El examen como espacio de aprendizaje: rediseño de la evaluación en la formación docente mediada por IA

The exam as a learning space: redesigning assessment in AI-mediated teacher training

Humberto Morales
Universidad de Flores, Argentina
humbertomorales@hotmail.com

Resumen

La incorporación de la inteligencia artificial (IA) en educación plantea interrogantes profundos sobre las prácticas de enseñanza y, especialmente, sobre los modos de evaluar. En la formación docente, estas preguntas adquieren un valor estratégico, ya que no solo se evalúa a estudiantes, sino que se modelan prácticas profesionales futuras. Este artículo analiza una experiencia desarrollada en un profesorado de inglés, en la asignatura Enseñar con Tecnología, donde un examen tradicional fue resignificado mediante el uso pedagógico de herramientas de IA. A partir de la comparación entre producciones estudiantiles y respuestas generadas por IA, el uso de esta herramienta como asistente de evaluación y mediadora de explicaciones alternativas se propone una concepción de la evaluación como instancia de aprendizaje significativo.

Palabras clave: evaluación educativa; inteligencia artificial; formación docente; aprendizaje significativo; tecnología educativa

Abstract

The integration of Artificial Intelligence (AI) into education raises critical questions about teaching and assessment practices. In teacher education, these issues are particularly relevant, as assessment not only measures learning but also shapes future professional practices. This paper analyzes an experience carried out in an English teacher training program where a traditional exam was redesigned through the pedagogical use of AI tools. The findings suggest that AI does not replace the teacher's role, but rather enhances it when integrated with clear pedagogical intentions.

Keywords: educational assessment; artificial intelligence; teacher education; meaningful learning; educational technology

Sección: Dossier “Desafíos de la Educación en la era de la Inteligencia Artificial”
Recibido: 09/02/2026
Aceptado: 20/05/2026

DOI: 10.63790/yvkxtd84

El Faro se encuentra bajo la licencia de Creative Commons Atribución-NoComercial-SinDerivadas 4.0 Internacional (CC BY-NC-ND 4.0)

1. Introducción

La evaluación constituye uno de los núcleos más sensibles del proceso educativo. A pesar de los avances en pedagogía y didáctica, continúa predominando una concepción de la evaluación asociada al control, la medición y la acreditación, más que al aprendizaje. La irrupción de la inteligencia artificial generativa tensiona estas prácticas no solo porque facilita el acceso a información y la producción de textos, sino porque interpela directamente el sentido de evaluar.

En la formación docente, esta discusión adquiere un valor estratégico. Los futuros profesores no solo aprenden contenidos disciplinares, sino que internalizan modelos de enseñanza y evaluación que luego reproducirán en sus propias prácticas. Por ello, revisar críticamente los dispositivos evaluativos implica también asumir una responsabilidad ética y profesional frente a los desafíos educativos contemporáneos.

2. Marco teórico: evaluación, tecnología y conocimiento

Desde una perspectiva pedagógica, la evaluación es parte constitutiva del proceso de enseñanza y aprendizaje, y no una instancia aislada al final del recorrido (Camilloni, 2007). Evaluar implica interpretar evidencias, emitir juicios fundamentados y tomar decisiones orientadas a mejorar los aprendizajes. Sin embargo, cuando la evaluación se reduce a la reproducción de información, pierde su potencial formativo.

Paulo Freire (1970) advierte sobre los riesgos de una educación bancaria, centrada en la acumulación de datos y la pasividad del estudiante. Esta crítica resulta especialmente pertinente en un contexto donde la IA puede acceder, organizar y producir información con gran eficacia. En este escenario, la pregunta no es si los estudiantes pueden usar IA, sino qué tipo de pensamiento promovemos al evaluar.

