



De la evaluación tradicional a la evaluación auténtica: estrategias para valorar el aprendizaje en contextos educativos mediados por inteligencia artificial

From Traditional Assessment to Authentic Assessment: Strategies for Evaluating Learning in Educational Contexts Mediated by Artificial Intelligence

Arminda Juliana Angulo Bolaños

juliangulo87@gmail.com

<https://orcid.org/0000-0002-1262-0990>

Investigadora Independiente

Resumen

La expansión de la inteligencia artificial en los entornos educativos está transformando no solo las formas de enseñar y aprender, sino también los mecanismos utilizados para valorar los aprendizajes. En este contexto, los modelos tradicionales de evaluación, centrados principalmente en pruebas estandarizadas, reproducción de contenidos y calificaciones cuantitativas, presentan limitaciones para evidenciar competencias complejas, procesos de razonamiento y desempeños auténticos del estudiantado. El presente artículo tiene como objetivo analizar la transición de la evaluación tradicional hacia enfoques de evaluación auténtica en contextos educativos mediados por inteligencia artificial, identificando estrategias que permitan valorar de manera integral, contextualizada y significativa el aprendizaje. Metodológicamente, se plantea una revisión teórico-documental de literatura académica reciente relacionada con evaluación educativa, inteligencia artificial, retroalimentación automatizada, competencias y evaluación auténtica. El análisis permite reconocer estrategias como el desarrollo de proyectos, resolución de problemas contextualizados, portafolios digitales, evaluación basada en desempeño, autoevaluación, coevaluación y uso crítico de herramientas de inteligencia artificial para la retroalimentación formativa. Asimismo, se advierte que la incorporación de estas tecnologías exige fortalecer la transparencia, la autoría académica, el juicio docente y la alfabetización en inteligencia artificial. Se concluye que la evaluación auténtica, apoyada de manera ética y pedagógicamente planificada por sistemas de inteligencia artificial, puede favorecer una valoración más profunda del aprendizaje, siempre que la tecnología complemente y no sustituya el criterio profesional docente.

Palabras clave: evaluación auténtica, evaluación educativa, inteligencia artificial, aprendizaje, retroalimentación formativa, competencias.

Abstract

The expansion of artificial intelligence in educational environments is transforming not only the ways in which teaching and learning take place, but also the mechanisms used to assess student learning. In this context, traditional assessment models, mainly focused on standardized tests, content reproduction, and quantitative grading, present limitations in demonstrating complex competencies, reasoning processes, and authentic student performance. This article aims to analyze the transition from traditional assessment toward authentic assessment approaches in educational contexts mediated by artificial intelligence, identifying strategies that enable learning to be assessed in a comprehensive, contextualized, and meaningful manner. Methodologically, the study adopts a theoretical-documentary review of recent academic literature related to educational assessment, artificial intelligence, automated feedback, competencies, and authentic assessment. The analysis identifies strategies such as project-based tasks, contextualized problem-solving, digital portfolios, performance-based assessment, self-assessment, peer assessment, and the critical use of artificial intelligence tools for formative feedback. Furthermore, the integration of these technologies requires strengthening transparency, academic authorship, teacher judgment, and artificial intelligence literacy. It is concluded that authentic assessment, when ethically and pedagogically supported by artificial intelligence systems, can foster a deeper and more meaningful evaluation of learning, provided that technology complements rather than replaces professional teacher judgment.

Keywords: authentic assessment, educational assessment, artificial intelligence, learning, formative feedback, competencies.

Introducción

La evaluación constituye uno de los componentes centrales del proceso educativo, debido a que permite obtener evidencias sobre los aprendizajes alcanzados, identificar dificultades, retroalimentar el proceso formativo y orientar decisiones pedagógicas. Sin embargo, durante décadas han predominado prácticas evaluativas sustentadas en exámenes escritos, cuestionarios, pruebas estandarizadas y actividades orientadas fundamentalmente a la reproducción de información. Aunque estos mecanismos pueden resultar pertinentes para comprobar determinados conocimientos, presentan limitaciones cuando se pretende valorar capacidades complejas como el pensamiento crítico, la resolución de problemas, la creatividad, la argumentación, la toma de decisiones o la transferencia del conocimiento a situaciones reales. Esta problemática adquiere especial relevancia en un escenario educativo caracterizado por la creciente incorporación de tecnologías digitales y sistemas de inteligencia artificial capaces de intervenir

directamente en la producción de contenidos académicos.

