/10.1207/s15327809jls0402_2

Bloom, B. S. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher, 13*(6), 4–16. https://doi.org/10.3102/0013189X013006004

Chen, J. J., y Perez, C. (2023). Enhancing assessment and personalized learning through artificial intelligence. *Childhood Education, 99*(6), 72–79. https://doi.org/10.1080/00094056.2023.2282903

Corbett, A. T., & Anderson, J. R. (1994). Knowledge tracing: Modeling the acquisition of procedural knowledge. *User Modeling and User-Adapted Interaction, 4*(4), 253–278. https://doi.org/10.1007/BF01099821

Csikszentmihalyi, M. (1990). *Flow: The psychology of optimal experience*. Harper & Row. https://archive.org/details/flowpsychologyof00csik

François-Lavet, V., Henderson, P., Islam, R., Bellemare, M. G., & Pineau, J. (2018). An introduction to deep reinforcement learning. *Foundations and Trends in Machine Learning, 11*(3–4), 219–354. https://doi.org/10.1561/2200000071

Iglesias, A., Martínez, P., Aler, R., y Fernández, F. (2009). Learning teaching strategies in an adaptive and intelligent educational system through reinforcement learning. *Applied Intelligence, 31*(1), 89–106. https://doi.org/10.1007/s10489-008-0115-1

Kulik, J. A., & Fletcher, J. D. (2016). Effectiveness of intelligent tutoring systems: A meta-analytic review. *Review of Educational Research, 86*(1), 42–78. https://doi.org/10.3102/0034654315581420

Lan, A. S., Studer, C., & Baraniuk, R. G. (2014). Time-varying learning and content analytics via sparse factor analysis. En *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 452–461). Association for Computing Machinery. https://doi.org/10.1145/2623330.2623631

Li, X., Xu, H., Zhang, J., & Chang, H. H. (2021). Optimal hierarchical learning path design with reinforcement learning. *Applied Psychological Measurement, 45*(1), 54–70. https://doi.org/10.1177/0146621620947171

Li, X., Xu, H., Zhang, J., & Chang, H. H. (2023). Deep reinforcement learning for adaptive learning systems. *Journal of Educational and Behavioral Statistics, 48*(2), 220–243. https://doi.org/10.3102/10769986221129847

McNamara, D. S., Levinstein, I. B., & Boonthum, C. (2004). iSTART: Interactive strategy training for active reading and thinking. *Behavior Research Methods, Instruments, & Computers, 36*(2), 222–233. https://doi.org/10.3758/BF03195567

OECD. (2023). *PISA 2022 results: Peru country note*. Organisation for Economic Co-operation and Development. https://www.oecd.org/en/publications/pisa-2022-results-volume-i-and-ii-country-notes_ed6fbcc5-en/peru_3e71791c-en.html

Piech, C., Bassen, J., Huang, J., Ganguli, S., Sahami, M., Guibas, L. J., & Sohl-Dickstein, J. (2015). Deep knowledge tracing. En C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, & R. Garnett (Eds.), *Advances in neural information processing systems 28* (pp. 505–513). Curran Associates. https://papers.nips.cc/paper_files/paper/2015/hash/bac9162b47c56fc8a4d2a519803d51b3-Abstract.html

Saavedra, J., & Molina, E. (2026). *La principal limitación de la IA en educación no es la tecnología. Es la cultura organizacional*.

https://blogs.worldbank.org/es/latinamerica/binding-constraint-on-ai-in-education-latin-america

Salas-Pilco, S. Z., & Yang, Y. (2022). Artificial intelligence applications in Latin American higher education: A systematic review. *International Journal of Educational Technology in Higher Education, 19,* Article 21. https://doi.org/10.1186/s41239-022-00326-w

Schulman, J., Wolski, F., Dhariwal, P., Radford, A., & Klimov, O. (2017). Proximal policy optimization algorithms. *arXiv*. https://doi.org/10.48550/arXiv.1707.06347

Steenbergen-Hu, S., & Cooper, H. (2013). A meta-analysis of the effectiveness of intelligent tutoring systems on K–12 students' mathematical learning. *Journal of Educational Psychology, 105*(4), 970–987. https://doi.org/10.1037/a0032447

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction* (2.ª ed.). MIT Press. http://incompleteideas.net/book/the-book-2nd.html

UNESCO. (2022). *K-12 AI curricula: A mapping of government-endorsed AI curricula*. United Nations Educational, Scientific and Cultural Organization. https://unesdoc.unesco.org/ark:/48223/pf0000380602

UNESCO. (2023). *Global education monitoring report 2023: Technology in education: A tool on whose terms?* United Nations Educational, Scientific and Cultural Organization. https://doi.org/10.54676/UZQV8501

VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. Educational Psychologist, 46(4), 197–221. https://doi.org/10.1080/00461520.2011.611369

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press. https://doi.org/10.2307/j.ctvjf9vz4

Wang, X., Huang, R. T., Sommer, M., Pei, B., Shidfar, P., Rehman, M. S., Ritzhaupt, A. D., & Martin, F. (2024). The efficacy of artificial intelligence-enabled adaptive learning systems from 2010 to 2022 on learner outcomes: A meta-analysis. *Journal of Educational Computing Research*, *62*(6), 1348–1383. https://doi.org/10.1177/07356331241240459

Wijekumar, K. K., Meyer, B. J. F., & Lei, P. (2012). Large-scale randomized controlled trial with 4th graders using intelligent tutoring of the structure strategy to improve nonfiction reading comprehension. *Educational Technology Research and Development, 60*(6), 987–1013. https://doi.org/10.1007/s11423-012-9263-4

Wijekumar, K. K., Meyer, B. J. F., Lei, P., Lin, Y. C., Johnson, L. A., Spielvogel, J. A., Shurmatz, K. M., & Ray, M. (2014). Multisite randomized controlled trial examining intelligent tutoring of structure strategy for fifth-grade readers. *Journal of Research on Educational Effectiveness*, *7*(4), 331–357. https://doi.org/10.1080/19345747.2013.853333

World Bank. (2019). *Learning poverty brief: Peru*. World Bank. https://thedocs.worldbank.org/en/doc/669911571223458621-0090022019/original/LACLCC6CPERLPBRIEF.pdf

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, Article 39. https://doi.org/10.1186/s41239-019-0171-0