Michel Serres (2013) plantea que la tecnología ha externalizado funciones cognitivas tradicionales, liberando al sujeto de ciertas cargas, pero obligándolo a desarrollar nuevas formas de pensamiento. Por su parte, Noriko Arai (2017) demuestra que la IA puede resolver exámenes complejos sin comprender, lo que obliga a replantear qué entendemos por saber y qué valoramos al evaluar.

Desde la tecnología educativa, Casablancas (2014) enfatiza que el problema no radica en la herramienta, sino en el uso pedagógico que se haga de ella. Integrar tecnología sin reflexión didáctica reproduce prácticas tradicionales con nuevos soportes, sin generar transformaciones significativas.

3. Contexto y diseño de la experiencia

La experiencia se desarrolló en la asignatura Enseñar con Tecnología, correspondiente al segundo año del Profesorado de Inglés. Participaron 22 estudiantes distribuidos en dos turnos: 16 en el turno mañana y 6 en el turno noche. La intervención tuvo lugar en una única clase, inmediatamente posterior a la corrección de un examen parcial domiciliario tradicional.

El examen se basaba en el análisis de tres materiales seleccionados intencionalmente: una conferencia de Michel Serres sobre tecnología y pensamiento, una charla TED de Noriko Arai sobre los límites de la IA en los exámenes académicos y una entrevista a Silvina Casablancas sobre la integración pedagógica de las tecnologías. Las consignas no apuntaban a la reproducción de contenidos, sino a la inferencia, la relación entre ideas y la construcción de sentido. Una condición formal relevante era que cada respuesta debía limitarse a cinco renglones como máximo.

4. La IA como espejo pedagógico

Una vez corregidos los exámenes, se replicó la consigna completa utilizando ChatGPT, proporcionándole exactamente los mismos materiales y preguntas que habían recibido los estudiantes. En una primera instancia, la IA produjo respuestas extensas y bien estructuradas, pero que excedían el límite de cinco renglones establecido. Al reformular la consigna con esa restricción explícita, la IA logró sintetizar las ideas principales de manera precisa, evidenciando que la claridad de la consigna también opera como variable pedagógica.

Al comparar estas respuestas con las producciones estudiantiles, emergió un hallazgo central: las respuestas de la IA eran similares, en calidad argumentativa, a las mejores producciones humanas. Esto generó sorpresa, pero sobre todo reflexión colectiva. Los comentarios relevados entre los estudiantes permiten ilustrar el impacto de la experiencia: “No sabía que la IA podía analizar y corregir exámenes”, “Me sorprenden las devoluciones que hace”, “Es impresionante el diálogo que construyó entre dos 'personas' hablando sobre los temas del examen, relacionándolos y tomando puntos de vista de cada uno de los autores”.

Estos comentarios revelan un desconocimiento inicial de las capacidades de las herramientas de IA, pero también una apertura reflexiva ante su potencial pedagógico. La

comparación habilitó debates orientados a construir una comprensión más profunda de los contenidos y de los criterios de evaluación.

5. La IA en rol docente: evaluación, rúbrica y metacognición

En una segunda instancia, se propuso a la IA asumir el rol de docente evaluador. Algunos estudiantes se ofrecieron voluntariamente para que sus exámenes fueran evaluados por la IA, que debía asignar una calificación y brindar una breve devolución. Las calificaciones otorgadas resultaron muy similares a las del docente, pero el valor principal residió en el proceso: los estudiantes pudieron contrastar su percepción del propio desempeño con una devolución externa, favoreciendo instancias de autoevaluación.

A partir de esta evaluación, se solicitó a la IA explicaciones alternativas, reformulaciones y ejemplos adaptados. Esta posibilidad evidenció el potencial de la IA como mediadora del aprendizaje, especialmente en contextos de diversidad cognitiva. Asimismo, se utilizó NotebookLM para generar un diálogo tipo podcast en el que dos voces discutían los materiales del examen desde distintos enfoques, lo que contribuyó a la comprensión de los contenidos por parte de los estudiantes.