La expansión de la inteligencia artificial generativa ha intensificado este debate. Herramientas capaces de producir textos, resolver problemas, generar código, sintetizar información o elaborar respuestas académicamente estructuradas cuestionan la capacidad de determinadas actividades tradicionales para evidenciar de manera inequívoca lo que un estudiante sabe, comprende y es capaz de hacer. La revisión realizada por Xia et al. (2024) advierte que el desarrollo de la inteligencia artificial generativa exige transformar los sistemas de evaluación y fortalecer competencias como el aprendizaje autorregulado, la responsabilidad y la integridad académica. Los autores sostienen, además, que las instituciones necesitan revisar sus políticas evaluativas y fortalecer las competencias docentes vinculadas con evaluación, inteligencia artificial y alfabetización digital.

Esta situación no significa que las tecnologías de inteligencia artificial deban ser concebidas exclusivamente como una amenaza para la evaluación. Por el contrario, pueden ampliar las oportunidades para generar retroalimentación, apoyar procesos de autoevaluación, diversificar actividades, ofrecer orientaciones personalizadas y favorecer nuevas formas de interacción con el conocimiento. No obstante, su incorporación también genera interrogantes relacionados con la autoría, la validez de las evidencias de aprendizaje, la fiabilidad de los resultados generados algorítmicamente, los sesgos, la privacidad y la integridad académica. UNESCO (2023) plantea que la incorporación de inteligencia artificial generativa en educación debe realizarse desde una perspectiva centrada en el ser humano, preservando la agencia de estudiantes y docentes, la inclusión, la equidad y el desarrollo de capacidades humanas. Desde esta perspectiva, el desafío educativo no consiste únicamente en incorporar nuevas tecnologías, sino en reconsiderar qué aprendizajes deben evaluarse y qué evidencias permiten demostrar realmente su adquisición.

En este escenario, la evaluación auténtica constituye una alternativa relevante frente a modelos evaluativos centrados predominantemente en la reproducción de contenidos. Este enfoque busca enfrentar al estudiante con situaciones, problemas, proyectos o desempeños que demanden la aplicación integrada de conocimientos, habilidades y actitudes en contextos cercanos a situaciones reales o profesionalmente significativas. La evaluación deja así de concebirse exclusivamente como un mecanismo destinado a medir resultados finales y adquiere una dimensión formativa, contextualizada y vinculada con el desempeño. La revisión sistemática desarrollada por Vlachopoulos y Makri (2024), basada en 94 estudios, evidencia que las prácticas de evaluación auténtica pueden contribuir al desarrollo de competencias asociadas con el pensamiento crítico, la resolución de problemas y la colaboración, aunque su implementación

también requiere formación docente, recursos institucionales y estrategias adaptadas a diferentes disciplinas.

La irrupción de la inteligencia artificial refuerza la pertinencia de esta transformación. Cuando un sistema puede generar una respuesta académica convencional en pocos segundos, la evaluación necesita desplazarse progresivamente desde la valoración exclusiva del producto hacia la consideración del proceso mediante el cual el estudiante construye, argumenta, aplica, contrasta y defiende sus conocimientos. Esto implica diseñar actividades en las que sea necesario justificar decisiones, resolver casos contextualizados, elaborar proyectos, explicar procedimientos, documentar procesos de aprendizaje, revisar críticamente información generada por inteligencia artificial o demostrar conocimientos mediante desempeños situados. En esta línea, Kizilcec et al. (2024) encontraron que docentes y estudiantes consideran que la inteligencia artificial generativa afecta especialmente tareas como la producción de ensayos y programación, mientras que el profesorado muestra una mayor preferencia por adaptar los sistemas de evaluación hacia actividades que incorporen pensamiento crítico, procesos de aprendizaje y aplicaciones auténticas.

Consecuentemente, la discusión contemporánea comienza a desplazarse desde estrategias destinadas únicamente a impedir el uso de inteligencia artificial hacia modelos de evaluación capaces de integrarla de manera pedagógicamente pertinente. Lye y Lim (2024) señalan que una respuesta basada exclusivamente en restringir o vigilar el uso de herramientas generativas resulta insuficiente ante la evolución constante de estas tecnologías. En su lugar, proponen reconsiderar el diseño de las evaluaciones a partir de sus propósitos educativos y de las posibilidades y limitaciones de la inteligencia artificial. Esta transformación exige determinar cuándo el uso de IA resulta incompatible con los resultados de aprendizaje, cuándo debe limitarse y cuándo puede incorporarse explícitamente como parte de una actividad que evalúe competencias de orden superior.

Desde esta perspectiva, la evaluación auténtica mediada por inteligencia artificial no implica delegar en los algoritmos el juicio evaluativo ni automatizar indiscriminadamente la valoración del desempeño estudiantil. Supone aprovechar determinadas capacidades tecnológicas sin desplazar la responsabilidad pedagógica del docente. La inteligencia artificial puede contribuir a generar retroalimentación preliminar, formular casos, apoyar procesos de revisión o facilitar experiencias adaptativas, pero la interpretación de las evidencias, la contextualización de los desempeños y la valoración integral de las competencias continúan requiriendo criterio profesional, conocimiento disciplinar y comprensión del contexto educativo. En consecuencia, resulta indispensable mantener una relación equilibrada entre innovación tecnológica y agencia humana, evitando que

la eficiencia de los sistemas automatizados se convierta en el principal criterio para decidir cómo y qué evaluar.

A pesar del crecimiento de investigaciones relacionadas con inteligencia artificial y educación, persiste la necesidad de profundizar en estrategias evaluativas que permitan distinguir entre la producción generada con asistencia tecnológica y el aprendizaje efectivamente construido por los estudiantes. La cuestión central ya no se limita a determinar si el alumnado utiliza inteligencia artificial, sino a establecer de qué manera pueden diseñarse experiencias evaluativas capaces de evidenciar comprensión, razonamiento, autonomía, capacidad crítica y aplicación contextualizada del conocimiento incluso en escenarios donde estas herramientas están disponibles. Esta problemática demanda revisar la función de la evaluación y avanzar hacia modelos que consideren tanto los resultados obtenidos como los procesos cognitivos, metacognitivos y reflexivos involucrados en su construcción.

En atención a esta problemática, el presente artículo tiene como objetivo analizar la transición de la evaluación tradicional hacia la evaluación auténtica en contextos educativos mediados por inteligencia artificial, identificando estrategias que permitan valorar el aprendizaje de manera integral, contextualizada y significativa. Asimismo, se busca examinar las posibilidades pedagógicas de la inteligencia artificial dentro de los procesos evaluativos, los principales desafíos asociados con su incorporación y la necesidad de preservar el juicio docente, la transparencia, la autoría académica y la integridad del aprendizaje. De esta manera, se pretende contribuir a la discusión sobre la construcción de modelos evaluativos acordes con las transformaciones contemporáneas de la educación, en los que la tecnología sea utilizada como recurso de apoyo y no como sustituto de los procesos humanos de enseñanza, aprendizaje y evaluación.

Metodología

La investigación se desarrolló bajo un enfoque cualitativo, con diseño documental y alcance descriptivo-analítico, mediante una revisión integrativa de la literatura orientada a examinar la transición de la evaluación tradicional hacia la evaluación auténtica en contextos educativos mediados por inteligencia artificial. La búsqueda bibliográfica se realizó en bases de datos y buscadores académicos como Scopus, Web of Science, ERIC, ScienceDirect, SpringerLink, SciELO y Google Scholar, utilizando descriptores en español e inglés relacionados con evaluación auténtica, evaluación educativa, inteligencia artificial generativa, evaluación formativa y competencias. Se consideraron principalmente publicaciones comprendidas entre 2020 y 2026, priorizando artículos científicos, revisiones y documentos institucionales con pertinencia directa para el objeto de estudio. Como criterios de inclusión se establecieron la relación temática con los

procesos de evaluación y aprendizaje mediados por IA, disponibilidad del texto completo y respaldo académico; se excluyeron documentos duplicados, estudios exclusivamente técnicos y publicaciones sin vínculo suficiente con el ámbito educativo.

El análisis de la información se realizó mediante una revisión inicial de títulos, resúmenes y palabras clave, seguida de la lectura completa de los documentos seleccionados. Posteriormente, se aplicó un análisis temático de contenido que permitió identificar coincidencias, diferencias y tendencias en torno a las limitaciones de la evaluación tradicional, los fundamentos de la evaluación auténtica, las estrategias para valorar competencias, el uso de inteligencia artificial en la retroalimentación y los desafíos relacionados con autoría, integridad académica, transparencia y juicio docente. La interpretación se realizó desde una perspectiva crítica y pedagógica, considerando tanto las oportunidades de la inteligencia artificial para diversificar las evidencias de aprendizaje como los riesgos derivados de su uso indiscriminado, con el propósito de establecer estrategias evaluativas que favorezcan una valoración integral, contextualizada y significativa del aprendizaje.