Un elemento central de esta fase fue el diseño colectivo de una rúbrica de evaluación, elaborada en clase con la asistencia de ChatGPT. Los criterios acordados, con sus descriptores de nivel excelente e insuficiente, se presentan en la Tabla 1.

Tabla 1

Rúbrica de evaluación construida colectivamente con asistencia de IA

Criterio	Descripción	Nivel Excelente	Nivel Insuficiente
Articulación Teórica	Vinculación orgánica de los tres autores	Logra vincular a Serres, Freire y Arai encontrando puntos de contacto y tensión.	Menciona a los autores de forma aislada, como lista de definiciones sin conexión.
Voz Propia y Posicionamiento	Postura crítica situada en el rol docente	El texto refleja una postura crítica y situada	El texto es impersonal o carece de una opinión

Criterio	Descripción	Nivel Excelente	Nivel Insuficiente
		en el rol del futuro docente de inglés.	pedagógica real (sesgo típico de IA).
Detección de Alucinaciones	Identificación de errores o conceptos inventados por la IA	El evaluador identifica si la respuesta inventa conceptos o usa términos fuera de contexto.	No se distinguen errores de contenido o se aceptan como válidos datos incorrectos.
Calidad de la Retroalimentación	Comentarios que profundizan la respuesta	El alumno propone comentarios que ayudan a profundizar, sugiriendo nuevas preguntas.	La corrección se limita a decir “está bien” o “está mal” sin justificar.

Nota. Elaboración con asistencia de ChatGPT.

La construcción colectiva de la rúbrica resultó especialmente significativa: al participar en la definición de los criterios, los estudiantes incorporaron una perspectiva evaluativa que trasciende la calificación y los posiciona como futuros diseñadores de propuestas pedagógicas.

6. Discusión

La experiencia permitió reformular una idea central: no se trata de evaluar contra la IA ni de ignorarla, sino de evaluar con ella. Este hallazgo dialoga directamente con el planteo de Arai (2017), quien advierte que si la IA puede resolver exámenes con niveles de desempeño equivalentes a los mejores estudiantes, entonces esos exámenes están midiendo capacidades que ya son automatizables. La restricción de cinco renglones resultó ser una consigna que la IA también pudo resolver con precisión, lo que invita a seguir indagando en qué tipo de pensamiento es genuinamente irreemplazable.

En este sentido, Camilloni (2007) señala que la evaluación debe orientarse a producir evidencias del pensamiento del estudiante, no solo de su capacidad de reproducir información. La rúbrica construida colectivamente operó en esa dirección: al incorporar criterios como “voz propia y posicionamiento” o “detección de alucinaciones”, se valoraron competencias que la

IA no puede acreditar por sí misma, ya que requieren un sujeto situado, con historia y con perspectiva pedagógica.

La advertencia de Freire (1970) sobre la educación bancaria cobra aquí nueva vigencia: si antes el riesgo era que los estudiantes acumularan datos sin comprenderlos, hoy el riesgo es que deleguen en la IA esa acumulación sin desarrollar criterio propio. Lejos de diluir el rol docente, la integración pedagógica de la IA lo fortalece: el docente se posiciona como diseñador de experiencias, mediador de sentidos y garante del enfoque pedagógico. La evaluación deja de ser un punto final para convertirse en un proceso continuo de aprendizaje, tal como propone Casablancas (2014) al distinguir entre el uso instrumental de la tecnología y su integración didáctica genuina.

7. Conclusiones e interrogantes para la práctica

La experiencia presentada demuestra que es posible resignificar un examen tradicional integrando la inteligencia artificial sin perder profundidad ni rigor académico. Lejos de promover el facilismo, la IA puede convertirse en una aliada para fomentar la reflexión, la autoevaluación y la construcción colectiva del conocimiento.