Resultados

El análisis temático de la literatura permitió identificar tendencias relacionadas con la transformación de los procesos de evaluación en escenarios educativos mediados por inteligencia artificial. Los hallazgos evidencian un desplazamiento progresivo desde modelos centrados en la reproducción de conocimientos y la valoración exclusiva del producto final hacia propuestas que consideran el desempeño, el proceso de aprendizaje, la reflexión, la aplicación contextualizada del conocimiento y el uso crítico de herramientas de inteligencia artificial.

Tabla 1. Principales hallazgos sobre la transformación de la evaluación educativa

Categoría de análisis	Hallazgo principal	Implicación para la evaluación
Limitaciones de la evaluación tradicional	Las pruebas centradas en reproducción de contenidos y productos finales presentan mayores dificultades para evidenciar el aprendizaje real cuando los estudiantes pueden utilizar herramientas de IA generativa.	Es necesario complementar o sustituir determinadas tareas tradicionales por actividades que permitan observar comprensión, razonamiento y aplicación del conocimiento.
Evaluación auténtica	Se fortalece la utilización de tareas vinculadas con situaciones reales, resolución de problemas, proyectos y desempeños contextualizados.	La evaluación debe valorar la capacidad del estudiante para movilizar conocimientos y competencias en situaciones significativas.
Diversificación de evidencias	El producto final deja de constituir la única evidencia relevante del aprendizaje.	Se requiere incorporar borradores, argumentaciones, exposiciones, procesos de revisión, portafolios y reflexiones metacognitivas.
Inteligencia artificial como apoyo	La IA puede utilizarse para generar retroalimentación, contrastar ideas, revisar producciones y apoyar procesos de mejora.	Su incorporación debe responder a objetivos pedagógicos definidos y no sustituir el proceso cognitivo del estudiante.

Autoría e integridad académica	La disponibilidad de IA genera dificultades para determinar qué partes de una producción corresponden al estudiante y cuáles fueron generadas tecnológicamente.	Se requieren mecanismos de transparencia, declaración del uso de IA y evidencias del proceso de elaboración.
Juicio profesional docente	La IA puede apoyar determinadas tareas evaluativas, pero presenta limitaciones para interpretar contextos, procesos y particularidades individuales.	La valoración final del aprendizaje debe mantener la intervención y responsabilidad profesional del docente.

Nota. Elaboración propia a partir de la revisión de la literatura.

Los hallazgos muestran que la evaluación auténtica adquiere especial relevancia en contextos mediados por inteligencia artificial debido a su capacidad para diversificar las evidencias mediante las cuales se demuestra el aprendizaje. En este sentido, las estrategias identificadas no se concentran únicamente en verificar un producto terminado, sino en hacer visibles los procesos intelectuales y las decisiones desarrolladas por el estudiante.

Tabla 2. Estrategias de evaluación auténtica identificadas en contextos mediados por IA

Estrategia	Evidencia de aprendizaje que permite valorar	Posible integración de IA
Resolución de problemas contextualizados	Aplicación de conocimientos, razonamiento, toma de decisiones y capacidad de justificar soluciones.	Generación o contraste de posibles soluciones que posteriormente deben ser evaluadas críticamente por el estudiante.
Aprendizaje basado en proyectos	Integración de conocimientos, planificación, creatividad, colaboración y resolución de problemas.	Apoyo en búsqueda inicial, organización de información, generación de alternativas o retroalimentación durante el proyecto.
Estudios de caso	Análisis, interpretación, argumentación y toma de decisiones frente a situaciones complejas.	Generación de escenarios alternativos o respuestas que el estudiante debe contrastar y justificar.
Portafolios digitales	Evolución del aprendizaje, producción progresiva, reflexión y capacidad de mejora.	Retroalimentación preliminar y apoyo durante procesos sucesivos de revisión.
Defensa oral	Comprensión, dominio conceptual, capacidad argumentativa y explicación de decisiones.	Permite comprobar y profundizar aprendizajes presentes en productos elaborados con apoyo tecnológico.
Autoevaluación	Capacidad metacognitiva, identificación de fortalezas y reconocimiento de aspectos por mejorar.	Comparación crítica entre la valoración personal y sugerencias generadas mediante IA.
Coevaluación	Capacidad de aplicar criterios, emitir juicios fundamentados y proporcionar retroalimentación.	Apoyo en la formulación inicial de criterios o sugerencias que deben ser revisadas por los estudiantes.
Análisis crítico de respuestas generadas por IA	Pensamiento crítico, verificación de información, argumentación y alfabetización en IA.	La propia respuesta generada por IA se convierte en objeto de evaluación, revisión y corrección.
Producción acompañada de evidencias del proceso	Autoría, razonamiento, revisión y evolución de la producción académica.	Registro y justificación de prompts, modificaciones, decisiones y aportes realizados por el estudiante.

Nota. Elaboración propia a partir de las estrategias recurrentes identificadas en la literatura analizada.

Los resultados también indican que la integración de inteligencia artificial en la evaluación no debe analizarse únicamente desde sus posibilidades tecnológicas. Su utilización introduce desafíos relacionados con la validez de las evidencias, la autoría, la dependencia tecnológica, la transparencia y el papel del docente. Por ello, la literatura revisada permite identificar condiciones necesarias para una implementación pedagógicamente responsable.

Tabla 3. Oportunidades, riesgos y criterios para el uso de inteligencia artificial en la evaluación

Dimensión	Oportunidad identificada	Riesgo asociado	Criterio pedagógico requerido
Retroalimentación	Facilita orientaciones inmediatas y procesos continuos de mejora.	Puede ofrecer información incorrecta, superficial o descontextualizada.	Supervisión y validación docente de la retroalimentación relevante.
Personalización	Posibilita adaptar determinadas orientaciones a necesidades particulares.	Puede generar dependencia tecnológica o reducir la autonomía del estudiante.	Utilizar la IA como apoyo y no como sustituto del esfuerzo cognitivo.
Producción académica	Puede apoyar la planificación, revisión y mejora de trabajos.	Dificultad para distinguir entre producción propia y contenido generado artificialmente.	Declaración transparente del uso de IA y documentación del proceso.
Pensamiento crítico	Las respuestas de IA pueden convertirse en objetos de análisis y contrastación.	Aceptación acrítica de información generada algorítmicamente.	Exigir verificación, argumentación, contraste de fuentes y justificación de decisiones.
Automatización de la evaluación	Puede apoyar tareas repetitivas y análisis preliminares.	Riesgo de sesgos, errores y pérdida de contextualización.	Mantener supervisión humana y criterios evaluativos explícitos.
Integridad académica	Permite replantear el concepto de autoría en ambientes digitales.	Uso no declarado o sustitución del desempeño del estudiante.	Establecer reglas claras sobre usos permitidos, restringidos y obligatoriamente declarados.
Juicio docente	La IA puede aportar información complementaria para la evaluación.	Delegación excesiva de decisiones pedagógicas a sistemas automatizados.	Preservar la responsabilidad profesional y la decisión final del docente.

Nota. Elaboración propia a partir del análisis temático de la literatura.

En síntesis, los resultados muestran que la transformación de la evaluación educativa en contextos mediados por inteligencia artificial no consiste simplemente en reemplazar instrumentos tradicionales por herramientas digitales. La tendencia identificada apunta hacia una evaluación más auténtica, procesual y contextualizada, en la que se diversifiquen las evidencias del aprendizaje y se valore no solo el producto obtenido, sino también el razonamiento, las decisiones, la capacidad crítica y el proceso seguido para construirlo. La inteligencia artificial puede contribuir a este propósito cuando su incorporación se encuentra pedagógicamente planificada, es transparente y permanece subordinada a los resultados de aprendizaje y al juicio profesional docente.

Discusión

Los resultados obtenidos evidencian que la expansión de la inteligencia artificial generativa está obligando a reconsiderar los fundamentos sobre los cuales tradicionalmente se ha construido la evaluación educativa. La posibilidad de que sistemas de IA produzcan ensayos, resuelvan problemas, elaboren explicaciones o generen respuestas estructuradas reduce la capacidad de ciertas actividades convencionales para demostrar, por sí solas, el aprendizaje alcanzado por el estudiante. Este hallazgo coincide con Xia et al. (2024), quienes, a partir de una revisión de 32 estudios empíricos, concluyen que los métodos tradicionales resultan progresivamente insuficientes para valorar con precisión los aprendizajes en contextos donde la inteligencia artificial generativa se encuentra ampliamente disponible. Los autores sostienen que la evaluación debe evolucionar hacia el desarrollo de aprendizaje autorregulado, responsabilidad, integridad y competencias relevantes para escenarios profesionales caracterizados por la coexistencia con tecnologías inteligentes.