En la formación docente, estas prácticas adquieren un valor estratégico ya que habilitan a los futuros profesores a repensar críticamente la evaluación y a diseñar propuestas pedagógicas acordes a los desafíos contemporáneos. Evaluar sigue siendo, más que nunca, una oportunidad para aprender.

La experiencia también deja abiertos interrogantes que merecen seguir siendo explorados: ¿cómo garantizar que la IA amplifique el pensamiento crítico sin reemplazarlo? ¿De qué manera pueden los docentes de distintas disciplinas adaptar este tipo de propuesta a sus propios contextos evaluativos? ¿Qué nuevas competencias requiere el docente para ser un mediador pedagógico efectivo en entornos con IA?

Como proyección inmediata, se prevé replicar y profundizar esta experiencia con los grupos del año en curso. La propuesta incorpora una variante metodológica relevante: el examen se tomará en formato presencial y escrito en papel. Los textos manuscritos serán digitalizados mediante herramientas de reconocimiento óptico de caracteres (OCR), lo que permitirá luego alimentar el proceso de evaluación mediada por IA. Esta modalidad busca que

los estudiantes conozcan y experimenten las posibilidades que ofrece la digitalización de documentos, en tanto futuros docentes que deberán integrar estas herramientas en sus propias prácticas. Asimismo, se incorporará el diseño colectivo de la rúbrica como parte del proceso evaluativo, profundizando la experiencia de aprendizaje colaborativo con asistencia de IA.

Esta propuesta no pretende ser un modelo cerrado, sino una invitación a repensar la evaluación como un espacio vivo, situado y en permanente diálogo con las transformaciones tecnológicas y culturales de nuestro tiempo.

8. Referencias

Arai, N. (2017, abril). *Can a robot pass a university entrance exam?* [Video]. TED Conferences. https://www.ted.com/talks/noriko_arai_can_a_robot_pass_a_university_entrance_exam

Camilloni, A. (2007). *La evaluación de los aprendizajes en el debate didáctico contemporáneo*. Paidós.

Casablancas, S. (2014). *Enseñar y aprender con tecnologías: un modelo teórico para las buenas prácticas con TIC*. Noveduc.

Freire, P. (1970). *Pedagogía del oprimido*. Siglo XXI.

Serres, M. (2013). *Pulgarcita*. Fondo de Cultura Económica.

Anexo: Consigna del examen domiciliario original

Primer Parcial, 2° año, Profesorado de Inglés, materia "Enseñar con Tecnología"

Para responder las siguientes preguntas es necesario haber visto los siguientes materiales:

- *Michel Serres, desde el minuto 30 en adelante: https://youtu.be/8qh44YFczto (activar subtítulos en español)*

- *Noriko Arai: ¿Puede un robot aprobar un examen de ingreso a la universidad? https://www.ted.com/talks/noriko_arai_can_a_robot_pass_a_university_entrance_exam?language=es*
- *Entrevista a Silvina Casablancas: https://www.studocu.com/es-ar/document/universidad-siglo-21/nuevas-tecnologias-en-entornos-de-aprendizaje/de-las-tic-a-las-tac-apuntes-de-apoyo/35457426*

Consignas (cada respuesta debe tener hasta 5 renglones, de lo contrario será considerada como no respondida):

1. *Serres habla del fuerte impacto de la aparición de la escritura y la imprenta. ¿Cuál es el cambio que según él esto produce y qué inferencia podés hacer respecto a la aparición de la inteligencia artificial?*
2. *"Prefiero una cabeza bien hecha que una cabeza bien llena". Esta frase que cita Serres puede ser tomada como un punto en común con el video de IA de TED. ¿Podés explicar por qué?*
3. *Si la tecnología nos obliga a ser inteligentes, como dice Serres, y la IA aún no puede comprender, como lo explica Noriko Arai, ¿cuál es la brecha que un docente debe cubrir cuando está enseñando? ¿Encontrás alguna referencia en la entrevista a Silvina Casablancas?*