En este sentido, los hallazgos no justifican la eliminación absoluta de las formas tradicionales de evaluación, sino su reconsideración dentro de sistemas más amplios y diversificados de obtención de evidencias. Determinados conocimientos fundamentales continúan siendo necesarios para el desarrollo de competencias superiores; sin embargo, una evaluación sustentada exclusivamente en preguntas reproductivas o productos finales presenta mayores dificultades para establecer qué procesos cognitivos fueron efectivamente realizados por el estudiante. Newton y Jones (2024) advierten precisamente que herramientas generativas pueden alcanzar elevados niveles de desempeño en actividades como ensayos y determinadas pruebas, lo que plantea problemas respecto de la validez de estas formas de evaluación cuando son utilizadas aisladamente. Esto refuerza la necesidad de combinar la comprobación de conocimientos con evidencias asociadas al razonamiento, la aplicación y la explicación del proceso seguido.

Desde esta perspectiva, la evaluación auténtica emerge como una alternativa especialmente pertinente. Los resultados de la revisión muestran que proyectos, resolución de problemas contextualizados, estudios de caso, portafolios, defensas orales y actividades basadas en desempeño permiten aproximarse de manera más directa a aquello que el estudiante comprende y puede aplicar. El valor de estas estrategias no reside únicamente en que sean más difíciles de resolver mediante IA, sino en que demandan movilizar conocimientos en situaciones significativas, justificar decisiones y hacer visible el proceso de construcción de una respuesta. De este modo, la autenticidad no debería entenderse simplemente como una estrategia destinada a impedir el uso tecnológico, sino como una cualidad del diseño evaluativo orientada a producir evidencias más

válidas y relevantes del aprendizaje.

Esta interpretación coincide con la propuesta de Lye y Lim (2024), quienes sostienen que responder a la inteligencia artificial mediante una competencia permanente entre tecnologías de detección y herramientas generativas resulta insuficiente. Su modelo de rediseño plantea que cada evaluación debería determinar explícitamente si la IA debe ser evitada, restringida o incorporada en función de los resultados de aprendizaje que se pretende valorar. Por tanto, la pregunta relevante deja de ser únicamente cómo impedir que el estudiante utilice inteligencia artificial y pasa a centrarse en qué tipo de desempeño debe demostrar y bajo qué condiciones el empleo de IA puede formar parte legítima del proceso de aprendizaje.

Los hallazgos relacionados con la utilización crítica de IA también permiten cuestionar una visión exclusivamente prohibitiva de estas tecnologías. Cuando el estudiante debe analizar una respuesta generada artificialmente, identificar errores, contrastar afirmaciones con fuentes confiables, reformular argumentos y justificar las modificaciones realizadas, la interacción con la tecnología puede convertirse en evidencia evaluable. Una investigación desarrollada en educación superior mostró que, cuando los estudiantes tuvieron libertad para utilizar diferentes herramientas de IA dentro de actividades evaluativas rediseñadas, desarrollaron criterios para seleccionar recursos tecnológicos según las necesidades específicas de cada tarea. Esto sugiere que una adecuada integración puede favorecer no solamente competencias disciplinares, sino también alfabetización en inteligencia artificial, discernimiento y toma de decisiones.

No obstante, estos resultados también ponen de manifiesto que integrar inteligencia artificial no garantiza automáticamente una evaluación de mayor calidad. La utilización de algoritmos para generar actividades, proporcionar retroalimentación o asignar valoraciones introduce interrogantes relacionados con validez, confiabilidad y equidad. Kaldaras, Akaeze y Reckase (2024) advierten que, cuando un sistema generativo participa en la puntuación o elaboración de evaluaciones, es indispensable determinar si está valorando los mismos conocimientos y habilidades que evaluaría un profesional capacitado. De manera semejante, Arslan et al. (2024) señalan que la personalización basada en IA puede ampliar oportunidades de participación y desempeño, pero también comprometer principios fundamentales de validez, fiabilidad y justicia si no existen mecanismos adecuados de supervisión.

La autoría académica constituye otro elemento relevante derivado de los resultados. La incorporación de IA modifica la manera tradicional de comprender la originalidad de una producción, dado que un trabajo puede integrar diferentes niveles de asistencia tecnológica sin que ello signifique necesariamente ausencia de aprendizaje. Luo (2024), al analizar las políticas de 20

universidades, identificó que una parte importante de las respuestas institucionales frente a la IA se concentra precisamente en el problema de la originalidad y en la dificultad para distinguir la contribución intelectual del estudiante de la asistencia externa proporcionada por sistemas generativos. Por esta razón, resulta más pertinente avanzar hacia mecanismos de transparencia que exijan declarar cómo fue utilizada la IA, qué decisiones permanecieron bajo responsabilidad del estudiante y de qué manera este verificó, modificó o rechazó las respuestas generadas.

Este problema adquiere mayor profundidad cuando la IA interviene en tareas reflexivas. Chan y Lee (2025), en un estudio realizado con estudiantes de educación secundaria, observaron que las herramientas generativas podían reducir determinadas cargas asociadas a la escritura reflexiva, pero también generaban dificultades para conservar profundidad, originalidad y autenticidad en las producciones. Este hallazgo resulta especialmente significativo porque demuestra que la autenticidad no depende únicamente de identificar si hubo o no utilización de IA, sino de analizar la calidad de la participación cognitiva que el estudiante mantiene durante el desarrollo de la actividad. En consecuencia, documentar borradores, justificar decisiones, explicar oralmente los productos o realizar reflexiones metacognitivas puede ofrecer evidencias complementarias de autoría y comprensión.

Otro aspecto central de los resultados es la permanencia del juicio profesional docente. Aunque la inteligencia artificial puede facilitar retroalimentación inmediata, organizar información o apoyar procesos preliminares de evaluación, no posee necesariamente una comprensión suficiente del contexto pedagógico, las trayectorias individuales o las circunstancias específicas en las que se construye un aprendizaje. Desde una perspectiva centrada en el ser humano, UNESCO (2023) establece que la incorporación de IA generativa en educación debe preservar la agencia humana, la inclusión, la equidad, la seguridad y la responsabilidad. Esto coincide con los resultados obtenidos, en cuanto la tecnología puede complementar determinados procesos evaluativos, pero no debería desplazar la responsabilidad profesional asociada con la interpretación y valoración integral del desempeño estudiantil.

En conjunto, la discusión permite sostener que el principal impacto de la inteligencia artificial sobre la evaluación educativa no consiste únicamente en facilitar la producción automatizada de respuestas, sino en cuestionar qué constituye una evidencia válida de aprendizaje. Frente a esta transformación, la evaluación auténtica ofrece un marco particularmente apropiado porque desplaza el énfasis desde la reproducción y el producto final hacia la aplicación, el proceso, la reflexión y la capacidad de justificar decisiones. En consecuencia, el reto contemporáneo consiste en diseñar evaluaciones en las que el estudiante pueda utilizar, cuando resulte

pedagógicamente pertinente, herramientas de inteligencia artificial sin que estas sustituyan el razonamiento que se pretende valorar. La calidad de la evaluación dependerá, por tanto, menos de la prohibición tecnológica y más de la capacidad docente para diseñar experiencias que hagan visible el pensamiento, la comprensión, la autonomía y la responsabilidad del estudiante.

Conclusiones

La transición de la evaluación tradicional hacia enfoques de evaluación auténtica constituye una respuesta pertinente frente a las transformaciones generadas por la inteligencia artificial en los procesos educativos. Las estrategias centradas exclusivamente en la reproducción de contenidos o en la valoración de productos finales presentan limitaciones para evidenciar con precisión el aprendizaje alcanzado por los estudiantes, especialmente cuando existen herramientas capaces de generar respuestas académicas de manera automatizada.

La evaluación auténtica permite ampliar y diversificar las evidencias de aprendizaje mediante proyectos, resolución de problemas, estudios de caso, portafolios, defensas orales, procesos de autoevaluación y actividades orientadas al desempeño. Estas estrategias favorecen la valoración de competencias como el pensamiento crítico, la argumentación, la toma de decisiones, la aplicación contextualizada del conocimiento y la capacidad de reflexionar sobre el propio proceso de aprendizaje.

La inteligencia artificial puede incorporarse de manera positiva dentro de los procesos evaluativos cuando su utilización responde a propósitos pedagógicos explícitos y se mantiene subordinada a los resultados de aprendizaje. Su potencial para apoyar la retroalimentación, la revisión de producciones y el análisis crítico debe complementarse con mecanismos de transparencia, declaración de uso, verificación de información y documentación de los procesos desarrollados por el estudiante.

Finalmente, la transformación de la evaluación en contextos mediados por inteligencia artificial requiere preservar el juicio profesional docente como elemento fundamental para interpretar las evidencias, valorar los desempeños y garantizar procesos evaluativos válidos, éticos y contextualizados. El principal desafío no consiste en impedir de manera absoluta el uso de inteligencia artificial, sino en diseñar experiencias de evaluación capaces de hacer visible la comprensión, el razonamiento, la autonomía y la responsabilidad del estudiante.

Referencias

- Arslan, B., Lehman, B., Tenison, C., Sparks, J. R., López, A. A., Gu, L., & Zapata-Rivera, D. (2024). Opportunities and challenges of using generative AI to personalize educational assessment. *Frontiers in Artificial Intelligence*, 7, 1460651. <https://doi.org/10.3389/frai.2024.1460651>
- Chan, C. K. Y., & Lee, K. K. W. (2025). The balancing act between AI and authenticity in assessment: A case study of secondary school students' use of GenAI in reflective writing. *Computers & Education*, 238, 105399. <https://doi.org/10.1016/j.compedu.2025.105399>
- Foung, D., Lin, L., & Chen, J. (2024). Reinventing assessments with ChatGPT and other online tools: Opportunities for GenAI-empowered assessment practices. *Computers and Education: Artificial Intelligence*, 6, 100250. <https://doi.org/10.1016/j.caeai.2024.100250>
- Kaldaras, L., Akaeze, H. O., & Reckase, M. D. (2024). Developing valid assessments in the era of generative artificial intelligence. *Frontiers in Education*, 9, 1399377. <https://doi.org/10.3389/feduc.2024.1399377>
- Kickbusch, S., Ashford-Rowe, K., Kemp, A., Boreland, J., & Huijser, H. (2025). Beyond detection: Redesigning authentic assessment in an AI-mediated world. *Education Sciences*, 15(11), 1537. <https://doi.org/10.3390/educsci15111537>
- Kizilcec, R. F., Huber, E., Papanastasiou, E. C., Cram, A., Makridis, C. A., Smolansky, A., Zeivots, S., & Radulescu, C. (2024). Perceived impact of generative AI on assessments: Comparing educator and student perspectives in Australia, Cyprus, and the United States. *Computers and Education: Artificial Intelligence*, 7, 100269. <https://doi.org/10.1016/j.caeai.2024.100269>
- Luo, J. (2024). A critical review of GenAI policies in higher education assessment: A call to reconsider the “originality” of students' work. *Assessment & Evaluation in Higher Education*, 49(5), 651–664. <https://doi.org/10.1080/02602938.2024.2309963>
- Lye, C. Y., & Lim, L. (2024). Generative artificial intelligence in tertiary education: Assessment redesign principles and considerations. *Education Sciences*, 14(6), 569. <https://doi.org/10.3390/educsci14060569>
- Miao, F., & Holmes, W. (2023). *Guidance for generative AI in education and research*. UNESCO. <https://doi.org/10.54675/EWZM9535>
- Newton, P. M., & Jones, S. (2025). Education and training assessment and artificial intelligence: A pragmatic guide for educators. *British Journal of Biomedical Science*, 81, 14049. <https://doi.org/10.3389/bjbs.2024.14049>
- Oliveira, M., Zednik, C., Bombaerts, G., Sadowski, B., & Conijn, R. (2025). Assessing students' DRIVE: A framework to evaluate learning through interactions with generative AI.

Computers and Education: Artificial Intelligence, 9, 100497.
<https://doi.org/10.1016/j.caeai.2025.100497>

Vlachopoulos, D., & Makri, A. (2024). A systematic literature review on authentic assessment in higher education: Best practices for the development of 21st century skills, and policy considerations. *Studies in Educational Evaluation*, 83, 101425.
<https://doi.org/10.1016/j.stueduc.2024.101425>

Xia, Q., Weng, X., Ouyang, F., Lin, T. J., & Chiu, T. K. F. (2024). A scoping review on how generative artificial intelligence transforms assessment in higher education. *International Journal of Educational Technology in Higher Education*, 21, 40. <https://doi.org/10.1186/s41239-024-00468-z>

